

Interactive comment on “Palaeoceanographic changes in Hornsund Fjord (Spitsbergen, Svalbard) over the last millennium: new insights from ancient DNA” by J. Pawłowska et al.

J. Pawłowska et al.

pawlowska@iopan.pl

Received and published: 4 January 2016

We would like to thank the Referee for constructive review, that will help us to improve the manuscript. Written below are our responses to the Referee's comments. The comments were reproduced and are followed by our responses. Based on the comments, we propose the changes of the manuscript. The revised version of the manuscript will be prepared based on the decision of the Editor.

Referee's comment: Abstract Page 3666 line 12: The early LIA: : .This is a strong claim since only one sample in this climate interval was analyzed for aDNA. Here, as well as throughout the manuscript there is a need to describe the growth or environ-

C2834

mental requirements of described species in more detail.

Response: We agree with Referee that this is a strong claim, considering that it was based on only one sample. On the other hand, the sequences of *Hippocrepinella hirudinea* and *Cedhagenia saltatus* made up approx. 23% of the monothalamid sequences in the above-mentioned sample, while in the other samples their abundance was rather minor. Therefore, the statement might be justified. Our knowledge about monothalamids ecology is limited, as they are usually not included in the studies of modern foraminifera assemblages. Therefore, the ecological interpretation of monothalamids data is often difficult or even not possible (see page 3680, lines 2-3; page 3682, lines 25-26). The need of increasing our knowledge about monothalamids ecology was one of the messages of the paper (page 3684, lines 6-7).

Referee's comment: Page 3666 line 17: Also here, only an expert would know what an increase in the relative abundance of these two species implies. In general: Are environmental sequences really that informative that you can say which exact species were present? I think that you can only describe environmental sequences at species-level if the corresponding microfossil is present. If not, it is a safer bet to stay at genus or family level.

Response: The taxonomic assignment of sequences strongly depends on the reference database. When this database is rich, as in the case of Arctic foraminifera, the number of assigned species might be high. Indeed, to the paleontological purpose of the study, using species names is accessory because it cannot be ascertained whether a sequence assigned to a given species originated from a microfossil assigned to this species, for which there might be no reference sequence available in the sequence database. We analyzed two different species entities: the morphospecies based on microfossils and the OTU based on the SSU rDNA fragment. Both species entities can be assigned to known species name and the several OTUs that could be assigned to the same known species name may in fact represent different strains or genotypes of this species name. Indeed, the 37f region does allow the assignment of environ-

C2835

mental sequences down to the species level, provided that the species is present in the reference database. We chose to amplify and sequence a short fragment of the SSU rDNA corresponding to the hypervariable region 37f, which only exists among the foraminifera. This region has been identified as an ideal barcode for species-level assignment of foraminiferal sequence (cf. Pawłowski and Lecroq, 2010; J. Eukaryot. Microbiol.). The exactness of the sequence assignment also depends on the quality of the sequencing data and on the completeness of the reference database. We agree that to some extent, the variation in the 37f region may be such that it may represent another species. This is why we accounted for the possibility of wrongly matching environmental sequences to distant species by taking only the consensus of the taxonomies of all the reference sequences that match within a threshold of 5 differences, which complies with the Referees' thought.

Referee's comment: Introduction Page 3667 Line 13: However, to fully understand: :
: This paragraph seems to be out of place. Namely, this study does not result in the full understanding of the consequences of climate changes in the Arctic. Please stick to claims and aims that you have studied and discussed. The first few paragraphs should only discuss what is known about past climate in the region. Then: what the big unknowns are, how forams can help, limitations of the analysis of fossils, how aDNA can help, followed by what you did here and a few lines about the major findings.

Response: The aim of the mentioned paragraph was to point at fjords' potential of providing high-resolution sedimentary record, which might provide valuable information about climate-driven environmental changes in the Arctic. We agree that the first sentence of the paragraph is irrelevant, therefore we propose to replace it with: "Fjords are unique form of coastline, that are affected from two directions: the glaciated land and the ocean, rendering the fjord system a sensitive indicator of climate change phenomena."

Referee's comment: Line 26: Therefore, it is crucial: : This is a very big claim since a complete model of past environmental changes in the Arctic fjords is not provided with

C2836

this study. Hence, the need for better structuring the introduction.

Response: The sentence emphasize the need of creation of a wide range of paleoceanographic proxies sourcing from different research methods for better understanding past climate and environmental changes. We have not claimed that we will provide a complete model of paleoenvironmental changes in the Arctic. We propose to change the sentence to: "To study accurately the climate-driven environmental variability in the past, it is crucial to create a network of proxies carrying different but complementary information."

Referee's comment: Page 3668 Line 19: Metagenetics (the analysis of many genes) is a cool but also vague term. Please be more specific about what "metagenomics" was performed (i.e., the identification of past foraminifera including non-fossilized taxa through PCR amplification and sequencing analysis of preserved sedimentary taxonomic marker genes. Response: Herein, the term metagenetics refers to genetic material obtained from environmental samples. Therefore, we will replace the term with 'ancient environmental DNA'.

Referee's comment: Page 3668 Line 25 and following: The ignorant reader might wonder why you can detect the DNA but not the microfossils. Please say a few words about why the DNA might still be present.

Response: We agree that there is a need to provide more information about the match between the fossil and molecular data. Therefore, we propose to add a more detailed description of previous findings in the Introduction: "To include monothalamids into palaeoecological studies of the Arctic foraminifera we analysed the ancient foraminiferal DNA record from the last millennium from Hornsund (Pawłowska et al., 2014). The study showed that aDNA record detected most of species reported for Hornsund from previous micropaleontological investigations (e.g., Hald and Korsun, 1997; Pogodina, 2005).including species that dominate fossil assemblage (i.e., *E. excavatum*, *C. reniforme*, *C. lobatulus* and *N. labradorica*; cf. However, the number of

C2837

aDNA sequence reads and fossil specimens differed considerably. The richness of the foraminiferal communities revealed by molecular analysis was much higher than in the fossil record, mainly due to the detection of high number of monothalamous species that were not preserved during the fossilisation process as well as small-size species that are not retained on micropaleontological sieves.”

Referee's comment: Page 3669 Line 3: The Pawłowska et al., 2014 seems to be very important to cross read to fully explore this study. I was unable to get an electronic version despite being able to use the online libraries of two major universities. I strongly suggest to describe major findings and relevant methods from this paper in more detail also in this paper.

Response: We agree with the Referee that more detailed description of previous findings and analytical methods will facilitate the understanding of the presented study. The adequate information will be added to the Introduction and Material and methods. However, the .pdf file of Pawłowska et al. (2014) paper can be obtained via Google Scholar.

Referee's comment: Page 3671 Line 14: Please describe in a bit more details what this statistical approach exactly does.

Response: The Principal Component (PC) analysis showed the contribution of each foraminiferal species in the assemblage, what enables us to identify the dominant species. The taxa that favor similar environmental conditions may have high scores on one PC, indicating their participation in the assemblage. The adequate explanation will be added to the Material and methods.

Referee's comment: Page 3671 Line 19: For reasons mentioned two comments ago: Please provide a brief summary of these methods here. I don't think that the reader needs to be able to cross read the 2014 paper to find out what methods have been used.

C2838

Response: As mentioned above, the additional information will be added to the text.

Referee's comment: Results Page 3673 Line 18: Spell out VPDB the first time.

Response: It has been corrected.

Referee's comment: Page 3674 Line 21 and following: It would have been nice to have seen a similar type of analysis to identify indicator taxa and their importance to explain environmental stages for the molecular data. However, to do so you would need a much higher sampling resolution such as was the case for the microfossil work. I am not sure why the sampling resolution for the aDNA data is not the same. Extracting DNA and subsequent sequencing has become very cheap. It would have been a month or so extra work to get all the DNA extracts, do the PCRs and to prepare the libraries for sequencing. I have more comments about this later on.

Response: Statistical analysis (in our case PC analysis) is based on the quantitative data (i.e. the absolute number of fossil specimens). The aDNA data is mainly qualitative, therefore, the results of statistical analyses of molecular data will be strongly biased. The quantitative analysis of aDNA is still a challenging issue. The aDNA data should be interpreted carefully, as it is not possible to establish the direct relationship between the number of specimens and the number of ribosomal sequences, due to the e.g., interspecific variability of number of rDNA copies. In ancient DNA studies major difficulties arise also from DNA degradation and chemical modification. Therefore, the absolute number of sequences should be interpreted with caution; however, it is possible to identify the dominant species based on the sequence proportion (Weber and Pawłowski, 2013; PLoS ONE). The aspects of qualitative and quantitative molecular data analysis were discussed in the last paragraph of the discussion (see page 3682, line 29 and following). We agree with the Referee, that higher sampling resolution and higher amount of data will provide more complete view of changes in foraminiferal assemblages. However, the presented as well as previous study (Pawłowska et al. 2014; Geobiol.) were the first attempts to analyze the ancient foraminiferal DNA in the Arctic.

C2839

We did not know if foraminiferal DNA is preserved in Arctic marine sediments. Therefore, the chosen sampling resolution is not as high as in case of micropaleontological analysis. As described in Material and methods, the samples for molecular analyses were taken onboard, directly after taking the sediment core. For aDNA analysis, sub-surface sediment samples were taken from the inner part of the core. In order to prevent the disruption of the core structure and therefore cross-contaminations, the core tube was bored at each selected depth and ~5 grams of sediment was sub-sampled using disposable spoons. Additional spoons were used to carefully remove the outer part of the core. This method does not allow to perform the subsampling with as high resolution as in case of micropaleontological analysis. Later, the core was cut into 1-cm slices and the material was used for micropaleontological and sedimentological analyses (page 3670, lines 11-15). Therefore, it was not possible to increase the resolution of molecular sampling thereafter. One of the possibilities to increase the amount of molecular data would be to extract larger volume of sediments, what shall be done in the future studies.

Referee's comment: Page 3675 Line 10 and following: Please provide more detail in the methods so that it becomes clear how the # of OTUs was determined. The reader should not have to get a copy of the 2014 paper to understand this study.

Response: As mentioned above, more detailed description will be added to the text.

Referee's comment: Same page line 24: Are these the only possible most similar sequences (i.e., top hit returns from BLAST)? Often several species or genera have the same sequence similarity. Please make sure to be precise about the true taxonomic level that can be revealed from the sequences. See also earlier comment about this.

Response: We would like to remind the we did not use BLAST but global sequence alignments using the Needleman-Wunsch algorithm, based on which we calculated the distance by counting each gap and substitution as a difference, allowing up to 5 differences to perform a "species" level assignment (Pawlowski et al., 2014, Biol. Bull.;

C2840

Esling et al. 2015, Nucleic Acids Res.). This "species" level is accessory as explained above, and thus do not refer to any formal species nomenclature but rather serves to provide a gross taxonomic information since (1) the reference sequence database is not complete for the Svalbard area, (2) species level assignment for monothalamous foraminiferans are bound to revision and (3) it suffice to reliably document diversity patterns observed based on the OTUs, that may entail a taxonomic information more precise than that conveyed by the microfossils (cryptic species resolution). Within 5 differences, each of the eleven OTU sequences implicated in the reviewer's question were matching one or several reference sequences, but when multiple matches arose, then the taxonomies of all the matching reference sequences were congruent at the formal "species" level. In fact, only one of these 11 OTU sequences matched to more than one reference sequence within 5 differences (one *Micrometula* reference sequence from Scotland and another from Skagerrak).

Referee's comment: Page 3676 Line 8 and following: As mentioned earlier: This claim is based on only one sample from that climate interval.

Response: As mentioned above – we agree that higher resolution of sampling will provide more data and more complete view of the foraminiferal community. However, in the early LIA (which encompasses two samples – at 150 m, dated to be ~ 1550 AD and at 125 m, dated to be ~ 1800 AD), the percentage of sequences of *Hippocrepinella hirudinea* and *Cedhagenia saltatus* was much higher than in the precedent and following periods. Therefore, the statement might be justified.

Referee's comment: Discussion: Page 3677 Line 9: I think that the sampling resolution is too low to make such claims. Please inform a bit more about what is known about the growth or environmental requirements of *Toxisarcon*.

Response: As mentioned above (see the response to the first comment), monothalamous foraminifera are important but largely ignored component of Arctic meiofauna. So far, three species of *Toxisarcon* have been described: *T. synsuicidica* (Cedhagen

C2841

and Pawlowski, 2002; J. Foramin. Res.), *T. alba* (Wilding, 2002; J. Foramin. Res.) and *T. taimyr* (Voltski et al., 2015; Mar. Biodiv.). However, only Wilding (2002) provided information about their possible environmental preferences: the specimens were found buried or semi-buried in well oxygenated sand. In Svalbard, Gooday (2005; Mar. Biol. Res.) found unidentified *Toxisarcon* specimens in the inner parts of three fjords, dominated by glaciomarine mud. To conclude, there is almost no data on *Toxisarcon* ecology and distribution patterns. Therefore, we were able to suggest only the link between *Toxisarcon* occurrence and phytoplankton-originated organic matter input (as presented in page 3677, line 15)

Referee's comment: Same page line 17: Is the $\delta^{18}O$ at 1600 AD really that different to link this to an increase of melt water delivery etc? Response: As written in the mentioned line, the peak was slight. Our interpretation was supported by other proxies (IRD, foraminiferal fauna) as well as previous findings from Hornsund fjord (Majewski et al., 2009).

Referee's comment: Page 3678 Line 8 and following: I don't see why this is obvious when looking at Fig. 3. When looking at the scale, *Islandiella* spp. seem to have never exceeded more than 3.5% of the total foram distribution.

Response: As mentioned in the manuscript, the peaks were slight. However, the percentage of *Islandiella* spp (*I. norcrossi* and *I. helenae*) made up 11% of the total assemblage. However, the *Islandiella* spp. peak was not interpreted separately, but was supported by other proxies.

Referee's comment: Page 3682 Line 25: This is true but with a substantially higher sampling resolution throughout the core, it would have been possible to perform an indicator species analysis to identify which taxa show a statistically significant response to the various environmental stages. This way even unnamed environmental sequences could potentially become proxies for certain conditions in comparable settings.

Response: Indeed, the statistically-supported analysis might allow to relate the pres-

C2842

ence of environmental sequences to certain environmental variables. However, as explained above, the proper statistical analysis of our aDNA data is not possible because of limited number of sequences. The results of such analysis will be strongly biased and the interpretation will be speculative. However, we hope, that further development of environmental aDNA research will overcome these limitations and will provide sufficient amount of data to perform statistical analyses.

Referee's comment: Page 3682 final paragraph: This paragraph about the problems with aDNA work is highly speculative. You don't actually have empirical proof that your DNA is degraded and if this differs between intervals. The sediments analyzed here are relatively young. A much higher sampling resolution (e.g. every other cm or so) combined with statistical approaches will most likely reveal highly significant changes in the species distribution as a result of major climate shifts. There will probably be less need to write a negative and speculative paragraph about the things that can go wrong with the aDNA approach. Right now this paragraph is totally out of place.

Response: We agree that higher sampling resolution and statistical analysis of data will provide more detailed paleoenvironmental information. As discussed above, the statistical analysis of non-quantitative data will be strongly biased and interpretation will be speculative. The final paragraph tackles the important issue of quantitation of aDNA data. In the modern environment, the establishment of the relationship between the number of sequences and number of specimens is possible through the normalization of the results, according to, e.g., the interspecific variability of number of rDNA copies (Pawlowski et al., 2014; Biol. Bull.). In case of aDNA data additional difficulties arise from the degradation of the material and the limited number of sequences, so the data quantitation is much more complicated. The paragraph will be modified to emphasize the fact, that the presented aDNA data can be only interpreted qualitatively and therefore, an ancient environmental DNA approach should be used as a complementary source of information, supported by other proxies.

Referee's comment: Table S2: Please make sure to identify the highest taxonomic level

C2843

for each OTU based on Blast results (e.g., if an OTU shows the same highest similarity with multiple species use genus or even family level).

Response: Again, we would like to remind that we did not use BLAST because the short length of the environmental sequence reads makes it possible to employ global alignments. As explained in the article Pawłowska et al. (2014; *Geobiol.*) and several others (Pawłowski et al. 2014, *Biol. Bull.*; Pawłowski et al. 2014, *Mol. Ecol. Resour.*; Esling et al. 2015, *Nucleic Acids Res.*; Pochon et al., 2015; *Marine Poll. Bull.*), we did account for the cases when an OTU shows the same highest similarity with multiple species, by assigning the OTU to the taxon that makes consensus among all the taxonomies of multiple species. In fact, our approach is even more conservative because we kept all the reference sequences that match within the threshold of 5 differences, and hence not only the reference sequences that would show the highest similarity. This means that if an environmental sequence is 100 % similar to a reference sequence belonging to genus A/species A but that there is a reference belonging to genus A/species B that is distant from A by less than 5 differences, then the environmental sequences would be assigned to genus A only, because of the conflict between species A and species B.

Interactive comment on *Clim. Past Discuss.*, 11, 3665, 2015.