Dear reviewers and editor

We thank the reviewers and editor for their helpful recommendations and have implemented them as the following:

We have implemented the comments that were made regarding biogenic calibrations by Reviewer 1. Regarding the comments given by Reviewer 1 for section 4.5, we have given an extensive point-by-point response and implemented the relevant suggestions and clarified our section 4.5.

As suggested by Reviewer 2, a brief paragraph on coccolithophore biomineralisation was added to the introduction and has been further implemented into relevant sections. Other sections regarding coccolithophore vital effects were adjusted following the reviewer's comments.

Our point-to-point responses are listed in blue as the following.

Reviewer 1

The authors have considerably revised the original manuscript, essentially overturing the study's conclusions and title. I would emphasize that this is not a problem in itself and actually shows a commendable willingness to reconsider one's initial interpretations. That being said, the manuscript still suffers from many problems, many of which are listed below in the line-by-line comments.

I see two major issues remaining at this point.

For one thing, the term "temperature calibration" would imply, to most readers at least, the absence of disequilibrium/vital effects on D47, or at least a quantitative scheme to correct these effects. This is not the case here. It is wrong to suggest that a "well-constrained" modern T-D47 relationship is a "calibration" (ie, has good predictive power) without explicitly demonstrating that the relationships applies equally well to different seawater chemistries. The authors do not explicitly make this claim, but point out repeatedly that their apparent D47 offset from equilibrium does not appear to vary (within precision limits, which is hard to quantify for the reader) with water chemistry, particularly pH or [CO2aq]. The range of pH investigated, however, is 8.0-8.7, which is clearly different from a Cenozoic range of 7.4-8.2 (Rae et al., 2021, Annual Review of Earth and Planetary Sciences). In my earlier review I called for explicitly listing the range of culture pH in the main text. In the revised manuscript, this information is provided but only in supplementary materials. Ultimately, in my opinion the authors do not convincingly argue that their culture observations are a good predictor of past coccolith D47-T relationships.

In other words: the revised manuscript appears to make the implicit statement that although we don't understand why cultured coccoliths have a different D47-T relationship from that observed in foraminifera, bivalves, slow-growing calcites, travertines, etc., this cultured coccolith calibration must apply equally well to all coccoliths, past and present. That is, again, a bold statement that calls for strong supporting evidence, that is simply missing at this point.

Re: In the field of palaeoceanography, empirical relationships between parameters, be it from laboratory or natural studies, are routinely termed calibrations, even when not in thermodynamic equilibrium or when the concept of thermodynamic equilibrium is not applicable i.e. TEX_{86}, Mg/Ca, Uk'37, and $\delta^{18}O$ foraminifera relationships with temperature; relationships between micropalaeontological abundance data and productivity. Such calibrations are then applied extensively in the field of palaeoceanography in predicting the past from present observations, with continuous community efforts extending our knowledge and assessing each calibration rigorously. For example, our coccolith calibration was tested and shown to be consistent with coccolith samples from recent sediment traps (Clark et al., 2024; EGU General Assembly 2024). Our approach closely follows the community practice in palaeoceanography.

Constraints on every single seawater parameter is the goal of every type of calibration study in palaeoceanography, and the community is constantly working on refining every calibration. The seawater chemistry of our coccolith calibration was constrained in controlled laboratory settings. While we did not adjust each parameter to every single potential Cenozoic value, we had more precise control over every parameter than most natural, ocean-derived samples. We did not have to infer parameters from neighbouring stations that are tens to hundreds of kilometres away or from previous, ambiguously constrained, empirical relationships such as is done for pH or $\delta^{18}O$, which have been subsequently used without question in for example foraminifera-based Δ_{47}-temperature calibrations. From the foundation laid by our data, we welcome future studies in constraining the other parameters that we did not explicitly study, in particular for coccolithophores.

As suggested by the reviewer, we have added the DIC, pH, and $CO_{2(aq)}$ ranges of our study in L20-21, L278-279, and L481-482 for further clarification.

Second major issue: Despite the overturned conclusions, the manuscript still appears to promote the idea that all/most biogenic calcites follow D47-T relationships different from the equilibrium calibration (which is neither biogenic nor inorganic by nature, it's just thermodynamics). A very long section 4.5 appears to argue that point indirectly, essentially illustrating that calibration studies rest on many ambiguous interpretations, and that making somewhat arbitrary changes will yield a wide range of D47 "calibrations". In my opinion this is the weakest part of the manuscript, because it needlessly dives into arcane details and fails to make clear, compelling points. Ultimately, whatever the success of this argument, my previous review's point remains that if we are to conclude that most/all biogenic carbonates have out-of-equilibrium clumped isotopes, the logical consequence is that we need to understand the processes at play, because their isotopic effects are unlikely to vary only with temperature by default. Arguing for a universally applicable "biogenic" calibration is irrational and willingly ignores the well-established fact that biomineralization strategies are hugely variable across genera. In my opinion, the manuscript would greatly benefit if it stopped pushing this appealing but deeply misguided idea.

Re: We have adjusted our discussion regarding biogenic carbonates, their Δ_{47}-temperature relationships and equilibrium. To make section 4.5 more readable, rather than listing the studies compared in the text, we have listed each discussed study regarding biogenic carbonates separately in Table 4 for clarification. We have also adjusted Figures 7 and 8 in removing the general biogenic calibration lines to make it easier to compare our results with those of individual previous studies. Please see further details in the line by line responses below.

Line-by-line comments:

- Lines [25-26]: "agree with a previous culture study that there are no species- or genus-specific vital effects on the D47-temperature relationship in coccolithophores". As written, this is presented as a general truth.

Re: We have rewritten this sentence to provide more nuance:

L22-24: *"Our well-constrained results agree with a previous culture study that there are no apparent species- or genus-specific vital effects on the Δ_{47}-temperature relationship in coccolithophores despite significant deviations from equilibrium in the C and O isotopic composition."*

- [32-33]: "All published biogenic studies fall within within ±1°C of our coccolith-specific calibration if [some arbitrary interpretative choice is made]". See my comments below regarding section 4.5.

Re: We have revised our discussion regarding biogenic calibrations and removed this sentence.

- [65-63]: "However, there are clear discrepancies between most inorganic calibrations [...] and an often used, generalised biogenic calibration [Meinicke et al., 2020]": First, as written, this sentence suggests that there are discrepancies between the inarganic calibrations, which is arguably no longer the case. Ambiguity could be avoided by adding "on one hand" after the inoraginc references and "on the other" when citing Meinicke et al. Second, I'm not sure why Meinicke et al. is described as a "generalised biogenic calibration" since is is exclusively based on foraminifera.

Re: We have implemented the suggestion:

L52-54: *"However, there are clear discrepancies between on the one hand most inorganic calibrations (Swart et al. 2019; Jautzy et al. 2020; Anderson et al., 2021; Fiebig et al., 2021) and an often used biogenic calibration (Meinicke et al., 2020)."*

- [fig 3] equilibrium and kinetic limit equations are switched.

Re: We thank the reviewer for pointing this out and it has been changed.

- [349-351]: "This approximation is derived from their experimental setup, which is not necessarily in equilibrium, and therefore potential small growth rate and pH effects are still present". This statement is incorrect. This equilibrium limit is actually tied to the Devils Hole observation of Coplen (2007). Watkins et al. (2013, 2014) quite reasonably assumed that the T sensitivity of equilibrium oxygen-18 fractionation between calcite and water is inherited from that for DIC species, and assumption which was further strengthened by a second, later observation from Laghetto Basso (Daeron et al., 2019).

Re: We had rewritten this section, as suggested previously by the reviewer, to provide more nuance to our statements regarding equilibrium. We have again rewritten this sentence:

L323-324: *"This approximation is derived from the assumed equilibrium of Coplen (2007), with potential small growth rate and pH effects present for carbonates not precipitated in equilibrium."*

- [361]: The Daeron citation does not seem relevant here. Guo (2020, GCA "Kinetic clumped isotope fractionation in the DIC-H2O-CO2 system") might be more appropriate.

Re: We agree with the reviewer and have changed the citation.

- [430-432]: "While the omni-variant generalised least squares regression would be better suited, as this incorporates the full error covariance (Daëron and Vermeesch, 2024), our data is standardised through reference materials in a moving time window and thus cannot be analysed through this method." Respectfully, may I suggest that this statement is irrelevant. As I noted in my earlier review, the temperature uncertainties are entirely negligible here, so that using York regression is already overkill. Using an even more complicated method is not warranted.

Re: While this is a good and valid point by the reviewer, even though our study has uniquely well-constrained temperatures and thus negligible temperature uncertainties, using the York regression is justified for consistency with studies with non-negligible temperature uncertainties such as most other non-laboratory calibration studies. Further, we also want to highlight the potential usefulness of the omni-variant generalised least squares regression for future studies that can apply it on their datasets.

- [442-444] "All regression lines fall within 0.0012‰ error of each other, which shows that with the available data there is no species- or genus-specific vital effect on the Δ47-temperature relationship.". Again, I am compelled to point out that this statement is overreaching. For one thing, the fact that three regression lines (or rather, two regression lines and one isolated data point) lie within 1ppm of each

other is one thing, but its significance depends quite a lot on the formal precision of the regression D47 values. If each of these lines had a D47 "precision" of 1 ppm, the agreement would be very strong, whereas for corresponding D47 precisions of 10 ppm, the statement would be much weaker. It would also be more fair to rephrase the second part of the sentence to "potential species- or genus-specific vital effects on the Δ47-temperature relationship remain undetectable at the X ppm level." (with X being the actual precision of the differences instead of the 0.0012‰ spread of regression lines).

Re: We have provided more nuance to this statement:

L414-416: *"All regression lines fall within 0.0012‰ error of each other, which shows that with the available data and at the current analytical precision there is no discernible species- or genus-specific vital effect on the Δ_{47}-temperature relationship."*

- [455-456] "there is no carbon isotope vital effect affecting the coccolith Δ47-values"; [447-448] "an important impact of the oxygen isotope vital effect on coccolith Δ47-values."; [581-582] "The vital effects observed in the coccolith carbon and oxygen isotopes do not have an impact on the Δ47": This wording is quite confusing. 13C and 18O "vital effects" do not "affect" D47 values. Instead, isotopic disequilibria affecting different isotopologues reflect underlying physical and chemical causes.

Re: We have clarified this in multiple instances:

L309-310: *"For our culture experiments, in order to evaluate whether processes promoting variable stable isotope effects would systematically affect the Δ_{47}-temperature relationship…"*

L426-427: *"The non-significant correlation in $\Delta^{13}C_{c\text{-}DIC}$ and $\Delta\Delta_{47,off}$ for all setups, shows that the processes responsible for the carbon isotope vital effect do not significantly influence the coccolith Δ_{47}-temperature relationship (Fig. S6)."*

L444-445: *"The similar culturing study of three coccolithophore species by Katz et al. (2017) also found that species-specific vital effects do not correlate with variations in the Δ_{47}-temperature relationship and also found a consistent Δ_{47}-temperature correlation."*

L483-484: *"The processes responsible for vital effects observed in the coccolith carbon and oxygen isotopes do not lead to corresponding variations in the Δ_{47}-temperature relationship…"*

- [459-462] "Thirdly, average Δ47 values were calculated for each species at every growth temperature. These temperature-weighted averages can highlight bias from a low number of measurement replicates at certain growth temperatures, such as at 6°C and 27°C. The resulting Δ47-temperature regression is indistinguishable from regressions using individual Δ47 sample datapoints ((±6.1 ppm; Table 2)." First, I don't understand. Are the regression not accounting for D47 uncertainties, which should scale with the inverse sqrt of the number of replicates? That would seem like the default method of performing such calibration regressions. Additionally, do I understand correctly that the quoted +/-6.1 ppm number is the difference between two different ways of performing the regressions? If so, that 6ppm difference seems quite large (equivalent to +/- 2 degrees) compared to the above statement that regressions agree to within 1.2 ppm.

Re: The regressions do account for Δ_{47} uncertainties since they are performed using the York regression. We illustrate that the low number of datapoints at one end can bias the regression even when incorporating the uncertainty. As Figure 4 shows, the average and uncertainty associated with it, which incorporates the inverse square of the number of replicates, do not fully capture all datapoints and at the high temperature end are almost the same average measured Δ_{47} values. The quoted offset is the maximum offset between the resulting regression through the average Δ_{47} value and relevant temperatures, and any one of the previous regressions for the different species.

- [492-493] "The biogenic data sets are combined into a general 'biogenic' calibration, excluding this study". This once again sweeps under the rug the fact that the two foram datasets were repeatedly shown to be consistent with one another, but published using different calcification temperature estimates, both of which are very likely flawed according to the reassessment of Daeron/Gray. What's more, Meinicke et al. (2020) also include many benthic foraminifera from Piasecki et al. (2019), which now appear to be either analytically compromised or far from equilibrium D47 values (cf Dearon & Gray, 2023).

Re: Our discussion regarding "general" biogenic calibrations has been changed and largely removed. As explicitly stated in our manuscript, we consider our coccolith-specific Δ_{47}-temperature calibration to only be applicable to coccolith samples. The reviewer points out that two foraminifera datasets discussed were shown to be consistent with each other, however this is the case with or without using the recalculated temperatures of Daëron & Gray (2023). The Piasecki et al. (2019) datapoints are not included in this study, only the planktic foraminifera from Meinicke et al. (2020; 2021) are used.

We acknowledge that the reviewer favours the recalculation approach of Daeron and Gray (2023), but there is still significant discussion and debate around this recalculation within the community. The debate exists as the core top foraminifera calibrations require an assumption about the actual calcification temperature of the samples. Our section 4.5 opens the discussion on the comparison of Δ_{47}-temperature calibrations from well-constrained experimental temperatures, such us our coccolith calibration, with the various unconstrained proxy derived, indirect calibrations derived from other biogenic carbonates. In Daëron & Gray (2023), the recalculations to the Δ_{47}-temperature calibrations are evaluated through comparison of Cenozoic temperatures derived from on the one hand, two empirical calibrations with Mg/Ca and $\delta^{18}O$ on benthic foraminifera by Cramer et al. (2011; Journal of Geophysical Research) and on the other, Meckler et al. (2022; Science) using Δ_{47} of benthic foraminifera and the recalculated Δ_{47}-temperature calibration. We note in our manuscript, that the recalculations by Daëron & Gray (2023) may impart a cold bias, one that approximates deep waters to be unrealistically cold for modern oceans; ~ -3°C as already shown by their Figure 18.

- [497] "Daeron & Gray 2023 orig" is a misnomer. Both N. Meinicke and M. Peral have published I-CDES versions of their respective datasets, which should be properly cited.

Re: These published datasets were used and the citations are now provided. The reason we are using the "orig" is for our comparison to the recalculations done by Daeron & Gray (2023), which includes some of the used datasets but recalculated temperatures. We have clarified this in Table 4.

- [fig 7] What is Daeron & Vermeesch ("MIT")? Is this a typo for Daeron & Gray ("MIT")?

Re: No, in Daeron & Vermeesch (2024) the inorganic carbonates of Anderson measured at MIT were recalculated and that regression is used.

- [499-519] I am truly sorry to say so, but with all due respect section 4.5 is neither a light nor a fun read. Juggling between original studies, the same studies recalculated with different T assumptions, partial versions of the same data by different authors, and datasets truncated below arbitrary T thresholds, yields a seemingly infinite number of semi-arbitrary options to chose from, detracting from the point(s) the authors' are trying to make. Is there something to be understood here beyond the fact that the cultured coccolith D47 values are greater than expected from non-coccolith calcite calibrations? If not, this can be stated much more simply. The section overall goes into bizarre tangents, such as a detailed critique of the earlier foraminifer studies, which I believe derails the discussion, only to conclude that "the use of calcification temperatures from oxygen isotopes need further testing, ideally on laboratory cultured specimens", a point that I believe is far from controversial today.

Re: As we point out and discuss in this section, for core top foraminifera the choice of calcification temperature, as performed by either the original authors or recalculated, can cause significant differences and thus interpretations to any calibration study. This we feel is an important matter to discuss with the palaeoceanography and clumped isotope communities. The matter of whether recalculations may or may not be convincing belies the fact that our study has complete control and constraint over the coccolith growth temperatures, while those from most other biogenic carbonate studies do not and infer them from other, ambiguous, and unconstrained parameters and proxies. Especially since there are studies, for example, the OGLS23 calibration of Daëron and Vermeesch (2024) that take two completely different types of biogenic carbonates, planktic foraminifera and marine bivalves that have different calcification mechanisms, and inorganic carbonates associated with equilibrium values together into the same calibration regression. We thus use this section to show that our coccolith-specific Δ_{47}-temperature calibration with well-constrained temperatures does indeed have a consistent offset from the inorganic carbonate Δ_{47}-temperature calibration. Yet, we also highlight for the general clumped isotope and palaeoceanography communities that there can be variability in other non-constrained, natural biogenic carbonate samples with ambiguous temperature constraints.

- [613-616] "Future studies with constrained and in situ temperature measurements such as in sediment traps (Clark et al., 2024) or cultures are recommended to disentangle and validate our findings of this biogenic disequilibrium." It would be misleading to suggest, as appears here, that "our findings of this biogenic disequilibrium" include non-coccolith biogenic carbonates. This is simply absent from this studies findings. Perhaps this is a remnant of the earlier version of the manuscript?

Re: We have adjusted this line to the following:

L548-549: *"...or cultures are recommended to disentangle and validate our findings of this coccolith disequilibrium."*

- [630-631] The revised conclusion states that "Our coccolith Δ47 data is largely consistent with a previous coccolith culture study (Katz et al., 2017), and indicates that coccolithophores precipitate coccolith calcite in clumped isotope disequilibrium". The original manuscript stated the opposite (clumped isotope equilibrium), but also claimed agreement with Katz et al. This highlights that the way(s) in which this study's results agree with those of Katz et al. are meant quite loosely (since direct I-CDES comparison remains impossible). I suggest that bringing up this highly flexible agreement does not truly strengthen the conclusion.

Re: We agree with the reviewer and have removed the ambiguity of this sentence:

L571-574: "*Our coccolith Δ_{47} data indicates that coccolithophores precipitate coccolith calcite in clumped isotope disequilibrium with their environment."*

- [634-635] "The discrepancies derived from the differences in calcification temperature render it difficult to conclusively state whether a general biogenic calibration should be used": bringing this up in this way in the conclusion would require this point to have been explicitly and convincingly argued in the discussion, which is far from the case. Perhaps another remnant of the earlier version of the conclusions?

Re: We have altered our discussion on biogenic calibration and have removed this sentence.

Reviewer 2

Clark et al. extensively revised their manuscript and answered the reviewers' questions. They make a convincing case that there are no species-specific effects across the analyzed coccolithophores. However, there are still two minor points that the authors might want to consider.

1) The paper would still benefit from a brief, one-paragraph description of coccolithophore biomineralisation. The current sentence, beginning on line 74, merely references other papers without explaining the process. Providing this explanation is necessary to understand where vital effects could occur in coccolithophores.

Re: We concur with the reviewer and have written a brief paragraph regarding coccolithophore biomineralisation.

L73-88: "*In part due to these CAPs, coccolithophores have a fine control on the formation of coccolith calcite. Calcite crystals are nucleated in a circular protococcolith ring upon an organic baseplate within the coccolith vesicle, which subsequently matures into a coccolith (Brownlee et al., 2015; Walker et al., 2019). The coccolith is then extruded towards the exterior of the cell, where it is adhered to the cell and forms an interlocking system of coccoliths known as a coccosphere (Brownlee et al., 2015; Taylor et al., 2017; Walker et al., 2018). CAPs and other organic compounds are found in abundance in all calcification steps. Intracrystalline CAPs from different species of coccolithophores can be crystal-inhibiting (such as for E. huxleyi; Henriksen et al., 2004; Gal et al., 2016; Walker et al., 2019) or promote calcite specifically even in unfavourable conditions (such as for G. oceanica; Walker et al., 2019). Extracrystalline CAPs can aid in adherence of the coccolith to the cell, of the coccoliths to each other, and maintain the coccosphere structure (Walker et al., 2018). Subsequently, there are few anion substitutions and a lack of lattice defects on the coccolith surface that further aid in a better preservation relative to foraminifera (Berman et al., 1993; Stoll et al., 2001; Frøhlich et al., 2015; Walker et al., 2019). Additionally, there are a multitude of specialised pathways that regulate the fluxes of ions such as* Ca^{2+} *and dissolved inorganic carbon (DIC) species into various intracellular compartments to allow for controlled calcification and photosynthesis (Brownlee et al., 2015; Gal et al., 2017; Taylor et al., 2017).*

Biogenic carbonates often feature carbon and oxygen isotopic compositions that differ from those expected for abiogenic carbonates near equilibrium, offsets informally called "vital effects", as a result of the complexity of coccolith calcification described above."

2) The authors conclude that "[…] thus while we can't fully rule out that vital effects are present on Δ47, the offset from the inorganic equilibrium calibration must be similar at all temperatures, systematic and unrelated to vital effects." The authors do not make a convincing case for this. Specifically, they base their conclusion on considering only pH and growth-rate dependent kinetic fractionation effects as "vital effects" and observing no correlation between Δ47 and some calcification parameters. What about metabolic effects, diffusion, or crystal surface effects? The possibility of any of these factors provides a more plausible explanation for the observed offset between the coccolithophores and the other Δ47-T calibrations than suggesting that every other calibration is biased in some way.

Re: The reviewer raises an excellent point about other potential vital effects that may be present and have not been addressed in this study. There is certainly a possibility that metabolic effects can affect Δ_{47}, however these would be consistent amongst species. Diffusion and crystal surface effects can also affect the Δ_{47} yet would be difficult to quantify and address directly in culture studies such as ours.

L542-545: "*Other effects may also be at play, such as species-specific metabolic pathways unique to coccolithophores, diffusion of cations or DIC species into the coccolith vesicle, surface speciation, or crystal surface interactions with cations from solutions (Hermoso, 2014; Sand et al., 2014; Gal et al.,*

2017; Taylor et al., 2017). However, while the offset from the equilibrium inorganic calcite is systematic across the three coccolithophore species cultured here, no definitive cause of this observed offset can be determined, and there is more work needed to identify the disequilibrium processes for other coccolithophore species."

Minor points:

Suggest removing "well-constrained" from the title. Although it may be true, this adjective unnecessarily lengthens the title without adding useful information.

Re: We agree that it is important to specify that the temperatures are the aspect which is "well-constrained" as this is the key difference from the sediment core top temperatures which need to infer a habitat temperature. We therefore adjust the title to:

L1-2: *"A clumped isotope calibration of coccoliths at well-constrained culture temperatures for marine temperature reconstructions"*

L15: "We thus…" This sentence reads strange as it contains two clauses (thus, because) referring to why the authors chose coccolithophores. Consider rephrasing.

Re: We agree with the reviewer and have reworded it:

L15-16: "*We thus determined the Δ_{47}-temperature relationship for coccoliths due to their relative ease of growth in the laboratory.*"

L31: It is confusing to highlight foraminifera here, as forms were not studied in the paper. Consider moving this statement to the outlook paragraph.

Re: We highlighted foraminifera here, as most of the previous biogenic carbonate calibrations have been performed on fossil foraminifera, yet lack any laboratory-based empirical calibration. We have removed the mention of foraminifera for clarification.

L28-30: *"Thus, we suggest the use of our coccolith-specific calibration for further coccolith palaeoceanographic studies and that calibrations derived from laboratory-grown biogenic carbonates are desirable to reinforce the confidence of clumped isotope-based temperature reconstructions in palaeoceanography."*

L34: Suggest removing "relatively new". Ghosh et al. 2006 was a long time and over 200 clumped papers ago.

Re: We agree with the reviewer.

Chapter 2.2: did the different cleaning procedures affect the δ18O or Δ47 values?

Re: For the data used in this study, there were no significant differences in $\delta^{13}C$, $\delta^{18}O$, nor Δ_{47} due to the cleaning procedures themselves. Any difference was within the standard deviation or error of each measurement and non-systematic; ±0.15‰ for $\delta^{13}C$ and $\delta^{18}O$ and ±0.016 for Δ_{47} respectively. When no cleaning procedure was performed, there were large positive offsets in Δ_{47} and these have not been included in this study. We added a sentence to the manuscript as the following:

L188-189: "*Any difference in isotope measurements as a result of cleaning protocols was within the standard deviation or error of each measurement (±0.15‰ for $\delta^{13}C$ and $\delta^{18}O$; ±0.016 for Δ_{47}).*"

Fig. 8: Consider explaining the difference between subplots a) and b) in the caption.

Re: We agree with the reviewer.