

Journal Club EcoEvol #01 - 06.06.2018

Cristian Dambros¹, Tiago Freitas², Bruno R. Ribeiro³, Gracielle Higino³, Pablo Silva⁴, and Adriana Diaz³

¹Affiliation not available

²Universidade Federal do Pará

³Programa de Pós-Graduação em Ecologia e Evolução, Universidade Federal de Goiás.

⁴Programa de Pós-Graduação em Recursos Naturais do Cerrado, Universidade Estadual de Goiás

April 28, 2020

Abstract

This is a review of the preprint entitled ” **Challenging the Raunkiaeran shortfall and the consequences of using imputed databases**”, by Lucas Jardim, Luis Mauricio Bini, José Alexandre Felizola Diniz-Filho and Fabricio Villalobos. The preprint was originally posted on bioRxiv on October 18, 2016 (<https://doi.org/10.1101/081778>). Our journal club reviewed this preprint in the meeting of June 6, 2018.

The preprint by Jardim *et al.* makes important advances in the assessment of missing trait data. They analyzed how the simulated target and auxiliary functional traits imputed in different sorts of missing data would influence the descriptive statistics, model parameters, and phylogenetic signal estimation from these databases. They simulated coalescent trees and missing data (missing completely at random, missing at random but phylogenetically structured, and missing at random but correlated with another variable) and found that the structure of the missing data, the evolutionary model used to simulate the phylogeny and the percentage of missing data were important factors determining estimation errors. The manuscript is well-written and makes a novel and sound contribution to the ecological literature. Trait data is rarely available for entire communities of species and most trait databases use imputation methods. In spite of the common use, previous studies have not evaluated the impact of data imputation in common metrics of trait distribution and phylogenetic signal. This paper shows that missing trait data and data imputation can create biases in common ecological and evolutionary metrics, and suggest ways to minimize the problem when only incomplete data are available. Our suggestions are outlined below.

Major concerns:

- Although the authors investigated the most advanced imputation methods, the most simple and common imputation methods were not tested: i) filling up trait gaps using data on congeneric (or closely related) species and ii) imputing data for entire large clades (eg. families). We suggest testing the performance of these simple methods.

Minor concerns:

- The results section could be shortened to highlight the main findings.
- The authors could avoid using jargons in the abstract and other parts of the manuscript (eg. is it necessary to mention “Raunkiaeran shortfalls?”).
- Table 2: the legend should describe better the models presented in the table. Our suggestion is: “Interaction of variables included in the seven models used to explain the estimation error of descriptive statistics (mean, variance and regression coefficient) and phylogenetic signal (Blomberg’s K and Moran Correlogram).”

Title

We are not sure whether authors challenged the shortfall and the consequences of using imputed databases. If they only challenged the shortfall (and not the consequences), the title could be: “Challenging the Raunkiaeran shortfall: consequences of using imputed databases”

Summary

- [L16] It is recommended to change the phrase “empirical investigations” by “empirical research” in the abstract
- [L21] It is recommended to change the phrase “coalescent phylogenies” by “coalescent trees” because phylogeny and coalescence are two different concepts.

Introduction

- The first two paragraphs are too specific (i.e., present information regarding methods to deal with data imputation). It would better describe the “big picture” before presenting specific details. So, we would suggest remove both paragraphs. Further, the first paragraph could be moved to the methods section.
- [L79-80] Suggestion to replace the sentence “**Thus, to facilitate research and make it reproducible and data more accessible (Reichman et al. 2011) [...]**” with “**Thus, to make reproducible and more accessible (Reichman et al. 2011) [...]**”
- [L82] remove parenthesis before “[...] Jones et al. 2009 [...]
- [L86]: Eltonian shortfall was mentioned as a “lack of knowledge about species’ traits and their ecological functions”, but this is an exclusively approach of Raunkiaeran shortfall. Eltonian shortfall addresses the lack of knowledge species’ interactions.
- [L89] To avoid confusing, we suggest replacing “them” by “database”
- [L92] Just a matter of style, but you could state “because closely related species generally resemble each other more than distantly related species”.
- [L94] “Consequently” is not need here.
- [L101] There is a space between “- burst”.
- [L114] Why is it mentioned as a mechanism?
- [L117-118] Suggestion to replace the sentence “**In such cases, if analysis were to be conducted on phylogenetically imputed data, results could be misleading given [...]**” with “**In such cases, phylogenetically imputed data could be misleading given [...]**”
- [L120] “[...] thus potentially inflating the level of phylogenetic signal.” with “[...] which potentially inflates the level of phylogenetic signal.”
- [L120] Suggestion to replace the sentence “**This potential issue can have important consequences for studies evaluating, for example [...]**” with “**This potential issue may be essential for studies that evaluate [...]**”.

Methods

- [L144] In the methods section, it can be highlighted that you simulated “**functional traits**” or you can leave the article more general. This could be useful for any type of trait, morphological with or without impact on fitness? How would the results change if the trait had sexual selection?

- [L196] What “this” is supposed to mean?
- [L266] Parenthesis can be removed.
- [L271] Comma is not necessary here.

Figures

- [Figure 2] X-axis name is missing. In the figure legend, would better include the meaning of each abbreviation (e.g., NRMSE, MSE). Just a matter of style: would better put closer to each square and its legend.
- [Figure 4] Idem to figure 3
- [Figure 5] Idem to figure 4
- [Figure 7] It is a very interesting and well-presented figure. In the figure legend, would better include the meaning of each abbreviation (e.g., NRMSE, MSE).

Concluding remarks

- Could be interesting to include some concerns about data imputation on phylogenies (due to a lack of evolutionary knowledge – addressed by the Darwinian shortfall) and how this could be driving toward more misleading studies on functional and/or phylogenetic diversity.