> We are grateful for these new and again thoughtful reviews. We are excited to present our improvements in light of those reviews. Below we detail our responses to each comment and in the manuscript text our new edits are highlighted in blue.

Reviewer 1

I thank the authors for taking on board my comments. In particular it helped that the authors clarified the relationship between beta and niche structure, so that (IIUC) the parameter beta should be interpreted just as one among multiple other traits that may plausibly be similar for organisms within the same taxonomic group.

> We thank the reviewer for their detailed evaluation of our work which has greatly improved its quality.

My remaining remarks relate to the model assumptions and I am going to reiterate one of my earlier comments that I hope the authors can address or comment on.

** there is still some ambiguity on whether I should think of there as being net diversification or not. The authors say in their supplementary information that "most fossil occurrences of a taxon come from the period of its history when it is dominant and in steady state. In a time slice of duration τ during such a period of steady state the latent per capita rates of origination and extinction would be equal (i.e. $\lambda = \mu \equiv \rho$)".

This seems at odds with the need (stated more clearly elsewhere) to zero-center the time series of species richness, thus in effect removing the effects of net diversification (i.e. removing any net increasing or decreasing richness within a taxonomic group).

The authors also say that

"We also note that the distributions of fluctuations for most clades are already very close to a mean of 0"

So perhaps this small difference is why this doesn't seem like a big deal either way—but I'd like to see it stated more clearly in the model description whether they authors are in effect considering a model with net lambda-mu non-zero, fitting this model, and then focusing on lambda+mu (the variance)—or if this is not the right way to think about their approach.

> This final statement is indeed very close to our thinking. We do not exactly fit a birth-death model and then investigate the sum of its parameters, instead we take a data-driven approach to first zero-center the time-series of richness fluctuations and then look at the variance of that zero-centered time-series. In as much as the fluctuations time series is a sufficient statistic of the birth-death model our data-driven approach is very much related to the model-driven approach outlined by the reviewer. We recognize the confusion around this key assumption of no net diversification in our analysis and have made efforts to further clarify our approach and assumptions, specifically lines 76–81 and supplemental section S1; see also the discussion below.

** The model described in S1 with net (per taxon) diversification rate lambda-mu =0 is in effect a model analogous to pure drift—correct? To derive the expected form of Gaussian fluctuations around the mean taxon abundance the authors state that

"Under the steady state assumption we can approximate N(t) by mean(N)"

On short timescales it would be reasonable to treat the number of taxa in a higher taxonomic group, N, as equal to its mean value, and hence I'd expect the authors' stated result. However, if the baseline of the time series relative to 1/lambda is large, then I would expect to see large excursions away from the mean value under this pure drift model. I think it would be helpful to clarify whether this is a self-consistent assumption given the parameter values the authors find.

We agree with the reviewer and have sought to clarify two key points to better demonstrate our agreement (see lines 76–81 and supplemental section S1):

1) We do not want to wed our analysis strictly to a linear birth-death process, because, for example, a stochastic process with an upper limit would also be consistent with our findings (analogous to a logistic-like 'top-hat' model [e.g., Liow, L. H., & Stenseth, N. C. (2007). The rise and fall of species: implications for macroevolutionary and macroecological studies. Proceedings of the Royal Society of London B. 274: 2745–2752.]). This is why we do not rely parametrically on the linear birth-death process, but instead use the data directly. We acknowledge that our previous version did invoke linear birth-death and we have now removed this limitation to the interpretation of our work.

2) The approximately Gaussian fluctuations arise because of time averaging, and not just the stationarity assumption, therefore we have made the importance of time averaging more clear in the supplemental section S1.

** Related to this—the authors use the terminology 'steady-state'. But there is no steady- state in the linear birth-death model (apart from the absorbing state when a family or order disappears altogether). So this terminology should be clarified. Or if I am missing what is meant by this, maybe the steady-state in this model could be defined.

We should have specified 'stationary' instead of steady state, this has been changed.

** In the birth-death model I was not quite sure what would be the "null" dependence of beta on family size, N—i.e. how would beta depend on family size if lambda and mu were identical for all families. I had thought that beta might be proportional to (or at least increasing with) N, which was somewhat born out by the three families shown in Fig 1A and C. I.e. the smaller mean richness in the Spondylidae goes with the smallest fitted value of beta. Etc.

However, Fig S6 shows the opposite relationship altogether—roughly this figure shows that beta~1/N, and this seems worthy of more comment! In fact, I can't even quite see where the Spondylidae family can fit into fig S6, given that they have a (standardized) mean richness of 1 genus and fitted beta ~ 10^{-1}. But maybe this is just down to the way the y-axis is defined in Fig S6 vs Fig 1A.

We should first note that we are not explicitly using a linear birth-death model, but rather a data-driven approach such that we can be agnostic about the specific dependence of per captita rates on standing diversity (we clarify this throughout, specifically lines 76–81 and supplemental section S1). Nonetheless, the likely key to interpreting the relationshipe between β and clade size is that in Fig S6 we look at the total richness of a clade, which includes turnover through time, and so while the Spondylidae tend to have 1 "standardized" genus in any given time interval, the identity of genera changes and so the group's total richness is greater than 1 standardized richness unit. This helps explain where Spondylidae might fall in Fig S6, but also helps understand why beta ~ 1/N: the time-series of fluctuations is not an average, but rather a summation of +1 and -1 for each genus origination and extinction, in the time intervals when those events took place. Viewing the fluctuation of each time interval as a sum of random variables means that the variance of that fluctuation should increase with increasing N. Because $\beta = 1/\sigma^2$ this is consistent with β decreasing with increasing N. Even if we looked at average clade richness on the x-axis in Fig S6 we would see this same negative correlation because fluctuations still result from a summation and not an averaging and average richness is related to total richness by a constant—the number of time intervals. We add this to the commentary in lines 209–217.

In any case, this is an interesting relationship that seems like it has some power in explaining the variability in fitted values of beta across families. However, the authors say it is not sufficient to explain the observed

variation in betas. I did not quite understand how the permutation test described in the main text was showing this. But maybe here is another way of thinking about it. Suppose I fitted beta ~ const/N using Fig S6. How much worse would this do than the gamma distribution that is fitted to observed betas in Fig 1C?

> The permutation test is trying to achieve exactly this suggestion: derive the set of β values that would result from N-dependence alone and see how much worse the gamma and resulting P(x) distributions are. We have added this interpretation of the permutation experiment at lines 209–217.

In summary I think Fig S6 deserves more comment and readers may want to understand better why we would see beta ~ 1/N.

> We agree and have added more commentary, following from the discussion above, at lines 209–217.

** Overall, one suggestion that might clarify some of the above questions might be to infer the best fit parameters for the birth-death process from the observed transitions for each time series, allowing both lambda and mu to be fitted parameters, and then these more clearly related to beta.

> We agree that such an analysis could indeed be illuminating and perhaps ripe for a future publication. We refrain from implementing it here because of the previously discussed desire to not confine our analysis to only the realm of a simple, linear birth-death process. We feel that trying to determine exactly which process(es) are best suited to which clades and then analyzing the inferred parameters, while interesting in its own right, distracts from the simplicity of our intended message about the applicability of superstatistics in describing fat-tailed fluctuations and directing the development about hypotheses for their macroevolutionary origin.

Reviewer 2

The authors have responded to comments in my original review by (1) adding an analysis of families, (2) revising parts that argue for the utility of the superstatistics approach, and (3) revising some of the process interpretations based on the results. I appreciate the family analysis, which is similar to that of orders in showing macroevolutionary coherence within taxonomic units.

> We thank the reviewer for their first round of comments and for again evaluating our revision, helping us to improve its clarity.

As for changes (2), I feel pretty similarly as I did about the first version of the manuscript. The methods are interesting and, as far as I can tell, valid. I am still not sure that have clear advantages over other approaches, but I could be wrong, and it would be good to release these ideas into the wild, so to speak, and see how they fare.

> We appreciate the reviewer's open mindedness in suggesting that the wilds of our scientific community can decide the ultimate utility or lack thereof of our approach.

In terms of interpretation, the revisions are fine, though I think the same limitation holds in that the main result — macroevolutionary dynamics are more within clades rather than between clades — is not very surprising. But there may be value in the specifics of how this work breaks down this heterogeneity.

> Here we agree that the result of macroevolution between clades is not revolutionary, but that such an established idea can be sufficient to explain an otherwise unexplained pattern (fat-tailed fluctuations) is more novel—a point we try to make clear in the text.

One side note about interpretation: the paper uses the metaphor of the adaptive landscape to frame the interpretations, but in places I lost the thread of this framing. The adaptive landscape in macroevolution is usually used in the interpretation of trait evolution, not speciation/extinction dynamics. I think the connection made here is the assumption that the clade volatilities (beta_k) are related to trait values, which may evolve by OU-like dynamics on an adaptive landscape. Does there need to be an additional assumption about the nature of this relationship - linear, perhaps?

> The connection of our work to adaptive landscapes indeed requires refinement. We should make clear that the traits themselves are not responsible for volatilities, but rather the macroevolutionary fitness conferred by those traits to the clade are what we hypothesize to determine the clade's volatility. Genera proliferate or decline due in part to the traits they posses and this is the fitness consequence of the traits. We make that clearer in lines 91–93. One other point to note is that Simpson's original fitness landscape was aimed not only at traits (which constitute the "geographic coordinates" of the landscape) but also macroevolutionary rates which come from the elevation of the landscape, making our narrative a natural extension of his. We solidify that interpretation in lines 61–62 and 91–95.

I was tripped up in interpreting this quote: "For macroevolutionary rates to vary across an adaptive landscape, this landscape cannot be flat, and thus niche conservatism around local optima in adaptive space interrupted by adaptive exploration is likely (27,47)."

What does elevation on this landscape represent? For OU dynamics, a flat landscape indicates Brownian motion. Brownian motion still produces trait variation, and If trait variation causes differences in clade volatilities (beta_k), then Brownian motion will still produce variation in macroevolutionary rates. So I am confused as to how the adaptive landscape is construed in this interpretation.

> Here the key, which we admitted we could (and now do) make more clear, is that it is not the traits per se which should correspond to volatility but rather the macroevolutionary fitness of those traits. Traits will indeed change in a flat landscape, but this would be be consistent with the super-statistical pattern we observe. We clarify in lines 246–249 and 91–95.