Our replies to the reviewers is indicated blue

Round #1

---

by Christina Belanger, 03 Sep 2023 17:22
Manuscript: **https://osf.io/fu7rp/** version 1

Invitation to revise preprint submitted to PCI Paleo

Greetings -
Your pre-print "Calibrations without raw data – A response to "Seasonal calibration of the end-Cretaceous Chicxulub impact event" has received two reviewers. Both reviewers were generally supportive of the pre-print, but also raised some important points that should be addressed before recommendation, thus my decision is to ask for revision.

1. The reviewers see many of the issues raised with the DePalma et al. paper as reasonable criticisms, however feel that alternative interpretations to intentional fabrication of the data should be acknowledged in the pre-print. Both Reviewers also express that the degree of offence (mistake, sloppy handling, manipulation, fabrication) cannot be definitively established with the present evidence. I agree with this point, and this uncertainty should be clearly expressed in the pre-print.

We agree with the suggested revision and have expressed the uncertainty in the preprint

1a. Reviewer 1 (who provides their comments as a .pdf) provides some alternative interpretations as well as benign explanations for why some observations made by the present manuscript. Reviewer 2 (who provides comments as plain text) points out an alternative explanation of the d13C values of the sturgeon. These explanations should be incorporated into the pre-print.

In general, we agree, but in detail we have responded to every individual comment.

1b. Some criticisms raised by the pre-print point out what may be called "best practices" – for example, similar analytical details are missing in many papers (see R1) but are expected by others (see R2). I think an impact of this pre-print could be to recommend community standards for reporting methods and data that will help future authors and reviewers.

2. Given these valid alternative explanations, I agree with Reviewer 1 that "softening" the tone of the pre-print is needed. Reviewer 1 provides examples such as changing definitive terms, like "demonstrates" to terms that allow for alterative interpretations, such as "suggests." R2 uses the phrases "gives the impression" and "suggests" in their review as well.

3. Both reviewers express that responsibility for the quality of the DePalma et al. article also lies with Scientific Reports and that many of the issues raised by the present manuscript should have probably been raised in the original review process. I agree with this point and shared responsibility should be more clearly expressed in pre-print.

We agree that the responsibility also lies with Scientific Reports, and have now emphasized in the paper that we have also asked these questions to the editors, sadly these remain unanswered.

In addition to the three themes summarized above, the reviewers also provide some detailed comments and suggestions that will improve this pre-print.
I invite you to respond to the reviewers' comments and revise the pre-print for the further consideration in revised form.
Thank you for your submission,
Christina Belanger

Reviews

Reviewed by Thomas Cullen, 17 Aug 2023 19:56

In this manuscript by During et al., the authors document various irregularities, omissions, and outright unusual aspects of the 'data' presented in a prior study by DePalma et al. which covered a very similar subject (and the same site). I think they present the evidence supporting their arguments very well, and nicely outline the issues in the DePalma et al. study. It is unfortunate that their submission was apparently not considered sufficient by Scientific Reports to be featured as an official reply/rebuttal to DePalma et al.'s study (or as sufficient grounds to have DePalma's study investigated and potentially retracted based on these issues) [or perhaps that is still ongoing?], but at least in the current form the manuscript will be available for others wishing to examine the issues in detail. I think this manuscript has merit, and have only minor comments/changes suggested.

Overall Review / Comments:
I agree with the authors in their criticisms and concerns with the primary data itself, as well as the lack of analytical outputs. I also agree with their concerns regarding the notably incomplete methods section, which lacks information on sample weights, pretreatment and dissolution procedures, or analytical standards (all of which would be considered a basic reporting requirement in a stable isotope geochemistry study). The amount of samples they obtained via microdrilling is unusually high given the small size of the specimens (as pointed out here by During et al.), and I agree that it too requires some corroboration, as such density of sampling would be very difficult (to somewhat implausible) if done using 'typical' microsampling approaches for d13C and d18O analysis of bioapatite structural carbonate on specimens of the size indicated.
The combination of an absence of primary data and the substandard methodological reporting alone should have been sufficient to call the results of DePalma et al.'s study into question and functionally invalidate their paper as a result of incompleteness, regardless of whatever post-hoc excuses may or may not have been since offered by DePalma and colleagues (e.g. collaborator who ran the

analyses is now deceased, no one kept analysis records, a dog ate their homework, etc). Indeed, one wonders how these issues escaped the notice of the original reviewers of the DePalma et al. study in Scientific Reports, as these problems should have been more than sufficient to reject their paper (or at least delay it until the missing information was provided to support their results/claims), going off of the editorial and data availability standards purported to be required for publication in that journal.
That above is before one even begins to evaluate the arguably more serious allegations made here that DePalma et al. manipulated or outright fabricated their data. Considering those more serious allegations, It is hard to disagree with the conclusions presented here by During et al., as there are far too many irregularities to dismiss as being the product of chance or some sort of image artefact from uploading/publishing. As the authors note, there are myriad examples of datapoints for d13C and d18O which are mismatched/misaligned in the plots of sampling location, of outright missing datapoints (despite line dips/angles suggesting the presence of a data point), of datapoints and error bars which are only partially present on the figure or off-centre, and of places where the same specimen analysis data are presented in the main figures and the supplement but somehow have different numbers of datapoints and isotopic composition patterns.
At best, it seems fair to say that the images presented as results in DePalma et al. are manipulated in some way which has led to an inaccurate/inconsistent reporting of the original data (which of course cannot be confirmed due to the primary data not being provided by DePalma et al.), and which does suggest that this manipulation may go beyond image/figure issues and represent direct manipulation of the data itself. To that end, I largely agree that it gives the impression that the plots were 'hand-made' rather than being representative of some original data output being plotted up by a program. I think the authors' points concerning the issues with the data overlays between individuals (and their remarkably implausible consistency), as well as the similarly unlikely 'coincidence' that curves from allegedly distinct individuals happen to align perfectly when stretched or compressed, are very well-explained and raise serious questions of the validity of the data plots presented in DePalma et al.'s study. The distinction between manipulation vs. outright fabrication of data cannot be firmly established of course since the primary data are not provided by DePalma et al. (which as noted above, should itself be enough of a reason for the study results to be considered questionable/invalid/irreproducible), but it is any case fairly damning.

Minor / more specific comments:
- a few places in the manuscript the authors write "carbon and oxygen isotopes" without including "stable", which I think should be corrected given the context of these discussions does not concern radiogenic isotopes such as 14C.

We are thankful for this suggestion and have incorporated this.

- concerning the d13C compositions of the sturgeon not showing a strong shift representative of feeding in marine vs freshwater settings (and particularly when the d18O does show a stronger cyclicity), I would mention that it is not impossible for a marine vertebrate to have d13C compositions in the –4 to –1 per mil range. For example, mosasaurs have been reported in the –12 to +2 per mil range, depending on location/time. That doesn't necessarily contradict the overarching concerns, just

noting that I don't think that lack of cyclicity is necessarily something that supports charges of data manipulation per se.

Thank you for this suggestion, we have added that this is not impossible.

- on lines 94-95 you note that one of the re-used samples (FAU.DGS.ND.755.57.T) has 43 sampling spots in Fig 2 but has 29 sampling spots in the Supp Materials. When I look at the graph for that specimen in SUP MAT 9 of DePalma et al.'s paper I see 35 sampling spots for d13C and 35-36 for d18O (the line angle suggests a point but one is not marked). I do agree with you, however, that the plots provided for the same specimen in Fig 2 and SUP MAT 9 do not match and show different numbers of samples, which is very irregular and requires some sort of explanation (particularly when it is identical for portions of the interval, but widely different in other select locations).

Thank you for noticing this error, there are indeed 35 (or 36 if we count the dip without a datapoint) points. We have corrected this.

Thank you for the opportunity to review this manuscript, which is a response to a recent paper published in Scientific Reports (DePalma et al 2022). The manuscript calls attention to a number of issues regarding the Scientific Reports publication, especially regarding the lack of data availability, methods details, and possible inconsistencies in the presentation of results. The manuscripts argues that these issues represent evidence that the results may be fabricated, a very serious allegation indeed. I approach this as someone with 10+ years of experience conducting stable isotope analysis on biological and paleontological samples, including the collection, handling, and (micro)sampling of samples, running (and customizing/repairing) isotope ratio mass spectrometers and associated peripheral devices (including those mentioned in this manuscript), and handling datasets ranging in size from tens to thousands of stable isotope measurements. My overall impression of this manuscript (perhaps shared by the authors), is that all of these issues ought to have been raised during the peer review and editorial process at Scientific reports prior to the publication of DePalma et al 2022, and thus are reasonable to raise in some form. However, a number of these issues are minor (not naming the analytical facility, not providing sample weights, not naming specific standards used) and do not either individually, or in combination, provide evidence one way or the other regarding the possibility of data fabrication. Some issues raised in this manuscript regarding the graphs in DePalma 2022 are potentially more serious, and are indeed worth raising, but I don't see a smoking gun. As such, I would ask the authors of this manuscript consider revising their manuscript such that it clearly acknowledges alternative interpretations of the issues raised, such as unintentional mistakes, database (copy/paste) errors, or graphing software misuse cannot be discounted.

We appreciate the suggestion and have put less emphasis on what the intensions could be that caused the errors.

Specific comments: Lines 67-69: I agree the lack of data availability is unfortunate, and that the authors of the Scientific Reports publication should have included results with their paper. Some fault here also lies on the editor and reviewers of that paper, and as such this issue does not itself constitute evidence of fault solely on the part of the authors of the Scientific Reports publication.

Naturally we agree with the reviewer that this should have been tackled by Scientific Reports, we added how we have also reached out to the editors, but sadly nothing came out of this. We have however tried to alter the tone of the article to emphasize that this is also an editorial error.

Lines 71-74: I agree that it is good practice to include this information, but many papers do not and this does not constitute a major anomaly as long as there is some clear indication where the analyses were conducted and by whom, which the original Scientific Reports publication does clearly provide.

We disagree with the reviewer that the original Scientific Reports publication clearly provides where the analyses were conducted and by whom. De coauthor listed as responsible for the analyses did not have access to the described facilities in the laboratory he is affiliated with.

Lines 77-78: I agree that it is good practice to include such information, but this does not constitute a major anomaly but rather a minor omission that is often caught in the course of the peer review/editorial process. The authors of the PCI preprint might specify what other information they would wish to know regarding the techniques. For instance, one might wish to see a statement explaining that phosphoric acid was used to analyze carbonate component of fossil samples, and the reaction temperature.

We wholeheartedly agree that this should have been caught in the course of peer review or the editorial process. Our first step therefore was to alert the journal, but they have not made any efforts to investigate the matter.
Since the resulting graphs cannot be explained by the described analyses, we are raising a question about everything "standard procedure" that is lacking here. Individually these could be considered minor omissions, but altogether the methodology as described does not explain the presented graphs.

Lines 83-89: This is a reasonable question to raise, and I agree here that additional information should have been provided by the authors of the Scientific Reports publication regarding their sampling strategy, especially regarding the typical area over which powder was collected for each analysis.

Thank you.

Lines 110-113: How do the authors define failure of either measurement? Do they mean the software does or does not provide a value? Regardless, it is not correct to say that situations where either carbon or oxygen analyses fail (however defined), the other cannot still be used. Rather, it depends on why how failure is defined. For instance, if high inter-peak variation is observed in d18O for an individual sample, the d13C value could still be used if the its inter-peak variation is 'normal'.

We disagree with the reviewer. It is common practice to dismiss both results when the uncertainty of one of the isotopic determinations is anomalously large as both C and O isotopes are measured together as CO_2. When isotopic values of one the elements are omitted then at least a threshold value should be given by DePalma et al. (2021), and a possible explanation for these failures. The (internal) standard deviations of the individual measurements as well as the isotopic mean and standard deviations of the isotopic calcite standards have been not provided by DePalma et al. (2021). Only a typical analytical uncertainty has been reported, which is about 2 times larger than can be achieved with the quoted setup, which requires further explanation. Given the conversion of only 2/3 of the oxygen from carbonate (CO_3^{2-}) into CO_2 that is measured by the IRMS, the oxygen isotope values are most sensitive to the conditions during acid

digestion, in particular to temperature. It is therefore striking that both carbon as well as oxygen isotopic values are omitted as data points in some the samples. Furthermore, the minute samples sizes that are required to obtain the presented isotopic records of DePalma et al. (2021) would be too small to allow for multiple measurements as is typically done for carbonates, which contain 10 times more carbonate than apatite samples.
The reviewer is correct that the inter-peak variation – of typically 9 peaks is used to 1) determine the internal standard deviation per measurement, and 2) determine the mean value, which would be the point marked in the graph. However, this is done by measuring the CO_2 molecule; thus C and O simultaneously.
Firstly, the minute amount of bone samples that is available for these measurements would never have been able to provide sufficient CO_2 for multiple measurements with the method as described. For carbonates, at least 9 sub-sample will be analysed for each sample. Secondly, the internal standard deviation was never provided, so the inter-peak variation is unknown and the error bars in the graphs are not based on data (they have the same length independent from scale of y-axes). Thirdly, if indeed such inter-peak variation would hypothetically have occurred in the oxygen measurement for an individual sample, this may justify using the carbon value if it had normal inter-peak variation. If this inter-peak variation occurs in the carbon measurement, it is almost impossible to maintain use of the oxygen measurement as oxygen is far more temperature-sensitive (with regards to fractionation) than carbon. However, when looking at DePalma et al., 2021, we can tell that there are at least 5 instances where the carbon value is absent, but the oxygen value is present.

We believe that it is safe to say that it is protocol to dismiss the other value if one of them has high inter-peak variation, and that it is highly unlikely that a measurement of oxygen can be successfully executed (over 9 peaks or even some less) when carbon has high inter-peak variability.
Lastly, the cause of the failures has not been explained, it is impossible to assume that the cause would be inter-peak variation. It could also be low amplitude measurements, which tend to be valid for both isotopic variations and would dismiss both. Or it could be due inadequate flushing of the exetainer vails with Helium resulting in the contamination with atmospheric gases which would affect both isotopes. Since the failure of either or both measurements was not described, this is something that we needed to raise.

Lines 113-117: I wonder if this could also be explained by repeated micro sampling of the same areas, measured multiple times, or potentially by errors in spreadsheet management and/or data use in graphing software.

Again, this would have to be explained in the methodology and there was no mention of this. Furthermore, there would not be enough material to sample twice and measure. As explained in the article and above, there would not be enough bone material and within this bone material there would not be enough CO_3^{2-} available for 2 (multi-peak) measurement of the same area – we already question whether they could have extracted enough for the presented sampling resolution.

Lines 117-118: Could this also be the result of ‘sloppy’ use of graphing software?

This all depends on what you define as sloppy, but we disagree with the reviewer that this is not a point worth raising. When you publish a scientific paper, your graphs should not have sloppy errors, even most undergraduate and graduate theses have higher standards this would have never met.

Lines 119-127: I agree the difference in error bar length is an issue worth raising, but as with the other issues raised here more innocent explanations such as simple sloppy graphing software use cannot be discounted. The parenthetical statement is not relevant here and should be removed.

We do not intend to imply that the explanations cannot be innocent, we are criticizing the work, and believe these are valid points to raise.

Lines 128-133: The conclusions here are one possibility, but their case is very far from conclusive. I do not mean to suggest that such errors are unimportant, but sloppy handling of data and graphic software (perhaps by a student) could very easily result in such issues, which indeed should be corrected but are nonetheless not equivalent to intentional forgery. Thus, the authors should soften their language a bit, especially by changing "demonstrate" to "suggests the possibility" and also by acknowledging other possible explanations.

We disagree with the reviewer. We have indeed demonstrated that these graphs have been drawn by hand. Furthermore, we have now been informed that these graphs were indeed drawn by hand – and the error bars were pasted as data points and do not reflect known errors or standard deviations at all. This information was provided by the first author to the Ethics Panel of the University of Manchester and to the editors of *Scientific Reports* and therefore these graphs are not the result of sloppy use of graphing software.

Lines 137-138: This sentence is too vague, please provide more information.

We thank the reviewer for pointing this out, we have elaborated to improve the clarity of the sentence.

Lines 141-154: These are interesting points raised here, which are perhaps the most (really, only) compelling evidence to even raise the possibility of data fabrication. Figures: Could the authors please define "misaligned" data points?

Thank you, we have added a definition of misaligned data points; which are data points that do not align with the same value of the X-axis, suggesting that these were not measured from the same CO_2 molecule as described. We added this explanation to lines 121-122.