Biogeosciences
Discussions

Open Access

EGU

BGD

Interactive
comment

Interactive comment on "Seed traits and phylogeny explain plant distribution at large geographic scale" *by* Kai Chen et al.

Anonymous Referee #2

Received and published: 8 August 2020

The authors focus on an interesting topic (how seed mass affects species range size). The Introduction is generally well written (can be improved by presenting clear predictions and I have a couple of minor comments below). However, the Methods section is so poorly written, with many flaws and errors. I got disappointed after reading the Methods section, and therefore cannot trust the Results section and therefore skipped the Discussion section (since I cannot judge whether the results are justified). The authors seemingly present many fancy analyses but are not really clear why to use them. Luckily, I know clearly all these methods and present my suggestions here.

First, please minimise errors, otherwise, audience cannot understand you. The URL of GBOWS in L85 cannot be opened (I guess you are saying "http://www.genobank.org/"?). L112 is wrong – "S.PhyloMaker" is not implemented in

the phytools package. They are separate things. Also, you should use the updated version, known as "V.PhyloMaker" (Jin and Qian 2019. Ecography Vol. 42). The most severe error is in L123 – there is no 'rda' function in ade4 package. This error makes me unable to examine whether your variation partition (the most important part of this study) is suitable or not, because the method is unexposed.

Second, please upload your data to make the study reproducible and meanwhile give clear descriptions. L86 – how did you choose these 1616 species from these two databases? This should be an important piece of information in this whole study, otherwise, it may produce biases. Will the species in China be overrepresented and cause potential biases to the analyses? Have you checked and standardised nomenclatures of species and how? L88 – two to 137 populations per species in China. Then, how many populations per species from the Kew SID? Another important issue is that how did you treat species with more than one type of morphological adaption (L104-110)? Many species can be dispersed by several modes, for example, by both endozoochory and exozoochory. What about species that are dispersed by water or ants? Please give a sample size of species in each category. I suggest uploading the compiled data you used in the analysis, since they are from two databases anyway.

Third, you mentioned many important issues in analysing data, but you stop at only a brief mention. You need to continue with why they are important issues and how you have solved them. That is, you need to explain what this analysis does and why it is relevant to your data, rather than just referring to an R function. Since I could not find the function you mentioned in doing variation partitioning (L123), I doubt whether it can really be suitable to your data structure where you have a continuous variable (seed mass), categorical variables (dispersal mode, genus) and proportion data (seed mass variability) all mixed? In L122-123, it is good that you realise multicollinearity, but how did you treat variables with high collinearity issue? L126 – the "assumptions of normality" refers to the residuals, but have you checked your model residuals? L89-91: Why did you use this index to show seed mass variability? Why not use (for example)

C2

coefficient of variation, which is more commonly used?

Fourth, there are some analytical flaws. L101 – 1 degree × 1 degree grids cannot be of the same size across the globe. This is a wrong procedure in calculating species range size, because it seems you did not consider projection. Since your grids have different sizes near the equator and at high latitudes, how could the species range size be comparable across species? In addition, I doubt the use of genus to surrogate phylogeny. In the variation partitioning, it seems you did not incorporate the phylogenetic correlation among species, which violates your previous sentence "closely related species tend to have similar traits and interspecific analyses can be compromised by phylogenetic correlation". In L125 – are you saying you did ANOVA with post-hoc Tukey HSD tests? What package did you use to do the tests? Again, you need to take phylogeny into account, otherwise, residuals do not fit model assumptions.

Minor comments:

The Introduction is generally written but can be improved with clear predictions. You only present study goals and predictions until the last paragraph (L74-75 is not a clear hypothesis). Why you make such hypotheses and what makes your study novel are essential throughout the whole Introduction section. For example, in L39-40 "few studies have explored this relationship", then what have they found? What makes your study different from these previous ones? L50 – what do you plan to do about this gap?

L66: Rephrase "have the same time to dispersal". No idea what this means – this line either has a grammar error or is delivered wrongly.

All references do not have years – how could they be matched with citations in the main text?

Figure 1 and Figure 2 actually tell little information. I suggest removing Figure 2 since no tip can be seen with these many species. For Figure 1, I suggest using grids to show numbers of specimen records (same unit as the grids used in range size), which

can avoid overlapping.

Figure 4: Is range size log-transformed (Figure 4 and Table A2) or log10-transformed (the main text)? What is the flat panel in each figure? Here, you standardised predictors, but this information is not given in the Methods section.

Figure 5: Seed mass and seed mass variability are two separate variables, but why are they combined in the variation partition?

The results in Figure 3 and Figure 5 are not reliable, due to the flawed methods.