Response to Reviewer#2 comments

In the following, the Reviewers' comments or questions on the manuscript are given in black italics, and our response is highlighted in blue and indented. We only consider points that provide information and clarifications that are of general interest to the readership. Minor points like " Line 258: add a comma (,) after samples" or "Break this sentence into two" will be performed during the manuscript revision, and hence are not to be discussed here.

The authors are going too far in their conclusions: in the abstract lines 25-27 "Our calibrations and the calculated partition coefficients… enable the direct quantification of metals in polluted and pristine environments" and in conclusion lines 801-802 "The presented DTE´s allow a direct quantification of metal concentrations in polluted and pristine areas". First, given the really high DTE ranges found in this study (including or not phase 3) and/or DTE (from linear regression) strongly based on the phase 3 data point where the seawater element concentrations are variable, it is not possible to maintain that "quantification" of metal pollution in natural environments is possible. Secondly, the authors are contradicting themselves since they explain in the introduction that a mix of metal may result in interactions that can lead to different incorporation of the metal. Therefore, the mix proposed in this study, which is peculiar since including 10 trace metals at a time (in polluted environment, most often only 1 or 2 metals are above the threshold limit, not 10 at the same time), is not representative of other type of pollutions. The authors should be more measured in their conclusion. The elemental concentration in the shell may definitely be used to look at relative variation of heavy metal concentrations in the seawater through time and space but definitely not to give quantitative data… and only for elements where a positive correlation has been found between TE/Ca foram and seawater.

- We respectfully disagree. Any approximation of environmental signals by using the elemental or isotope composition of calcareous shells grown in the respective environment is only that precise as the variability of the calibration data set. The uncertainties are discussed in the paper indeed. Furthermore, the heavy metal mixture we applied was found very often in natural environments, for example in the vicinity of harbors or bigger cities. Some examples are given in Table 5 of our manuscript. This is only a brief overview and much more work on this is available. Therefore, our metal mix is definitely representative for heavy metal pollution in near-shore environments. Indeed, it was intended in this study too investigate the impact of metals that do co-occur and are potentially interacting.

I think that, instead of describing and discussing each trends or absence of trends observed, they should maybe realize that the absence of systematic tendency (within one element or one species) is unexpected and might be the result of multi-metal experiment since singe metal culture exhibit usually positive correlation between shell and seawater element ratios (cf literature). I would also advise to elaborate more the interspecific differences and maybe on the new elements that have never been measured before.

- We have specifically chosen to conduct experiments that mimic natural conditions as much as possible. This brings added complexity but by carefully monitoring the changes in culturing medium metal concentrations the results are robust. It is important scientifically to report non-results but we will focus more on the elements with linear partitioning and species differences in the revised manuscript.

To my opinion, there is confusion between the toxicity of metals to the organism and their incorporation into the shell (cf line 27 "This in turn allows monitoring of the ecosystem 27 status of areas"). What are expected from environmental/governmental studies is to evaluate the impact of heavy metal concentrations on the organism life (ability to survive, grow, reproduce…). Here the authors measure the elemental concentration in the shell. The speciation of the metal incorporated in the shell might be different from the one causing toxicity and bioaccumulation in the cell. The elemental concentration in the shell may help to reconstruct variations of seawater elemental concentration but for the moment, the link between this concentration and its effect on marine life is still unknown (and may depends on elements!). I think that the authors should discuss more precisely about this aspect.

- We take the message to more precisely distinguish between toxicity, bioavailability and incorporation into the shell at respective sections in the manuscript and will rephrase the sentence.

I have problem to understand different aspects regarding the metal mix solution:

1. *How and where was added this solution? Was it added in the supply tank located on top of the system? In this case, knowing that the pump is flowing at 0.017ml/s, how long would it take to replace and reach the same metal concentration in the culturing vessel (ie Tupperware) as in the tank?*

- The solution was added in the supply tank and according to the flow rate, the metal concentration in the culturing vessels would be the same as in the supply tank after a few hours. Not only the flow rate and water exchange in the vessels but also sorption processes can have an effect on the metal concentration, which is addressed in the manuscript in detail.

2. *When was added the solution? According to line 206, we understand that this is added once before each phase. But on line 229-230, it is written "For keeping the metal concentration at the same level over the different culturing phases, water with elevated heavy metal concentrations was fed into the system bi-weekly.". I'm lost, what is this "water with elevated heavy metal concentrations" you are referring to? Is it the stock solution? This need to be clarified…*

- The stock solution was added at the beginning of each phase to reach the targeted concentration. Additionally, a smaller aliquot of the same stock solution, termed "water with elevated metal concentration", was introduced twice a week during the three week of a phase, because a loss of metals during the culturing phase was expected, e.g., because of the uptake by foraminifera or algae or because of adsorption to surfaces (see Response to Reviewer 1 above).

3. *Where are taken the samples for trace metal analyses in the seawater? I don't think the information is given (Lines 215-216, line 234 "from both systems")… Are they taken at the outflow of the vessels so that it really corresponds to the concentration of the seawater in which the foraminifera are growing? Or they were sampled in one of the tank, which would be of course less precise…*

- The water samples were taken from the supply tank. This information is to be added in the Methods chapter. This point in the system was considered appropriate because the high flow rate and hence well-mixed system facilitated a representative sampling.

4. *How often was measured the metal concentrations in seawater over the course of the experiment. This should be indicated in the material and method part. For the moment, it is written "frequently" (line 216 and 234) however, when looking at Figure B1, only 1 to 4 data point are available within each phase. This is to my opinion problematic when applying individual curve fit for every phase to calculate the weighted mean value… (see comment after)*

- Fourteen samples were taken from the metal system over the course of the whole experiment, which can be seen in Table A1. From beginning of phase 1, sampling took place twice a week. Indeed, the metal concentration was expected to be more stable during the culturing phase, which was why sampling twice a week was considered as appropriate. Nevertheless, the application of a fit curve for every phase is in our opinion the only way to approach a representative mean value for a given phase especially when taking into account that the metal concentration in the culturing medium varied.

5. *I don't understand the calculation in table 1. The factor between each phase is 10 times but on line 207, it is written "phase 1 = 1 ml, phase 2 = 10 ml, phase 3 = 150 ml". How were calculated these target values?*

- In phase one, 1 ml of the stock solution was added and in phase 2, 10 ml of the same stock solution were added, which is the factor of 10 mentioned above. The target values were not calculated they were taken from the literature (see line 208 – 213). Based on this, we calculated how much of the stock solution was needed to be added to the system for each phase.

6. *The authors used "stock solution" all over the manuscript when referring to the metal mix solution. However, I think it would help to clarify this on line 205 when you first used this term so that it is clear in the discussion (on line 487 and after) that you talk about this stock metal mix solution.*

- This will be clarified in the revised manuscript.

The culturing system description is very precise but also very long and all the details makes it difficult to understand the general principle. I think that the authors would gain clarity if they explain earlier that a different vessel, freshly filled with calcein labelled forams (if this is correct), is incubated for each different phase of the experiment. At the moment, this essential information appears only (if I'm correct!) on lines 222-223 whereas it should already be said in chapter 2.2.2 or at least beginning of 2.2.4. The description would also be clearer by keeping the same term to describe the same "object".

- We agree and will move this information to an earlier point in the manuscript. Furthermore, the description is to be rewritten taking into account that the same terms are used for corresponding parts of the culturing system.

Abstract

Line 17: "Seawater analysis… between culturing phases". This sentence is not 100% correct since the increases between phase 0, 1 and 2 are not very obvious for all elements (e.g. Cu, Mn). This is however clear for phase 3.

- This needs to be specified.

Line 24-25: I have the idea that Zn and Cd are showing variations that are more or less similar to other elements (eg. Cd like Pb), no?

- Indeed, the small variation applies to Sn only but not to Cd and Zn.

Introduction

There is confusion between "heavy metals" and "trace metals" throughout the manuscript. To my knowledge, the 10 metals studied here are not all considered as "heavy" metals, some are trace metals. I think that this depends on the atomic weight of the element… Please check and use the appropriate terms.

- "Heavy metal" as term is not clearly defined. This issue is nicely described in Duffus et al (2002). The atomic weight is one possible criterion, but no threshold is set for a minimum weight an element must have to be considered as "heavy metal". Some authors pretend that the atomic weight needs to be greater than sodium, which would apply for all of our metals, and others take Hg or Ca as a boundary weight. Another criterion is the density. The boundary value for this parameter is ranging between 3.5 and 7.0 g/cm^3 depending on the author. Other criteria involve their behavior as Lewis acids… It is therefore difficult to apply an "appropriate" term, but we will define our use of the term in the outset of the revised manuscript.

Lines 69-75: here you talk about the physiological effects. This is interesting but you are looking at the incorporation in the shell which is different (cf comment earlier). The information is correct but it has to be clear that you will not have a look at this aspect yourself in this study.

- See Response to Reviewer 1.

Line 86: "bioavailability". I guess that this is correct to say that if the element is found in the shell, it is bioavailable since it might (depending on the biomineralisation process involved) goes through the cell. However, I would say that this is different from toxicity effect (cf comment earlier).

- Bioavailability and toxicity are definitely different and the sentence needs to be reformulated to appropriately discriminate these two terms.

Line 46: check in Kotthoff et al. (2017) that Mn/Ca is actually used for O2 or redox reconstructions and not for contamination.

- This is true and should be added at this place.

Line 53: These species are also dominant in intertidal mudflats, not only subtidal areas.

- This is a misunderstanding. "near-shore" does not mean subtidal as the "shore" is legally defined by the Mean High Water level. As such, mudflats are well near the shore. Perhaps it is more precise to say "intertidal and shelf environments".

Material and Methods

It would be nice to document with SEM pictures and light pictures the 3 species of this study. I think it is even more important knowing that Ammonia and Elphidium are species rising lot of identification discussions! Whatever the name given, it is essential to have to possibility to look at the picture and compare it to literature and also recent DNA papers.

- This is not necessary. The species from our sampling locations are already well documented in the literature (Lutze, 1965; Nikulina et al., 2008; Schweizer et al., 2011; Francescangeli et al., 2021; Schmidt and Schönfeld, 2021). We will add this information and citations to the revised version of the manuscript.

Lines 118-121: I am wondering if this information is relevant for the manuscript. Since the text is too long, I would suggest to delete this part. Also, the authors mention cores sampled for ecological study which are not presented in the manuscript. This is maybe not necessary?

- We will delete these lines.

Lines 138-150: There are too many details here (eg the size of the petri dish). Some information is repeated several times. For example, the fact that the authors checked several times to be sure that the forams were alive (lines 142-143: "glossy, transparent and undamaged test… cytoplasm present", line147 "structural infill of cytoplasm", line 151 "the color of the cytoplasm was checked"). I don't think the precision of this check at each step is necessary… The important information is that the forams used at the end in the experiment were labelled with calcein and exhibited a green cytoplasm proving that they were active.

- We will shorten this section where possible but avoid losing any important information in the revised manuscript.

Lines 151-156: I had some difficulties to understand (when I first read the manuscript) when this labelling step happened? Is it only once at the beginning of the entire experiment (before phase 0)? But in this case, the forams added for example at phase 3 could have calcified new chambers in the meantime… Or is it done before each phase in order to add freshly labelled forams in the new introduced vessels? Here the authors should precise this aspect.

- The labelling took place before each phase to ensure that freshly labelled foraminifera are inserted in the well plates. This will be clarified in the revised manuscript.

Although the culturing system is well described, it is difficult to not get lost since everything is described with lots of details. Therefore I would recommend to always use the same term when describing one part of it (eg "vessels" for the box containing the well plates, that you should name this way on line 186). On line 195-196, the term chamber is used but we do not really know to what it refers to: well-plate cavities? Vessel? Please try to keep it simple and clear.

- We will clarify this section and we will also consistently use the same terms for respective parts in the culturing system.

Lines 223-224: One vessel was left from phase 0 to phase 4 (84 days). What was the interest of this vessel? Were the forams from this vessel analysed? If this is not the case, you should say it to avoid any confusion!

- The interest was to have a look at the metal incorporation during all four phases in one individual specimen, but the foraminifera have not been analysed yet. We will delete

this sentence to avoid any confusions.

Lines 286-290 "the total number of chambers was counted before and after the experiment for every specimen (Table 2)": I don't see the interest of counting all the chambers of each foraminifera before and after the phase since the authors used calcein. And this information is not given in Table 2. Moreover, I agree that this is possible to count the total number of chamber in Ammonia species since they are trochospiral. However, this is not the case for Elphidium species since spires of new chambers recover the initial chambers! Therefore, if forams where indeed labelled with calcein just before their introduction into the culture system, I would keep it simple and only mention calcein to identify newly formed chambers.

- Chamber counting was to double check if foraminifera grew during the experiment, because calcein staining may eventually fail. This needs to be stated in the manuscript.

Line 312-313: Could you explain why you chose to use NIST612 for calibration and monitoring of instrument drift since the elemental concentrations in this standard are way above the concentrations found in the forams? Moreover, you chose to use a glass standard as quality control whereas it would be more appropriate, to my opinion, to use a carbonate standard with similar matrix to your forams. Moreover, the conditions are similar between carbonate standards and forams (I guess) whereas NIST standards are measured with higher energy and frequency. Please explain.

- The glass standard was chosen because all elements of interest but Hg are reported in the literature, which is not the case for carbonate standards. For further quality control, a variety of carbonate-based reference materials have been measured. All values can be found in Table A3 in the appendix. Furthermore, Dueñas-Bohórquez et al. (2009) demonstrated that different energy density between the foraminiferal calcite and the glass standard does not affect the Laser ablation analyses.

*Line 334: The authors considered the data as usable if above LOD. However, the limit above which the data can be used for quantitative purposes is commonly defined as the LOQ (limit of quantification). This is defined as 10*SD of the blank. How many data would be excluded from the dataset if the authors use LOQ instead of LOD?*

- We can check the LOQ.

Lines 364-365: It is not described in the Material and Methods how the living forams were differentiate from the dead ones at the end of each phase. Did the forams lost the colour of the cytoplasm (or their cytoplasm itself) so quickly that you could see it?

- Indeed, the foraminifera loose the color of their cytoplasm quickly. Furthermore, they do not gather particles or food any more, thus are lacking a detritus cyst before their aperture.

Line 102: what is Hallig Hooge? Is it still on the field?

- "Hallig Hooge" is an island in the North Frisian Wadden Sea and yes, it is yet still there.

Line 225: Use PSU everywhere or even no unit at all for salinity.

- We agree, this needs to be unified.

Figure 2a: If I understood properly, there were only 2 vessels per incubator so, to avoid confusion, you should remove 6 of the 8 vessels drawn in figure 2a.

- This of course needs to be adjusted.

Figure 2e: this picture is not very clear. Is the shell of the foram entirely fluorescent (ie born in calcein bath)? Otherwise, how many chambers are labelled here? I have the feeling that this is the cytoplasm that exhibit high fluorescence at the bottom since the fluorescence is patchy and fill half of the last chamber...Could you try to show a better picture?

- No, the shell is not entirely labelled. Only the last 2 ½ chambers are labelled. It can be excluded that only the cytoplasm is fluorecenting because the specimen was dead, cleaned and dried. Therefore, no cytoplasm should be there anymore.

Line 152: Why did the authors used a concentration of 16mg/L which is different from the recommended concentration given by Bernhard et al. (2006)?

- See Reply to Reviewer 1.

Line 156: To my opinion, this is not enough time to remove the calcein from the vesicles in the cytoplasm. Anyway, if this seawater is used to calcify 1 new chamber in your experiments, you can hope that this new chamber would exhibit a small fluorescence.

- Reviewer 2 is right. A sufficient time is needed to remove the calcein from seawater vesicles in the cytoplasm. If the foraminifera are taken directly from the calcein staining bath for incubation, all subsequent chambers will be stained (see Haynert et al., 2011). In our case, the youngest chambers were not stained in that a purification time of 1 or 2 days was sufficient.

Line 160: Dagan et al., 2016 is a report. Is it available online somewhere?

- The report is not available to the public but Woehle et al., 2018 reported the experimental setup as well in the online supplement. Dagan et al., 2016 is therefore to be deleted.

Line 171: it is the air that was filtered?

- Yes.

Line 171: The authors do not mentioned pH or alkalinity measurements. Did they measure carbonate chemistry during the experiments? At least pH has been measure since it is mentioned in discussion on line 580 "As the pH during the experiment was stable around 8.0 ± 0.1 (measured twice a week)". This information should arrive in material and methods.

- Carbonate chemistry was not measured during the experiment. We will add the information, that pH was measured in the "Material and Methods" chapter.

Results

Table 2: In C2 for A. aomoriensis, does it mean that on the 10 forams recovered, 2 were dead but all of them (10) had calcified new chambers?

- Yes.

Line 368: Since the Ammonia calcified usually more than 4 new chambers, is it possible to see the evolution of seawater metal concentration in the successive chambers of 1 given individual? At least in phase 3? This could help to gain precision in the estimated DTE…

- The evolution of the metal concentration in seawater of phase 3 was only indicated in some individuals *of Ammonia aomoriensis* and *Ammonia batava*. Particularly, the first high concentration of certain heavy metals could be found in the first chambers after the staining (i.e. the first chamber built in culture). But this was not the case for all individuals, which is most likely due to the individual timing of calcification. It also cannot be determined, at which point in time the foraminifera calcified within one phase. Therefore, a mean value over the whole culturing phase is most representative.

Figure B1: Could you indicate the error of the measurement on the graph? ON line 340-344, the authors explained that they fit a regression curve on the data to calculate a weighted mean per phase. This seems a good idea when 4 data points are available within a given phase and that a trend can be seen (eg phase 3 for Cr, Ag, Sn). However, this seems difficult when only 2 data points are available and very different (eg Cu) or when the trend is not regular (eg phase 3 for Mn, Ni…). Actually, did you realise that Mn, Ni, Zn and Cd show similar variation though time in phase 3 (lower value at the second sampling time) compared to Cr, Ag, Sn or Pb which show decreasing trends?

- As these are single measurements, the error that could be provided would be based on frequent measurements of the seawater reference materials. The respective values are given in Table A2 but will also be added to this figure.
- When only 2 data points are available a linear regression was made, which is in our opinion the only way to account for the different concentrations because we do not know at which time within a phase the new chamber was built. If no clear trend was observed, the regression with the highest fit (highest p-value) was chosen.
- It is indeed interesting that Mn, Ni, Zn and Cd show similar patterns in phase 3. We could think about possible mechanism affecting all of these metals at the same time.

Figure 3: I have the idea that the use of weighted means and standard error of the mean instead of standard deviations, the authors reduce artificially a lot the real elemental variations that they have, mainly in phase 3. Maybe the figure could be completed showing the range of values actually measured in shadow or use box plot to better represent the variability of this artificially created dataset...

- It is true, that the variation carry less weight in this figure and this is why we added figure B1 in the appendix. Nevertheless, the variability in the seawater during one phase can be added to this graph or an extra Box-Plot can be created to clarify this.

Figure 4:

How are calculated the statistics of the correlations? These correlations should not be based only on the mean values per phase but on the all data set. For example for Ag and Pb, the R^2 and p values are really good but the D is only based on the Phase 3 data which has a high variability! Therefore the D value is not precise and robust.

- The statistics of figure 4 are indeed based on the mean value per phase and not on the entire data set. The plots were made using the software Grapher, which is calculating statistics along plotting. Furthermore, the program PAST was used to calculate statistics. We actually also calculated R^2 and p-value based on the whole data set, which was e.g., for *A. aomoriensis* with phase 3 comparable to the statistics based on the mean. This is why we decided to go with the mean values. We can add this information to the manuscript.

Figure 4 and Table 4: The authors have no objective reasons to fit the correlation through 0 for some elements and not for others. It could be decided on statistical arguments but I have the idea that the authors did not check this.

- We tried to fit all element correlations in all species through the origin, because a real correlation would also include the origin. Only in cases where this was clearly not possible (Mn of *A. batava* with phase 3 and Hg of *E. excavatum* without phase 3), because the course of the regression line changed significantly or the R^2 value decreased, no forcing through the origin was applied. We can clarify this in the revised version of the manuscript.

For A. aomoriensis Mn/Ca, there is a problem with the correlation line. This is not possible that the line don't go through the phase 3 datapoint. Please check.

- The line is not going through the data point from phase 3 because the line is forced through the origin, which is changing the course of the line minimal. It should be clear to the readership that not only the data point from phase 3 but also the data from the other phases are driving the course of the regression line. Furthermore, the R^2 – value of the regression line did not decrease when forcing through the origin, this is why we decided to include the origin.

The graphs for this figure should have similar y axis range for a given element for the 3 species so that the difference of incorporation between species is highlighted. All graphs should start at 0on the y and x axis. I think that the main (and most robust) output of this study is the difference of incorporation between Ammonia and Elphidium species and this is at the moment only shortly discussed and observable in graphs. This is a shame.

- The axis can be adjusted and the differences between species will be more of a focus in the revised manuscript.

This is a really good idea that the authors also analysed their data without the data from phase 3. To my opinion, this phase is important to get a trend because the problem when you remove it is that you have no correlation anymore, probably because the range of seawater elemental concentration is not wide enough. On the other hand, when phase 3 is considered, then a more relevant D value can be calculated but the correlation are only based on this data points and therefore the correlation is not statistically robust.

- Yes, it is true that the correlation gets lost because the range of the metals in the seawater is very narrow without phase 3. Furthermore, it is also true that point three makes the correlation statistically more robust but nevertheless, figure B2 also shows that the general trend is still visible without phase 3 for some elements. Forcing through the origin further adds a fix point, which provides at 3 points, though artificially, and not 2 only. We can clarify this in the revised version of the manuscript.

Line 473: Now authors are removing phase 3 and 2?

- We did not remove phase 2 from the calculation of the regression line but if one has a look at every data point from a phase individually (meaning without any calculation of regression), the D_{Cd} and D_{Cr} values from phase 0 and 1 are >1 while the D_{Cd} and D_{Cr} from phase 2 and 3 are <1. This needs to be clarified and rephrased.

Line 464 to 481: this is very descriptive and difficult to follow…

- As this is part of the "Results" section, a description of the data is appropriate. We rearrange the paragraph to make it easier to follow for the reader and focus more on the results that will be discussed later in the manuscript.

Discussion

It is not possible to discuss the significance and meaning of partitioning coefficient that are showing a very high range since this variation is meaningless to my opinion in terms of biomineralisation processes… For example, DCd are varying from values below 1 to values such as 10-20 even 50 in all species (lines 678-679). In terms of incorporation mechanisms, that would mean that some specimens are fractionating against Cd whereas some others (from the same species and in the same condition) would concentrate this element! I would suggest to the author to rather focus on:

Elements were a positive correlation is found but instead of using the mean TE/Ca value (eg line 557-559), they should take into account the variability of the data and give a SD for the slope (ie for the DTE). They should also be aware and acknowledge in the manuscript that these correlations are driven by the phase 3 data and might be imprecise.

- It is probably useful to add the SD for the slope of the partition coefficients and to go more into detail concerning the uncertainties of the calculated D_{TE}. It is also reasonable to note that the correlation is driven by the data point from phase 3. We will add this to clarify the circumstances for the reader.

Elements were the range is relatively low so that a general tendency/interpretation might be given.

- We can separate the elements with a smaller variability and discuss the behavior of those elements individually.

Finally, do not discuss further forward the other elements that exhibit vary wide DTE also if no literature is available on this element, it is interesting to know that this is incorporated and measurable in foraminiferal calcite.

- We may shorten the discussion of the elements with higher variability. But nevertheless, a proxy is only as good as it´s variability and therefore we think, that it is important to mention variable D_{TE}´s too.

I have the feeling that the authors use the DTE with or without phase 3 when it helps them to compare with the literature. This is bothering me: is phase 3 really usable to calculate a partitioning coefficient knowing that the seawater concentration of the metal was not stable during this phase and the regression line is totally driven by this single condition?

- The regression is not only driven by the data point from phase 3, because other points and the origin also play a role, which is already demonstrated above and can be seen in figure B2. But nevertheless, phase 3 is very much driving the slope of the regression line. Even though the seawater concentration was not as stable as during phases 1 or 2, we are convinced that it is appropriate and justified to use a mean value calculated from the individual fit curve for every element and to create it to the mean value of the foraminiferal calcite. It is possible, that the variability of the seawater concentration in our study is higher because we measured more often than other studies did. This means that other studies simply not monitor the variability. Furthermore, pollution events in nature are also transient events rather than stable once. We can discuss this maybe in the "Experimental Uncertainties" section a little further.

Line 506: The authors mentioned the growth of algae as a reason for element concentration changes in the seawater but I understood that the algae were given dead. Therefore, one would not expect algal growth in the experimental set up?

- This is a misunderstanding. The algae that were fed were dead, but germs of other algae were introduced without purpose together with the living foraminifera and grew during the experiment. These algae preferentially grow on plastic surfaces and create biofilms. Therefore, it is well possible that these films also took up metals.

Line 523-528: this paragraph should be more or less upside down. Since you used calcein prior to the experiment, you do not have to worry that this probe could have impacted the elemental concentration in your forams. This paragraph could therefore be shorten.

- We agree.

Lines 551-552: according to Erez endocytosis biomineralisation, I thought that the composition of the seawater vesicle (ie Mg content) was also modified somehow?

- Yes, this is partially correct. "Endocytosis" as such describes only the uptake of a seawater vacuole, which is subsequently modified during their pathway in the cell. This needs to be clarified.

Lines 559-561: this is interesting but where can we see this information (ie. D vs seawater trace element concentration)?

- Figure 4 shows this indirectly, we will refer to this figure.

Line 559: if D>1, this means that the foram is concentrating the element inside its shell. Therefore, I would not define this as a "non-selective uptake", no?

- "Non-selective" at this point referrers to an uptake that is not driven by the chemical property of the ion size of the metal ion itself. This can be clarified in the revised version of the manuscript.

Line 561: other studies have observed the same trend of decreasing D with increasing seawater concentrations: Mewes et al. (2015) for Mg and Barras et al. (2018) for Mn.

- The references will be taken into account.

Figure 5: The authors refer to this figure for each element but I think that this is also interesting to observe that there is apparently no trend between D and the ionic radius to charge ratio.

- Yes, this is true, but this figure should mainly provide information whether the D_{TE} is higher or lower than 1 to the reader. Nevertheless, we can add that no clear trend between D_{TE} and the ionic radius is observed. It would maybe make sense to remove this figure.

Figure 5: it is strange to me that the author used a single data point for each LogD value. Is it the mean of all measurements? In this case, it would be nice to see the SD since D might be highly variable.

- Yes, this is the mean value in cases were no significant correlation between the heavy metal concentrations in seawater and calcite was found. In cases were a correlation was significant, the slope of the regression line was used. Indeed, D_{TE} is variable and the SD will be added in the revised manuscript.

The authors compare their D values to the literature. Sometimes they compare these values to tropical symbiont-bearing large benthic forams (high Mg content species) or miliolids (line 635, 671, 708) which are known to incorporate much more elements than Ammonia for example and other small benthic foraminifera (low Mg species) (cf van Dijk et al.,2017). This should be specified and discussed.

- See Reply to Reviewer 1.

Chapter 4.3: as mentioned before I think that Figure 4 should be reworked (or a new figure) in order to observe more easily the differences between species (e.g. similar axis for Ammonia species and different (if needed) for Elphidium). Maybe differences between species would be even better observed when considering only phases where the seawater elemental concentrations are stable?

- See Response to Reviewer 1. Stable concentrations occurred in phases 0 to 2 and only phase 3 had higher variations in the trace element concentrations. Therefore, figure B2 in the appendix, which is showing TE/Ca in calcite versus TE/Ca in seawater without phase 3, can clarify the species-specific differences in the heavy metal incorporation. This figure will be adapted in the same way as Figure 4 and we will add a figure comparing different species.

Line 724-725: Food is added quite regularly during the experiment. Could the deposition of a layer of food at the surface of the "sediment" could create microenvironments within the hole of the weel-plate? Indeed, the food would be degraded and could influence pH and O2 conditions for exemple…

- It is indeed possible that the food deposited as a thin layer on top of the sediment, which could have created a microhabitat. This effect would be the same for all cavities and therefore for all three species. In account of this, species –specific differences in the heavy metal incorporation cannot be caused by this effect. This will be mentioned in the revised version of the manuscript.

Line 727: Read van Dijk et al. (2017) paper but I don't think that the hhigh elemental incorporation of symbiont bearing forams is due to the presence of symbionts but rather to the fact that they are high-Mg content species. Other symbiont barren large benthic forams exhibit high elemental incorporation.

- We agree, the high Mg-content of the calcium carbonate of the species in the tropics at high temperatures and salinities could play a role, which will be discussed in the revises version of the manuscript.

Line 735-737: be aware that there is a difference between number of chamber added (individual growth rate) and calcification rate (crystal growth rate). Depending on the element, one could expect that slower calcificiation would give more time to remove (or discriminate more against) the element as it is the case for example for Mg.

- This is an interesting aspect, which we will include in the discussion.

Line 738-742: this is an interesting point. I think the authors could potentially unravel this problem if they compare Elphidium data with the first chambers calcified after the calcein stained chamber. Indeed, that would be the forst chamber calcified in the experiment when the seawater elemental concentration was probably the highest. Moreover, as previously mentioned, you could have a look at successive chamber composition to see if you can observe a decreasing elemental composition for the elements exhibiting decreasing trend in seawater.

- *Elphidium* mostly build only one chamber, which means that the data presented here are already from the first chamber calcified after staining. This makes a tracking of the decreasing concentration impossible. For both *Ammonia* species, see comment above.

Table 5: this table is very interesting and complete but to my opinion, it could be moved in supplementary materials.

- We can move the table to the supplements.

Table 5 : how were the metals analysed in these studies? Analytical techniques used? Extractions? Speciation of the metal?

- This can add the information, which analytical technique was used to determine the heavy metal concentratin in the comparing studies.

Line 707: this paper from Remmelzwaal refers to post-depositional overprinting. I don't kown this study but are you sure that this DCr corresponds to primary calcite values?

- Yes. They performed culturing experiments with different foraminiferal species and calculated this D_{Cr} based on these experiments.

Conclusion

I think that the authors could highlight the interest to use fossil records (or regular sampling of living forams through time or space) to determine the relative variations of seawater metal concentrations in porewater through time. Although quantitative reconstructions are to my opinion not feasible at the moment, relative variations are usable for elements where a correlation was observed between shell and seawater ratios (not for all elements). The authors should be more realistic in their conclusions.

- We agree, reference to the fossil record is to be given.

The authors could also highlight the interest of forams as they are integrating in their shell the metal concentration over a certain period of time. Indeed, dissolved metal concentrations measured directly in seawater (for monitoring purposes) give the concentration the day of the sampling but this concentration may vary very rapidly… Both aspects should even be mentioned already in the introduction.

- We agree that foraminifera offer the opportunity for long- and short-term monitoring of changes in the heavy metal concentration, because they are recording the environmental signal. This will be added in the introduction and in the conclusions.

Line 795-796: ok but there is no impact on survival or growth in your experiments.

- Any organism reacts in a protective way before harmful or lethal effects do occur. This is also why a reduced incorporation of a certain metal could point towards the onset of a protective mechanism prior to damage of the organism and may also prior to a reduced growth and following death.

New References

Dueñas-Bohórquez, A., da Rocha, R. E., Kuroyanagi, A., Bijma, J., & Reichart, G. J. (2009). Effect of salinity and seawater calcite saturation state on Mg and Sr incorporation in cultured planktonic foraminifera. *Marine Micropaleontology*, *73*(3-4), 178-189.

Duffus, J. H. (2002). " Heavy metals" a meaningless term? (IUPAC Technical Report). *Pure and applied chemistry*, *74*(5), 793-807.

Francescangeli, F., Milker, Y., Bunzel, D., Thomas, H., Norbisrath, M., Schönfeld, J., & Schmiedl, G. (2021). Recent benthic foraminiferal distribution in the Elbe Estuary (North Sea, Germany): A response to environmental stressors. *Estuarine, Coastal and Shelf Science*, *251*, 107198.

Lutze, G. (1965). Zur Foraminiferen-Fauna der Ostsee. *Meyniana*, *15*, 75-142.

Nikulina, A., Polovodova, I., & Schönfeld, J. (2008). Environmental response of living benthic foraminifera in Kiel Fjord, SW Baltic Sea. *eEarth*, *3*, 37-49.

Schweizer, M., Polovodova, I., Nikulina, A., & Schönfeld, J. (2011). Molecular identification of Ammonia and Elphidium species (foraminifera, Rotaliida) from the Kiel Fjord (SW Baltic Sea) with rDNA sequences. *Helgoland Marine Research*, *65*(1), 1-10.

Schmidt, S., & Schönfeld, J. (2021). Living and dead foraminiferal assemblage from the supratidal sand Japsand, North Frisian Wadden Sea: distributional patterns and controlling factors. *Helgoland Marine Research*, *75*(1), 1-22.