

27 August 2018

*Revised Submission of bioRxiv 152215*

Dear Dr Mooers

Thank you for the opportunity to resubmit a revised version of our manuscript **Range size dynamics can explain why evolutionarily age and diversification rate correlate with contemporary extinction risk in plants** for consideration by *PCI Evol Biol*. We are very grateful for the comments provided by yourself and in the two reviews. The reviews have been very helpful in improving our MS and I am sorry for the delay in responding to them.

Our point-by-point responses to the reviews follow this letter. We understood that the main concerns necessitating revision of our manuscript were the need to: (1) expand the rationale for our contrast-based test and PGLS of proportion data; (2) justify the use of GBIF data for measuring niche filling; and (3) clarify the environmental mismatching idea. We have now added extensive text in the Methods that explains why a sister-species analysis can complement a regression of range size on species age, as requested by you and Reviewer #2. We have also added text to the Methods that explains why a PGLS analysis is entirely appropriate for proportion data, addressing a concern of both Reviewers. Finally, we have removed the analysis of niche filling and environmental mismatching idea given that you and both Reviewers found these points unclear or poorly justified.

We also want to draw your attention to another major change in the MS. While preparing our statistical scripts for upload, we noticed that some of the data were being filtered out of the analyses. Correcting this has roughly doubled our sample size, resulting in slightly different results from our original submission. These results are however more logically consistent.

Again, we thank you and the Reviewers for the insightful feedback and hope that our revisions, which comply with all of the Reviewers' suggestions, will make the manuscript clearer and of broader interest and we look forward to your response.

Yours sincerely,

Dr Andrew J Tanentzap

University of Cambridge

Decision by Recommender Arne Mooers:

The authors are interested in relating several macroevolutionary quantities of angiosperms -- stem age of a genus, diversification rate of a genus, and (for palms and conifers) the pendant edge length (or "age") of a species -- with the probability of being listed by the IUCN. As potential moderators of possible relationships, they measure (for palms and conifers) both the number of cells with GBIF records as a measure of range size and the difference between this number and the number of climatically suitable cells as a measure of range "filling." There are several motivations: patterns might offer possible mechanisms for (or even predictors of) extinction risk in plants, perhaps via a very old hypothesis linking the age of a lineage and its total range size. As the authors state, there are few known predictors of whether a particular plant species is imperiled, though Vamosi and Wilson did find that species in smaller families (especially monotypic families) were more likely to be at risk of extinction, while Davies et al. found that younger species of plants in the Fynbos of South Africa were more likely to be at risk of extinction, perhaps due to the small ranges associated with recent peripatric speciation in this group.

**We thank the Recommender for drawing attention to Davies et al. and Vamosi and Wilson. Our study builds upon their work by considering links between taxon age and range size dynamics across a broader suite of species. As we already cite both of these sources in the Introduction, no further changes are made.**

As pointed out by reviewer 2, the framework of the study might need a bit more work. Older species may indeed be more generalist for various reasons (sampling, past adaptation), and so be buffered against certain threats. Older species may also have larger ranges, or may occupy a larger fraction of available range as a function of time to disperse, though this latter idea would need some justification on the large temporal and geographic scales considered here. And (or maybe but) generalist plant species may also have larger ranges as a function of their generalist lifestyle. And, of course, threatened species have smaller ranges almost by definition. This means that picking these four measures (age, generalism, range size and range filling) apart as predictors of threat may be tough.

**In response to Reviewer 2, and shared comments from the Recommender, we have extensively re-written the MS to improve the framework of the study. Our changes have involved: 1) expanding on the role of niche generalization in the Introduction (lines 48-58); 2) adding a verbal model linking time since speciation, speciation mode, and range filling on lines 59-70 (full text of changes provided below in response to this specific comment); and 3) better explaining how age, generalism, and range size can be picked apart in the Introduction (lines 90-93) and Methods (lines 175-177, 187-191).**

**In response to the Recommender's specific comment above, we have also given some justification to the idea that older species have had more time to disperse.**

*Added on lines 40-43:*

**"While differences in the time for dispersal may weaken over long time scales (i.e. millions of years), younger taxa may also face less available space and resources as niches fill through time irrespective of dispersal ability (Tanentzap et al. 2015)."**

Consistent with some previous work, the authors do find evidence ( $r=0.15$ ) that faster-diversifying genera indeed have a higher proportion of IUCN-listed species. The relationship with genus age was consistent, with younger genera also having a higher proportion of species at risk, though weaker (the authors state "no relationship", but the difference in the r-

value of the fit of the two predictors is only 0.02 – I would definitely remove this from the abstract).

This sentence has been removed from the Abstract as we now find a significant association with our larger dataset.

Unfortunately, small sample sizes made the test of this relationship within the conifers and palms inconclusive, though no formal power analysis is offered.

We now offer a formal power analysis and refer to it in the main text.

*Added on lines 207-209:*

“Repeating our analyses with only the genera from the more complete conifer and palm species-level datasets was also inconclusive (table B3), potentially because of small sample sizes ( $n < 70$ ; fig. B1).”

*Added on pg 54 of Electronic Supplemental Material:*

“**Figure B1.** Power analysis of association between threat status and diversification rate. We used the dataset of 509 genera with high-quality divergence time estimates and IUCN data. We sampled between 25 and 500 genera at a time, repeating each sampling intensity 100 times at random. For each replicate, we fitted the same PGLS model described in the main text and extracted the  $p$ -value associated with the effect of diversification rate on threat status to generate a distribution of values at each sampling intensity. Points are mean  $p$ -values  $\pm$  95% confidence intervals. Dashed line denotes  $p = 0.05$ .”

On the bright side, consistent with these patterns, threatened conifer species are also younger on average than non-threatened species, though in palms, threatened and non-threatened species are indistinguishable in age.

All these are interesting patterns that deserve consideration, and the fact that palms and conifers differ is intriguing. In a final set of experiments, though, the authors consider whether total range size or their measure of range filling can help explain the conifer pattern. They contrast threatened vs. non-threatened sister species of conifer, and look at the difference in range attributes as a function of the age of the contrast. Just looking at the plots, it seems that older pairs of threatened vs. non-threatened species differ more in the proportion of their range that is filled, but less in the size of their ranges, with equal explanatory power (again,  $\text{abs}(r) \sim 0.15$ ). The authors report that the null model, which considers pairs of species that do not differ in risk status, returns an expected  $r$  of -0.12 for both these contrasts vs. age. In the null model, the threat status is the same, such that there is no directionality to the contrasts, and so this negative relationship with age is based on the absolute difference in these quantities. It might also make sense that young pairs of species could differ more in range size or range filling than older pairs of species, e.g. under a simple peripatric model of speciation. The observed data are, of course, signed (NT-T, though it seems that the NT species always has the larger range, and more of its range filled because all y-axis values are positive on both plots. While I believe the first range-size pattern, the second range-filling pattern is itself perplexing.) Generally, I am not quite sure how the unsigned null model is to be used to interpret the signed real data. The second reviewer and I found this difficult to interpret, and so I suggest the methods here need to be clarified a bit more.

We have now extensively tried to clarify the Methods and Results.

*Added on lines 170-195:*

“We also tested how potential range size was associated with species age in both conifers and palms. ... One limitation with this analysis is that it does not compare threatened and non-threatened species of the same age, and so can introduce biases if there are systematic differences in the ages of these two groups.

To further analyse how potential range size was associated with species age, we undertook a second comparison that focused on pairs of sister species with contrasting threat status. For each pair, we calculated the difference in potential range size between the sisters, so as to avoid pseudoreplication, and correlated this with their age. We compared this association to when sisters had the same threat status to test the null hypothesis that being threatened with extinction does not change age-range associations. Focusing on sister pairs was desirable because it can minimize factors that confound age-range associations, such as unobserved extinctions (Hodge and Bellwood 2015). Range differences can also shed light on the underlying mode of speciation. For example, there may be greater disparity in the ranges of young species pairs under peripatric as opposed to allopatric speciation (Gaston 1998; Hodge and Bellwood 2015), resulting in a negative correlation between age and range asymmetry. We tested if this correlation was different from randomly sampling the same number of sister pairs 1,000 times, but choosing those where both members of the pair had the same threat status. We chose both species to be non-threatened for conifers and both to be threatened for palms as most identical species pairs in the two clades fell into these two categories (70/85 and 51/68, respectively).”

*Added on lines 217-222:*

“A smaller potential range size increased the extinction risk of older conifers, supporting the specialism hypothesis. We specifically found that non-threatened conifers had narrower ranges as their age increased relative to sister species that were threatened (fig. 3); ranges in neither threat status independently changed with age (table B1). As the age of conifers increased, this difference between sister-species pairs of contrasting threat status was larger than expected if sisters had the same threat status ( $r = -0.27$ ,  $p = 0.025$ ; fig. 3a).”

We also struggle to understand the concern about an unsigned null model. This sort of scenario would occur by definition through conventional significance testing of a correlation. Specifically, in the absence of our null model test, we might wish to assess whether the slope of the correlation is significantly different from zero. This determination would be made by randomly sampling the data and calculating slopes just like we have done here. Again, just like we have done here, the model would be effectively unsigned. We would still calculate NT-T, but the difference would be effectively randomly assigned through the permutation of values across threat statuses. The problem with this conventional null model test is that we are not only interested in whether there is an association between age and range attributes. We also want to know how this association compares to the situation when sisters do not differ in extinction risk to test the null hypothesis that extinction risk does not result in something “special” occurring.

More generally, tests based on contrasts are often preferable (since they control so well for many other aspects of biology), but this approach does require more justification in this instance because contrast tests are designed to control for age, and age is a specific part of the model here – reviewer 2 raises the same issue.

As quoted above, we have added text to justify our approach further. Specifically, we are trying to control for biases in the estimation of age *within* sister pairs. We do not control for age more generally in our model as we are estimating contrasts among pairs of different ages.

There would seem to be other, more straightforward things to test in order to get a better understanding of what is going on: we know from these data that threatened species have smaller ranges, but do younger species of conifers also have smaller ranges (or perhaps less filled ranges, which might be a proxy for long-term dispersal)? If so, might phylogenetic path analysis be useful - e.g. compare the model [age -> positive range attributes -> decreased risk] with the simpler direct models that have [age-> decreased risk] or [positive range attributes -> decreased risk]? If, however, age does not predict range attributes across conifers, then this more focused contrast-based test might be the way to go, if explained more clearly.

*Added on lines 170-185:*

“We also tested how potential range size was associated with species age in both conifers and palms. First, we used PGLS to test whether older ages correlated with larger range sizes, which by definition reduce extinction risk (IUCN 2016), and allowed the effect to vary with threat status (i.e. statistical interaction). We expected threatened species would, by definition, always have relatively small ranges, producing an invariant or weak age-range association. By contrast, non-threatened species should reach larger ranges with time if the age-and-area hypotheses was supported, whereas the reverse could be expected under the specialism hypothesis. One limitation with this analysis is that it does not compare threatened and non-threatened species of the same age, and so can introduce biases if there are systematic differences in the ages of these two groups.

To further analyse how potential range size was associated with species age, we undertook a second comparison that focused on pairs of sister species with contrasting threat status.”

All that aside, it is still not clear to me how the main observation - that the difference in range filling between non-threatened and threatened species increases with age - should be best interpreted. One possibility is that non-threatened species fill their range faster than threatened species (e.g. due to dispersal). Under the peripatric model of speciation (which I invoked to explain the null model negative correlations), the smaller-ranged budded species is also the one with the less filled range at the outset, and time should make sister species more similar for both measures. But under this model, the larger-ranged species is ancestral, and so should be closer to its equilibrium range filling, and it should be the small-ranged species that starts filling suitable habitat. Because threat is something relatively new, the pattern might imply that species that cannot fill their suitable climatic range (perhaps due to biotic interactions or "specialism"), and actually fill it less and less through time are more likely to become threatened. Is this what you are invoking? Likely not, because range expansion and contraction is evoked for conifer speciation. But such a model does not make strong predictions about the difference in range size or range filling in young vs. old pairs of species, does it? (It might predict that older species have larger or more filled ranges, but that is not what is presented). A very clear verbal model that links time since speciation, speciation mode, and range filling would help me. Overall, while I like much of the data and the patterns, I am not yet sure of the mechanistic explanations proposed in this paper.

We thank the Recommender for this comment as it has helped us phrase how we think time since speciation, speciation mode, and range filling might mechanistically interact.

*Added on lines 59-70:*

“The potential for species to expand their range and reduce extinction risk with time may ultimately depend on their mode of speciation. Repeated range expansion and contraction (i.e. “taxon cycles”) that isolate peripheral populations consistent with centrifugal or peripatric speciation can produce small ranges in descendent taxa (Gaston 1998). Consequently, older species may have a lower extinction risk because they have had more time to disperse and expand their range, and experience less niche pre-emption from earlier evolving competitors (Tanentzap et al. 2015). Lineages with high diversification rates under this mode of speciation can similarly face greater extinction by producing species that have small ranges (Schwartz and Simberloff 2001). By contrast, any signature of time in extinction risk distributions may be absent with vicariant speciation because asymmetry in the ranges of ancestors and daughter species is consistently smaller and ancestral species often disappear via cladogenesis (Gaston 1998).”

*Added on lines 187-191:*

“Range differences can also shed light on the underlying mode of speciation. For example, there may be greater disparity in the ranges of young species pairs under peripatric as opposed to allopatric speciation (Gaston 1998; Hodge and Bellwood 2015), resulting in a negative correlation between age and range asymmetry.”

Both reviewers highlight some technical issues concerning GLMs and proportion data, and reviewer 2 asks for evidence that the measure of range filling is reliable. It is interesting that point data like GBIF are usually used as input to produce SDMs, not to compare with SDMs to get a measure of range filling. I am not expert in this area, especially for mechanistic SDMs, and so cannot comment further, but do wonder if there is a way to support the reliability of this measure, especially given how haphazard GBIF data are.

To avoid repeating large sections of text, we please ask you to refer to pages 8-9 of this reply, where we have addressed the point about proportion data in full. We have also removed the measure of range filling from the analysis, eliminating the concern of Reviewer #2.

The idea of environmental "mismatch" as a mechanism for a positive age-risk connection likely needs some extra justification as well: both reviewers picked up on this. Cycads and long-term global cooling might be a better example than ice ages and rhinoceroses (if narrower available niche space is indeed a driver of cycad decline). I add this at the end because, of course, you do not find any such relationship, and so perhaps you do not have to set it up at all.

We have removed this idea from the MS, so the comment is no longer relevant.

PS: Additional remark by the Managing board.

-Details of the quantitative analyses (e.g. data treatment and statistical scripts in R, bioinformatic pipeline scripts, etc.) in the recommended preprints must be available to readers in the text or as appendices or supplementary materials, for example.

*Added on line 195:*

“R code to perform our analyses is in Data S1.”

Review by Dan Greenberg:

Tanentzap et al. in *Greater range filling can explain why evolutionarily older and slower diversifying plants are less threatened by extinction* present an intriguing analysis examining patterns of contemporary threat with respect to evolutionary dynamics across plant genera. The authors do a good job of setting up the alternative expectations between lineage age and extinction risk according to different hypotheses and use geographic data to test the hypothesis that range filling may contribute to these patterns of threat. I like the overall story and approach of this paper, and I believe their analysis to assess a potential mechanism by examining range filling in two well studied clades represents a step forward for understanding why macroevolutionary dynamics may influence contemporary species threat. My comments are largely minor in nature, although I do believe that one aspect of the current analysis has a methodological issue that should be addressed and I've included a few suggestions and code on how to move forward with this.

We thank the Reviewer for their positive feedback. We have addressed each of the comments below, including their methodological concern with the analysis.

Lines 13 – 14: Important to be clear that it's a higher proportion of species threatened with extinction. Small distinction, but some might dismiss this as a moot point as a larger clade would be expected to have threatened species.

Thanks for pointing this out. We were trying to save words in the Abstract.

*Changed on lines 14-15:*

“We found that a greater proportion of species were threatened by extinction in younger and faster-diversifying genera.”

Line 15: I'm a bit confused by the statement “they were older, on average, than expected” where does the expectation of age derive from? I presume (after reading the paper) you mean in terms of the distribution of lineage ages from the entire plant phylogeny, but it's not clear here.

This sentence has been removed from the MS so the comment is no longer relevant.

Line 31-32: I think this statement is a bit strong. Non-random extinction patterns across the tree could (in theory) be due to chance/human activities if there's simply a phylogenetic signal in where species occur (ie. close relatives are geographically close). I would just re-word this to eg. ‘suggesting that chance events and human activities alone may not be fully responsible for species losses’.

The Reviewer makes an excellent point.

*Changed on lines 30-31:*

“...suggesting that chance events and human activities alone may not be fully responsible for explaining species losses”

Line 57-62: Is the example of the woolly rhinoceros evidence that evolutionary age may influence extinction risk? Were more recently diverged rhinoceros species less extinction prone in this case? I'm not convinced this is a good example to illustrate that evolutionary age increases extinction risk unless you can contrast how other, younger, species escaped the climate trap. Another example could be cycads – once widely distributed but now restricted to a much narrower niche space (and consequently quite threatened in the present, see comment below). Beyond this particular example this paragraph sets up the dichotomy of higher or lower extinction risk in early split lineages nicely.

We have taken the Recommender's advice and have removed this text from the MS.

Line 78-79: Though not a direct test of lineage 'age' (but largely analogous) – a recent paper suggests threat is concentrated in evolutionarily distinct cycad species – see Yessoufou et al. 2017, *Ecology & Evolution*.

Thanks for point us to this fine paper. We now mention it on lines 254-256.

Line 137 – 142: I like this approach, but one potential concern is spatial barriers. To what degree is range filling influence by spatial disjunction between realized and potential range size? It seems like the primary hypothesis is in terms of eg. biotic interactions limiting range filling, so if there are some species with considerable spatial barriers (oceans, mountain ranges, etc.) could this influence your conclusions on range filling? I don't think it invalidates the method, but it's something to discuss and consider at least.

We have removed the analysis of range filling given the concerns about this measure from the Recommender and Reviewer #2, so this comment is no longer relevant.

Lines 145 – 147: Technically this violates the underlying assumptions of the generalized least squares model, as the error isn't normally distributed (it's bounded within 0 to 1, a binomial error distribution). This is a fairly common mistake due to the complexity of phylogenetic models with different distributions. While I suspect this ultimately doesn't bear upon the results, it would be better to do this in a generalized linear model (GLM) with a binomial error distribution.

Unfortunately, this is a bit complicated due to the phylogenetic structure in the data. There are a few options however:

1. You could run this as a binomial GLM with random effects for taxonomy (as a surrogate for phylogeny). This isn't ideal, but it's reasonable.
2. You could transform these proportions to the logit or arcsine square root. There are issues with both of these (logit – cannot have a 0 or 1, as these tend towards infinity on the logit scale; arcsine – can handle a 0 or 1, but has less than ideal properties as a distribution see Warton & Hui 2011, *Ecology*). You could then run this in a pglS framework.
3. You could implement a phylogenetic GLM, there's two packages that handle this well and they both use a Bayesian MCMC sampling framework: 'MCMCglmm' (see Hadfield et al. 2010) or 'BRMs' (Buerkner). (see links below). These would be the ideal approaches, but will require some careful consideration of interpreting Bayesian models, ensuring model fit and chain mixing, and choice of priors. I've included some example code that would implement this sort of phylogenetic model in 'MCMCglmm' at the end of this review to could help get you started, but you should familiarize yourself with either method and tune the model to your dataset.

We thank the Reviewer for their helpful suggestion and generous provision of code. The choice to use PGLS was not however a "mistake". It was decided upon after extensive testing of all the potential options (including MCMCglmm) and discussion with Tony Ives, who is one of the leading developers of phylogenetic models.

Generally, there were two reasons for our decision to treat binomial data as normally distributed for hypothesis testing via PGLS:

- 1) **PGLS performs similarly and sometimes better in terms of type I errors and power than phylogenetic GLM or MCMCglmm** (Ives and Garland 2014). Ives (2015) reported similar results from comparing LS and GLMs more generally, especially when the latter were misspecified. Importantly, while parameter estimates may differ between PGLS and other methods, PGLS is entirely appropriate for testing the null hypothesis of no statistically significant effect of an independent variable, as we do here, rather than being used to predict future events (Ives and Garland 2014; Ives 2015). These conclusions are consistent with the simulation results of O'Meara et al. 2015 for a binary dependent variable, which found phylogenetic GLM poorly recovered known distributions and signs of dependent variables.
- 2) **Only PGLS can weight observations by sample size.** Weighting is necessary to account for the different number of IUCN assessments that each genus received.

We apologise that these issues were not explained in the Methods and we have now done so.

*Added on lines 149-156:*

“Although the least squares model assumed normally distributed errors, and the response variable was a proportion with binomial errors, PGLS is appropriate for testing the null hypothesis of no statistically significant effect of an independent variable on a non-Gaussian response (Ives 2015). We also fitted the PGLS regression using the *gls* function in R because this approach, unlike other model fitting functions that incorporated phylogenetic information (e.g. *phyloglm*), could account for different sample sizes across genera by weighting observations with the inverse square-root of the number of IUCN assessments that they received (Garamszegi and Møller 2010).”

*References used in this response*

- Ives AR & Garland T Jr. 2014. Phylogenetic regression for binary dependent variables. *In: Modern phylogenetic comparative methods and their application in evolutionary biology* (ed. LZ Garamszegi), pp. 231–261. Springer-Verlag: Berlin.
- Ives AR. 2015. For testing the significance of regression coefficients, go ahead and log-transform count data. *Methods in Ecology and Evolution* 6: 828-835.
- O'Meara BC *et al.* 2015. Evolutionary models for the retention of adult–adult social play in primates: The roles of diet and other factors associated with resource acquisition. *Adaptive Behavior* 23: 381-391.

Review by Anonymous Reviewer #2:

There is much to like about this analysis, which explores the relationship between diversification, range size and extinction risk across vascular plants. However, limits to the available data and some rather weak conceptual links somewhat detract from the paper overall.

We thank the Reviewer for their positive comments. We have tried to strengthen the datasets and conceptual links as per their comments in order to improve the paper.

Previous work has demonstrated that, at least in some clades, more rapidly diversifying lineages tend to be characterised by a greater proportion of threatened species and suggest that this is likely a reflection of the mode of speciation, whereby new species are often range restricted and of low abundance, and thus at high risk of extinction (or at least appear to be at high risk, given IUCN Red Listing criteria). The present study adds qualified support to these findings. More rapidly diversifying (but not younger) genera tend to contain a greater proportion of threatened species. Further, in conifers, older taxa are less threatened with extinction, but there is no relationship between species age and extinction risk in palms. The authors suggest these clade differences can be explained by the contrasting modes of speciation that characterise the respective clades – conifers have diversified historically via range-shift dynamics, whereas palms have diversified by exploiting new niche space. Here the Reviewer summarises our results rather than providing specific comments on the MS. Nonetheless, we have added text in response to their point about the appearance of extinction risk in relation to IUCN Red Listing criteria.

*Added on lines 111-114:*

“Threat status is jointly determined from abundance, recent temporal change in population size, and various measures of geographic distribution, such as occupancy and fragmentation (IUCN 2016). Therefore, metrics of range size alone may not entirely predict extinction risk despite the potential to use these terms interchangeably.”

I am broadly sympathetic to the authors' arguments and find the results highly believable. However, I was not convinced by the link with range filling, mostly because I do not think this metric is reliable as calculated. Occupied range is estimated using GBIF occurrence data (incidentally, summing non-equal area cells), but obviously species occur in many more areas/cells than those recorded in GBIF, and it is for this reason we frequently use SDMs to 'fill-out' species distributions. Without independent empirical evidence, I am unconvinced that the difference between SDM predictions and GBIF occurrence data captures range filling.

We have removed the analysis of range filling.

*We have also added the following on lines 141-143:*

“We then summed the total number of equal-area (Mollweide projected) 0.25 decimal degree grid cells occupied by each species.”

Relatedly, there is a discrepancy in the timescales of environmental change (i.e. the tens or hundreds of thousands of years over which glacial cycles occur) and species ages (which are measured in the millions of years), which undermines the authors' environmental 'mismatch' hypothesis. The example of the woolly rhinoceros is illustrative: if it lost its habitat after the

retreat of the last ice age ~18K years ago, it is not obvious how this should relate to its evolutionary age.

We agree with this point, which was made by the other Reviewer as well. Following the Recommender's advice, we have removed the associated text from the MS.

Further, the alternative scenario, which suggests that older species might be more generalist, and thus better 'matched' to the environment, ignores the potential trade-off between specialist and generalist. I would, for example, hypothesise that specialist not generalist would show a better 'match' to the environment.

*Added on lines 49-58:*

“Older species may have survived long-term environmental changes because they are more generalist (Liow 2007). As broader niches are positively associated with larger ranges (Slatyer et al. 2013), this explanation would result in another positive age-and-area association. By contrast, there may be a negative correlation between age and extinction risk if older species are more specialised and have smaller ranges. We term this idea the evolutionary specialism hypothesis. Older species can appear more specialised because traits that were once advantageous became less adaptive as environments diverged from past selection regimes (Wilson 1959; Žliobaitė et al. 2017). More specialist species with narrower niches and geographic ranges may only persist over long time periods in refugia or by having large local population sizes (Williams et al. 2009).”

Evaluating range correlations of extinction risk in plants is particularly challenging as the Red List status of many species is, more or less, defined by their range size; it can, therefore, be difficult to tease apart these attributes and there is a danger of circularity. For the most part, the authors escape such pitfalls, but it might be useful to remember that extinction risk and range size may be interchangeable.

*Added on lines 111-114:*

“Threat status is jointly determined from abundance, recent temporal change in population size, and various measures of geographic distribution, such as occupancy and fragmentation (IUCN 2016). Therefore, metrics of range size alone may not entirely predict extinction risk despite the potential to use these terms interchangeably.”

### **Minor Points**

A PGLS model might not be the most appropriate for analysing proportion data (e.g. in the regression of taxon age against proportion of threatened species).

There were two reasons for why we analysed proportion data using PGLS:

- 1) **PGLS performs similarly and sometimes better in terms of type I errors and power than phylogenetic GLM or MCMCglmm** (Ives and Garland 2014). Ives (2015) reported similar results from comparing LS and GLMs more generally, especially when the latter were misspecified. Importantly, while parameter estimates may differ between PGLS and other methods, PGLS is entirely appropriate for testing the null hypothesis of no statistically significant effect of an independent variable, as we do here, rather than being used to predict future events (Ives and Garland 2014; Ives 2015).
- 2) **Only PGLS can weight observations by sample size.** Weighting is necessary to account for the different number of IUCN assessments that each genus received.

We apologise that these issues were not explained in the Methods and we have now done so.

*Added on lines 149-156:*

“Although the least squares model assumed normally distributed errors, and the response variable was a proportion with binomial errors, PGLS is appropriate for testing the null hypothesis of no statistically significant effect of an independent variable on a non-Gaussian response (Ives 2015). We also fitted the PGLS regression using the *gls* function in R because this approach, unlike other model fitting functions that incorporated phylogenetic information (e.g. *phyloglm*), could account for different sample sizes across genera by weighting observations with the inverse square-root of the number of IUCN assessments that they received (Garamszegi and Møller 2010).”

The analysis of range size, range filling and species age using sister-species is rather odd as sister species are, by definition, the same age. P-values are estimated using random draws of sister pairs, which I assume is why the positive slope between difference in range filling and age, with  $r = 0.15$ , is highly significant, whereas the negative slope between difference in range size and age, with  $r = -0.16$ , is not at all significant. I am not sure what to make of these analyses.

*Added on lines 170-195:*

“We also tested how potential range size was associated with species age in both conifers and palms. ... One limitation with this analysis is that it does not compare threatened and non-threatened species of the same age, and so can introduce biases if there are systematic differences in the ages of these two groups.

To further analyse how potential range size was associated with species age, we undertook a second comparison that focused on pairs of sister species with contrasting threat status. For each pair, we calculated the difference in potential range size between the sisters, so as to avoid pseudoreplication, and correlated this with their age. We compared this association to when sisters had the same threat status to test the null hypothesis that being threatened with extinction does not change age-range associations. Focusing on sister pairs was desirable because it can minimize factors that confound age-range associations, such as unobserved extinctions (Hodge and Bellwood 2015). Range differences can also shed light on the underlying mode of speciation. For example, there may be greater disparity in the ranges of young species pairs under peripatric as opposed to allopatric speciation (Gaston 1998; Hodge and Bellwood 2015), resulting in a negative correlation between age and range asymmetry. We tested if this correlation was different from randomly sampling the same number of sister pairs 1,000 times, but choosing those where both members of the pair had the same threat status. We chose both species to be non-threatened for conifers and both to be threatened for palms as most identical species pairs in the two clades fell into these two categories (70/85 and 51/68, respectively).”

*Added on lines 217-222:*

“A smaller potential range size increased the extinction risk of older conifers, supporting the specialism hypothesis. We specifically found that non-threatened conifers had narrower ranges as their age increased relative to sister species that were threatened (fig. 3); ranges in neither threat status independently changed with age (table B1). As the age of conifers increased, this difference between sister-species pairs of contrasting threat status was larger than expected if sisters had the same threat status ( $r = -0.27$ ,  $p = 0.025$ ; fig. 3a).”

Finally, the authors should refer back to earlier papers, for example, those by Schwartz and coauthors, who explore similar patterns in extinction risks across diverse plant clades and propose similar mechanistic explanations.

*Added on lines 65-67:*

“Lineages with high diversification rates under this mode of speciation can similarly face greater extinction by producing species that have small ranges (Schwartz and Simberloff 2001).”

*References used in this response*

- Ives AR & Garland T Jr. 2014. Phylogenetic regression for binary dependent variables. *In: Modern phylogenetic comparative methods and their application in evolutionary biology* (ed. LZ Garamszegi), pp. 231–261. Springer-Verlag: Berlin.
- Ives AR. 2015. For testing the significance of regression coefficients, go ahead and log-transform count data. *Methods in Ecology and Evolution* 6: 828-835.