Dear Editor,

Please find enclosed the revised version of the manuscript entitled “Planktonic foraminifera-derived environmental DNA extracted from abyssal sediments preserves patterns of plankton macroecology” modified after the comments of the reviewers. The revised version of the paper is provided as a supplement.

We are thankful to both reviewers for their constructive comments. The comments made by the reviewers were complementary which is why we decided to reply jointly to them. We have supplied additional information in the figures 1, 5, 6 and also modified the figure 4 to make its interpretation easier. We have also rewritten part of our conclusion to nuance statements and also clarify technical points following the comments of the reviewers. We have tried to answer the issues raised by the reviewers as constructively as possible and hope that it helps to further clarify our paper. The detailed answer to the comments are below.

We believe that the modifications we have further improve the manuscript such that it could be deemed suitable for publication.

Sincerely yours,

Raphael Morard

In the following, the comments of the reviewers Indicated by these symbols ***…*** and our responses are indicated by these symbols >>> …. <<<. We specify always the line number in the modified version of the manuscript with track changes that is attach at the bottom of the response.

Reviewer #1

*****The manuscript entitled ”Plankton-derived environmental DNA extracted from abyssal sediments preserves patterns of plankton macroecology” address relevant scientific question within the scope of BG. The manuscript focuses on the identification of planktonic foraminifera from marine surface sedimentary DNA using a metabarcoding approach. The manuscript uses already available data and focuses on a data subset from data in which the entire foraminifera diversity (including benthic taxa) was analysed. The authors are specifically interested in the planktonic taxa, because this offer the opportunity to investigate imprints of the ocean surface biota and its processes of transport and deposition of planktonic eDNA to the oceans bottom. I like the idea of the manuscript, because there are especially in the marine realm only a few papers that focus on sedimentary DNA. Especially in deep ocean sediments it is a valid question to ask how much eDNA from planktonic organisms reaches the bottom and which is then also usable as an archive for planktonic taxa diversity changes. As the author have specific knowledge about the planktonic taxa and a well documented reference database, the detection of specific and rare taxa seems to be possible and valid.**

However, I have the feeling that the small subset of the entire data might be problematic in terms of the complete diversity of planktonic taxa. I have the feeling that the dominance of benthic taxa in their re-analysed data set is caused by the effect that DNA from benthic organisms is less degraded and is therefore preferentially amplified by PCR. By this the reads are dominated by benthic taxa, which reduces the number of planktonic sequences. ***

>>> The reviewer correctly highlights the fact that the majority of the amplified DNA belongs to benthic taxa. This is to be expected because only benthic foraminifera are alive on the seafloor and thus deliver intact "live" DNA. As a result, the number of reads from the environmental samples that could be assigned to plankton varied and in some samples was small. This could theoretically affect the results but only if it can be shown that the patterns we observe result from unequal number of plankton-assigned reads among the samples. We are aware of this problem and highlighted this in the discussion, page 13 on lines 1-16 in the modified version. We observe that the patterns are replicable with assemblage composition in the Caribbean samples consistent with other tropical sites despite extremely low number of reads that were recovered (see Fig. 4). The lack of specificity of the primers used for the amplification of foraminiferal DNA was because the sequence libraries were originally generated to study the distribution of benthic taxa, not to study the deposition of planktonic eDNA on the seafloor. Future studies of planktonic eDNA could use more specific primers increasing the yield of the target sequences.<<<

*****I would be interested to see next to the relative proportion also the absolute number of sequence reads of benthic compare to planktonic reads.*****

>>> We have added to the figure 1 three histograms to show (1) the absolute number of reads generated for each sample (2) the relative proportions of benthic and planktonic taxa and (3) the total number of reads retained for each sample. <<<

*****If you have strong variations between the total read numbers of the samples, this might be the reason that in some cases you detected only a very small fraction of planktonic reads, which is then not representative for the sample in terms of abundance and diversity. I think this makes it difficult to see global trends in planktonic foraminifera taxa in this dataset. *****

>>> We are perfectly aware that in some samples the full spectrum of the diversity is not recovered. The present article does not intend to characterize the diversity of foraminifera, but rather by taking advantage of our knowledge of it, to show that the transfer of eDNA from the plankton to deep-sea sediment preserves community structure and known macroecological patterns. However, we point out that despite the strong variation in the total number of reads between Japan and the Caribbean regions (Figs. 1, 2), the dominant species remain the same (Fig. 3), which explains why the structure of both sets of communities is replicated within each of these regions (Fig. 4d). We acknowledge that we are not trying to demonstrate that the full spectrum of planktonic Foraminifera diversity is recovered in the re-analyzed dataset. We specify in the discussion section that a certain minimal sequencing effort would be necessary to recover it (page 13, lines 4-7 in the modified version).<<<

***** Further I would suggest that the author show that the genetic marker used is equally specific to benthic as well as planktonic taxa. *****

>>>This analysis can be easily done using the reference database for planktonic taxa by Morard et al. (2015). This was indeed the first step in our analysis to check the specificity of the marker and the results are presented in the modified manuscript on page 6 lines 13-27 and page 7 lines 1-4. Because of the generally higher substitution rates among planktonic taxa (de Vargas, C., Zaninetti, L., Hilbrecht, H. & Pawlowski, J. (1997). Phylogeny and rates of molecular evolution of planktonic foraminifera: SSU rDNA sequences compared to the fossil record. Journal of molecular evolution 45, 285–294), we observe (and report on page 4 lines 21-25 in the modified

version) that the genetic marker has a specificity among planktonic taxa which in most cases allows recognition of genetic divergence below the level of morphological species, i.e. better than for benthic taxa.<<<

Detailed comments of Reviewer #1:

*****Page 2, Line 11: "i would not use this term" (for the word "foreign") *****

>>>We have removed the word "foreign". Page 2, line 13 in the modified version.<<<

*****Page 2, Line 22 "planktonic foraminifera are found in the sediment? Please give some more details. is this fraction representing the diversity of the planktonic taxa? Does taphonomy lead to the preservation of only a small fraction?" *****

>>> At this stage of the introduction, we are only providing background information and we are not specifically explaining the case of planktonic foraminifera. We chose to provide the information requested by the reviewer later in the text where we present the planktonic foraminifera at the page 4, lines 11-25 in the modified version.<<<

*****Page 3, Lines 19-21 "I would suggest to place this at the beginning of the section, because this is a major fact and all designed barcodes and their limitations depend mostly on this fact." *****

>>>We have restructured this paragraph accordingly to the reviewer's suggestion. Page 3, line 19-20 in the modified version.<<<

*****Page 4, Line 2. "I suggest to use allothonus rather than foreign" *****

>>> We have made the change thorough the text.<<<

*****Page 4, Lines 9-11. "this can only be done when the barcode is phylogenetically informative, please add a comment on that." *****

>>>We have detailed our argumentation in a restructured paragraph. We have specified that among the planktonic foraminifera, there are segments of the rDNA which are highly informative despite short length. The chosen barcode is one of such segments. Page 4, lines 19-25 in the modified version.<<<

*****Page 4, Line 20. "a dot is missing." *****

>>> Dot added.<<<

*****Page 5, Lines 7-11 "please add a sentence if you sequenced also negative controls or did extractions blanks, if yes please add a comment on the cleanliness of the controls" *****

>>> The negative controls have not been sequenced in the initial study. In order to control the cleanness of the procedure, one library consisted in the extraction of a freshwater foraminifera (So not present in the marine samples). The sequenced library consisted only of reads derived from this foraminifera and no potential contaminant were found, confirming the cleanness of the dataset. These procedural steps were explained in the study of Lecroq et al. (2011) and we refer to this study accordingly. Nonetheless, we added those details on page 6, lines 5-11 in the modified version. <<<

*****Page 5, Lines 14-16 "I don't understand this sentence? Do you mean rank as a taxonomic unit? Please clarify"*****

>>> We have provided additional information to explain the organization of the database. We have specified that the classification scheme is organized hierarchically, just like classical taxonomy. Page 6, lines 15-23 in the modified document.<<<

*****Page 5, Lines 19-24 "how do you differentiate the genetic types? How do you differentiate genetic types from cryptic types? Do you use a threshold/cut-off (% of sequence similarity) to identify cryptic types and group them to known sequences? How do you differentiate authentic types from possible sequencing or PCR errors?" *****

>>> We here reply successively to the answers of the reviewer:

- We differentiate the genetic types in the plankton using phylogenetic inferences and/or automated delimitation methods to delineate clusters of sequences which are then compared to biogeographic and ecological data (See Morard et al. (2015) and references herein). We provide those details on Page 6, lines 19-23 of the modified document.

- Genetic types or cryptic species are equivalent. We specify it on page 6 line 27 of the modified document.

- We use a phylogenetic approach to group sequences into genetic types in the present study (Page 8, Lines 1-16 in the modified documents) to aggregates the e-ribotypes into cryptic species. We use a phylogenetic approach because no unique threshold exists to delineate cryptic species in planktonic foraminifera (Morard et al., 2016). In doing this, we aggregate variants (genuine and potential artefacts) into single units.<<<

*****Page 6, Lines 11-13. "an abundance of ten only occurring once in a dataset is a very small number. I would suggest that this is only valid for sequence types which have a 100% match to your well-curated references.**
I think it is always more important that sequence types occur indepently in two PCR reactions. Do you mean by samples your sediment samples or also the replicates of the PCR?" ***

>>> The majority of the sequence motives we analyzed were found in multiple samples and thus were generated by independent PCR. The referee is right to point out that we also include in our analysis sequences that only occur in one PCR product (albeit at least 10 times). These sequences collectively represent 7,6% of the diversity and 1,17 % of the volume of the analyzed dataset. It would have indeed been advisable to use only sequences that were replicated among independent PCR reactions. However, our sampling coverage is far from global so that it is conceivable that some sequences were originally indeed only present in one sample. This is why we decided to keep them in the analysis. We believe the retained sequences are unlikely to represent sequencing errors, because they were retained due to high similarity to the reference database. The referee agrees that this is a valid approach but indicates that a more safe approach would be to use only motives with 100% similarity. This is certainly correct and would be important if the purpose of the study would be to describe or characterize the patterns of eDNA. However, we only use the data to test if the patterns are congruent with large scale plankton macroecology. In future studies, the best approach would be to increase the level of replication and only retain sequences occurring in independent PCR reactions. Since the production of the dataset we use, sequencing costs have dropped hugely and such replication is now a standard.<<<

*****Page 7, Lines 20-24. "13 genotypes were then new genotypes, which were detected by phylogenetic clustering? What do you mean by known genotypes? please clarify" *****

>>>Known genotypes as those that have been described from material obtained from single-cell DNA extraction on living foraminifera. We have rewritten this section to make it clearer. Page 9, lines 8-10 in the modified version.<<<

*****Page 8, Line 5. "please provide the absolute numbers" *****

>>>We have added histograms to the figure 1 to show the absolute numbers of reads obtained for each sample. <<<

*****Page 8, Lines 11-15. "I would rather write that the abundance of planktonic reads seemed not to be related to the latitutde or depth of the sample location, as planktonic types were found at different locations and in different depth showing no significant trends. I also think that you might need more sample localities to investigate such questions at global scale. " *****

>>>We have rewritten this section, however we kept the interpretation that the relative amount of planktonic eDNA does not seem to decrease with depth, because it represents a specific hypothesis on DNA preservation (less DNA preserved at deeper sites due to longer time for degradation in the water column during sinking). Page 9, lines 19-26 in the modified version.<<<

*****Page 10, Lines 1-2. "could you please clarify: what do you mean with fossil record, you mean morphologically identified taxa? How can it be "identical" and "with only a partial coverage" at the same time? that means both records eDNA and morphology do not represent or only partially represent the real planktonic diversity?" *****

>>>To avoid any confusion, we have changed "identical" to "similar" in the paper. The fossil record refers here to the species identified by morphology. We have also better explained census counts at the end of the method section after a request of the second reviewer. Page 8, lines 21-23 in the modified document.<<<

*****Page 10, Line 8. "please explain, what do you mean with three taxonomic levels? I think you mean genotype, ribotype and morphospecies, but in my opinion this are no taxonomic levels. As all of the three can be assigned e.g. to species level. This are rather taxonomic units or groups." *****

>>> In this particular sentence we highlight the fact that the diversity gradient is present among taxa at all three hierarchical levels. Page 11, line 24 in the modified version.<<<

*****Page 11, Lines 19-21. "it is not only a matter of abundance but also a matter of primer specificity to the benthic and planktonic species. Please comment on this. Please provide more information on the primer specificity between benthic and planktonic taxa." *****

>>> This discussion is important and is present later in the manuscript. At this point in the text, we prefer to restrict our discussion to the variation of planktonic reads among the samples, which are equally affected by the primer biases. However, we then expand on the impact of the primer biases later in the text p13, lines 3 to 18 in the modified version.<<<

*****Page 12, Lines 5-7 "they were also not identified via phylogenetic analyses? Are these species only well-represented in the database or are they also abundant as fossils in your data set?" *****

>>> These species are both abundant in the fossil record and well represented in our database. we have rewritten this part of the text to make it clearer. Page 14, line 4 in the modified version.<<<

*****Page 12, Line 24 "if you propose primer mismatch as the major bias in your analyses, could you please provide a short notice on the specificity of the primer. Please check for possible annealing preferences of your primers. I think this could be easily done, if you have a well-curated database. Software like ecoPCR could be helpful for such a test see Ficetola et al 2010" *****

>>> In the case of foraminifera, the number of sequences affected is small so we could carry out this analysis manually. The number of mismatches between the common species and the primer (up to 5 mismatches for one of the two primers for each of the two species) and is reported on page 14, lines 8-10 in the modified version.<<<

*****Page 13, Lines 19-21 "Do you argue specifically for foraminifera records or generally for ancient sedimentary DNA? Maybe you should include to very recent papers on marine sedimentary DNA in your discussion: Solomon et al. 2016 DOI 10.1007/s13205-016-0482-y and Kirkpatrick et al. 2016 doi:10.1130/G37933.1*****
"

>>>Here we meant ancient sedimentary record in general. We have slightly modified the first sentence of the conclusion to clarify that we talk about the ancient sedimentary DNA in general (Page 15, line 15 in the modified document). We have integrated the study of Kirkpatrick et al. (2016) in the introduction of the paper (Page 3, lines 12-13 in the modified document). The paper of Solomon et al. (2016) is about enriching sediment sample in bacteria to further isolate their DNA. This has nothing to do with our work, therefore we did not integrated it. <<<

*****Page 13, Line 22. "what do you mean with taxonomic bias? different preservation dependent on different taxa? Primer bias...?Please clarify." *****

>>> Here we mean preferential preservation that would induce taxonomic bias. We now specify it in the sentence page 15 line 23 in the modified version.<<<

*****Page 21 line 5. "Does this mean that not all samples were replicated? What was the reason for this?** " *******

This comment is directly related to the following comment and both are addressed together:

***** "do the replicates show also similarities in the genetic types obtained? Or are there large differences? Are these replicates PCR replicates of the same DNA extract? " *****

>>> Not all the samples have been replicated but we have to keep in mind that these data are now more than 6 years old and that the standards in metabarcoding have drastically evolved since then. At the time these data were generated, the access to sequencing technology was limited and the costs prohibitive, limiting the possibility of full replication, forcing the authors to make choices between replication and sequencing depth.

The replicates are DNA isolates extracted from subsamples of sediments recovered from the same deployment (now quoted in the caption of the figure). Discrepancies can exist between replicates (see Lejzerowicz, F., Esling, P. & Pawlowski, J. (2014). Patchiness of deep-sea benthic Foraminifera across the Southern Ocean: Insights from high-throughput DNA sequencing. Deep Sea Research Part II: Topical Studies in Oceanography. Elsevier 108, 17–26.). However these are lower than the inter region comparisons. Page 23, lines 3-11.<<<

*****Page 23 line1. "what do you mean with geneotypes here? Do you refer to e-ribotypes or genetic types?" *****

>>>Here genotypes means genetic types. We have made the change in the text to avoid confusion. Page 25, line 1 in the modified version. <<<

*****Page 25, line 2. "Please explain the mid-domain effect in the manuscript, it has never been mentioned before." *****

>>>This was an oversight on our side. We have replaced "mid-domain effect" by "latitudinal gradient of diversity" which is used in the text and appropriate to describe the pattern. Page 27, line 4 in the modified version. <<<

*****Page 28, line 2. "verstehe ich nicht? dass die grünen clustern".*****

>>>The green stars are clustering because they represent comparisons of sampels from different climatic regions, where the communities are dissimilar, and thus have a low score in pairwise comparison, both for fossil and eDNA assemblages. Page 28.<<<

Reviewer #2

*****This paper explores the potential of using eDNA from sediments to infer plankton community structure. The authors analyze already available foraminifera sequence data to address an ecological relevant and timely question. The idea of using the aDNA preserved in sediments as an archive to explore planktonic biogeographic patterns is interesting; however, as the authors focus only on planktonic foraminifera, I do not think that the authors can conclude that the approach is valid for all planktonic taxa, and the title and some conclusions should then be accordingly revised. *****

>>> The only way to prove that planktonic signature is preserved in sediments is to use a group for which fossil record exists, a curated reference database is available and where a short barcode have enough resolution to allow assignation of environmental read at the species level for direct comparison. To our knowledge, planktonic foraminifera is the only group meeting these criteria. In our opinion we do not conclude in the paper that the results we observe apply to all plankton. We avertedly used the word "hypothesize" in the conclusion to propose that if the conservation of the ecological patterns is true for foraminifera it should be the case for the rest of the pelagic community because we have no reason to believe that the transfer of organic matter from the top to the bottom of the ocean acts differently depending on taxonomic group. We hope that our study will motivate other studies to exploit this potential source of information still locked into the sedimentary archive.<<<

*****Overall, the manuscript is well written and the methodology description is detailed and precise, although there are some important concerns that should be clarified or**

revised before publication. Furthermore, the results do not so strongly support the main conclusion, as eDNA and fossil record only produce somewhat similar patterns. The only clear reproducible pattern is the separation between high and low latitude samples. A major issue relates to the large variation in the number of retained reads per sample after filtering (48 to 124,355). ***

>>> We have discussed this point in our paper (Page 13, line 3-18 in the modified version). It is actually remarkable that despite the large variation in the number of reads among samples, it is still possible to recover the structure of the community. The Caribbean and Japan samples have respectively the lowest and highest amounts of reads on average, and are found clustered together when comparing the relative proportion of their community, regardless to the taxonomic level that is considered (Fig. 4d-f). In addition, there are clear patterns which are reproducible at the regional scale, which is a strong result since it shows that the preservation of planktonic eDNA in sediments is not random.<<<<

*****The authors conduct all the analyses without subsampling to the lowest number of reads, which undoubtedly bias community comparisons and diversity estimates. *****

>>>We have considered all reads in our primary analysis of the dataset, and these data are available in the supplementary information 2. We have not identified a higher diversity in the reads in low abundance compared to the dataset we retained. We have concluded that those reads were variants of abundant reads rather than genuine sequences. Their inclusion in the final dataset would have rendered its analysis more difficult, but not more complete. We point out that reviewer #1 made the exact opposite comment on our strategy, that retaining e-ribotype with an abundance of 10 only is a too low number. This signifies that there is no clear consensus on the strategy to follow and that the strategy we chose is a balanced compromise between two opposite philosophies. <<<

*****As some samples have an extremely low number of reads belonging to planktonic foraminifera, I suggest excluding those samples and reanalyze the dataset equalizing the number of reads per sample. *****

>>> This is an important issue and we have therefore considered the data from several points of view, successively reducing the effect of sequencing depth. We have done an analysis including the total number of reads recovered, then analysed the relative proportion of the reads only, which removes the information on read number and finally an analysis of diversity which removes the information on read number in a different way. These analyses are showed in figure 4, where we have performed NMDS successively on the dataset using Bray-curtis indices on the absolute number and relative proportions, and finally the Dice indices that consider only presence/absence data. As said above, it is remarkable that despite the low number of reads retrieved in some samples it was possible to retrieve the community structure (relative proportions). We have no reason to believe that re-analyzing the dataset by doing another form of equalization of sequencing depth would bring a different conclusion. <<<

*****In addition, the authors should clarify how they analyzed data from census counts of microfossils (details of sampling, number of individuals per sample, normalization, etc.) as they use these data to validate the use of eDNA from sediments to infer planktonic foraminifera community structure. *****

>>> We have provided further information to contextualize the data on census counts. Page 8, lines 21-23 in the modified document <<<

*****Another suggestion is to exclude all the sequences belonging to small (<150 microns) foraminifera from the eDNA dataset, as the census counts are not including this fraction of the community. *****

>>> Some microperforate species also occur in the larger fraction, but in different proportions than in the small fraction (Brummer et al. (1986), quoted in the text). Removing them from the data would in our opinion decrease the number of taxa in the analyses too much.<<<

Specific comments

*****Title. I suggest revising the title as the authors focus only on planktonic foraminifera. *****

>>>Whilst we maintain that it is extremely unlikely that the results we obtained apply to the plankton at large, we have to acknowledge that the results are only based on planktonic foraminifera and so we have revised the title accordingly.<<<

*****Page 2, lines 13-17. These final statements are too strong. *****

>>> We have hedged these statements along the lines described above with regard to the likelihood that the results we obtain apply to plankton at large. Page 2, lines 16-21 <<<

*****Page 2, lines 26-27. Please explain why high concentrations of DNA in sediments indicates that part derives from planktonic/pelagic organisms. *****

>>> Here, we are just stating that the vast majority of sedimentary DNA is extracellular, we do not say that this DNA originate from plankton. Which is why we state in the next sentence that the DNA "survives" even after the death of an organism, therefore at least a part of this extracellular pool derives from the organisms inhabiting the water column. In addition there are good evidences that part of the DNA present in the abyss derive from phototrophic taxa that occur in plankton only (Pawlowski, J., Christen, R., Lecroq, B., Bachar, D., Shahbazkia, H. R., Amaral-Zettler, L. & Guillou, L. (2011). Eukaryotic richness in the abyss: Insights from pyrotag sequencing. PLoS ONE 6).<<<

*****Page 3, lines 25-26 and Page 4, lines 1-2. The authors should keep in mind throughout the manuscript that they are focusing on a taxonomic group that seems to be particularly suited for validating their hypothesis and thus extrapolating to all planktonic or even the entire spectrum of pelagic organisms is not straightforward. *****

>>>As said in the reply to the general comment, we use foraminifera only to demonstrate that the imprint of planktonic DNA is preserved in sediments. We only hypothesize that if it is true for planktonic foraminifera, it should be the same for other organisms that leave no fossil record.<<<

*****Page 7, lines 10-11. Please, provide details on the census counts dataset. *****

>>>We have provided further details. Page 8, lines 21-23 of the modified document.<<<

*****Page 7, line 18. Change "ascribed" to "ascribe". *****

>>>We have made the change. Page 9, line 2 in the modified document.<<<

*****Page 9, lines 21-23. I do not see that the eDNA dataset reproduce separation between Caribbean and Japan samples (only when using absolute numbers of reads due to several order of magnitude difference in number of reads). I do not find appropriate the analyses conducted with absolute read numbers considering the extreme differences among samples. Please consider excluding panels a-c in figure 4. *****

>>> We disagree with the interpretation of the figure 4 by the reviewer. We have modified the figure 4 to make the pattern more obvious. We have reduced the size of the data points and colored the area they are covering (See the modified version of the manuscript page 26). The Caribbean and Japan areas (red and purple) are perfectly disjoint with the Dice index and using E-ribotype (Fig. 4i) and Genotypes (Fig. 4j) taxonomic levels. They are also disjoint with the Bray-Curtis index when using the absolute number (Fig. 4a and 4b). It is true that they are partially overlapping at the morphospecies level, but it is caused by a single sample. We consider that the panel a-c are useful because they show that the patterns observed are reproducible at the regional level when considering the "raw data" and this is one of the messages of our paper.<<<

*****Page 10, line 1. I do not see that the patterns are identical, please revise. *****

>>>We have replaced "identical" with "similar" to moderate our statement. Page 11, line 21 in the modified version.<<<

*****Page 10, lines 12-13. It looks from data in figure 6 that the correlation is not significant. Moreover, many data are well above or below the 1:1 line. Please provide p-value. *****

>>> We have added the r^2 and p value on the figure 6. The p value is highly significant in both cases. We have also added the p value on the correlation shown on the figure 5 to be consistent. The reviewer is right that even if the correlations are statistically supported for both graph, the relationships are not 1:1.<<<

*****Page 11, lines 4-5. As already commented, the patterns are not so consistent. *****

>>>As mentioned above, we prefer to stick with our interpretation of the dataset and we consider that our evidences are strong enough to support the statement. <<<

*****Page 12, line 5. I suggest changing "discover" to "detection". *****

>>>We have made the change. Page 14, line 3 in the modified version.<<<

*****Page 12, lines 5-7. Are these species present in the fossil record? Please, add a comment on that. *****

>>>Yes, the species we are mentioning are present in the fossil record, we have change the sentence to make it clearer. Page 14, line 4 in the modified version.<<<

*****Page 12, lines 18-25. If microperforate species are not represented in the fossil record, I suggest excluding these sequences from the eDNA archive in order to compare both datasets. *****

>>> The microperforates are present in the fossil record but they have different relative abundance depending of the size fraction which is considered. In the large fraction, used in micropaleontology, they have a modest abundance, whilst in the small size fraction, which is rarely used, they are dominant. We provide this information in the text “Alternatively, the higher abundance of reads assigned to microperforate taxa could represent a genuine pattern, questioning the representativeness of census counts of fossil foraminifera, which ignore specimens smaller than 150 μm (Kucera et al., 2005). Microperforate species tend to be small and are therefore disproportionately abundant in the size fraction smaller than 150 μm (Brummer et al., 1986)”.<<<

*****Page 13, line 25. I suggest smoothing this statement. *****

>>>Here we are only making a hypothesis, however we have slightly modified the statement to moderate its strength. Page 15, Lines 22-24 in the modified version.<<<

Planktonic foraminifera-derived environmental DNA extracted from abyssal sediments preserves patterns of plankton macroecology

Raphael Morard[1], Franck Lejzerowicz[2], Kate F. Darling[3,4], Béatrice Lecroq-Bennet[5], Mikkel Winther Pedersen[6], Ludovic Orlando[6], Jan Pawlowski[2], Stefan Mulitza[1], Colomban de Vargas[7,8] and Michal Kucera[1]

[1]MARUM Center for Marine Environmental Sciences, University of Bremen, Leobener Strasse, 28359 Bremen, Germany

[2]Department of Genetics and Evolution, University of Geneva, Geneva, CH1211, Switzerland

[3]School of GeoSciences, University of Edinburgh, Edinburgh EH9 3FE, UK

[4]School of Geography and GeoSciences, University of St Andrews, Fife KY16 9AL, UK

[5]Okinawa Institute of Sciences and Technology Graduate University, 1919-1 Tancha, Onna-son 904-0495, Japan

[6]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, Copenhagen 1350K, Denmark

[7]Centre National de la Recherche Scientifique, UMR 7144, EPEP, Station Biologique de Roscoff, 29680 Roscoff, France

[8]Sorbonne Universités, UPMC Univ Paris 06, UMR 7144, Station Biologique de Roscoff, 29680 Roscoff, France

Keywords: Planktonic foraminifera, metabarcoding, DNA preservation, eDNA, protists.

Corresponding author: Raphaël Morard, MARUM Center for Marine Environmental Sciences, University of Bremen, Leobener Strasse, 28359 Bremen, Germany,Phone: +49 (0) 421 218 -65973, Fax: +49 (0) 421 218 – 9865974, email: rmorard@marum.de.

Abstract

Deep-sea sediments constitute a unique archive of ocean change, fueled by a permanent rain of mineral and organic remains from the surface ocean. Until now, paleo-ecological analyses of this archive have been mostly based on information from taxa leaving fossils. In theory, environmental DNA (*e*DNA) in the sediment has the potential to provide information on non-fossilized taxa, allowing more comprehensive interpretations of the fossil record. Yet, the process controlling the transport and deposition of *e*DNA onto the sediment and the extent to which it preserves the features of past oceanic biota remains unknown. Planktonic foraminifera are the ideal taxa to allow an assessment of the *e*DNA signal modification during deposition because their fossils are well preserved in the sediment and their morphological taxonomy is documented by DNA barcodes. Specifically, we re-analyze foraminiferal-specific metabarcodes from 31 deep-sea sediment samples, which were shown to contain a small fraction of sequences from planktonic foraminifera. We confirm that the largest portion of the metabarcode originates from benthic bottom-dwelling foraminifera, representing the *in-situ* community, but a small portion (< 10%) of the metabarcodes can be unambiguously assigned to planktonic taxa. These organisms live exclusively in the surface ocean and the recovered barcodes thus represent an allochthonous ~~"foreign"~~ component deposited with the rain of organic remains from the surface ocean. We take advantage of the planktonic foraminifera portion of the metabarcodes to establish to what extent the structure of the surface ocean biota is preserved in sedimentary *e*DNA. We show that ~~the~~ planktonic foraminifera DNA is preserved in a range of marine sediment types, the composition of the recovered *e*DNA metabarcode is replicable and that both the similarity structure and the diversity pattern ~~of the pelagic environment~~ are preserved. If these observations apply to the rest of the pelagic community, it would ~~These observations~~ pave the way for surveys of seafloor sedimentary eDNA covering the entire spectrum of pelagic biodiversity and its interaction with the climatic history of the oceans.

1 Introduction

With over two thirds of the planet covered by oceans, ~~deep~~ deep-sea deposits form the most extensive archive of the Earth's recent history. These deposits preserve mineralized skeletons of marine nano- and micro-plankton, which serve as an record of past climate (e. g., Hillaire-Marcel and de Vernal, 2007) and biodiversity (Yasuhara et al., 2015) changes. However, planktonic groups leaving fossilized remains only represent a small fraction of the marine diversity (de Vargas et al., 2015). In theory, environmental DNA (*e*DNA) buried in marine sediments can provide information on the history of marine organisms that do not produce fossils (Pedersen et al., 2015). Deep sea sediments are rich in

DNA with 0.31 ± 0.18 g of DNA per m² in the surface layer, and more than 90% of this DNA is extracellular (Dell'Anno and Danovaro, 2005). This means that DNA from many organisms is preserved after their death in the sediment and the high abundance of the DNA indicate that at least a part of the DNA pool derives from organisms living in the water column above the sediment (Lejzerowicz et al., 2013). Part of this DNA pool remains preserved in ancient sediments, and can be extracted and analyzed using metabarcoding to reveal the molecular diversity of past ecosystems (Lejzerowicz et al., 2013; Pawłowska et al., 2014). This potential has been demonstrated in a range of other depositional environments, such as cave sediments, lake and ice cores where the dynamics of plant and animal communities could be followed over 50 ka (Pedersen et al., 2015).

In marine sediments, the presence of *e*DNA sequences has been reported from organic-rich layers in the Mediterranean dating back to 217 ka (Coolen and Overmann, 2007) and 125 ka (Boere et al., 2011), in sediments covering the last 11.4 ka in the Black Sea (Coolen et al., 2013), and in up to 32.5 ka old deposits in the Atlantic (Lejzerowicz et al., 2013; Pawłowska et al., 2014). Recently, Kirkpatrick et al. (2016) showed that the abundance of planktonic DNA was decreasing within 100-200 ka in sediments of the Bering sea but traces were still detected in sediments up to 1.4 Ma. Direct comparison with co-occurring fossils showed that the sequenced *e*DNA pool exceeds the taxonomic spectrum of the fossils, but many of the taxa preserved as fossils were not identified in the *e*DNA (Pawłowska et al., 2014; Pedersen et al., 2013). This raises the question of how well the sedimentary DNA pool reflects the autochthonous (in situ origin) or allochthonous (~~foreign~~ external origin) community composition, whether there is any differential DNA preservation across taxa and whether the metabarcode marker selected is fully representative of the entire taxonomical diversity, regardless of its origin. The extensive fragmentation of *e*DNA (Pedersen et al., 2015) makes incompatible the amplification of sequences longer than ~100 bp, preventing the access to long and informative barcodes.

The primary difficulty in the analysis of the sedimentary DNA pool is to separate the local and ~~foreign~~ allochtonous origin of the sequenced material (Torti et al., 2015). This can be done with certainty only when the ecological origin of the sequenced eDNA is unambiguously resolved. Potential bias could arise from a range of factors including preferential amplification (Taberlet et al., 2012), inconsistent taxonomic resolution of the sequenced barcodes (Pawlowski et al., 2012) and insufficient coverage of the barcode reference database (Pawlowski et al., 2014b). ~~In~~

~~addition, the extensive fragmentation of *e*DNA (Pedersen et al., 2015) makes incompatible the amplification of sequences longer than ~100 bp, preventing the alleviation of bias by targeting longer, more informative barcodes.~~

Here we take advantage of the possibility to unambiguously ascribe sequences of foraminifera to benthic and planktonic lineages. By analyzing the planktonic portion of foraminiferal metabarcodes from deep sea sediments, we provide evidence that the structure and diversity of surface ocean communities is preserved in *e*DNA molecules and that the preservation is not limited to specific depositional environments. We focus our analysis on the Foraminifera because of access to highly resolving short barcodes (Pawlowski and Lecroq, 2010) and the availability of a taxonomically well resolved barcode database for the planktonic taxa (Morard et al., 2015). It allows the unambiguous separation of the benthic ~~(~~, autochthonous~~)~~ , component of the dataset from its planktonic ~~(~~, allochthonous~~)~~ , component.

Foraminifera are single-cell eukaryotes (protists) belonging to the phylum Rhizaria (Adl et al., 2012). Most Foraminifera lineages occupy benthic ecological niches. Their ~ 5000 morphospecies inhabit the bottom of shallow coastal environment to deep abyssal plains. In contrast,- the planktonic lineages only include 50 morphospecies, living mostly in the photic part of the water column. They are found from tropical to polar water masses and spend their entire life cycle in the plankton (Hemleben et al., 1989). After their death, planktonic foraminifera sink to the bottom of the ocean where they are found in the calcareous ooze, ranging from ~1 to 4,5 km water depth and distributed from low to high latitude (Dutkiewicz et al., 2016). The fossil planktonic assemblages are preserved without taxonomic bias above the lysocline and become increasingly affected by the preferential dissolution of thin-shell species below this limit (Berger and Parker, 1970). Foraminifera are known for their unusually high rate of evolution (de Vargas et al., 1997) resulting in highly resolving barcodes even in fragments shorter than ~100 bp thus allowing unambiguous species identification with relatively short barcodes (Pawlowski and Lecroq, 2010). In addition planktonic foraminifera harbor considerable cryptic diversity (Darling and Wade, 2008; Morard et al., 2016), which offers an additional layer of taxonomic information that can be exploited in eDNA studies. Therefore, planktonic foraminifera possess barcodes with resolution that is equal or higher than their benthic counterparts. This facilitates the taxonomic identification of short, potentially degraded, *e*DNA sequences.

~~column and spend their entire life cycle in the plankton (Hemleben et al., 1989). Foraminifera are known for their unusually high rate of evolution (de Vargas et al., 1997) resulting in highly resolving barcodes even in fragments~~

~~shorter than ~100 bp (Pawlowski and Lecroq, 2010). This circumstance facilitates identification of short, potentially degraded, extracellular *e*DNA sequences. In addition planktonic foraminifera harbor considerable cryptic diversity (Darling and Wade, 2008; Morard et al., 2016), which offers an additional layer of taxonomic information that can be exploited in eDNA studies.~~

In the present study we perform new analysis on *e*DNA libraries generated by Lecroq *et al.* (2011), which comprise metabarcodes from 31 abyssal sediment samples containing ~ 78 million foraminiferal sequences derived from the 37f foraminiferal specific barcode of the 18S rDNA. The ~~main part~~major portion (>99%) of the sequences could be assigned to benthic taxa and their composition was analyzed to unravel the patterns of benthic diversity on the sea floor. However, a tiny portion of the barcodes (<1 %) could be assigned to planktonic foraminifera. These sequences represent eDNA exported to the seafloor from the plankton. With the recent development of the *Planktonic Foraminifera Ribosomal Reference* Database (PFR², (Morard et al., 2015)), the environmental sequences belonging to planktonic foraminifera in the *e*DNA libraries generated by Lecroq *et al.* (2011) can now be for the first time thoroughly analyzed and assigned to the morphological and cryptic species ~~taxonomic~~ levels

The extensive knowledge on the distribution and abundance of planktonic foraminiferal shells in surface sediments (Kucera et al., 2005), enabled the *e*DNA data to be directly compared with data derived from classical taxonomy. We thus assess to what extent the eDNA originating from plankton is representative of the source community which is an essential prerequisite for interpretation of the eDNA archive in the sediment.

2 Material and Methods

The 31 surface sediment samples analyzed were taken at water depths ranging from 1,745 to 5,338 m and cover sediment types from calcareous ooze in the Caribbean Sea to fine clastic sediments in the Arctic Ocean (Fig. 1a, Supplement 1). All analyses are based on the Illumina Solexa GAII datasets generated by Lecroq et al. (2011) and registered at the NCBI's Short Read Archive under the BioProject number PRJEB2682. The original sequencing data include 78,613,888 reads covering the 36 positions starting 3' of the "GACAG" motif delimitating the foraminifera-specific hypervariable region 37f region (Pawlowski and Lecroq, 2010).

We used the unique sequences obtained for each library in Lecroq et al. (2011) after the strict dereplication step and the removal of singletons associated with only one read occurrence in a library. For each DNA library, we parsed sequencing reads passing the default base calling of GAPipeline v 1.0 and reads showing a single base quality or averaged base qualities inferior to 10 and 20, respectively as well as sequencing reads presenting ambiguities (N) or homopolymers over 30 positions. This resulted in a total of 204,704 unique and filtered 36 bp-long sequences representing 39,210,426 reads (Supplement 1, Fig 1b). During the generation of the data, one sample was used as a control to check for potential cross-contamination. This sample consisted in the DNA extract of a single cultured species: *Reticulomyxa filosa*. The sequencing of this sample produced 2,416,756 reads, corresponding to 1,689 dereplicated tags with at least 2 reads per tag. After filtering and clustering, we recovered only one OTU, which was identical to the 37f hypervariable sequence of *R. filosa* previously obtained by using classical Sanger technology, thus showing the absence of cross contamination.

~~reads.~~

We compared the retained reads to the *Planktonic Foraminifera Ribosomal Reference* database (PFR², (Morard et al., 2015)), which represents a compilation of 3,322 curated partial SSU rDNA sequences of planktonic foraminifera groups associated with a 6-ranks taxonomy. The ranks reflect taxonomic units and are organized into a hierarchal framework, with the basal ranks being the coarsest units and the terminal ranks corresponding the finest taxonomic levels. The first three basal ranks correspond to the level of assignation comparable to that achievable using morphological data, and is thus analogous to fossil data. The three terminal levels correspond to the molecular taxonomy accessible using molecular data only. The PFR² taxonomic framework derives from single-cell genetic studies where the molecular taxonomy (definition of genetic types) was based on phylogenetic inferences and or/automatic delimitation methods. The delimited cluster of sequences were then compared to ecological and biogeographical data to validate their status as genuine biological species (see Morard et al., 2015 and references herein). Of the 3,322 sequences available in PFR², ~~Of these,~~ 2,418 sequences covered the fragment of the region 37f. These sequences were downloaded from the PFR² database (http://pfr2.sb-roscoff.fr/) and trimmed to the 36-nt fragment corresponding to the environmental sequences, which resulted in a total of 463 unique homologous reference sequences (Supplement 2). Initially, we evaluated the taxonomic resolution of the 36 nt barcoding region and found that it was variable enough to discriminate the genetic types (equivalent to cryptic species) within morphological

species of almost all planktonic foraminifera taxa referenced in *PFR²*. We observed a lack of genetic resolution (different taxonomic entities yielding identical barcodes) ~~taxonomic conflicts~~ for only two species pairs belonging to *Globorotalia* (*tumida* and *ungulata*) and *Globigerinella* (*calida* and *siphonifera*) and three pairs of genetic types among *Globorotalia truncatulinoides* (type III and IV), *Pulleniatina obliquiloculata* (types I and II) and *Globigerinita glutinata* (types III and IV).

We then individually aligned the 4,466 to 27,578 unique sequences obtained for each of the 31 samples against the 461 reference sequences using the Needleman-Wunsch global sequence alignment algorithm (Needleman and Wunsch, 1970), to separate the portion of the dataset belonging to the planktonic foraminifera (allochthonous origin) from the portion belonging to the benthic foraminifera (autochthonous origin). Pairwise genetic distances were calculated as the number of differences (counting successive indels and terminal gaps as one difference), and an iterative clustering of the unique environmental sequences with the reference sequences was performed, allowing 1, 2, 3, 5 and 10 differences as thresholds for the average linkage algorithm. We then extracted all environmental sequences found within each cluster containing a planktonic reference sequence in an iterative manner, by screening from the most stringent (1 difference threshold) to the most permissive (10 differences threshold) clusters. As a post hoc verification, we compared these sequences with the extensive benthic foraminifera sequence database used in Pawlowski *et al.* (2014a) together with the sequences of the *Protist Ribosomal Reference Database* (*PR²*, Version based on Release 203 of Genbank, (Guillou et al., 2013)) and additional undescribed benthic specimen sequences to ensure that the extracted sequences do not belong to benthic foraminifera. No match was found. We assigned to each extracted environmental sequence the taxonomy of the planktonic reference sequences in the cluster. Finally, we retained only the sequences occurring in at least 2 samples or having a minimal abundance of 10 for downstream analysis. The final product was then considered an individual e-ribotype (Supplement 3). E-ribotypes are unique environmental sequences (not cluster) originating from planktonic foraminifera and thus transferred from surface ocean to the bottom (allochthonous origin). The relative proportions between e-ribotypes (planktonic reads) and the benthic reads of each sample are shown on Fig. 1B. We calculated the rarefaction curves of each individual samples using PAST v 2.17 (Hammer et al., 2001) to estimate to what degree the full taxonomic spectrum of each sample was recovered by *e*DNA (Fig. 2).

Genuine sequences of planktonic foraminifera representing species not yet registered in the reference database may have been omitted. We therefore structured our analyses to account for the detection of possibly unknown genetic types. To this end, we used the phylogenetic signal contained in the 36bp-reads to build a taxonomic framework within each morphospecies. In contrast to strict annotation approaches using arbitrary similarity thresholds, a phylogenetic approach can identify novel genetic type, not represented in the reference comparative database. The retained e-ribotypes were automatically aligned using MAFFT v.7 (Katoh and Standley, 2013), with reference sequences of the complete 37f region. The complete 37f region was used at this step instead of the 36-bp fragment to avoid possible read alignment shifts caused by artificial mismatches with trimmed 36-bp sequences during the assignment process. A single alignment was produced per morphospecies. For each resulting alignment, a phylogenetic tree was inferred using PhyML (Guindon et al., 2010) implemented in SEAVIEW 4 (Gouy et al., 2010) with default option using aLRT for branch support estimation. The resulting trees were visualized with ITOL (Letunic and Bork, 2011) and all visually distinct clusters were considered as unique genotypes (Supplement 4). The reads clustering with reference sequences were assigned at the genetic type level, the sequences clustering without a close reference received an artificial genetic type attribution (Supplement 3). These assignments were used to prepare three datasets with different degrees of taxonomic resolution (at the level of e-ribotype, genetic types and morphological species). The occurrences of the defined genetic types in the samples are shown on the Figure 3.

In order to compare the taxonomic richness and structure of the dataset, Non-Metric Distribution Scaling (NMDS) as implemented in PAST v 2.17 (Hammer et al., 2001) was applied on the dataset at the different taxonomic resolution. We used the Dice distance to consider only the presence/absence data and the Bray-Curtis distance to compare absolute and relative abundances of reads among the samples (Fig. 4). To compare the similarity structure and diversity in the samples based on the *e*DNA reads with census counts of microfossils, we used the MARGO database . The census count represent the relative abundance of species observed in a fossil assemblage based on the count of typically 300-500 specimens (Kucera et al., 2005). We calculated the~~(Kucera et al., 2005a) and calculated~~ fossils-based diversity (Shannon-Wiener) and similarity (Dice and Bray-Curtis) matrices using PAST v 2.17 for all surface samples within the regions outlined in Fig 1A (between 6 and 13 per region, Fig. 5 and 6).

3 Results

After quality filtering and collapsing of identical reads into single sequences, the comparison of the entire dataset with reference databases (Supplement 2) allowed to ascribe~~d~~ with certainty 1,373 unique sequence patterns representing 488,291 reads to planktonic foraminifera (Supplement 1, 3). Because we required reads to be present in a minimum of two samples or to show a minimal abundance of 10 in the entire dataset, the retained dataset was reduced to 697 unique sequences of planktonic foraminifera (e-ribotypes), which are representing a total of 486,435 reads (~0.63% of the total dataset, Supplement 1). Diversity was then assessed using a phylogenetic approach and the 697 e-ribotypes were found to represent 37 genotypes (Fig. 3, Supplement 4). Of these, 675 e-ribotypes (representing ~ 99 % of the planktonic reads) were attributed to 24 ~~known~~ genotypes already detected in plankton and assigned to 17 morphological species (Supplement 3, 4). The remaining 22 e-ribotypes clustered into 13 genotypes with no apparent affinities with the genotypes detected in plankton. These e-ribotypes represent only ~0.5% of the planktonic reads.

After this filtering, between 48 (Library #SFA-17) and 124,355 (Library #SFA-15) reads were retained in 28 samples (Supplement 1, Fig. 1, 2), representing between 0.003 and 9.412% of the total foraminifera reads in the libraries from these samples (Fig. ~~1b~~1c). Three Arctic samples did not yield any sequences that could be assigned to planktonic foraminifera (Fig. ~~1b~~1d, 3). The total number of reads per sample is a function of sequencing effort and is therefore not related to initial community density. However, the relative abundance of reads assigned to planktonic foraminifera in the DNA accumulated on the sea floor should reflect the relative proportion of the foraminiferal DNA produced by planktonic communities and the DNA produced by the in-situ benthic community. Whilst the absolute number of planktonic reads varied among the samples and replicates (Fig. 1d), we did indeed observe a higher reproducibility of the relative number of planktonic reads recovered from replicates at the same location (Fig. ~~1b~~1e). The relative abundance of planktonic reads seems unrelated to the latitude or depth of the sample location (Fig. 1c, e). The samples with the highest relative proportions originate from Japan (0.790% to 9,412%) whilst the lowest abundances are observed in the Caribbean samples (0.003 to 0.032%). The high latitude samples (Arctic, North Atlantic and South Atlantic) show relative abundances ranging from 0.011 to 1.204% when excluding samples without planktonic reads. The relation between water depth and relative sequence abundance is not clear (Fig. 1b). It does not seem that the proportion of planktonic DNA reads decreases with increasing depth, suggesting that bentho-pelagic flux exporting planktonic DNA do not weakens compared to the *in situ* community. ~~No obvious correlation is observable between the relative number of planktonic reads and the latitudinal position of the samples. The samples with the highest relative proportions originate from Japan (0.790% to 9,412%) whilst the lowest abundances are observed in the~~

~~Caribbean samples (0.003 to 0.032%). The high latitude samples (Arctic, North Atlantic and South Atlantic) show relative abundances ranging from 0.011 to 1.204% when excluding samples without planktonic reads. In contrast, there appears to be a trend of increasing abundance of planktonic reads with depth (Fig. 1b), suggesting that the organic matter flux delivering planktonic DNA becomes relatively more abundant compared to DNA from in situ community with depth. This trend is relatively weak and influenced by outliers, but it clearly appears to be the case that the proportion of planktonic DNA reads does not decrease with depth.~~

Rarefaction analysis has been used to assess the degree to which the retained planktonic reads cover the diversity they contain (Fig. 2). As expected, the general trend indicates a higher degree of saturation in samples with more reads. For example, samples with the highest number of reads (Japan) had saturated diversity (Fig. 2a), whereas the Caribbean samples representing a similar geographical province but with fewer reads are clearly under-saturated (Fig. 2c). However, we observe that samples from high latitude regions are also saturated (Fig 2b, c), despite having a lower number of reads than the samples from Japan, implying lower diversity.

With respect to the composition of the reads, we observed that e-ribotypes attributed to the microperforate species *Globigerinita glutinata* dominated the dataset (~77% of the reads) and were particularly abundant in subtropical communities (Fig. 3). E-ribotypes of common subtropical species *Orbulina universa*, *Globorotalia menardii*, *Globorotalia hirsuta*, *Hastigerina pelagica Neogloquadrina dutertrei*, *Globigerina falconensis*, *Globigerinella siphonifera, Pulleniatina obliquiloculata, Galitellia vivans* and *Candeina nitida* were found in subtropical samples, whereas e-ribotypes assigned to the species *Globigerinita uvula* and *Neogloboquadrina pachyderma* appeared to dominate subpolar and polar samples. Among these, e-ribotypes belonging to the genotype IV of *N. pachyderma* were mostly found in the Southern Ocean (>99.99%) whereas e-ribotypes of the genotype I were only observed in the subpolar samples from the northern hemisphere. Additionally, different *Globigerina bulloides* e-ribotypes were detected either in subtropical samples (type I) or in subpolar assemblages (type II). Similarly, type II e-ribotypes of *Globigerinita uvula* were found more frequently in subpolar samples from both hemispheres, whereas e-ribotypes of type I were also abundant in low-latitude samples (Fig. 3).

Next, we calculated similarity matrices among samples at the level of morphological species, genotypes and e-ribotypes in order to identify patterns of community structure in the planktonic reads in each sample. Visualization was based on Nonlinear Multi-Dimensional Scaling (NMDS, Fig. 4) with either Dice (Presence/absence) or Bray-

Curtis similarity metrics computed from relative as well as absolute read abundances. Calculation performed on absolute number with Bray-Curtis (Fig 4a-c) showed high reproducibility for samples from the same regions, best expressed at the e-ribotype taxonomic level. The high latitude communities are more similar (Fig 4c), whilst we observe a strong separation between Caribbean and Japan samples, probably reflecting the several order of magnitude of difference in the~~substantially~~ different number of reads between localities (Figs. 1d, 2). Indeed, this difference vanishes when relative proportions are considered (Fig 4d-f). This shows that despite the enormous differences in read numbers between the Caribbean (48 to 212 reads per sample) and Japan (3,630 to 124,355 reads per sample) samples, the relative proportions of the major taxa are the same between these regions (Fig. 3). We also observe a clear separation between low and high latitude samples (Fig 4f) at morphospecies level, which is analogous to the structure represented by fossil assemblages in nearby samples (Fig 4g). Calculation performed on Dice indices (Fig 4i-k) tends to reproduce the same structure as the calculation performed with Bray-Curtis calculated on absolute read number (Fig 4a-c), especially at e-ribotypes level (Fig 4i), but the patterns are noisier ~~(Fig 4a-c)~~. This is most likely due to the different level of taxonomic saturation between the samples (Fig. 2). For example, less than 15 e-ribotype have been observed in every Caribbean samples whilst more than 200 were detected for 5 of the 6 Japan samples, thus maintaining the separation between these regions. Remarkably, despite the different numbers of reads and the associated different level of taxonomic saturation, the recovered pattern of taxonomic composition of the reads is so strong that the opposition between the high and low latitude samples, clearly observed with fossil assemblages, remains even in the eDNA assemblages when considering the morphospecies level (Fig 4k-l). The signal in the eDNA data is noisier because only a fraction of the morphospecies have been detected by the eDNA (1 to 11 morphospecies per sample, Supplement 1), whilst 1 to 24 morphospecies are observed in the census counts. This means that the relative proportions of the reads carry enough information to reproduce ~~identical~~ similar patterns between eDNA and fossil record despite only a partial coverage of the morphological diversity (Fig 4f-g).

These observations taken together imply that the relative abundance of planktonic eDNA reads in the sediment samples contains exploitable information at all three taxonomic- (morphological species, genotypes and e-ribotypes) levels. To further explore the diversity patterns implied by eDNA data, we calculated the Shannon-Wiener diversity index within each sample (Fig. 5a). Despite differences in sediment type and sequencing depth, eDNA in the analyzed samples reproduces the latitudinal diversity gradient based on morphospecies abundances in surface sediment samples

(Rutherford et al., 1999). The latitudinal diversity gradient is present at all three taxonomic levels, but is most pronounced at the e-ribotype level (Fig. 5b).

Finally, since census counts of planktonic foraminifera morphospecies in surface sediments are available from the same regions as those analyzed for *e*DNA (Fig. 1a), we assessed whether the e-ribotype abundances reflect the same community turnover pattern as that indicated by fossil assemblages (Fig. 6). To this end, we compared pairwise distances between *e*DNA MOTU assemblages with pairwise distances between fossil assemblages. This comparison reveals that *e*DNA and morphospecies community turnover rates are significantly correlated (Fig. 6), with highest similarity among samples from the same region and lowest similarity among samples from different climatic regimes. This pattern emerges both when relative abundances and presence/absence data are considered. This implies that the proportionality of *e*DNA reads abundance is consistently scaled with the proportionality of plankton flux to the seafloor. The analysis based on relative abundances yields a pattern with highly consistent results for comparisons between climatic zones and more scatter when comparing samples within a region or within one climatic zone. This is likely due to the fact that the eDNA data only cover a part of the morphological diversity of the foraminifera combined with differential distortion of the original abundance signal due to variation in gene copy number (Weber and Pawlowski, 2013) and primer bias (Bradley et al., 2016).

4 Discussion

Here, we provide evidence that *e*DNA originating from planktonic foraminifera is indeed preserved in the DNA pool of abyssal marine sediments irrespective of water depth, geographic region and sediment type. Earlier *e*DNA studies on marine sediments assumed that DNA preservation is proportional to the preservation of organic matter and, thus, prioritized sampling in organic-rich sediment layers (Coolen et al., 2009). Yet, recent experimental research and field studies suggest that the primary structure of DNA molecules is adsorbed to solid particles and molecules preserved in this way may form an archive of extracellular DNA regardless to the organic content of the sediment (Corinaldesi et al., 2007, 2011, 2014; Torti et al., 2015). We also show that the *e*DNA composition consistently reflects the composition of the pelagic planktonic communities from which it was derived (Figs. 5, 6). The high reproducibility of reads diversity (Dice index) and relative abundance (Bray-Curtis index) within a single region (Fig. 4) suggests that

the taphonomic process governing the transfer and preservation of extracellular DNA from surface to bottom ocean are similar at regional scale and do not differentially impact DNA from species within different ecological groups.

Although the number of planktonic foraminifera reads recovered differed by three orders of magnitude between the Caribbean (62 to 212 reads per samples, representing ~0.003 to 0.03 % of the dataset) and Japan (3,620 to 124,355 reads per sample, representing 0.8 to 9.4 % of the dataset), the information recovered was sufficient to unveil the structure of foraminifera communities across the whole range of environments investigated (Figs. 3-6). However, since the taxonomic richness in *e*DNA data increased with sequencing efforts (Fig. 2), the recovery of the full taxonomic diversity requires a certain minimum sequencing effort. From the analyzed dataset, it is not possible to explain the large variation in the numbers of reads ascribed to planktonic foraminifera among regions (Fig. 2). This could represent DNA differential preservation conditions, or an imbalance between flux from the surface, ~~(~~allochthonous~~)~~ community and the abundance of DNA from the ~~autochthonous~~ benthic, autochthonous community. The latter is a likely explanation because the analyzed *e*DNA material was amplified by PCR primers annealing to all foraminiferal sequences (Lecroq et al., 2011). During the PCR, the DNA of planktonic foraminifera might well be outcompeted by the autochthonous DNA of benthic foraminifera, which is potentially more abundant, less damaged and more easily extracted from cells than when tightly absorbed to sediment particles (Ceccherini et al., 2009; Torti et al., 2015). It is noteworthy that the relative proportion of sequence reads may reflect the relative proportion of DNA molecules – but not necessarily that of cells – as shown in the case of a mock foraminiferal DNA community amplified using foraminiferal-specific primers (Esling et al. 2015)

Consistent with earlier studies (Capo et al., 2015; Lejzerowicz et al., 2013; Pawłowska et al., 2014; Pedersen et al., 2013), the taxonomic diversity revealed by the analyzed *e*DNA barcodes overlaps only partly with the diversity based on fossils present in the sediment. One part of the observed difference could be ascribed to the limited coverage of the reference database. Because of the way we assigned reads to planktonic foraminifera, we cannot assess the portion of the planktonic foraminifera diversity not represented in the reference database, although all major planktonic foraminifera taxa making >90% of tests larger than 150 µm are present in the reference database (Morard et al., 2015). We note, however, that our method allowed the discovery of unknown e-ribotypes clustering within e-ribotypes of known morphological species. Despite the discovery of the new e-ribotypes, the vast majority (99 %) of the retained

reads could be associated with known genetic types. This exemplifies that the overlap of the *e*DNA reads library is large for well-studied taxa.

However, there might be a PCR bias that impairs the ~~discovery~~ detection of some species. Indeed, none of the recovered barcodes could be attributed to four common species in the fossil record and well represented in the reference database: *Globorotalia truncatulinoides*, *Turborotalita quinqueloba*, *Trilobatus sacculifer* and *Globigerinoides ruber*. This observation is consistent with preferential PCR amplification. The rDNA of planktonic foraminifera is characterized by high and variable substitution rates (de Vargas et al., 1997), and two of the four above species exhibit some of the highest mutation rates (Aurahs et al., 2009). The manual inspection of a multiple sequence alignment containing the reference database sequences (Morard et al., 2015) revealed the presence of up to 5 mismatches between these species sequences and the primer sequences used to generate the dataset. Hence, such mutations in the conserved regions of the gene where the primers anneal may be responsible for detection failures. Another preferential PCR amplification could also explain the strong skew dataset towards microperforate species sequences, which represent 55 to 99% of the reads (Fig. 3), but only 0 to 30% of the morphological assemblages (Kucera et al., 2005). The microperforate clade appears to have significantly lower rDNA substitution rates (Aurahs et al., 2009) and here we observe no mismatch between the primer and the reference sequences within this clade.

Alternatively, the higher abundance of reads assigned to microperforate taxa could represent a genuine pattern, questioning the representativeness of census counts of fossil foraminifera, which ignore specimens smaller than 150 µm (Kucera et al., 2005~~b~~). Microperforate species tend to be small and are disproportionately abundant in the size fraction smaller than 150 µm (Brummer et al., 1986). This is significant because the *e*DNA archive comprises information on all planktonic foraminifera irrespective of size and is thus potentially a more comprehensive recorder of species proportions in the plankton.

Overall, our results indicate PCR/primer bias as the ~~most~~ important limitation of planktonic foraminiferal community surveys based on metabarcoding. Alleviating them will allow detection of the full taxonomic spectrum, provided that sufficient sequencing effort is achieved, as recently discussed for fungi (Adams et al., 2013a, 2013b). To our knowledge, the dataset we re-analyzed represents the largest sequencing data for a given taxonomic group. Yet, it seems to indicate that the main ecological pattern can be extracted even from metabarcodes found at relatively modest frequencies (< 1000 reads, Figs. 3, 4: Caribbean samples). This conclusion underlines the importance of

comprehensive reference datasets and barcoding efforts to facilitate the development of specific and effective probing techniques to recover the signal of individual key groups (Pawlowski et al., 2012).

Metabarcoding surveys of marine sediments offer a powerful alternative to study marine plankton ecology and biogeography. Plankton *e*DNA diversity observed in sea floor sediments represents a continuous flux of biomass, averaged over seasons and throughout the entire water column. Unlike plankton sampling, sea-floor deposits are not affected by the seasonality, reproductive cycle or habitat depth of the plankton at the time of sampling. They offer a spatiotemporally archive of the overlying water column, which contains an integrated record of the maximum range of taxa that is realized at least at some point during the seasonal cycle. In this way, it is possible to constrain biogeographical patterns like endemism or ecological exclusion across oceanic gradients, without the need for highly time-resolved sampling. Importantly, *e*DNA data can be used to test the stability of biotic interactions inferred from the plankton (Lima-Mendez et al., 2015) simultaneously across a large range of environmental conditions represented in the sediment.

5 Conclusion

Assuming that *e*DNA deposited on the sea floor is also preserved through time, marine sediments should contain a remarkable ancient DNA (*a*DNA) archive of the history of the complete plankton communities. There is growing evidence that *e*DNA is preserved in marine sediments old enough to cover the previous ice age (Lejzerowicz et al., 2013). Until now, the interpretation of *a*DNA datasets from marine sediments suffered from insufficient sequencing depth (Coolen et al., 2009) or insufficient coverage of the reference database (Pawlowski et al., 2014a). As a result, to which degree the observed *a*DNA patterns reflect genuine past ecological shifts remained contentious. If sedimentary DNA is incorporated into marine sediments without preferential preservation that would induce taxonomic bias, as is the case in other environments (Pedersen et al., 2015), our data would support previous claims of DNA survival in deep-sea sediments, even where the sequencing depth was limited (Coolen et al., 2009). We hypothesize that if the ~~molecular imprint~~basic biogeographic and ecological information of foraminifera is preserved in surface sediments, the same should apply to the rest of the pelagic community. If confirmed, this, together with the latest developments in sequencing technologies, would open new avenues for paleoceanography and paleoecology, including the investigation of the impact of major past climate crises on oceanic communities, and the genetic detection of organisms not preserved in the fossil record. This is extremely important, now that the Tara-Oceans global

metabarcoding survey has shown that the largest portion of plankton biodiversity is composed of heterotrophic protists, parasites and symbionts that do not fossilize (de Vargas et al., 2015). In these regards, the information potentially preserved in deep sea sedimentary *a*DNA will likely revolutionize our understanding of the past ecology of marine plankton.

6 Data availability

Raw sequence data generated by Lecroq et al. (2010) and used in the present study are registered at the NCBI's Short Read Archive under the BioProject number PRJEB2682.

7 Supplements

Supplement 1

Detailed information on environmental samples and sequence data.

Supplement 2

Fasta files of the reference database and of the e-ribotypes attributed to planktonic foraminifera.

Supplement 3

Occurrences and taxonomic assignation of the e-ribotypes. Field explanations are given in the file.

Supplement 4

Individual alignments, phylogenetic trees and interpretation used to cluster the e-ribotype into genotypes.

7 Author contribution

R.M., F.L., M.K., L.O, J.P., S.M., and C.d.V conceived the project. B. L-B generated the original data. R.M., K. F. D., C.d.V. and M.K generated the reference database. F.L, R.M and M.K analyzed the dataset. R.M., F.L and M. K. wrote the manuscript and M.P., C.d.V., L.O., J.P., S.M., K.F.D., B. L-B., provided critical discussions and editions to the manuscript.

8 Competing interests

The authors declare no competing interests.

9 Acknowledgements

The study was supported by Swiss National Science Foundation grants 31003A-140766 and 313003A-159709 and by the DFG-Research Center/Cluster of Excellence "The Ocean in the Earth System".

10 References

Adams, R. I., Amend, A. S., Taylor, J. W. and Bruns, T. D.: A Unique Signal Distorts the Perception of Species Richness and Composition in High-Throughput Sequencing Surveys of Microbial Communities: A Case Study of Fungi in Indoor Dust, Microb. Ecol., 66(4), 735–741, doi:10.1007/s00248-013-0266-4, 2013a.

Adams, R. I., Miletto, M., Taylor, J. W. and Bruns, T. D.: Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances, ISME J., 7(7), 1262–1273, doi:10.1038/ismej.2013.28, 2013b.

Adl, S. M., Simpson, A. G. B., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., Le Gall, L., Lynn, D. H., McManus, H., Mitchell, E. a D., Mozley-Stanridge, S. E., Parfrey, L. W., Pawlowski, J., Rueckert, S., Shadwick, R. S., Shadwick, L., Schoch, C. L., Smirnov, A. and Spiegel, F. W.: The revised classification of eukaryotes., J. Eukaryot. Microbiol., 59(5), 429–93, doi:10.1111/j.1550-7408.2012.00644.x, 2012.

Aurahs, R., Göker, M., Grimm, G. W., Hemleben, V., Hemleben, C., Schiebel, R. and Kucera, M.: Using the Multiple Analysis Approach to Reconstruct Phylogenetic Relationships among Planktonic Foraminifera from Highly Divergent and Length-polymorphic SSU rDNA Sequences., Bioinform. Biol. Insights, 3, 155–177, 2009.

Berger, W. H. and Parker, F. L.: Diversity of planktonic foraminifera in deep-sea sediments., Science, 168(3937), 1345–1347, doi:10.1126/science.168.3937.1345, 1970.

Boere, a C., Rijpstra, W. I. C., De Lange, G. J., Sinninghe Damsté, J. S. and Coolen, M. J. L.: Preservation potential of ancient plankton DNA in Pleistocene marine sediments, Geobiology, 9(5), 377–393, doi:10.1111/j.1472-4669.2011.00290.x, 2011.

Bradley, I. M., Pinto, A. J. and Guest, J. S.: Design and Evaluation of Illumina MiSeq-Compatible, 18S rRNA Gene-Specific Primers for Improved Characterization of Mixed Phototrophic Communities, edited by G. Voordouw,

Appl. Environ. Microbiol., 82(19), 5878–5891, doi:10.1128/AEM.01630-16, 2016.

Brummer, G. A., Hemleben, C. and Michael, S.: Planktonic foraminiferal ontogeny and new perspectives for micropalaeontology, Nature, 319, 50–52, 1986.

Capo, E., Debroas, D., Arnaud, F. and Domaizon, I.: Is Planktonic Diversity Well Recorded in Sedimentary DNA? Toward the Reconstruction of Past Protistan Diversity, Microb. Ecol., 70(4), 865–875, doi:10.1007/s00248-015-0627-2, 2015.

Ceccherini, M. T., Ascher, J., Agnelli, A., Borgogni, F., Pantani, O. L. and Pietramellara, G.: Experimental discrimination and molecular characterization of the extracellular soil DNA fraction, Antonie Van Leeuwenhoek, 96(4), 653–657, doi:10.1007/s10482-009-9354-3, 2009.

Coolen, M. J. L. and Overmann, J.: 217 000-year-old DNA sequences of green sulfur bacteria in Mediterranean sapropels and their implications for the reconstruction of the paleoenvironment, Environ. Microbiol., 9(1), 238–249, doi:10.1111/j.1462-2920.2006.01134.x, 2007.

Coolen, M. J. L., Saenz, J. P., Giosan, L., Trowbridge, N. Y., Dimitrov, P., Dimitrov, D. and Eglinton, T. I.: DNA and lipid molecular stratigraphic records of haptophyte succession in the Black Sea during the Holocene, Earth Planet. Sci. Lett., 284(3–4), 610–621, doi:10.1016/j.epsl.2009.05.029, 2009.

Coolen, M. J. L., Orsi, W. D., Balkema, C., Quince, C., Harris, K., Sylva, S. P., Filipova-Marinova, M. and Giosan, L.: Evolution of the plankton paleome in the Black Sea from the Deglacial to Anthropocene., Proc. Natl. Acad. Sci. U. S. A., 110(21), 8609–14, doi:10.1073/pnas.1219283110, 2013.

Corinaldesi, C., Dell'Anno, A. and Danovaro, R.: Viral infection plays a key role in extracellular DNA dynamics in marine anoxic systems, Limnol. Oceanogr., 52(2), 508–516, doi:10.4319/lo.2007.52.2.0508, 2007.

Corinaldesi, C., Barucca, M., Luna, G. M. and Dell'Anno, A.: Preservation, origin and genetic imprint of extracellular DNA in permanently anoxic deep-sea sediments., Mol. Ecol., 20(3), 642–54, doi:10.1111/j.1365-294X.2010.04958.x, 2011.

Corinaldesi, C., Tangherlini, M., Luna, G. M. and Anno, A. D.: Extracellular DNA can preserve the genetic signatures of present and past viral infection events in deep hypersaline anoxic basins, Proc. R. Soc. B Biol. Sci., doi:10.1098/rspb.2013.3299, 2014.

Darling, K. F. and Wade, C. M.: The genetic diversity of planktic foraminifera and the global distribution of ribosomal RNA genotypes, Mar. Micropaleontol., 67(3–4), 216–238, doi:10.1016/j.marmicro.2008.01.009, 2008.

Dell'Anno, A. and Danovaro, R.: Extracellular DNA Plays a Key Role in Deep-Sea Ecosystem Functioning, Science, 309(5744), 2179–2179, doi:10.1126/science.1117475, 2005.

Dutkiewicz, A., O'Callaghan, S. and Müller, R. D.: Controls on the distribution of deep-sea sediments, Geochemistry, Geophys. Geosystems, 17(8), 3075–3098, doi:10.1002/2016GC006428, 2016.

Esling, P., Lejzerowicz, F. and Pawlowski, J.: Accurate multiplexing and filtering for high-throughput amplicon-sequencing, Nucleic Acids Res., 43(5), 2513–2524, doi:10.1093/nar/gkv107, 2015.

Gouy, M., Guindon, S. and Gascuel, O.: SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building., Mol. Biol. Evol., 27(2), 221–4, doi:10.1093/molbev/msp259, 2010.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., Mahe, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A.-L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P. and Christen, R.: The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy, Nucleic Acids Res., 41(D1), D597–D604, doi:10.1093/nar/gks1160, 2013.

Hammer, Ø., Harper, D. a. T. and Ryan, P. D.: Paleontological statistics software package for education and data analysis, Palaeontol. Electron., 4(1), 9–18, doi:10.1016/j.bcp.2008.05.025, 2001.

Hemleben, C., Michael, S. and Anderson, O. R.: Modern Planktonic Foraminifera, Springer-Verlarg, New York., 1989.

Hillaire-Marcel, C. and de Vernal, A.: Proxies in Late Cenozoic Paleoceanography, Elsevier., 2007.

Katoh, K. and Standley, D. M.: MAFFT multiple sequence alignment software version 7: Improvements in performance and usability, Mol. Biol. Evol., 30(4), 772–780, doi:10.1093/molbev/mst010, 2013.

Kirkpatrick, J. B., Walsh, E. A. and D'Hondt, S.: Fossil DNA persistence and decay in marine sediment over hundred-thousand-year to million-year time scales, Geology, 44(8), 615–618, doi:10.1130/G37933.1, 2016.

Kucera, M., Weinelt, M., Kiefer, T., Pflaumann, U., Hayes, A., Weinelt, M., Chen, M.-T., Mix, A. C., Barrows, T.

T., Cortijo, E., Duprat, J., Juggins, S. and Waelbroeck, C.: Reconstruction of sea-surface temperatures from assemblages of planktonic foraminifera: multi-technique approach based on geographically constrained calibration data sets and its application to glacial Atlantic and Pacific Oceans, Quat. Sci. Rev., 24(7–9), 951–998, doi:10.1016/j.quascirev.2004.07.014, 2005.

Lecroq, B., Lejzerowicz, F., Bachar, D., Christen, R., Esling, P., Baerlocher, L., Osterås, M., Farinelli, L. and Pawlowski, J.: Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments., Proc. Natl. Acad. Sci. U. S. A., 108(32), 13177–13182, doi:10.1073/pnas.1018426108, 2011.

Lejzerowicz, F., Esling, P., Majewski, W., Szczucinski, W., Decelle, J., Obadia, C., Arbizu, P. M. and Pawlowski, J.: Ancient DNA complements microfossil record in deep-sea subsurface sediments, Biol. Lett., 9(4), 20130283–20130283, doi:10.1098/rsbl.2013.0283, 2013.

Letunic, I. and Bork, P.: Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy, Nucleic Acids Res., 39(suppl), W475–W478, doi:10.1093/nar/gkr201, 2011.

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J. C., Roux, S., Vincent, F., Bittner, L., Darzi, Y., Wang, J., Audic, S., Berline, L., Bontempi, G., Cabello, A. M., Coppola, L., Cornejo-Castillo, F. M., D'Ovidio, F., De Meester, L., Ferrera, I., Garet-Delmas, M.-J., Guidi, L., Lara, E., Pesant, S., Royo-Llonch, M., Salazar, G., Sanchez, P., Sebastian, M., Souffreau, C., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Gorsky, G., Not, F., Ogata, H., Speich, S., Stemmann, L., Weissenbach, J., Wincker, P., Acinas, S. G., Sunagawa, S., Bork, P., Sullivan, M. B., Karsenti, E., Bowler, C., de Vargas, C. and Raes, J.: Determinants of community structure in the global plankton interactome, Science, 348(6237), 1262073–1262073, doi:10.1126/science.1262073, 2015.

Morard, R., Darling, K. F., Mahé, F., Audic, S., Ujiié, Y., Weiner, A. K. M., André, A., Seears, H. a., Wade, C. M., Quillévéré, F., Douady, C. J., Escarguel, G., de Garidel-Thoron, T., Siccha, M., Kucera, M. and de Vargas, C.: PFR^2: a curated database of planktonic Foraminifera18S ribosomal DNA as a resource for studies of plankton ecology, biogeography, and evolution, Mol. Ecol. Resour., 49, 1–14, doi:10.1111/1755-0998.12410, 2015.

Morard, R., Escarguel, G., Weiner, A. K. M., André, A., Douady, C. J., Wade, C. M., Darling, K. F., Ujiié, Y., Seears, H. A., Quillévéré, F., de Garidel-Thoron, T., de Vargas, C. and Kucera, M.: Nomenclature for the Nameless:

A Proposal for an Integrative Molecular Taxonomy of Cryptic Diversity Exemplified by Planktonic Foraminifera, Syst. Biol., 65(5), 925–940, doi:10.1093/sysbio/syw031, 2016.

Needleman, S. B. and Wunsch, C. D.: A general method applicable to the search for similarities in the amino acid sequence of two proteins., J. Mol. Biol., 48(3), 443–453 [online] Available from: http://www.ncbi.nlm.nih.gov/pubmed/5420325, 1970.

Pawłowska, J., Lejzerowicz, F., Esling, P., Szczuciński, W., Zajączkowski, M. and Pawlowski, J.: Ancient DNA sheds new light on the Svalbard foraminiferal fossil record of the last millennium., Geobiology, 12(4), 277–88, doi:10.1111/gbi.12087, 2014.

Pawlowski, J. and Lecroq, B.: Short rDNA barcodes for species identification in foraminifera., J. Eukaryot. Microbiol., 57(2), 197–205, doi:10.1111/j.1550-7408.2009.00468.x, 2010.

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., Bowser, S. S., Cepicka, I., Decelle, J., Dunthorn, M., Fiore-Donno, A. M., Gile, G. H., Holzmann, M., Jahn, R., Jirků, M., Keeling, P. J., Kostka, M., Kudryavtsev, A., Lara, E., Lukeš, J., Mann, D. G., Mitchell, E. a D., Nitsche, F., Romeralo, M., Saunders, G. W., Simpson, A. G. B., Smirnov, A. V, Spouge, J. L., Stern, R. F., Stoeck, T., Zimmermann, J., Schindel, D. and de Vargas, C.: CBOL Protist Working Group: Barcoding Eukaryotic richness beyond the Animal, Plant, and Fungal Kingdoms, PLoS Biol., 10(11), e1001419, doi:10.1371/journal.pbio.1001419, 2012.

Pawlowski, J., Esling, P., Lejzerowicz, F., Cedhagen, T. and Wilding, T. A.: Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities, Mol. Ecol. Resour., 14(6), 1129–1140, doi:10.1111/1755-0998.12261, 2014a.

Pawlowski, J., Lejzerowicz, F. and Esling, P.: Next-Generation Environmental Diversity Surveys of Foraminifera : Preparing the Future, Biol. Bull., 227, 93–106, 2014b.

Pedersen, M. W., Ginolhac, A., Orlando, L., Olsen, J., Andersen, K., Holm, J., Funder, S., Willerslev, E. and Kjær, K. H.: A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa, Quat. Sci. Rev., 75, 161–168, doi:10.1016/j.quascirev.2013.06.006, 2013.

Pedersen, M. W., Overballe-Petersen, S., Ermini, L., Sarkissian, C. Der, Haile, J., Hellstrom, M., Spens, J.,

Thomsen, P. F., Bohmann, K., Cappellini, E., Schnell, I. B., Wales, N. A., Caroe, C., Campos, P. F., Schmidt, A. M. Z., Gilbert, M. T. P., Hansen, A. J., Orlando, L. and Willerslev, E.: Ancient and modern environmental DNA, Philos. Trans. R. Soc. B Biol. Sci., 370(1660), 20130383–20130383, doi:10.1098/rstb.2013.0383, 2015.

Rutherford, S., D'Hondt, S. and Prell, W.: Environmental controls on the geographic distribution of zooplankton diversity, Nature, 400, 749–753, doi:10.1038/23449, 1999.

Taberlet, P., Prud'Homme, S. M., Campione, E., Roy, J., Miquel, C., Shehzad, W., Gielly, L., Rioux, D., Choler, P., Clément, J.-C., Melodelima, C., Pompanon, F. and Coissac, E.: Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies, Mol. Ecol., 21(8), 1816–1820, doi:10.1111/j.1365-294X.2011.05317.x, 2012.

Torti, A., Lever, M. A. and Jørgensen, B. B.: Origin, dynamics, and implications of extracellular DNA pools in marine sediments, Mar. Genomics, 24, 185–196, doi:10.1016/j.margen.2015.08.007, 2015.

de Vargas, C., Zaninetti, L., Hilbrecht, H. and Pawlowski, J.: Phylogeny and rates of molecular evolution of planktonic foraminifera: SSU rDNA sequences compared to the fossil record., J. Mol. Evol., 45(3), 285–294, 1997.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horak, A., Jaillon, O., Lima-Mendez, G., Luke, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Acinas, S. G., Bork, P., Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Raes, J., Sieracki, M. E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P., Karsenti, E., Boss, E., Follows, M., Karp-Boss, L., Krzic, U., Reynaud, E. G., Sardet, C., Sullivan, M. B. and Velayoudon, D.: Eukaryotic plankton diversity in the sunlit ocean, Science, 348(6237), 1261605–1261605, doi:10.1126/science.1261605, 2015.

Weber, A. A. T. and Pawlowski, J.: Can Abundance of Protists Be Inferred from Sequence Data: A Case Study of Foraminifera, PLoS One, 8(2), 1–8, doi:10.1371/journal.pone.0056739, 2013.

Yasuhara, M., Tittensor, D. P., Hillebrand, H. and Worm, B.: Combining marine macroecology and palaeoecology in understanding biodiversity: microfossils as a model, Biol. Rev., doi:10.1111/brv.12223, 2015.

Figures

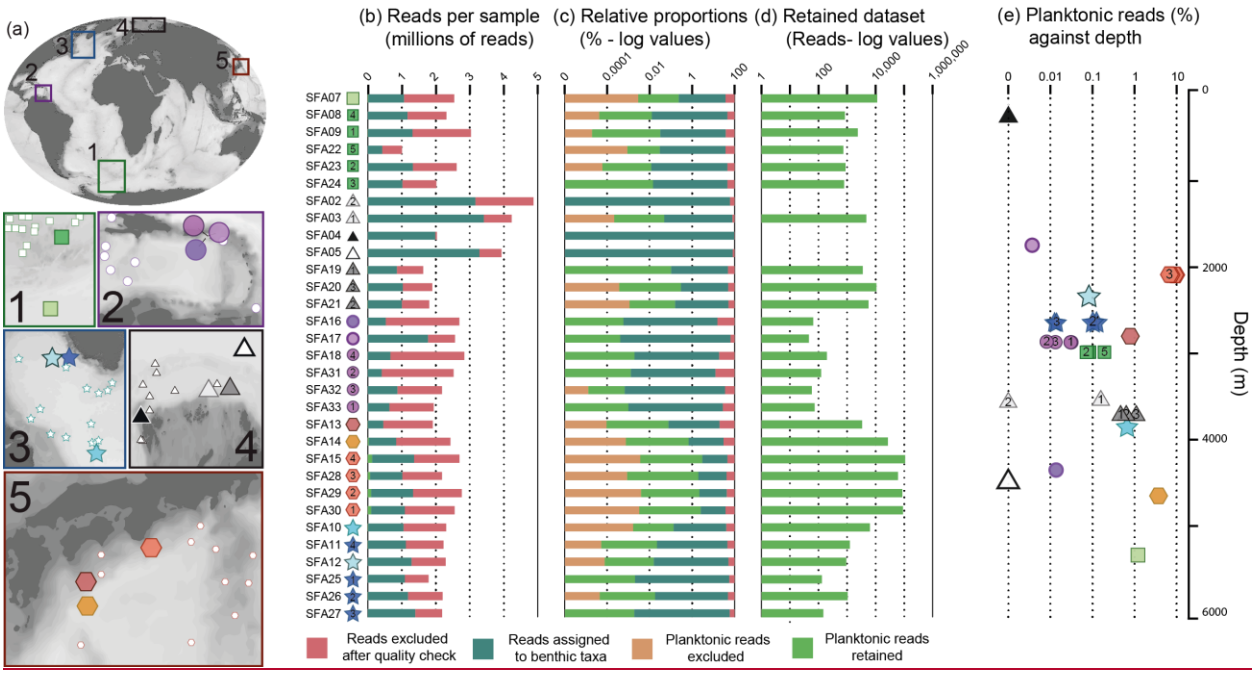

Figure 1. Occurrences of planktonic foraminifera in abyssal sedimentary *e*DNA. (a) Geographic location of the samples. The boxes indicate the location the core top samples in the 5 sampled regions. The larger symbols indicate the location of the samples used for *e*DNA analysis generated by Lecroq et al. (2011) and the smaller open symbol the location of the census count from the MARGO database (Kucera et al., 2005). (b-d) Results of the filtering and assignation of the dataset. The symbols with numbers correspond to the replicates of a single location shown on (a), next to the libraries name. The replicates are subsamples originating from the same gear (Lecroq et al., (2010)). (e) Relative proportions of planktonic reads in the individual samples in logged values plotted against depth.~~(b) Relative proportions of planktonic reads in the individual samples in logged values plotted against depth. The symbols with numbers correspond to the replicates of a single location shown on (a).~~

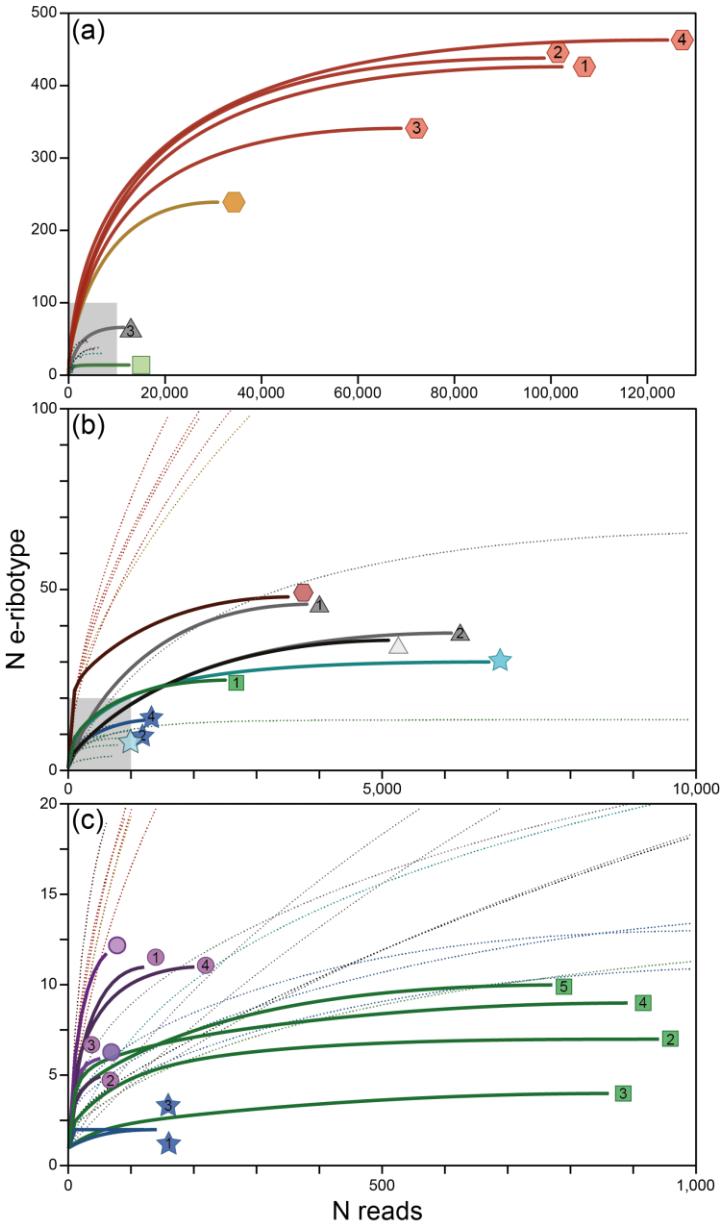

Figure 2. Rarefaction curves. e-ribotypes rarefaction curves of each of the 28 samples containing planktonic foraminifera sequences. The three boxes show the same rarefaction curves at 3 different scales highlighted by grey rectangles. ame symbols as~~The symbols with numbers correspond to the replicates of a single location shown on Fig. 1~~. For each magnification, the curves which are out of range are drawn in dashed lines to ease the reading to the figure.

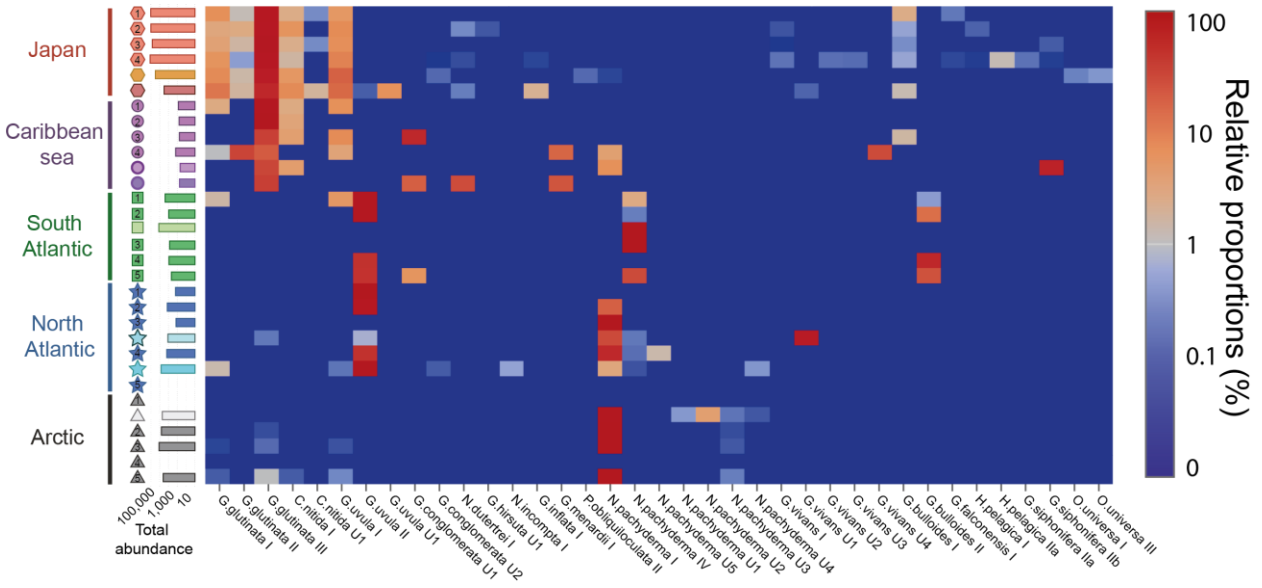

Figure 3. Heat Map of the relative proportions of the genetic types ~~genotypes~~ detected in the 31 samples. The histogram on the left side of the heat map indicates the total abundance in log-value of the reads belonging to planktonic foraminifera. Symbols as on Figs 1-2.

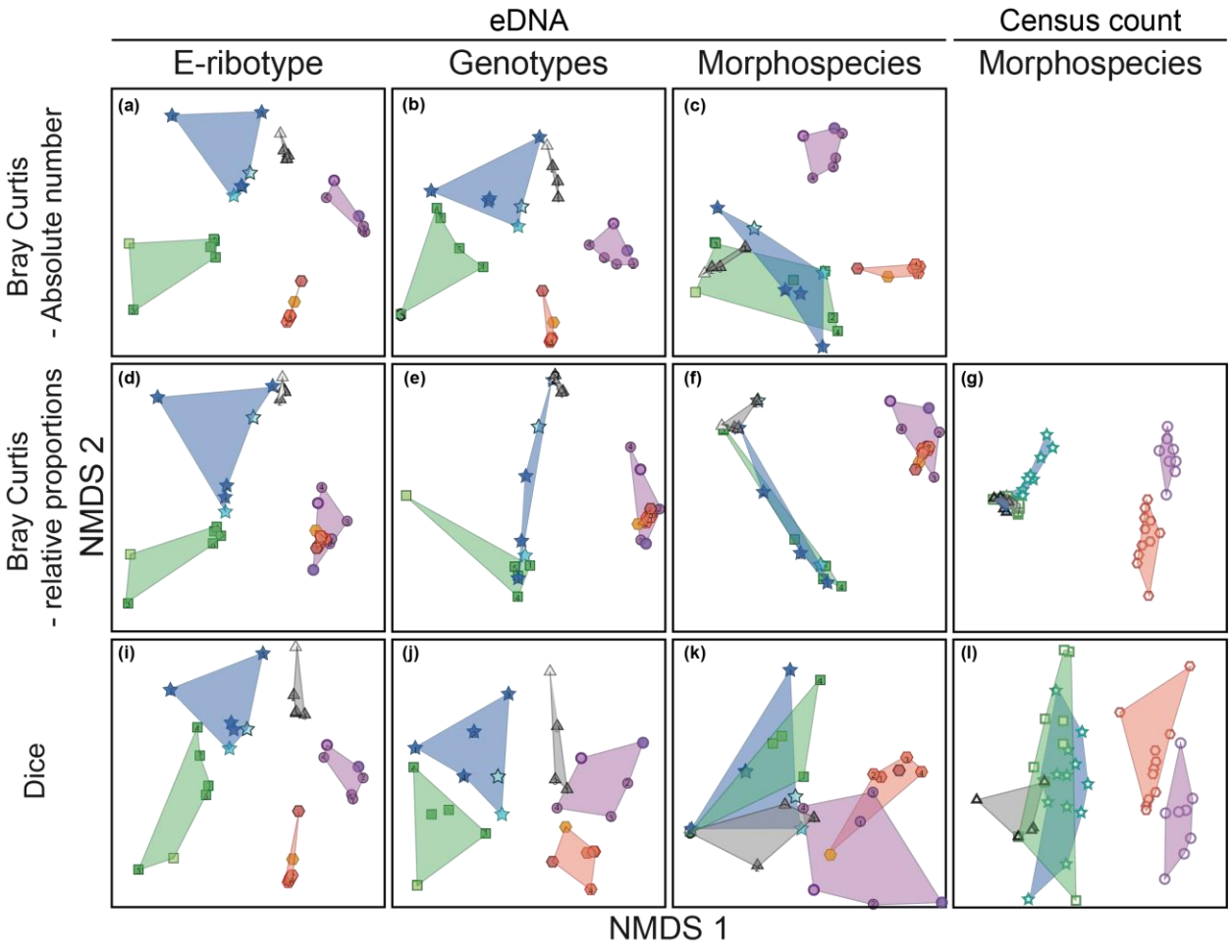

Figure 4. Community structuring of planktonic foraminifera in sedimentary *e*DNA. Grouping of *e*DNA and census count samples according to their taxonomic composition using Non-linear Multi-Dimensional Scaling based on Bray-Curtis (Absolute number (a-c) and relative abundances (d-g)) and Dice distances (i-l). The NMDS are provided for the three different degrees of taxonomic resolution (ribotypes, genotypes and morphospecies) for the *e*DNA samples. As the census count are relative abundances, the Bray-Curtis on absolute value is not provided for the census count assemblages. The area covered by the samples of each region is highlighted. Symbols as in Fig. 1b and Fig. 2.

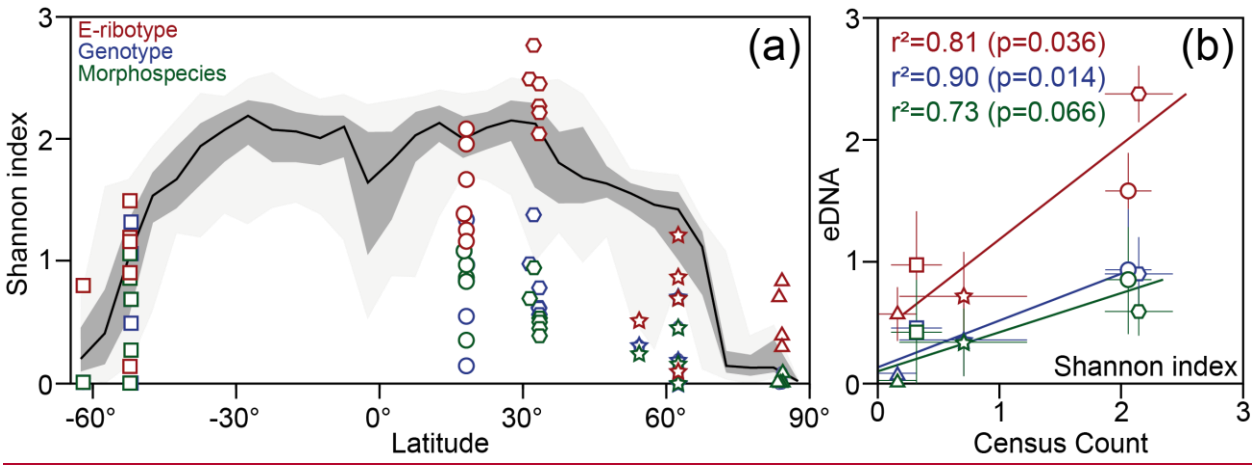

Figure 5. Macro ecological pattern of spatial diversity known as the latitudinal gradient of diversity~~mid-domain effect~~. (a) The grey areas represent the distribution of the Shannon index calculated on the census count of planktonic foraminifera in core top samples from the MARGO database (Kucera et al., 2005) against latitude. The dark grey area represents the 1st-3rd quartiles (50% confidence interval), light grey the 5th-95th percentile (90 % confidence interval), and the black line is the median. The same similarity measure has been calculated at each location for the *e*DNA samples based on the relative abundances with the three levels of taxonomic resolution. **(b)** Relationships between mean Shannon index calculated at regional levels (symbols as on Figure 1) for census count and *e*DNA assemblages, vertical and horizontal lines indicate the standard deviation. Coefficient of correlation and p values are provided for the three taxonomic levels but are only indicative because the number of data point is too low to draw definitive conclusion.

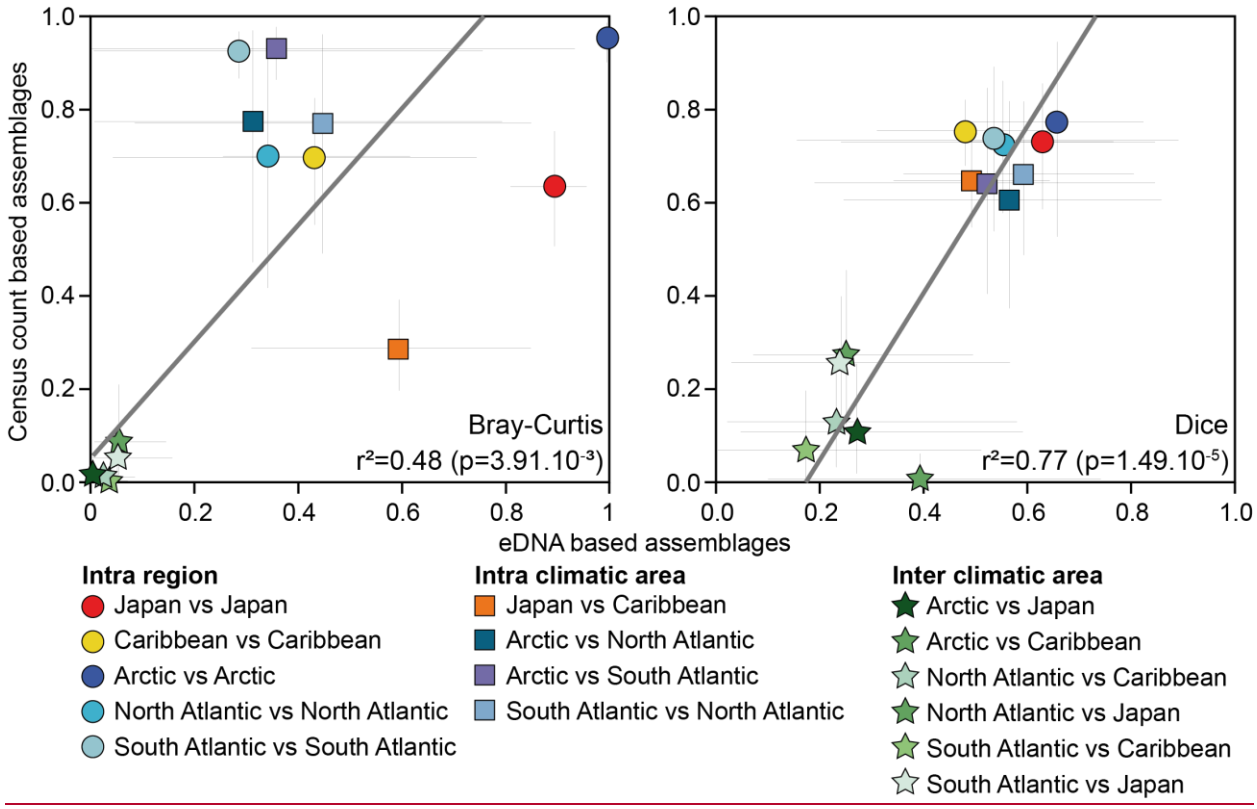

Figure 6. *e*DNA vs census counts. Similarity pairwise comparison of the community composition inferred from relative abundances of morphospecies based on *e*DNA and census counts among and between the five sampled regions based on the Bray-Curtis and Dice indices. ~~.~~ Each symbol corresponds to the average between all pairwise distances of each category and the lines represent the standard deviation. The gray lines represent the linear regression, with r^2 and p values provided in the right bottom corner of each graph. ~~The distances are based on the Bray-Curtis and Dice indices.~~