

Support in Area Prioritization Using Phylogenetic Information

Daniel Rafael Miranda-Esquivel

Abstract Human activities have accelerated the level of global biodiversity loss. As we cannot preserve all species and areas, we must prioritize what to protect. Therefore, one of the most urgent goals and crucial tasks in conservation biology is to prioritize areas. We could start by calculating ecological measures as richness or endemism, but they do not reflect the evolutionary diversity and distinctness of the species in a given area. The conservation of biodiversity must be linked to the understanding of the history of the taxa and the areas, and phylogeny give us the core for such understanding. In such phylogenetic context, evolutionary distinctiveness (ED) is a feasible way for defining a ranking of areas that takes into account the evolutionary history of each taxon that inhabits the area. As our knowledge of the distribution or the phylogeny might be incomplete, I introduce Jack-knife re-sampling in evolutionary distinctiveness prioritization analysis, as a way to evaluate the support of the ranking of the areas to modifications in the data used. In this way, some questions could be evaluated quantitatively as we could measure the confidence of the results, since deleting at random part of the information (phylogenies and/or distributions), would help to quantify the persistence of a given area in the ranking.

Keywords Phylogenetic conservation • Taxonomic distinctiveness • Jack-knife

Conservation Planning

The biodiversity is at risk, therefore decisions must be made in order to tackle the biodiversity crisis. In the process of conservation planning, one or maybe the most important task is to evaluate the quality and importance of a given area. To fulfill this task there are many metrics, from species richness to endemism, but these two values do not consider the evolutionary uniqueness of a species (Purvis and Hector

D.R. Miranda-Esquivel (✉)

Escuela de Biología, Universidad Industrial de Santander, Santander, Colombia

e-mail: dmiranda@uis.edu.co

© The Author(s) 2016

R. Pellens, P. Grandcolas (eds.), *Biodiversity Conservation and Phylogenetic*

Systematics, Topics in Biodiversity and Conservation 14,

DOI 10.1007/978-3-319-22461-9_11

2000). Any useful metric must include the evolutionary value of the species (Rolland et al. 2012), where the most important and therefore the selected area is the one that harbors the highest biodiversity, but this does not mean the highest number of species but the highest number of unique species or evolutionary fronts.

There are many approaches in the context of phylogenetic diversity and conservation, from community ecology to taxon or area conservation. Given this broad spectrum, the questions are different and vary a lot. In the context of community ecology and phylogeny, the approach is to evaluate whether there is structure in the community given the phylogeny (Cavender-Bares et al. 2009), and therefore the null model approach is used to present the null hypothesis. The species by area matrix is shuffled (see: Gotelli and Graves 1996), or the species or area labels are shuffled. Here the “support” is closer to the traditional confidence limits and error evaluation.

To evaluate the diversity of an area using phylogenies as a general frame, two main perspectives could be used, evolutionary distinctiveness (ED) or phylogenetic diversity (PD). Evolutionary distinctiveness refers to species-specific measures developed to assign scores to the species and therefore the areas they inhabit (Vane-Wright et al. 1991). The measures are topology-based indices, calculated as “the sum of basic taxic weights, Q , and the sum of standardised taxic weights, W .” (Schweiger et al. 2008), and therefore are also known as Taxonomic distinctiveness indices. Phylogenetic diversity (PD) is a distance-based index using minimum spanning path of the subset in the tree (Faith 1992). Redding et al. (2008) identified some of the major differences between ED and PD. PD is effective only if all the species within the optimal subset are protected, otherwise other optimal subsets are possible; unlike ED, PD is not species-specific and thus does not offer priority species rankings, which are important to species conservation approaches as the IUCN Red List of Threatened Species. Furthermore, topologies are more stable than branch lengths. Increasing the number of characters or changing the set of characters seldom leads to entire shifts in the relationships among species, whereas branch lengths change considerably from one set of characters to another and permit only to state about the evolution of the data set that generated the topology and the branch lengths (Brown et al. 2010).

Indexes Used

I present the general protocol to evaluate species or areas in a phylogenetic context in Fig. 1. The different indices for each species are calculated to obtain the species phylogenetic values, while the sum of the indices of all species in a given area produces the areal phylogenetic values.

I used the traditional I & W indices created by Vane-Wright et al. (1991), along with the modifications introduced by Posadas et al. (2001) to consider endemism and widespread species (I_e/W_e), the size of the topology (I_s/W_s) or both variables at the same time (I_{es}/W_{es}). The standardization of the indices I and W enables the

comparison of topologies with different number of species. In a topology with three species (I (II III)), distributed as taxon I in area A, II in B, and III in C. The taxon I and therefore the area A will have a value of 2.0 for indices I and W , while in a five species topology (I (II (III (IV V))))), the taxon I will have an index value of 8.0 for index I and 4.0 for index W , while the standardized I_s for this taxon and the area it inhabits will be 0.5 for both topologies.

If we consider the distributional pattern of the species, it could be endemic or widespread. We could apply the same index value to all areas where the species is present, but areas inhabited by widespread species will be selected, as we will sum the index values for each taxon, while an area inhabited only by an endemic taxon will be valued just for the single taxon it contains.

In a five taxa topology (Fig. 1), with four widespread species in the areas F, G, and H. If we use index I these three areas are as important as the area A, while using W index they are more important than the area A, as each area obtains the final index value because of the sum of all species inhabiting the area. Areas F, G and H are selected not because they are inhabited by unique species as area A but by widespread species. Using I_e/W_e or I_{es}/W_{es} the most important area is A, as it contains an evolutionary unique species, which is not found elsewhere.

Given the plethora of indices to choose, Winter et al. (2013) presented an important question: “We also call for a comprehensive guideline through the jungle of

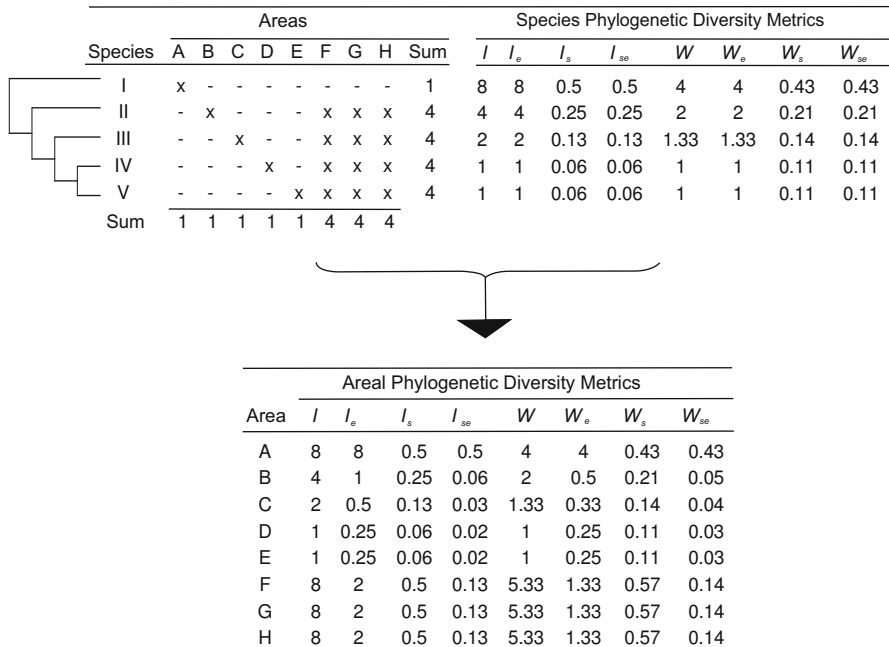


Fig. 1 An example for determining phylogenetic diversity metrics at species and area levels for a hypothetical topology with five species (four widespread), distributed in eight areas (Modified from Lehman (2006))

available phylogenetic diversity indices, with particular respect to the needs of conservationists – which index helps to protect what?”. Part of the answer to this question is given by the support to the decisions made, but in species or areas prioritization the literature does not present any kind of support measure (Whiting et al. 2000; Posadas et al. 2001; Pérez-Losada et al. 2002; López-Osorio and Miranda-Esquivel 2010; Prado et al. 2010), neither the most recent revisions cite any measure to evaluate the stability, confidence or support to the results (Schweiger et al. 2008; Vellend et al. 2011).

Jack-Knife

In a jack-knife analysis, given a sample of observations and a parameter to evaluate, a subsample is made by eliminating a proportion of the original data and the parameter is calculated for the subsample. This procedure is repeated n times and summarized. Since the introduction of the jack-knife (Quenouille 1949), researchers have used it, to define limits of confidence in many sorts of analyses, from statistics (Efron 1979; Smith and van Belle 1984) and ecology (Crowley 1992) to phylogeny. It has been used not only as a measure of support (Lanyon 1987), but as a way to obtain the best solution for large data sets (Farris et al. 1996), to test competing hypotheses (Miller 2003), to generalize the performance of predictive models or for cross-validation to estimate the bias of an estimator. As the bootstrapping, it could be seen as “a measure of robustness of the estimator with regard to small changes in the data” (Holmes 2003).

I use this re-sampling approach to evaluate the support of the area ranking in the context of conservation and phylogeny. Therefore, some questions could be evaluated quantitatively.

Jack-Knife in Conservation

The use of a meta-criterion to define an optimal parameter value has been used widely in phylogenetic analysis, i.e. the incongruence length difference test to define the ts/tv/gap costs (Wheeler 1995) or jack-knife frequencies to evaluate whether concavity parsimony outperforms linear parsimony (Goloboff et al. 2008).

In conservation biology, there must be a measure of the confidence and robustness of the results. A sensitivity analysis, deleting at random part of the information, helps to understand the support of the data as the persistence of a given area in the ranking. Therefore, jack-knife is the appropriate tool to explore the behavior of the results to perturbations in the data set (Holmes 2003).

In a conservation phylogenetic based analysis, there are three different items to evaluate, as we have three input parameters: the topology, the species in a given topology, and the distribution of a species.

The first question arises when we ask about the distributional pattern of the species: what if a locality (therefore all or some species in that area) is not included in the analysis? A species could not be included in a given locality for three reasons, because (1) it was never present there; (2) it is locally extinct; or (3) it was not sampled, although the species is present in the area. To evaluate such situation, the species can be deleted from a number areas to quantify the effect of missing information.

The second question arises when a species included in the phylogenetic analysis is not considered in the conservation analysis: what if a species is not included? A species not included in the analysis will affect the index value as this depends on the species included on the calculation. In this context, the presence of a species is deleted from all the areas it inhabits.

The third question arises when we do not include a given phylogeny: what if a phylogeny is not included? The whole topology might not be available for the conservation analysis. We could depend on a limited subset of phylogenies to the ranking of an specific area. Here, the topology, therefore the species and their distributions are deleted.

Given the three questions we can decide whether a phylogeny, a taxon or an area is deleted, with different probability values:

- j.topol is the probability to choose a topology ($= p$)
- j.tip is the probability to choose a species ($= q$)
- j.area is the probability to choose an area ($= r$)

In the first scenario, an area is deleted from the distribution of a species with a probability of $p \times q \times r$ ($0 < p, q, r < 1$), that is, the probability to select the topology and then select the species and then select the area. An area could be removed from the whole analysis, and this has to be run only the number of areas times, eliminating a single area each time. It would show the position of the area in the ranking of the areas and is equal to delete the area from the final results.

In the second scenario, a species is deleted from a single topology with a probability $p \times q$ ($0 < p, q < 1, r = 1.0$), therefore all areas inhabited by this species will not be included.

In the third scenario, the whole topology is not included in the analysis with a probability p ($0 < p < 1, q = r = 1.0$), all the species and areas, belonging to that topology, will not be included in the analysis.

The first decision in the three scenarios, is made on the topology. As the number of topologies NOT included increases with the value of p , the absolute indices values would be small and inversely proportional to the value of p .

Those areas prioritized because of its position in a single or just a few topologies would change, the indices values would be lower, and the position of the area in the ranking might change. If an area is supported by all or most of the topologies, its position in the ranking must be stable, although the index value would be small in all the replicates, therefore the index values *per se* are meaningless, but the ranking is informative.

There is a fourth question, not considered here, related to the length of the branch. This question is valid in the context of Phylogenetic Diversity [PD] (Faith 1992), Genetic Diversity [GD] (Crozier 1992), or total lineage divergence (Scheiner 2012) [a metric similar to PD]. These methods require the precise estimation of the length, therefore the accuracy of the index value depends heavily on the length estimation.

Although Krajewski (1994) considers that the debate of the use and calculations of divergence in systematics and conservation are two topics, I consider that the same criticisms to the accuracy estimation of the length in systematics will have a profound impact in the decision made when the topology and its branch lengths are used in conservation. And as this quotation from Brown et al. (2010) states, “in any phylogenetic analysis, the biological plausibility of branch-length output must be carefully considered”. Therefore, we must be well aware of the methodological approach used to construct the phylogeny (Rannala et al. 2012).

Additionally, in some cases we must consider the sensitivity of PD value to intra-specific variation (Albert et al. 2012). Therefore, we must take into account the source of the tree (species vs. gene trees) [see for example Spinks and Shaffer (2009)].

Optimal Scenario

Given a data set and n random perturbations on this data, if the index is robust, all (or most) perturbations would yield the same general ranking. Therefore, in the context of conservation in an optimal situation, we would prefer areas that:

1. *Have the same position in the ranking (original and re-sampled), no matter if we delete areas, species, or phylogenies*
= same ranking or position, insensitive to changes in the item(s) deleted.
 2. if not, at least must be the same position in the ranking but considering just a subgroup (e.g. be first or second, or first to third).
 3. *Have the same position in the ranking (original and re-sampled), no matter the delete probability used (from 0.01 to 0.5).*
= same ranking or position, insensitive to changes in the delete probability.
 4. or, have the same position for most of the probabilities used, but not counting extreme situations as a delete probability of 0.5.
- = not too sensitive to the probability values used.

In a real world, an scenario to meet the requirements of the first and third conditions is too strict and maybe impossible to fulfill. Therefore, my decision rules to select the best index and the best ranking are based in the second and fourth situations. The area must have the same position in the ranking considering just a subgroup, from the first to the third position in the ranking, no matter the type of item deleted, and for most of the probability values.

An alternative measure is to evaluate the behavior of an index and its success as the number of times that a replicate recovers part of the original ranking (e.g.

1st/2nd/3rd), but in any order. The researcher could consider only the first position in the ranking and evaluates the persistence of this area, or could consider the whole ordered ranking. These measures could be too strict and will be sensitive to the smallest perturbation to the data set, while the first to third position would be enough in terms of conservation planning.

Given any measure of success, the re-sampling approach in conservation have some possible applications as:

1. Which is the best index? that will answer also, what do we want to conserve/use to prioritize?

The best index would be defined as the most supported index, while the area used would be that found for most of the probabilities used.

2. How stable is the ranking (e.g. 1st/2nd/3rd position)?

This is a variation of the previous question, but focused in the ranking, as we prefer a supported ranking, we might evaluate the support for the original ranking.

Proposed Protocol

Following the expected behavior in an optimal condition, first I evaluated the index. I considered the best index as the one that recovered most times the same original ranking -first to third areas-, as an ordered ranking. Then, using the selected index, I evaluated the best area, as the one found most often in the first place.

I tested six scenarios by modifying j_{topol} and j_{tip} values as follows: j_{topol} values of 0.50 and 0.32, and j_{tip} values of 1, 0.50 and 0.32. These values are just used to introduce the concept, but they are similar to strong, mild and relaxed tests. A value of 1 to delete a species means that all areas for that species will be deleted, while a value of 0.32 means that one out of three will be deleted. Smaller values as 0.01 are discarded, it would make no difference, as the perturbation to the data would be unimportant.

The effect of deleting areas is related to the number of areas inhabited. If the species is in an endemic area, the effect of deleting an area would be as deleting the whole species, while in a widespread species, the effect should be minimal with indices as I_e/W_e or I_{es}/W_{ess} , but we can not define which is the best index as the four indices have similar properties. In all cases the probability of deleting areas was 1, therefore I tested the effect of the topology and species but not the effect of the distribution.

Number of Replicates

Hedges (1992) presented the number 1825 as the number of replicates needed to obtain an accuracy of $\pm 1\%$ for a bootstrapping proportion of 95%. Although the higher the number of replicates the higher the accuracy of the estimation of the bootstrap or jack-knife value, Pattengale et al. (2010) introduced a stopping criteria that yield lower figures as 500 replicates to get robust bootstrapping values for a 2500 taxa analysis. I randomized each scenario 10,000 times, that could be considered intuitively an appropriate number of replicates to estimate the jack-knife proportion for conservation purposes.

For these analyses, I used a modified version of the program *Richness* (Posadas et al. 2001) to randomize the data and to perform the index calculations [Jrich: available from <https://github.com/Dmirandae/jrich>], while the data analyses were performed using the software R (R Core Team 2013) and the figures were prepared using the library ggplot2 (Wickham 2009).

Empirical Examples

First Case: The Original Ranking Does Not Mean Support

Posadas et al. (2001) evaluated the conservation ranking in southern South America areas. They found that depending on the index used, the selected area changed, as the best area could be: Santiago (D), Ñuble (F), Valdivia (H), or the Malvinas islands (K). Also, for a single index, the values could be misleading, as the differences between the *W* index values are quite small, and the ranking could be an artifact rather than a real result (Table 1). I reanalyzed their dataset and found that the best index for this analysis is I_s (Fig. 2) as this index that has the highest jack-knife frequency.

The most stable area using I_s or raw *W* (the second best index), was the Malvinas islands, a candidate to be the best area (Fig. 3). The high uncertainty in the area chosen is eliminated when the support is included in the selection of the best area. Santiago has the highest number of species and harbors the highest number of endemic species, but it was not placed as the highest priority, while Malvinas island, the second most endemic area has the highest priority. The inferences based on the

Table 1 First area in the ranking proposed by Posadas et al. (2001). For raw *W* the index values are 52.62/52.58/52.05. Labels follow Posadas et al. (2001)

	<i>I</i>	<i>W</i>
<i>raw</i>	Malvinas islands (K)	Valdivia (H) or Santiago (D) or Ñuble (F)
<i>e</i>	Malvinas islands (K)	Malvinas islands (K)
<i>s</i>	Santiago (D)	Ñuble (F)
<i>es</i>	Malvinas islands (K)	Ñuble (F)

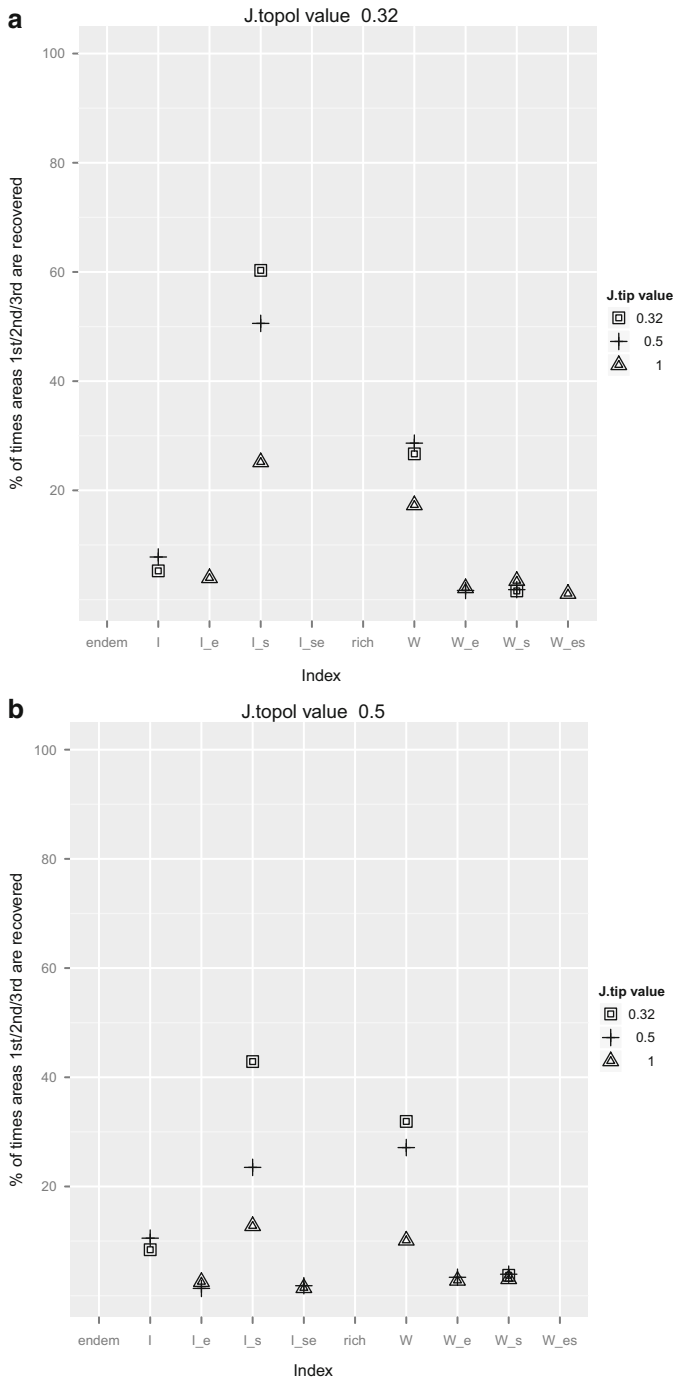


Fig. 2 For each index, the number of hits with a delete value of 0.32 or 0.5 for j.topol and three j.tip values of 0.32, 0.5, and 1 (in the last situation, the whole topology is deleted). Species endemism and richness (number of species) are included for comparative purposes (Data from Posadas et al. (2001)). (a) Number of hits with a j.topol value of 0.32 (b) Number of hits with a j.topol value of 0.5

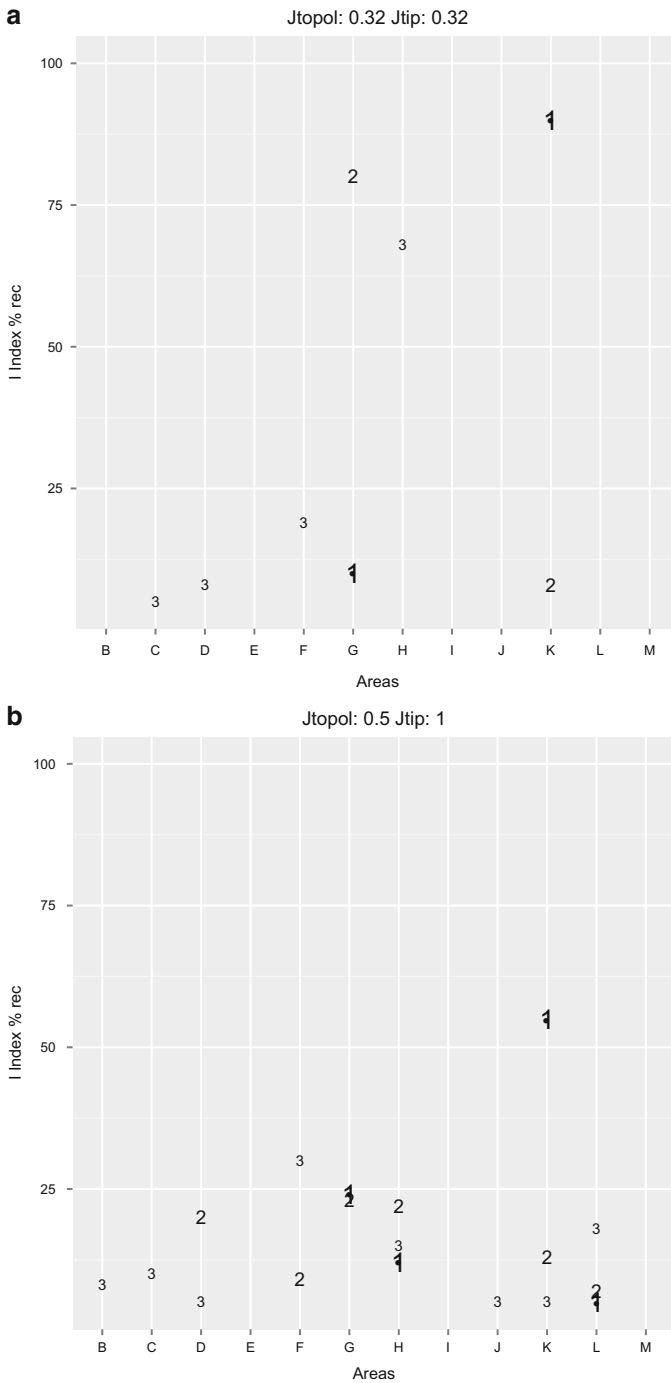


Fig. 3 Number of times an area is recovered as the first (number 1), second (number 2), or third area (number 3). These are the lowest (a) and the highest (b) delete probabilities used in this analysis. Acronyms for areas follow Posadas et al. (2001) and table 1. (a) Probabilities of $j.topol=0.32$, $j.tip=0.32$ (b) Probabilities of $j.topol=0.5$, $j.tip=1$

un-sampled data set might be misleading, while jack-knifing could help to decide which is the most supported solution.

Second Case: The Support for the Original Ranking

There are two main approaches to define amazonian areas of endemism, eight areas from Bates et al. (1998) and Da Silva et al. (2005) or 16 areas from Da Silva and Oren (1996). López-Osorio and Miranda-Esquivel (2010), used both ways to establish conservation priorities for Amazonia's areas of endemism.

Using Bates et al. (1998) areas, they found that Guiana and Inambari are the first and second priority areas. Inambari is the richest area while Guiana presents the highest endemism value. Their inferences were based on W_{es} , on theoretical grounds as the index includes endemism and standardization (López-Osorio and Miranda-Esquivel 2010).

The reanalysis showed that the best index is either W_{es} , W_e or W_s (Fig. 4). These three indices select Guiana as the first area and Inambari as the second area (Fig. 5), as stated in the original paper. In this example the re-sampling reinforces the original findings, giving a stronger support to the areas chosen as first and second in the ranking.

Using the areas from Da Silva and Oren (1996), López-Osorio and Miranda-Esquivel (2010) found that depending on the index, either Guiana2 or Rondonia could be the highest priority area, while the second area could be Guiana3, Inambari2 or even Rondonia or Guiana2. Therefore, the first question is, which is the best index for conservation in Amazonia? and given that index, which are the areas chosen as the first and second priority?.

López-Osorio and Miranda-Esquivel (2010) found that most indices selected the same area Guiana2, which could be seen as there is no difference given the index. The reanalysis showed that in general I_s and W_s are more stable than any other index, and I_s behaves better than W_s . As the size of the topologies is different and some large topologies with more nodes may have more impact than smaller topologies, standard I and W indices are not stable (Fig. 6). The first area is Guiana2 in all indices used, while the second area varies: Rondonia, Guiana3 or Inambari2 (Fig. 7). These results are similar to those found by López-Osorio and Miranda-Esquivel (2010). Here the re-sampling helped to resolve the initial discrepancy as the highest priority is Guiana2 and not Rondonia, that could be a possible candidate. The second area could be any of the three initially considered, so the evidence is not misleading but inconclusive to define the second area, even after re-sampling the data.

These brief examples show that the confidence of the original ranking should be evaluated using re-sampling, as an un-sampled ranking analysis could be unstable when some information (phylogenies or species) is deleted. The results may render any output, from a different answer from the original ranking to a congruent answer with the original ranking. Only after the re-sampling analysis, the quality of the answer could be stated without hesitation. Even if we only calculate the support for

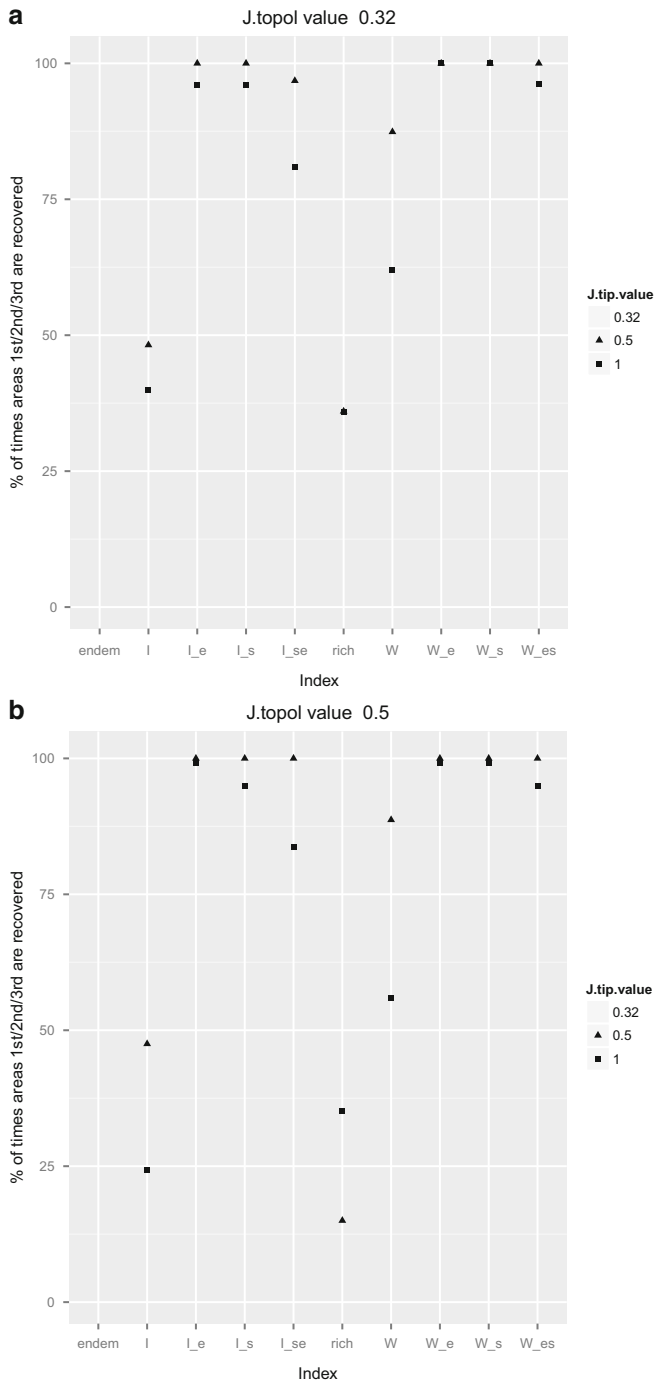


Fig. 4 Number of hits with a delete value of 0.32 or 0.5 for j.topol and three j.tip values of 0.32, 0.5, and 1 (in the last situation, the whole topology is deleted) (Data from López-Osorio and Miranda-Esquivel (2010). Areas from Bates et al. (1998)). (a) Number of hits with a j.topol value of 0.32 (b) Number of hits with a j.topol value of 0.5

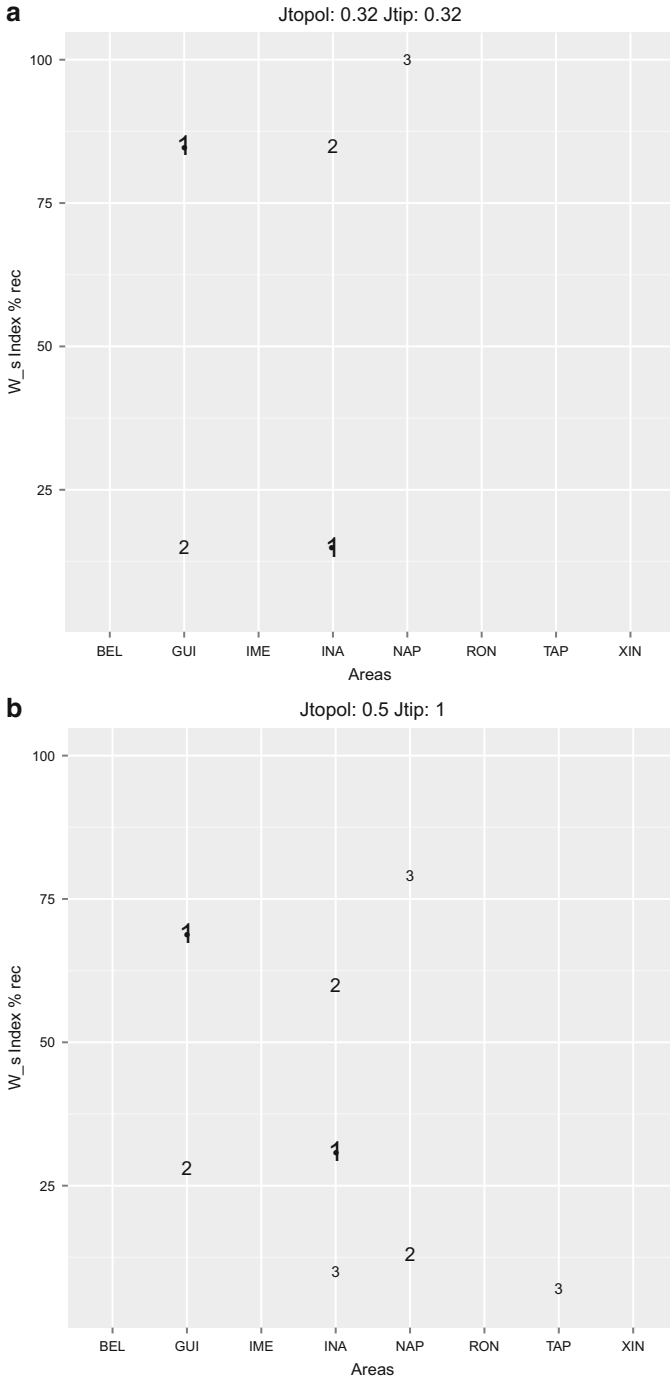


Fig. 5 Number of times an Area is recovered as the first (number 1), second (number 2), or third area (number 3). These are the lowest (a) and the highest (b) delete probabilities used in this analysis. (a) Probabilities of $j.topol=0.32, j.tip=0.32$ (b) Probabilities of $j.topol=0.5, j.tip=1$

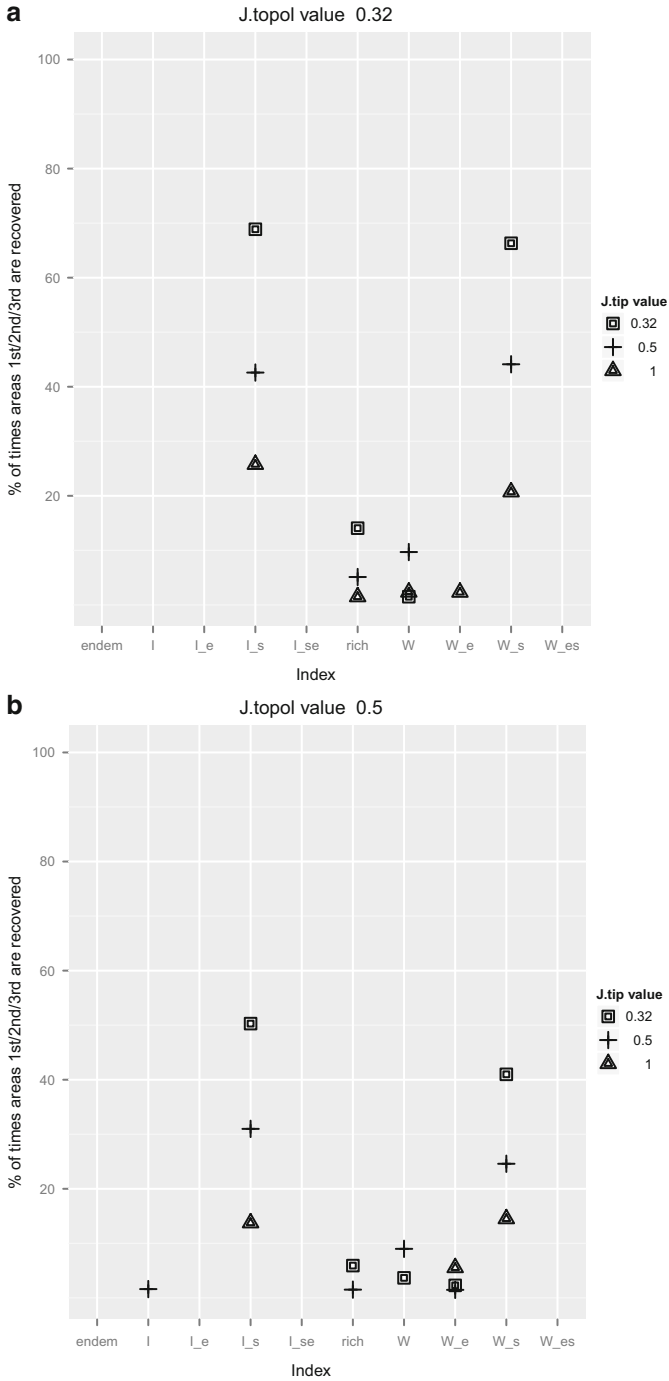


Fig. 6 Number of hits with a delete value of 0.32 or 0.5 for j.topol and three j.tip values of 0.32, 0.5, and 1 (in the last situation, the whole topology is deleted) (Data from López-Osorio and Miranda-Esquivel (2010). Areas from Da Silva and Oren (1996)). (a) Number of hits with a j.topol value of 0.32 (b) Number of hits with a j.topol value of 0.5

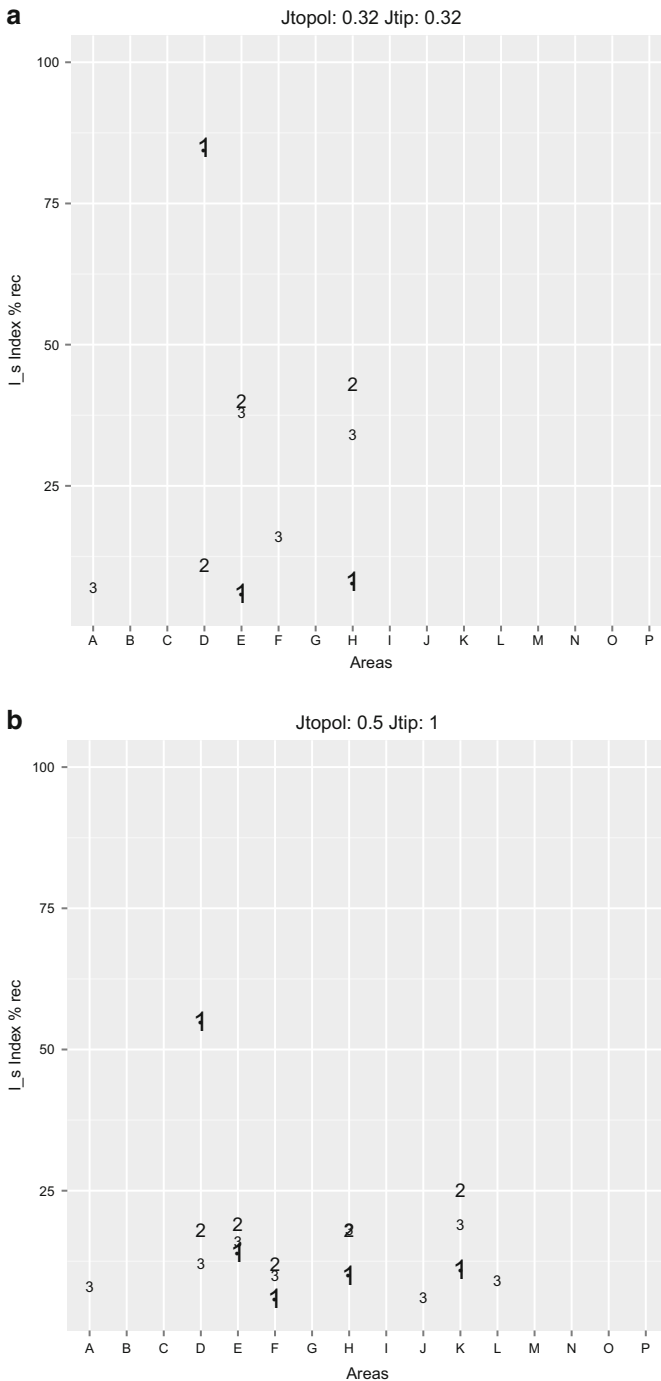


Fig. 7 Number of times an Area is recovered as the first (number 1), second (number 2), or third area (number 3). These are the lowest (a) and the highest (b) delete probabilities used in this analysis (Data from López-Osorio and Miranda-Esquivel (2010). Areas from Da Silva and Oren (1996)). (Acronyms *A* = Guiana3, *B* = Guiana2, *C* = Guiana1, *D* = Imeri2, *E* = Imeri1, *F* = Napo3, *G* = Napo2, *H* = Napo1, *I* = Inambari1, *J* = Inambari2, *K* = Inambari3, *L* = Inambari4, *M* = Rondonia, *N* = Tapajos, *O* = Xingu, *P* = Belen). (a) Probabilities of $j.topol=0.32$, $j.tip=0.32$ (b) Probabilities of $j.topol=0.5$, $j.tip=1$

a given ranking, the results after re-sampling would give a clue of the situation when the information is perturbed.

Acknowledgements I am grateful to Roseli Pellens for her kind invitation to participate in the book. Two anonymous referees helped to improve my perspective about Jack-knife and Conservation. I am indebted to División de Investigación y Extensión, Facultad de salud (project 5658), and Facultad de Ciencias, Universidad Industrial de Santander (project 5132) for their financial support.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

- Albert CH, de Bello F, Boulangeat I, Pellet G, Lavorel S, Thuiller W (2012) On the importance of intraspecific variability for the quantification of functional diversity. *Oikos* 121(1):116–126
- Bates JM, Hackett SJ, Cracraft J (1998) Area-relationships in the neotropical lowlands: an hypothesis based on raw distributions of passerine birds. *J Biogeogr* 25(4):783–793
- Brown J, Hedtke S, Lemmon A, Lemmon E (2010) When trees grow too long: investigating the causes of highly inaccurate bayesian branch-length estimates. *Syst Biol* 59(2):145–161
- Cavender-Bares J, Kozak KH, Fine PVA, Kembel SW (2009) The merging of community ecology and phylogenetic biology. *Ecol Lett* 12(7):693–715
- Crowley P (1992) Resampling methods for computation-intensive data analysis in ecology and evolution. *Annu Rev Ecol Syst* 23:405–447
- Crozier R (1992) Genetic diversity and the agony of choice. *Biol Conserv* 61(1):11–15
- Da Silva JMC, Oren D (1996) Application of parsimony analysis of endemism in amazonian biogeography: an example with primates. *Biol J Linn Soc* 59(4):427–437
- Da Silva JMC, Rylands AB, Da Fonseca GAB (2005) The fate of the amazonian areas of endemism. *Conserv Biol* 19(3):689–694
- Efron B (1979) Bootstrap methods: another look at the jackknife. *Ann Stat* 7(1):1–26
- Faith D (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61(1):1–10
- Farris J, Albert V, Källersjö M, Lipscomb D, Kluge A (1996) Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 12(2):99–124
- Goloboff PA, Carpenter JM, Arias JS, Miranda-Esquivel DR (2008) Weighting against homoplasy improves phylogenetic analysis of morphological data sets. *Cladistics* 24(5):758–773
- Gotelli N, Graves G (1996) Null models in ecology. Smithsonian Institution Press, Washington, DC
- Hedges SB (1992) The number of replications needed for accurate estimation of the bootstrap *P* value in phylogenetic studies. *Mol Biol Evol* 9(2):366–369
- Holmes S (2003) Bootstrapping phylogenetic trees: theory and methods. *Stat Sci* 18(2):241–255
- Krajewski C (1994) Phylogenetic measures of biodiversity: a comparison and critique. *Biol Conserv* 69(1):33–39

- Lanyon S (1987) Jackknifing and bootstrapping: important “new” statistical techniques for ornithologists. *Auk* 104(1):144–146
- Lehman SM (2006) Conservation biology of Malagasy strepsirrhines: a phylogenetic approach. *Am J Phys Anthropol* 130(2):238–253
- López-Osorio F, Miranda-Esquivel D (2010) A phylogenetic approach to conserving amazonian biodiversity. *Conserv Biol* 24(5):1359–1366
- Miller JA (2003) Assessing progress in systematics with continuous jackknife function analysis. *Syst Biol* 52(1):55–65
- Pattengale N, Alipour M, Bininda-Emonds O, Moret B, Stamatakis A (2010) How many bootstrap replicates are necessary? *J Comput Biol* 17(3):337–354
- Pérez-Losada M, Jara C, Bond-Buckup G, Crandall K (2002) Conservation phylogenetics of chilean freshwater crabs *Aegla* (anomura, aeglidae): assigning priorities for aquatic habitat protection. *Biol Conserv* 105(3):345–353
- Posadas P, Miranda-Esquivel DR, Crisci J (2001) Using phylogenetic diversity measures to set priorities in conservation: an example from southern South America. *Conserv Biol* 15(5):1325–1334
- Prado A, Hawkins J, Yesson C, Bárcenas R (2010) Multiple diversity measures to identify complementary conservation areas for the Baja California peninsular cacti. *Biol Conserv* 143(6):1510–1520
- Purvis A, Hector A (2000) Getting the measure of biodiversity. *Nature* 405(6783):212–219
- Quenouille M (1949) Approximate tests of correlation in time series. *R Stat Soc Ser B* 11:18–84
- R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org>
- Rannala B, Zhu T, Yang Z (2012) Tail paradox, partial identifiability, and influential priors in bayesian branch length inference. *Mol Biol Evol* 29(1):325–335
- 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(4):606–615
- Rolland J, Cadotte MW, Davies J, Devictor V, Lavergne S, Mouquet N, Pavoine S, Rodrigues A, Thuiller W, Turcati L, Winter M, Zupan L, Jabot F, Morlon H (2012) Using phylogenies in conservation: new perspectives. *Biol Lett* 8(5):692–694
- Scheiner SM (2012) A metric of biodiversity that integrates abundance, phylogeny, and function. *Oikos* 121(8):1191–1202
- Schweiger O, Klotz S, Durka W, Kühn I (2008) A comparative test of phylogenetic diversity indices. *Oecologia* 157(3):485–495
- Smith E, van Belle G (1984) Nonparametric estimation of species richness. *Biometrics* 40:119–129
- Spinks P, Shaffer H (2009) Conflicting mitochondrial and nuclear phylogenies for the widely disjunct *Emys* (testudines: Emydidae) species complex, and what they tell us about biogeography and hybridization. *Syst Biol* 58(1):1–20
- Vane-Wright R, Humphries C, Williams P (1991) What to protect? – systematics and the agony of choice. *Biol Conserv* 55(3):235–254
- Vellend M, Cornwell W, Magnuson-Ford K, Mooers A (2011) Measuring phylogenetic biodiversity in Biological Diversity: Frontiers in Measurement and Assessment. eds A.E. Magurran B.J. McGill. Oxford University Press, New York, pp 194–207, Chap. 14
- Wheeler WC (1995) Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. *Syst Biol* 44(3):321–331
- Whiting A, Lawler S, Horwitz P, Crandall K (2000) Biogeographic regionalization of Australia: assigning conservation priorities based on endemic freshwater crayfish phylogenetics. *Anim Conserv* 3(2):155–163
- Wickham H (2009) *ggplot2: elegant graphics for data analysis*. Springer, New York
- Winter M, Devictor V, Schweiger O (2013) Phylogenetic diversity and nature conservation: where are we? *Trends Ecol Evol* 28(4):199–204