

We would like to thank the editor for allowing us to address the reviewer's comments and suggestions. We are also grateful for the comments and suggestions that have, without doubt, improved the quality of the manuscript. We address the individual comments below: the reviewer's comments are in black, **our actions are in purple** (also marked with response).

Dear William Foster and colleagues,

As you can see in the online discussion, we have received one additional review report on your revised manuscript. The reviewer is supportive of your work but raises a couple of points that they believe you should address before acceptance. I agree with the reviewer, and am therefore happy to provisionally accept your manuscript barring minor revisions. Before definitive acceptance, I would like to ask you for one final round of revision to improve the contents of the manuscript based on the reviewer's suggestions. Please submit a revised version and a reply to the reviewer. I will review this revised version before my final decision, so there will be no extra review round.

The reviewer raises several valuable points, most of these seem like they could be implemented quite straightforwardly. Two main issues stand out to me:

Of specific interest seems to be alignment of results between supplement S1 and the main paper and the potential to include some of the outcomes in that supplement in the main story. I believe that would be worthwhile to try, but if the supplement is not integrated in the main text an effort may be needed to resolve the apparent differences in the outcomes between S1 and the main discussion flagged by the reviewer.

RESPONSE: We have checked for consistency and also incorporated direct and relevant results into the text, rather than the Table S1.

The reviewer also suggests to conservatively limit the analyses to taxa with at least 20 observations, but I personally would urge caution. If my understanding of your analysis is correct, taxa with lower amounts of observations are already weighed down statistically in your analysis. In that case I would personally refrain from removing these datapoints from the analysis at the risk of unnecessary throwing away data.

RESPONSE: I think there is misunderstanding here or at least we misunderstood the reviewer's comment if this is what they meant. Instead, we have added all results to the Table S1. We agree that results and data should not be omitted.

I look forward to receiving your revised manuscript and your final response to the review comments.

Kind regards,
Niels de Winter

Report 1

Thank you for addressing our suggestions. The availability of the script and provided output of supporting analyses in supplementary material makes the results and interpretations easier to reproduce. The added information on the paleogeographic reconstruction, potential constraints of the impact of growth rates, data availability during crisis or facies changes also make the manuscript easier to follow. I feel you made a significant effort to accommodate as much as possible of raised points. As you provide critical new data and raise interesting point related to the discrepancy between species and genus responses, I would love to see this interesting work published. There remain however some crucial points which need additional addressing:

1) Sample size and significance of changes mentioned in supplementary materials: I greatly appreciate the inclusions of significance and sample size in the supplementary material. I do not completely understand why some crucial analyses such as those in table S1 are not included in the main manuscript. These are critical to understand your results, so it would in my opinion be critical to include this table and/or values in the main manuscript. I also agree that it makes sense to err on the side of caution (e.g., only listing those comparisons with a sample size of 20). However, I was also a bit confused by the listed significant changes in text versus those available in table S1, which often deviate as some comparisons claimed to be significant are not available in this table neither are some of the pooled samples (e.g., *Neoschizodus*/*Schizodus*) which suggest they are not significant or are considered of too low samples size to be considered meaningful. Please make sure to add pooled samples mentioned in text also in this table and make sure the mentioned significance (and sample size) are consistent between text and table.

RESPONSE: (1) We think it is okay to leave the table in the supplement, but bring out some of the test results when specific cases are mentioned. E.g., We are trying to hit a balance between too much info that the reader stops following the main results vs. as much detail as possible. (2) We have added the test result for *Neoschizodus* and *Schizodus* to the text. This was simply overlooked, so thank you. (3) We also checked for consistency between the text and results, as you are right that there should not be discrepancies!

2) Rejection of the lilliput effect and overprint of poor sampling of surviving species: I feel the wording is too broad sweeping given the paucity of information for surviving species. At least this potential difficulty and its impact on such analyses as well as the need for additional high-resolution studies like yours should at least be highlighted more both for the dolomites as well as the understanding of the lilliput effect more generally. In this respect, additional discussion on the differing results when using relative abundances may also be critical. It may be not surprising that within lineage effects seem more pronounced during the crisis and less pronounced during the transition from the Permian to the Triassic. By your decision to not focus on abundance, there seems to be a clearer within-lineage before (between Lo4 and Lo5) as well as a more pronounced effect appearance and disappearance of taxa (between Lo5 and In1) which would be expected as you are focusing on the most dominant species rather than treating each species equally (in your approach you are effectively penalizing against undersampled taxa which would not yield significant or spurious results due to low sampling issues in your approach).

RESPONSE: This is an interesting comment and every approach, of course, has strengths and weaknesses. We already toned down our rejected of the Lilliput effect "Summarised, our data reject, or

at least question, the Lilliput effect“ so we will not make further adjustments there. Regarding the differences with the Rego diagrams, we have added more text to add the more nuanced interpretations highlighted by the reviewer.

When revising there are also some additional points which should be considered.

Line 165: I greatly appreciate the inclusion of the approach using specimen abundances. I feel pointing some of the difference with the approach when correcting for abundances, would be critical as the differences are more than just extreme fluctuations. There is clearly a stronger relative effect of within-lineage changes between Lo4 and Lo5 and a different/stronger effect of appearance and disappearance between Lo5 and In1. Discussing it however is rather something for the results or discussion section.

RESPONSE: Changes made in the results section and not here in the methods, see also main comment 2 from the reviewer above. We did, however, delete this comparison from the methods.

Line 169: you mention extinction, but can you be sure in each case that a disappearance represents a global extinction rather than a local extirpation. It would make sense to discuss this nuance if relevant to some of the taxa (e.g., at least those temporally disappearing from Italy dolomites relative to Hungary).

RESPONSE: This is a misreading of our text. “In this region, only Bakevellia ceratophaga, survived the extinction event.” Our original text conveys the reviewers request and no changes were made.

Line 169: it seems this result is not significant (or sample size considered too low) as B. ceratophaga is not listed in table S1. This needs to be more clearly stated in main manuscript. As mentioned previously, it would be critical to add table S1 to the main text and make sure the discussion of significance is consistent between text and table.

RESPONSE: It is clear in the text that it is a comparison between the Middle Changhsingian and extinction aftermath, because there are no late Changhsingian occurrences, which is why there is no value in Table S1 which only compares time bins with immediate pre- and post-data. We also never wrote that it was a significant increase.

Line 181: the size increase in Eumorphotis lorigae seems not significant (or sample size too low to be considered meaningful) as it is not listed in table S1. This needs to be more clearly stated. It raises the question, if differing size responses on species level could not just relate to poor(er) sampling.

RESPONSE: p-value added to the text, and text changed to insignificant size increase.

Line 182: I am confused as according to table S1 only Towapteria and Unionites show significant size decreases (not the other 2 genera). Please make sure discussion of significance in text is consistent with table.

RESPONSE: p-value added to the text, and text changed to show that 3 genera have a significant size decrease.

Line 183: According to your table S1, Bakevellia shows no significant size decrease

RESPONSE: We wrote that it shows a significant increase, which is correct. P-value added to the text.

Line 186: Please add these pooled analyses result to your table as it is currently not listed in table S1 making it impossible to verify.

RESPONSE: We have added the result to the text.

Line 207-208: Given “The changes are driven by the increase in the number of smaller-sized species and the abundance of smaller specimens.”, it would make even more sense to discuss the discrepancies between the specimen abundance approach versus the one corrected for it. This may potentially be the place to do it rather than method section (see comments related to Line 165).

RESPONSE: Changes made, see also main comment 2 from the reviewer above.

Line 221: Bakevellidae are not listed in table S1, please add or do you mean Bakevellia? Please make sure the results and their possible significance is significant.

RESPONSE: The text is correct and the text does not suggest any statistical test was applied here. Likewise no members of the Bakevellidae show a size decrease, which is already in Table S1. No changes were made.

Line 222-223: The significance is missing for the pooled sample, please add it to table S1. More generally, I feel it is critical to add this table and significance to the main manuscript.

RESPONSE: This p value is now already reported in the text.

Line 250: “Species-level the Lilliput effect is not observed”: as state before, it would be crucial to discuss potential constraints in picking up these effects on species level versus genus level given your conservative approach which may inadvertently overemphasize spurious data of rarer species, obvious sampling issues in surviving species or difficulties in phylogenetic relationships or agreement of taxonomic assignment on species or even genus level.

RESPONSE: The text is accurate, no size decrease at the species-level was observed. The situation that is being described by the reviewer here suggests that we got lucky and all our post-extinction data happen to be the larger representatives, whereas the pre-extinction data is more even or even the smaller representatives of the species, which is not a parsimonious explanation. We agree that more data and sampling is required, but this database is based on decades of work and is unlikely to change with sampling campaigns in the near future.

The outcomes are robust and no caveat is required.

Line 251: Given none of these taxa is listed in Table S1, these changes may relate to sample size and are not significant, so should be interpreted with caution and this needs to be more clearly highlighted.

RESPONSE: We wrote “retained or slightly increased”, not that there were significant changes. This is also the Discussion section, and the p-value and test results are reported in the results section.

Line 252: If such species are pre-adapted, one may wonder why they are or do not become more abundant during this crisis.

RESPONSE: We do not explore this point here, as anything would just be pontificating. However, the evolution of new species would have led to intraspecific competition and extinction. So a species may survive the event, but later show a demise following the evolution of better adapted species that outcompete the survivors.

Line 258: It is true that many studies report changes at genus level, but this may at least in part also be related to the focus, design or data availability of the studies or around extinction events more critically. Extinction events often coincide with larger stratigraphic gaps or obvious sampling issues for particularly rare and surviving taxa so these may hamper or actually pick up species-level changes if they were there at all. It would be worth underlining this issue (e.g., the need for high-resolution and intensive sampling).

RESPONSE: We thought about adding this, but decided that it would divert the focus. The study by Metcalfe et al. (2011) for example who investigated genus-level changes, was an undergraduate MSc project who collected body size data over 1 week in the Dolomites, or Twitchett (2007) was done at the genus-level and does not have the taxonomic expertise for specific identifications. It feels unnecessary to highlight details like this here. We also introduced the requirement for high-resolution studies in the introduction and we will keep the text concise.

Line 262: It is not entirely clear which publication is referenced here as it is not cited in reference list. I assume this may be a typo. Only Rita et al. 2018 states this more clearly (see annotated pdf for additional discussion and full bibliographic details).

RESPONSE: Corrected to 2019.

Line 277: It is reassuring that there is no clear change in juvenile specimens, but the lack of data on growth rates and ontogeny may hamper recognizing specimens that may be maturing at different sizes. If there is strange impact on growth rates, they may even coincide with switches in allometry/morphology (e.g., Nätscher et al. 2021 which you already cite).

RESPONSE: The suggestion is fair, but the reality is that this is something that we cannot investigate with our specimens.

Line 300: *E. lorigae* seems not be listed in table S1 which suggests these changes may be non-significant or too low of a sample size to be considered. Please clarify this.

RESPONSE: P-value added to Table S1 and the text in the results section.

Line 307: In my opinion, "reject" is too strong of a wording. I feel stating there is more support for the impact of turnover than within-species changes. At least with the large sampling gaps or low sample size of some of the rare surviving species and difficulties/disagreements on assignment on the species level of others which seems to relate at least in part to preservation, such patterns could at least be partially an artefact of sampling and taxonomic issues. It would be critical to acknowledge that further data is needed for these rare species in the dolomites as well as more generally for other groups or regions or crises.

RESPONSE: See response to main comment.

Line 413: With the available data and provided analyses, I feel you cannot entirely rule out sampling issues particularly for rare species as opposed to genera where more data is available. This should at least be stated more clearly before trying to resolve the underlying abiotic and biotic factors.

RESPONSE: We agree that sampling issues cannot be ruled out, in that certain species have a low number of individuals. But not that we didn't try hard enough to find more specimens, i.e., further sampling is unlikely to uncover more specimens that would change the results of this study.

Line 428: If you use this r-/K-concept that is not entirely uncontroversial, the way it used in your work as well as its history should at least be discussed and supporting reference(s) provided (see suggestions in annotated pdf).

RESPONSE: We have removed the sentence from the discussion.

Line 435: "may create" would be more appropriate than create given this seems to be a hypothesis rather empirically evidenced.

RESPONSE: Text changed.

Line 452: I feel that it may be appropriate to thank the reviewers particularly if you feel it helped to improve your work.

RESPONSE: Of course, we are very grateful! We wouldn't normally write this in until a manuscript is accepted. For Biogeosciences, since our discussions are online, the reader can also see that we acknowledge the reviewer's time! Shown by the fact that we actually engaged with the comments that improved the manuscript.