

# Title page

**Article title:** Assessing the Effects Imputation on ED Values

**Running head:** Assessing the Effects Imputation on ED Values

**Authors:** K. Bodie Weedop<sup>1</sup>, William D. Pearse<sup>1</sup>

<sup>1</sup> 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](mailto:will.pearse@usu.edu)

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

<sup>9</sup> **Abstract**

<sup>10</sup> **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 to reverse 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 “Noah’s Ark problem” (Weitzman 1998). The concept of conservation prioritization, or conservation triage, has provided an efficient method of allocating resources to confront this issue (Bottrill et al. 2008).

Conservation triage requires decision making supported by some metric quantifying the importance and degree of endangerment urgency for a set of species. By using such a metric, researchers are able to remove themselves and any potential bias from the allocation of time and resources for conservation. 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: Evolutionary Distinctiveness (ED) and Global Endangerment (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. These numerical values are assigned through transforming the IUCN Red List categories into probabilities of extinction (Mooers et al. 2008). As species become increasingly threatened and are placed into more concerning categories (*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.

EDGE has been applied in order to prioritize a number of species groups in the recent past since its' initial deployment. Some of these applications include applying EDGE to mammals (Isaac et al. 2007), amphibians (Isaac et al. 2012), and more recently for bird species (Jetz et al. 2014), corals (**Curnick2015**), and sharks (Stein et al. 2018). Alongside EDGE there have been a number of similar metrics which have been developed and some discussion of which should be used over others (Steel et al. 2007; Pearse et al. 2015). Apart from discussion of preferred metrics, there is also discussion of whether likelihood of success in conservation (Wilson et al. 2007; McBRIDE et al. 2007), relative cost of certain interventions(Naidoo et al. 2006), and complementarity of interventions(Pressey et al. 1993; Myers et al. 2000) should also be accounted for when designing metrics used going into the future. EDGE is limited in this regard; it does not include these nuanced factors from which it may benefit. However, EDGE has formed the basis for a successful program which addresses conservation in a quantitative manner. EDGE has provided actionable insights for conservation effort to be focused upon widely unknown species. It has established that phylogenetic conservation prioritization metrics can be used by conservation biologists and policy makers alike. Nonetheless, each application of EDGE has dealt with the uncertainty presented by missing species data.

In the event of missing DNA sequence data, species are often difficult or not able to be placed onto a phylogeny. Even in the face of such uncertainty, it is understandable that conservation biologists still want to make prioritizations. In fact, most cases where EDGE has been applied, taxonomic information, rather than sequence data, alone is what is used to place and impute species into phylogenies (**Curnick2015**; Isaac et al. 2007; Isaac et al. 2012; Jetz et al. 2014; Stein et al. 2018). By using the taxonomic information alone, researchers are able to produce fully resolved phylogenies using a birth-death polytomy resolver (Kuhn

et al. 2011). However, if we are using a quantitative method for prioritizing species, we should remain objective as possible and understand how these methods affect conservation priorities. 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 rather 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). 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.

Our aim is to test of how missing species data affects EDGE rankings and assess whether imputing species is a defensible method for dealing with species not easily placed on the phylogeny. The uncertainty of missing species data was replicated by removing species from simulated trees in two ways: at random and in a phylogenetically biased manner. By doing so, we hope to provide a model of how EDGE values are affected when species difficult to place on a phylogeny are excluded. Also, we assess the extent to which EDGE rankings based upon imputed phylogenies can be used within applied conservation biology. To do this, we replicate the imputation approach by randomly removing, replicating, and replacing one

particular clade all under the same model of evolution. In doing so, we hope to understand the effect that imputation has on ED values and offer some discussion of whether it is a viable solution for dealing with missing data species. We found that ED values throughout the tree are affected by excluding missing species from the phylogeny (Fig. 1). Also, we found no correlation between imputed and true ED values for a species within an imputed clade (Fig. 2; Table 2).

## 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. 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. EDGE is the product of both ED and GE, any true GE values assigned will remain biased by any inaccuracy in ED values due to imputation.

All simulations 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` function `geiger` R package (Harmon et al. 2007). This particular model was chosen because it is the simplest model possible; each branch is associated with only one value: birth rate. If excluding or imputing missing species shows to be ineffective under this particular model of evolution, it will most likely not be effective in more complex biological simulations. In order to calculate ED values, the function `ed.calc` within the R package `caper` was used to calculate ED values for each tree (Orme 2013).

## Assessing the impact of missing species on EDGE-listing

Our first set of simulations assesses the impact of random loss and phylogenetically-biased loss of species on ED scores. We test both manners of species loss in order to simulate both random selectivity and phylogenetic selectivity in missing species data. 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 ( $\sigma^2=0.5$ , starting root value = 1). Tips were dropped if their character trait values exceeded a relative threshold and placed them into the upper quantile for the set of trait values. More specifically, if the fraction to be dropped was 10%, species within the 90th quantile of character trait values were dropped. By doing so, we replicated the threshold model (Felsenstein 2005). This allowed us to take continuous character trait values, treat them as discrete character traits given the relative threshold, and assess how species which go missing due to a discrete trait would affect ED values of remaining species. Both sets of simulations were carried out using phylogenies of different sizes (number of taxa: 64, 128, 256, ..., 2048, 4096), removing constant fractions of tips from the tree (0%, 1%, 2%, ..., 19%, 20%).

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 coefficient of correlation between the original tree and its' manipulated counterpart.

## 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) which originated from a tree of a particular size (number of taxa = 64, 128, 256, 512, 1024). To simulate the effect that phylogenetic imputation has upon EDGE rankings, we randomly selected one clade to be removed from the original tree, simulated a new separate tree of the same size under the pure-birth model used before, and placed the newly simulated clade 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 assigning all the species of the resolved clade the mean ED value obtained from all possible or numerous resolutions of the clade. Therefore, our method is being generous because the numerous resolutions of that clade would be produced under a model which would have to be estimated. By doing this, we replicated the process of imputation of a clade which has been resolved.

To assess whether clades, once imputed, had similar ED scores, we correlated the imputed ED scores against true ED scores. We also looked at ranks to understand the amount of error introduced when imputing species. We statistically modelled these as a function of the number of total species and number of species within an imputed clade, hypothesizing that each would matter because the erroneous ranking of imputed species would range across the total amount of species depending on the amount of species imputed.

## Results

Under both random and phylogenetically-patterned loss, ED values of remaining species are affected by the total number of species removed (Table 1; Fig. 1). In either case, a similar effect on the correlation between remaining ED values before and after species are removed



161 is seen.

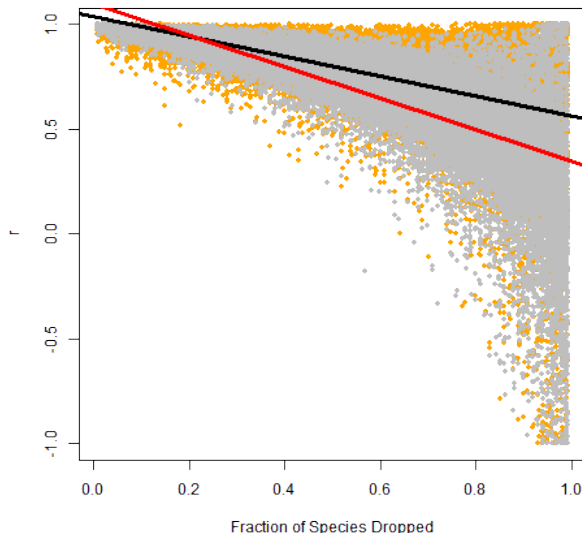
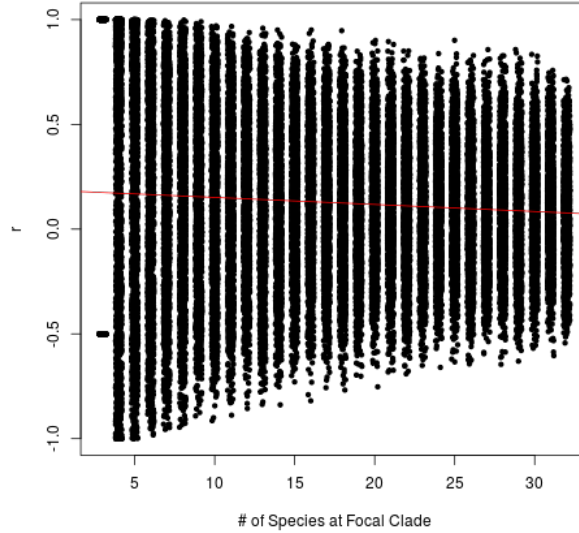


Figure 1: **R-values plotted against the fraction of species dropped at random versus clustered manner.** The color of data points denote whether species were dropped at random (orange;  $n = 100$ ) or in clustered manner (grey;  $n = 100$ ). The regression lines are demonstrating the relationship when 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 phylogeny after other species are dropped.

162 We find no support for a correlation between the imputed and true ED values for a species  
 163 within an imputed clade (Fig. 2, table 2). We found that measures of the true phylogeny  
 164 such as phylogenetic diversity (PD), lambda, Colless' Index, skew, and kurtosis do not provide  
 165 any indication that imputation would negatively affect ED values (Appendix A). We do find  
 166 evidence that, when imputing larger clades, the variation in the correlation is lesser (Fig. 3),  
 167 but the correlation between true and imputed ED values are converging to zero correlation  
 168 (Table 3). Just as imputed ED values did not reflect true ED values, the rankings of species  
 169 within the focal clade were altered significantly under imputation (Fig. 4; Table 4). Our  
 170 model suggests that with increases in the size of the imputed clade and overall number of  
 171 species, species within the clade are ranked farther from their true ranking (Table 4). Also,  
 172 this same effect is seen in the error rate within top ranking species (Appendix B).

	Estimate	Std. Error	t value	Pr(> t )
(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

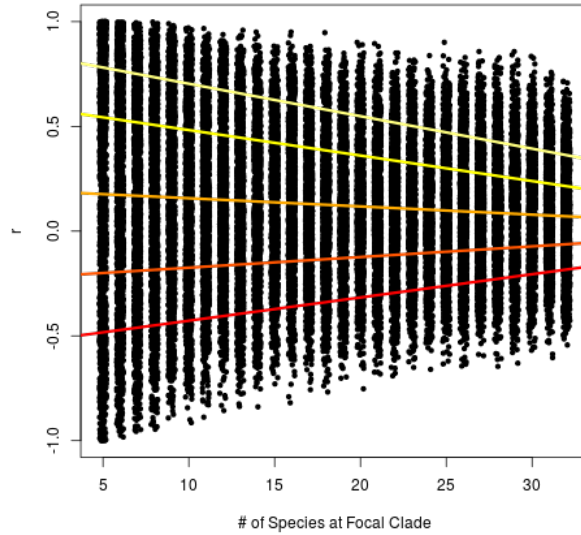
**Table 1: ANCOVA model summary describing the effect of dropping species on remaining species ED Values.** The fraction of species dropped significantly affects the the remaining ED values. Dropping the fraction both at random and in clustered manner both have negative effects on the remaining ED values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



**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.

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.1691	0.0500	3.38	0.0007
Size of Focal Clade	-0.0029	0.0002	-14.15	0.0000
Size of Phylogeny	-0.0001	0.0001	-1.01	0.3128
PD	0.0001	0.0001	0.97	0.3339
Lambda	0.0051	0.0492	0.10	0.9179
Colless' Index	0.0020	0.0021	0.96	0.3388
Skew	0.0039	0.0083	0.47	0.6409
Kurtosis	-0.0005	0.0008	-0.64	0.5247

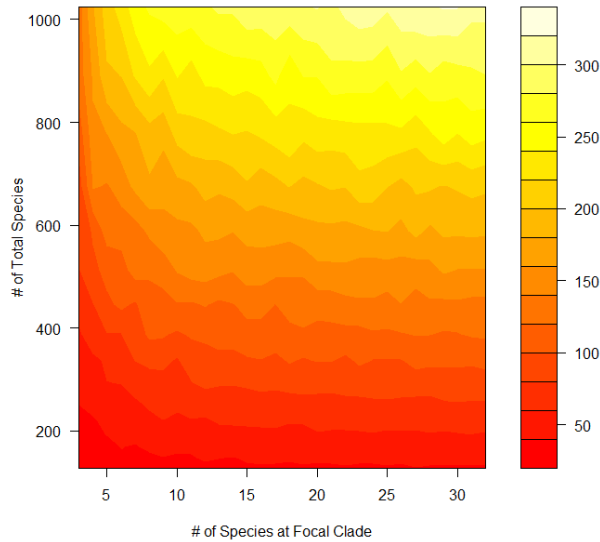
**Table 2: Effect of Clade Size on Imputed ED Values.** The intercept describes that the correlation between the true and imputed values begins quite low. As the clade size increases, this correlation only tends toward zero. The total number of species in the full phylogeny along with measures of the true phylogenetic diversity, lambda, Colless' Index, skew, and kurtosis show no significant effect. ( $F_{47992,7} = 29.38$ ,  $R^2 = 0.005$ ,  $p < 0.0001$ ).



**Figure 3: Quantile regression of r-values against size of imputed clades.** Each data point denotes a correlative comparison between ED values within the focal clades where imputation has occurred. Each regression line (top to bottom) represent quantile regressions from highest to lowest, respectively. Each of the regression lines demonstrate a convergence of the variation in r-values around zero.

	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
(Intercept)	-0.54	-0.23	0.20	0.60	0.86
Size of Focal Clade	0.01	0.01	-0.00	-0.01	-0.02

**Table 3: Quantile Regression of Clade Size and Total Species on Ranking Error.** Quantile regression model demonstrating the effect of clade size on the correlation between true and imputed ED values. The quantile regression estimates demonstrate statistical significance that as imputed clade size increases, variation in coefficient of correlation ( $r$ ) between ED values center around zero (all p-values are  $<0.0001$ ).



**Figure 4: Mean ranking error of species within the focal clade.** The gradient on the right demonstrates average number of positions 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 4: 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 conservation prioritization is an emerging tool providing a much needed, objective measure for conservation decision-making and policy. However, phylogenetic uncertainty is a major obstacle for EDGE and similar metrics (Collen2015). Uncertainty in phylogenies could mean that species in desperate need of conservation effort are overlooked. In order to address such uncertainty, we aim to determine how removing missing species from a phylogeny affects ED values for the remaining species and demonstrate how imputation affects ED values of species where imputation is performed. Our results demonstrate that missing a proportion of overall species both at random and in a phylogenetic-biased manner have different yet significant effects on remaining ED values throughout the tree and imputation does not recover the ED value or ED rank of an imputed species.

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. While we do not have empirical data, imputing species under this model should be much easier given the simplicity. We have not been given any implication through this investigation that a more complex model would produce any different result. Therefore, we suggest that our results generalize from the simplest case to 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.

Normally, imputation and artificially, fully resolved clades are achieved by generating a pseudo-posterior distribution of numerous trees, using a birth-death model of evolution, for clades where there are unsampled species using a polytomy resolver (Kuhn et al. 2011). However, our results demonstrate that there is uncertainty in that pseudo-distribution of trees being used to set priorities (Fig. 2).

## 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. In previous research, incomplete phylogenies were shown to produce nearly the same result as the later, more complete trees (Curnick2015). While demonstrating EDGE scores derived in the face of phylogenetic uncertainty still perform quite well in setting accurate priorities, they did not explicitly test the effect which missing species, nor imputing missing species, has on ED values. Our results support the finding that missing species do affect ED scores but the effect is relatively insignificant compared to percentages of species missing (Table 1).

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. This effect should be considered when debating whether under other, more complex models of evolution a different effect might be seen. There could be cases where species could be missing from the phylogeny with the result that an entire clade is unsampled. This would be a case of species radiating in situ and have persisted in one, restricted area. A situation such as this is seen with corals in the Indian Ocean region (Arrigoni2012).

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% or 3,330 species) missing (Jetz et al. 2014).

While imputation does include missing species, relying upon taxonomic information and constraints under some estimated model of evolution could become subjective. Kuhn et al. 2011 suggests that when using their polytomy resolver, the application of the of the resolver will be biased by the model of evolution estimated. More recent research confirms, that when under the prior of a birth-death model, imputation methods consistently bias evolutionary rates and estimates of phylogenetic signal used in downstream 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 (Fig. 2; Fig. 4). As the size of the imputed clade increases, ED values of imputed species do not correlate with their true values (Table 2). 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 also lead to misprioritizing species based on our results. In the event that a clade of 25 species in a phylogeny of 850 species, our results show those imputed species would be, on average, 250 ranks from their true rank (Fig. 2; Table 4). Analyzing the performance of imputation of a clade less than five species was not performed because we can't report a correlation reliably with so little data (Crawley 2012). Even still, we cannot see how the trend in our results would suddenly change such that, when imputing only one species, it improves.

## Guidelines for the use of imputation

We suggest there is a straightforward synthesis of these results that should be useful in applied conservation biology. Both random and phylogenetically-patterned loss of species

247 affect ED values throughout the tree. However, we found that ED values of non-imputed  
248 species remain constant under imputation. Therefore, imputation could be used to avoid  
249 missing species biasing ED values of non-missing species. We have shown that imputing  
250 missing species would not provide accurate ED values for missing species. However, it would  
251 provide a method of avoiding the loss of species affecting ED values throughout the remainder  
252 of the tree. Basing conservation priorities upon the ED values of imputed species would lead  
253 to inaccurate prioritization. Nevertheless, imputation may be useful to stop missing species  
254 from biasing species easily placed on the phylogeny.

255 Given these results, we now present guidelines for how missing species should be dealt with  
256 and when imputation might be appropriate when calculating EDGE. Our aim is to provide a  
257 rule of thumb for the use of imputation and how imputed species should be handled. In the  
258 event of missing species, researchers should be investigating the exact amount of species that  
259 are missing and consider whether imputation is necessary. Our analyses show that even with  
260 a large proportion of species missing, ED values of remaining species are highly correlated.  
261 For some context, in our analysis we found that 30% of species missing at random or in  
262 a phylogenetic-biased manner from the phylogeny leads to a correlation coefficient of 0.8  
263 and 0.89, respectively. Therefore if species are missing, we should verify that the amount  
264 of missing species does not exceed a percentage which we have found to provide poor ED  
265 values for remaining species. While below an acceptable percentage of missing species,  
266 EDGE should be carried out without attempting to impute the missing species. However, if  
267 a larger proportion of species are missing, imputation may be used but with some caution.  
268 ED values for species easily placed on the phylogeny are relatively unaffected and can be  
269 trusted when setting conservation priorities. Nevertheless, ED values and ranks of species  
270 which have been imputed should either be ignored or used cautiously within EDGE. Our  
271 results demonstrate that an individual species within an imputed clade of 25 species on a  
272 phylogeny of 1024 species is, on average, ranking  $\pm 318$  positions from its true rank (Fig.  
273 4). While imputation does not affect ED values of known species, includes missing species



274 in the phylogeny, it does not help understand the true ranking of an imputed species. By  
275 following these guidelines we avoid biasing species which are easily placed on the phylogeny  
276 and confounding the rankings for the sake of including missing species.

## 277 **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. (2005). Using the quantitative genetic threshold model for inferences between and within species. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 360.1459, 1427–1434.
- 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.
- Kuhn, T. S., Mooers, A. Ø., & Thomas, G. H. (2011). A simple polytomy resolver for dated phylogenies. *Methods in Ecology and Evolution* 2.5, 427–436.
- 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.
- McBRIDE, M. F., Wilson, K. A., Bode, M., & Possingham, H. P. (2007). Incorporating the effects of socioeconomic uncertainty into priority setting for conservation investment. *Conservation Biology* 21.6, 1463–1474.
- 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.
- Mooers, A. Ø., Faith, D. P., & Maddison, W. P. (2008). Converting endangered species categories to probabilities of extinction for phylogenetic conservation prioritization. *PloS one* 3.11, e3700.
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403.6772, 853–858.
- Naidoo, R., Balmford, A., Ferraro, P. J., Polasky, S., Ricketts, T. H., & Rouget, M. (2006). Integrating economic costs into conservation planning. *Trends in ecology & evolution* 21.12, 681–687.
- Orme, D. (2013). The caper package: comparative analysis of phylogenetics and evolution in R. *R package version* 5.2, 1–36.

Pearse, W. D., Chase, M. W., Crawley, M. J., Dolphin, K., Fay, M. F., Joseph, J. A., Powney, G., Preston, C. D., Rapacciuolo, G., Roy, D. B., et al. (2015). Beyond the EDGE with EDAM: prioritising British plant species according to evolutionary distinctiveness, and accuracy and magnitude of decline. *PloS one* 10.5, e0126524.

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.

Pressey, R., Humphries, C., Margules, C. R., Vane-Wright, R., & Williams, P. (1993). Beyond opportunism: key principles for systematic reserve selection. *Trends in ecology & evolution* 8.4, 124–128.

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.

Steel, M., Mimoto, A., & Mooers, A. Ø. (2007). Hedging our bets: the expected contribution of species to future phylogenetic diversity. *Evolutionary bioinformatics online* 3, 237.

Stein, R. W., Mull, C. G., Kuhn, T. S., Aschliman, N. C., Davidson, L. N., Joy, J. B., Smith, G. J., Dulvy, N. K., & Mooers, A. O. (2018). Global priorities for conserving the evolutionary history of sharks, rays and chimaeras. *Nature Ecology & Evolution*, 1.

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.

Wilson, K. A., Underwood, E. C., Morrison, S. A., Klausmeyer, K. R., Murdoch, W. W.,  
Reyers, B., Wardell-Johnson, G., Marquet, P. A., Rundel, P. W., McBride, M. F., et al.  
(2007). Conserving biodiversity efficiently: what to do, where, and when. *PLOS biology*  
5.9, e223.

## A. Effect of Measures of the True, Full Phylogenies

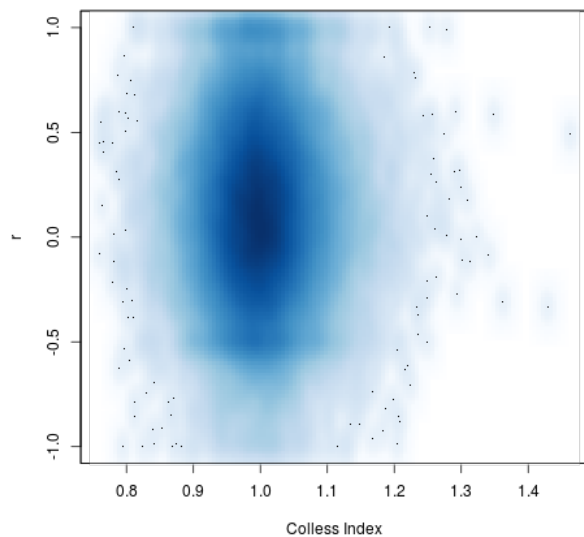


Figure 5: **Effect of the True Colless Index of FullPhylogeny.**

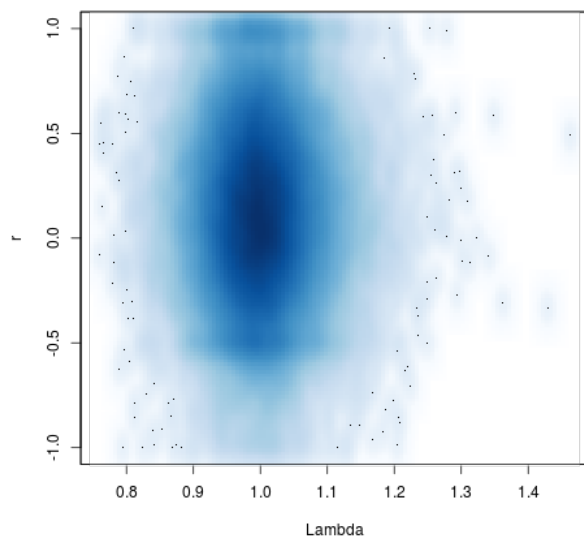


Figure 6: **Effect of the True Lambda of Full Phylogeny.**

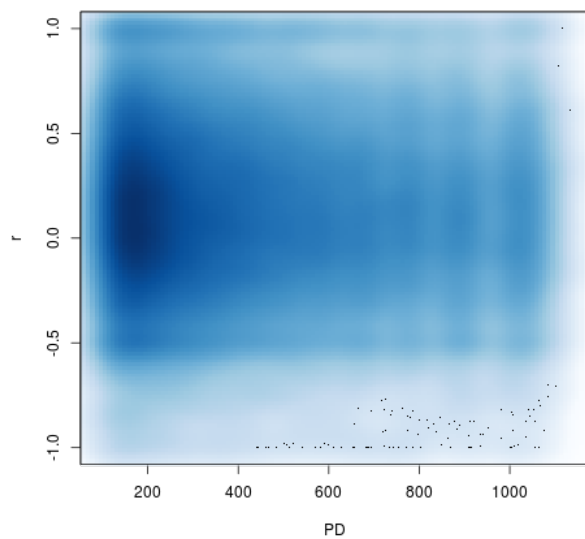


Figure 7: **Effect of True PD of Full Phylogeny.**

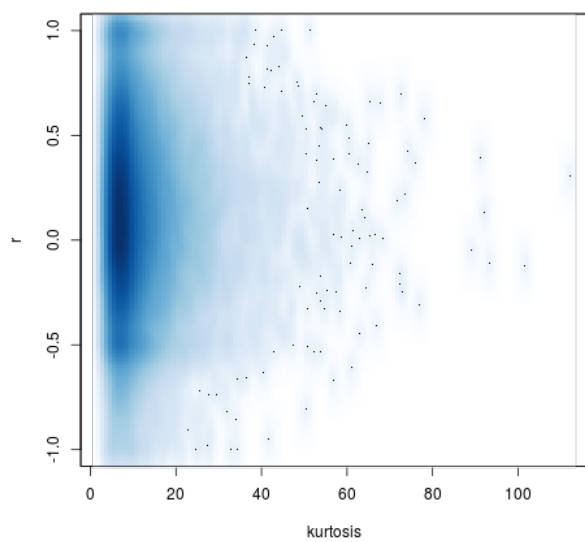


Figure 8: **Effect of the True Kurtosis of Full Phylogeny.**

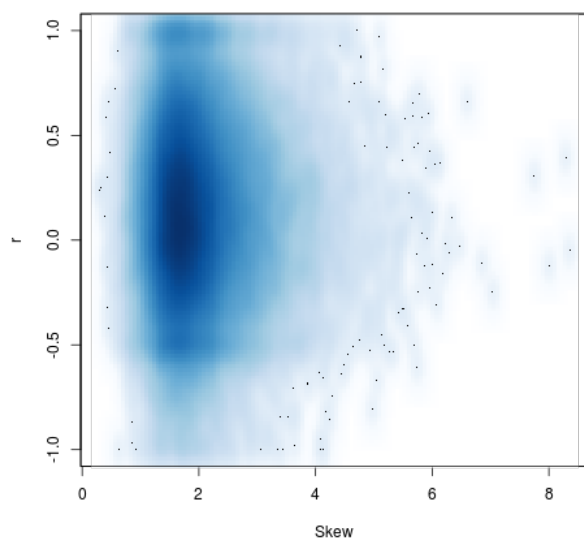


Figure 9: **Effect of the True Skew of Full Phylogeny.**



## 366 B. Error Rate in Top Rankings

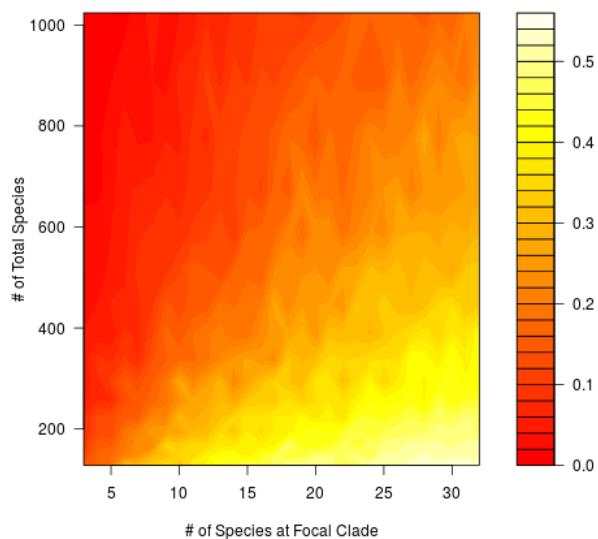


Figure 10: Mean error rate in the ranking of top 50 species.

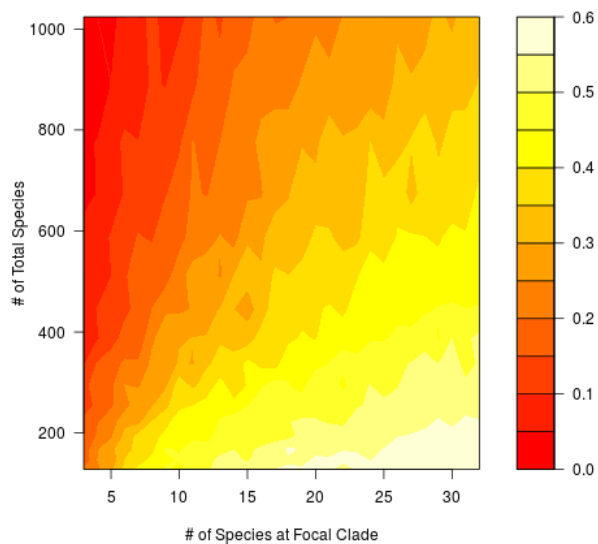


Figure 11: Mean error rate in the ranking of top 100 species.

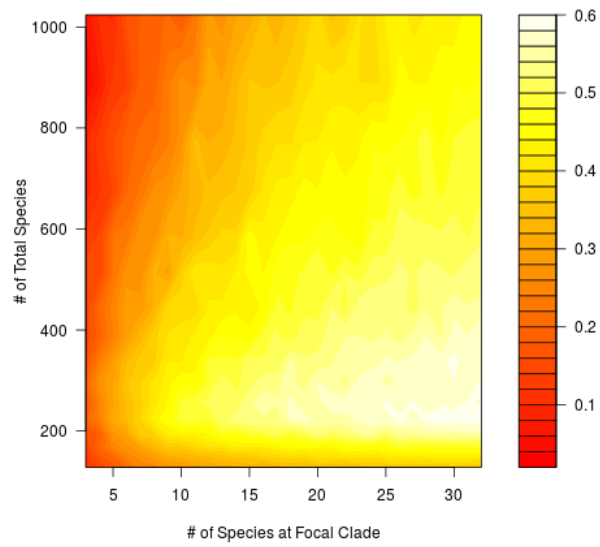


Figure 12: Mean error rate in the ranking of top 200 species.

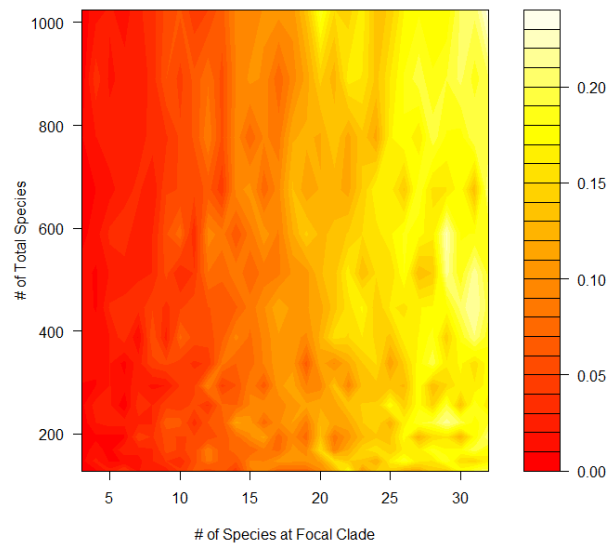


Figure 13: Mean error rate in the ranking of top 5% of species.

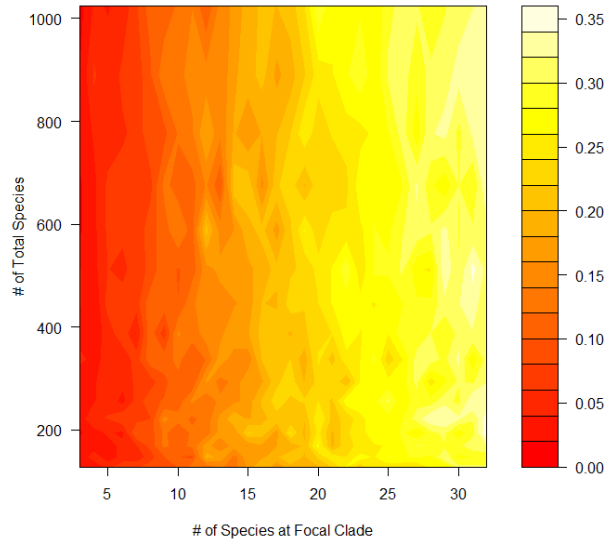


Figure 14: Mean error rate in the ranking of top 10% of species.

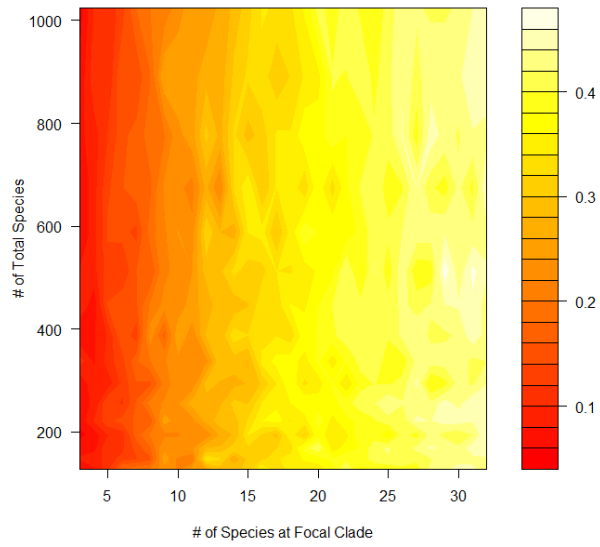


Figure 15: Mean error rate in the ranking of top 20% of species.