

Exploring the efficiency of Evolutionary Distinctness in conservation prioritization

by
Vahab Pourfaraj

M.Sc. Tehran University, 2006

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Biological Sciences
Faculty of Science

© Vahab Pourfaraj
SIMON FRASER UNIVERSITY
Spring 2017

Copyright in this work rests with the author. Please ensure that any reproduction
or re-use is done in accordance with the relevant national copyright legislation.

Approval

Name: **Vahab Pourfaraj**

Degree: **Master of Science**

Title: Exploring the efficiency of Evolutionary Distinctness in conservation prioritization

Examining Committee: **Chair: David Green**
Professor

Arne Mooers
Senior Supervisor
Professor

Nicholas Dulvy
Supervisor
Professor

Matthew Pennell
External Examiner
Assistant Professor
Department of Zoology & Biodiversity
University of British Columbia

Date Defended: April 19, 2017

Abstract

In this Age of Extinction, we must prioritize the species we want to conserve.

Conservation programs use different metrics for species prioritization, but more work is needed linking these metrics to particular aspects of biodiversity value. Here, I focus on the species-specific conservation metric of Evolutionary Distinctness (ED) designed to identify species with few close relatives. I first explore the relationship between ED and a presumed valuable attribute, the average rarity of traits. Using simulations, I find high degrees of association between ED and trait rarity; however unlike another metric of isolation (Average Pairwise Distance) this ability decreases as higher gamma clades are sampled. I then examine, under different scenarios of extinction, how well ED captures a related touted value, total phylogenetic diversity (PD). I find a very strong correlation between PD and ED across all surveyed trees. Overall, ED is not perfect, but shows some promise as a simple conservation metric, capturing at least two related measures of biodiversity value.

Keywords: Conservation prioritization; evolutionary distinctness; rarity of traits; phylogenetic diversity

Acknowledgements

First of all, I thank Arne for his immense support throughout these years, especially for his intellectual contribution to this research. I have always admired his thirst for science and his incredible level of motivation. I also thank Nick Dulvy for being on my committee. His excellent research skills and his involvement in conservation decision-making made him an excellent committee member for me.

I am very grateful to Carl Schwarz whom I consulted for my statistical analysis and whose office is always open to students who are seeking help. I thank Florent Mazel, Mike Steel, Giulio Valentino Dalla Riva and the FAB* LAB for their valuable intellectual input. I would like to acknowledge Abhishek Chaudhary for posing the question that prompted the third Chapter of this thesis. Gordon Smith also provided me with lots of computational support for running time-labouring simulations on servers. Karen Gordon and Melanie Monroe updated the phylogeny of birds that I used in Chapter 3. I also thank Abdollah Safari, Michelle Jones and the SciProg at SFU who provided coding support for this thesis. And finally, to everyone in the Mooers lab, it was a great experience sharing a workplace during these years. I'll miss you all!

Table of Contents

Approval	ii
Abstract	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	vii
List of Figures	viii
Chapter 1. Introduction.....	1
Chapter 2. Are the most Evolutionarily Distinct species the most original?	4
Abstract	4
Introduction	4
Methods	7
Simulating trees and traits	7
Calculation of isolation scores	9
Calculation of trait rarity	9
Distribution of rarity	10
Empirical data	10
Results	10
Discussion	13
Chapter 3. Is Evolutionary Distinctness a robust proxy for Phylogenetic Diversity?	17
Abstract	17
Introduction	17
Methods	20
Simulated trees	20
Pruning process	21
Real Trees	21
Results	21
Random loss of species	21
Non-random loss of species	24
Discussion	24
Chapter 4. Conclusion	28
Figures and Tables	31
References	53
Appendix A	60
Update to the three vertebrate trees used	60
Mammal trees	60

Bird trees.....	60
Amphibian trees	61
Appendix B	62
Spectral density profile of phylogenetic trees.....	62
Appendix C	69
Annotated R code for simulating traits and calculating rarity	69
Annotated R code for calculating loss of ED and loss of PD	71

List of Tables

Table 1.	Results of the linear model performed on three sizes of imbalanced high gamma trees. The response variable is mean rarity and predictors are tree size, number of states (K), rate of trait change (q) and the interaction between K and q.....	50
Table 2.	Effect of average rarity, number of trait states (K), rate of trait change (q) and the interaction between K and q on the observed variation of rarity among species; obtained from a linear model performed on three size of imbalanced high gamma trees	50
Table 3.	Estimates of slope between loss of PD and loss of ED under random extinction in simulated and empirical phylogenies.....	51
Table 4.	Results of two linear regression models, performed on simulated and empirical phylogenies	51
Table 5.	Results of linear regression model, performed on 4000 simulated trees in order to test the effects of asymetry and peakedness on the slope of PD loss and ED loss.....	52
Table 6.	Estimates of slope between loss of PD and loss of PD under high ED extinction in seven simulated and empirical phylogenies	52

List of Figures

Figure 1.	The concept of originality and uniqueness in functional space. Based on these two traits, species A is both original and unique while B and C are original but not unique.	31
Figure 2.	Two metrics of tree shape used in this study: a) Colless' measure of tree balance (I_c), smaller value represents more balanced tree (left) as apposed to higher values (right) and b) Distribution of nodes ages (γ) with higher values indicating younger speciation events (lef) and lower values representing older speciation events.....	32
Figure 3.	Mean rarity for 500 traits across three size of phylogenies and four rate of character change in case of: a) two-state traits, b) four-state traits c) eight-state traits and d) sixteen-state traits.....	32
Figure 4.	Standard deviation of mean rarity among species. Each point represnts the SD of mean rarity among species of the corresponding tree.	33
Figure 5.	Distribution of average rarity of 500 traits with 8 states on imbalanced high gamma trees: a) rate of 0.008 (Mean rarity=0.7, skewness=-0.85) and b) rate of 0.016 (Mean rarity=0.79, skewness=-1.06).....	33
Figure 6.	Correlation between each isolation metric and average rarity in: a) balanced high gamma trees, b) balanced low gamma trees c) imbalanced high gamma trees and d) imbalanced low gamma trees.	34
Figure 7.	Effect of tree size on the performance of isolation metrics in terms of capturing trait rarity. Top panel: 5 traits simulated on 100 Yule trees of different size and bottom panel: 5 traits simulated on 100 imbalanced high gamma trees of different size.	35
Figure 8.	Correlation between each isolation metric and average rarity in imbalanced high gamma trees of size 64 with 500 traits having: a) Two, b) Four c) Eight and d) Sixteen trait states, at different rates of evolution...	36
Figure 9.	Correlation between each isolation metric and average rarity in imbalanced high gamma trees of size 256 with 500 traits having: a) Two, b) Four c) Eight and d) Sixteen trait states, at different rates of evolution.	37
Figure 10.	Correlation between each isolation metric and average rarity in imbalanced high gamma trees of size 1024 with 500 traits having: a) Two, b) Four c) Eight and d) Sixteen trait states, at different rates of evolution.	38
Figure 11.	Correlation between each isolation metric and average rarity in imbalanced high gamma trees of size 64 with 250 traits evolving at low rate (0.0002) and 250 traits evolving at high rate (0.016) both having 16 character states.	39
Figure 12.	Relationship between trait difference and normalized phylogenetic distance on an imbalanced high gamma tree (n=64) whereby: a) 500 traits with 16 states evolving at the rate of 0.002 and b) 500 traits with 16 states evolving at the rate of 0.016.....	39

Figure 13.	Effect of trait number on the efficiency of isolation metrics in capturing average rarity in: a) imbalanced high gamma tree with 64 species, b) imbalanced high gamma tree with 256 species and c) imbalanced high gamma tree with 1024 species.	40
Figure 14.	Distribution of the two isolation scores and average rarity across 1073 species of bats: a) APD and b) ED	41
Figure 15.	Relationship between Sum of ED loss and PD loss under random extinction in 1000 a) low-death ($d=0.1$) trees b) high-death ($d=0.9$) trees..	41
Figure 16	Relationship between Loss of ED and Loss of PD in case of 15% random extinction in: a) Yule trees, b) Imbalanced trees, c) Low death trees and d) High death trees.	42
Figure 17.	Linear relationship between loss of ED and loss of PD up to 15% species loss under random extinction in: a) 1000 trees of mammals, b) 1000 trees of birds and c) 1000 trees of amphibians.	43
Figure 18.	Effect of the tree size on the goodness of fit regarding the relationship between PD loss and ED loss in: a) 100 Yule trees of size 64 and b) 100 Yule trees of size 1024.	44
Figure 19.	The distribution of variation in pendant edges (standardized by PD of the corresponding tree) in three different sets of phylogenies.....	44
Figure 20.	Values of Colless index in: a) four types of simulated trees and b) three empirical phylogenies	45
Figure 21.	The distribution of Gamma values in: a) four types of simulated trees and b) three empirical phylogenies.....	46
Figure 22.	The scatter plot of slope and two metrics of tree shape; a) Colless index and b) Gamma in simulated trees.	46
Figure 23.	The Scatter plot of slope against two metrics of tree shape; a) Asymmetry and b) log transformed Peakedness.....	47
Figure 24.	Relationship between Sum of ED loss and PD loss in 1000 a) low-death ($d=0.1$) trees b) high-death ($d=0.9$) trees under high ED loss scenario. .	47
Figure 25.	Relationship between Loss of ED and Loss of PD in case of high ED extinction (up to 15% loss) in: a) Yule trees, b) Imbalanced trees, c) Low death trees and d) High death trees.....	48
Figure 26.	Loss of 10 species in 1000 Yule trees of size 64 under: a) random extinction and b) high-ED extinction where I controlled for number of extinct species.	49
Figure 27.	Relationship between degree of imbalance and the sum of normalized external edges in 2000 simulated trees.....	49

Chapter 1.

Introduction

The magnitude of biodiversity loss over past centuries and millennia is comparable to that of past five mass extinction events (Barnosky et al. 2011). The latest estimates for number of threatened species of Vertebrates ranges from 13% to 42% for birds and amphibians respectively (Hoffman et al. 2010).

Given the fact that we are far from having enough funding in order to address the biodiversity loss crisis, prioritization seems inevitable. This means that we are likely only ever going to save a portion of the biodiversity that is at the risk of extinction. Since species is the most commonly studied unit of biodiversity (Agapov et al. 2004) and conservation resource allocation usually occurs at the species level, various species-specific metrics have been suggested for conservation prioritization. One class of those metrics, evolutionary isolation, is broadly considered to capture non-redundant genetic information of species across the tree of life (Redding and Mooers 2006; Faith 2008). These metrics capture the extent to which a species has more or fewer, and more distant or closer, relatives. Some of these metrics have been designed for conservation prioritization and some have been borrowed from other disciplines (Redding et al. 2014). Evolutionary distinctness (ED), however, is the only metric from that pool that is being actively used for species prioritization, by the *Zoological Society of London* in its *Edge of Existence* (EDGE) program. Under that scheme, the EDGE score is calculated as a function of species ED and its risk of extinction (Zoological Society of London 2008). ED itself is the sum of branch length from a focal species to the root of tree, with each branch length divided by the number of taxa it subtends (Redding 2003).

The sum of the ED values across the tips is the total phylogenetic diversity of that tree: the sum of branch lengths of a species subset (Phylogenetic Diversity (Faith 1992)) is a second popular set metric of biodiversity that has been purported to explain difference in ecosystem functioning and ecosystem services (Srivastava et al. 2012; Cadotte 2015). However, PD has yielded mixed results in explaining ecosystem function at different scales (see, e.g. (Genung et al. 2014)). This has made another argument more valuable, that of “option value.” Option value argues for the unknown values and

benefits of biodiversity for future generations (Faith 1992; Faith 2013). From a biological standpoint this can be translated to unique characteristics of a species that can be lost forever due to extinction. Related to this concept is "originality," defined in ecology as the distance from a focal species to the centre of a trait space (Buisson et al. 2013)(see Figure 1). Originality is recommended as an essential criteria in conservation planning (Pavoine et al. 2005), and the rarity of traits (the proportion of species in a set that express a trait value) has been suggested as a proxy for originality (Pavoine et al. 2005). Unique character states are, of course, as rare as they can be. Because studying biological traits is challenging, researchers usually consider a limited subset of traits and use this subset for assessing the functional diversity of an assemblage (see Buisson and Grenouillet 2009; Buisson et al. 2013). Inclusion of functional diversity into conservation programs has been suggested from a utilitarian perspective to ensure ecosystem services (Díaz et al. 2007). It has also been argued that functional diversity will severely suffer from the impacts of ongoing environmental changes (Buisson et al. 2013). The infeasibility of studying large sets of biological traits across taxa impedes the inclusion of trait rarity in conservation programs, such that evolutionary isolation may be a useful proxy. Here, I investigate to what extent we can use evolutionary isolation (and specifically, the ED measure of this isolation) to capture what we think is of conservation value, namely trait rarity.

The main aim of this thesis is to help evaluate whether ED has any essential characteristics for being a preferred measure of prioritization. The rising availability of complete dated phylogenies for a wide range of taxa and the simplicity of calculations has made it very convenient to calculate ED scores. I first employ a combination of simulated tree shapes and sizes and modes of character evolution to investigate if ranking species based on ED scores leads to preserving trait rarity or not. I use different numbers of discrete traits simulated on the phylogenies and calculate average rarity as a measure of originality (Pavoine et al. 2005; Huang et al. 2011). For comparison I also look at the ability of average pairwise distance (APD) in capturing average rarity (Ricotta, 2007); it has been shown that these two isolation metrics are found at two ends of an evolutionary isolation spectrum (Redding et al. 2014) (detailed in chapter 2).

I then ask a second simple question motivated by a query from A. Chaudhary to A. Mooers: how does the loss of ED relate to the loss of phylogenetic diversity (PD)? The motivation here is to see if loss of ED can be used as a proxy to loss of PD across

different types of phylogenies. Unlike ED, PD is a property of a set and its calculation is not always straightforward. Given the justifications for application of PD in conservation planning and prioritization (see Faith, 1992, 2013), testing this link may be important to users of ED (e.g. the ZSL) and to conservation managers looking for a simple-to-calculate conservation metric with flexibility.

Chapter 2.

Are the most Evolutionarily Distinct species the most original?

Abstract

While trait diversity is considered as a crucial component of biodiversity and funding shortages for conservation programs make prioritization likely, metrics being used for species prioritization do not explicitly take trait diversity into account. Here, I consider a species-specific metric of conservation worth, evolutionary distinctness (ED). I use the concept of average trait rarity, which has been suggested as a measure of "originality", whereby original species contribute more to total trait richness, and investigate to what extent prioritizing species based on ED captures trait rarity. Using simulated phylogenies and traits simulated under a simple Markov model of evolution at various rates, I find that mean trait rarity increases with the rate of evolution as well as the number of traits' states. Importantly, while ED is associated with trait rarity in stemmy phylogenies (low gamma), this relationship deteriorates in higher gamma trees. Another common metric of distinctness used in community ecology, average pairwise distance (APD), showed both a higher positive correlation with trait rarity and also seemed fairly robust to tree shape and size. The findings can be seen as a cautionary message for conservation planning: different facets of diversity may not covary and hence conservation programs must explicitly delineate facet(s) of biodiversity they aim to preserve and choose the metric of prioritization accordingly.

Keywords: Evolutionary Distinctness, Average Pairwise Distance, trait rarity.

Introduction

The unprecedented loss of biodiversity and ongoing funding shortages for conservation makes prioritization inevitable (Thuiller et al. 2011; Waldron et al. 2013). To address this challenge, various currencies such as threat status, economical cost and social value have been used to rank species for conservation actions (Weitzman 1998; Joseph et al. 2009). Here, we focus on evolutionary isolation as another currency that is

gaining traction for species prioritization (Isaac et al. 2007). As the name suggests, metrics of evolutionary isolation consider the position of species on phylogenetic trees, where species with fewer close relatives are ranked as more isolated and hence of higher conservation rank. More than a dozen different isolation metrics exist, some created especially with conservation in mind and some adopted from other literatures, with varying degrees of redundancy (Redding et al. 2014).

The Fair Proportion index (Redding 2003) is one commonly-considered conservation-specific metric. It is now generally called evolutionary distinctiveness or evolutionary distinctness (ED) after its adoption by the Edge of Existence conservation program in 2007 (Isaac et al. 2007). ED is calculated as the sum of edge lengths from a focal species to the root of phylogenetic tree, with each edge being divided by the number of taxa it subtends. Taxa with longer branches, and taxa that share fewer close relatives, are more evolutionarily isolated. In addition to the convenience of calculation, ED has been used to help prioritize conservation actions in three classes of animals: *Mammalia*, *Amphibia*, and *Aves*, as well as the *Scleractinia* corals (Zoological Society of London 2008).

Average Pairwise Distance (APD) (Ricotta 2007), calculated as the average patristic distance between a given species and all other species on a phylogeny, is another isolation metric that was adapted for conservation from the community ecology literature (Ricotta 2007; Buisson et al. 2013). In one recent study on North American birds, APD showed a positive correlation with its spatially restricted version (i.e. high correlation of APD calculated from a global tree correlated with APD calculated from a community tree) and some ability to capture measures of trait uniqueness (Redding et al. 2015). Interestingly, APD at the community level is the tree-based counterpart of functional originality, defined as the proximity of species to the center in multidimensional functional space derived from a set of measured traits (Buisson et al., 2013). The further a taxon is away from the center of multidimensional functional space, the more original it is in terms of surveyed features (see Figure 1). Its counterpart, ecological "uniqueness" is the distance to the nearest neighbour in functional space (Buisson et al. 2013). On an ultrametric phylogeny, this would be 2*the Pendent Edge (PE) of a focal species: PE and ED scores are highly correlated across many tree shapes (Redding et al. 2008; Redding and Mooers 2015). While PE is too simplistic to be considered an applicable evolutionary isolation metric, APD and ED were shown to

be opposite points of an external-internal branch range: APD measures the phylogenetic information that is deeper on a tree while ED gives more weight to information nearer to the tips (Redding et al. 2014). Although an increasing number of studies have begun to evaluate the effects of change in functional diversity at the ecosystem level (see, e.g., Thuiller et al. 2006; Buisson and Grenouillet 2009), we know of only one study that has compared the ability of different isolation metrics to capture functional diversity for conservation (Redding et al. 2015).

Researchers have been trying to link functional diversity and ecosystem functions and provide stronger justifications to extend criteria for conservation prioritization such that conservation procedures capture functional diversity as well (Cadotte et al. 2011). In parallel, the importance of saving rare features has been emphasized more explicitly through the concept of "option value" (Faith 1992; Faith 2008). Since rare features are not possessed by many (or, at the extreme, any) other species and extinction is forever, one rationale for saving rare features originates from uncertainty of the need for those features in the future. An often-cited example of such rare trait is found in the gastric brooding frog *Rheobatrachus silus*. This species had a specific mechanism to inhibit acid secretion in the female's stomach where the young frogs develop (Tyler et al. 1983). Since the species is now extinct (Meyer et al. 2004), we have missed the opportunity of exploring that feature in more detail as well as any potential benefit that could have been yielded from it.

More formally, the "average rarity of characters" has been suggested as a measure of "originality," whereby original species contribute more to total feature richness (Pavoine et al. 2005; Huang et al. 2011). Character rarity is the probability that a focal taxon shares a character state with other species of a set (here a phylogenetic tree). Expected character rarity varies from 0 (where all taxa are expected to share the same character state) to 1 (where no other taxa are expected to share the observed character state). We use this concept of trait rarity to investigate to what extent prioritizing species based on isolation metrics will concomitantly save rare features.

It has been shown that isolation scores in trees smaller than 250-300 species are not absolute and need standardization (e.g. by the depth of the tree). However after this threshold most surveyed isolation metrics reach an asymptote and hence are comparable among taxonomic groups (Redding et al. 2014). Given the effect of tree

size, I designed my study to involve tree sizes below, at, and above that threshold. Intuitively, tree shape can also impact the relationship between isolation metrics and trait rarity and supposedly a metric that is robust to tree shape and hence can be used effectively in wide range of phylogenies is likely to be preferred. In order test this property, I employed a range of tree shapes in terms of their balance and average node age in my study.

More concretely, I investigate the relationship between two isolation metrics and the rarity of evolving traits under very simple process models. My main analyses employ simulated phylogenies and simulated character data under simple Markov models. I ask three main questions:

- 1) Do APD and ED capture rarity differently? How do tree size, number of character state and rate of character change affect the relationship between each metric and rarity?
- 2) Which of the two isolation metrics performs best at capturing rarity in the class of trees where APD and ED show highest divergence?
- 3) How do these two metrics perform in real life situations?

Methods

Simulating trees and traits

In order to get a range of tree shapes, I first simulated 3 sets of 5,000 pure-birth (Yule) trees using the function `pbtrees` in R package *phytools* (Revell 2012). The depth of all trees was set to 50 and the three sets of phylogenies had 64, 256 and 1024 species. Gamma for each tree was calculated using the `gammaStat` function in R package *APE* (Paradis et al. 2004) while the tree balance value was obtained by means of `Colless` function in *apTreeshape* (Bortolussi et al. 2006). The values of I_c were standardized to the Yule model in order to allow comparison among trees with different numbers of tips (Blum et al. 2006).

I then randomly chose 1000 trees with I_c values smaller than 0 and 1000 trees with I_c values greater than 0. I performed the same procedure for each of the three tree

sizes, resulting in 6000 Yule trees (2000 trees of each size) with overall I_c values from -1.99-3.69. I then applied the delta transformation (Pagel 1999) (using the rescale function in *geiger* package (Pennell et al. 2014) to produce a wide range of gamma values on both balanced (trees with I_c below 0) and imbalanced (trees with I_c above 0) set of trees. The range of gamma values for the transformed trees was -6.33 to 12.30. This produced 4000 trees of each size, 1000 in each of the four combinations of low/high gamma and low/high I_c . Hereon, I refer to trees with gamma values greater than 0 and trees with gamma values less than 0 "high" and "low" gamma trees respectively (see Figure 2).

For traits, I used the Markov model of trait evolution (Mk) to simulate 500 discrete traits on each simulated tree using the *sim.history* function in *phytools* (Lewis 2001; Revell 2012), with k (number of states) set to 2, 4, 8 or 16 states. In order to test the effect of number traits on the ability of APD and ED in capturing average rarity, I also included simulation of 5 and 50 traits. Based on the Markov model the probability of transitions can be calculated as follows:

$$P_{ii}(t) = \frac{1}{k} + \frac{k-1}{k} e^{-kqt}$$

$$P_{ij}(t) = \frac{1}{k} - \frac{1}{k} e^{-kqt}$$

Here, t is the elapsed time and q is the instantaneous rate of character change (Lewis 2001). In order to set q , the expected distance between two randomly selected tips under the Yule model was calculated (Steel and McKenzie 2001) and four rates of character change (0.002, 0.004, 0.008 & 0.016) were set using the above equations so that it includes slow and fast evolving characters while avoiding full saturation. In order to make sure those rates would result in reasonable probability of change and my simulation involves slow and fast evolving traits, I measured the average pairwise patristic distance across my trees. Low gamma trees ought to exhibit higher patristic distances on average for a given tree depth. The average pairwise patristic distance in balanced low gamma trees with 64 tips was ~ 90. For a binary trait this would mean a chance of change equal to 0.15, 0.26, 0.38 and 0.47 for rates of 0.002, 0.004, 0.008 and 0.016 respectively between two randomly chosen tips.

Calculation of isolation scores

Evolutionary Distinctiveness. I calculated the Fair Proportion index (Redding 2003) (hereafter ED) as a globally well-known measure of evolutionary isolation using *caper* (Orme et al. 2013) package of R. For each species on a tree, ED is the weighted sum from the root to the tip of phylogeny (Redding 2003):

$$ED(T, i) = \sum_{e \in (T, i, r)}^r \frac{\lambda_e}{c_e}$$

Where i is the species of interest, λ_e is the edge length, which is being divided by number of species (c_e) it subtends.

Average Pairwise Distance. This metric of isolation quantifies the mean pairwise distance between a focal species and all other species on a phylogeny (Ricotta 2007). I calculated this metric using *caper* (Orme et al. 2013) package of R.

$$APD(T, i) = \frac{1}{N-1} \left[\sum_{\substack{i \in f(T, x) \\ j \in g(T, x - \{i\})}} d_{ij} \right]$$

Where i is the species of interest, x is the set of species, N is the size of the set (i.e. $|x|$) and d_{ij} is the patristic distance between focal species and other species on the tree (see Redding et al. 2014).

Calculation of trait rarity

The rarity of a trait for a species is calculated as:

$$R = 1 - \sum_{i=1}^{s-1} \frac{p_i}{(s-1)}$$

Where s is number of species and p_i is the presence (1) or absence (0) of the character state in other species of a phylogeny compared to the target species (Huang et al. 2011). After simulating 500 traits on each tree for all combinations of rate of

character change (four rates) and number of traits' states (four levels), I calculated the average rarity across the 500 traits to produce a mean rarity value for each species in each phylogenetic tree for each combination of rate and number of states.

Distribution of rarity

After estimating the average rarity of simulated traits for each species, I calculated the mean of those values for each single tree and extended it to all 12000 trees of this study. This gives us an insight into the behaviour of average rarity regarding changes in tree size, rate of character change and the number of character change.

Empirical data

In order to evaluate the behaviour of ED and APD in real life situations, I employed the Elton traits dataset for mammals (Wilman et al. 2014). I chose the Chiroptera clade (1073 species) as a test case since they represent a diverse order of mammals especially in regards to diet and activity-time. I calculated rarity for 15 discrete traits for each species, including foraging and activity-time data. I converted the continuous body size trait to a discrete trait by log-transforming body mass and assigning each one unit interval to a different state, which resulted in nine states for log of body mass. The two isolation metrics were then calculated for each bat species using a recent supertree of mammals (Martyn et al. 2012). Appendix A provides more detail on how the employed empirical trees were updated.

Results

Distribution of rarity. In order to investigate the effect of number of character states and rate of character change on mean rarity, I calculated mean rarity value for all species in each tree across four rates of character change and four levels of character state. In all tree sizes there is a rising pattern in mean rarity values when the rate of character change increases. Also, for any given rate of character change, increasing the number of character states yields a broader range of mean rarity values (Figure 3). Those findings are also reflected in the results of the linear model where K and q and trees size all positively affect mean rarity and there is also an interaction between K and q (Table 1).

Figure 4 shows the impact of tree shape on observed variation of rarity values among species of each tree. Trees with high gamma contained species that are more different in terms of their rarity values. In other words, trees with more recent node distribution are expected to have species with higher variation in trait rarity scores compare to phylogenies with deeper node distribution. In low gamma trees I observed lower variation in rarity values, although if controlled for the rate and number of trait state, those trees contain species with higher average rarity. While at first glance the impact of tree balance on variation in rarity might be overlooked, the results of linear models performed at a given tree size and a specific rate of change and K suggest both metrics of tree shape are significant predictors of variation in rarity. The impact of rate of change on the mean rarity and skewness is plotted in Figure 5 where higher rates of character changes results in larger mean rarity values as well as higher skewness.

As depicted in Table 2 when I controlled for tree shape, the variation in average rarity scores could be explained by average rarity, number of trait states, rate of trait change and their interactions. As expected, increases in average rarity, number of trait states and rate of trait change result in more variation in rarity scores.

Correlations of rarity with isolation metrics. Figure 6 shows the correlation between the two isolation metrics (ED and APD) and average rarity across the four tree-shape classes and three sizes of phylogenies. Increasing tree size doesn't substantially change the ability of APD to capture average rarity. In contrast, ED's ability to capture rarity deteriorates as trees gets larger. Larger trees of this study happened to have higher gamma values (median gamma increased from 6.07 in imbalanced high gamma trees with 64 species to 8.07 in trees with 1024 species). In order to test if gamma is the main driver of this pattern in large trees I simulated Yule trees of different size and compared with imbalanced high gamma trees. As depicted in Figure 7, when gamma is kept constant (ie. in the Yule trees), different sizes of phylogenies exhibit similar patterns for the performance of ED and APD in terms of capturing trait rarity. Consistently, in high gamma trees where the pendant edges are relatively short, APD captures average rarity notably better than ED, while in low gamma trees the two isolation metrics perform fairly similar.

Real world phylogenies are generally imbalanced (Blum and François 2006) and as one of the main objectives of this research was to explore the application of these

metrics in real-life situations, hereon I focus my attention primarily on the performance of each metric at different rates and numbers of trait states in imbalanced trees. Also, since I found higher degrees of divergence in high gamma trees in terms of isolation metrics performance I will be focusing on imbalanced high gamma trees specifically.

Figure 8 illustrates the ability of the two isolation metrics in capturing average rarity at all sixteen combinations of rate of change and number of traits state evolved on the smallest employed trees ($n=64$). APD is shown to be more highly correlated with average rarity in 13 out of 16 combinations. However, this correlation drops as trait dimensionality and rate increases, while ED seems less affected; when the trait dimensionality was at the maximum, ED performed better than APD at capturing average rarity. A similar pattern was observed in bigger phylogenies (Figure 9 and Figure 10) and the two isolation metrics diverged increasingly with increased size of phylogenies. In order to explore plausible explanations for better performance of ED at high rates of trait evolution, I performed two tests: first I checked if mixed rates of trait evolution (i.e. slow and fast rates) result in an intermediate performance of the two isolation metrics (Figure 11) and if there is any evidence of saturation in high rates of trait change (Figure 12). As shown in Figure 12, there is evidence of saturation at high rates of changes and high trait dimensionality, where, beyond a certain phylogenetic distance, more distantly related species are not more divergent in terms of simulated traits.

I also showed that increasing the number of traits has a positive effect on both metrics' performance (Figure 13). The mean correlation between isolation metrics and average rarity is 0.58 and 0.52 for APD and ED respectively when 5 traits are considered on the smallest trees. This correlation increases to 0.88 and 0.78 when 500 traits are considered on the same tree size. Increasing tree size weakens the ability of both metrics so that in the case of the biggest tree size ($n=1024$) with 5 traits the correlation drops to 0.44 and 0.12 for APD and ED respectively, but once again it improves when number of simulated traits is 500 (0.86 and 0.25).

Empirical data. The two isolation metrics showed different degrees of association with average rarity in 1073 species of bats, with a correlation of 0.36 with APD and only 0.02 for ED (Figure 14). The Pel's pouched bat (*Saccolaimus peli*) was ranked first based on APD scores while Madagascar sucker-footed bat (*Myzopoda aurita*) had the

highest score of median ED; interestingly both species were ranked very low in terms of average rarity of studied traits (0.11 and 0.15). Sister species: Straw-colored fruit bat (*Eidolon helvum*) and Madagascan fruit bat (*Eidolon dupreanum*) have the highest score (0.35) for average rarity (e.g. they are most original bats) and APD gave a higher rank to these species (241st and 242nd ; same APD score) compared to ED (566th and 567th; same ED scores).

Discussion

The fact that evolutionary history has not been extensively applied to conservation has been attributed to the lack of empirical justifications (Winter et al. 2013) and also confusion around selecting the proper metric from the ‘Jungle’ of indices (Winter et al. 2013; Tucker et al. 2016). It should be noted that metrics of functional diversity haven’t been used for conservation prioritization either, with one barrier being the subjectivity of those measures, which makes it difficult to decide about species value (see Srivastava and Vellend 2005; Mouquet et al. 2012). In parallel, an increasing amount of research is being directed towards studying loss of functional traits as a crucial facet of biodiversity that is at stake in a rapidly changing environment (McGill et al. 2006; Buisson et al. 2013). In functional space, two metrics of diversity have been suggested: uniqueness, which is defined as the distance to the nearest neighbour as apposed to functional redundancy (Fonseca and Ganade 2001; Buisson et al. 2013) and originality, which represents the average rarity of species traits and is measured as the average distance to the center of functional space (Pavoine et al. 2005; Mouillot et al. 2008; Buisson et al. 2013). The explicit inclusion of functional diversity in conservation has been suggested specifically as a response to deal with a changing environment (Dalerum 2013; Thuiller et al. 2014). Conservation of uncommon biological traits has been also advocated strongly from an ecosystem functioning perspective (Purvis et al. 2000; Mouillot et al. 2008; Mouillot et al. 2013) . Regardless of the suggested justifications for including functional diversity in conservation planning, this facet of biodiversity is not yet been considered explicitly in conservation. To help address this gap, I investigated if ED, as a simple and species-based metric of evolutionary isolation, captures this facet of biodiversity, and whether its application should be encouraged due to it capturing this particular value.

I employed the concept of average rarity, applied it to a discrete trait framework, and explored drivers of average rarity. All three predictors (rate of character change, number of character states and tree size) significantly affect mean rarity. Tree size however showed a very small effect size compare to the other two predictors of mean rarity. From a conservation standpoint this can be interpreted as expecting higher average rarity in cases where traits evolve rapidly and where they can take more states. Mean rarity itself turned out as a powerful predictor of variation in rarity, besides the other two predictors (rate of character evolution and number of character states). Those analyses were insightful as they provided a better understanding of factors that are in play for determining a species' average rarity value and suggests that in cases where assemblages represent high average rarity, we expect to find more variation in the rarity of traits as well. This makes the choice of which species we preserve more important.

I also showed that the ability of ED to capture trait rarity is highly impacted by properties of the phylogeny. While ED's performance in capturing average rarity in small trees ($n=64$) is encouraging and comparable to that of APD, its performance deteriorated with increasing size of the phylogeny in my dataset. This pattern was more profound in high gamma trees. Unfortunately, the effect of tree size is confounded with gamma in my dataset. In other words, as the employed phylogenies get larger in size they tend to show higher gamma values (i.e. the pendant edges become relatively short). Under the Markov model I used, this would lead to higher probability of change on internal vs. pendant edges. While APD is an effective metric in capturing changes in internal edges, ED is highly correlated with the pendant edge (Redding et al. 2014). So, in cases where pendant edges are short and are less informative, ED performs poorly.

The robust performance of APD might be further explained by the fact that it is an average measure of distance between species, as is the measure of rarity used here. I also simulated large number of traits under Markov model of trait evolution: this set of characters could be used to recover the original tree under Maximum Parsimony: this would lead us to predict that the full tree measure APD might perform better than ED on average. The modest performance of both isolation metrics in low gamma trees on the other hand, may be explained by the distribution of rarity values in different tree shapes. As illustrated in Figure 4, lower gamma trees embody species with more similar rarity scores. Lower variation in trait rarity in low gamma trees could be explained by the fact that in those trees where pendant edges are relatively long, species can embody

different character states depending on trait dimensionality. Therefore (and particularly at high trait dimensionality) we will usually find species with rarity values close to maximum and so lower variation overall, compared to high gamma trees where we find rare as well as unrare species.

I also identified examples where ED performed better than APD. When I controlled for tree shape and looked at performances of the two metrics at different rates of character change, then on small trees and high rate of change and many state ($K=8$ and $K=16$), ED was more highly correlated with average rarity than was APD. This could be due to getting some degrees of saturation at these high rates of evolution coupled with many autapomorphies on the pendant edges. To test this idea, I replicated the simulation process that was done to create Figure 8d with the only difference being that I had half of traits evolving at the slowest and half of them at highest rate. Consistent with what I hypothesized, I found that the two isolation metrics behave very similarly in this case. I also showed that there is evidence of saturation in my data when traits evolve at the highest applied rate and the trait dimensionality is also at the maximum. The fact that ED is a tippy metric and is highly correlated with pendant edge (Redding et al. 2014), makes it more effective in cases of multiple hits whereby the most recent changes occur likely on the pendent edge; APD that can not capture information nearer the tips as efficiently. In other words, in cases of saturation, information near the tips is more informative and intuitively the metric that is more weighted towards the tips (i.e. ED) is expected to perform better. Measures of phylogenetic distance in general tend to underestimate trait rarity when saturation fades the phylogenetic signal (Pavoine et al. 2005; see also Wake et al. 2011).

APD shows very high degree of correlation (close to one) with average rarity when high numbers of traits are simulated. Another study also reported good performance of APD in terms of capturing PD at local scale as well as some metrics of trait uniqueness (Redding et al. 2015). Whether these qualities of APD can be explained by its correlation with the tree shape at different spatial scales or its ability in capturing deep time, or whether the performance is simply an artefact of surveyed traits and their model of evolution needs further investigation.

In the case of the empirical dataset, lists of high ED bats did not conform to that of bats with rare traits at all. In contrast, APD exhibited some degrees of correlation with

rarity. Similar results were found in a study of Nearctic and Neotropical birds where unlike ED, APD showed degrees of association with a community metric of trait diversity (Redding et al. 2015). This might imply that APD on average is a better representative of trait rarity. However more scrutiny is needed in order to advocate for its application in species prioritization.

While there are many justifications for saving isolated species (Redding et al. 2010; Jetz et al. 2014) and conservation dollars are spent to save evolutionary distinct species (Zoological Society of London 2008) it is not yet clear which facet(s) of biodiversity is being preserved. Here, I showed that ED captures trait rarity on low gamma trees and also in case of fast evolving traits in simulation. The efficiency of ED is impacted by tree shape and underlying properties of trait evolution (e.g. doing better at higher rate and higher numbers of trait state). The inconsistent performance of ED in terms of capturing trait rarity presented here should be taken into account by conservation planners. ED might effectively capture trait rarity in some (i.e. if the target clade is mostly consisted of old lineages (i.e. low gamma) or if the functional traits of interest saturate) but not all, real life situations. Aside from some unappealing properties of APD (i.e. double counts), there is evidence for its efficiency in terms capturing some facets of biodiversity (Redding et al. 2015 and this chapter). If future studies confirm its promising performance, then conservation managers should consider its application in conservation prioritization.

Those findings can be seen as a cautionary message for conservation planning: different facets of diversity may not covary. This in turn emphasizes that different currencies of prioritization should be selected based on the explicit aims of any given conservation program. I take up more specific suggestions and ideas for future research following on from my MSc. work in Chapter 4.

Chapter 3.

Is Evolutionary Distinctness a robust proxy for Phylogenetic Diversity?

Abstract

A rising number of studies are recommending the inclusion of phylogenetic information in conservation planning. The oft-cited Phylogenetic diversity (PD) metric is problematic for direct inclusion into species based conservation planning. In contrast, Evolutionary Distinctness (ED), which is actively being used for species prioritization, has been criticized for not taking complementarity into account. Here, I show a strong correlation between loss of PD and loss of ED under two scenarios of extinction across different sizes and shapes of phylogenies. I also report a robust performance of ED in capturing phylogenetic diversity when I controlled for number of dropped species. My results suggest that ranking species based on ED safeguards PD very well. This may offer a strong argument for focusing species-specific conservation efforts on highly ED species.

Keywords: Evolutionary Distinctness, Phylogenetic diversity, Extinction, species prioritization

Introduction

The criteria for species prioritization in conservation biology range from cultural preferences to extinction risk (Metrick and Weitzman 1996; Possingham et al. 2002). Phylogenetic diversity (PD) – the sum of the lengths of branches in a phylogeny connecting species of a set – has been presented as a framework for preserving sets of species and areas for conservation (Faith 1992), justified by the argument that subsets of species that represent more phylogeny are likely to embody more evolved features, something that may impart value from community to ecosystem levels (see, e.g. Cadotte et al. 2008; Faith 2016).

Evolutionary isolation may be considered a general term for species-specific measures of a species' contribution of branch length to a phylogenetic tree. Many exist (see Redding et al. 2008; Vellend et al. 2011; Redding et al. 2014), with the most commonly used being Fair Proportion (Redding 2003), which, called Evolutionary Distinctiveness or ED, serves as the basis for the Zoological Society of London's Edge of Existence (EDGE) conservation program (Zoological Society of London 2008). Indeed, ED scores are now available globally for mammals (Isaac et al. 2007), amphibians (Isaac et al. 2012), birds (Jetz et al., 2014), squamate reptiles (Tonini et al. 2016), Scleractinian corals (Huang 2012), Sharks and Rays (Stein et al., pers. comm.) and some smaller groups (see, e.g., Redding et al. 2010; Bennett et al. 2014). Given the convenience of getting and using ED scores, researchers have begun to use the sum of ED scores as a surrogate for PD based on the premise that these two should be closely related (see, e.g., Dalerum 2013). Indeed, it has been shown that sets of species with high ED scores produce a high PD phylogeny on small simulated trees (Redding et al. 2008), and ranking on ED can produce a near-optimal approach to conserving threatened PD (for birds, see Jetz et al., 2014). However, it is easy to produce cartoon trees where high ED species do not contribute highly to PD because ED scores fail to completely incorporate "complementarity" (Faith 2016). Such cases can be imagined when two subsets of species represent similar ED but embody different PD values due to different relatedness on the phylogeny.

In this chapter, I investigate the strength and the shape of relationship between sum of ED loss and PD loss for different types (i.e. shapes) of simulated trees and several real phylogenies under different scenarios of extinction. This work complements recent analytical work predicting the relationship between sum of ED loss and PD loss in the case of birth-death trees (Steel et al., pers. comm.). Both studies stem from a query about the form of this relationship from A. Chaudhary to A. Mooers.

In order to investigate what factors might be contributing to the shape and the strength of relationship between ED loss and PD loss, I measured two typical measures of tree shape: Pybus' gamma, γ (Pybus and Harvey 2000) and Colless' index of imbalance, I_c (Colless 1982; Mooers and Heard 1997) and tested how they contribute to the above mentioned relationship. The γ is an indicator for timing of branching events on a phylogeny with $\gamma > 0$ denoting more recent and $\gamma < 0$ indicating deeper speciation events than expected on a equal-rates birth tree (Pybus and Harvey 2000). Colless'

index on the other hand, measures variation in clade size throughout the tree, and is calculated as the sum of differences in number of species between sister clades at each internal node (Colless 1982; Mooers and Heard 1997). Greater Colless index is found on phylogenetic trees that are more imbalanced (Blum et al. 2006). I also considered two new metrics of tree shape introduced by Lewitus and Morlon (2016a, b) based on spectral density profile: "skewness" and "kurtosis". Skewness is a measure of the asymmetry of a tree's spectral density profile, and is a metric of relative timing of branching events from the root to the present. Positive skewness is an indicator of branching events being recent and negative skewness denotes deep speciation events. Kurtosis quantifies the peakedness of trees' spectral density profile where small values of this measure is interpreted as branch length being homogeneous, which in turn can be inferred as the tree being balanced (Lewitus and Morlon 2016a; Lewitus and Morlon 2016b). However, skewness and kurtosis were shown not to be tightly correlated with I_c and γ , respectively (Lewitus and Morlon 2016b). I therefore tested whether variation in the relationship linking ED loss to PD loss is captured by the variation in any of these four features of tree shape.

Evolutionarily isolated species are shown to embody unique traits in some taxa (Magnuson-Ford et al. 2009; Redding et al. 2010). It has also been suggested that sets of evolutionarily distinct species may capture more total phylogenetic diversity (Redding et al. 2008). Those results suggest that isolated species may represent some special value of biodiversity. Therefore, I included both simulations where species loss is random with respect to ED, and simulations where extinctions events occur based on the ED ranking of the species, with high ED species going extinct first. This latter approach helps assess the efficiency of ED as a surrogate of PD under scenarios of extinction where high ED species are more likely to be threatened.

Given the challenges of applying PD into species-specific conservation programs, if on average the relationship between ED loss and PD loss is found to be tight, and robust across different phylogenies, then conservation planners can benefit from the convenience of calculating summed ED scores and using this as a proxy for PD. A strong link between ED loss and PD loss would also support the use of ED in the ZSL's edge of existence program.

Methods

Simulated trees

I simulated three sets of 1000 trees with $n=64$ tips using the function *pbtree* in R package *phytools* (Revell 2012) representing Yule, low and high extinction regimes ($d=0$, $d=0.1$ and $d=0.9$). I also simulate pure birth trees with evolving rates (Beaulieu and O'Meara 2015) and selected trees with normalized Colless index greater than zero so that my simulated trees covered a wider range of balances. For each tree I calculated the four metrics of tree shape. Using *gammaStat* function in *ape* package (Emmanuel Paradis et al. 2004) and *colless* function in R package *apTreeshape* (Bortolussi et al. 2006) I calculated γ and normalized I_c for each tree respectively. γ value greater than zero indicates recent speciation while negative value of γ is an indicator of old speciation (long pendant edges). Normalized I_c on the other hand, can vary between $-\infty$ and $+\infty$ with higher positive values indicating more heterogeneity in diversification rates among clades. In order to explore the phylogenetic space in more scrutiny and to be able to better capture the variation in tree shape among phylogenies I also measured two metrics of the spectral density profile of each tree using the *RPANDA* package of R (Morlon et al. 2016). Using the *spectR* function I computed the skewness and kurtosis of the spectral density profile for each phylogeny. Skewness (aka asymmetry) is another measure of the node distribution from the root to the tips in a tree. Positive skewness value is an indicator of recent branching events (stemmy trees) while negative skewness is an indicative of deep branching events (tippy trees) (Lewitus and Morlon 2016). These authors reported weak relationships between skewness and gamma ($R^2=0.19$) which they interpreted as suggesting that their proposed metric measured nodes distribution differently from gamma (Lewitus and Morlon 2016a). Lastly, I measured the kurtosis for each tree, which quantifies homogeneity/heterogeneity of branch lengths. Low kurtosis value is a sign of homogeneity in branch lengths, which in turn indicates a balanced tree shape, while high values represent imbalanced trees where there is higher degree of heterogeneity in branch lengths. Conceptually, The kurtosis value of a phylogeny should overlap with that of Colless index; however these two measures of imbalance are reported to be uncorrelated (Lewitus and Morlon 2016a). More details on the spectral density profiles of my simulated and empirical phylogenies can be found in Appendix B, where I conclude that the behaviour of these measures are not yet well established.

Pruning process

I randomly pruned species iteratively from each tree using the *drop.tip* function in the R package *ape* (Paradis et al., 2004), from one up to 64 species, recording the summed ED loss and the attendant PD loss from that tree. A similar procedure was followed when I performed high ED extinction with the only difference being that species were dropped based on their ED scores starting from the highest to the lowest. All loss values were divided by the total length of the corresponding full tree, so that all loss values range between 0 and 1. The form of the relationship over the first 15% (10 species) lost (simulating the current proportion of species at risk of extinction in major groups (Hoffmann et al. 2010)) and over the 100% (64 species) lost were tested separately using *lm* function in stats package of R (R Core Team 2016).

Real Trees

In order to extend the results to real world situations, I also considered updated sets of 1000 trees of birds (Jetz et al. 2014; BirdLife International 2015), $n = 10284$ species, of mammals (Martyn et al. 2012), $n = 5139$ species; and amphibians (Isaac et al. 2012), $n=5713$ species. Details of the updates for each phylogeny set can be found in Appendix B. I randomly dropped 1543, 771 and 857 species (15%) from the birds, mammals and amphibians trees respectively. The exact same protocol of pruning was applied to real trees in order to generate the slope of relationship between ED loss and PD loss for each taxon over 15% of random extinction. I then repeated this choosing the top 15% ED species for pruning. All four metrics of tree shape were also calculated for empirical phylogenies.

Results

Random loss of species

The shape of relationship between ED loss and PD loss (from 0 to 100% loss) under random scenario of extinction in birth-death trees is presented in Figure 15. The relative death rate of simulated trees has a significant effect on the shape of the curve. In trees where on average terminal edges are longer than internal edges, relatively high amounts of PD is lost initially. In contrast, in high death trees short terminal edges cause

the PD loss to be low until a certain tipping point. As can be seen from Figure 15, the expected curve of such relationship can be well predicted using the equation from Steel (Steel et al., pers. comm.).

In order to make the simulations more comparable to real life situations, I turn my focus to the relationship between ED loss and PD loss when only a moderate proportion of species (15%) are removed from each phylogeny. For each empirical and simulated phylogeny, I generated the slope of ED loss and PD loss. As indicated by r^2 values in Table 3, loss of cumulative ED and loss of PD under random extinction are very closely related to each other in all surveyed phylogenies, and specifically in the empirical phylogenies. Estimates of slope for Yule and low-death trees are very similar (~ 0.5) where on average each unit of ED loss corresponds with about 0.5 unit of PD loss. This finding is more intuitive when we consider the fact that Yule trees are expected to have equal internal and pendant edge lengths and they are also expected to exhibit an intermediate degree of balance. Under minor random extinction, individual pendant edges are most commonly lost, and total pendant edge length is roughly half the entire PD of a tree (Mooers et al. 2012).

High- I_c and high-death trees had the highest and lowest estimated slopes among all surveyed phylogenies respectively. The shallow slope of high-death (high gamma) trees is not unexpected given that other research has shown negligible amount of loss of PD as species are lost on high gamma (coalescent) trees (Nee and May 1997). Among empirical phylogenies, birds showed the steepest, and amphibians showed the shallowest, slopes (Table 3). The phylogeny of birds also exhibited the lowest amount of variation along the best fit line (standard error of the slope=0.0002 as apposed to 0.0007 for mammals and amphibians). A similar improvement in goodness of fit resulted when I compared Yule tree of different size (64 vs. 1024) where the r^2 increased from 0.85 to 0.98 (Figure 18). This can be partly due to the fact that external edges in larger trees are more similar in length compare to smaller trees. This difference is reflected in Figure 19 where external edges show very negligible amount of variation in the super tree of birds compare to those of smaller trees.

In order to explore drivers of observed variation in slope of different phylogenies, I considered the four metrics of tree shape. I first considered Colless' I_c , which quantifies the degree of variation in diversification rate. Three out of four types of simulated trees

had median Colless values centered on zero (no variation) while the median Colless index for high- I_c trees was 1.51. Empirical phylogenies occupied a much wider range, with birds showing the highest (Median = 12.03) and mammals showing the lowest (Median = 6.39) values (Figure 20).

The result of gamma statistics distribution is shown in Figure 21 where high- I_c trees ($I_c > 0$) were the only group of simulated trees with negative gamma values (Median = -1.03). Low death and Yule trees show almost the same values of median gamma (not significantly different from zero), and high death trees exhibit the highest values of gamma (Median gamma = 4.86). Among empirical phylogenies, amphibians displayed the highest (Median = 27.25) and mammals showed the lowest (Median = 5.33) gamma values: amphibians have the most relatively recent distribution of nodes and mammals the deepest relative distribution of nodes among empirical phylogenies.

As shown in Figure 22, there seems to be roughly linear relationships between the PD-ED slope and each of these two metrics of tree shape. I therefore fit a linear model in order to estimate the effect size of gamma and Colless index for explaining variation in slope. The outputs from the linear model performed on 4000 simulated trees showed that Colless index positively affects the slope, meaning that in phylogenies where there is high variation in diversification rates among clades, higher estimates of slope are expected. Gamma, on the other hand inversely affects the slope: trees with smaller values of gamma (so a, deeper distribution of nodes) are expected to lose higher amounts of PD on average when one unit of ED is lost. Importantly, the effect size of gamma is more than twofold bigger than that of I_c . The same linear model was performed on the 3000 empirical phylogenies. The general inferences are the same, with the model fitting the empirical trees better (Table 4).

I also considered the new measures of tree shape (Figure 23), while these scatterplots are messier, I tested a linear model with skewness and kurtosis as predictors in order to check whether these two metrics of tree shape can better predict the slope of PD loss and ED loss relationship. The goodness of fits for the models using the two sets of measures are very similar, indicating that the two newly-introduced metrics of tree shape (Lewitus and Morlon 2016a) are as powerful, but no more powerful, as classic measures of tree shape in describing the variation in slope across different phylogenies.

Non-random loss of species

Figure 24 depicts the shape of relationship between ED loss and PD loss (from 0 to 100% loss) in birth-death trees where species are dropped based on the descending rank of their ED values. Overall and importantly, the fit is much improved and the slopes are nearer unity. Because of this, and unlike the random scenario of extinction, here the relative death rate of simulated trees does not have a strong effect on the slope (0.94 vs. 0.91 in low death and high death trees respectively). Given that the initial slope is near unity, the quadratic relationship between the two variables under the random extinction has become a nearly linear relationship

Similar to random scenario of extinction, I focused on 15% extinction here. Figure 25 shows the relationship between loss of ED and loss of PD in the four types of simulated trees where top 15% high ED species are dropped. The slopes are much higher in this case comparing to that of random extinction scenario for all surveyed phylogenies. Among simulated trees, imbalanced trees showed the steepest slope, as expected. Due to the shape of these phylogenies, loss of highly isolated species, which occur in the species-poor areas of the tree will produce this steep slope. Unlike random extinctions, empirical phylogenies exhibit slopes similar to simulated trees (i.e. near unity). My results show that in this type of non-random extinction, tree shape and size seem not to be crucial determinants of how much phylogenetic diversity is lost per each unit of ED loss.

Discussion

Following the introduction of phylogenetic diversity (PD) by Faith in his landmark paper (Faith 1992), subsequent research has focused on what is effectively being conserved by preserving PD (Faith 2013). Preserving feature diversity is one of the arguments for PD-based conservation approaches (Faith 1992). Some researches have been able to confirm that prioritizing conservation efforts based on PD translate into better-than-random capture of feature diversity and subsequently option-value (see, e.g., Forest et al. 2007).

Apart from arguments for and against PD as a measure of feature or trait diversity, a central barrier for application of PD-based metrics in conservation arises

from the fact that PD is a set-specific measure and conservation planners may find it challenging to apply such metrics for species conservation. In other words, although many conservation plans are designed at species level there is been no straightforward way of calculating the expected PD contributed by a single species to some future or hypothesized set of species. This shortfall has encouraged development of species-specific isolation metrics (see Redding et al. 2014). These metrics differentially quantify the degree of isolation of each tip on a phylogenetic tree (Redding et al. 2014). Among several existing isolation metrics, ED (Redding 2003) has been the only one known to be used actively in conservation prioritization, by The Zoological Society of London's Edge of Existence programme (Zoological Society of London 2008). The simplicity of calculating ED and the ever-increasing availability of complete dated phylogenies facilitates the use of this metric for conservation prioritization purposes. One implicit argument for the use of such isolation metrics is that sets of high-ranking species capture non-random amounts of PD.

In this chapter, I studied the relationship between loss of ED and loss of PD across different types/sizes of simulated and real trees. Using a combination of simulated and real phylogenies, I show for the first time that the loss of summed evolutionary distinctness is very closely correlated with loss of phylogenetic diversity under the two scenarios of extinction performed in this thesis.

Under random extinction, tree shape significantly determines how much phylogenetic diversity is lost for each unit loss of ED. I find that tree balance (Colless index) has a positive effect on amount of PD loss, meaning that in trees with higher degrees of imbalance where clades are more dissimilar in number of species, loss of ED is more closely related to loss of PD. This can be attributed to two things: imbalanced trees have more of the total PD represented by exterior (pendant) branches (Figure 27), and also have the loss of internal branches when the few representatives of a small clade are lost.

Gamma is shown to be negatively related to the loss of phylogenetic diversity. Trees with younger speciation events (higher gamma) lose less phylogenetic diversity per unit of ED lost. This is due to the fact that minor extinction events prune the pendant edges of such trees that tend to be short in average. Similar results were suggested by Nee and May (1997) in which they argued that a big portion of tree of life would survive a

mass extinction event. This claim was based on the fact that they employed coalescent trees that have very short pendant edges (Mooers et al., 2012).

The association between loss of PD and loss of ED under random extinction was always greater than 0.65. However I detected some degree of variation, with bigger trees showing stronger correlations between the two variables. As demonstrated in figure 15 bigger trees tend to have less variation in the length of pendant edges and that is perhaps the underlying reason for getting tighter relationship for empirical phylogenies.

Although slow-down in diversification rates of clades is common in real-life (Moen and Morlon 2014), there is no unequivocal agreement on the mode of evolution through time (Lewitus and Morlon 2016b). All three real phylogenies were found to be high gamma trees, with nodes closer to the tips of the phylogeny, with amphibians' standing out as the "tippiest" phylogeny. More relatively recent diversification in amphibians (shorter pendant edges) is reflected in the shallower slope of ED loss and PD loss relationship compare to birds and mammals. As one can predict from the shape metrics of the three empirical phylogenies, extinction of one amphibian species leads to much lower loss of amphibian PD compared to the extinction of a bird or mammal species and the loss of their total PD.

Interestingly, under extinction scenarios where high ED species are chosen for extinction, the relationship between loss of PD and loss of ED was very similar among all surveyed phylogenies. Unlike the observed variation among different types and sizes of phylogenies under the random scenario of extinction, here I report very steep slopes (above 0.9) and tight associations between loss of phylogenetic diversity and ED, regardless of tree shape and size. Successive loss of highly isolated species can result in the loss of internal branches associated with them and under such circumstances a high proportion of PD is lost from early on. Evolutionarily distinct species have been shown to contribute disproportionately to overall phenotypic diversity for several taxa (Redding et al. 2008; Magnuson-Ford et al. 2009; Redding et al. 2015, see also Chapter 2). In sharks and rays, though not in other vertebrate groups (Verde Arregoitia et al. 2013; Jetz et al. 2014) evolutionary isolated species are also at higher risk of extinction (Stein et al., pers. comm.). Although much more work is needed testing the overall value

of PD, the strong relationship I report here may offer a strong argument for focusing species-specific conservation efforts on the highest ED species.

Despite the arguments for saving evolutionarily isolated species, the ED metric has been criticized vigorously for not taking species complementarity into account (Faith 2008), and more sophisticated forms of ED have been suggested that explicitly take complementarity into account (Jensen et al. 2016). The tight correlations I report here argue against the need for more complex metrics.

A correlation between PD and cumulative ED is somewhat expected because PD is highly correlated with species richness (Davies and Buckley 2011, F. Mazel, pers. comm.), and so is cumulative ED (results now show). In order to test to what extent species richness is driving the strong correlations we see, I employed the two discussed scenarios of extinction in Yule trees where I controlled for number of species pruned. In both cases I report high degrees of association between PD and ED ($r^2 \sim 0.8$) for any given number of extinct species. Consistently, loss of 10 highly isolated species is associated with loss of more phylogenetic diversity compared to the loss of 10 random species (Figure 26). This final test is an indication that ED may be a much more reliable surrogate of PD than has been appreciated by both advocates and critics of the approach. I take up more specific suggestions for future research following on from my MSc. work in Chapter 4.

Chapter 4.

Conclusion

The unprecedented biodiversity loss has invoked responses from scientists and decision makers in hopes of mitigating this crisis. As one of the most internationally engaged organizations in conservation, the Convention on Biological Diversity (CBD) has set the following as one of its goals (www.cbd.int/sp/targets/#GoalC):

“By 2020, at least 17 per cent of terrestrial and inland water, and 10 per cent of coastal and marine areas, especially areas of particular importance for biodiversity and ecosystem services, are conserved through effectively and equitably managed, ecologically representative and well connected systems of protected areas and other effective area-based conservation measures, and integrated into the wider landscapes and seascapes.”

There is probably a consensus on the need for more conservation effort and resources but what different scientists and policy makers may find “*of particular importance for biodiversity*” varies noticeably. This challenge, combined with funding limitations, can hinder implementation of comprehensive conservation plans to a great extent.

Aside from those challenges, conservation programs are inconsistently allocating the scarce funding across ecosystems/species/populations (see Joseph et al. 2009; Veron et al. 2015). One species-based currency for conservation prioritization; Evolutionary Distinctness (Redding 2003) has been adopted by EDGE of EXISTENCE initiative (Zoological Society of London 2008). In this thesis I rigorously tested the effectiveness of this metrics from two perspectives:

Since evolution is assumed to be a conservative process, phylogenetic distance should correlate with functional distance. However this relationship may decay very fast under different models of trait evolution (see Cadotte et al. 2013; Kelly et al. 2014; Letten and Cornwell 2015). I employed the concept of average trait rarity which was suggested as a measure of originality (Pavoine et al. 2005) and showed that ED highly correlates

with average rarity of traits in low gamma trees but its efficiency declines as tree gets tippy. However, when traits evolve very fast and there is evidence of saturation, ED actually captures trait rarity quite well. Those result echoes the fact that ED is an effective measure in capturing changes that occur near the tip of a phylogeny so that in low gamma trees where pendant edges are relatively longer on average as well as cases of multiple changes along the external edges of a phylogeny it is positively correlated with the employed index of originality (i.e. average rarity). Further studies (simulation and real life data) should scrutinize the behaviour of continuous traits, and consider different measures of functional diversity, as has been suggested by others (see Buisson and Grenouillet 2009; Buisson et al. 2013)

An analytical critique about ED is its failure for taking complementarity into account (Faith 2008). However, ED has favourable properties because it is a species-specific metric. This latter property is an asset since many conservation planning and resource allocation occur at species level. Here, I showed that loss of ED is strongly correlated with loss of PD regardless of size and shape of the phylogeny. This holds under both random and high ED scenarios of extinction. Notably, this relationship remained tight when I controlled for number of lost species. In spite of rigorous criticism about ED lacking complementarity (Faith 2008), my results suggest that, on average, ED is a robust proxy for phylogenetic diversity across several types and different sizes of phylogenies.

In summary, results of this research can be applied for informing conservation programs, and particularly the programs that are applying ED for species prioritization. These conservation programs should consider the possibility that saving evolutionarily isolated species might not always lead to conservation of high degree of trait rarity. In case of high gamma phylogenies and also instances where traits of interests are known to evolve rather slowly, ED shouldn't be the first choice for prioritization purposes. Indeed, the poor performance of ED in capturing rarity of life history traits in the order of *Chiroptera*, which was also detected in birds at different spatial scales (Redding et al. 2015) should be of particular concern for its users. The good news however, is the robust performance of another metric of isolation (APD) in capturing trait rarity. APD showed very high degrees of association with rarity in simulation and some degrees of positive correlation with trait rarity in bats which is in accord with its performance in case of Nearctic and Neotropical birds (see Redding et al. 2015) . These results should

prompt further studies to test APD's performance in capturing other facets of biodiversity before its application in conservation prioritization is advocated.

That said, if a conservation program's goal is to preserve the tree of life (i.e. preserve PD), it can conveniently benefit from application of ED as a species-specific metric that highly correlates with PD on a wide range of phylogenetic sizes and shapes. Taken together, conservation programs should choose their metric of prioritization based on the explicit aim of the program. Otherwise, the limited resources available for conservation will not be used efficiently.

Figures and Tables

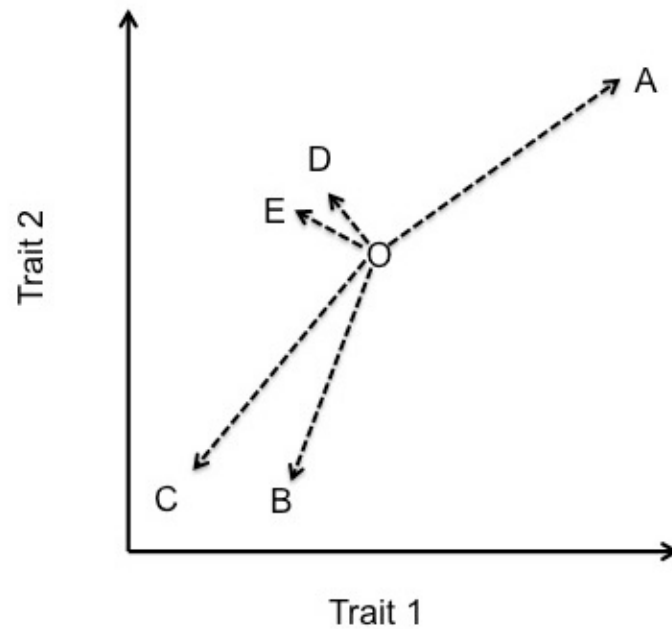


Figure 1. The concept of originality and uniqueness in functional space. Based on these two traits, species A is both original and unique while B and C are original but not unique. Adapted from Buisson et al. 2013.

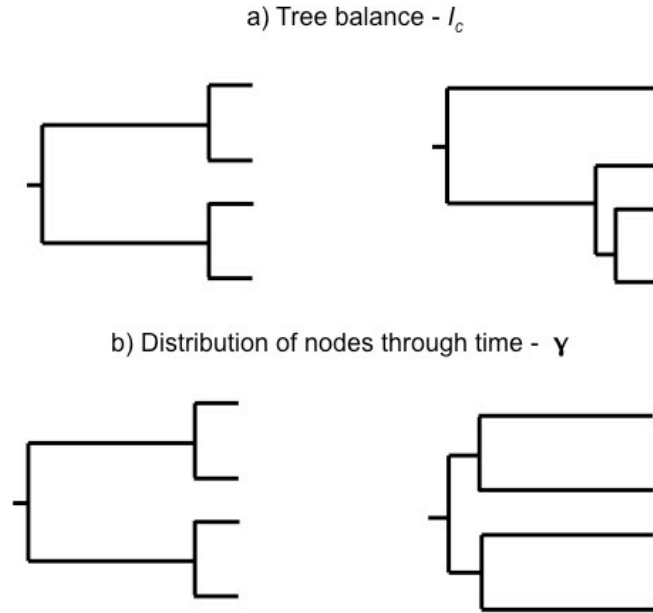


Figure 2. Two metrics of tree shape used in this study: a) Colless' measure of tree balance (I_c), smaller value represents more balanced tree (left) as apposed to higher values (right) and b) Distribution of nodes ages (γ) with higher values indicating younger speciation events (lef) and lower values representing older speciation events.

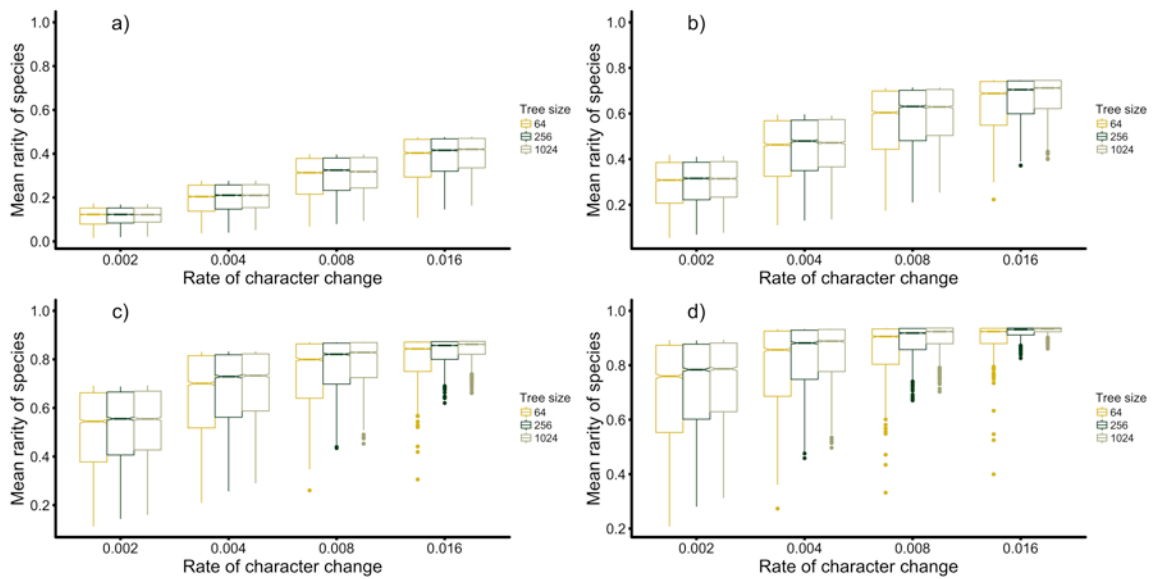


Figure 3. Mean rarity for 500 traits across three size of phylogenies and four rate of character change in case of: a) two-state traits, b) four-state traits c) eight-state traits and d) sixteen-state traits.

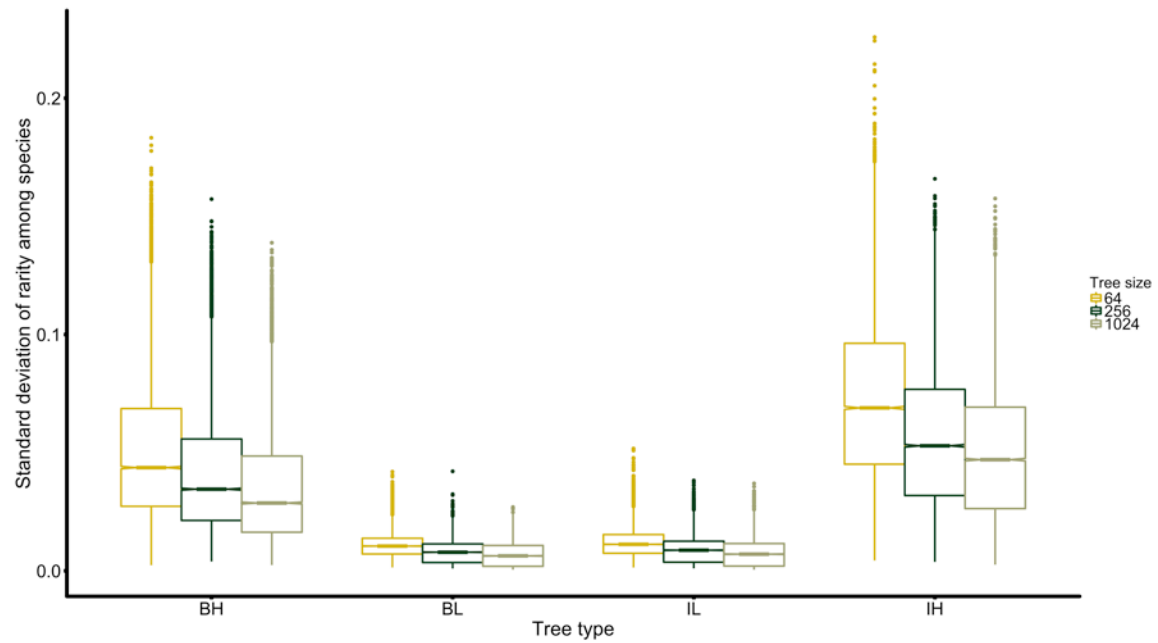


Figure 4. Standard deviation of mean rarity among species. Each point represents the SD of mean rarity among species of the corresponding tree.

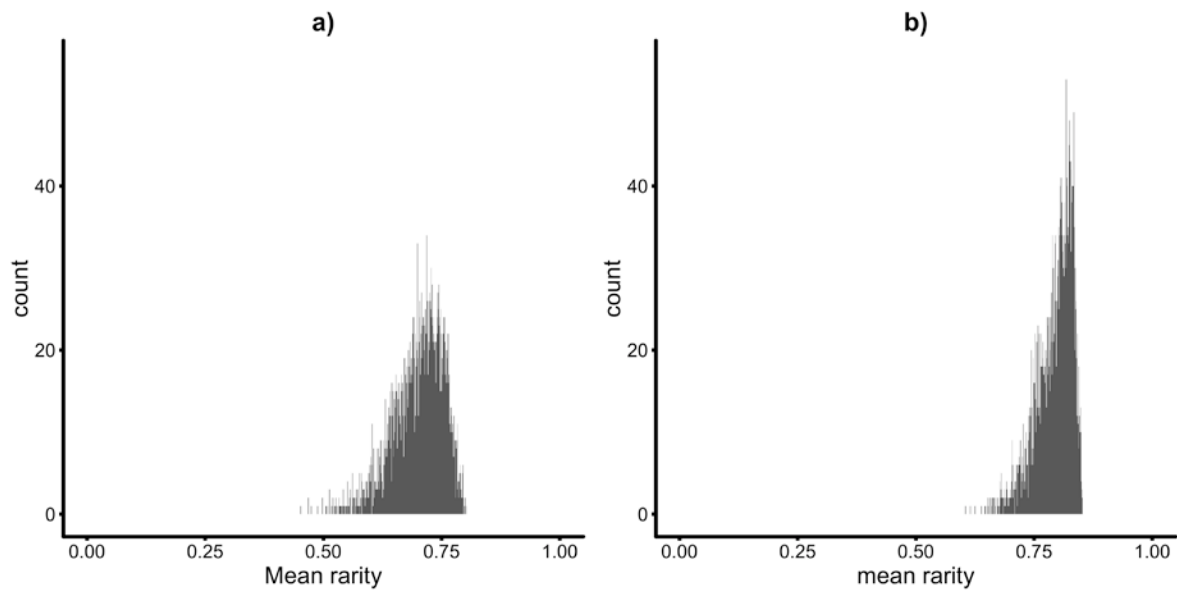


Figure 5. Distribution of average rarity of 500 traits with 8 states on imbalanced high gamma trees: a) rate of 0.008 (Mean rarity=0.7, skewness=-0.85) and b) rate of 0.016 (Mean rarity=0.79, skewness=1.06).

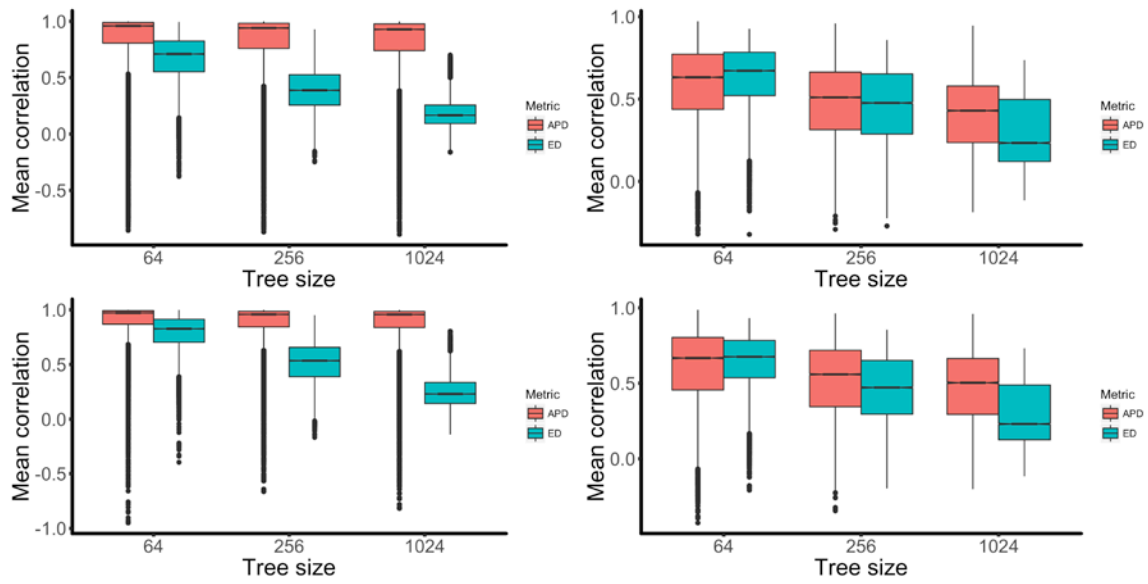


Figure 6. Correlation between each isolation metric and average rarity in: a) balanced high gamma trees, b) balanced low gamma trees c) imbalanced high gamma trees and d) imbalanced low gamma trees.

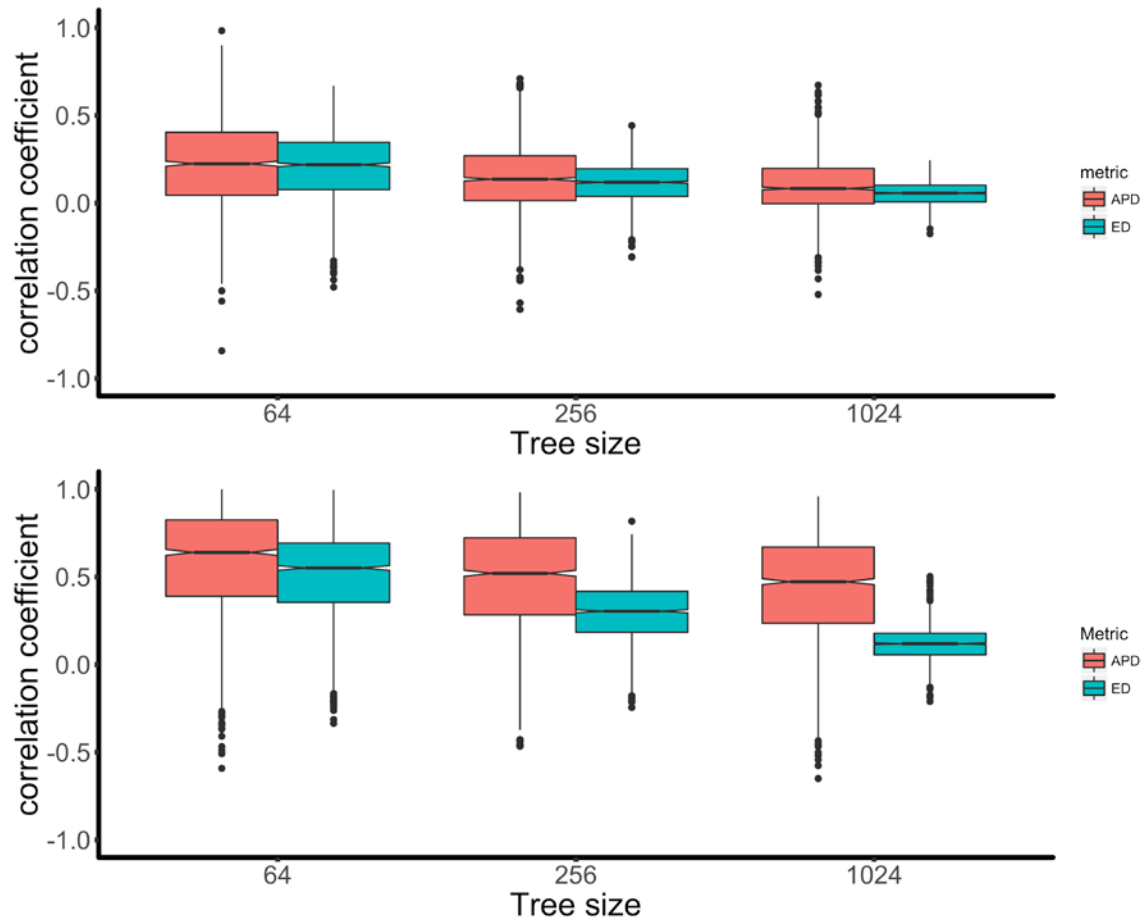


Figure 7. Effect of tree size on the performance of isolation metrics in terms of capturing trait rarity. Top panel: 5 traits simulated on 100 Yule trees of different size and bottom panel: 5 traits simulated on 100 imbalanced high gamma trees of different size.

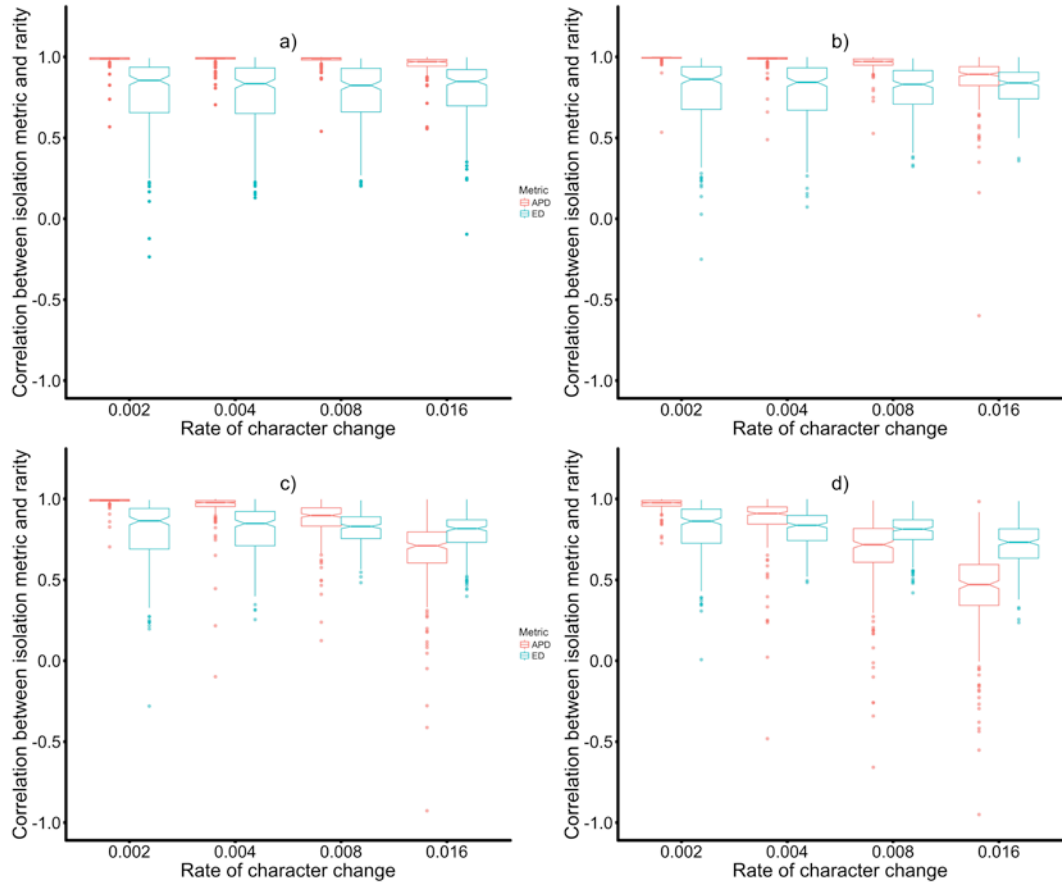


Figure 8. Correlation between each isolation metric and average rarity in imbalanced high gamma trees of size 64 with 500 traits having: a) Two, b) Four c) Eight and d) Sixteen trait states, at different rates of evolution.

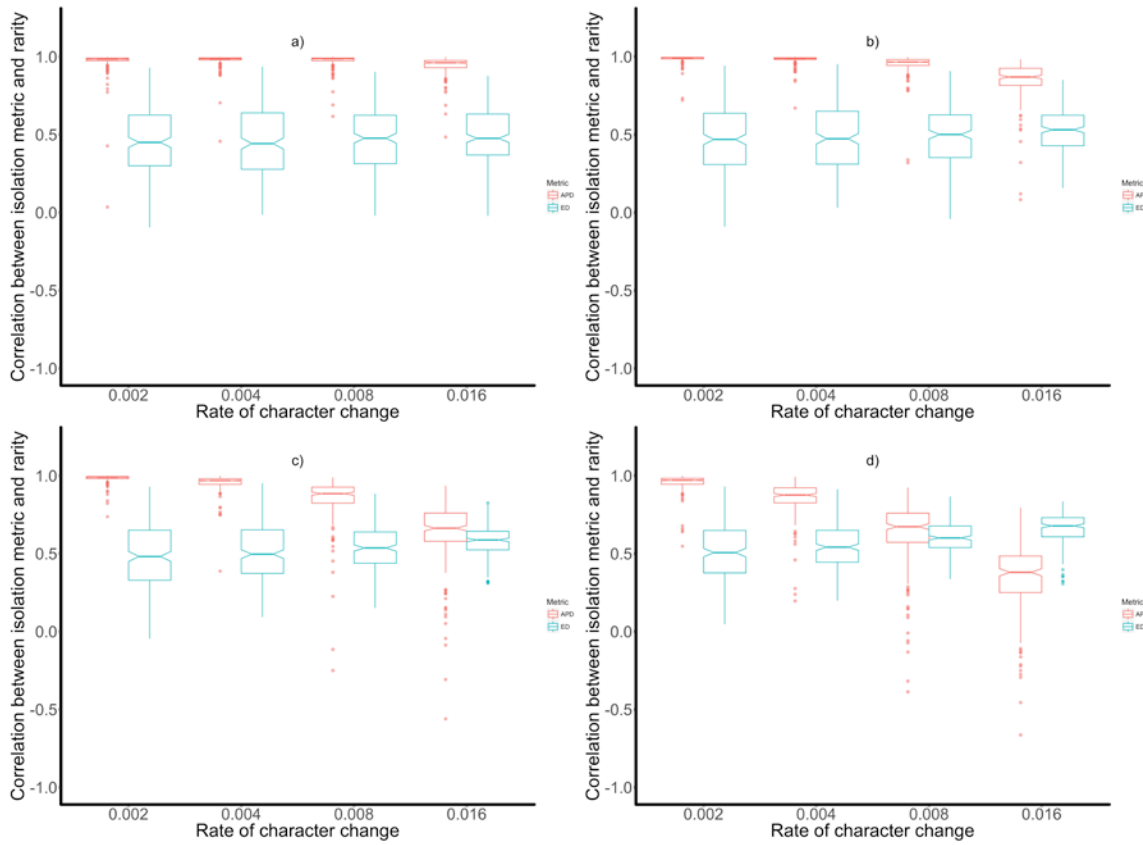


Figure 9. Correlation between each isolation metric and average rarity in imbalanced high gamma trees of size 256 with 500 traits having: a) Two, b) Four c) Eight and d) Sixteen trait states, at different rates of evolution.

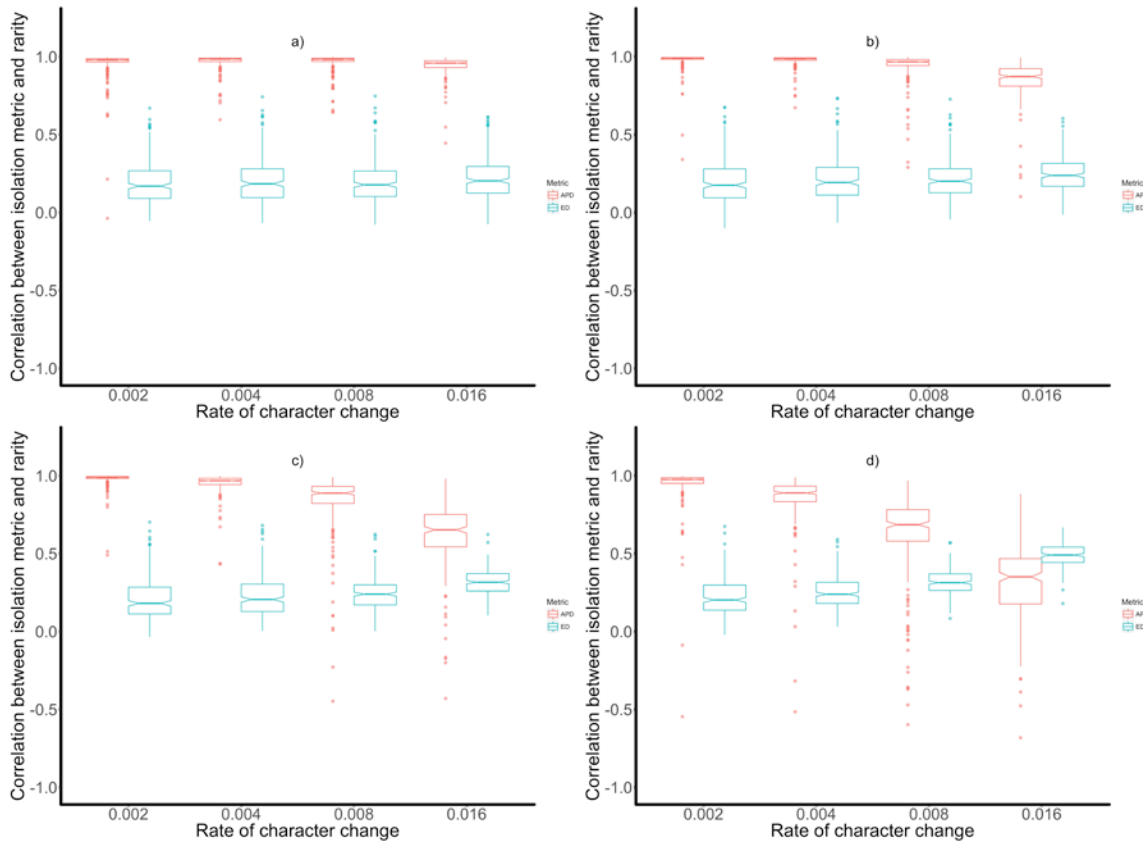


Figure 10. Correlation between each isolation metric and average rarity in imbalanced high gamma trees of size 1024 with 500 traits having: a) Two, b) Four c) Eight and d) Sixteen trait states, at different rates of evolution.

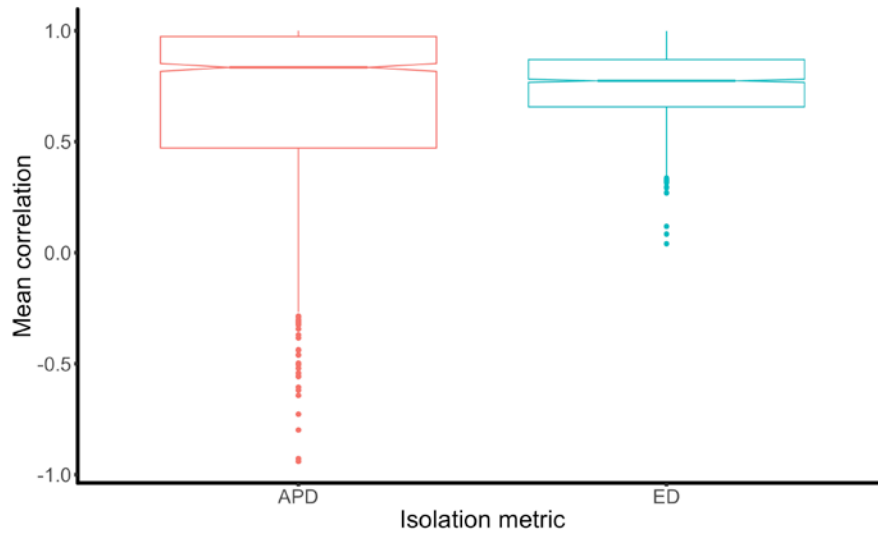


Figure 11. Correlation between each isolation metric and average rarity in imbalanced high gamma trees of size 64 with 250 traits evolving at low rate (0.0002) and 250 traits evolving at high rate (0.016) both having 16 character states.

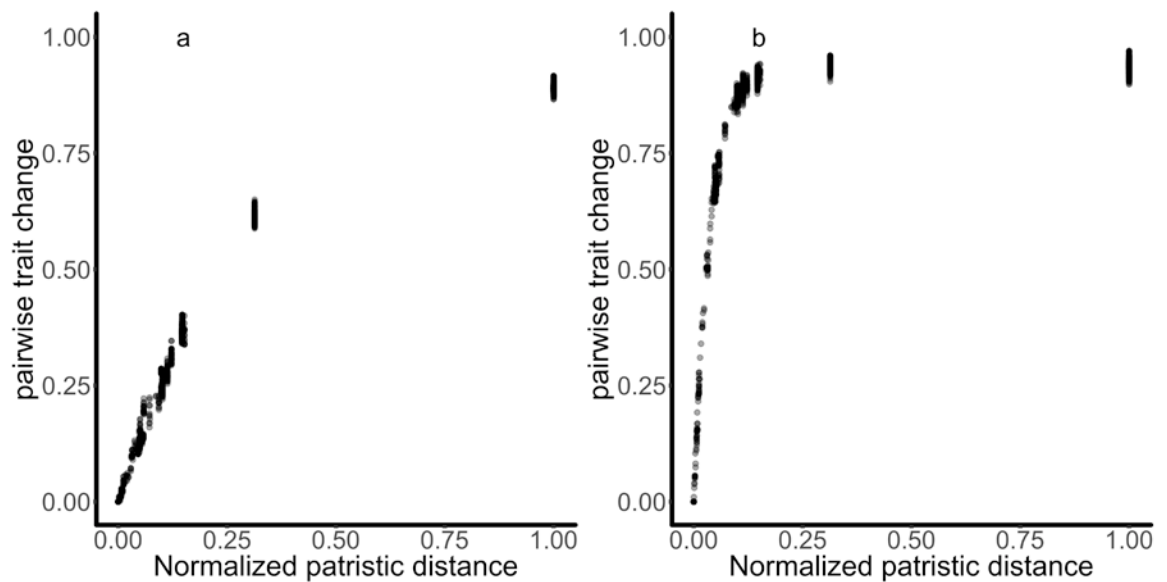


Figure 12. Relationship between trait difference and normalized phylogenetic distance on an imbalanced high gamma tree ($n=64$) whereby: a) 500 traits with 16 states evolving at the rate of 0.002 and b) 500 traits with 16 states evolving at the rate of 0.016.

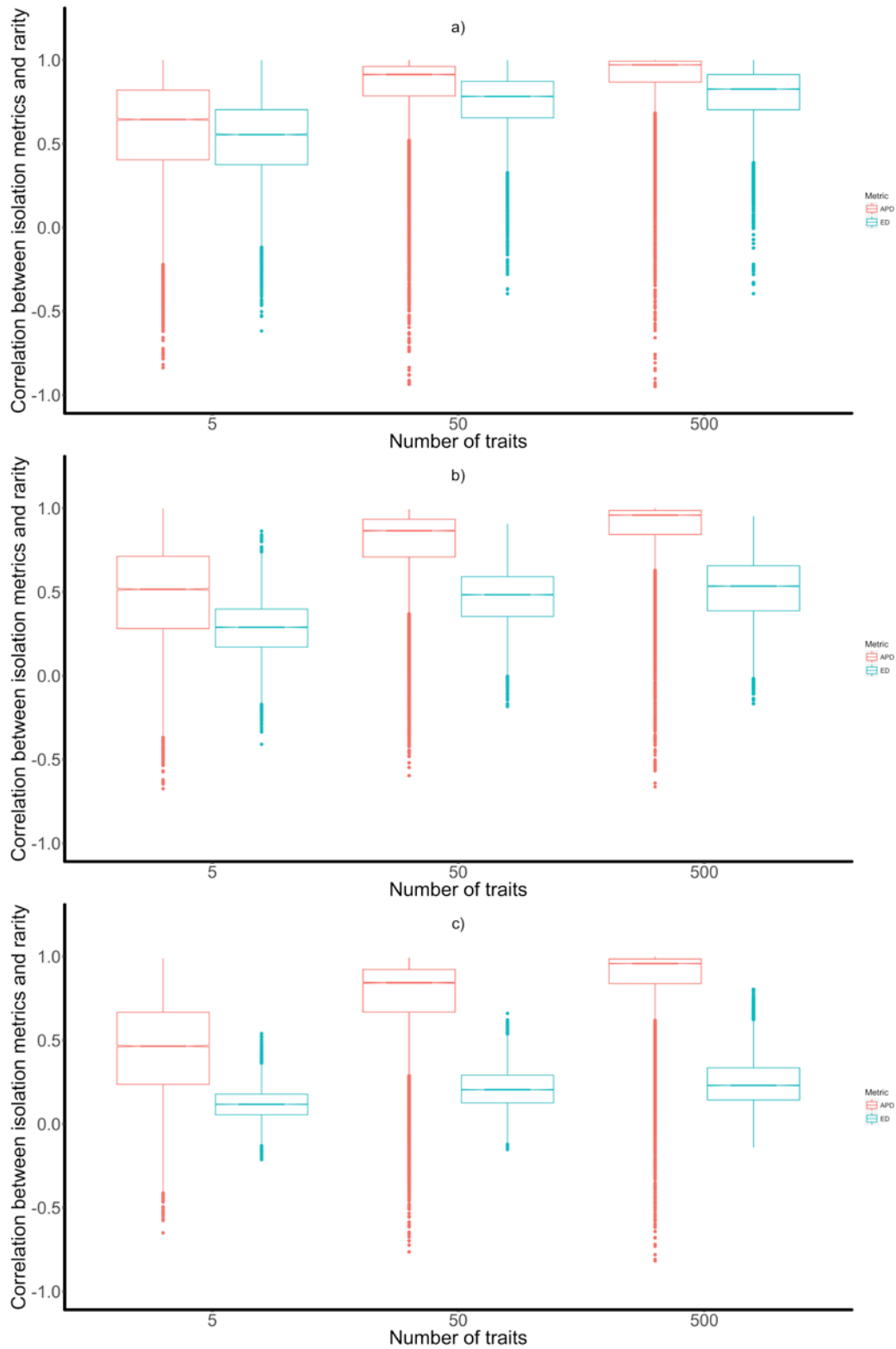


Figure 13. Effect of trait number on the efficiency of isolation metrics in capturing average rarity in: a) imbalanced high gamma tree with 64 species, b) imbalanced high gamma tree with 256 species and c) imbalanced high gamma tree with 1024 species.

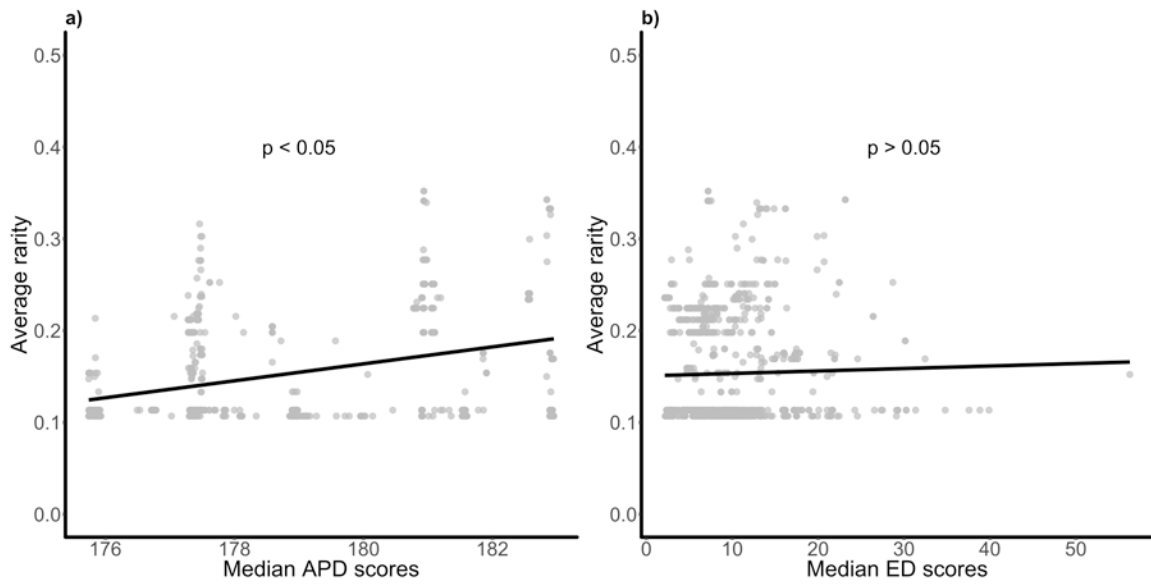


Figure 14. Distribution of the two isolation scores and average rarity across 1073 species of bats: a) APD and b) ED

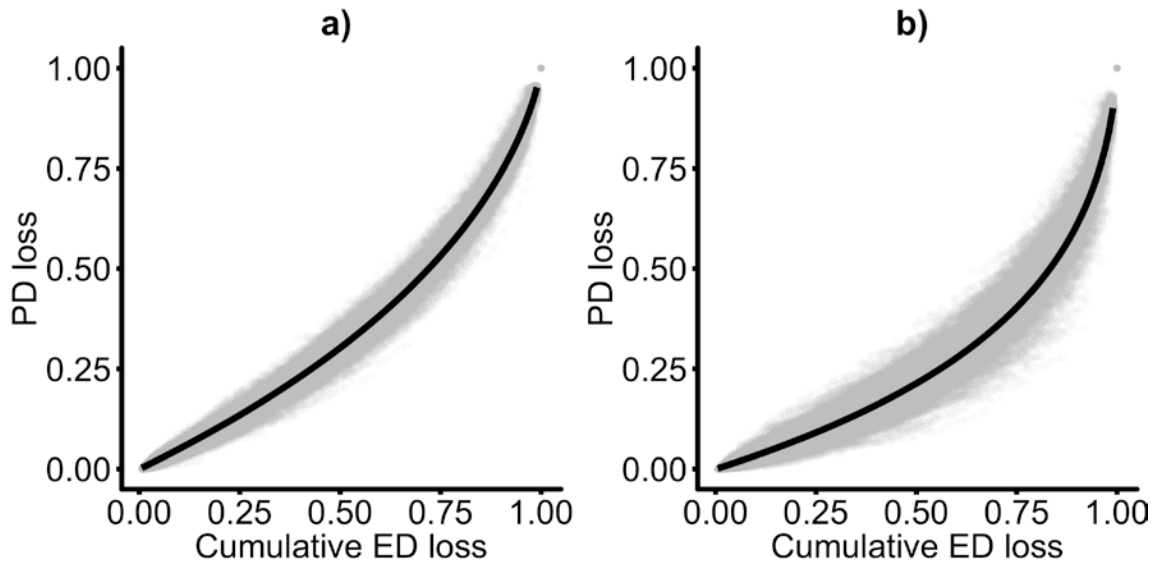


Figure 15. Relationship between Sum of ED loss and PD loss under random extinction in 1000 a) low-death ($d=0.1$) trees b) high-death ($d=0.9$) trees. The black curve in each panel depicts the expected curve from Mike Steel et al., pers. comm.

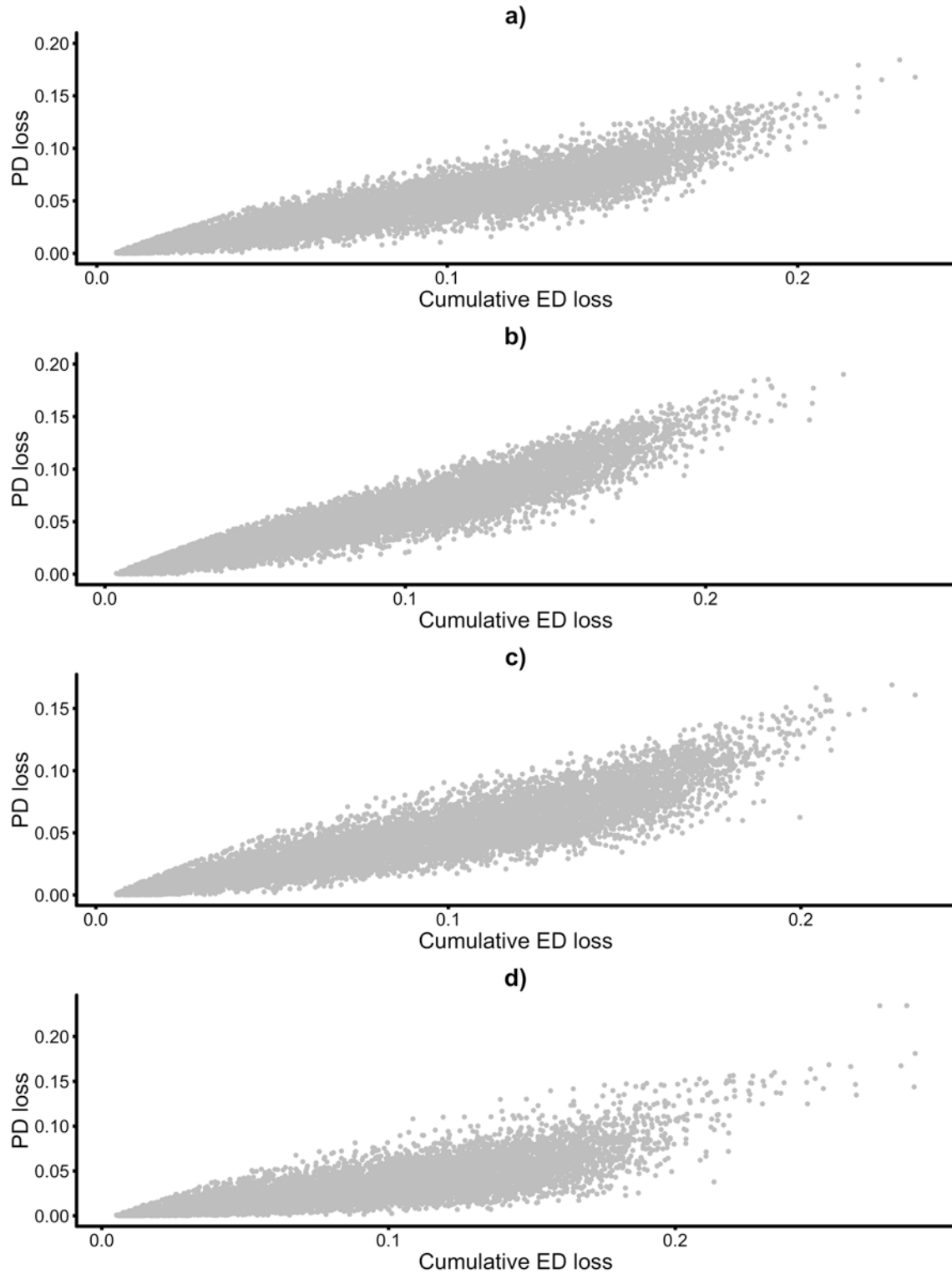


Figure 16 Relationship between Loss of ED and Loss of PD in case of 15% random extinction in: a) Yule trees, b) Imbalanced trees, c) Low death trees and d) High death trees.

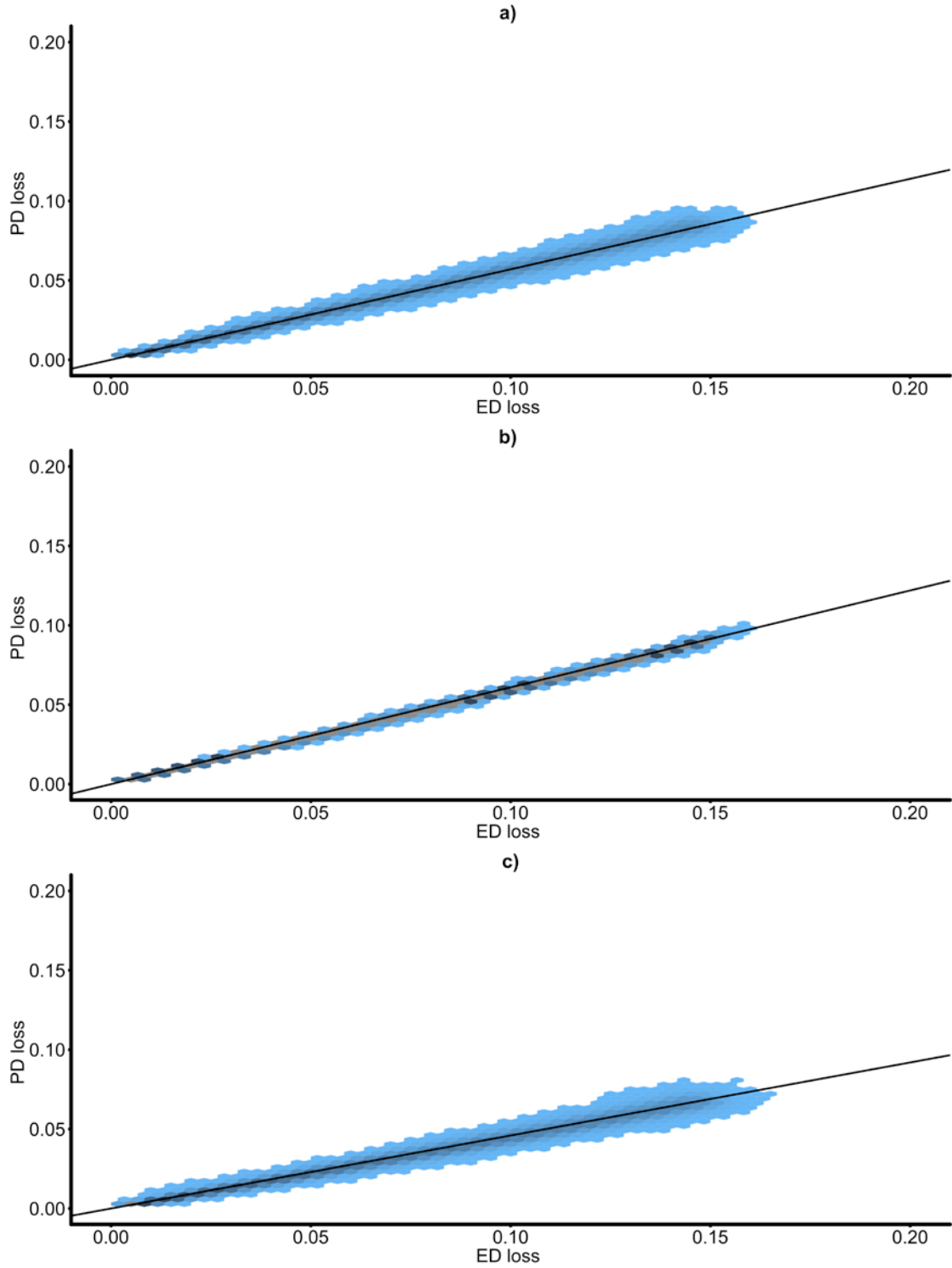


Figure 17. Linear relationship between loss of ED and loss of PD up to 15% species loss under random extinction in: a) 1000 trees of mammals, b) 1000 trees of birds and c) 1000 trees of amphibians.

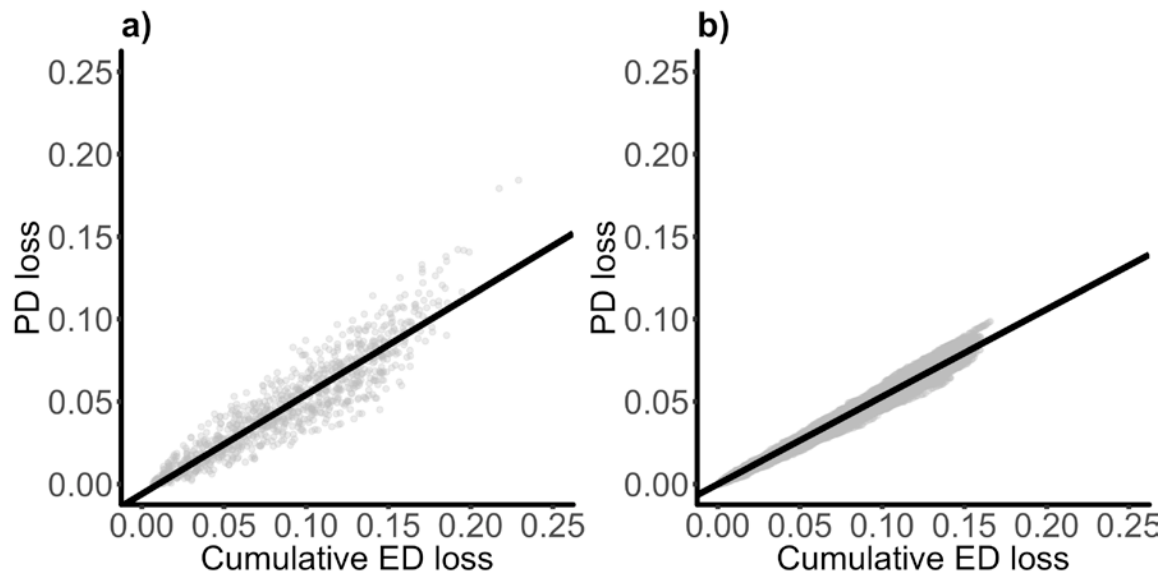


Figure 18. Effect of the tree size on the goodness of fit regarding the relationship between PD loss and ED loss in: a) 100 Yule trees of size 64 and b) 100 Yule trees of size 1024.

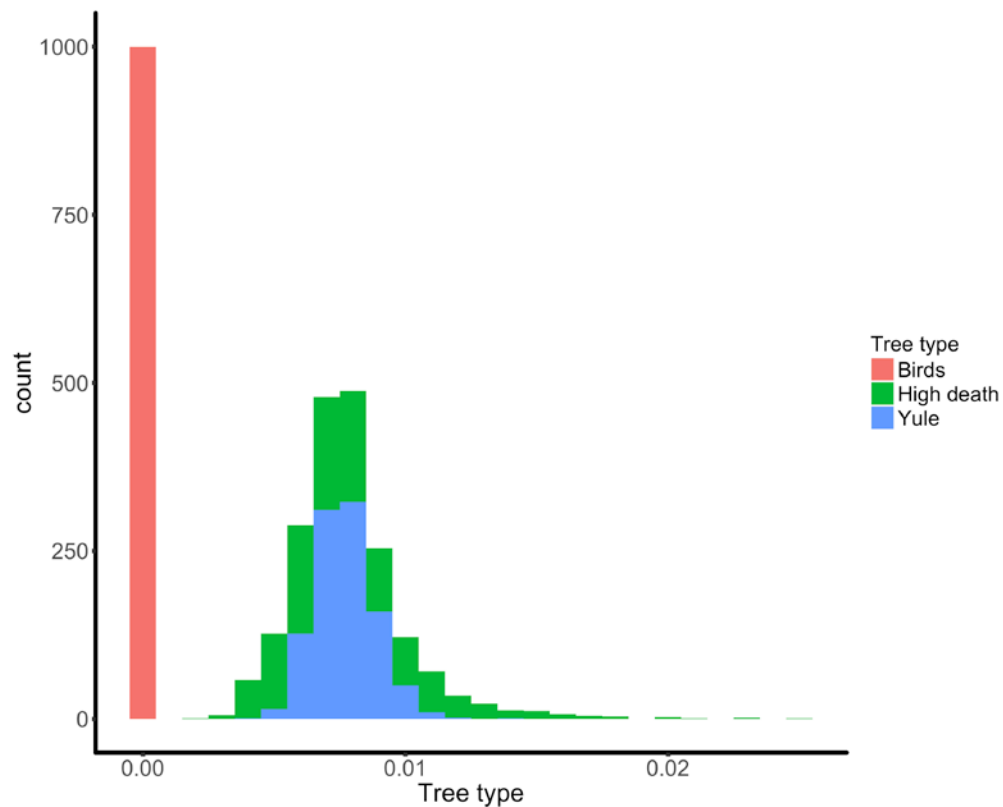


Figure 19. The distribution of variation in pendant edges (standardized by PD of the corresponding tree) in three different sets of phylogenies.

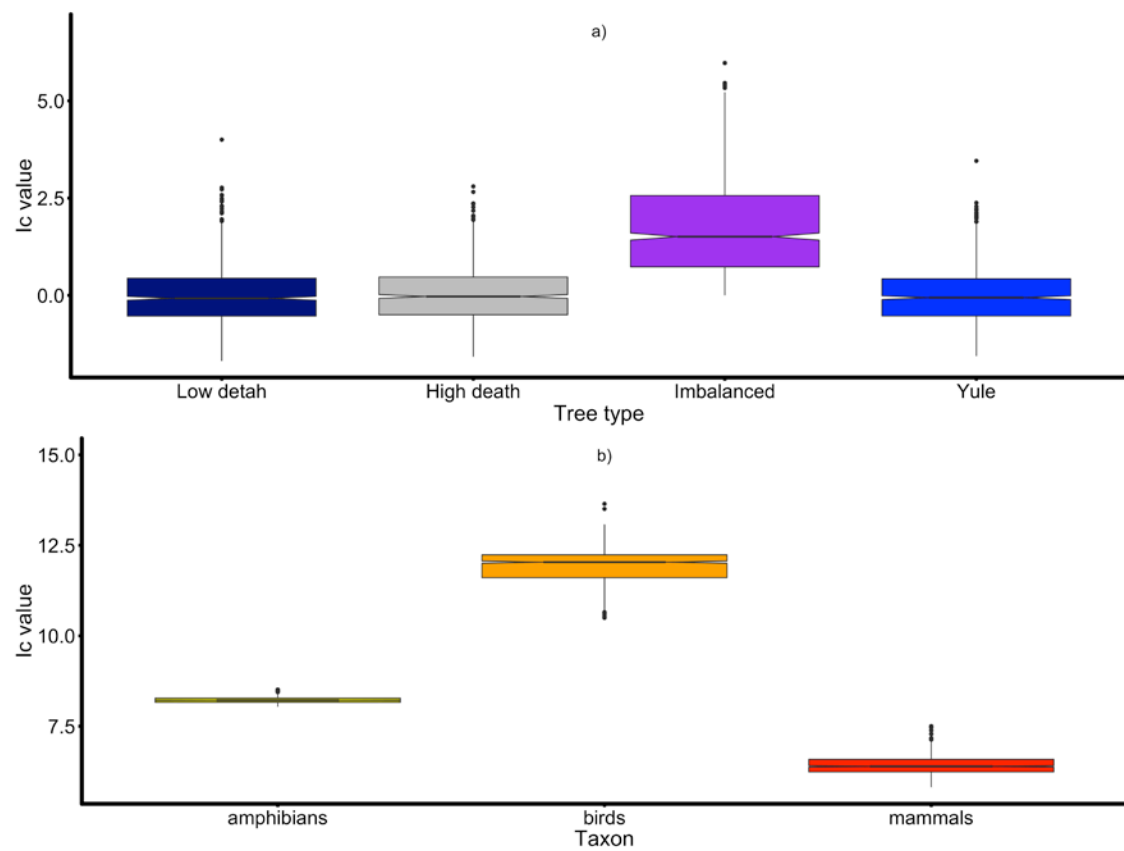


Figure 20. Values of Colless index in: a) four types of simulated trees and b) three empirical phylogenies

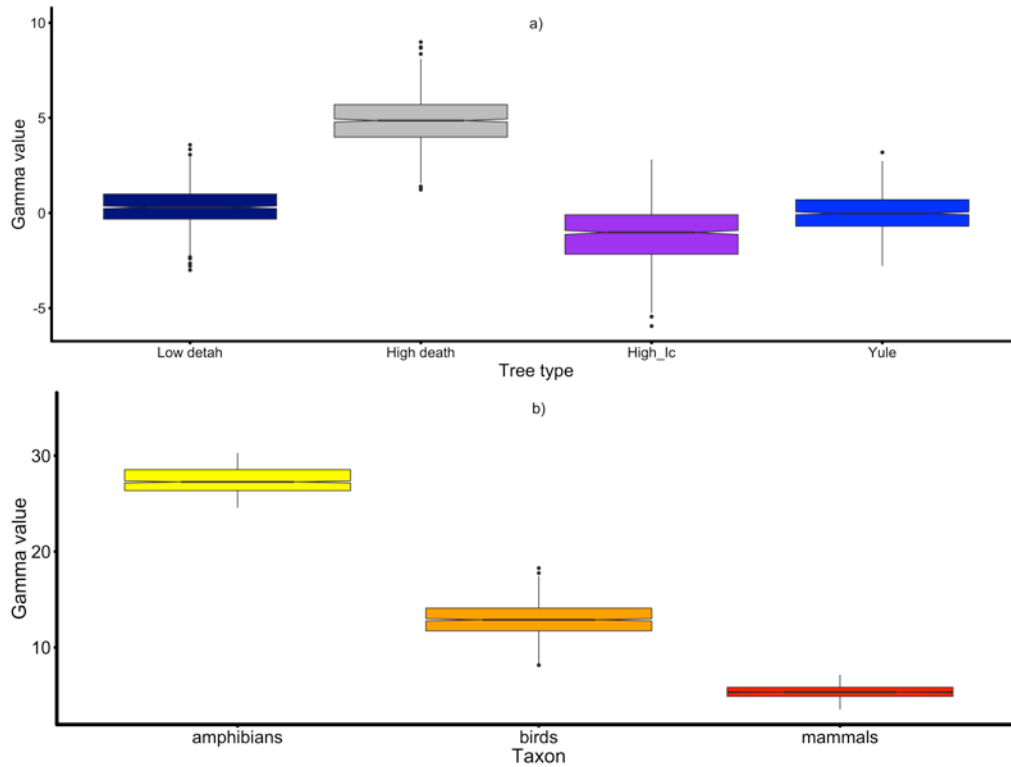


Figure 21. The distribution of Gamma values in: a) four types of simulated trees and b) three empirical phylogenies.

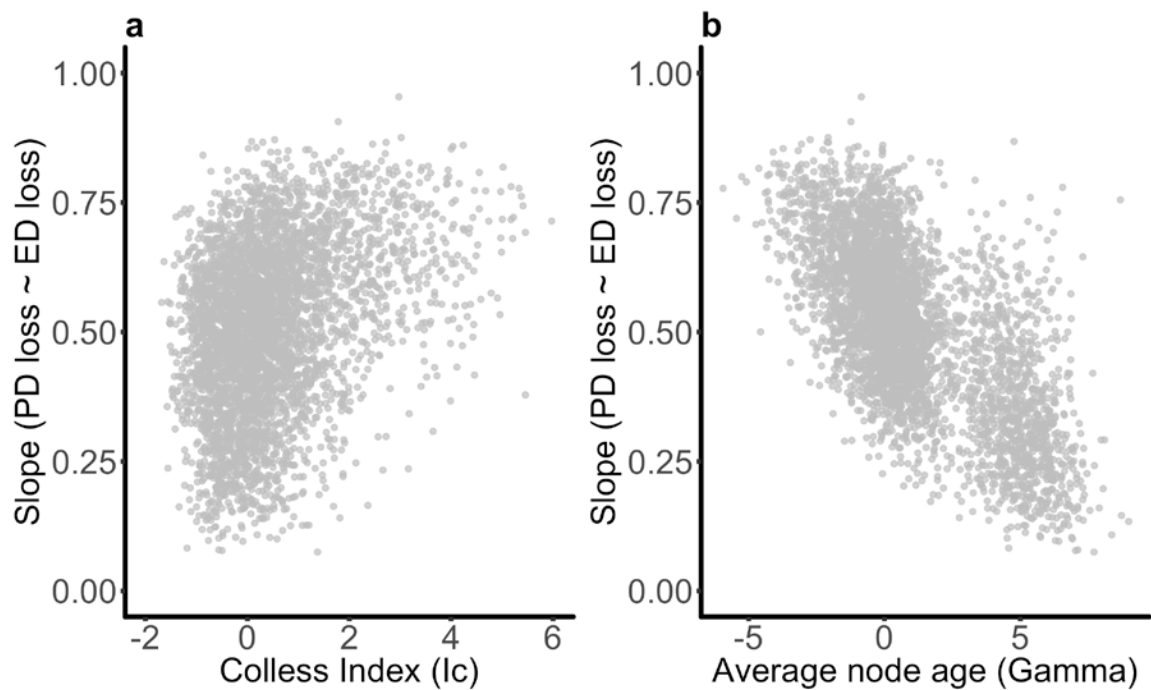


Figure 22. The scatter plot of slope and two metrics of tree shape; a) Colless index and b) Gamma in simulated trees.

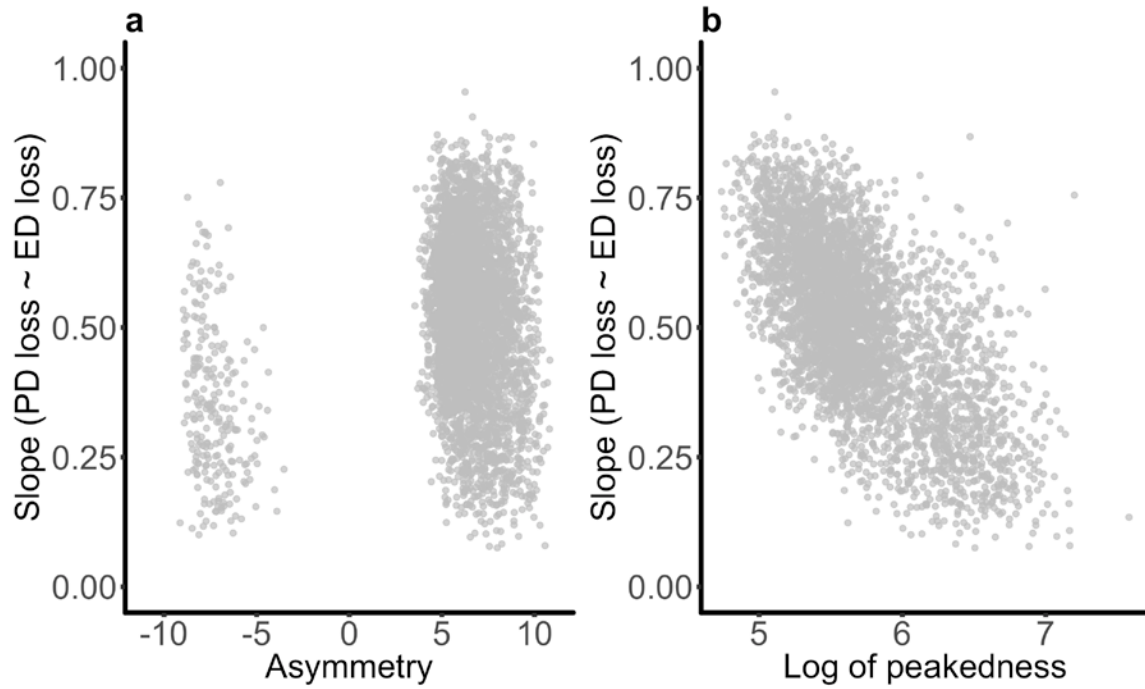


Figure 23. The Scatter plot of slope against two metrics of tree shape; a) Asymmetry and b) log transformed Peakedness.

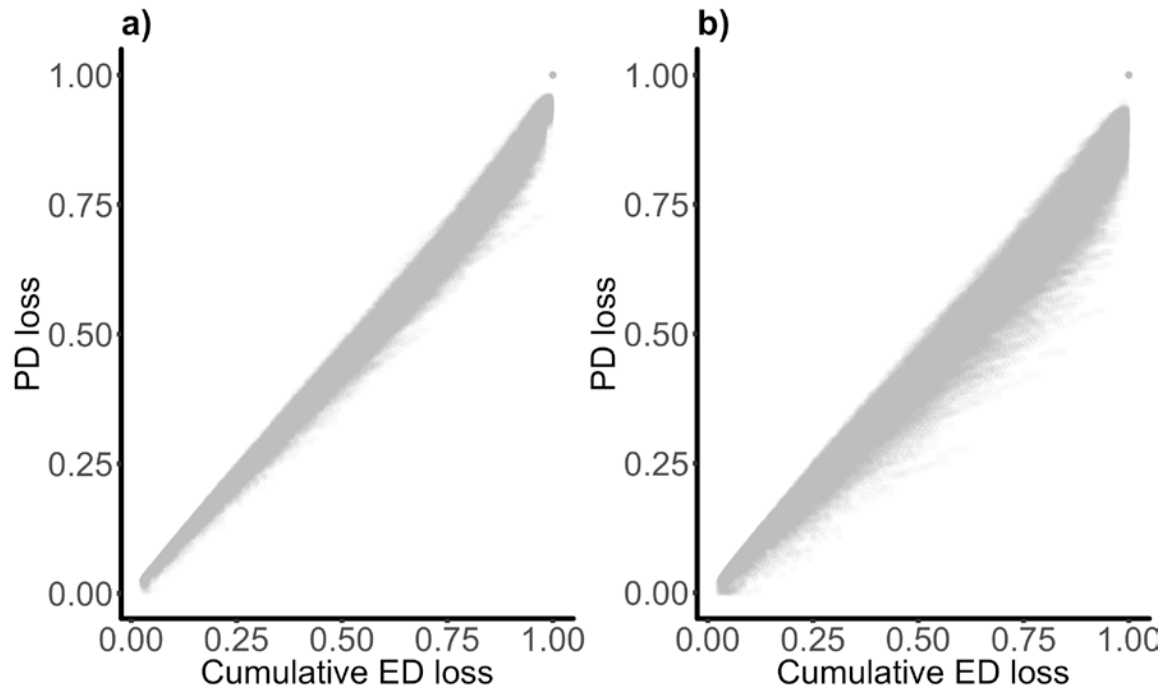


Figure 24. Relationship between Sum of ED loss and PD loss in 1000 a) low-death ($d=0.1$) trees b) high-death ($d=0.9$) trees under high ED loss scenario.

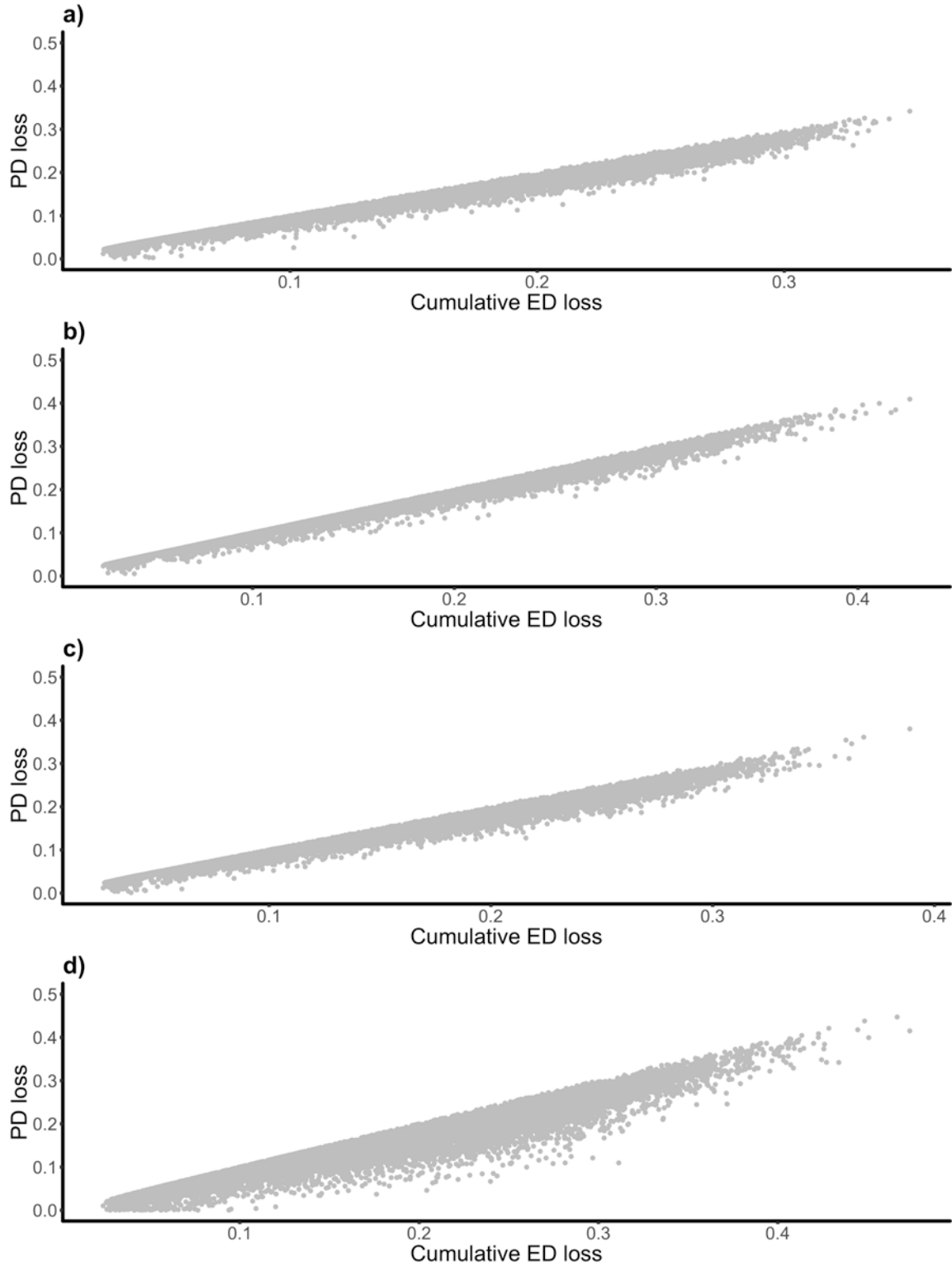


Figure 25. Relationship between Loss of ED and Loss of PD in case of high ED extinction (up to 15% loss) in: a) Yule trees, b) Imbalanced trees, c) Low death trees and d) High death trees

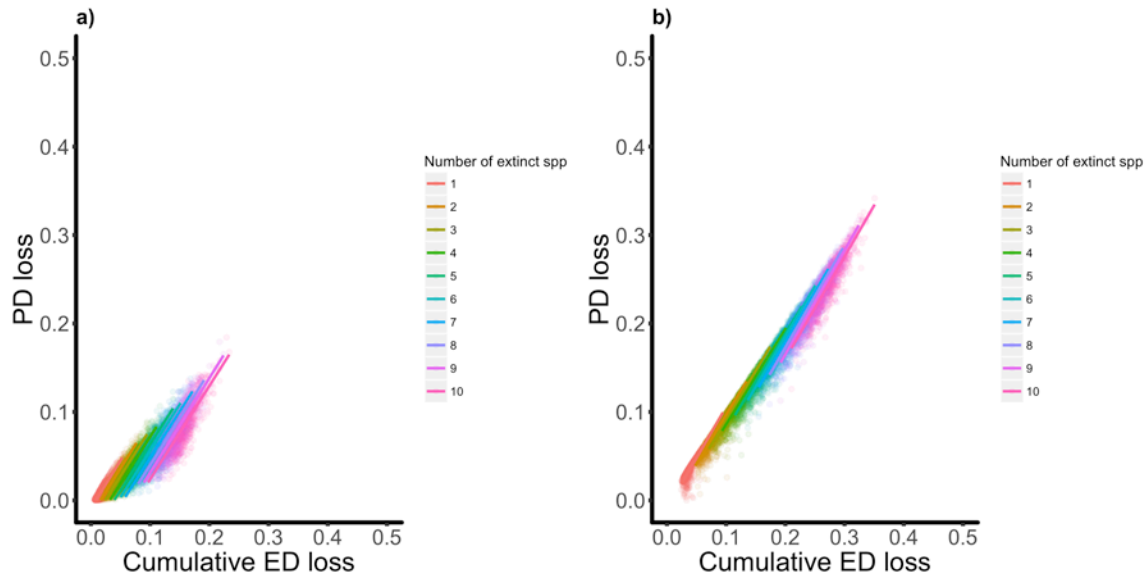


Figure 26. Loss of 10 species in 1000 Yule trees of size 64 under: a) random extinction and b) high-ED extinction where I controlled for number of extinct species.

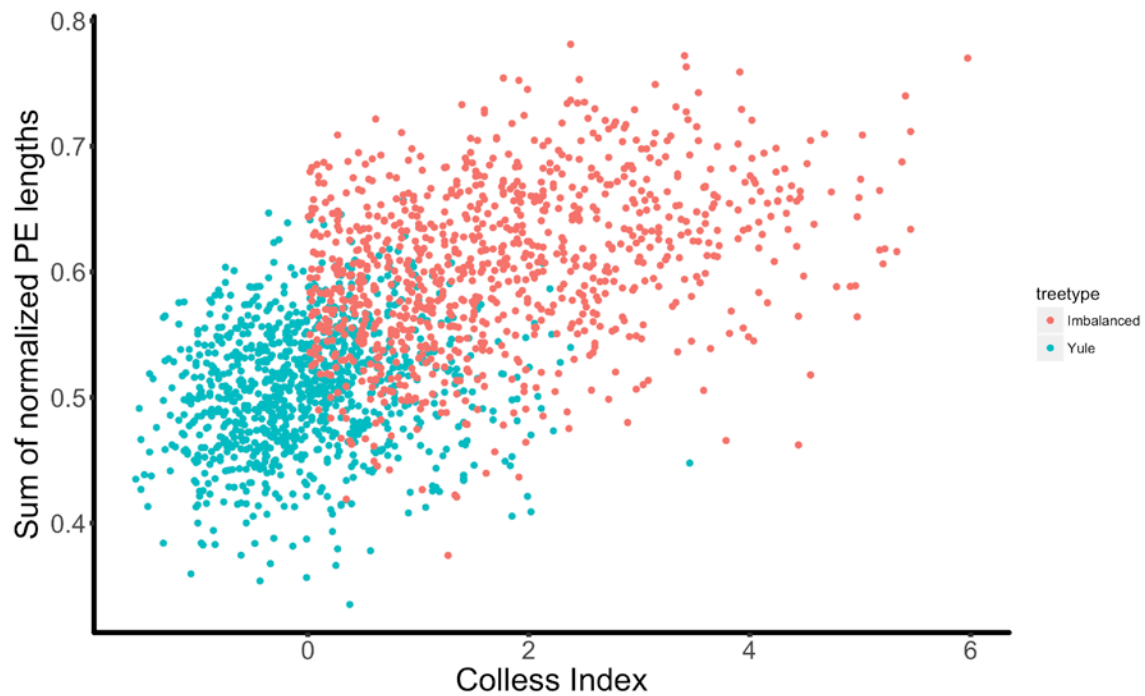


Figure 27. Relationship between degree of imbalance and the sum of normalized external edges in 2000 simulated trees

Table 1. Results of the linear model performed on three sizes of imbalanced high gamma trees. The response variable is mean rarity and predictors are tree size, number of states (K), rate of trait change (q) and the interaction between K and q ($R^2=0.93$).

Predictor	Estimate (Standard Error)
Tree size (256)	0.04(0.000)
Tree size (1024)	0.065(0.001)
K=4	0.124(0.001)
K=8	0.3(0.001)
K=16	0.51(0.001)
q	15.1(0.12)
(K=4)*q	10.16(0.17)
(K=8)*q	12.7 (0.17)

Table 2. Effect of average rarity, number of trait states (K), rate of trait change (q) and the interaction between K and q on the observed variation of rarity among species; obtained from a linear model performed on three size of imbalanced high gamma trees ($R^2=0.69$).

Predictor	Estimate (Standard Error)
Average rarity	0.08 (0.004)
K=4	0.04 (0.001)
K=8	0.1 (0.001)
K=16	0.17 (0.001)
q	7.25 (0.081)
(K=4)*q	4.31 (0.11)
(K=8)*q	8.15 (0.13)
(K=16)*q	11.1(0.15)
Average rarity* (K=4)	-0.062(0.005)
Average rarity* (K=8)	-0.17(0.005)
Average rarity* (K=16)	-0.26(0.005)
Average rarity* q	-20.47(0.18)

Table 3. Estimates of slope between loss of PD and loss of ED under random extinction in simulated and empirical phylogenies

Phylogeny	#Species	Slope $\text{PD loss} - \text{ED loss}$ (s.e.)	r^2
Yule	64	0.58 (0.002)	0.85
High- I_c trees ($I_c > 0$)	64	0.69 (0.002)	0.89
Birth-death ($d=0.1$)	64	0.57 (0.002)	0.83
Birth-death ($d=0.9$)	64	0.43 (0.003)	0.67
Mammals	5139	0.57 (0.000)	0.99
Birds	10284	0.61 (0.000)	0.99
Amphibians	5713	0.46 (0.000)	0.98

Table 4. Results of two linear regression models, performed on simulated and empirical phylogenies

Simulated trees			Empirical trees	
Predictor	Estimate (s.e.)	p -value	Estimate (s.e.)	p -value
I_c	0.022 (0.002)	$< 2e-16^{***}$	0.014 (0.000)	$< 2e-16^{***}$
Gamma	-0.038 (0.001)	$< 2e-16^{***}$	-0.006 (0.000)	$< 2e-16^{***}$
Multiple R-squared of the model: 0.47 RSME: 0.11			Multiple R-squared of the model: 0.87 RSME: 0.02	

Table 5. Results of linear regression model, performed on 4000 simulated trees in order to test the effects of asymmetry and peakedness on the slope of PD loss and ED loss

Slope ~ Asymmetry + Log (peakedness) Multiple R-squared of the model: 0.44, RSME: 0.12		
Predictor	Slope (Standard error)	p-value
Asymmetry	0.005(0.000)	<2e-16***
Log (peakedness)	-0.23(0.004)	< 2e-16 ***

*** Significant at 0.000,

Table 6. Estimates of slope between loss of PD and loss of PD under high ED extinction in seven simulated and empirical phylogenies

Phylogeny	#Species	Slope PD loss ~ ED loss (s.e.)	r ²
Yule	64	0.90 (0.000)	0.98
High- <i>Ic</i> trees (<i>Ic</i> >0)	64	0.94 (0.001)	0.99
Birth-death (d=0.1)	64	0.90 (0.001)	0.98
Birth-death (d=0.9)	64	0.89 (0.000)	0.92
Mammals	5139	0.92 (0.000)	0.99
Birds	10284	0.94 (0.000)	0.99
Amphibians	5713	0.90 (0.000)	0.99

References

- Agapov P-M, Bininda-Emonds ORP, Crandall KA, Gittlemann JL, Mace GM, Marshall JC, Purvis A. 2004. The Impact of Species Concept on Biodiversity Studies. *Q. Rev. Biol.* 79:161–179.
- Barnosky AD, Matzke N, Tomiya S, Wogan GOU, Swartz B, Quental TB, Marshall C, McGuire JL, Lindsey EL, Maguire KC, et al. 2011. Has the Earth's sixth mass extinction already arrived? *Nature* 471:51–57.
- Beaulieu JM, O'Meara BC. 2015. Extinction can be estimated from moderately sized molecular phylogenies. *Evolution* (N. Y). 69:1036–1043.
- Bennett JR, Elliott G, Mellish B, Joseph LN, Tulloch AIT, Probert WJM, Di Fonzo MMI, Monks JM, Possingham HP, Maloney R. 2014. Balancing phylogenetic diversity and species numbers in conservation prioritization, using a case study of threatened species in New Zealand. *Biol. Conserv.* 174:47–54.
- BirdLife International. 2015. The BirdLife checklist of the birds of the world: Version 8. [accessed 2016 May 1]. http://www.birdlife.org/datazone/userfiles/file/Species/Taxonomy/BirdLife_Checklist_Version_80.zip [.xls zipped 1 MB]
- Blum MGB, François O. 2006. Which Random Processes Describe the Tree of Life? A Large-Scale Study of Phylogenetic Tree Imbalance. *Syst. Biol.* 55:685–691
- Blum MGB, François O, Janson S. 2006. The mean, variance and limiting distribution of two statistics sensitive to phylogenetic tree balance. *Ann. Appl. Probab.* 16:2195–2214.
- Bortolussi N, Durand E, Blum M, François O. 2006. apTreeshape: Statistical analysis of phylogenetic tree shape. *Bioinformatics* 22:363–364.
- Buisson L, Grenouillet G. 2009. Contrasted impacts of climate change on stream fish assemblages along an environmental gradient. *Divers. Distrib.* 15:613–626.
- Buisson L, Grenouillet G, Villéger S, Canal J, Laffaille P. 2013. Toward a loss of functional diversity in stream fish assemblages under climate change. *Glob. Chang. Biol.* 19:387–400.
- Cadotte M, Albert CH, Walker SC. 2013. The ecology of differences: Assessing community assembly with trait and evolutionary distances. *Ecol. Lett.* 16:1234–1244.
- Cadotte MW. 2015. Phylogenetic diversity-ecosystem function relationships are insensitive to phylogenetic edge lengths. *Funct. Ecol.* 29:718–723.
- Cadotte MW, Cardinale BJ, Oakley TH. 2008. Evolutionary history and the effect of biodiversity on plant productivity. *PNAS* 105:17012–17017.

Cadotte MW, Carscadden K, Mirotchnick N. 2011. Beyond species : functional diversity and the maintenance of ecological processes and services. *J. Appl. Ecol.* 48:1079–1087.

Colless DH. 1982. Phylogenetics : The Theory and Practice of Phylogenetic Systematics. *Syst. Zool.* 31:100–104.

Dalerum F. 2013. Phylogenetic and functional diversity in large carnivore assemblages. *Proc R Soc B* 280: 20130.

Davies TJ, Buckley LB. 2011. Phylogenetic diversity as a window into the evolutionary and biogeographic histories of present-day richness gradients for mammals. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 366:2414–2425.

Díaz S, Lavorel S, de Bello F, Quétier F, Grigulis K, Robson TM. 2007. Incorporating plant functional diversity effects in ecosystem service assessments. *Proc. Natl. Acad. Sci.* 104:20684–9.

Faith DP. 1992. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* 61:1–10.

Faith DP. 2008. Threatened species and the potential loss of phylogenetic diversity: conservation scenarios based on estimated extinction probabilities and phylogenetic risk analysis. *Conserv. Biol.* 22:1461–1470.

Faith DP. 2013. Biodiversity and evolutionary history: useful extensions of the PD phylogenetic diversity assessment framework. *Ann. N. Y. Acad. Sci.* 1289:69–89.

Faith DP. 2016. The PD Phylogenetic Diversity Framework: Linking Evolutionary History to Feature Diversity for Biodiversity Conservation. In: Pellens R, Grandcolas P, editors. *Biodiversity Conservation and Phylogenetic Systematics*. Vol. 14. Springer International Publishing. p. 57–80.

Fonseca CR, Ganade G. 2001. Species functional redundancy, random extinctions and the stability of ecosystems. *J. Ecol.* 89:118–125.

Forest F, Grenyer R, Rouget M, Davies TJ, Cowling RM, Faith DP, Balmford A, Manning JC, Procheş S, van der Bank M, et al. 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* 445:757–760.

Genung M a., Schweitzer J a., Bailey JK. 2014. Evolutionary history determines how plant productivity responds to phylogenetic diversity and species richness. *PeerJ* 2:e288.

Hoffmann M, Hilton-Taylor C, Angulo A, Böhm M, Brooks TM, Butchart SHM, Carpenter KE, Chanson J, Collen B, Cox NA, et al. 2010. The impact of conservation on the status of the world's vertebrates. *Science* 330:1503–9.

Huang D. 2012. Threatened Reef Corals of the World. *PLoS One* 7:1–12.

- Huang J, Mi X, Ma K. 2011. A genome evolution-based framework for measures of originality for clades. *J. Theor. Biol.* 276:99–105.
- Isaac NJB, Redding DW, Meredith HM, Safi K. 2012. Phylogenetically-Informed Priorities for Amphibian Conservation. *PLoS One* 7:1–8.
- Isaac NJB, Turvey ST, Collen B, Waterman C, Baillie JEM. 2007. Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS One* 2(3): e296.
- Jensen EL, Mooers A, Caccone A, Russello MA. 2016. I-HEDGE: determining the optimum complementary sets of taxa for conservation using evolutionary isolation. *PeerJ* 4:e2350.
- Jetz W, Thomas GH, Joy JB, Redding DW, Hartmann K, Mooers AO. 2014. Global Distribution and Conservation of Evolutionary Distinctness in Birds. *Curr. Biol.* 24:919–930.
- Joseph LN, Maloney RF, Possingham HP. 2009. Optimal Allocation of Resources among Threatened Species: a Project Prioritization Protocol. *Conserv. Biol.* 23:328–338.
- Kelly S, Grenyer R, Scotland RW. 2014. Phylogenetic trees do not reliably predict feature diversity. Austin J, editor. *Divers. Distrib.* 20:600–612.
- Letten AD, Cornwell WK. 2015. Trees, branches and (square) roots: why evolutionary relatedness is not linearly related to functional distance. *Methods Ecol. Evol.* 6:439–444.
- Lewis P. 2001. A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data. *Syst. Biol.* 50:913–925.
- Lewitus E, Morlon H. 2016a. Characterizing and Comparing Phylogenies from their Laplacian Spectrum. *Syst. Biol.* 65:495–507.
- Lewitus E, Morlon H. 2016b. Natural constraints to species diversification. *PLoS Biol.* 14:e1002532.
- Magnuson-Ford K, Ingram T, Redding DW, Mooers AØ. 2009. Rockfish (Sebastes) that are evolutionarily isolated are also large, morphologically distinctive and vulnerable to overfishing. *Biol. Conserv.* 142:1787–1796.
- Maire E, Grenouillet G, Brosse S, Villéger S. 2015. How many dimensions are needed to
- Martyn I, Kuhn TS, Mooers AO, Moulton V, Spillner A. 2012. Computing evolutionary distinctiveness indices in large scale analysis. *Algorithms Mol. Biol.* 7:6.
- McGill BJ, Enquist BJ, Weiher E, Westoby M. 2006. Rebuilding community ecology from functional traits. *Trends Ecol. Evol.* (Personal Ed. 21:178–85.
- Metrick A, Weitzman ML. 1996. *Patterns of Behavior in Endangered Species*

Preservation. *Land Econ.* 72:1–16.

Meyer E, Newell D, Hines H, May S, Hero J-M, Clarke J, Lemckert F. 2004. *Rheobatrachus silus*. IUCN Red List Threat. Species 2004 e.T19475A8896430. [accessed 2016 Mar 1]. <http://dx.doi.org/10.2305/IUCN.UK.2004.RLTS.T19475A8896430.en>.

Moen D, Morlon H. 2014. Why does diversification slow down? *Trends Ecol. Evol.* 29:190–197.

Mooers A, Gascuel O, Stadler T, Li H, Steel M. 2012. Branch lengths on birth-death trees and the expected loss of phylogenetic diversity. *Syst. Biol.* 61:195–203.

Mooers AØ, Heard SB. 1997. Inferring Evolutionary Process from Phylogenetic Tree Shape. *Q. Rev. Biol.* 72:31–54.

Morlon H, Lewitus E, Condamine FL, Manceau M, Clavel J, Drury J. 2016. RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. Fitzjohn R, editor. *Methods Ecol. Evol.* 7:589–597.

Mouillot D, Bellwood DR, Baraloto C, Chave J, Galzin R, Harmelin-Vivien M, Kulbicki M, Lavergne S, Lavorel S, Mouquet N, et al. 2013. Rare Species Support Vulnerable Functions in High-Diversity Ecosystems. *PLoS Biol.* 11(5): e10.

Mouillot D, Culioli JM, Pelletier D, Tomasini J a. 2008. Do we protect biological originality in protected areas? A new index and an application to the Bonifacio Strait Natural Reserve. *Biol. Conserv.* 141:1569–1580.

Mouquet N, Devictor V, Meynard CN, Munoz F, Bersier L-F, Chave J, Couteron P, Dalecky A, Fontaine C, Gravel D, et al. 2012. Ecophylogenetics: advances and perspectives. *Biol. Rev.* 87:769–785.

Nee S, May RM. 1997. Extinction and the loss of evolutionary history. *Sci. New Ser.* 278:692–694.

Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N, Pearse W. 2013. caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 0.5.2.

Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877–884.

Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20:289–290.

Pavoine S, Ollier S, Dufour AB. 2005. Is the originality of a species measurable? *Ecol. Lett.* 8:579–586.

Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, Fitzjohn RG, Alfaro ME,

- Harmon LJ. 2014. Geiger v2.0: An expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* 30:2216–2218.
- Possingham HP, Andelman SJ, Burgman M a., Medellín R a., Master LL, Keith D a. 2002. Limits to the use of threatened species lists. *Trends Ecol. Evol.* 17:503–507.
- Purvis A, Agapow P, Gittleman JL, Mace GM. 2000. Nonrandom Extinction and the Loss of Evolutionary History. *Science* (80-.). 288:328–330.
- Pybus OG, Harvey PH. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc. R. Soc. B* 267:2267–2272.
- R Core Team. 2016. R: A Language and Environment for Statistical Computing.
- Redding D. 2003. Incorporating Genetic Distinctness and Reserve Occupancy into a Conservation Prioritization Approach. Master Thesis. University Of East Anglia, Norwich UK.
- Redding D, Mooers AØ. 2015. Ranking mammal species for conservation and the loss of both phylogenetic and trait diversity. *PLoS One* 10:e0141435.
- Redding DW, DeWolff C V, Mooers AØ. 2010. Evolutionary Distinctiveness, Threat Status, and Ecological Oddity in Primates. *Conserv. Biol.* 24:1052–1058.
- Redding DW, Hartmann K, Mimoto A, Bokal D, Devos M, Mooers AØ. 2008. Evolutionarily distinctive species often capture more phylogenetic diversity than expected. *J. Theor. Biol.* 251:606–615.
- Redding DW, Mazel F, Mooers AØ. 2014. Measuring Evolutionary Isolation for Conservation. *PLoS One* 9:e113490.
- Redding DW, Mooers AØ. 2006. Incorporating evolutionary measures into conservation prioritization. *Conserv. Biol.* 20:1670–1678.
- Redding DW, Mooers AØ, Sekercioglu ÇH, Collen B. 2015. Global evolutionary isolation measures can capture key local conservation species in Nearctic and Neotropical bird communities. *Phil. Trans. R. Soc. B* 370: 20140.
- Revell LJ. 2012. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3:217–223.
- Ricotta C. 2007. A semantic taxonomy for diversity measures. *Acta Biotheor* 55:23–33.
- Srivastava DS, Cadotte MW, Macdonald a. AM, Marushia RG, Mirotchnick N. 2012. Phylogenetic diversity and the functioning of ecosystems. *Ecol. Lett.* 15:637–648.
- Srivastava DS, Vellend M. 2005. Biodiversity-Ecosystem Function Research: Is It Relevant to Conservation? *Annu. Rev. Ecol. Evol. Syst.* 36:267–294.

- Steel M, McKenzie A. 2001. Properties of phylogenetic trees generated by yule-type speciation models. *Math. Biosci.* 170:91–112.
- Thuiller W, Lavergne S, Roquet C, Boulangeat I, Lafourcade B, Araujo MB. 2011. Consequences of climate change on the tree of life in Europe. *Nature* 470:531–4.
- Thuiller W, Lavorel S, Sykes MT, Araújo MB. 2006. Using niche-based modelling to assess the impact of climate change on tree functional diversity in Europe. *Divers. Distrib.* 12:49–60.
- Thuiller W, Pironon S, Psomas A, Barbet-Massin M, Jiguet F, Lavergne S, Pearman PB, Renaud J, Zupan L, Zimmermann NE. 2014. The European functional tree of bird life in the face of global change. *Nat. Commun.* 5:3118.
- Tonini JFR, Beard KH, Ferreira RB, Jetz W, Pyron RA. 2016. Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. *Biol. Conserv.*
- Tucker CM, Cadotte MW, Carvalho SB, Davies TJ, Ferrier S, Fritz SA, Grenyer R, Helmus MR, Jin LS, Mooers AO, et al. 2016. A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biol. Rev.*
- Tyler MJ, Shearman DJC, Franco R, Brien PO, Seamark RF, Kelly R. 1983. Inhibition of Gastric Acid Secretion in the Gastric Brooding Frog, *Rheobatrachus silus*. *Sci. New Ser.* 220:609–610.
- Vellend M, Cornwell WK, Magnuson-Ford K, Mooers AØ. 2011. Measuring phylogenetic biodiversity. In: Magurran AE, McGill BJ, editors. *Biological diversity: frontiers in measurement and assessment*. New York: Oxford University Press. p. 194–207.
- Verde Arregoitia LD, Blomberg SP, Fisher DO. 2013. Phylogenetic correlates of extinction risk in mammals : species in older lineages are not at greater risk. *Proc R Soc B* 280.
- Veron S, Davies TJ, Cadotte MW, Clergeau P, Pavoine S. 2015. Predicting loss of evolutionary history : Where are we ? *Biol. Rev.*:000–000.
- Wake DB, Wake MH, Specht CD. 2011. Homoplasy: From Detecting Pattern to Determining Process and Mechanism of Evolution. *Science* (80-.). 331:1032–1035.
- Waldron A, Mooers AØ, Miller DC, Nibbelink N, Redding D, Kuhn TS. 2013. Targeting global conservation funding to limit immediate biodiversity declines. *Proc. Natl. Acad. Sci.* 110:1–5.
- Weitzman ML. 1998. The Noah's Ark Problem. *Econometrica* 66:1279–1298.
- Wilman H, Belmaker J, Simpson J, de la Rosa C, Rivadeneira MM, Jetz W. 2014. EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology* 95:2027.

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

Zoological Society of London. 2008. Edge of Existence programme. [accessed 2016 Jun 23]. <http://www.edgeofexistence.org>

Appendix A

Update to the three vertebrate trees used

Mammal trees

The list of species in 1000 tree of mammals used in this study conformed to that of Martyn et al. (2012). We calculated ED scores for 5319 species of mammals, which is highly correlated, with scores provided by Martyn et al. (2012).

Bird trees

I calculated ED scores using an updated distribution of phylogenies incorporating a total of 10,284 extant bird species as recognized by the IUCN, based on the distribution of "Hackett backbone" trees used by Jetz et al. (2014). Updates were done by Karen Magnusson-Ford, and primarily involved the splitting of one named species into two or more.

The phylogenies used by Jetz et al. (2014) were modified to be consistent with the 2016 IUCN Red List taxonomy (BirdLife International 2015). Thus (1) the scientific names of 561 species on the phylogeny were changed to match those of the IUCN Red List, (2) 136 species that were not recognized by the IUCN Red List were removed from the phylogeny, and (3) 427 species that are currently recognized by the IUCN Red List (BirdLife International 2015) as extant but are not on the Jetz *et al.* (2014) tree were added (details below). These changes were applied to the complete distribution of 10,000 trees, each which now include 10,284 species, fully consistent with the IUCN Red List taxonomy. This set of updated trees is available upon request.

Species additions: Of the 427 species added, 405 were species that had recently been elevated from subspecies to species status, i.e. splitting from a known sister species. In these cases, the new species was inserted into the phylogeny halfway down the branch of its sister. In cases where one species was split into more than two species, the topology of the new clade (including the sister already existing in the tree) was generated randomly (rtree in the R package ape) and all branch lengths were set as equal. This clade was then inserted into the tree, replacing the original sister branch and

the branch lengths of the pendant edges were increased such that the tree remained ultrametric.

The remaining 22 species added to the tree included 11 newly described species, and 11 species not recognized by Jetz *et al.* (2014). These 22 species were added to the tree as described above, where sister species were identified based on BirdLife species factsheets or recent scientific literature. The only exception was *Heliangelus zusii*, which was inserted halfway down the branch leading to the clade containing the *Taphrolesbia* and *Agelaiocercus* genera: BirdLife states in its fact sheet “*Heliangelus* spp. typically occur in cloud-forest and shrubbery at elevations of 1,200-3,400 m, mostly at 1,400-2,200 m. This species is probably more closely allied to *Agelaiocercus* and *Taphrolesbia*, and should be sought in humid or semi-arid habitat as high as 3,200 m from northwestern Venezuela to northern Peru (Kirchman *et al.* 2010).”

Amphibian trees

I used the amphibian tree from Isaac *et al.* (2012), which incorporated 5713 species. I applied the method described by Kuhn *et al.* (2011) to resolve polytomies in the original tree. Using BEAST version 1.7.5. (Drummond *et al.* 2012), sampling every 1000 trees and setting the burnin rate at 10%, 10,000 trees were retained as a posterior distribution of trees for further use.

Appendix B

Spectral density profile of phylogenetic trees

Lewitus and Morlon (2016a,b) have recently suggested that the spectral density profile may be a powerful description of tree shape. In particular, the authors suggest that, given the profile captures both among clade and through time patterns, it has distinct advantages over other more commonly applied but ad hoc metrics such as lc and gamma. I therefore explored if these metrics of tree shape might better predict the slope of ED loss and PD loss relationship across various phylogenies. In addition to two well-known metrics of tree shape (Colless index and γ statistics), I measured the spectral density profile for each phylogeny using the *RPANDA* package of R. The *spectR* function computes the distribution of eigenvalues for each tree and returns eigenvalues, principal eigenvalue, asymmetry, peakedness (kurtosis and peak height) and eigengap of a tree (Morlon et al. 2016). Here I focus on asymmetry, kurtosis and the y-axis height of the spectral density profile. These three measures can be computed in normalized and raw versions, with the former normalizing scores based on the corresponding degree matrix so it can be compared among phylogenies.

Asymmetry (skewness) was introduced as a measure of node distribution through time (Lewitus and Morlon 2016a). Unlike the Gamma statistic, asymmetry did not show much variation in four groups of simulated trees, though high-death trees showed more positive skewness (median = 0.62). Among empirical phylogenies, the birds' tree showed the highly positive skewness (median = 86.34).

As noted by Lewitus and Morlon (2016), asymmetry and gamma scores are weakly but positively correlated. I found the same pattern in my simulated trees whereby gamma could explain only a portion of variation in normalized asymmetry (Slope=0.14, $r^2=0.14$). Although gamma and raw asymmetry were also significantly related (Slope=-0.29, $r^2=0.04$), it was negatively: the goodness of fit was very weak in any case (Figure B3).

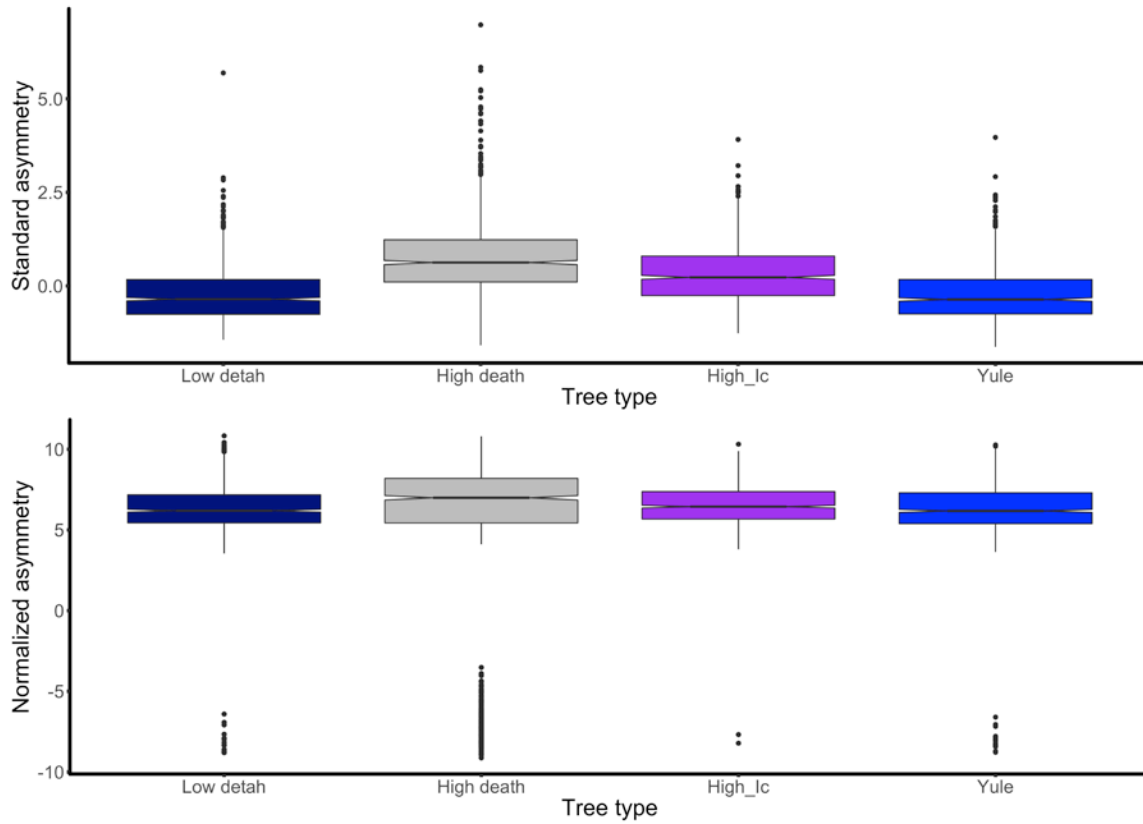


Figure B1. Box-plots of asymmetry (skewness) in four types of simulated trees; upper panel depicts the raw asymmetry and bottom panel shows the normalized version

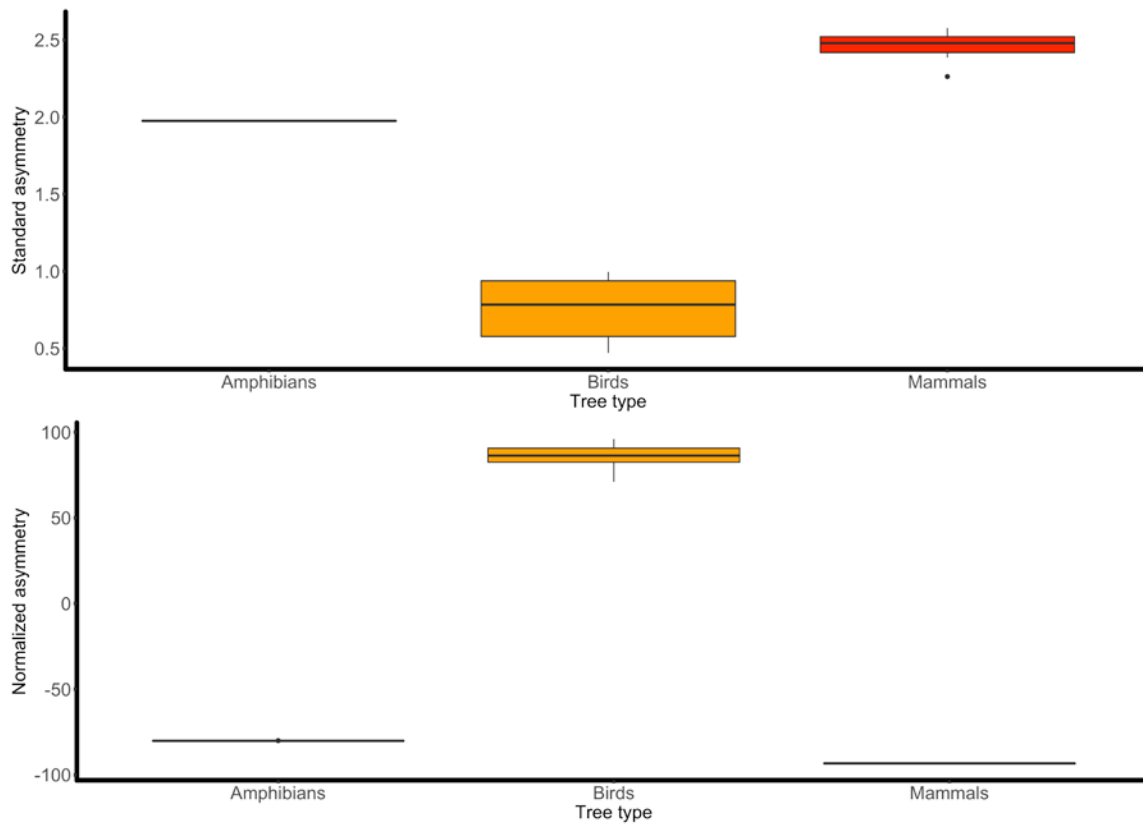


Figure B2. Values of asymmetry (skewness) in empirical phylogenies (10 trees from each group). The top panel shows raw values and the bottom panel depicts normalized values of asymmetry

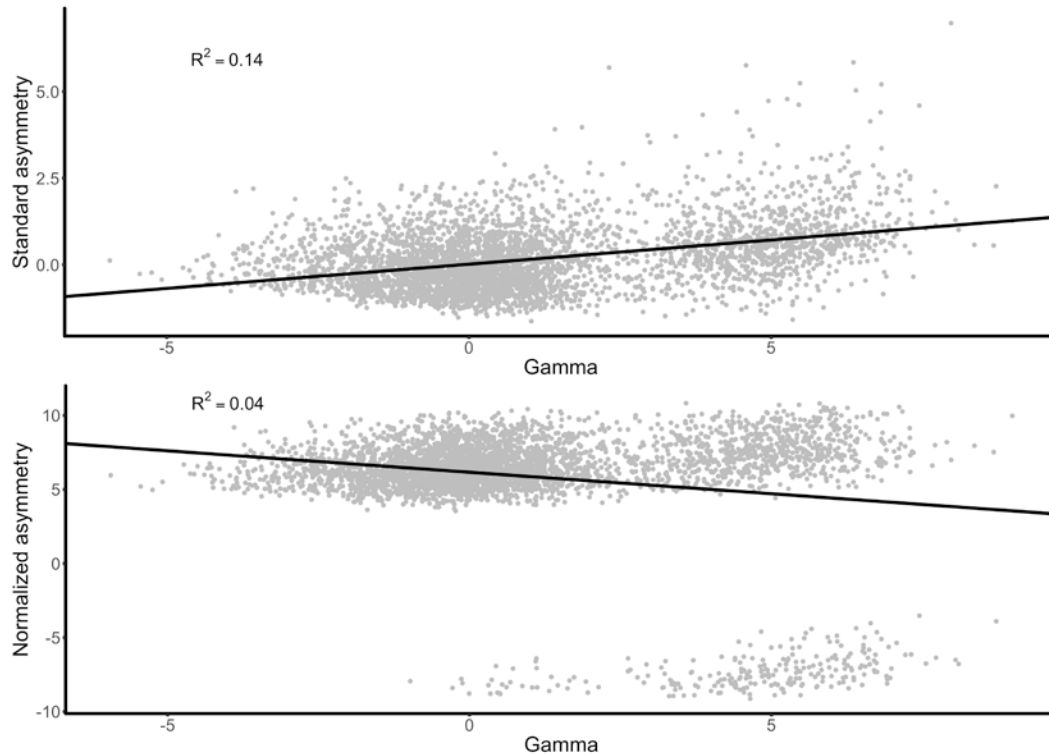


Figure B3. Scatter-plots of asymmetry and gamma values in four types of simulated trees; the upper panel depicts the raw asymmetry and the bottom panel shows the normalized version

Two measures of imbalance were proposed by Lewitus and Morlon (2016): The kurtosis of the spectral density, which quantifies the peakedness of a phylogenetic tree where higher peakedness implies heterogeneity of branch lengths; and the largest y-axis value of the spectral density profile (or peak height). Lewitus and Morlon (2016) showed that kurtosis was not correlated with the Colless index of tree imbalance. Here, however, it seems that kurtosis (both the raw and normalized versions) are significantly related to Colless index across my simulated tree set, though the strength of this relationship is very weak (Figure B4). On the other hand Colless index is more strongly and negatively correlated with the raw peak height ($r^2=0.31$).

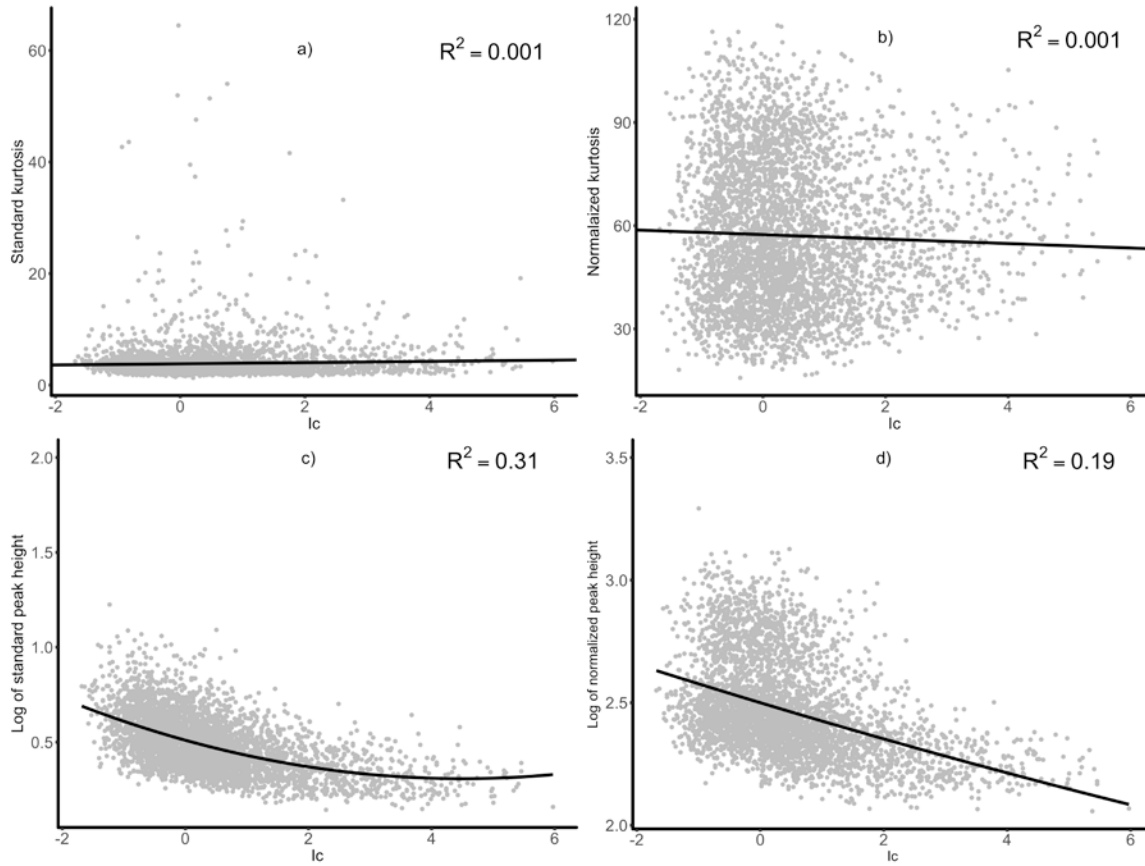


Figure B4. Relationship between Colless index and various versions of imbalance introduced by Lewitus and Morlon (2016): a) Raw kurtosis, b) Normalized kurtosis, c) Log-transformed raw peak height and d) Log-transformed normalized peak height, each across four types of simulated trees (high and low gamma, high and low I_c , $n=64$, $n=4000$ trees)

Interestingly, I found a very strong correlation between *gamma* and normalized peak height of the spectral density profile across my simulated trees ($r^2 = 0.85$). I also found kurtosis scores to be very strongly related to gamma despite the fact that kurtosis was primarily presented as a measure of imbalance (Lewitus and Morlon 2016a). Plots of kurtosis and peak scores in simulated trees vs. gamma are depicted in figure B5. All four variants of peakedness were found to be at their maximum in high-death trees (Figure B6) and as shown in Chapter 3 (Figure 2), high death trees exhibited the highest values of gamma among simulated trees.

These findings suggest that peakedness (i.e. normalized peak height) is more representative of tree stemminess, Lewitus and Morlon (2016) who suggested it as a measure of imbalance. These preliminary patterns highlight that more work is sorely

needed to characterize what the shape of the spectral density profile of a phylogenetic tree actually captures.

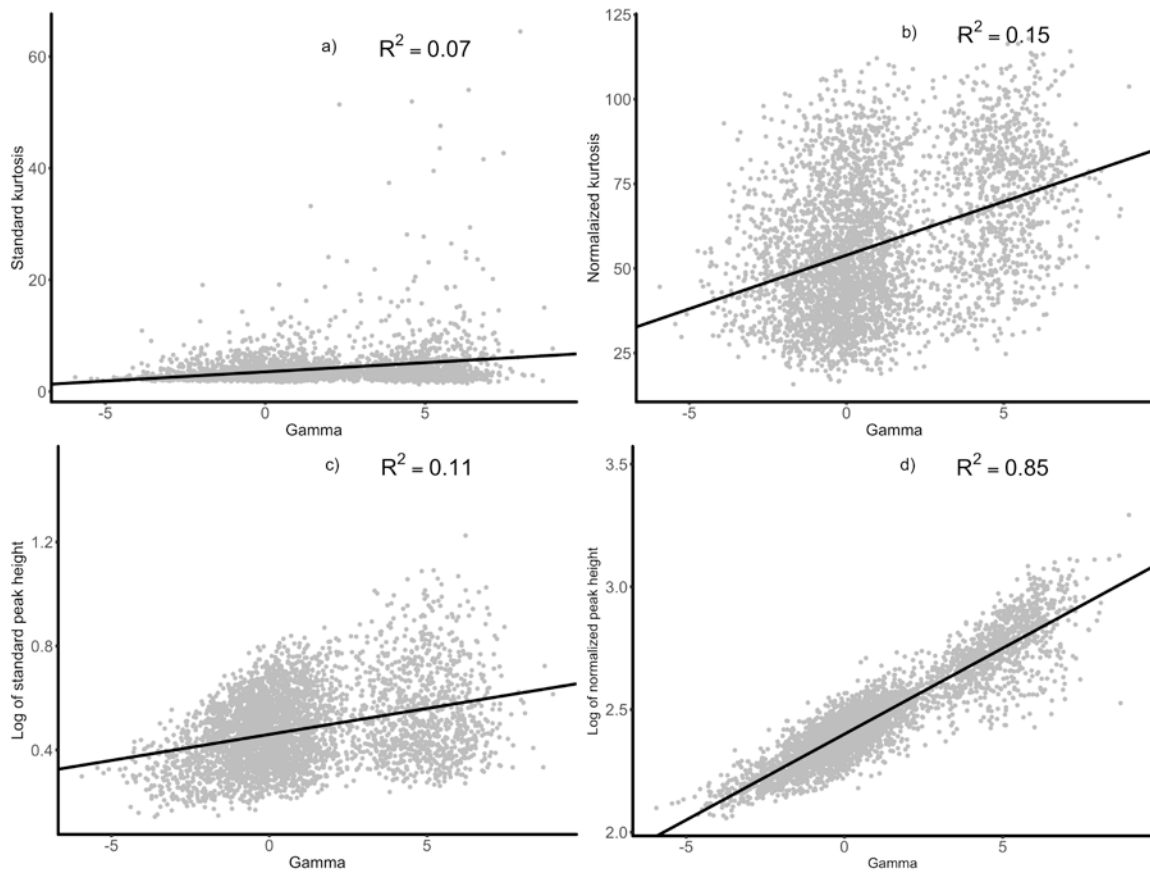


Figure B5. Relationship between gamma values and various versions of imbalance introduced by Lewitus and Morlon (2016): a) Raw kurtosis, b) Normalized kurtosis, c) Log-transformed raw peak height and d) Log-transformed normalized peak in 4000 simulated tree

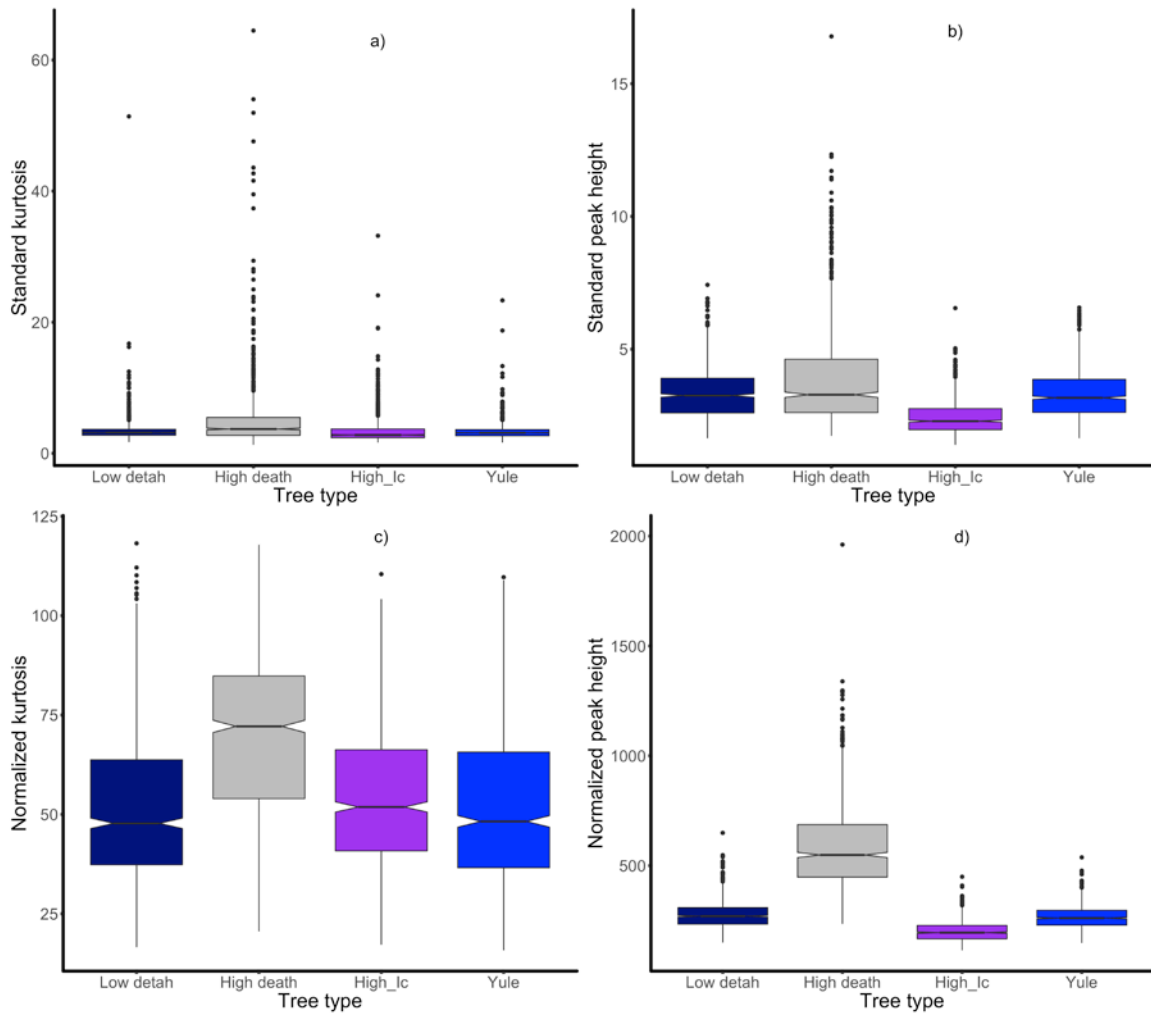


Figure B6. Distribution of two metrics of imbalance in four groups of simulated trees a) Raw kurtosis, b) raw peak height and, c) Normalized kurtosis and d) Normalized peak height

Appendix C

Annotated R code for simulating traits and calculating rarity

```
# Written by Vahab Pourfaraj and optimized by Karen Gordon
#Updated on April 2017 (Vahab Pourfaraj)

library("phytools")
Library("caper")

trees<- read.tree("") #import the multiphylo object here

#simulation of traits under Markov model
##Define number of states (k) and the rate of character change

## 2 states q=0.002
Q1<- matrix (c(-0.002,0.002,0.002,-0.002),2,2)
##2 states q=0.004
Q2<- matrix (c(-0.004,0.004,0.004,-0.004),2,2)
###2 states q=0.008
Q3<- matrix (c(-0.008,0.008,0.008,-0.008),2,2)
###2 states q=0.016
Q4<- matrix (c(-0.016,0.016,0.016,-0.016),2,2)

Q.all<- list(Q1, Q2, Q3, Q4))

avg_rarity.multi<- as.data.frame(matrix(nrow=length(trees[[1]]$tip.label),ncol=length(trees)))
row.names(avg_rarity.multi)<- trees[[1]]$tip.label[order(trees[[1]]$tip.label)]
filenames.rarity<- paste("Rarity_Q", 1:length(Q.all), ".csv", sep="")

for(l in 1:length(Q.all)){
  for(k in 1:length(trees)){

    tree<- trees[[k]]
    Q<- Q.all[[l]]

    file1<- sim.history (tree, Q, nsim=n) # n is the desired number of
traits to be simulated

    df<- getStates(file1, type="tips")

    df2 <- as.data.frame(matrix(nrow=nrow(df),ncol=length(file1)))
```

```

    for(i in 1:length(file1)){ ### changes according to the number of
traits (file1,nsim)
      for(j in 1:nrow(df)){ ###changes according to the number of speci
es in the phylogeny
        x <- df[j,i]
        num <- sum(df[,i] == x) - 1 #this counts the number of similar
states for that trait, discounting the species of interest
        rarity <- 1 - num/(nrow(df)-1) #formula; the number at the bott
om changes according to the number of taxa (N-1)
        df2[j,i] <- rarity

      }
    }
    average_rarity <- rowMeans(df2, na.rm=TRUE)

    if (l==1) {

      total<- as.data.frame(cbind(tree$tip.label, average_rarity))
      names(total)<- c("Tip.Label","Rarity")
      total<- total[order(total[, "Tip.Label"]),]

      avg_rarity.multi[,k]<- total[, "Rarity"]

    } else {
      total<- as.data.frame(cbind(tree$tip.label, average_rarity))
      names(total)<- c("Tip.Label","Rarity")
      total<- total[order(total[, "Tip.Label"]),]
      avg_rarity.multi[,k]<- total[, "Rarity"]

    }

  }

  write.csv(avg_rarity.multi, file=filenames.rarity[[1]])
}

```

Annotated R code for calculating loss of ED and loss of PD

```

library(phytools)

library(data.table)

EDtable<-read.table("") #Import ED scores
ED_df<- as.data.frame.matrix(EDtable) #converts it to df format
idx <- lapply(integer(ncol(ED_df)-1), function(...) sample(ED_df$V1, n)
)#n should change according to the number extinct species
setDT(ED_df)
setkey(ED_df, V1) # Set the key for the data.table to V1

Results = lapply(1:length(idx), function(i){
  ED_df[idx[[i]], lapply(.SD, cumsum), .SDcols = i + 1]
})
cumEDloss = do.call(cbind, Results)
write.csv(cumEDloss, "") # The output file for cumulative loss values
#####loss of phylogenetic diversity#####

#gets PD scores and stores it, this is before any pruning.
full_trees<- read.tree("") #Import the multiphylo object
PD_FULL<-matrix(rep(NA,length(full_trees)[1],1),nrow=length(full_trees)
[1],ncol=1) # place to store the PD values
for (s in 1:length(full_trees)[1]) #goes through all trees and gets the
PD value for each tree before pruning
{
  full_tree<-full_trees[[s]]
  PD_FULL[s] <-sum(full_trees[[s]]$edge.length)
}
write.csv(PD_FULL, "") #The output file for the PD of each tree

changed_multiphylo=list()

for (a in 1:length(full_trees))
{
  one_original_tree<- full_trees[[a]]
  changed_multiphylo[[a]]<- bind.tip(full_trees[[a]],"outgroup1",edge.l
length = 0, position = 0) #adding an outgroup so when we lose all ingrou
p spp we dont get errors
  changed_multiphylo[[a]]<- bind.tip(changed_multiphylo[[a]],"outgroup2
",edge.length = 0, position = 0) #adding an outgroup so when we lose al
L ingroup spp we don't get errors
}

#storage for PD values after extinction events
PD_loss<-matrix(rep(NA,length(idx[[1]])*length(changed_multiphylo)),nro

```



```

w=length(idx[[1]]),ncol=length(changed_multiphylo)) # place to store the PD differences
for (j in 1:length(changed_multiphylo)) { #goes through all trees
  full_tree<-changed_multiphylo[[j]]
  for (i in 1:length(idx[[1]])) { #goes through each pruning event
    PD_loss [i,j]<- sum(full_tree$edge.length) - sum(drop.tip(full_tree, idx[[j]][1:i])$edge.length)
  }
}
write.csv(PD_loss,"") #The output file for cumulative PD loss values

```