Title page

Article title: Assessing the Effects Imputation on ED Values

Running head: Assessing the Effects Imputation on ED Values

Authors: K. Bodie Weedop[1], William D. Pearse[1]

[1] Department of Biology & Ecology Center, Utah State University, 5305 Old Main Hill, Logan UT, 84322

*To whom correspondence should be addressed: `will.pearse@usu.edu` or `kbweedop@gmail.com`

Word-count: 5680 (abstract, main text, acknowledgements, and references)

Abstract

Keywords:

Introduction

Evidence from the fossil record and present-day studies argue we are in the midst of, or entering, a sixth mass extinction (Barnosky et al. 2011; Ceballos et al. 2015), such that more species than ever are declining and/or in danger of extinction across a range of environments (Wake & Vredenburg 2008; Thomas et al. 2004). Habitat destruction (Brooks et al. 2002), invasive species (Molnar et al. 2008), climate change (Pounds et al. 2006), and disease (Lips et al. 2006) are some of the leading causes of species declines globally. Conservation biologists seek~~s~~ to reverse~~overcome~~ these declines and their detrimental effects on species populations, but in reality they have limited resources with which to do so.. This challenge, termed the ~~Even still, researchers and conservationists are confronted with~~ "Noah's Ark problem" (Weitzman 1998), is the basis for modern conservation prioritisation and triage (XXX define triage).~~or an unfortunate reality of insufficient, finite resources to confront the increasing amount species requiring conservation effort.~~

Conservation triage has provided an efficient decision making process for allocating finite resources to obtain the greatest return (Bottrill et al. 2008). A sentence about how triage requires decision-making, and so you need a metric to guide that decision process. One of these triage strategies which have been introduced and used most widely is the EDGE metric (Evolutionary Distinction and Globally Endangered; Isaac et al. 2007). This method prioritizes species according to two metrics: E~~e~~volutionary EvolutioDdistinctiveness (ED) and G~~g~~lobal E~~e~~ndangerment (GE). ED measures relative contributions to phylogenetic diversity made by each species within a particular clade (Isaac et al. 2007). Such contributions are assessed by quantifying the amount of branch length which is unique to each species within the overall phylogeny. GE values are assessed by assigning numerical values to each of the World Conservation Union (IUCN) Red List Categories. As species become increasingly threatened and are placed into more concerning categories (e.g.,e.g. from Vulnerable to Endangered), the GE numerical value increases. Increases in either ED or GE place a particular species at a higher priority for conservation effort.

In the event of missing DNA or trait data, species are often difficult or not able to be placed onto a phylogeny. Even in the face of such uncertainty and missing data, it is understandable that conservation biologists want to make prioritizations. However, if we are using a quantitative method for prioritizing species, we should remain consistent even when uncertainty arises. To our knowledge, a proper and efficient method for prioritizing species where there is missing data is still untested. This issue pertains mainly to the calculation of ED than GE. The IUCN has collected data on most major clades, and has a strategy for assigning Red Listing values these species which we know little information and are considered Data Deficient (DD). IUCN and other conservation organizations support focus on DD species just the same as Critically Endangered and Endangered species to ensure consistency (Rodrigues et al. 2006). The major area of uncertainty in phylogenetic prioritisation is phylogenetic data. In the past, missing species data and poorly resolved trees have been addressed using imputation (Collen et al. 2011; Isaac et al. 2012; Jetz et al. 2014). However, to our knowledge, there has been no systematic investigation of the efficacy of such imputation, both in terms of the accuracy with which imputed ED values are estimated, and the effect on other known species' scores. Indeed, it is unclear whether any significant information on ED is gained by imputing species which cannot be placed on the phylogeny. It is also not well understood how simply removing missing species, compared to performing imputation, would effect ED values. It may be that simply excluding missing species may be less intrusive than imputation. In searching for a solution for missing species, we may be negatively affecting correct ED values and disrupting EDGE rankings in the process. As the desire to use ED and phylogenies for conservation triage grows, the importance of such tests and a consensus on how to resolve cases of phylogenetic uncertainty becomes more urgent.

Here we assess the extent to which EDGE rankings based upon imputed phylogenies can be used within applied conservation biology. To do this, we use an imputation approach... ~~Here,~~

~~we assess and compare the impact that missing species versus phylogenetic imputation has upon ED values.~~ In doing so, we hope to understand the effect that both methods have on ED values and offer a viable solution for dealing with missing data species. Missing species were simulated and removed from trees in two ways: randomly and in a phylogenetically biased manner. Additionally, we tested how ED values were affected by resolving and imputing polytomies of varying sizes on a phylogeny. We found that ED values are by removing missing species. Here is a reference to figure.

Methods

Here we use a simulation approach to test the effect of removing and imputing species on a phylogeny on subsequent ED (Evolutionary Distinctiveness) scores of species.~~We attempted to demonstrate the effects of removing or imputing species which cannot be placed onto a phylogeny.~~ Since empirical studies do not (to our knowledge) impute GE (Global Endangerment) scores for species, and our focus here is on the importance of phylogenetic structure, we focus on the impact of imputing ED values.~~While assessing the impact of dropping species and phylogenetic imputations, we were primarily focused on ED values. In testing the effects on these values, we remain focused upon ED as a single variable. We exclude GE because it would only add complexity while not providing any additional information to our particular investigation.~~

~~In each test, we simulated 100 phylogenetic trees and manipulated each tree's tips or clade.~~ All simulations~~, calculations,~~ and analyses were performed using R (version 3.4.0; R Core Team 2017). For each combination of parameter values in a simulation, we performed 100 replicate simulations. Original and manipulated trees were simulated under a pure-birth Yule model using the `sim.bdtree`~~sim.bdtree~~ function `geiger`~~‘geiger’~~ R package (Harmon et al. 2007). This particular model was chosen to maintain simplicity. Results from this simple model should be applicable to other more complex scenarios. Also, to reduce uncertainty, we used the same model throughout each of the simulations. In reality, we would be estimating the parameters of the model which the phylogeny is built upon. The function `ed.calc`~~ed.calc~~ within the R package `caper`~~'caper'~~ was used to calculate ED values for each tree (Orme 2013).

~~We assessed the impact that removing missing species has upon ED values using the correlation of all ED values for the tips remaining within both trees. To evaluate the effect that imputation has upon ED values, we calculated ED for all tips in both the original and manipulated trees while excluding the focal clade where imputation has occurred. These ED values were compared using a correlation. Additionally, we did the same calculations and comparison using only the original focal clade and its' simulated replacement. If missing species have no effect upon ED values, we expect a high, positive correlation coefficient between the original tree and its' manipulated counterpart.~~

Assessing the impact of missing species on EDGE-listing

Our first set of simulations assesses the impact that species missing from a phylogeny have on estimated ED scores. If, when species are missing from a tree, the ED scores of the remaining species in the tree are XXX this implies XXX. If missing species has a negative effect on ED values , then the correlation between the ED of the species in the original tree and the same tree with a fraction of tips removed in some manner should be significantly different from 1. We performed simulations on phylogenies of different sizes (number of taxa: XXX, XXX), removing constant fractions of tips from the tree (0%, 1%, 2%, ..., 19%, 20%). ~~To investigate the degree of this effect, we removed tips from simulated trees (Number of taxa = 64, 128, 256, 512, 1024, 2048, 4096) at random and by phylogenetic clusters. To assess the effect under varying amounts of uncertainty, fractions of tips dropped ranged from~~

~~0 to 0.2 of total tree tips.~~

Missing species at random was simulated by selecting species at random without replacing, and removing those species from the tree. This randomization had no regard for phylogenetic structure. Missing species related by some character trait was tested by simulating character trait values for each tip. These simulations were all performed under a constant rate Brownian-motion model (σ^2=0.5~~par = 0.5~~, starting root value = 1). Tips were dropped if their character trait values place them into the upper quantile which had been selected to be dropped. More specifically, if the fraction to be dropped was 0.1, species within the 90th quantile of character trait values are dropped. This is equivalent to Felsenstein's threshold model (Felsenstein & Felenstein 2004) state why this model is a useful one—re-state the propoerty it has, linking it back to why this is a useful set of simulations to be doing.

Assessing the impact of phylogenetic imputations

We tested the impact of imputing missing species onto a clade of a particular size (sizes 3, 4, 5, ..., 30, 31, 32~~3 through 32~~) which originated from a tree of a particular size (N~~n~~umber of taxa = 64, 128, 256, 512, 1024). To simulate the effect that phylogenetic imputation has upon these simulated trees, we randomly chose clades within each tree and treated it as a polytomy to be resolved. The clade selected was removed from the original tree and a new separate tree of the same size was simulated under the pure-birth model used before and placed back where the original clade was removed. Thus we have imputed each clade under the same model used to generate it. In an empirical study, this would be done by... and so our method is being generous because... By doing this , we replicated the process of imputation of a clade which has been resolved. ~~To ensure that this is representative of cases where imputation is used, 100 repetitions of this simulation were performed across different parameter settings.~~

To assess whether clades, once imputed, had similar ED scores, we ... We also looked at ranks, because... We statistically modelled these as a function of ..., hypothesising that each would matter because...

Results

While the random loss of species from a phylogeny does not appear to affect the ED values of the remaining species, phylogenetically-patterned loss does (Fig. 1 and Table **??**). Under both random and phylogenetically patterned loss, XXX increases with XXX, although the effect is much more extreme (XXX times; Table **??**) for XXX.~~ED values for remaining species were significantly affected by the fraction of species which were removed (Table 1). However, different effects are seen when dropping species at random and in clustered manner (Fig. 1). Dropping species at random has a reduced effect when compared to the effect which dropping species in a clustered manner has on remaining ED values (Fig. 1).~~

Results

Assessing the impact of missing species on EDGE-listing

Our results demonstrate that the comparison between ED values before and after species were removed was negatively affected by the fraction of species removed in both treatments (Table 1). Similar but slightly different effects are seen when dropping species at random and in clustered manner (Fig. 1). Interestingly, the rate at which ED values deviate from their true value is increased with species missing at random.

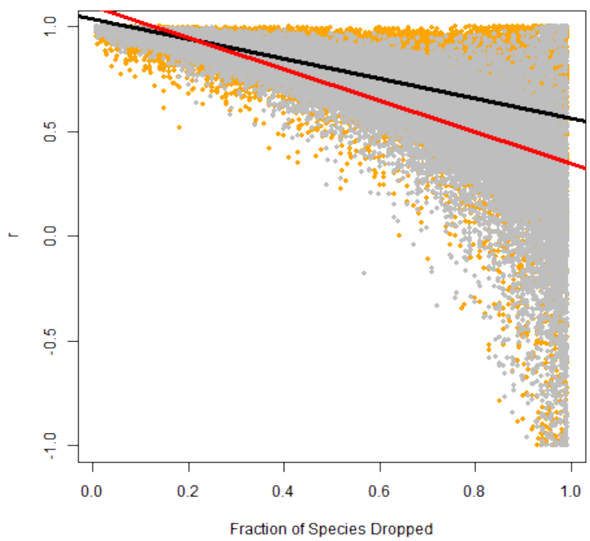

Figure 1: **R-values plotted against the fraction of species dropped at random versus clustered manner.** The color of data points denote whether species were dropped at random (orange; n = 100) or in clustered manner (grey; n = 100). The regression lines are demonstrating the relationship when species are dropped at random (red) and in a clustered manner (clustered). The correlations represent a comparison of the ED values (before and after species are dropped) of species which remain on the on the phylogeny after other species are dropped.

We find no support for a correlation between the imputed and true ED values for a species within an imputed clade (Fig. 2, table XXX). We do find evidence that, when imputing larger clades, the variation in the correlation is lesser (quantile regression), but this could be due to XXX. ~~ED values for the full tree while excluding the focal clade remain at 1 and unaffected. However, ED values for the imputed clades are significantly affected by the use of imputation. As the size of the focal clade increases, the informative value of the ED values within the clade decreases (Fig. 2). However, even when imputing smaller clades, ED values did not regularly reflect the true ED values (Table 2).~~ We found~~Our analysis demonstrates~~ that measures of the true~~original~~ phylogeny such as phylogentic diversity (PD), lambda, Colless' Index, skew, and kurtosis do not provide any indication that imputation would negatively affect ED values (Appendix A). ~~Additionally, j~~Just as imputed ED values did not reflect true ED values, the rankings of species within the focal clade were altered significantly

	Estimate	Std. Error	t value	$\Pr(>\lvert t\rvert)$
(Intercept)	1.0315	0.0013	821.39	<0.0001
Fraction of Species Dropped	-0.4696	0.0020	-233.16	<0.0001
Random Treatment	0.0630	0.0018	35.47	<0.0001
Number of Species Overall	0.0000	0.0000	7.89	<0.0001
Fraction of Species Dropped:Random Treatment	-0.2774	0.0028	-97.45	<0.0001
Random Treatment:Number of Species Overall	0.0000	0.0000	-4.38	<0.0001

Table 1: ANCOVA model summary describing the effect of dropping species on remaining species ED Values. The fraction of species dropped significantly affects the the remaining ED values. Dropping the fraction both at random and in clustered manner both have negative effects on the remaining ED values ($F_{139696,5} = 40350$, $R^2 = 0.5908$, p<0.0001).

under imputation (Fig. 3; Table XXX). Our model suggests that with increases in the size of the imputed clade and overall number of species, species within the clade are ranked farther from their true ranking (Table 3). For example~~More specifically~~, our model suggests that by imputing a clade of three species within a phylogeny of 128 species, the species within the clade would be 60 rankings from their true rank on average. ED values outside of the focal clade were not affected by imputation. While ED values within the focal clades were affected exclusively by imputation, a notable error rate in ranking crucial species correctly is present (Appendix B).

	Estimate	Std. Error	t value	$\Pr(>\lvert t\rvert)$
(Intercept)	0.1852	0.0533	3.47	0.0005
Size of Focal Clade	-0.0034	0.0002	-16.56	0.0000
Size of Phylogeny	-0.0001	0.0001	-0.41	0.6855
PD	0.0001	0.0001	0.37	0.7108
Lambda	-0.0012	0.0524	-0.02	0.9812
Colless' Index	0.0016	0.0022	0.72	0.4687
Skew	0.0043	0.0088	0.48	0.6288
Kurtosis	-0.0005	0.0009	-0.63	0.5269

Table 2: Effect of Clade Size on Imputed ED Values. The intercept describes that the correlation between the true and imputed values begins quite low. As the clade size increases, this correlation only tends toward zero. The total number of species in the full phylogeny along with measures of the true phylogenetic diversity, lambda, Colless' Index, skew, and kurtosis show no significant effect. ($F_{47992,7} = 39.57$, $R^2 = 0.006$, p<0.0001).

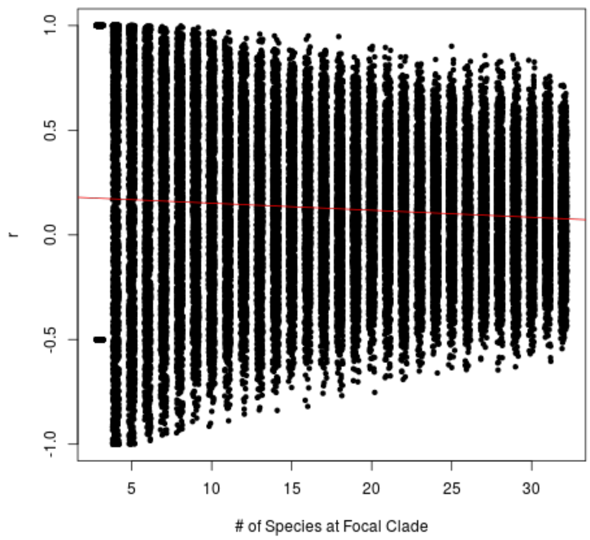

Figure 2: **R-values plotted against the number of species at focal clade.** Each data point denotes a correlative comparison between ED values within the focal clades where imputation has occurred. The regression line (red) and trend even closer to zero demonstrates the decrease in informative value of the imputed ED values. This is reinforced by the visual narrowing of r-values around zero.

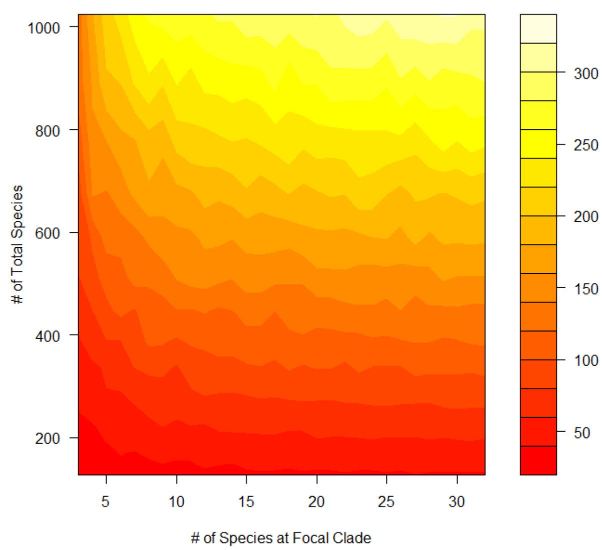

Figure 3: **Mean ranking error of species within the focal clade.** The gradient on the right demonstrates average number of posistions within the full ranking that focal clade species shifted from their true rank. While controlling for the size of the full phylogeny and focal clade, species within the focal clade were, on average, ranked far from the true rank.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.6344	0.0332	-49.29	0.0001
Size of Focal Clade	0.0900	0.0010	91.22	0.0001
Size of Phylogeny	0.5179	0.0013	383.99	0.0001

Table 3: Effect of Clade Size and Total Species on Ranking Error. Model demonstrating the relationship between focal clade species ranking error and the size of imputed clade and overall phylogeny. Square-root transformations have been applied to both ranking error and size of phylogeny. Significant increases ranking error are seen when increasing sizes of both the imputed clade and phylogeny ($F_{47997,2} = 77890$, $R^2 = 0.7644$, p<0.0001).

Discussion

Phylogenetic uncertainty and missing species are complications commonly encountered when applying ED for prioritization (Isaac et al. 2007). The aims of our investigation were to (1 determine how removing missing species from a phylogeny affects ED values for the remaining species and (2 demonstrate how imputation affects ED values of species where imputation is performed. Our results demonstrate that (1 missing a proportion of overall species both at random and in a phylogenetic-biased manner have different yet significant affects on remaining ED values throughout the tree and 2) imputation does not recover the ED value or ED rank of an imputed species.

[Moved; see comments] Our results are derived solely from simulations under a simple model of diversification—the Yule model. We do acknowledge that, in the real world, lineages evolve in more complex ways than are captured by such a simple model.~~We acknowledge that a Yule model of evolution is a simple model and more complex models may be present in empirical data~~. While we do not have empirical data, our simulations are performed under the same evolutionary model and therefore able to generalize for more complicated cases which might be seen in empirical data. However, our results demonstrate that even under a simple model imputing species lead to a misrepresentation of true ED values. Also, we have not been given any implication through this investigation that a more complex model would give any different results. Nomrally, imputation is averaged across all numerous trees to get the closest estimate of true ED values. However, imputed trees still deviate from the true trees and therefore the average would also be far from the true average.

Uncertainty in imputed species

Missing species and poor phylogenetic resolution have been identified as causes of uncertainty when calculating ED (Isaac et al. 2007). Prior to our investigation, we could not find any assessment of how missing species might affect ED values of species which are not missing. Our results demonstrate that it matters not just how many species are missing from a phylogeny when calculating an EDGE score, but also how those missing species are distributed across the phylogeny~~missing species at random or in a phylogenetically-biased manner can have different effects on ED values 1~~. Both manners of missing species cause ED values of species easily placed on the phylogeny to stray from their true values (Table 1). ~~Missing species at random~~ However, the manner (either randomly or phylogenetically-patterned) in which species are missing from the phylogeny is, to our knowledge, not normally investigated when deciding how to address them. In light of our results, this should be a concern when deciding how to move forward with missing species. We realize that we have provided just two ways in which species could be missing from a phylogeny and there are more that could occur. Missing species could be biased by some phylogenetic pattern other than Brownian motion evolution. Nevertheless, our investigation shows that missing species cause ED values of species remaining in the phylogeny to deviate from the true value.

In the past, we have included missing species into the EDGE framework using different methods. Collen et al. assigned the mean ED score of presumed congeneric species to the missing species 2011. More frequently, missing species and poorly resolved clades have been dealt with by imputing the missing species and assigning all the species of the resolved clade the mean ED value obtained from all possible or numerous resolutions of the clade (**Isaac2007; Isaac2012**). This method has been adapted by others and applied where there were large percentages (30%) of species missing (Jetz et al. 2014). While imputation does include missing species, research has shown imputed phylogenetic data leads to biases in some ensuing analyses (Rabosky 2014). When considering the analysis of ED, our results show that imputation does not recover true ED values nor ED rank of missing species **??**. Additionally, as the size of the imputed clade increases, ED values within the clade stray, on average, further from their true values. Even though we are including missing species into calculating ED, we are not obtaining accurate information about those species. While being uninformative, these ED values would lead to mispriortizing species based on our results 3. Analyzing the performance of imputation of a clade less than five species was not performed due to limitations of statistical validity (Crawley 2012).

Even so, our results provide no indication that the effects of imputation seen in our results would improve when applied at smaller clades.

Guidelines for the use of imputation

We suggest there is a straightforward synthesis of these results that should be useful in applied conservation biology.~~Combining the previous arguments provides an interesting case of when imputation could be useful.~~ Both random and phylogenetically-patterned loss of species affect ED values throughout the tree. However, we found that ED values of non-missing species remain relatively constant under imputation. Therefore, imputation provides an interesting solution to missing species biasing ED values of non-missing species. We have shown that imputing missing species would not provide accurate ED values for missing species. However, it would provide a method of avoiding the loss of species affecting ED values throughout the remainder of the tree. Basing conservation priorities upon the ED values of imputed species would lead to inaccurate prioritization. Nevertheless, imputation may be useful to stop missing species from biasing species easily placed on the phylogeny.

Given these results, we now present~~have developed several~~ guidelines for how missing species should be dealt with and when imputation might be appropriate when calculating EDGE. In the event of missing species, we should be investigating the amount of species that are missing and consider whether imputation is necessary. For some context, in our analysis we found that if 30% of species are missing at random or in a phylogenetic-biased manner from the phylogeny, respectively, 80% and 89% of the remaining ED values remained accurate. Therefore if species are missing, we should verify that the amount of missing species does not exceed a percentage which we have found to provide poor ED values for remaining species. While below an acceptable percentage of missing species, EDGE should be carried out without attempting to impute the missing species. However, if a larger proportion of species are missing, imputation may be used but with some caution. ED values for species easily placed on the phylogeny are relatively unaffected and can be trusted when setting conservation priorities. Nevertheless, ED values and ranks of species which have been imputed should either be ignored or used cautiously within EDGE. By following these guidelines we avoid biasing species which are easily placed on the phylogeny even in the event that imputation is used.

Acknowledgments

References

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., Maguire, K. C., et al. (2011). Has the Earth's sixth mass extinction already arrived? *Nature* 471.7336, 51–57.

Bottrill, M. C., Joseph, L. N., Carwardine, J., Bode, M., Cook, C., Game, E. T., Grantham, H., Kark, S., Linke, S., McDonald-Madden, E., et al. (2008). Is conservation triage just smart decision making? *Trends in Ecology & Evolution* 23.12, 649–654.

Brooks, T. M., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., Rylands, A. B., Konstant, W. R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G., et al. (2002). Habitat loss and extinction in the hotspots of biodiversity. *Conservation biology* 16.4, 909–923.

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science advances* 1.5, e1400253.

Collen, B., Turvey, S. T., Waterman, C., Meredith, H. M., Kuhn, T. S., Baillie, J. E., & Isaac, N. J. (2011). Investing in evolutionary history: implementing a phylogenetic approach for mammal conservation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 366.1578, 2611–2622.

Crawley, M. J. (2012). *The R book*. John Wiley & Sons.

Felsenstein, J. & Felenstein, J. (2004). *Inferring phylogenies*. Vol. 2. Sinauer associates Sunderland, MA.

Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., & Challenger, W. (2007). GEIGER: investigating evolutionary radiations. *Bioinformatics* 24.1, 129–131.

Isaac, N. J., Redding, D. W., Meredith, H. M., & Safi, K. (2012). Phylogenetically-informed priorities for amphibian conservation. *PLoS one* 7.8, e43912.

Isaac, N. J., Turvey, S. T., Collen, B., Waterman, C., & Baillie, J. E. (2007). Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PloS one* 2.3, e296.

Jetz, W., Thomas, G. H., Joy, J. B., Redding, D. W., Hartmann, K., & Mooers, A. O. (2014). Global distribution and conservation of evolutionary distinctness in birds. *Current Biology* 24.9, 919–930.

Lips, K. R., Brem, F., Brenes, R., Reeve, J. D., Alford, R. A., Voyles, J., Carey, C., Livo, L., Pessier, A. P., & Collins, J. P. (2006). Emerging infectious disease and the loss of biodiversity in a Neotropical amphibian community. *Proceedings of the national academy of sciences of the United States of America* 103.9, 3165–3170.

Molnar, J. L., Gamboa, R. L., Revenga, C., & Spalding, M. D. (2008). Assessing the global threat of invasive species to marine biodiversity. *Frontiers in Ecology and the Environment* 6.9, 485–492.

Orme, D. (2013). The caper package: comparative analysis of phylogenetics and evolution in R. *R package version* 5.2, 1–36.

Pounds, J. A., Bustamante, M. R., Coloma, L. A., Consuegra, J. A., Fogden, M. P., Foster, P. N., La Marca, E., Masters, K. L., Merino-Viteri, A., Puschendorf, R., et al. (2006). Widespread amphibian extinctions from epidemic disease driven by global warming. *Nature* 439.7073, 161–167.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Rabosky, D. L. (2014). No substitute for real data: phylogenies from birth-death polytomy resolvers should not be used for many downstream comparative analyses. *Xx2*, 1–23. arXiv: `1503.04978`.

Rodrigues, A. S., Pilgrim, J. D., Lamoreux, J. F., Hoffmann, M., & Brooks, T. M. (2006). The value of the IUCN Red List for conservation. *Trends in ecology & evolution* 21.2, 71–76.

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F., De Siqueira, M. F., Grainger, A., Hannah, L., et al. (2004). Extinction risk from climate change. *Nature* 427.6970, 145–148.

Wake, D. B. & Vredenburg, V. T. (2008). Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proceedings of the National Academy of Sciences* 105.Supplement 1, 11466–11473.

Weitzman, M. L. (1998). The Noah's ark problem. *Econometrica*, 1279–1298.

A. Effect of Measures of the True, Full Phylogenies

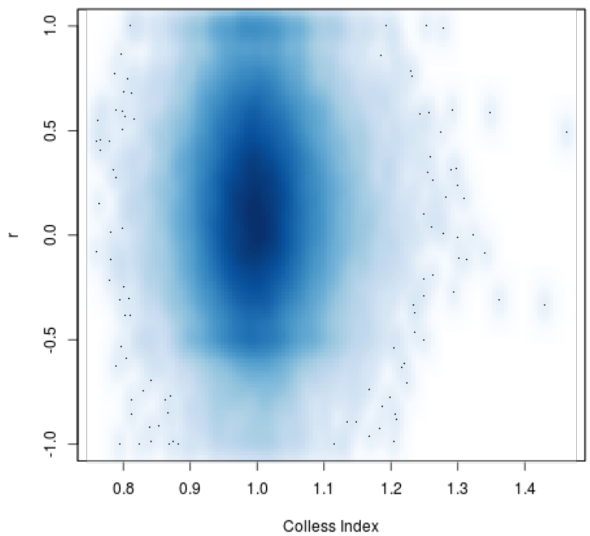

Figure 4: **Effect of the True Colless Index of FullPhylogeny.**

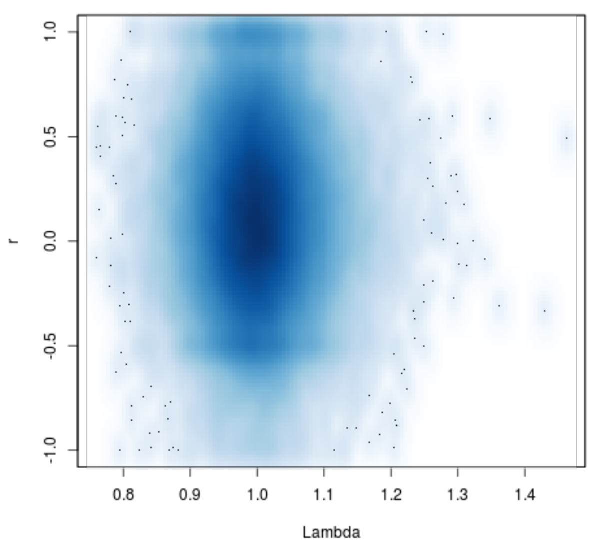

Figure 5: **Effect of the True Lambda of Full Phylogeny.**

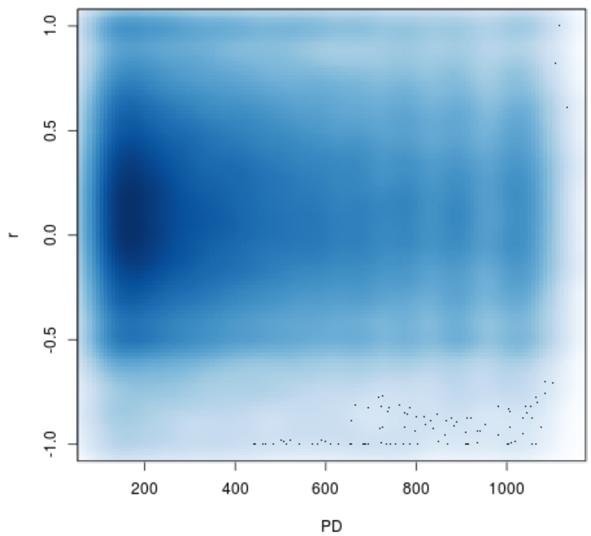

Figure 6: **Effect of True PD of Full Phylogeny.**

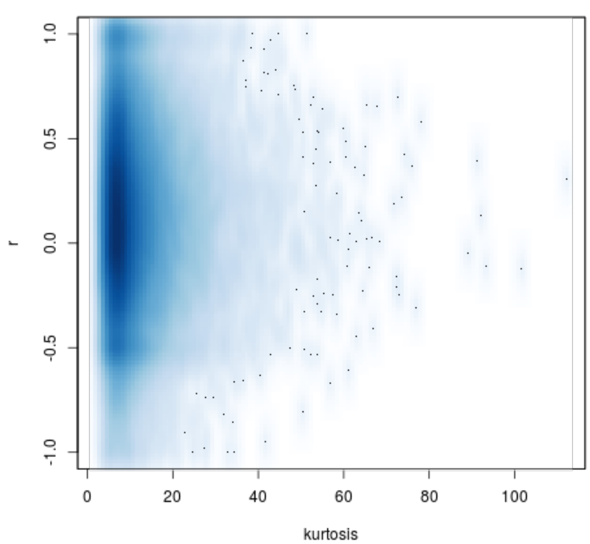

Figure 7: **Effect of the True Kurtosis of Full Phylogeny.**

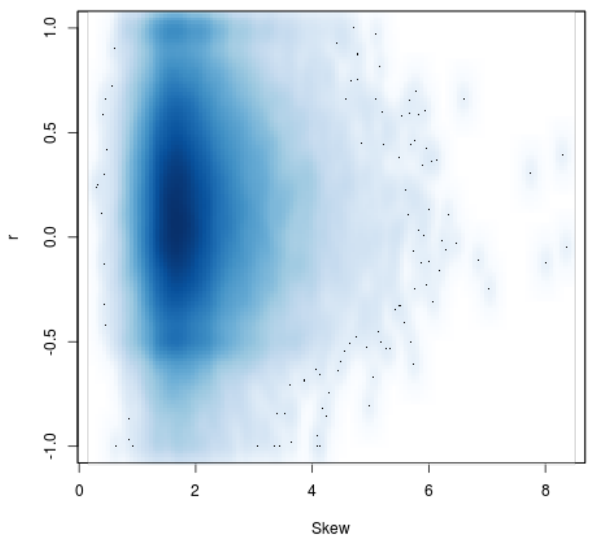

Figure 8: **Effect of the True Skew of Full Phylogeny.**

318 # B. Error Rate in Top Rankings

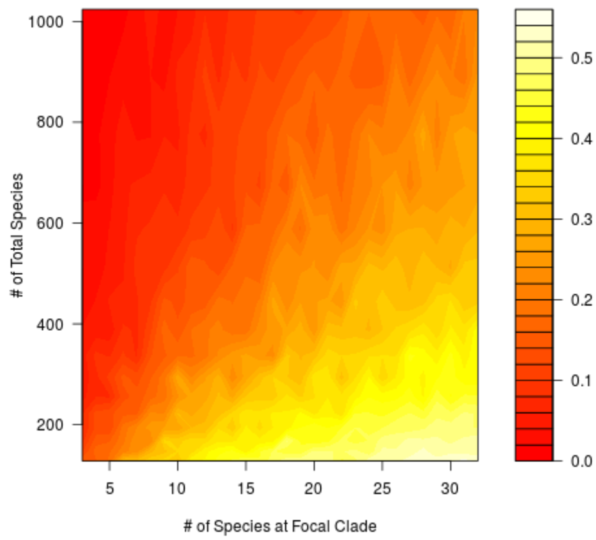

Figure 9: **Mean error rate in the ranking of top 50 species.**

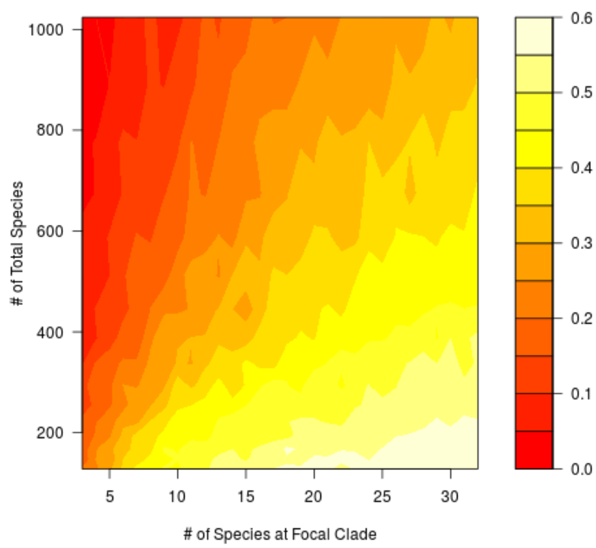

Figure 10: **Mean error rate in the ranking of top 100 species.**

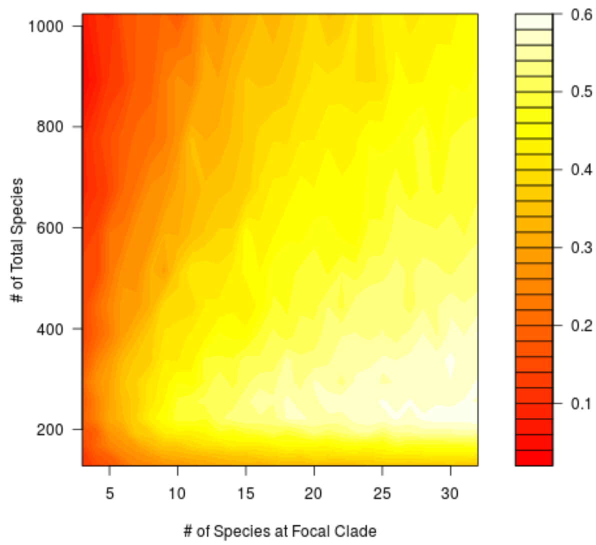

Figure 11: **Mean error rate in the ranking of top 200 species.**

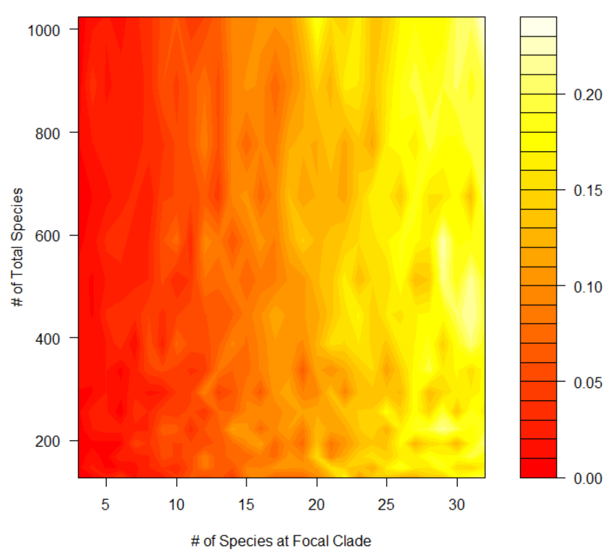

Figure 12: **Mean error rate in the ranking of top 5% of species.**

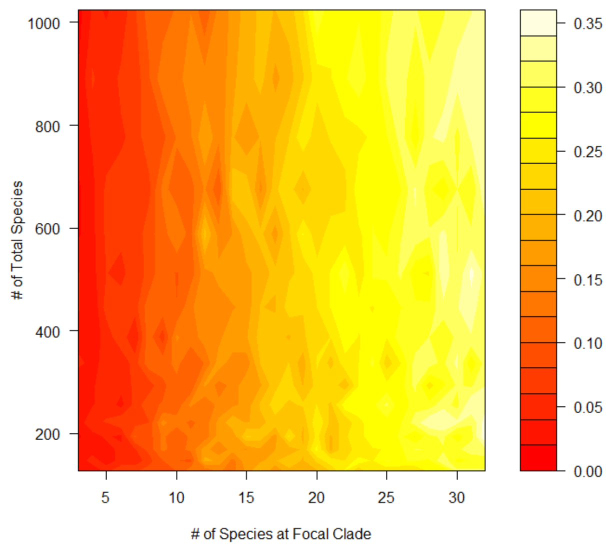

Figure 13: **Mean error rate in the ranking of top 10% of species.**

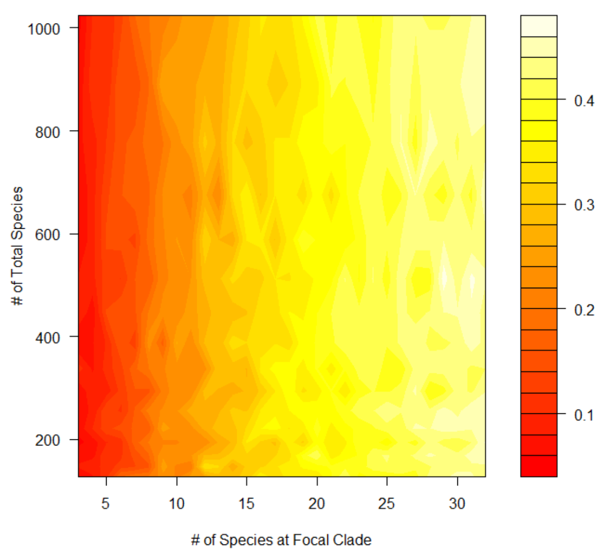

Figure 14: **Mean error rate in the ranking of top 20% of species.**