

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.

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.

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.

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'.

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.

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*.

Line 137 – 142: I like this approach, but one potential concern is spatial barriers. To what degree is range filling influenced 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.

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 help get you started, but you should familiarize yourself with either method and tune the model to your dataset.

```
##Basic code for MCMCglmm implementation##
```

```
library(MCMCglmm)
```

```
##The tip label identifier in your dataset needs to be named 'animal' (referring to an animal model)
```

```
data$animal<- as.character(tree$tip.label)
```

```
row.names(data)<- make.names(data$animal,unique=TRUE) #Sets the tip label to each data entry
```

```
#Set a standard (uninformative prior) – R = residual variance and G = variance attributed to phylogenetic autocorrelation; it's highly likely you will have to adjust this (usually increasing nu to sample a broader parameter space)
```

```
prior<- list(R=list(V=matrix(1),nu=0.02), G=list(G1=list(V=matrix(1),nu=0.02)))
```

```
MCMCglmm(cbind(N.threatened,N.not_threatened)~covariate,random=~animal,pedigree=tree,data=data, family="multinomial2",nitt=110000,burnin=10000,thin=100,prior=prior)
```

#You will want to play around with the thin parameter (and the number of iterations (nitt)) to make sure you have a high enough effective sample size for each parameter and variance estimate (~1000). Make sure to check for autocorrelation in chains and adjust the thinning accordingly. Follow the advice in <https://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf>

#You can calculate Pagel's lambda as the ratio of variance attributed to animal/ (animal + residual variance) – see Hadfield & Nakagawa (2010), *J. of Evol. Biol.*

For an alternative, check out 'BRMs'

<https://github.com/paul-buerkner/brms>

[https://cran.r-project.org/web/packages/brms/vignettes/brms\\_phylogenetics.html](https://cran.r-project.org/web/packages/brms/vignettes/brms_phylogenetics.html)