



Phylogenetic diversity metrics from molecular phylogenies: modelling expected degree of error under realistic rate variation

Andrew M. Ritchie¹  | Xia Hua^{1,2} | Marcel Cardillo¹ | Keaghan J. Yaxley¹ |
Russell Dinnage¹ | Lindell Bromham¹

¹Research School of Biological Sciences, Australian National University, Canberra, ACT, Australia

²Mathematical Sciences Institute, Australian National University, Canberra, ACT, Australia

Correspondence
Andrew M. Ritchie, Research School of Biological Sciences, Robertson Building, 134 Linnaeus Way, Australian National University, Canberra, ACT 2600, Australia.
Email: Andrew.Ritchie@anu.edu.au

Funding information
This work was supported by the Australian Research Council grant DP190101916.

Editor: Chris Burridge

Abstract

Aim: Phylogenetic diversity or phylo-diversity measures use information about evolutionary history and relationships to inform conservation priorities. These metrics are usually derived from the branches of molecular phylogenies. But inferring phylogenetic timescale from molecular data relies on many assumptions about the evolutionary process, most of which are based on statistical convenience rather than biological information. Here we ask whether known patterns of variation in rate of molecular evolution can lead to errors in phylo-diversity measures.

Location: Global.

Methods: We generated sequences with biologically realistic rate variation, parameterized by empirically well-supported relationships between species traits, macroevolutionary patterns and rate of molecular evolution. We then tested how well commonly used phylo-diversity measures from these phylogenies reflected the true evolutionary history.

Results: We show that reconstructed Faith's phylogenetic diversity (FPD) measures differ from true values by an average of about 10% and up to 38%. Species rankings based on evolutionary distinctness (ED) are also affected by rate variation, with the ranks of taxa changing by up to 10–11 positions after estimation.

Main conclusions: We have shown that realistic levels of rate variation can generate error in FPD and ED measures that could potentially influence prioritization ranking. Studies using metrics based on molecular phylogenetic branch lengths should consider the likely effect of uncertainty in phylogenetic reconstruction on their conclusions.

KEY WORDS

Phylogenetic diversity, evolutionary distinctiveness, relaxed clocks, conservation, EDGE of existence, species assemblage

1 | Introduction¹

Conservation planning may involve quantifying the contribution to conservation goals made by different areas, populations, species or other biological groups, and comparing the effectiveness of different conservation strategies in achieving those goals (Brooks et al., 2006). Conservation goals may be aimed at maximizing representativeness and persistence of biodiversity across a planning region (Margules & Pressey, 2000). But “biodiversity” is a rich and complex concept that is challenging to express in a way that can be easily measured and compared across ecosystems, areas or time periods (Lean, 2016; Maclaurin & Sterelny, 2008).

Phylogenetic measures of diversity, or phylo-diversity, have been developed as a way of deriving an objective, quantitative metric that reflects different intrinsic values of species or areas for the purposes of conservation prioritization (Crozier, 1997; Tucker et al., 2017). These measures have in common that they are derived from some function of the branch lengths of a phylogeny, as distinct from taxonomic measures that depend on the number of nodes or species in the tree (Tucker et al., 2017). Metrics may use just a single edge from a phylogenetic tree—such as the tip length connecting a species to its nearest relative—or a path between tips in the phylogeny, or the sum of edges for a clade or sample of tips (Kondratyeva et al., 2019).

Faith's phylogenetic diversity (FPD; Daniel P Faith, 1992) is one of the most widely used phylogenetic measures of biodiversity. It is calculated as the sum of the branch lengths of the minimum spanning tree connecting the members of a community or assemblage. This measure was proposed as a means of representing the diversity of features among a group of species. A feature is any aspect of a species which may in turn reflect its ecological role, distinct characteristics or traits of value to human society (Forest et al., 2015; Maclaurin & Sterelny, 2008). Feature diversity is difficult to measure directly, hard to compare among taxonomic groups and requires decisions about which features should be selected as salient for conservation worth (Kelly et al., 2014). Moreover, there is a desire not only to conserve known feature diversity but to maximize the probability of retaining features that could be recognized as important in the future, referred to as “option value” (Faith, 2007), though this usage differs somewhat from the economic origins of the term (Maier, 2012). Note that the unqualified term “phylogenetic diversity” may refer to either FPD in particular or the more general concept. To avoid confusion, we use the term “phylo-diversity” for the latter throughout.

Branch lengths have been considered an indicator of feature diversity on the assumption that a longer branch represents more opportunity for evolutionary change and the development of unique characteristics (Faith, 2018). This view has attracted some controversy. Empirical evidence for the relationship between

phenotypic or ecological diversity and FPD remains conflicted (Cadotte et al., 2017; Devictor et al., 2010; Flynn et al., 2011; Fritz & Purvis, 2010; Tucker et al., 2019), and some empirical studies have shown that species sets chosen to maximize FPD may not reliably conserve more feature diversity than random species sets in practice (Kelly et al., 2014; Mazel et al., 2018). Studies based on theoretical models of trait evolution are also divided on this issue (Letten et al., 2014; Mazel et al., 2017; Tucker et al., 2018). Partly in response to these controversies, authors have proposed alternative interpretations of FPD reflecting other qualities relevant to conservation. For example, some scholars have argued that “evolutionary heritage” is worth protecting independently of any link with feature diversity (Mooers, Heard, & Chrostowski, 2005; Rosauer & Mooers, 2013; Winter et al., 2013). Alternatively, FPD is sometimes interpreted as an indicator of future evolutionary potential (Muenchow et al., 2018).

FPD is typically used to compare the relative value of biological assemblages or communities, by summing over all branch lengths connecting the species. An extension of the notion of evolutionary history as a target for conservation is that phylogenetic branch lengths can be used to quantify the unique phylo-diversity contribution of particular species. Species with few close relatives, and therefore long branch lengths, are frequently identified as conservation priorities on the grounds that they represent a unique suite of traits (Isaac et al., 2007; Safi et al., 2013). For example, during the recent Australian bushfire crisis, extraordinary efforts were expended to save the Wollemi pine, a critically endangered species separated from its nearest living relative by over 150 million years of evolution (Laity et al., 2015). Phylogenetic measures of evolutionary distinctiveness (ED) have also been combined with indicators of global endangerment (GE) to produce the EDGE score, which is used to identify priority species for conservation in mammals (Isaac et al., 2007), amphibians (Isaac et al., 2012), corals (Curnick et al., 2015; Huang et al., 2016), birds (Jetz et al., 2014), sharks, rays and chimaeras (Stein et al., 2018), and reptiles (Gumbs et al., 2018). These efforts have increased public awareness of many lesser-known threatened species through top 100 priority lists maintained by the Zoological Society of London's (ZSL) “EDGE of Existence” programme (ZSL, 2014).

FPD and associated measures such as ED have become increasingly popular in conservation biology and are being promoted as potential tools for conservation decision-making (Brooks et al., 2015; Hendry et al., 2010; Isaac & Pearse, 2018). Despite the aforementioned debates regarding the relationship of FPD with various salient qualities for conservation, the degree to which we can rely upon molecular branch lengths to represent any of these qualities has received less attention. While few studies have addressed this question, some have found that FPD estimation can be affected by factors such as whether branch lengths are resolved in units of time or genetic difference (Elliott et al., 2018) and whether the tree is subsampled from a larger phylogeny (Park et al., 2018). These findings suggest that the nature of phylogenetic inference can impact metrics used in conservation prioritization.

¹Abbreviations: FPD: Faith's Phylogenetic Diversity; ED: Evolutionary Distinctiveness; EDGE: Evolutionarily Distinct and Globally Endangered; UCLN: Uncorrelated Lognormal relaxed clock model; NPRS: Non-Parametric Rate Smoothing.

Time-resolved branch lengths from molecular data are historical inferences based on a complex set of assumptions, and their interpretation depends on the interplay of time, rates and genetic differences (Bromham, 2019). Even where the rate of molecular evolution is assumed to be the same in all lineages, uncertainty in branch lengths arises from the need to infer past changes that cannot be directly observed and to infer the age of lineages using fossils whose age and position in the phylogeny cannot be known with absolute precision (Yang & Rannala, 2005). This uncertainty leads to a background level of error in estimating evolutionary quantities even where rates are constant (Wertheim & Sanderson, 2011). But the situation is much more complicated when rates of molecular evolution vary among lineages. Given the growing evidence that evolutionary rates are influenced by a wide range of species traits, as well as by environmental factors and macroevolutionary processes, we should expect that rates will vary across the phylogeny (Bromham, 2009; Bromham et al., 2015; Lanfear et al., 2010).

There are a large number of phylogenetic methods that allow rates of molecular evolution to vary over the phylogeny, such as "relaxed clock" models that allow limited random variation in rates among lineages (dos Reis et al., 2016; Ho, 2014; Kumar, 2005; Lepage et al., 2007). These methods make many assumptions that allow for statistical tractability. For example, most methods rely on modelling substitution rates based on simple parametric distributions, either independently for each lineage (Drummond et al., 2006) or determining the rate for each new lineage from the rate of its parent (Ho et al., 2005; Thorne et al., 1998). Changing these assumptions can affect the branch lengths produced by the method (Battistuzzi et al., 2010), and the effect is larger when there is more overall variation (Sarver et al., 2019). Error in inferred branch lengths can lead to erroneous conclusions about macroevolutionary processes, such as the speciation rate, where rates are associated with an evolving continuous trait (Duchêne et al., 2017) or discrete character (Shafir et al., 2020). Therefore, we might reasonably expect that choices made in phylogenetic analyses could influence measurements of phylo-diversity.

Does the inherent uncertainty of phylogenetic branch length estimation produce uncertainty in phylo-diversity measurements? We examine the performance of relaxed clock methods in estimating Faith's PD (FPD) and Evolutionary Distinctiveness (ED) from nucleotide data simulated under biologically realistic models of rate variation. We ensure that our results are relevant to current practice in the field using four strategies. Firstly, we conduct a systematic study of the literature to inform both our simulation design and our analysis, selecting the most common phylogenetic methods and phylo-diversity metrics. Secondly, since correlates of variation in substitution rates have been well-studied in birds (Berv & Field, 2018; Eo & DeWoody, 2010; Lanfear et al., 2010; Nabholz et al., 2016), we use published studies of birds to set realistic parameters for rate variation. Thirdly, we simulate biologically realistic patterns of rate variation driven based on observed patterns of association between substitution rates and species traits and diversification rates. Fourthly, we want our phylogenies to represent the kind of

trees typically analysed in phylo-diversity studies, so we generate sample phylogenies by drawing assemblages of taxa from known locations, then deriving the tree connecting those taxa from a global phylogeny. In this way, our simulations are conditioned on realistically sized data sets, realistic tree shapes and realistic patterns of rate variation.

We investigate two biologically informed models of rate variation. The first is a species-based trait model. There is growing evidence that particular species traits are associated with rate variation, for example with generation time, longevity, body size and fecundity (Bromham, 2011; Hua et al., 2015; Qiu et al., 2014; Thomas et al., 2010; Welch et al., 2008; Wong, 2014). Therefore, rates of molecular evolution may evolve along phylogenies as species traits. The second model is based on the observed association between net diversification rate and rate of molecular evolution, a relationship that has been found for a wide range of taxa, including birds, reptiles and plants (Barraclough & Savolainen, 2001; Bromham et al., 2015; Eo & DeWoody, 2010; Lanfear et al., 2010; Webster et al., 2003). While the causation of this relationship remains unknown, proposed mechanisms include elevated substitution rates speeding the development of reproductive isolation in populations (Hua & Bromham, 2017) or founder effects generating elevated rates following speciation (Pagel et al., 2006).

Once we have evolved DNA sequences under both trait-based and speciation-associated rate variation, we then apply the most commonly used phylogenetic methods to reconstruct the phylogenetic history from the sequences and apply a number of phylo-diversity metrics to the reconstructed phylogeny. This allows us to describe the likely scale of errors when estimating FPD and ED for real data and considers the implications for potential applications of these measures to conservation planning and prioritization.

2 | Methods

2.1 | Survey of common practice

We used a literature survey to gauge common practice in recent studies that use measures of phylo-diversity to provide information relevant to conservation (Table 1; Figure 1). This was not an exhaustive review of all published papers, but a representative sample of 131 recent studies (2015–2018), in order to gather information on metric use, size of phylogeny, amount of data, phylogenetic methods used and type of phylo-diversity measures (see Table S4 available as Supplementary Information for the full list of studies).

Many of the studies use previously published supertrees to derive a phylogeny for the taxa in the study; for example, a mammal supertree (Bininda-Emonds et al., 2007), a global bird phylogeny (Jetz et al., 2012) and various angiosperm phylogenies included in the package Phylocom (Webb et al., 2008). Some studies use the TimeTree of Life database, which is based on summaries of published dates (Hedges and Kumar, 2009). Others use molecular phylogenies freshly inferred from sequence data. The two most common approaches to phylogeny

TABLE 1 Uses of phylo-diversity metrics from a survey of 150 conservation biology studies published between 2015 and 2018: see Table S4 available as Supplementary Information for details

Phylo-diversity metric use	Number of Studies
Quantifying environmental impacts (e.g. agriculture, forest clearance)	41
Identifying geographic patterns and hotspots	39
Optimizing reserve designs/Prioritizing potential conservation sites	25
Adequacy of existing reserves	20
Testing against other metrics	17
Identifying priority species	9
Capacity to capture salient feature (e.g. ecosystem function, niche space)	5
Identifying drivers of biodiversity	5
Quantifying loss due to extinctions	4
Association with threat status	2
Predicting environmental impacts (e.g. climate change)	2
FPD-area relationship as a measure of ecosystem response	2
Evaluating collection strategies	2
Comparing different data sources	2
Other	6

reconstruction across both pre-published trees and newly inferred phylogenies were the Bayesian inference package BEAST (Bouckaert et al., 2014) with branch lengths estimated using an uncorrelated log-normal clock (UCLN), and the maximum likelihood package RAxML (Stamatakis, 2014) with branch lengths estimated using non-parametric rate smoothing (NPRS), for example using r8s (Sanderson, 2003) or treePL (Smith & O'Meara, 2012).

Across all 131 studies, we found over 50 different ways of measuring some form of phylo-diversity (Figure 1b). The four most commonly used metrics were Faith's PD (FPD), mean pairwise distance (MPD), Evolutionary Distinctness (ED, including Evolutionary Distinctiveness and Global Endangerment, EDGE) and Net Relatedness Index (NRI). Because MPD and NRI are similar to PD in using aggregates of branch lengths, we chose to focus on FPD and ED in our simulation study as two distinct ways of using phylogenetic branch lengths in conservation prioritization.

We also examined the intended use of the phylo-diversity metric (Table 1). The modal use was to quantify biodiversity changes due to environmental impacts, such as agriculture and habitat loss. However, the next three most frequent uses (identifying hotspots, optimizing reserve designs and assessing adequacy of existing reserves) all related to ranking or comparing phylo-diversity values among either discrete sites or among grid cells on a map, with the ultimate aim of informing prioritization of conservation resources. Meanwhile, studies using ED were largely focused on identifying priority species for conservation. Our evaluation of phylo-diversity metrics reflects these primary proposed uses of the relevant metrics.

2.2 | Design of simulation studies

We investigated the effects of realistic rate variation on the estimation of phylo-diversity metrics by simulating DNA sequences under

empirically informed conditions. We conditioned our simulations on studies in which a phylogeny is constructed directly, rather than drawn from a pre-existing tree. Of the studies that used BEAST, the median number of species was 147, so we set our simulation tree size to 150 tips. Of the studies that provided information on alignment lengths, the median value was 4733 nucleotide sites. As a practical limitation, we chose to simulate sequence lengths of 2000 variable sites. This number can be thought of as equivalent to a protein-coding alignment of about 6000 bases where the majority of signal comes from the third codon position.

Randomly generated trees for 150 taxa might not have realistic distributions of branch lengths, particularly for FPD measures that are applied to geographic assemblages which are not all closely related members of a single clade. Therefore, to construct our simulations we first generated 100 "community assemblages" by randomly selecting locations from the world map and sampling 150 species from that area. We used bird species to generate our sample datasets as a data-rich case study. We then extracted a phylogeny for each assemblage by taking the subtree that incorporated all sampled taxa from a published time-scaled supertree and simulated average substitution rates for each branch of each of the resulting 100 trees under each of two biologically informed patterns of rate variation (Figure 2). Finally, we simulated DNA sequences along these phylogenies under these rate variation patterns. A full description of our assemblage sampling and sequence simulation procedures is available as Supplementary Information (Extended Methods).

2.3 | Biologically informed patterns of rate variation

We have two ways of modelling realistic patterns of rate variation. One models the influence of species traits on rates of molecular evolution. Another models the association between rate of molecular

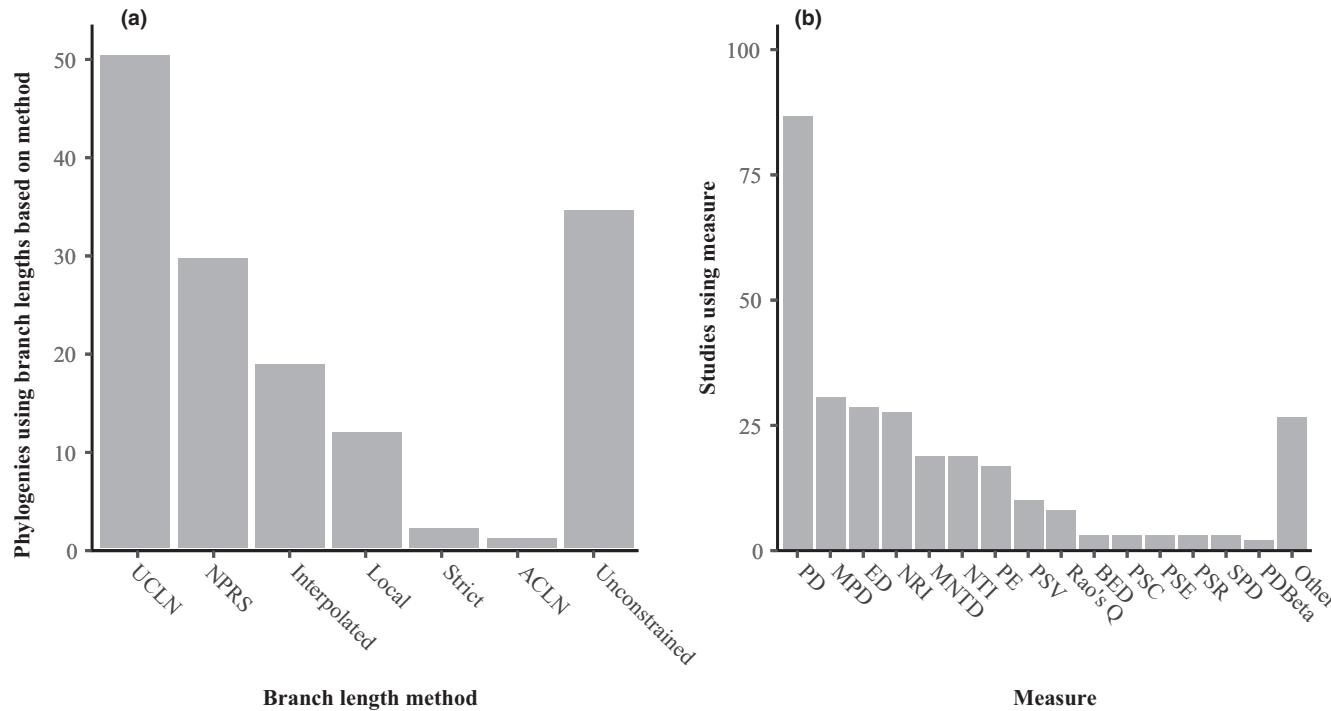


FIGURE 1 Summary of findings from a survey of 150 studies published between 2015 and 2018 that apply phylo-diversity measures to conservation biology. (a) Methods used to reconstruct branch lengths for use in phylo-diversity estimation. Bar heights show the number of times a phylogeny using a particular method is used among all of the surveyed studies. We do not distinguish whether the study itself used this method or whether it made use of pre-published branch lengths estimated by this method. UCLN: Uncorrelated lognormal clock. ACLN: Autocorrelated lognormal clock. NPRS: Non-parametric rate smoothing. The “interpolated” category refers to methods that distribute undated nodes uniformly between dated nodes. The “unconstrained” category refers to methods that estimate branch lengths in units of genetic divergence rather than separating rates and dates. (b) Phylo-diversity measures used and number of studies using each measure. The full list of studies is available as Supplementary Information (Table S4). Abbreviations are those used in Table 1. Note that many studies used multiple measures. Both regular and standardized versions of FPD are listed under FPD, while “sesMNTD” is grouped together with NTI and “sesMPD” is grouped with NRI

evolution and net speciation rates (Hua & Bromham, 2017; Webster et al., 2003). All simulations were undertaken in the R programming environment (R Core Team, 2019). Details of the simulation methods are available as electronic Supplementary Material.

In the first scenario (trait-based variation), we start with our phylogeny of 150 taxa. We evolve a continuous trait value as a Brownian motion along the tree and then use this to assign substitution rates to branches. Substitution rates are generated from trait values via a linear relationship between rates and body mass, which is derived from a published regression analysis of bird body sizes and third codon substitution rates (Nabholz et al., 2016).

In the second scenario (speciation-based variation), we model evolutionary rate variation that is correlated with variation in the rate of lineage diversification. We first use a combination of methods to infer speciation rates for every lineage in the bird phylogeny from which our sampled assemblages are drawn (see Supplementary Methods). Each branch in the sampled assemblage phylogenies is given the speciation rate of the branch in the full bird phylogeny that shares the same ancestral node. As in the trait-based scenario, we then simulate an average substitution rate for each branch under the assumption that the substitution rate is linearly related to the speciation rate.

2.4 | Reconstructing phylogenies from simulated alignments

The two simulation methods (trait-based rates and speciation-based rates) each produce 100 alignments of 150 sequences along a known phylogeny. How accurately can we infer the true history by reconstructing the phylogeny using only the sequences from the tips? We selected the most commonly used phylogenetic methods from our literature survey (Figure 1): a Bayesian relaxed clock method (BEAST 2; Bouckaert et al., 2019) and a penalized likelihood “rate smoothing” procedure (treePL; Smith & O’Meara, 2012). The topology for the latter was produced by maximum likelihood in RAxML (Stamatakis, 2014). Details of the analysis are available as Supplementary Information.

2.5 | Phylo-diversity metrics

For each of 200 alignments (100 under each of two simulation methods), these procedures give us four reconstructed phylogenies, using two different methods (Bayesian/BEAST and maximum likelihood/NPRS) for each of two calibration schemes (Figure 2). We

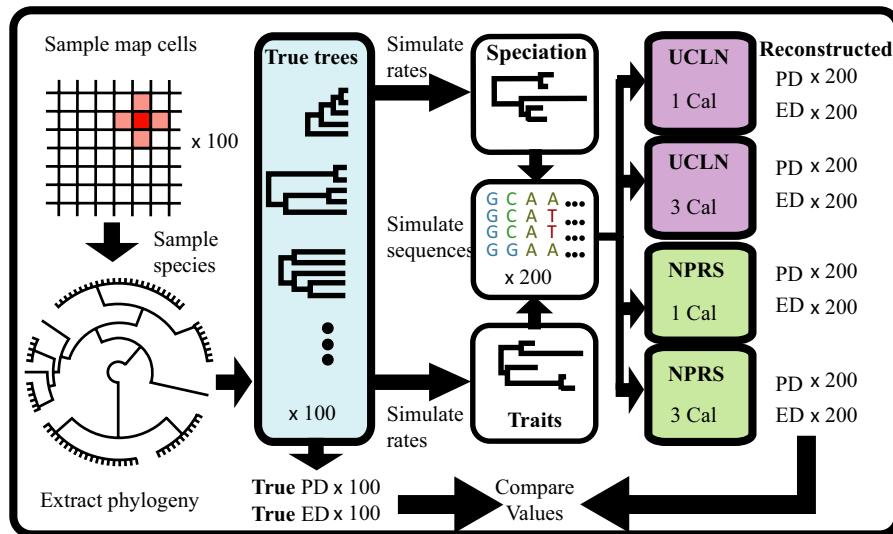


FIGURE 2 Schematic for the simulation study design. 100 assemblages are sampled from randomly chosen grid cells on a global map with occurrence data. The phylogeny for each assemblage is then extracted from a supertree, generating 100 trees. Patterns of evolutionary rate variation are simulated for each tree under two scenarios: one in which molecular rates are correlated with speciation rates, and one in which molecular rates are correlated with a life-history trait. Each of these two patterns is applied to each tree, and sequences are then simulated along the phylogeny under each rate scenario, producing 200 sequence data sets. Each data set is then used to reconstruct a phylogeny, including tree topology and branch lengths, using either Bayesian inference with an uncorrelated lognormal relaxed clock (UCLN) or maximum likelihood reconstruction using non-parametric rate smoothing (NPRS). Faith's phylogenetic diversity (FPD) and taxon-specific evolutionary distinctiveness (ED) values are calculated for each of the resulting 800 reconstructed trees. Finally, these “reconstructed” FPD and ED values are compared with FPD and ED values calculated from the original 100 “true” trees

TABLE 2 Correlation results for Faith's Phylogenetic Diversity estimation on alignments simulated under a speciation- or trait-based rate variation pattern. The correlation is between true and reconstructed FPD values. The error is the absolute difference between true and reconstructed values as a percentage of the true value

Rate variation	Method	Calibration	Correlation (Spearman)	Mean Absolute Percentage Error (% of true PD)	Max % Error
Speciation	Bayesian	1	0.91	6.43	23.55
		3	0.86	7.51	32.80
	NPRS	1	0.58	13.97	38.31
		3	0.74	9.14	27.47
Trait	Bayesian	1	0.73	10.50	31.43
		3	0.84	7.92	31.30
	NPRS	1	0.50	10.38	37.07
		3	0.75	7.60	31.50

now have 100 phylogenies that represent the history of our samples (along which sequences were simulated), which we will refer to as the “true tree,” and 100 phylogenies reconstructed from the simulated sequences for each combination of simulation scenario, reconstruction method and calibration condition, which we refer to as “reconstructed trees.” Now we apply two different phylo-diversity measures to each of these reconstructed phylogenies and compare the values to those calculated on the true tree used to simulate each data set.

We calculate Faith's PD score (FPD; Daniel P Faith, 1992) for each reconstructed tree and each corresponding true tree. This score is the sum of the branch lengths of the entire tree (the tree

length). Evolutionary distinctiveness (ED) values were calculated for each terminal branch in each true and reconstructed phylogeny. The ED of a tree tip is a weighted sum of all branch lengths connecting that tip to the root, where the weight of each branch is given by the reciprocal of the number of tips descending from that branch. FPD and ED calculations used the “picante” package for R (Kembel et al., 2010). We used the “pd” function for FPD and the “evol.distinct” function for ED using the “fair proportions” metric, which distributes the length of each branch evenly among its descendants. We used these values to compare phylo-diversity measures made on reconstructed trees to the values on the corresponding true trees and determine the degree of associated error.

3 | Results

3.1 | Estimation accuracy

FPD estimates were correlated with the true values with Pearson coefficients of 0.5 or higher across all simulation scenarios, calibration availabilities and reconstruction methods, but the mean error in estimated FPD was in all cases between 6% and 14% (Table 2). Correlation coefficients were higher for Bayesian reconstructed trees than maximum likelihood/NPRS trees. The slope of the linear regression of reconstructed FPD values on true FPD values was greater than one for all simulation scenarios under Bayesian analysis, overestimating larger FPD values and underestimating smaller values (Figure 3), while NPRS slopes were less than one (Figure S2 available as Supplementary Information). No other systematic differences were observed between results from Bayesian and ML approaches and only Bayesian results are shown henceforth (NPRS results available as Supplementary Information).

For ED, the slope of the linear mixed-effects models varied between 0.95 and 1, showing a tendency to overestimate small ED values or underestimate larger ED values (Table 3). Pseudo- R^2 values for reconstructed ED ranged from 0.90 to 0.93. The Bayesian approach produced slightly higher R^2 values for the speciation than the trait simulation scenario. The number of calibrations did not have a consistent effect across simulation scenarios.

3.2 | Rank order stability

Ranking errors are also observable in a comparison of rankings generated from true and reconstructed FPD values among our 100 true trees (Figure 4; Figure S7 available as Supplementary Information). Because FPD is often used to rank areas or assemblages by their relative diversity, here we are ranking all 100 sampled assemblages by their FPD values, hypothetically in order to identify the top-ten assemblages that have the highest phylo-diversity values. Across all of our simulation scenarios and reconstruction methods, the ranking of the assemblages based on reconstructed FPD differed from that using the true FPD of the assemblage, leading to some assemblages not being included in an arbitrary set of top-ranked assemblages when they should have on the basis of the true FPD. For the speciation simulations, four assemblages were erroneously excluded from the top 10 when they would have been included based on true FPD. Five assemblages were erroneously excluded under the trait-based simulation scenario with three calibrations, and three assemblages were excluded when there was one calibration.

We also ranked individual taxa within each assemblage on the basis of reconstructed ED values. Here we use taxa excluded from the top 100 as our measure, by analogy with the top 100 EDGE lists. The mean number of taxa mistakenly excluded from the top 100 when ranked based on reconstructed ED was between 10 and 11 for each combination of simulation scenario, reconstruction method and calibration condition (BEAST, Figure 5; NPRS, Figure S3

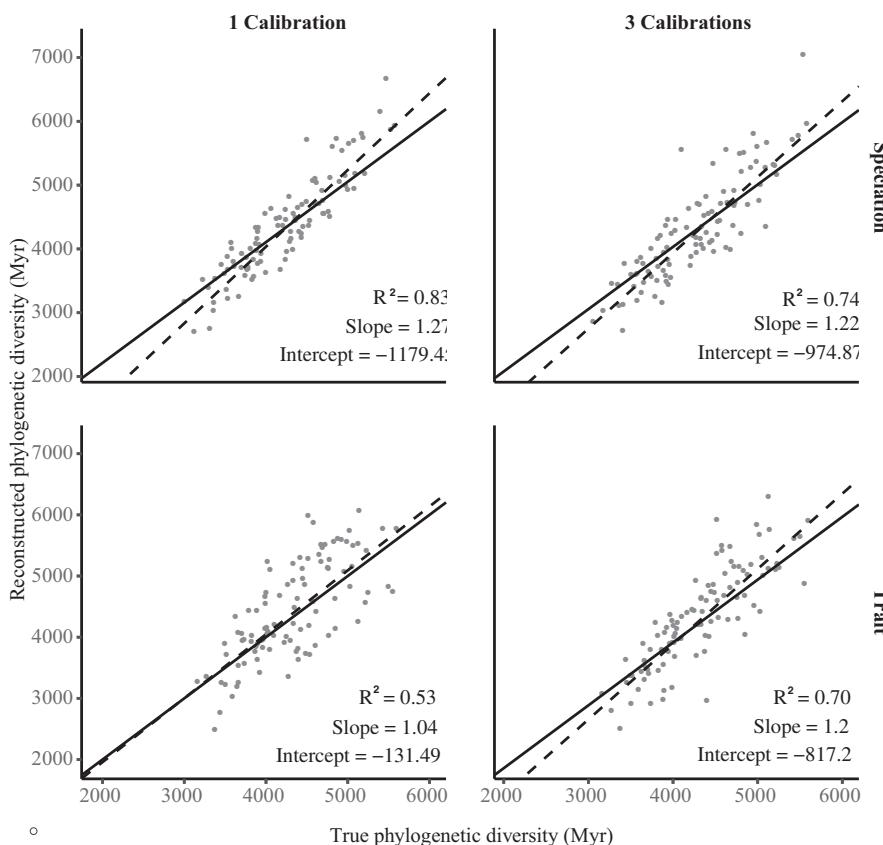
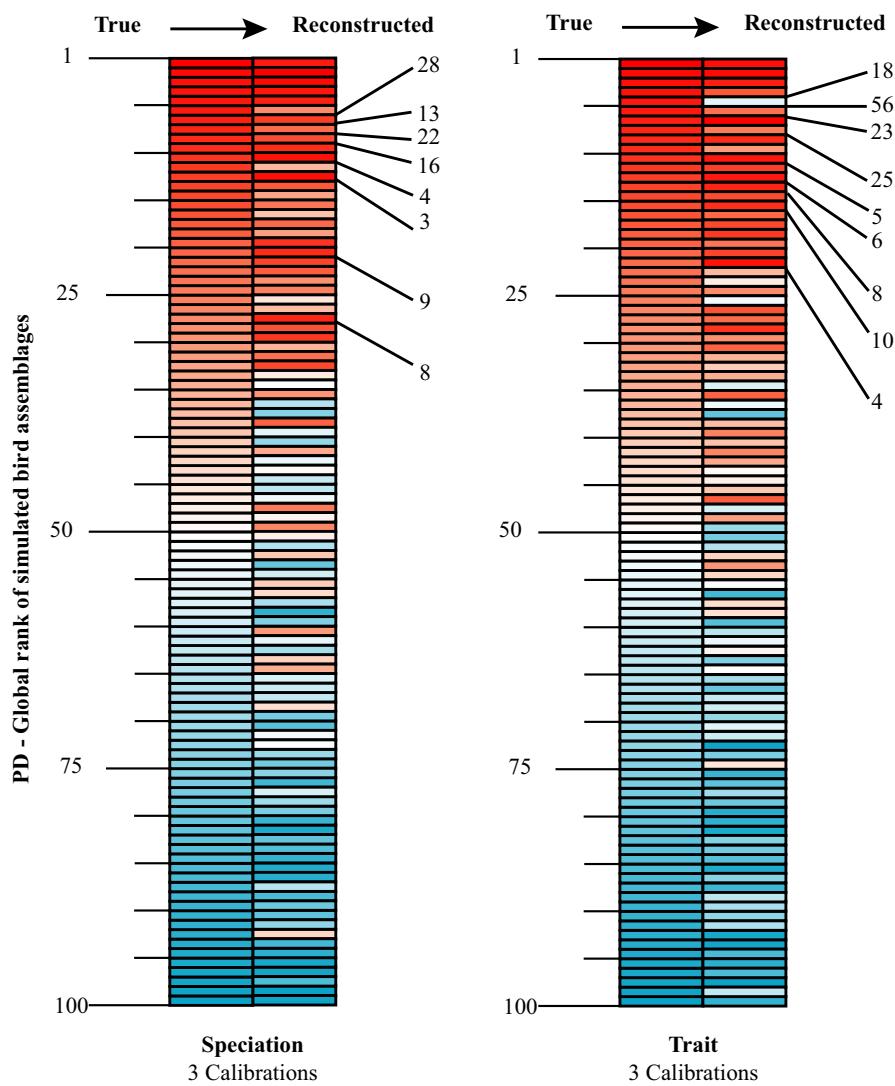


FIGURE 3 Linear relationships between simulated and reconstructed Faith's Phylogenetic Diversity, simulated under speciation-based and continuous trait-based patterns of rate variation and reconstructed via Bayesian inference in BEAST 2. Each point represents a single alignment, with the true FPD value based on the known (simulated) history and the reconstructed FPD from the phylogeny estimated from the sequences using Bayesian UCLN and either 1 or 3 calibrations. Solid lines indicate the line of perfect inference. Dashed lines indicate the slopes of the fitted models

TABLE 3 Slope and fit of a linear mixed-effects model relating simulated to estimated ED values across all trees in each studied scenario, accounting for variation among the means of individual trees as a random effect

Rate variation	Method	Calibration	Slope	SD of Slope	Residual SD	Pseudo-R ²
Speciation	Bayesian	1	0.96	0.0023	4.43	0.93
		3	0.95	0.0022	4.37	0.93
	NPRS	1	1.0	0.0029	5.61	0.91
		3	1.0	0.0028	5.32	0.91
Trait	Bayesian	1	0.96	0.0028	5.39	0.90
		3	0.96	0.0026	5.18	0.90
	NPRS	1	0.97	0.0026	5.06	0.91
		3	0.97	0.0025	4.84	0.92

FIGURE 4 PD ranking errors with three calibrations. 100 simulated community assemblages are ranked by FPD value when FPD is derived from branch lengths simulated under a speciation rate or trait-based scenario and reconstructed using Bayesian analysis with an uncorrelated lognormal relaxed clock and three calibrations. Each rectangle represents one phylogeny and is shaded to represent its true rank, as shown on the left of each diagram. The right of each diagram shows the rank based on reconstructed FPD, with the true ranks of taxa incorrectly included in or excluded from the top 10 indicated by black numbers on the right



available as Supplementary Information). Looking at how the position of each taxon changed when we used reconstructed ED, we found that taxa were mis-ranked by 10-11 positions on average and 20-40 positions at the 95th percentile compared to their rankings based on true ED values (BEAST, Figure 6; NPRS, Figure S4 available as Supplementary Information). Taxa that were top-ranked in the true tree were substantially more likely to be correctly ranked

than those that had ED values in the middle of the ranking. An alternative way to interpret this data is to compare the proportion of the top 10, 50 (and so on) ranked species that are correctly identified under estimation. The above results are then equivalent to saying that 83-87% of the top 10 or top 50 species are correctly identified by estimation, whereas about 90% of the top 80 are correctly identified.

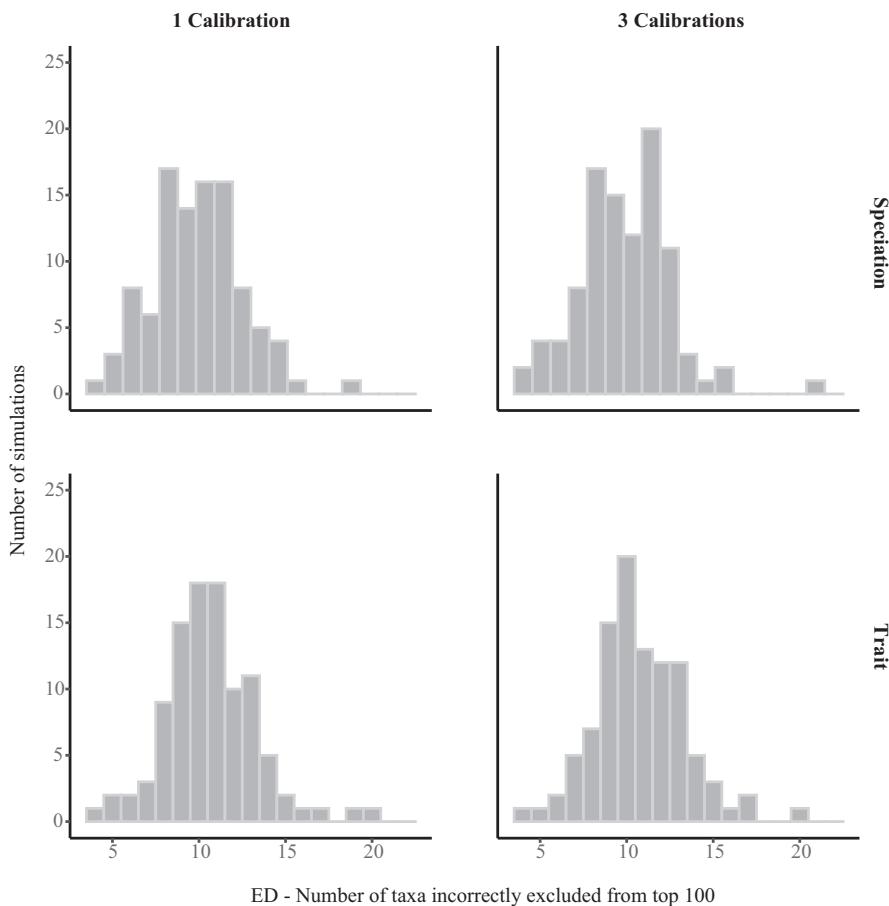


FIGURE 5 Histograms showing the number of taxa incorrectly excluded from the top 100 ED rankings when ED is calculated from reconstructed branch lengths. The number of excluded taxa is calculated from 100 replicate simulations for each of 4 combinations of simulation scenarios and calibration regimes. All branch lengths were estimated by Bayesian reconstruction in BEAST 2 (for NPRS results see Figure S3 available as Supplementary Information)

3.3 | Analysis of the largest errors

We asked whether the cases with the largest FPD and ED errors were due to failures of phylogenetic inference, in which errors in topology or divergence times might be large enough to be noticed by inspection of the tree. To test this, we asked whether the trees with the greatest error in FPD and ED were those for which the reconstruction was very different from the true tree, as indicated by larger tree distances between the true and reconstructed trees. We therefore plotted topological and branch score errors for each reconstructed phylogeny (Figure 7; see Supplementary Methods). While these largest FPD errors did cluster at the upper end of the distribution of branch length errors, for topology errors they are observed throughout the plot. The greatest ED ranking errors were not clustered within the upper tail of either the distribution of topology errors or the distribution of branch length errors.

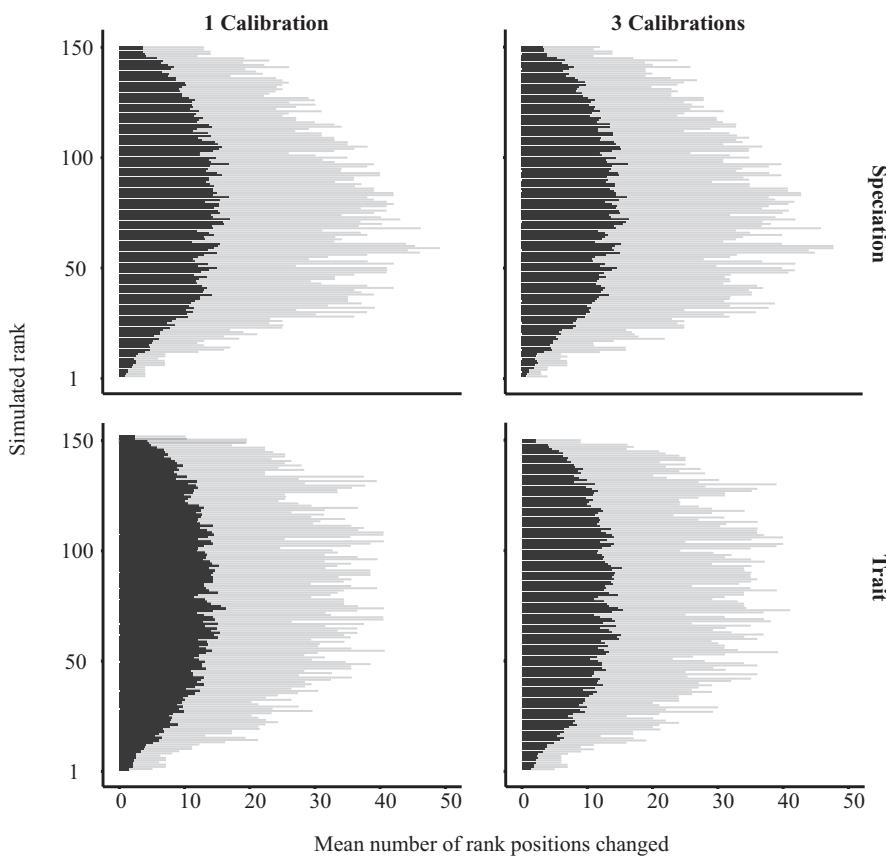
4 | Discussion

Phylogenetic diversity (FPD) or Evolutionary Distinctness (ED) measures aim to produce universal metrics that reflect important properties for conservation prioritization. These measures do not directly target species characteristics or assemblage properties that are considered desirable for conservation, but instead aim to produce objective measures that will scale with important characteristics of

biodiversity. These metrics are commonly taken to represent evolutionary time and/or opportunity for acquisition of unique traits. There has been a vigorous debate on whether phylo-diversity measures are appropriate tools for conservation prioritization, and the accuracy of various metrics at making representations of phylogenetic properties has been tested. But most such tests assume that the phylogeny is known without error. Given that phylo-diversity measures are usually based on molecular phylogenetic branch lengths, it is important to consider the accuracy of phylogenetic reconstruction. Here we focus on only one potential source of uncertainty, arising from lineage-specific patterns of molecular rate variation, but other issues such as calibration error or gene tree incongruence could also lead to inaccuracy. Ideally, researchers would consider the variance arising from all aspects of molecular phylogenetic inference (e.g. Bromham et al., 2018).

In order to examine how well inferred FPD represents true evolutionary history, we have simulated data sets that mimic realistic studies, based on a survey of the literature and known patterns of rate variation. We then ask how close FPD measures on phylogenies inferred for these data are to the true underlying history. We have found that the two phylogenetic reconstruction methods most commonly used in studies of phylo-diversity are expected to lead to average levels of error in FPD estimates of 6 to 14% under realistic models of molecular rate variation, and up to 24–38% error, depending on the reconstruction method and number of calibrations (Table 2). These levels of error can impact ranking of assemblages

FIGURE 6 Number of rank changes based on Evolutionary Distinctiveness when ED is derived from estimated branch lengths. ED is calculated from branch lengths reconstructed through a Bayesian analysis under an uncorrelated relaxed clock, for two rate variation scenarios and two calibration conditions. Dark grey bars indicate the mean number of positions the taxon at each rank position has moved in either direction after estimation, across 100 reconstructed rankings for each combination of treatments. 1 is the highest rank position while 150 is the lowest. Light grey bars indicate the empirical upper 95% confidence bound of the number of rank changes (i.e. 95% of reconstructed rankings had each rank position change by this number or less)



by phylo-diversity and lead to changes in rankings based on evolutionary distinctness. The mean errors are comparable to those of up to $\pm 0.5\%$ reported in one study of species richness estimators for sample sizes greater than 20 (Gwinn et al., 2016), despite the greater number of parameters in phylogenetic models.

How much impact would phylogenetic error of this magnitude have on conservation prioritization? If prioritization of limited conservation resources, or selection of relatively few sites for protection, was based on the selection of top-ranked areas based on FPD alone, then we would in some cases miss making the optimal decision due to error in the reconstruction of branch lengths. While some recommendations for priority conservation areas have been made using simple scoring and ranking schemes, typically FPD is considered in combination with other facets of biodiversity such as taxonomic and functional diversity (e.g. Mazel et al., 2014; Soutullo et al., 2005). Recommendations for incorporating phylo-diversity metrics into conservation plans commonly make use of complementarity-based reserve selection algorithms that aim to achieve a pre-determined biodiversity target with maximum efficiency (e.g. Brum et al., 2017; Morales-Barbero & Ferrer-Castán, 2019; Pollock et al., 2015; Rosauer et al., 2017). Whether measures of phylo-diversity are used as the sole input to conservation ranking or combined with other information to set priorities, it is appropriate to ask whether the expected degree of error in these measures could impact on prioritization processes. Our literature survey suggests that the aim of many is to provide relative ranking of conservation priorities either of individual species or species assemblages by their inferred

phylogenies (Table 1). However, it is unclear the degree to which such fine distinctions would impact on real-world conservation planning.

For example, FPD and ED may form part of frameworks that incorporate estimates of extinction risk and prioritize conservation action based on the expected loss of phylo-diversity. For FPD, this is formalized in the “expected” or “probabilistic” diversity framework, which prioritizes combinations of taxa by the average amount of phylo-diversity that could be lost if they are not protected. For ED, the “EDGE of Existence” project combines phylogenetically derived ED scores with the conservation status of the taxon, as represented by its classification under the IUCN scale of extinction risk. In essence, a taxon with few close relatives (high ED) that is also globally endangered (GE) will get the highest ranking. Because EDGE ranking combines phylogenetic information with extinction risk evaluation, error in phylogenetic branch length is unlikely to change the ranking of most cases. But we have shown that uncertainty in ED can change the relative weighting of species based on phylogenetic information, which leaves open the possibility for error in cases where threat status is less of a factor.

One way of seeing this issue is that FPD or ED can serve as a way to order the relative conservation value of species within a given category of threat. The Endangered Ganges river dolphin (*Platanista gangetica*), the only living member of its family, has a higher EDGE score than the Endangered Amazon river dolphin (*Inia geoffrensis*), which is co-familial with two other living South American river dolphins (*I. araguaiae* and *I. boliviensis*). Another example is provided by the Hornbills. Three species (*Rhinoploax vigil*, *Anthracoceros montani* and *Rhabdotorrhinus waldeni*) have the same IUCN Red List status (Critically

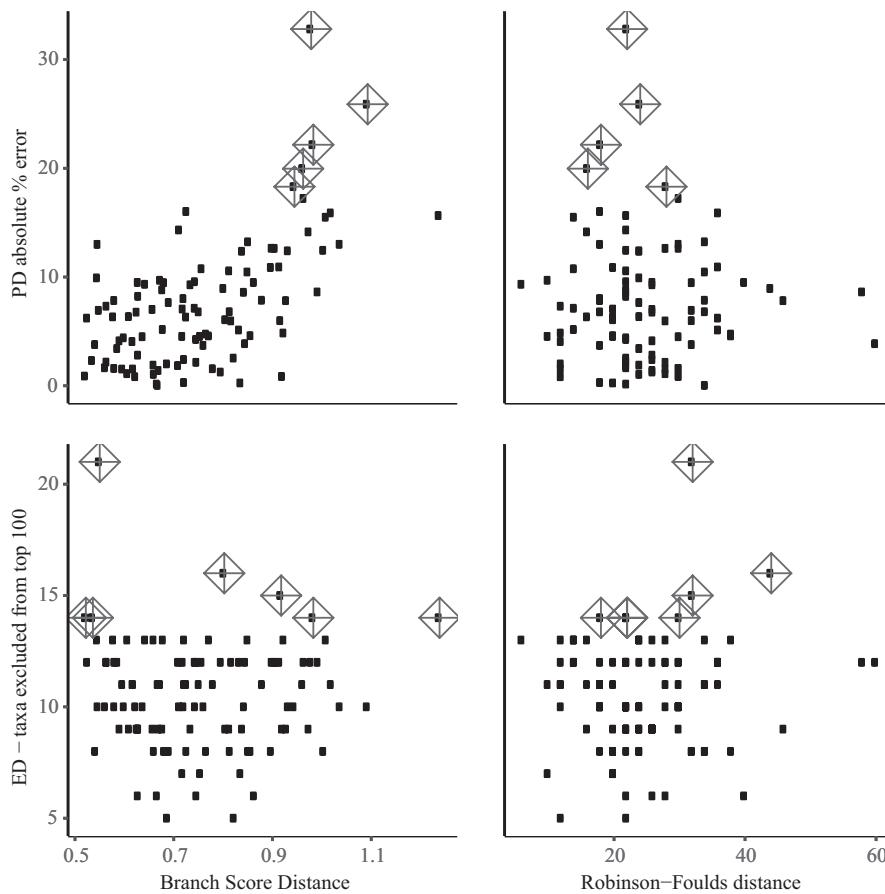


FIGURE 7 Distribution of the 95th percentile of most severe FPD and ED estimation errors across different levels of phylogenetic error. Phylogenetic error is calculated either by the Robinson–Foulds distance, emphasizing topological differences or the branch score distance, emphasizing branch lengths. The 95th percentile of most severe FPD and ED errors are marked by diamonds, severity is given as the absolute percentage error for FPD and as the number of taxa incorrectly excluded from the top 100 for ED. Data are shown for the speciation scenario under BEAST 2 with 3 calibrations

Endangered), but have quite different EDGE rankings (26, 41 and 50, respectively) on the basis of branch lengths in published phylogenies (Jetz et al., 2014). These rankings could change if EDGE scores were based on phylogenies inferred under different models or assumptions.

Levels of phylogenetic error reported in this study are unlikely to impact high profile cases concerning the most evolutionarily distinct species in a given EDGE list, such as the Wollemi pine, and however, they may impact which critically endangered species with lower ED scores are included in top 100 lists, with implications for recognition and prioritization. We have shown that the ED scores in the middle of the distribution are more vulnerable to ranking error (Figure 4). In fact, many EDGE species with higher global endangerment values have ED values in the middle of the range for their taxon group (Figure 8). This could mean that rankings among these more highly endangered species are especially vulnerable to ED estimation error caused by branch length reconstruction. The assemblage phylogenies used in this study are much smaller and less taxonomically complete than the large trees used in many EDGE studies (e.g. Bininda-Emonds et al., 2007). But here we demonstrate that realistic levels of error in ED scores can lead to changes in rankings and are therefore worth investigating when using ED scores to set conservation priorities.

A theme of conservation planning is the necessity of properly characterizing the risk arising from potential worst-case outcomes, given the irreplaceability of biological resources (Brooks et al., 2006). Accounting for this risk requires accounting not only for expected average level of error, but also for the worst-case scenarios (Daniel

P. Faith, 2008; Daniel P. Faith, 2015). Since the goal of this study is to examine the potential for error in FPD and ED outcomes arising from common molecular phylogenetic inference on realistic data sets, it is interesting to study the characteristics of the largest errors we encountered. When estimating FPD, the least accurate cases based on Bayesian phylogenies involve errors of up to 32.8% of the true FPD value (Table 2). In the data sets with the greatest ED ranking errors, one fifth of the membership of a top 100 list could be incorrectly included, if most or all of the species were at the highest level of endangerment (Figure 5). Whether errors of this magnitude will have significant real-world impact on prioritization will depend on how they are estimated and applied in any given case study. Where the outcomes of such studies are considered to have important implications for conservation, the potential impact of phylogenetic error should be considered. Notably, FPD and ED scores are commonly applied to very large trees (Arregoitia et al., 2013; Jetz & Pyron, 2018) or to smaller phylogenies extracted from supertrees (approximately 60% of the sampled studies used supertrees: Table S4 available as supplementary information). Errors in branch lengths may thus be propagated through many studies that all sample the same supertree.

It is important to note that errors in phylo-diversity measures may be significant even where there are no obvious errors in the inferred phylogeny. Our results show that the largest errors in FPD and ED are not necessarily associated with major topology errors or large-scale errors in inferring branch lengths. This suggests that it will not always be possible to detect errors from obvious inconsistencies in

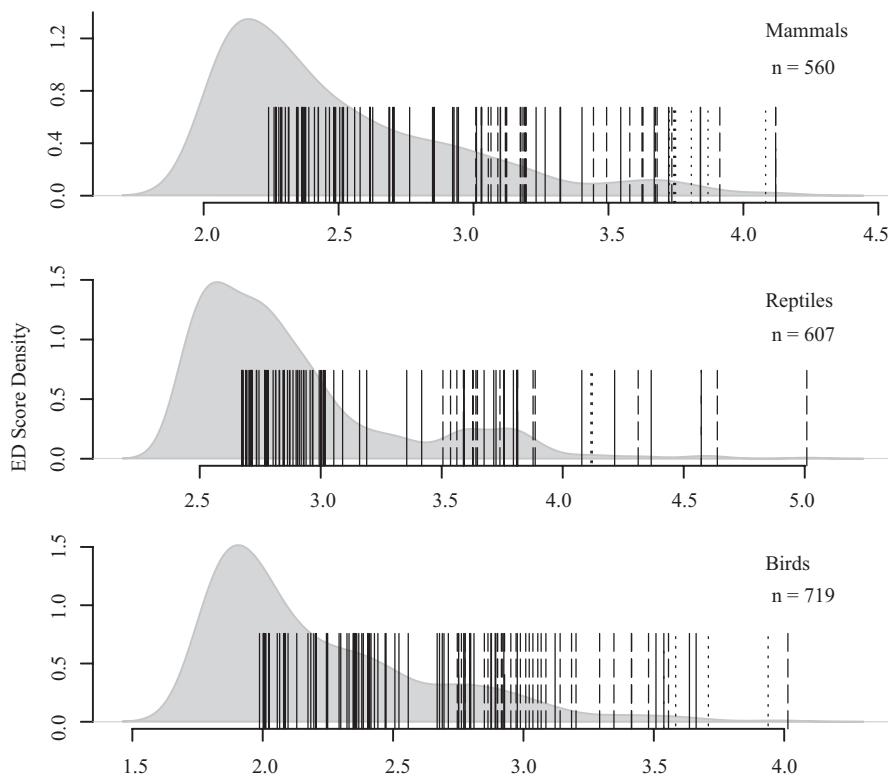


FIGURE 8 Illustration of the log ED values for EDGE top species lists. These represent the top 100 EDGE rankings for mammals, reptiles and birds. The grey curve indicates the distribution of log ED values for all taxa in the respective phylogeny. Vertical lines mark the position in the distribution of top 100 species. Line style indicates endangerment status: solid, critically endangered (CR); long dashed, endangered (EN); and dotted, vulnerable (VU). Each EDGE list includes a large component of endangered and critically endangered species with ED scores that fall within the middle of the distribution. This is an example of a situation in which critically endangered taxa could be excluded from the rankings if scores used in the EDGE lists were to suffer from ranking errors like those observed in our study. Data were sourced from <https://www.edgeofexistence.org/edge-lists> (accessed 13 Dec 2019)

the reconstructed topology. More methodological work is needed to determine the precise characteristics of trees that bear on this.

This discussion also touches on a broader issue in the testing and validation of evolutionary models. An average rate of error that is considered acceptable for some uses, such as phylogeny estimation or molecular dating, might be considered unacceptably high for practical applications, such as conservation prioritization. Here, the number of measurements is generally small; for example, community metrics may be calculated for a small number of potential reserves in a particular locality, or species may be ranked based on their individual EDGE scores from a single estimate of a phylogeny. Therefore, a level of error such as that found in these simulations could have a significant impact on conclusions of a prioritization study. We need to know not only which qualities of a phylogeny most align with the things we wish to conserve, but also how reliably our measurements are able to determine those qualities. Where the outcomes of phylogenetic studies are expected to have real-world impacts on conservation, it is vital that clear statements can be made about the accuracy and precision of such measures. Importantly, we must model the error arising from all stages of the estimation process, including the uncertainty in phylogenetic reconstruction, because even “best practice” phylogenetic methods are not perfect.

5 | Conclusions and recommendations for future research

We have characterized the likely levels of error in estimating phylo-diversity metrics when rates of molecular evolution vary between species, even when a rate-variable method is used to infer the phylogeny. This level of error may be acceptable for many practical applications of biodiversity metrics, but demonstrates that phylogeny should be considered as a potential source of error when designing prioritization methods. The levels of phylogenetic error reported under biologically realistic levels of rate variation, using common phylogenetic methods, could be sufficient to have an impact on prioritization rankings based on phylogenetic measures. Future studies could consider the potential impact of other types of clock model misspecification on phylo-diversity estimates, as well as studying the impacts of this and other error sources in the practical application of phylogenetic diversity to conservation.

ACKNOWLEDGEMENTS

The authors wish to thank Professor Kim Sterelny and Dr. Chris Lean for input in the early stages of the project.

PEER REVIEW

The peer review history for this article is available at <https://publon.com/publon/10.1111/ddi.13179>.

DATA AVAILABILITY STATEMENT

Data and code associated with this analysis are available via Mendeley Data at <https://data.mendeley.com/datasets/52ht8hr93p/draft?a=a1face2f-9d75-4075-b8a6-e9f503cf145d>.

ORCID

Andrew M. Ritchie  <https://orcid.org/0000-0001-9716-0649>

REFERENCES

Arregoitia, L. D. V., Blomberg, S. P., & Fisher, D. O. (2013). Phylogenetic correlates of extinction risk in mammals: species in older lineages are not at greater risk. *Proceedings of the Royal Society B-Biological Sciences*, 280(1765), 20131092. <https://doi.org/10.1098/rspb.2013.1092>.

Barraclough, T. G., & Savolainen, V. (2001). Evolutionary rates and species diversity in flowering plants. *Evolution*, 55, 677–683.

Battistuzzi, F. U., Filipski, A., Hedges, S. B., & Kumar, S. (2010). Performance of relaxed-clock methods in estimating evolutionary divergence times and their credibility intervals. *Molecular Biology and Evolution*, 27(6), 1289–1300. <https://doi.org/10.1093/molbev/msq014>.

Berv, J. S., & Field, D. J. (2018). Genomic signature of an avian Lilliput effect across the K-Pg extinction. *Systematic Biology*, 67(1), 1–13. <https://doi.org/10.1093/sysbio/syx064>.

Bininda-Emonds, O. R. P., Cardillo, M., Jones, K. E., MacPhee, R. D. E., Beck, R. M. D., Grenyer, R., ... Purvis, A. (2007). The delayed rise of present-day mammals. *Nature*, 446, 507–512.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., ... Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, 10(4), e1003537. <https://doi.org/10.1371/journal.pcbi.1003537>.

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., ... Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, 15(4), e1006650. <https://doi.org/10.1371/journal.pcbi.1006650>.

Bromham, L. (2009). Why do species vary in their rate of molecular evolution? *Biology Letters*, 5, 401–404.

Bromham, L. (2011). The genome as a life-history character: why rate of molecular evolution varies between mammal species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1577), 2503–2513.

Bromham, L. (2019). Six impossible things before breakfast: assumptions, models, and belief in molecular dating. *Trends in Ecology & Evolution*, 34(5), 474–486. <https://doi.org/10.1016/j.tree.2019.01.017>.

Bromham, L., Duchêne, S., Hua, X., Ritchie, A. M., Duchêne, D. A., & Ho, S. Y. W. (2018). Bayesian molecular dating: opening up the black box. *Biological Reviews*, 93(2), 1165–1191. <https://doi.org/10.1111/brv.12390>.

Bromham, L., Hua, X., Lanfear, R., & Cowman, P. (2015). Exploring the relationships between mutation rates, life history, genome size, environment and species richness in flowering plants. *American Naturalist*, 185, 507–524.

Brooks, T. M., Cuttelod, A., Faith, D. P., Garcia-Moreno, J., Langhammer, P., & Pérez-España, S. (2015). Why and how might genetic and phylogenetic diversity be reflected in the identification of key biodiversity areas? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1662), 20140019. <https://doi.org/10.1098/rstb.2014.0019>.

Brooks, T. M., Mittermeier, R. A., da Fonseca, G. A. B., Gerlach, J., Hoffmann, M., Lamoreux, J. F., ... Rodrigues, A. S. L. (2006). Global biodiversity conservation priorities. *Science*, 313(5783), 58–61. <https://doi.org/10.1126/science.1127609>.

Brum, F. T., Graham, C. H., Costa, G. C., Hedges, S. B., Penone, C., Radeloff, V. C., ... Davidson, A. D. (2017). Global priorities for conservation across multiple dimensions of mammalian diversity. *Proceedings of the National Academy of Sciences*, 114(29), 7641. <https://doi.org/10.1073/pnas.1706461114>.

Cadotte, M. W., Davies, T. J., & Peres-Neto, P. R. (2017). Why phylogenies do not always predict ecological differences. *Ecological Monographs*, 87(4), 535–551. <https://doi.org/10.1002/ecm.1267>.

Crozier, R. H. (1997). Preserving the information content of species: genetic diversity, phylogeny, and conservation worth. *Annual Review of Ecology and Systematics*, 28(1), 243–268. <https://doi.org/10.1146/annurev.ecolsys.28.1.243>.

Curnick, D. J., Head, C. E. I., Huang, D., Crabbe, M. J. C., Gollock, M., Hoeksema, B. W., ... Redding, D. W. (2015). Setting evolutionary-based conservation priorities for a phylogenetically data-poor taxonomic group (Scleractinia). *Animal Conservation*, 18(4), 303–312. <https://doi.org/10.1111/acv.12185>.

Devictor, V., Mouillot, D., Meynard, C., Jiguet, F., Thuiller, W., & Mouquet, N. (2010). Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. *Ecology Letters*, 13(8), 1030–1040. <https://doi.org/10.1111/j.1461-0248.2010.01493.x>.

dos Reis, M., Donoghue, P. C. J., & Yang, Z. (2016). Bayesian molecular clock dating of species divergences in the genomics era. *Nature Reviews Genetics*, 17(2), 71–80. <https://doi.org/10.1038/nrg.2015.8>.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLOS Biology*, 4(5), e88. <https://doi.org/10.1371/journal.pbio.0040088>.

Duchêne, D. A., Hua, X., & Bromham, L. (2017). Phylogenetic estimates of diversification rate are affected by molecular rate variation. *Journal of Evolutionary Biology*, 30(10), 1884–1897.

Elliott, M. J., Knerr, N. J., & Schmidt-Lebuhn, A. N. (2018). Choice between phylogram and chronogram can have a dramatic impact on the location of phylogenetic diversity hotspots. *Journal of Biogeography*, 45(9), 2190–2201. <https://doi.org/10.1111/jbi.13399>.

Ho, S. H., & DeWoody, J. A. (2010). Evolutionary rates of mitochondrial genomes correspond to diversification rates and to contemporary species richness in birds and reptiles. *Proceedings of the Royal Society B: Biological Sciences*, 277, 3587–3592.

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

Faith, D. P. (2007). Biodiversity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved from <http://plato.stanford.edu/archives/fall2008/entries/biodiversity>

Faith, D. P. (2008). Threatened species and the potential loss of phylogenetic diversity: conservation scenarios based on estimated extinction probabilities and phylogenetic risk analysis. *Conservation Biology*, 22(6), 1461–1470. <https://doi.org/10.1111/j.1523-1739.2008.01068.x>.

Faith, D. P. (2015). Phylogenetic diversity, functional trait diversity and extinction: avoiding tipping points and worst-case losses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1662), 20140011. <https://doi.org/10.1098/rstb.2014.0011>.

Faith, D. P. (2018). Phylogenetic diversity and conservation evaluation: perspectives on multiple values, indices, and scales of application. In R. A. Scherson, & D. P. Faith (Eds.), *Phylogenetic Diversity: Applications and Challenges in Biodiversity Science* (pp. 1–26). Springer International Publishing.

Flynn, D. F. B., Mirochnick, N., Jain, M., Palmer, M. I., & Naeem, S. (2011). Functional and phylogenetic diversity as predictors of biodiversity-ecosystem-function relationships. *Ecology*, 92(8), 1573–1581. <https://doi.org/10.1890/10-1245.1>.

Forest, F., Crandall, K. A., Chase, M. W., & Faith, D. P. (2015). Phylogeny, extinction and conservation: embracing uncertainties in a time of urgency. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 370(1662), 20140002. <https://doi.org/10.1098/rstb.2014.0002>.

Fritz, S. A., & Purvis, A. (2010). Phylogenetic diversity does not capture body size variation at risk in the world's mammals. *Proceedings of the Royal Society B-Biological Sciences*, 277(1693), 2435–2441. <https://doi.org/10.1098/rspb.2010.0030>.

Gumbs, R., Gray, C. L., Wearn, O. R., & Owen, N. R. (2018). Tetrapods on the EDGE: Overcoming data limitations to identify phylogenetic conservation priorities. *PLOS ONE*, 13(4), e0194680.

Gwin, D. C., Allen, M. S., Bonvechio, K. I., Hoyer, V., & Beesley, L. S. (2016). Evaluating estimators of species richness: the importance of considering statistical error rates. *Methods in Ecology and Evolution*, 7(3), 294–302. <https://doi.org/10.1111/2041-210X.12462>.

Hedges, S. B., & Kumar, S. (2009). *The Timetree of Life*. Oxford University Press.

Hendry, A. P., Lohmann, L. G., Conti, E., Cracraft, J., Crandall, K. A., Faith, D. P., ... Donoghue, M. J. (2010). Evolutionary biology in biodiversity science, conservation, and policy: a call to action. *Evolution*, 64(5), 1517–1528. <https://doi.org/10.1111/j.1558-5646.2010.00947.x>.

Ho, S. Y. W. (2014). The changing face of the molecular evolutionary clock. *Trends in Ecology & Evolution*, 29(9), 496–503. <https://doi.org/10.1016/j.tree.2014.07.004>.

Ho, S. Y. W., Phillips, M. J., Drummond, A. J., & Cooper, A. (2005). Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. *Molecular Biology and Evolution*, 22(5), 1355–1363. <https://doi.org/10.1093/molbev/msi125>.

Hua, X., & Bromham, L. (2017). Darwinism for the genomic age: connecting mutation to diversification. *Frontiers in Genetics*, 8, 12.

Hua, X., Cowman, P., Warren, D., & Bromham, L. (2015). Longevity is linked to mitochondrial mutation rates in rockfish: a test using Poisson regression. *Molecular Biology and Evolution*, 32(10), 2633–2645. <https://doi.org/10.1093/molbev/msv137>.

Huang, D. W., Hoeksema, B. W., Affendi, Y. A., Ang, P. O., Chen, C. L. A., Huang, H., ... Chou, L. M. (2016). Conservation of reef corals in the South China Sea based on species and evolutionary diversity. *Biodiversity and Conservation*, 25(2), 331–344. <https://doi.org/10.1007/s10531-016-1052-7>.

Isaac, N. J. B., & Pearse, W. D. (2018). The use of EDGE (Evolutionary Distinct Globally Endangered) and EDGE-like metrics to evaluate taxa for conservation. In R. A. Scherson, & D. P. Faith (Eds.), *Phylogenetic Diversity: Applications and Challenges in Biodiversity Science* (pp. 27–39). Springer International Publishing.

Isaac, N. J. B., Redding, D. W., Meredith, H. M., & Safi, K. (2012). Phylogenetically-informed priorities for amphibian conservation. *Plos One*, 7(8), <https://doi.org/10.1371/journal.pone.0043912>.

Isaac, N. J. B., Turvey, S. T., Collen, B., Waterman, C., & Baillie, J. E. M. (2007). Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLOS ONE*, 2(3), e296. <https://doi.org/10.1371/journal.pone.0000296>.

Jetz, W., & Pyron, R. A. (2018). The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. *Nature Ecology & Evolution*, 2(5), 850–858. <https://doi.org/10.1038/s41559-018-0515-5>.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, 491, 444. <https://doi.org/10.1038/nature11631>.

Jetz, W., Thomas, G. H., Joy, J. B., Redding, D. W., Hartmann, K., & Mooers, A. O. (2014). Global distribution and conservation of evolutionary distinctness in birds. *Current Biology*, 24(9), 919–930. <https://doi.org/10.1016/j.cub.2014.03.011>.

Kelly, S., Grenyer, R., & Scotland, R. W. (2014). Phylogenetic trees do not reliably predict feature diversity. *Diversity and Distributions*, 20(5), 600–612. <https://doi.org/10.1111/ddi.12188>.

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., ... Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26(11), 1463–1464. <https://doi.org/10.1093/bioinformatics/btq166>.

Kondratyeva, A., Grandcolas, P., & Pavoine, S. (2019). Reconciling the concepts and measures of diversity, rarity and originality in ecology and evolution. *Biological Reviews*, 94(4), 1317–1337. <https://doi.org/10.1111/brv.12504>.

Kumar, S. (2005). Molecular clocks: four decades of evolution. *Nature Reviews Genetics*, 6(8), 654–662. <https://doi.org/10.1038/nrg1659>.

Laity, T., Laffan, S. W., Gonzalez-Orozco, C. E., Faith, D. P., Rosauer, D. F., Byrne, M., ... Newport, K. (2015). Phylodiversity to inform conservation policy: an Australian example. *Science of the Total Environment*, 534, 131–143. <https://doi.org/10.1016/j.scitotenv.2015.04.113>.

Lanfear, R., Ho, S. Y. W., Love, D., & Bromham, L. (2010). Mutation rate influences diversification rate in birds. *Proceedings of the National Academy of Sciences*, 107(47), 20423–20428.

Lean, C., & Maclaurin, J. (2016). The value of phylogenetic diversity. In R. Pellen, & G. Philippe, (Eds.), *Biodiversity conservation and phylogenetic systematics: Preserving our evolutionary heritage in an extinction crisis* (pp. 19–37). Springer.

Lepage, T., Bryant, D., Philippe, H., & Lartillot, N. (2007). A general comparison of relaxed molecular clock models. *Molecular Biology and Evolution*, 24(12), 2669–2680. <https://doi.org/10.1093/molbev/msm193>.

Letten, A. D., Keith, D. A., & Tozer, M. G. (2014). Phylogenetic and functional dissimilarity does not increase during temporal heathland succession. *Proceedings of the Royal Society B-Biological Sciences*, 281(1797), 20142102. <https://doi.org/10.1098/rspb.2014.2102>.

Maclaurin, J., & Sterelny, K. (2008). *What Is Biodiversity?*. University Of Chicago Press.

Maier, D. S. (2012). *Theories of Biodiversity Value What's So Good About Biodiversity?*. Springer.

Margules, C. R., & Pressey, R. L. (2000). Systematic conservation planning. *Nature*, 405(6783), 243–253. <https://doi.org/10.1038/35012251>.

Mazel, F., Guilhaumon, F., Mouquet, N., Devictor, V., Gravel, D., Renaud, J., ... Thuiller, W. (2014). Multifaceted diversity-area relationships reveal global hotspots of mammalian species, trait and lineage diversity. *Global Ecology and Biogeography*, 23(8), 836–847. <https://doi.org/10.1111/geb.12158>.

Mazel, F., Mooers, A. O., Dalla Riva, G. V., & Pennell, M. W. (2017). Conserving phylogenetic diversity can be a poor strategy for conserving functional diversity. *Systematic Biology*, 66(6), 1019–1027. <https://doi.org/10.1093/sysbio/syx054>.

Mazel, F., Pennell, M. W., Cadotte, M. W., Diaz, S., Dalla Riva, G. V., Grenyer, R., ... Pearse, W. D. (2018). Prioritizing phylogenetic diversity captures functional diversity unreliably. *Nature Communications*, 9(1), 2888. <https://doi.org/10.1038/s41467-018-05126-3>.

Mooers, A. O., Heard, S. B., & Chrostowski, E. (2005). Evolutionary heritage as a metric for conservation. In A. Purvis, J. L. Gittleman, & T. Brooks (Eds.), *Phylogeny and Conservation* (Vol. 8, pp. 120–138).

Morales-Barbero, J., & Ferrer-Castán, D. (2019). Using a goal programming approach to design and evaluate protected areas for the conservation of multiple dimensions of biodiversity. *Journal for Nature Conservation*, 49, 54–62. <https://doi.org/10.1016/j.jnc.2019.01.007>.

Muenchow, J., Dieker, P., Kluge, J., Kessler, M., & von Wehrden, H. (2018). A review of ecological gradient research in the Tropics: identifying research gaps, future directions, and conservation priorities. *Biodiversity and Conservation*, 27(2), 273–285. <https://doi.org/10.1007/s10531-017-1465-y>.

Nabholz, B., Lanfear, R., & Fuchs, J. (2016). Body mass-corrected molecular rate for bird mitochondrial DNA. *Molecular Ecology*, 25(18), 4438–4449. <https://doi.org/10.1111/mec.13780>.

Pagel, M., Venditti, C., & Meade, A. (2006). Large Punctuational contribution of speciation to evolutionary divergence at the

molecular level. *Science*, 314(5796), 119. <https://doi.org/10.1126/science.1129647>.

Park, D. S., Worthington, S., & Xi, Z. (2018). Taxon sampling effects on the quantification and comparison of community phylogenetic diversity. *Molecular Ecology*, 27(5), 1296–1308.

Pollock, L. J., Rosauer, D. F., Thornhill, A. H., Kujala, H., Crisp, M. D., Miller, J. T., & McCarthy, M. A. (2015). Phylogenetic diversity meets conservation policy: small areas are key to preserving eucalypt lineages. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 370(1662), 20140007. <https://doi.org/10.1098/rstb.2014.0007>.

Qiu, F., Kitchen, A., Burleigh, J. G., & Miyamoto, M. M. (2014). Scombrid fishes provide novel insights into the trait/rate associations of molecular evolution. *Journal of Molecular Evolution*, 78(6), 338–348. <https://doi.org/10.1007/s00239-014-9621-4>.

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org>.

Rosauer, D. F., & Mooers, A. O. (2013). Nurturing the use of evolutionary diversity in nature conservation. *Trends in Ecology & Evolution*, 28(6), 322–323. <https://doi.org/10.1016/j.tree.2013.01.014>.

Rosauer, D. F., Pollock, L. J., Linke, S., & Jetz, W. (2017). Phylogenetically informed spatial planning is required to conserve the mammalian tree of life. *Proceedings of the Royal Society B: Biological Sciences*, 284(1865), 20170627. <https://doi.org/10.1098/rspb.2017.0627>.

Safi, K., Armour-Marshall, K., Baillie, J. E. M., & Isaac, N. J. B. (2013). Global patterns of evolutionary distinct and globally endangered amphibians and mammals. *Plos One*, 8(5), <https://doi.org/10.1371/journal.pone.0063582>.

Sanderson, M. J. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*, 19(2), 301–302. <https://doi.org/10.1093/bioinformatics/19.2.301>.

Server, B. A. J., Pennell, M. W., Brown, J. W., Keeble, S., Hardwick, K. M., Sullivan, J., & Harmon, L. J. (2019). The choice of tree prior and molecular clock does not substantially affect phylogenetic inferences of diversification rates. *PeerJ*, 7, <https://doi.org/10.7717/peerj.6334>.

Shafir, A., Azouri, D., Goldberg, E. E., & Mayrose, I. (2020). Heterogeneity in the rate of molecular sequence evolution substantially impacts the accuracy of detecting shifts in diversification rates. *Evolution*, <https://doi.org/10.1111/evo.14036>.

Smith, S. A., & O'Meara, B. C. (2012). treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics*, 28(20), 2689–2690. <https://doi.org/10.1093/bioinformatics/bts492>.

Soutullo, A., Dodsworth, S., Heard, S. B., & Mooers, A. O. (2005). Distribution and correlates of carnivore phylogenetic diversity across the Americas. *Animal Conservation*, 8, 249–258. <https://doi.org/10.1017/s136794300500226x>.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.

Stein, R. W., Mull, C. G., Kuhn, T. S., Aschliman, N. C., Davidson, L. N. K., Joy, J. B., ... Mooers, A. O. (2018). Global priorities for conserving the evolutionary history of sharks, rays and chimaeras. *Nature Ecology & Evolution*, 2(2), 288–298. <https://doi.org/10.1038/s41559-017-0448-4>.

Thomas, J. A., Welch, J. J., Lanfear, R., & Bromham, L. (2010). A generation time effect on the rate of molecular evolution in invertebrates. *Molecular Biology and Evolution*, 27(5), 1173–1180. <https://doi.org/10.1093/molbev/msq009>.

Thorne, J. L., Kishino, H., & Painter, I. S. (1998). Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution*, 15(12), 1647–1657. <https://doi.org/10.1093/oxfordjournals.molbev.a025892>.

Tucker, C. M., Aze, T., Cadotte, M. W., Cantalapiedra, J. L., Chisholm, C., Díaz, S., ... Mooers, A. O. (2019). Assessing the utility of conserving evolutionary history. *Biological Reviews*, 94(5), 1740–1760. <https://doi.org/10.1111/brv.12526>.

Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A., ... Mazel, F. (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological Reviews*, 92(2), 698–715. <https://doi.org/10.1111/brv.12252>.

Tucker, C. M., Davies, T. J., Cadotte, M. W., & Pearse, W. D. (2018). On the relationship between phylogenetic diversity and trait diversity. *Ecology*, 99(6), 1473–1479. <https://doi.org/10.1002/ecy.2349>.

Webb, C. O., Ackerly, D. D., & Kembel, S. W. (2008). Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics*, 24(18), 2098–2100.

Webster, A. J., Payne, R. J. H., & Pagel, M. (2003). Molecular phylogenies link rates of evolution and speciation. *Science*, 301, 478.

Welch, J. J., Bininda-Emonds, O. R. P., & Bromham, L. (2008). Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evolutionary Biology*, 8(1), 53. <https://doi.org/10.1186/1471-2148-8-53>.

Wertheim, J. O., & Sanderson, M. J. (2011). Estimating Diversification Rates: How Useful Are DIVERGENCE TIMES? *Evolution*, 65(2), 309–320. <https://doi.org/10.1111/j.1558-5646.2010.01159.x>.

Winter, M., Devictor, V., & Schweiger, O. (2013). Phylogenetic diversity and nature conservation: where are we? *Trends in Ecology & Evolution*, 28(4), 199–204. <https://doi.org/10.1016/j.tree.2012.10.015>.

Wong, A. (2014). Covariance between testes size and substitution rates in primates. *Molecular Biology and Evolution*, 31(6), 1432–1436. <https://doi.org/10.1093/molbev/msu091>.

Yang, Z., & Rannala, B. (2005). Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution*, 23(1), 212–226. <https://doi.org/10.1093/molbev/msj024>.

ZSL (2014). EDGE of Existence. Retrieved from <https://www.edgeofexistence.org>.

BIOSKETCH

The Macroevolution and Macroecology Group studies the evolutionary and ecological processes responsible for generating patterns of biodiversity. Our research covers a wide range of questions, taxonomic groups and spatial scales. Current directions include the diversification and biogeography of Proteaceae; predicting range and extinction risk under climate change; investigating the link between molecular evolution, life history and diversification; and unravelling the drivers of human language diversity. L.B. conceived the study with input from K.J.Y and M.C. K.J.Y. conducted the pilot study. X.H. and R.D. contributed data and code for simulations. A.M.R. performed the final analyses. L.B. and A.M.R. led the writing, to which all authors contributed.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Ritchie AM, Hua X, Cardillo M, Yaxley KJ, Dinnage R, Bromham L. Phylogenetic diversity metrics from molecular phylogenies: modelling expected degree of error under realistic rate variation. *Divers Distrib*. 2021;27:164–178. <https://doi.org/10.1111/ddi.13179>