RESPONSE TO REVIEWER 1

The authors have strengthened their analyses by introducing new filtering criteria and refining their methodological approach. Notable improvements include the application of transformation-centred log-ratio transformations and the Aitchison distance dissimilarity measure for ordination and PERMANOVA. Additionally, the study now integrates complementary analytical tools such as PICRUSt2 and ALDEx2, allowing for a more comprehensive inference of differentially abundant metabolic pathways within the foraminiferal microbiome in relation to the surrounding water column.

The results are clearly presented, and the methodology is well-documented. These refinements contribute to a more robust and informative study.

However, addressing the following points would further enhance the clarity, coherence, and reproducibility of the manuscript.

SPECIFIC COMMENTS:

1. On page 8, lines 165–168, the authors mention performing rarefaction and calculating alpha diversity metrics. However, corresponding rarefaction curves do not appear to be included in the supplementary material. Adding these curves is necessary as it would help assess whether the sampling effort was sufficient to capture the diversity of the prokaryotic community in both foraminifera and water samples.

Response: These curves (Shannon, observed-features, Faith PD) are now added to the appendix (Figure A1).

2. In their response, the authors state:

"As reported under point 2 above, in our new analyses we have shown that "Station" drives 2.3% of the differences in ASV composition (rather than 49%), and is not significant (PERMANOVA, Adonis). However, the provenances "foraminifera" and "water column" are significantly different in both the FW and 101 datasets. Here then, we have been able to reject the null hypothesis that there were no differences in the ASV compositions between "Provenances" and accept the null hypothesis that there are no differences between "Station", indicating Provenance rather than Station (location) has more influence on the ASV composition"

However, the multivariate analysis performed on the 'Forams only' subset indicates that 48.3% of the variation in foraminifera microbiome composition is attributed to station (lines 315-318, p. 16). In contrast, the same analysis on water samples (stations 101, 115, and 323) does not show a significant relationship with station (Fig. A1). Furthermore, when the 'FW' subset (including both water samples and foraminifera) is analyzed, station is no longer a significant factor.

Based on these findings, the authors conclude that 'provenance' rather than location is the main factor shaping the foraminifera microbiome. However, these results appear to be somewhat contradictory. If station explains a substantial proportion of variation in the 'Forams only' subset but not in the water samples or the combined 'FW' subset, additional clarification/discussion is needed in the manuscript.

To further investigate this discrepancy, I recommend adding another subset and repeating the analysis using only the foraminifera microbiome data from the same stations as the water samples (which likely corresponds to the 'FW' subset minus the water samples). This additional analysis could help determine whether the observed patterns in the 'Forams only' subset are driven by differences in water column composition, ultimately clarifying the influence of microbial community assembly processes in the surrounding water on the foraminifera microbiome.

Given that the authors define the core microbiome using all foraminifera data, it is crucial to discuss the ecological and biological factors shaping its composition to ensure a comprehensive interpretation of the results.

Response: We have done a series of additional analyses to tease these out. The additional/improved methods are reported in Section 2.4. These include:

- Using adonis2 from the Vegan package as an improved version of Adonis in QIIME2 which better handles unequal sample sizes.
- Using the R package *pariwise.Adonis* to investigate which Stations were driving the statistical differences observed in the foraminiferal dataset
- Testing for homogeneity of multivariate dispersions among groups using bestdisper on vegan to test whether the assumptions of PERMANOVA are violated.
- Implementing Wilcoxon test in ALDEx2 where PERMANOVA assumptions are violated.

The additional tests are reported in a new section "3.2.1 Station and Depth as factors influencing microbial assemblages"

- In the water samples because Depth was significant but Station was not, we considered that the multiple depths within each sample might be masking the effect of station. We therefore carried out a test on 50m water samples only.
- Because station was significant for foraminifera despite not being significant for water, we did several things.
    - First we took a more statistically robust set of foraminiferal samples excluding stations where n=2. Station remained significant in this dataset.
    - We also did pairwise analysis on both the full and the robust datasets to identify which pairs of stations were significantly different.
    - As suggested by the reviewer we made a small dataset including only foraminiferal samples from the stations where water was also taken so we could directly compare Station significance on the water and the forams independently. This small dataset violated the assumption of homogenous dispersions for PERMANOVA and therefore we ran ALDEx2 Wilcoxon tests as mentioned above.

We have discussed the results in a new brief section "4.1 The influence of station and depth on microbial assemblages". The main conclusions are

- Depth is the strongest driver of differences in microbial assemblages
- Only a small subset (4/21 pairs of stations in the foraminiferal dataset) are significantly different and these drive the global PERMANOVA results.
- Station has a more subtle influence than Depth, and those stations furthest apart are those that show significantly different foraminiferal ASV composition.

- The co-sampled (water and foram) stations, 101, 115 and 323, are not significantly different by any test.
- Therefore, the difference in significance between the foraminiferal and water column datasets likely reflects the broader geographic coverage of the former, along with host-mediated retention of a stable core microbiome through selective feeding.

3. The results support the previously hypothesized POM feeding mode and suggest that *N. pachyderma* also preys on living diatoms.

A recent study by Meilland et al. provides experimental evidence for this feeding behavior in cultured *N. pachyderma* from high latitudes (Type I). Citing this reference and incorporating its findings into the discussion would provide a more comprehensive context for interpreting the feeding ecology of *N. pachyderma*.

Meilland, J., Siccha, M., Morard, R., & Kucera, M. Continuous reproduction of planktonic foraminifera in laboratory culture. Journal of Eukaryotic Microbiology, e13022. *https://doi.org/10.1111/jeu.13022*

Response: Great idea, thank you. At lines 529-534 we have added this reference and some discussion on the species that best support growth in culture versus those identified in the natural environment.

4. Throughout the manuscript, microbiome community structure in foraminifera and water samples is described using expressions such as "% of all ASVs" (e.g., line 329, p. 16) and "contained >50% chloroplast ASVs" (e.g., line 359, p. 18). These phrases describe the number of ASVs rather than their relative abundance (i.e., read proportions) but are referenced alongside figures that likely represent read proportions.

To avoid potential misinterpretation, I recommend explicitly clarifying that these values refer to read proportions and updating the y-axis labels in Figures 3 and 6 to specify "% reads."

Response: This is a good point and does require clarity. We have added a line at lines 176- 179 in the methods to explain that the ASV counts were converted to relative abundances (percentage reads) for the analysis to make figures 3 and 6 and that all percentages discussed are relative abundances (read proportions) converted from read counts. We have changed the y-axis labels to “% Reads” it the figures as requested.

5. The statement regarding raw sequence data (line 248, p. 12) should be moved to the "Data Availability" section. Additionally, to enhance reproducibility, I strongly encourage making the environmental contextual data, ASV table, and sequence files publicly accessible, as previously suggested.

Response: The statement at line 248 has now been moved to the data availability section. A DOI giving access to the CTD data, pairwise analysis, ASV table and sequence data on Figshare (DOI: 10.6084/m9.figshare.28915598e) is also now provided in this section. At time of writing, this DOI has been reserved and will be released when the dataset is published.

RESPONSE TO REVIEWER 2

1. The revision of the manuscript by Bird et al. is quite improved over the original. The revised contribution, with updated figures and Conclusion paragraph, is more focused and clearer, although I am not a specialist in molecular approaches/methods. As with all manuscripts, there remain a few points of concern to consider.

Response: Thanks go to both our reviewers for giving generously of their time and their steely determination to improve this manuscript considerably.

2. The Abstract (line 17) and Introduction (line 70) both imply this species was studied by these authors "throughout the annual cycle of the Arctic", but they only sampled in summer.

Response: Thank you for pointing this out. It certainly was not our intention to imply this, so we have rewritten these lines to make clearer that our work covers only the summer season. In the abstract we indicated that we studied only the summer populations at line 20, but have moved this statement earlier (line 17), to make this clear immediately. At line 70, since this is a broad introductory section, we have just tweaked the wording a little to make it clearer that this is something that will need to be done, rather than something we are reporting. So it now says "To model the impending environmental consequences for this important high latitude species going forward it will be vital to investigate the modern ocean community structure throughout the annual cycle of the Arctic to understand the inter-dependencies of *N. pachyderma*."

3. The statement on line 345 about ASV956 being the most abundant chloroplast ASV and assignment to class Bacillariophyceae is not supported by Fig. 6 in that the dominant taxon shown in Fig. 6 is Chaetoceros (dark green), not Bacillariophycaea (light green). I understand Chaetoceros is a genus of Bacillariophycaea. The authors should check their wording to be sure it corresponds to the figure. If the figure and text are congruent, the authors are urged to check the passage (lines 344-347) for clarity.

Response: Lines 395-387 have been reworded to make this clearer. Since ASV956 was identified to the species level it was grouped in the *Chaetoceros* genus in Fig 6 and so makes up the dark green bar.

4. The authors insist on use of the term "shell" vs "test" (response page 12), yet they use the term "test" to describe the foraminiferal hard part on line 486 and perhaps elsewhere. Why the inconsistency? The external cytoplasm (which is why it is called a test, not a shell) is discussed by them on line 397. Further, a google search of "What is the hard part of a foraminifera called?" results in the term "test", not "shell". The authors can use an incorrect term but let the record show that I strongly object to such terminology.

Response: The reviewer makes a compelling and sensible argument, so we have changed the term shell to test throughout.

5. Also regarding terminology, the authors use an odd abbreviation for foraminifer ("fm"; Table 1, labels in Figs. 3 & 6, etc). If the authors are so concerned about misinterpretations of the term "test" (response document pg 12) then why use "fm", which to most oceanographers, marine scientists, and English-speaking people means fathoms?

Response: This had not occurred to us! We request to leave the abbreviation Fm in place at this time, as it is deeply embedded in the data set from the sequence identities in the original files generated by the sequencer and within all the code throughout the analysis. It would be a massive challenge to change it at this stage, and we hope the reviewer will be content to allow its use on this occasion. We will not use it again.

5. Unfortunately, there are some statements in the authors' response document that simply are not true. The response notes they would add "e.g.," to instances where the cited papers are examples (see page 18 of response document) versus cases where a complete list of existing publications is cited. The "e.g.," appears a few times (lines 38, 48, 72, 74, 636) but there are many more where it should appear, including lines 59, 555, 580-581, 586, 608, 611, 619, 670, 672, 678, and 685, at a minimum. The authors' response also noted (pg 18) they would cite the benthic foraminiferal kleptoplasty review by Bernhard and Bowser (1999), but they did not. A good place to cite it is line 555 and/or lines 608-609.

Response: Please forgive these oversights. We have amended our "e.g." additions accordingly which can be seen throughout the newly marked up submission. We have finally added reference to the paper by Bernhard and Bowser 1999 at lines 646 and 681. We are sorry for this omission - we were convinced we had added this paper in the last round of edits as it is very important for the discussion on kleptoplasty and was one of the first to show how abundant kleptoplasts can be.

6. There continue to be at least a few missing references (aside from not adding the one noted above), including Salonen et al. 2021 (cited on line 647) and Westgard et al. 2023 (cited on line 667). Clearly one of the authors must do a careful and complete check for such omissions through the entire manuscript.

Response: The authors did do a careful and rather painstaking check through the references, including checking the journal abbreviation consistency etc. We are sorry that some references still slipped through the gaps and have checked again. As well as adding a couple of additional references for additional R packages used in the stats analysis and discussion added as requested by reviewer 1, we have added two further references that we missed previously (Quast, and Qu).

7. Regarding TEM images of lipids, I must reiterate that the lipids in Fig. 9a are extremely dark compared to lipids of most benthic foraminifers. For excellent examples of what benthic foram lipids should look like, see LeKieffre et al. 2018 Mar. Micropaleo., Figs. 7 C, F; Fig 10 D, E; Fig. 12 D. The issue is that "electron opaque bodies" that truly are black (electron opaque) exist in benthic forams and the authors should be aware that these differ from lipids. Why are there no (black) lipids in Fig. 9B (no black structures)? Such dark lipids do not occur in most of the images in Fig A6 either. None of these comments on lipids require any changes to the contribution; the comments are only to provide the authors with information.

Response: Thank you for the clarification on "electron opaque bodies" this is useful to understand, and we have modified the manuscript at lines 468-469 and in the Figure 9 legend at line 473, to take this possibility into account, as we feel that it is important based on the reviewers feedback.

Minor points

"Microbiome" should be defined at first use in the Introduction. Now, it is defined in the Discussion (line 429).

Response: This has now been added at line 76.

The proper term for localized "extinction" is "extirpation" (line 61).

Response: OK, we have changed this.

Define "TEM" (line 78).

Done

Remove the blank line (line 84).

Done

Define "interactome" (line 90). Is its use valid in this work? Interactions were not investigated.

Response: This is the term used by Greco et al., 2021 which is cited here, but we have changed it to 18S microbiome, as Greco et al are referring to the results of the 18S metabarcoding dataset, so microbiome seems appropriate and we agree with the reviewer that interactomes is less appropriate as it usually refers to molecules.

Consider changing the antiquated, generally disfavored "prokaryote", e.g., on lines 92, 319.

Response: Good point. This has been changed to "bacterial and archaeal" to reflect the targeting of these two very different groups.

Lines 114, 376 and elsewhere: do not start a sentence with a number or abbreviation.

Response: This has been checked and changed as necessary. However, there are several sentences that start with PCR and with ASV, and we feel that these two abbreviations should be left as such, even at the start of sentences just for ease of reading. There are plenty of manuscript examples where both PCR and ASV have been used at the start of sentences. I am happy to change these, however, if the editor feels it is necessary.

Change "samples" to "sample" on line 254.

Done.

Add comma after "ASV" on line 284.

Done

Change "Chloroplasts" to the singular in header to section 3.3.2 (line 326).

Done.

Omit the errant "d" near beginning of line 372; add comma after Fragilariopsis (before Synedra).

Response: This is not an errant “d” -it is part of the sentence: “Except for Fm176b which had only 3 chloroplast ASVs, Fm176a, c, and d contained a much higher relative proportion of…” so we have made this more clear by stating “Fm176a, 176c and 176d”

The passage on lines 398-400 belongs more in the Discussion than in the Results.

Response: The sentence referred to is : “These observations support the previous literature indicating that *N. pachyderma* consumes diatoms and given the intact nature of the diatoms observed (Fig. 8a), it is not only detrital (dead) diatoms that are consumed, as reported by Greco et al. (2021).”

It has been deleted from this location and rejigged into line 426-427 of the discussion.

Omit comma after “al.” on line 441.

Done.

Omit comma after “diatoms” on line 442.

Done.

Line 445 says “sits”, which is colloquial and anthropocentric. Rewrite. Consider “is situated”.

Response: We have changed “sits” for “remains”

Shouldn’t the two “or” on lines 468-469 be “and/or”?

Response: Yes- that makes more sense -thank you.

Check the math for statement on lines 471-472. Should the average be 0.89% instead of 8.93%?

Response: No, this is correct -so we have amended the wording to state “Six bacterial ASVs each contributed small percentages between 0.09 %- 2.96 %, averaging a total of just 8.93% of the core microbiome between them.

Is it proper to report values to the hundredths (lines 471-472, 552, and elsewhere)?

Response: We have chosen to report two decimal places where the numbers are very small (such as in the core microbiome) to distinguish between the ASVs at these lower levels of relative abundance.

Omit comma after “Cellvibrionales” on line 488.

Done

Omit comma after “C” on line 496.

Done

Remove the third “the” on line 547 (to read “…beginning of vertebrae evolution…”

Done

Add comma after “diversa” on line 575.

Done

Line 598 is underlined but other sections of that stature are not underlined. Be consistent.

Done

Capitalize the “k” in LeKieffre (line 605 and, perhaps, elsewhere).

Done

Omit comma after “cytosol” on line 619.

Done

Add citation at the end of sentence ending on line 627.

Done

Omit comma after “paleoenvironments” on line 668. Why use the US spelling given European spelling / words are used elsewhere (e.g., “anticlockwise” in figure caption)?

Response: Good point, spelling has been amended, and comma removed.

Add space in proper location on line 696 (“of2017”).

Response: There is already a space here -although it doesn’t look like there is!

Suggest changing “this” to “that” on line 697, to read …microbiome at that time.”

Response: Good idea, done.

Hyphenate “nutrient rich” (line 698).

Done.

The supplemental tables continue to use different fonts (see Table A1 vs Table A2).

Response: They are now consistent in Times New Roman size 9.