

We would like to thank the editor for allowing us to address the reviewer's comments and suggestions. We are also grateful to the three sets of 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).

Reviewer 1: Renato Posenato

Foster and co-authors investigate the size changes of bivalves across the end-Permian mass extinction by analyzing a large amount of specimens mostly from the Dolomites. The novelty of this study is that the analysis is performed between bivalve taxa occurring before and after the extinction. The comparison is based on data from the literature. The data repository indicated in the manuscript is empty/missing (!).

RESPONSE: This is not just a literature review, much of the data is new (having taken more than a decade to collect) and some is already published. The key pre-extinction data has not been published elsewhere.

All data and code was pushed to github and was checked before submission. The repository has been updated, and we will also upload the scripts and data to Zenodo upon acceptance of the manuscript.

The results and discussion of the "Lilliput effect" are, however, biased by merging the data of bivalve taxa occurring in the "Extinction interval" (i.e., lower Tesero Member and Gerennavár Limestone Formation) with those occurring in the Mazzin Member of the Werfen Formation. The bivalve assemblage of the Mazzin Member comprehends the species recording the peak of the extinction after the most severe and devastating effects on the marine biota, whereas those bivalve species from the "Extinction interval" belong to the last Permian-type assemblages with articulate brachiopods and calcareous algae. These brachiopods and algae extinguished close to the base of the Triassic, around the first appearance of *H. parvus*. I think that the bivalves occurring in the "Extinction Interval" (latest Changhsingian-basal Induan) must be separated from those occurring in the Mazzin Member (lower Induan). Consequently, the exceptions of the Lilliput effect suggested throughout in the manuscript must be reconsidered as well as the number of the Permian survivors in the Triassic.

RESPONSE: (1) the bivalves from the boundary clay of the Bükk Mts. most likely corresponds to the Tesero Member in the Dolomites and prior to the Permian/Triassic boundary, as suggested by the reviewer. The samples are binned according to the stratigraphic sequence and the occurrences are correctly placed with the In1 sequence. We agree that these should then belong to the extinction interval, and we will re-write the text to capture this. We did not, however, edit Fig. 2 or separate the data in the analysis, because that would cause an oversplit of the data. (2) This does not affect the text nor conclusions,

as there are still no examples of a size decrease at the species-level, i.e., the so-called Lilliput effect, and the exceptions, remain exceptions.

*Stutchburia tschernyschewi* is proposed as a Permian survivor which does not show the size reduction with respect to the Permian individuals, and thus it would be another exception to the Lilliput effect. However, this species occurs in the late Induan when the size is increased in all the bivalves. This species, not yet discovered in the Mazzin Member, is so rare (cf. Hofmann et al.) that a statistical comparison has no value.

RESPONSE: We agree that you need to be careful with overinterpreting the significance of *Stutchburia tschernyschewi*, because we cannot say with 100% confidence (based on the preservation of the Werfen Formation specimens) that those specimens before and after the extinction are definitely the same species. We also acknowledge there is a dearth of specimens recovered from after the mass extinction, which means that rigorous statistical comparisons for changes in body size, something that we did not do in this article, are not possible. We can amend the text to capture this issue.

The final statements of the Introduction declare that the Dolomites are “one of the best-studied regions ... for the Permian-Triassic crisis”. Although dozens of articles on this topic have been indeed published in high-impact journals over the last 40 years, only an extended abstract by the authors (Kustatscher et al.) is cited to support this claim. This abstract is also cited in line 94, even though the eastward deepening of the basin is well known in literature since the 1970s. References are also missing about the well-known Triassic carbonate platforms for which the Dolomites are renowned throughout the world. Only a field guide book is cited (Stefani et al., 2004). The Authors omit previous papers dealing with the topic of their study. These serious oversights of literature should be corrected by the authors.

RESPONSE: We cannot cite every article from the Dolomites for the Permian-Triassic transition, but we can add a few more key citations from the region. We aimed to include papers that summarise the history or capture the onshore-offshore gradient.

There is no indication of the referred age of the paleogeographic map of the Dolomites (Fig. 1 has already been published in other papers which must be necessarily referred to). The paleogeography of the Southern Alps remarkably changed during the Late Permian, with repeated transgressive and regressive cycles. Nonetheless, this happened even during the Early Triassic when the sea level rose tens of meters and the sea invaded the Lombardy.

RESPONSE: It is difficult to capture the dynamics of sea-level change with a single stagnant figure. The aim of the figure is to give an idea of palaeogeography during the interval and show relative differences between the investigated sites. We have added information to show which interval the map best represents, i.e., the Permian/Triassic boundary.

In the abstract, Authors write: “These measurements come from the Bellerophon and Werfen formations of the Dolomites in Italy, representing relatively shallow marine environments”. However, in the paleogeographic map the material was collect in sedimentary successions ranging from shoreface to basin settings. What age (Changhsingian, Induan or Olenekian?) does the “relatively shallow marine environments” refer to? Since this presumed great difference in depth, why is it not considered in the Discussion?

**RESPONSE:** In the western Dolomites, where all studied material was collected, both the Bellerophon (Changhsingian) and Werfen (Induan–Olenekian) formations consist of subtidal to lower-shoreface deposits above or close to the storm wave base, representing a shallow-marine ramp environment. Although the formations extend eastward into more distal, basinal facies, these settings were not sampled in our study. Consequently, all analysed material derives from shallow-marine successions, and differences in water depth among the investigated sections are minor and do not influence our interpretations. However, please also see the other reviewers comment and our response relating to water depth.

Editor: Niels de Winter

During the discussion of your manuscript, some of the reviewers highlighted that the link in your Code and Data availability section (<https://github.com/wjf433/DBiv>) leads to an empty GitHub repository. We at Biogeosciences are very keen on keeping all aspects of the review process open, and this includes making the code and data you use in your study available to the reviewers so they can judge the entire methodology and outcomes of your work. This includes code and data availability, and I wanted to stress that manuscripts without open availability of crucial elements of the methodology (such as code) will not be acceptable for publication. I therefore urge you to make your code repository available for the reviewers and indeed the rest of the community, preferably with a DOI.

The latter means that, if you prefer to keep using GitHub to share your code, you will have to link your GitHub repository to an online storage medium such as Zenodo which offers stable DOI's for GitHub releases. For more information on how to do this, please refer to the following guide on Zenodo: <https://help.zenodo.org/docs/github/>.

Needless to say, you are free to use any other repository as long as you can guarantee that the version of your code used to produce the outcomes exposed in your manuscript is accessible through a DOI to the readers and reviewers.

**RESPONSE:** All data and code was pushed to github and was checked before submission. The repository has been updated, and we will also upload the scripts and data to Zenodo upon acceptance of the manuscript.

Reviewer 2: Kenneth de Baets

First of all, I want to congratulate you on amassing this unprecedented data and analyses to studying these crucial aspects in Permian-Triassic bivalves from the dolomites. Given the scope and great interdisciplinary relevance of this analysis, I would love to see this published. However, I feel there are some crucial aspects which need to be addressed before publication (part of these can also be found in my online comment):

**RESPONSE: We really appreciate the positive and constructive feedback!**

1) Size changes across mass extinctions and/or extreme environmental perturbations: It is refreshing to see you tackling size changes in bivalves in higher resolution across a mass extinction as well as the associated issues such almost complete turnover, genus versus species. It may be worth pointing out (e.g., around lines 50-52) that similar difficulties may arise also during smaller crises, particularly when marked environmental perturbations occur (e.g., secondary extinction events such as Pliensbachian-Toarcian crisis). For example, it was challenging to study the TOAE as belemnite become rarer at particularly localities (e.g., Peniche) and a complete turnover is visible linked with a belemnite size increase (Rita et al. 2019). However, more comprehensive data obtained from a latitudinal gradient shows that there is consistent evidence for a size increase driven by appearance of larger species although the mechanisms behind it may be multifold warranting the need for additional studies (De Baets et al. 2021).

**RESPONSE: We also see latitudinal migrations during the Permian-Triassic mass extinction (*sensu* Bomfleuer et al., 2017; Foster et al., 2023). The impact of high magnitude events on understanding body size from a single location has been added the text. Hopefully this makes the text more directly relevant to other hyperthermal events.**

2) Mechanisms of the Lilliput effect *sensu stricto*: I feel it would be helpful to even more explicitly discuss about a Lilliput effect *sensu stricto* (e.g., original discussion by Urbanek 1993) versus *lato* when other factors are involved. In this respect, it is worth highlighting that the mentioned ammonoid body size pre-extinction decrease by Kiessling et al. (2018) occurs above species-level and not at the species level (line 273). The Lilliput effect *sensu stricto* was attributed to smaller adult size at the species level. However, in the modern and fossil assemblages, both smaller adult size and higher proportion of early (and therefore smaller) ontogenetic stages (e.g., juveniles) may contribute to smaller body sizes (Daufresne 2009; Ortega et al. 2016; Rita et al. 2019). In the fossil record, smaller specimens may be rarer or less preserved than larger specimens under particular circumstances which could also cause the opposite effect (e.g., we suggested it as a possible explanation for increase in belemnite size distribution during TOAE: De Baets et al. 2021). I would like to see a discussion on these factors in the discussion.

RESPONSE: It is very important to distinguish between, what the reviewer calls here, the Lilliput effect s.s and s.l. We can make the distinctions much clearer in the text. In the introduction we have specified “e.g., *is it the size-reduction at the species-level or genus-level that is being observed (Brayard et al., 2010, 2015), is it the preferential loss of larger species (Nützel et al., 2010), or is no body size reduction actually observed at all (Nätscher et al., 2023).*” and in the discussion we have also made sure we are clear. We have also done the same in the Discussion also referring to species- or genus-level when discussing the Lilliput effect. We have not added the s.s and s.l additions because that can be confusing in its self.

The second part is looking at changes in the population structure to explain the observed trends and we have added this aspect to the discussion too.

3) Impact of changes of facies and sea-level changes: Mass extinctions are often associated with marked stratigraphic changes associated with sea-level changes which may not only impact diversity but also preservation and organismal traits such as size (Holland 2000, 2020; Holland and Patzkowsky 2015). De Baets et al. 2022 for example found a relationship between changes in facies/lithology and body size distribution of Devonian ammonoids during background conditions with smaller sizes in black shales as opposed to carbonates or marls cautioning to interpret body size changes during marked changes in facies/lithology. Correcting for it, interestingly we found specimens at lower latitudes to be bigger rather than smaller which was the opposite effect as expected. Rita et al. 2019 in belemnites or Nätscher et al. 2023 in ostracods found no strong effect of lithology although these authors focused on individual sections which may not represent the full spectrum of (background) environmental/depositional conditions (compare De Baets et al. 2021). You seem to have a unique opportunity here to see the effect across multiple localities with the same region/basin and I feel this could be exploited more or at least discussed in greater detail (compare also online comment by Renato Posenato). I would at least like to see a small discussion in the discussion on the potential effect of these factors on your dataset. In the aftermath of crises, sometimes unlikely groups may benefit (e.g., deepwater brachiopods with low metabolism: Uhlmann et al. 2023).

RESPONSE: It has already been shown that controlling for different water depths is important for the post-extinction body size trends (*sensu* Foster et al., 2020), where the larger bivalves are typically found in the sediments deposited below wave base. We did not feel that we had enough data in this study to include this approach here, and given that Foster et al. (2020) have already shown this for the Triassic, we can just add the importance of this point to the discussion. Regarding our analysis, each sequence (except LOP3 and OI3), captures a similar range of water depths, meaning each sequence should include sediment deposited from supratidal to below wave base settings, allowing for our trends to remain a fair test. Regarding lithology specifically, bivalves are found in the same lithologies across the Bellerophon/Werfen Fm. boundary.

A second reason to not investigate water depth in this study, is because the sedimentological data for the Permian successions is not complete. However, it is possible to compare the body size between location along an onshore-offshore gradient, providing new insights into the Permian taxa. For the Permian taxa, we found that body sizes from the most proximal and likely shallowest succession (Tramin) recorded significantly smaller body sizes, but there were no significant differences between the shallow water section in the Dolomites and the deeper water sections in the eastern Dolomites. We have added this new information and figure to the manuscript.

4) Sample sizes and pooling of samples: from our own analyses, I am quite understanding of difficulties of obtaining sufficiently large sample sizes across extinct events as well as the impact on statistical power when sample size is lower. 3983 body sizes are unprecedented and impressive for this time interval and region. However, when we subdivide it by the number of studied species and time bins, there seem to be species/genera which are only represented by a single data point while other will be presented by tens to hundreds of data points. When looking at size distributions – 20-30 data points per species per bin would be minimum (likely more when large ontogenetic changes in size are expected: compare Nätscher et al. 2021) to draw meaningful conclusions. In this context, I would be crucial to explicitly provide sample sizes (per species or genus) and designate significant changes in

Figure 2. I also feel violin plots may be more appropriate to show the actual distribution of data (is their evidence multimodal distributions in case of multiple species). When dealing with changes within genera, it would be particularly interesting to see if there a change from one species to next species in the next bin (e.g., anagenetic lineages with different species assigned to different genera) or rather from multiple species from one bin to the next time bin. I also feel it is equally crucial for other figures (4-6) to add information on significance (e.g., by connected changes bracketing boundaries and adding as asterisk/different symbol for it) as well as sample sizes in the graphs. Also, consideration of violin plots more generally in these graphs may also be helpful if not in main text, at minimum in the supplementary material. In addition, choosing time bins and pooling specimens at such levels may also affect interpretations (see also online comment by Renato Posenato).

**RESPONSE:** This is indeed a very important data exploration step, but what would be important for the article is to focus on the important taxa in our record. We already discuss this at the genus-level, and we can add the point for the species-level.

We have added box plots, with the number of observations above each box, for each species to the supplemental dataset to show the distribution of data. We did not utilise violin plots, due to personal preference, but also due to scaling issues.

Regarding showing significant changes, we tried adding an \* and arrows to the plots to show significant changes and direction of body size changes, but it didn't look good. Instead, we have put a table in the supplement.

5) Rego et al. approach: I completely agree that changes in components may be sensitive to changes in relative abundance within the time interval using the Rego approach (line 155; see also De Baets et al. 2021). However, the question remains if focusing on dominant taxa where more (or at least sufficient) data is available is not more reliable or at least more significant particularly when attempts were made to sample and include all available specimens (compare Rita et al. 2019). In this respect, I would still like to see additional discussion in the main manuscript on the differences when correcting or not correcting for abundance.

**RESPONSE:** We thought we did this, and a comparison figure can be included in the supplement between the two approaches. The original approach basically leads to more exaggerated changes, although the general trends in our study stayed the same.

This figure and comparison can be found in the supplement.

6) Impact of taxonomic practices and preservation: It is great to hear that you standardized taxonomy which is crucial for such studies, and I have little doubt with bivalve experts involved that you thought long and hard about it. For the reader, it would be helpful to see some additional explanation on how this standardization was achieved. More generally, I also wonder if stronger size changes at genus level versus weaker changes in species may not be a partial artefact of stratigraphic gaps, preservation and taxonomic practices at such critical intervals. In an ideal world, we would have large samples of well-preserved specimens at all ontogenetic stages without (major) stratigraphic gaps allowing to make meaningful subdivision based on known phenotypic plasticity, ontogenetic variation and fluctuations over time between species based on morphology. Unfortunately, this is rarely the case across extinction events associated with major environmental perturbations such as End-Permian Mass extinction (or even secondary extinction like the TOAE - sometimes even growth may be affected and in different ways in different species: compare Nätscher et al. 2021). In this context, it may be tempting (or collide with conservative taxonomic principles) to assign specimens (particularly those showing marked differences in size and morphology at individual ontogenetic stages or high phenotypic plasticity) separated by large stratigraphic gaps to different species versus lumping them within genera. It would be crucial to see a larger discussion on how taxonomic challenges may contribute to the patterns you observed. You already hint at some of these aspects in text.

**RESPONSE:** This is a great suggestion, and a combination of life expectancy and growth rate obviously has an important role on body size. Unfortunately, as highlighted by the reviewer, the preservation does not allow for discerning these fascinating aspects. Even if it did, it would not be feasible to go back and gather this information. We can, however, go back to the text and do more than just hint at the importance of understanding these aspects.

Data availability: I greatly appreciate that it is planned to make the script and data available through Github. From the available figures and discussion, you did thorough analyses. Unfortunately, I could not access script/data at the link provided (e.g., Github folder seems to be empty: see also comment by editor). I feel providing your data through a repository with a digital object identifier such as Zenodo would be crucial (it is quite easy to connect and publish Github dataset with Zenodo). Providing additional figures and used data tables also as supplementary materials would also benefit reproducibility. I feel it is crucial such data such be shared at latest upon publication but ideally and as best practice already during reviewing phase (possibly in another way if you have concerns code and data to be re-used before publication).

**RESPONSE:** All data and code was pushed to github and was checked before submission. The repository has been updated, and we will also upload the scripts and data to Zenodo upon acceptance of the manuscript.

Additional suggestions:

Line 50-52: I would argue that in some (particularly less extreme) crises - more detailed studies at bed-level are possible without having to go to coarse level and may be worth pointing this out (for example: belemnite - Rita et al. 2019; De Baets et al. 2021; Nätscher et al. 2021; or bivalves - Piazza et al. 2019, 2020; Ullmann et al. 2020 during the TOAE: Rita et al. 2019; Piazza et al. 2019, 2020; Nätscher et al. 2021; De Baets et al. 2021; Ullmann et al. 2020). Nätscher et al. 2025 (which you already cite) also found no consistent association changes in taxonomic level or stratigraphic grain in available data.

**RESPONSE:** We have added this element in.

Line 105: On which reference(s) is this reconstruction and inferred paleoenvironments in Figure 1 based (see also online comment by Renato Posenato)? Did these locations markedly change their environments across the studied interval?

**RESPONSE:** We have added that this is the interpretation for the time of the Permian/Triassic boundary.

Line 130-132: You allude to biases related taxonomic practices. Could you explain in greater detail how you tried to resolve these and how they may influence your results and interpretations.

**RESPONSE:** We believe that we already do this by giving *Claraia* as an example.

Line 175: Could you explain in greater detail how *Schizodus* and *Neoschizodus* are closely related? Are they part of an anagenetic lineage or is it more complicated and difficult to resolve at the moment?

RESPONSE: We have added this element in.

Line 240: can it be falsified fully? Could taxonomic practices also play a role in this?

RESPONSE: We softened the wording. Regarding taxonomic practises, it is really unlikely giving who identified the fossils (mostly by us).

Line 271-273: “Pre-extinction body size reductions have often been considered a precursor signal of a mass extinction event, and one expectation is that species would have reduced their size prior to their extinction. This has been observed for ammonoids and podocopid ostracods from Iran (Kiessling et al., 2018; Nätscher et al., 2024) and brachiopods from South China (He et al., 2010; Zhang et al., 2016).” I suggest rephrasing as within species changes were not observed in Permian ammonoids (e.g. it rather takes place at genus or higher level: compare data in Kiessling et al. 2018). I also suggest to add these cited changes take place during the Permian to avoid confusion and potentially also cited work of similar observations during other crises (e.g., Piazza et al. 2019)

RESPONSE: We have amended the text.

Line 279: you allude to difficulties related with stratigraphic gaps in studying pre-extinction size reductions. Would such gaps or stratigraphic changes no also not cause wider difficulties in interpreting the lilliput effect or even partially cause it around mass extinction events?

RESPONSE: We clearly lack data from the Bulla Mb., but anecdotally we know that these assemblages (when looking at the brachiopods) are not showing a Lilliput effect. We do not want to add this element in here, but leave the suggestion open for future studies with data.

Line 280: you mention “our data reject the “Lilliput effect” in the sense of a temporal within-species size reduction during an extinction/hyperthermal event”. I feel it would be more appropriate to state it more specific and less firm (given the constraints you discussed earlier) as: “our data seem to reject the “Lilliput effect” in the sense of a temporal within species-reduction during the Permian-Triassic extinction/hyperthermal event in bivalves”

RESPONSE: We added a caveat.

Line 284: Can you sure it is at least not partially an artefact of taxonomic level?

RESPONSE: Body size is not used as a taxonomic trait, if it was, then we could really understand your concern!

Line 286: I would argue taxonomic practice, degree of ontogenetic allometry and representation of juvenile versus adults (and changes therein) may potentially also play into it. I would like to see this discussed in greater detail.

RESPONSE: We agree, but given that the juvenile stages were likely microscopic, and we are dealing with bivalve macrofossils, it seems unlikely that we are not comparing adults with adults. Such a suggestion would require a large shift in population structure, which would not be such a parsimonious explanation. We have this element in the text to suggest how population changes could be driving our signals.

Line 287: I feel influence of changes in depositional environment, facies and/or sea-level changes also need to be discussed under abiotic changes.

RESPONSE: We have added this element in.

Line 338-340: Can you be sure these small forms are adults or rather may reflect an increase of the proportion of juveniles versus adults? It would be crucial to discuss this aspect.

RESPONSE: We could never be certain, but given that the juvenile stages were likely microscopic, and we are dealing with bivalve macrofossils, it seems unlikely that we are not comparing adults with adults. Such a suggestion would require a large shift in population structure, which wouldn't be such a parsimonious explanation. To address this question, you would need exceptional preservation. A caveat is to mention that we know we have fossils of different ages, which you can see especially from the publication on the taxonomy by Prinoth and Posenato (2023), but whether these younger ages actually represent the juvenile stage is another question.

Line 344-346: This discussion is highly interested and potential contribution to such issues on size patterns more generally could be discussed in this context.

RESPONSE: I'm sorry, but we do not fully understand the comment.

Line 350: Could lack of ribs or ornamentation also related to differences in preservation?

RESPONSE: Certainly, but even in poorly preserved specimens, we would expect to see some indication of external ornamentation.

Line 388: I would stress Lilliput effect sensu stricto (or how it was traditional seen) here for clarity.

RESPONSE: The sentence as a whole is clear in our view.