Dear Recommender,

Following your suggestions, I have addressed several changes across the manuscript. The reviewers of the previous version of the manuscript also raised important and valid criticisms, which have led me to improve this preprint. Below, I give the individual comments of the editor and reviewers in black and my responses in blue, preceded by a '>>' notation.

Sincerely,

Ricardo A. Segovia
Adjunt Researcher
Instituto de Ecologia y Biodiversidad (Chile)

Dear author

Sorry for the time taken to reach a decision about your preprint. Briefly, it has been difficult to find reviewers due to the holiday season, and some expected reviews have been delayed. In any case, now we do have two reviews, and both agree with my assessment that the paper may merit a recommendation in PCI, once some key problems with the current version are solved.

>> I am thankful to receive this revision with comments and suggestions from the editor and reviewers. And, also I am sorry for the long time I took to send a reviewed version. Pandemic times have been difficult to maintain the rhythm of work.

More precisely, for the preprint to be recommendable, you need to:

(a) provide a much better theoretical explanation linking the environmental temperature suffered by plants and mean annual temperature, as well as other descriptors of "harshness" such as Freezing Days (by the way, Humboldt first proposal of a mechanism for the latitudinal diversity gradient was precisely harshness; I think Hawkins TREE 2001 highlighted that).

>> I have rewritten the introduction of the manuscript addressing this suggestion.

(b) Assess the effects of regional variations on the richness/MAT relationship, ideally using differences between biomes and/or ecoregions, realms and glaciated/unglaciated areas.

>> I have added the distribution of the residual variation by biomes (WWF). And I have estimated the slope of the variation by biomes.

(c) Pay special attention to the conversion of units, as it determines the slope values, and make a clearer formulation of your hypothesis about the slope that allows identifying the actual slope that is assessed. Current information in the methods is not enough so as to ascertain the exact way you may reach a comparable -0.65 slope.

>> I have added an explanation for the expected value of the slope in the section Material and Methods (Lines 382-388).

(d) Provide estimates of the goodness of fit of the models. It could be argued that within an information-theory-based hypothesis testing framework goodness of fit is not needed - because you effectively assess whether/to which extent some hypotheses are informative or not. However, you use AIC for comparing between alternative models, which leaves the reader with no information about which is the power of these models to "explain" the data. If goodness-of-fit lies below, say, 5%, we are talking about massive residuals and limited explanation of the overall phenomenon of richness. If, on the contrary, such percentage goes above 30 or even 40% of variation, that is really a lot. If you account for (b) and you end up having final models that include MAT, Frost Days, CWD, biome and realm, for example, and that accounts for more than half of the variation, your results will be much more convincing that if you explain one third of richness variations, and most of it is due to regional efects.

>> I am providing a measure of goodness of fit for quantile regression, based on R1, suggested by Koenker, R and Machado, J. (1999, Goodness of Fit and Related Inference Processes for Quantile Regression, Journal of the American Statistical Association, 94, 1296-1310). Although this measure has been criticized even for their authors, who have even not added in their packages ("qantreg"), I decided to use it to show the 95% quantile fit well for both linear and segmented models. R1 is estimated as 1 minus the ratio between the sum of absolute deviations in the fully parameterized models and the sum of absolute deviations in the null (non-conditional) quantile model. The values are useful for comparisons between quantile models, but they are not comparable to standard coefficients of determination. The latter is based on the variance of squared deviations, whereas goodness of fit values for quantile regression is based on absolute deviations. The goodness of fit values for QR (or R1) will always be smaller than R2 values.

In concrete, I added R1 measures in table 1 for each quantile, and added the discussion about 95% quantile is a good model, and in the methods section, I present the rationale behind R1 measure.

Also, I corrected the usage of AIC. Model comparison between 0,1 or 2 breakpoints is a hot topic, and no definitive response exists about the best way to discriminate the best model. While in mean models the BIC appears to perform reasonably well, for QR the SIC (Schwartz Information Criterion), similar to BIC for a mean model, performs even better. Thus, I am now using SIC with the following function (AIC(yourQRmodel, k=log(n)) #n=sample size).

In concrete, I have replaced the names of columns in table 1 and presented properly the SIC criterium.

See the reviewer's assessments for more details on these four points, and several other issues. Among these, let me highlight that you should avoid using MAT as abbreviation in the title (Mean Annual Temperature or simply Temperature would be more clear, and of course informative), and also that this manuscript

desperately needs maps with richness and residual values, to allow the readers to assess your results in a wider extent.

I am looking forward to receive a revised version of the preprint, together with a detailed answer to the comments provided. I'm convinced that your research has enough quality so as to finally merit a recommendation in PCI Ecology.

>> I have changed the title and added two maps: One for the distribution of species richness and other of the distribution of the residuals of the .95 quantile.

Reviews

Reviewed by Rafael Molina-Venegas, 2020-01-08 16:04

The author presents an interesting piece of work that aimed to prove that the kinetic hypothesis of biodiversity may not satisfactorily explain the central tendency in species richness (as it has been previously stated with empirical evidence for and against), but the upper bound (i.e. maximum richness) of the relationship (which according to the author's claims, it has been hypothesized but empirical evidence remains obscure). To do so, he used either linear or segmented quantile regressions, a statistical tool that has been proved useful to provide comprehensive descriptions of biological response patterns in observational studies of limiting factors. Overall, the manuscript is well written, including clearly stated hypotheses/results and fairly transparent descriptions of the methods.think there is potential for an interesting publication in this draft. Yet, I would like the author to address few comments and suggestions before recommending the article, specially those concerning the hypotheses.

>> Thanks for the clear description of this MS.

Firstly, a formal comment on the authorship. While the draft is signed by one single author, the text is written in plural tense (we). Please, fix the text to first person singular, otherwise add the omitted authors.

>> I have changed the tense of the MS writing.

Title: I think the title is too general given that the study focuses just on tree species.would suggest something like "Mean annual temperature drives the variation in maximum potential tree species richness and frost organizes the residual variation".would also avoid the use of acronyms (i.e. MAT) in the title.

>> I have taken this suggestion and replaced the title.

The author state that the model derived from the exponential Boltzmann temperature relationship predicts a negative slope of -0.65 between the inverse of temperature (1.000/kK) and the natural logarithm of species richness, and then cites Allen et al. (2002). These authors fitted linear regressions between the inverse of ambient temperature and the natural logarithm of species richness for multiple taxonomic groups (including

North American tree species), and they presented both the slope of each relationship and the averaged slope across all groups. The only numerical difference between the models of Allen et al. (2002) and the ones presented by the author is that the former used 1000/K instead of 1000/kK, and thus the slopes reported by Allen et al. (2002) are fully comparable to the ones shown by the author if divided by 10. Given the data provided by Allen et al. (2002) (see Figure 1 and Table 1 below), I cannot see where the -0.65 slope is coming from (for example, in the case of North American trees, the slope provided by these authors is -1.005 if temperature is scaled at 1000/kK).

>> I have corrected the citation and I have taken the expected values for the slope proposed Brown et al. (2004) and Allen et al. (2007), and later tested by Hawkins et al 2007a. Also, I have explained the differences between the different expected values in the section Material and Methods.

The text reads "Currie (2007) hypothesized that the model proposed by Allen et al. (2002) only fits the upper bound of the relationship between species richness and MAT in plants, but cannot explain the species richness variation in general". I wish I could have checked such statement in Currie (2007), but unfortunately the citation is a book chapter to which I have no access.

>> I can provide a pdf with the chapter.

In line 33, the author introduces the case study. I would suggest expanding a bit the description of the data, given that "the Americas" is a rather vague term. Also, I think the manuscript will much benefit from a figure showing a map of the study area including the plots.

>> I have added more specific information for what I mean by "the Americas" concept (Lines 79-81). Also, I have added a couple of maps (Fig. 3) to represent the distribution of the inventories, species richness, and residual variation. Therefore, I think the area of study is this version clearer than in the previous version.

In line 48, the text reads "segmented models are more robust in all of the quantiles analyzed", and Figure 1 caption reads "The red lines represents the stronger models [...]". I would suggest using more specific descriptions of the results, because while "robust" and "strong" are rather synonyms, they refer to completely different things in the article (i.e. the most supported models by the AIC criterion and the t-test on the two slopes of segmented models, respectively).

>> I have changed the usage of "stronger". I replaced it with "best supported" models (Legend Fig 1). In addition, I have added an explanation of the way I define best supported in Material and Methods (Lines 373-381).

The author provided slopes, p-values for the t-test on the slopes of segmented models and AIC values. However, such descriptors do not provide information on the goodness-of-fit of the models. I am not particularly familiar with quantile regression, yet I suspect the classical R2 used for models that are based on the conditional mean are not valid for quantile regression. However, a quick search in google suggests that some alternatives exist (e.g. Koenker, R and Machado, J. 1999. Goodness of Fit and Related Inference Processes for Quantile Regression, Journal of the American Statistical Association, 94, 1296-1310). I think the manuscript will much benefit from including any suitable goodness-of-fit metric for the models. After all, the AIC criteria serves to choose the best model, but it does not tell anything about whether the selected model is actually a good descriptor of the data (as the devil's advocate, I may argue that the author is choosing the best model among very bad ones, which might invalidate the conclusion of the study).

>> As suggested by Reviewer 1, I have added the R1 estimate to show the goodness of fit of the models. The goodness of fit (goodfit function from R package "WRTDStidal") measure for quantile regression is estimated as 1 minus the ratio between the sum of absolute deviations in the fully parameterized models and the sum of absolute deviations in the null (non-conditional) quantile model. The values are useful for comparisons between quantile models, but they are not comparable to standard coefficients of determination. The latter is based on the variance of squared deviations, whereas goodness of fit values for quantile regression is based on absolute deviations. The goodness of fit values will always be smaller than R2 values.

In addition, I have explained R1 estimate in the section Material and Methods (Lines 367-372).

Figure 1. Just a suggestion. If the author is going to present this figure in colour, consider using a colour palette instead of just red for a better visual impact.

>> I prefer to maintain the colors of Figure 1 as in the revised version because I need to highlight the model for quantile 0.95, and only secondarily showing if segmented or linear models are best supported toward lower quantiles.

The results of the residual analysis are very interesting. I wonder if the author has considered splitting the dataset between tropical (including both humid and dry tropics, where frost days may not be a big issue but water availability does) and temperate (where frost days may be an important issue rather than water availability) regions to test the hypothesis separately for each group of plots. This analysis should not imply much of an effort, and it may serve to get further insight on the biological response of tree species to environmental thresholds.

>> I have done the analyses separated by biomes, which are organized in tropical and temperate (Figures 2 and 4). Specifically, I show the distribution of the residuals in a "by biome" plot in Figure 4. Also, I have discussed extensively the differences in residual distribution across tropical and extratropical biomes (Lines 256-307).

The final sentence of the article reads: "The structure of wet-tropics rich versus dry-tropics and extratropics poor seems more feasible than the famous pattern of the latitudinal gradient of species richness". I do not get the meaning of such statement.

>> I have rewritten that last paragraph.

Figure S1. While I have no problem with this figure, I think it is unnecessary.

>> I have removed this figure.

Abstract. Please, describe the fact that you analyzed upper bonds rather than central tendencies when introducing the hypothesis. This is indeed the main novelty of the manuscript, and yet it is presented when introducing the results instead. I think this section requires some rewriting.

>> The abstract have been rewritten

Reviewed by anonymous reviewer, 2020-02-17 11:50

This is a nicely conducted, straightforward macroecological assessment of which are the actual effects of temperature on species richness. As any good macroecological work, it has the strengths and limitations of the discipline: while it makes a massive use of data – thus providing robustness, it also makes some oversimplifications – so statements about the real meaning of the relationships found need to be made with caution.

One of these oversimplifications is how you refer to temperature throughout the text (BTW, please avoid using the abbreviation in the title!). Perhaps the most worrying is in the opening paragraph, where you state that mean annual temperature would be the main driver of richness. This is not what Brown, Allen and colleagues said, for the average of the monthly averages means nothing in biological terms. This variable is a proxy for the general temperature conditions in each place throughout the year; as such it tells a lot about the (lack of) harshness of the conditions for life in each place, but it tells nothing about the mechanisms you are talking about here. As Brown and colleagues lined them up in their Metabolic Theory, the limit to life growth (and richness) is given by the temperature of activation of enzymatic reactions, which gives the approximate slope of - 0.65. Therefore, the important factor is the temperature experimented by the cells, and more precisely for how long this temperature is available throughout the year in environmental conditions. This would be the physiological limit imposed by temperature to population growth and several aspects of diversity, including richness. In absence of such value (time with environmental temperatures suitable for enzymatic reactions), or of other a priori more adequate proxies such a Growing Degree Days, Frost Days, or the like, mean annual temperature could be used as a substitute, assuming that the higher the average temperature throughout the year, the more time with temperatures adequate for normal cell functioning, the lower the energetic costs for organic functioning, and the higher the population growth, diversification rates and the conundrum of mechanisms leading to higher species richness.

Well, I'm missing at least part of this theoretical introduction in your text. Perhaps not all of this, but at least a sharper justification for the use of mean temperatures as proxies, indicating the mechanisms linking temperature with richness, and why MAT could be a good proxy for that. In fact, to me the most important of your results (the relationship between residuals and Frost Days) would be highlighted much better if this theoretical framework was properly outlined. And so would be Water Deficit, another measure of the quantity of time available for proper biological activity.

>> I really thank you for this comment. I have rewritten the introduction to both giving a better theoretical background (Lines 26-46) and highlighting the importance of freezing temperatures (Lines 62-72).

Also, I'm missing two key papers by Hawkins et al (2007a,b) from the original evaluation of the Metabolic Theory. In the first a number of authors assessed whether the relationships between richness and temperature followed the -0.65 slope for many groups and regions (almost none of them did follow it;, and in the second a subset of them provided several arguments of why this should be an expected outcome; their main argument was that even if such limit to enzymatic activity was universal, the course of evolution would lead most of biodiversity to escape from the limitations it imposes through evolving many different strategies (endothermy, cell wall structures to create microclimates, etc.) that following my argument above would lead to longer times with adequate cell functioning throughout the year, escaping from environmental temperature conditions. The fact that you find that MAT could be a limit to maximum richness and frost alters these numbers is somehow related with this realisation in Hawkins et al 2007b, although it contradicts to some extent the findings in Hawkins et al 2007a. Referring to such literature, and some other key papers on physiological limits (e.g. Peck et al. 2009, Peters et al. 2016, Brodie 2019, or some papers in a recent Phil Trans issue, see Spicer et al. 2019).

>> I have reviewed the literature suggested for reviewer 2 and I have added some of them in both sections Introduction and Discussion. About the two key papers from Hawkins et al (2007 a/b), they were included in the previous version but absent in the list of references by a mistake from me. Actually, a portion of the Methodology of this manuscript is inspired by Hawkins et al (2007a).

Besides that, your work is undermined by the lack of accounting for regional effects. The slope of the relationship between species richness and temperature is well known to vary in space (e.g. Cassemiro et al. 2007, Hortal et al. 2011, etc.) according to several factors such as glaciations and differences among regional communities (sensu Ricklefs 2008, 2015). So while reading it I feel constantly guessing that the alternative hypothesis that the relationship between richness and temperature varies significantly in space by factors other than frost.

>> I have split the dataset into biomes (Figure 2) and, therefore, I attempt to account for regional influences and delimitate the influence of frost. In addition, I have discussed extensively in the section Discussion the

non-stationarity of the variation of observable species richness below the upper bound defined by Temperature.

To assess this, all time I'm missing a map to locate studies with their richness and another with their residual values. This paper desperately needs these two maps.

>> Thanks for the suggestion. I added the maps (Figure 3) and I agree the manuscript is now easier to discuss and I hope it is easier to read.

But also, the robustness of your results is compromised by the lack of assessments of effect differences between different types of regions. Accounting for this could be easy by adding as cofactors biogeographic realms, biome maps, glaciated/unglaciated areas, etc. I would be really surprised to find that there are no differences between regions, biomes, ecoregions or areas that suffered glaciations or not. But if that is the case, then the importance of your paper would increase a lot.

>> Rather than modeling the regional influence as a cofactor, I did an estimation of the slope in different biomes. This way allows me to discuss the regional effect and put the predictions of Metabolic Theory on a global scale. Thus I avoid going farther than predicted by the Metabolic theory. I agree with the idea that modeling in a sophisticated way the upper bound of the variation in species richness would be a great advance, but I see that this future work needs to be built above the results presented in this manuscript if the criticism allows it.

Besides these major problems, there are several typos (e.g, in the abstract "attempts to explain (…) richness HAVE focused"), but most importantly there are a few idiomatic problems here and there, so a light English revision is needed. In particular, the following sentences need to be reworked:
- Last sentence of the abstract is confusing, it needs to be rewritten
- Lines 67-69. This sentence is also confusing ("significantly further" does not work well), rephrase and clarify. While doing so, try to relate it with the sentence in lines 90-94, which states more or less the same; right now the text feels repetitive, so it may be worth splitting the information between the two paragraphs, linking them in terms of content.
- Lines 103-104. I can't tell what does "by drop in" mean in this context, and most readers will not understand either. Please rephrase.
>> Thanks for these minor comments. I have reviewed the text and asked for a third part of reading to improve the writing.

Brodie, J.F. (2019) Environmental limits to mammal diversity vary with latitude and global temperature. Ecology Letters, 22, 480-485.

Cassemiro, F.A.d.S., Barreto, B.d.S., Rangel, T.F.L.V.B. & Diniz-Filho, J.A.F. (2007) Non-stationarity, diversity gradients and the metabolic theory of ecology. Global Ecology and Biogeography, 16, 820-822.

Hawkins, B.A., Albuquerque, F.S., Araújo, M.B., Beck, J., Bini, L.M., Cabrero-Sañudo, F.J., Castro Parga, I., Diniz-Filho, J.A.F., Ferrer-Castán, D., Field, R., Gómez, J.F., Hortal, J., Kerr, J.T., Kitching, I.J., León-Cortés, J.L., Lobo, J.M., Montoya, D., Moreno, J.C., Olalla-Tárraga, M.Á., Pausas, J.G., Qian, H., Rahbek, C., Rodríguez, M.Á., Sanders, N.J. & Williams, P. (2007a) A global evaluation of metabolic theory as an explanation for terrestrial species richness gradients. Ecology, 88, 1877-1888.

Hawkins, B.A., Diniz-Filho, J.A.F., Bini, L.M., Araújo, M.B., Field, R., Hortal, J., Kerr, J.T., Rahbek, C., Rodríguez, M.Á. & Sanders, N.J. (2007b) Metabolic theory and diversity gradients: Where do we go from here? Ecology, 88, 1898–1902.

Hortal, J., Diniz-Filho, J.A.F., Bini, L.M., Rodríguez, M.Á., Baselga, A., Nogués-Bravo, D., Rangel, T.F., Hawkins, B.A. & Lobo, J.M. (2011) Ice age climate, evolutionary constraints and diversity patterns of European dung beetles. Ecology Letters, 14, 741-748.

Peck, L.S., Clark, M.S., Morley, S.A., Massey, A. & Rossetti, H. (2009) Animal temperature limits and ecological relevance: effects of size, activity and rates of change. Functional Ecology, 23, 248-256.

Peters, M.K., Hemp, A., Appelhans, T., Behler, C., Classen, A., Detsch, F., Ensslin, A., Ferger, S.W., Frederiksen, S.B., Gebert, F., Haas, M., Helbig-Bonitz, M., Hemp, C., Kindeketa, W.J., Mwangomo, E., Ngereza, C., Otte, I., Röder, J., Rutten, G., Schellenberger Costa, D., Tardanico, J., Zancolli, G., Deckert, J., Eardley, C.D., Peters, R.S., Rödel, M.-O., Schleuning, M., Ssymank, A., Kakengi, V., Zhang, J., Böhning-Gaese, K., Brandl, R., Kalko, E.K.V., Kleyer, M., Nauss, T., Tschapka, M., Fischer, M. & Steffan-Dewenter, I. (2016) Predictors of elevational biodiversity gradients change from single taxa to the multi-taxa community level. Nature Communications, 7, 13736.

Ricklefs, R.E. (2008) Disintegration of the ecological community. American Naturalist, 172, 741-750.

Ricklefs, R.E. (2015) Intrinsic dynamics of the regional community. Ecology Letters, 18, 497-503.

Spicer, J.I., Morley, S.A. & Bozinovic, F. (2019) Physiological diversity, biodiversity patterns and global climate change: testing key hypotheses involving temperature and oxygen. Philosophical Transactions of the Royal Society B: Biological Sciences, 374, 20190032.

\>>