

Review PCI-EvolBiol 542

2022-04-25

Dear Editor,

Please find below my comments on the manuscript entitled “Do closely related species interact with similar partners? Testing for phylogenetic signal in bipartite interaction networks”. In this manuscript, the authors investigate the ability of different statistical methods to recover phylogenetic signal of species interactions on a set of simulated networks and then propose an empirical example with best practices guidelines stemming from their findings.

Overall, this paper proposes a very useful benchmarking of the false positive error for species interactions analysis and would be of great help to ecologists and evolutionary biologists. I especially like the clarity of the results as presented in Figure 2: the Mantel test approach very often outperforms the Phylogenetic Bipartite Linear Model method by having a much lower false positive rate.

I have however several major comments and some minor ones (often semantic) that I think the authors should address prior to publication but I have little doubts that they will affect the general findings (Mantel test > PBLM for network interactions).

Major suggestions

Simulation protocol

The findings of the authors essentially relies on the simulation protocol and mainly on the `RPANDA::build_network.BipartiteE` function. I suggest the authors spend more time explaining how this function works (even though it is already described in Maliet et al 2020). Furthermore, I found the methods section on simulations quiet complicated to read. There is a lot of information and at the time of the reading, the reader doesn't know which one is essential and which ones are just parameters details. I suggest making a figure or a table summarising the simulations: 1) what are the specific simulated scenarios for (i.e. what aspect are the authors trying to simulate), what are the important parameters used to approximate that and, 3) what are the expected results from that scenario (e.g. no phylogenetic signal, etc. . .).

Also, I believe the authors could strengthen their results by demonstrating the robustness or realism of their simulation pipeline. For example, part of their work is looking at the effect of tree uncertainty on measuring signal in interaction networks (l. 316-328). Here the authors propose to simulate a network (using `RPANDA::build_network.BipartiteEvol`), then simulate DNA sequences (using `HOME::simulate_alignment`), inferring a tree from the simulate sequences (using `ape::nj`), and then measuring the phylogenetic signal from the resulting networks to compare them with the results without tree uncertainty. Although this is a perfectly good approach for measuring it, it relies on the effect of each intermediary algorithm which, from experience, although they are the best algorithms available can perform poorly at simulating realistic data in some conditions. This makes the authors' results slightly less convincing since the reader has to put faith in the different intermediary simulations.

Maybe one way to solve this problem would be to benchmark the author's results with some already published empirical data for which the answer is known? The authors could replicate the results of Martos et al 2012 not only for the application/guideline section but also for benchmarking of the simulations. For example, in the tree uncertainty scenario, they could compare the range of their simulated results to the one they would obtain from measuring the phylogenetic signal on the tree distribution from Martos et al 2012.

Data availability and reproducibility

The authors seem to provide all their code and data for reproducibility (I haven't checked all scripts though) which is a great! I have two minor suggestions on that point though:

- Although the authors provide their code for the simulations in the GitHub repository (and the link to the original publication containing the data), I think the repo would benefit of having an additional entry in the README to explain which script does what (and not just for `script_phylogenetic_signal_network.R`).
- I suggest adding a link to the (`mpfr` help page)[<https://www.mpfr.org/mpfr-current/mpfr.html>] in the README since `mpfr` (and dependent `Rmpfr`) might not be straightforward to install.

Minor suggestions

- 1.34: Although “Type I error” is a perfectly valid term here, I would suggest changing it to “false positive” throughout the manuscript to make it easier for the reader (I personally always need to google which type error is which one even after years working with it).
- 1.154: Is the opposition of the terms “modular” and “nested” common in the interactions field? They seem slightly confusing to me as opposed terms in the context of this sentence (can modular interactions not be neatly nested as well?).
- 1.163: I am not sure about the coining of the term “anti-phylogenetic signal” throughout the manuscript. This particularly led to confusion in the conclusions (1.672-678). Although I do understand the naming of it as the inverse of a phylogenetic signal, in the broader sense of phylogenetic comparative methods it could introduce some confusion where phylogenetic signal is broadly understood as being a link between a trait and the tree structure and no phylogenetic signal being the absence of it. Then an anti-phylogenetic signal would be the non-link between a trait and a tree? Which doesn't make much sense to me and I believe is not what the authors mean here. Maybe they could change it to something like “phylogenetic signal for dissimilarity” or something like that.
- 1.168-169: It is unclear to me how the traits are generated here and why six traits are generated. Under what distribution(s) (and which parameters) are these traits generated? Why did the authors chose to use six traits? And what is the correlation between these generated traits?
- 1.167-197: I think the reader could benefit of a table recapitulating the different simulation scenarios and parameters (number of interacting pairs, α values, expected phylogenetic signal, etc. . .).
- 1.202: I suggest adding a comparison of computational time and resources between both methods in the supplementary materials to convince the reader to which degree a PGLMM is more “computationally intensive” than a Mantel test.
- 1.221: Why did the author chose to use so many permutations? I think 1000 would be plenty (or even 100 since they consider that enough for the Kendall correlations). I am not entirely sure if the amount of permutations have an effect on the calculations of the statistics in this specific case but from experience, I know that it can induce some false negative error due to the pseudo-replication (so in doubt I would suggest reducing the number of permutations). Same on line 353.
- 1.254: Although the unit is interpreted as exactly the same, I suggest changing “Go” to “Gb” which I believe is more popularly used.
- 1.265-266: I don't understand what “the absolute differences in degrees” are.
- 1.354: I suggest using “False positive” instead of “alpha-risk” to avoid introducing yet another technical term.
- 1.355: Similarly here, I suggest changing “we generated a synthetic network” to “we simulated a network” to keep in line with the vocabulary used in the rest of the methods.
- 1.374-378: I suggest adding more information to this section: how did the authors chose the calibration(s) for the tree? (I suggest following Parham et al 2012 Syst. Biol. 61(2):346–359 recommendations). How did the authors obtained species-level trees (do they mean a fully resolved tree?) with polytomies at 10 Mya? (and why choosing this arbitrary time?)? Or do the authors mean they resolved polytomies by adding arbitrary 10 Mya branches to make the nodes binary? To which side of the node where these branches added (i.e. did that made the unresolved node more ancient - 10 Mya extra per multifurcation)

or did make the species splits arbitrarily 10 Mya younger? Regardless, this process seems to generate a stochastic distribution of trees. Did the authors used a tree distribution for the subsequent analyses or did they just generated a single tree?

- 1.448: Did the authors meant “guilds A and B” or “clades A and B”?
- Figure 3: I suggest adding the function’s package name in the examples of the guideline (i.e. `RPANDA::phylosignal_network`) and maybe changing the font to courier to make it clear this is an example code snippet.

Best regards,

Thomas Guillaume