Referee comment 3

We would like to thank anonymous referee 3 for his/her thoughtful comments and suggestions which will improve the clarity and the quality of the paper. We are especially thankful that he/she shared his/her knowledge about chloroplast genomes in pollen in the discussion forum. Below, we addressed all comments and questions.

I read with interest this manuscript that offers interesting results in good agreement with similar recent studies conducted by other researchers on similar settings from lakes and peat. The manuscript presents the results of ancient pollen and sedimentary DNA study extracted from permafrost from the Buor Khaya Peninsula. It shows several important methodological results of interest for readers working with plant ancient DNA from bulk sediments. It shows first that combination of classical palynological analyses is often not sufficient to resolve flora composition and that DNA is an important complementary tool to investigations of this kind. It confirms (and this should perhaps be strengthened even more by the authors) that DNA signal is local in origin and this fact helps when combined with pollen proxies that we know may have limitations linked to the fact that pollen is transported over long distances. Among papers that should be cited for this result are: Pedersen et al 2016, Sjögren et al 2016 and Parducci et al 2013.

We thank Referee 3 for his/her positive feedback. We will strengthen that our results suggest the local origin for sedaDNA in this study and its benefits in combination with palynological analyses. We will include the recommended papers.

I am less familiar with the type of statistical analyse the authors used and I cannot comment on that in details. I can perhaps comment generally that they seem to me to overanalyse and over interprete their data. For example I didn't understand why to calculate terrestrial-aquatic and Poaceae-Cyperaceae ratios as well as the PCA analysis if these calculations are not later even used and discussed.

We understand this concern and will discuss this now thoroughly in the manuscript. We are aware that it is difficult to use sedaDNA in a quantitative way due to technical biases. However, we were conservative in our bioinformatic filtering, in our threshold to include taxa in the statistical analysis and in comparing taxonomic richness between samples. Furthermore, we do not rely entirely on sedaDNA but compare it to pollen. Nevertheless, we will discuss this issue and will describe our inferences more carefully.

In some palynological studies the ratio of Poaceae to Cyperaceae is used to assess temporal changes in humidity. We applied this ratio to trace the hydrological development along the core since our results suggested the local presence of a shallow water body in several depths. However, since some Poaceae (e.g. *Arctophila fulva*/*Dupontia fisheri*) are associated with wet conditions while the Cyperaceae *Kobresia* is an indicator for (cryo-) arid climate we chose to compare this ratio with the terrestrial-aquatic ratio. Additionally, the ratios were an instrument to visualize and compare the contribution of terrestrial and swamp/aquatic taxa between samples as well as between the two proxies. We hope that we were able to make our intent clearer.

I reply here also to one of the other referee to state that pollen contains all three plant DNA genomes (mitochondrial, nuclear and chloroplast) regardless of the type of inheritance of the organelle DNA of the mother plant (paternal or maternal). There is however an increase or a decrease of the mitochondrial or chloroplast organelles in pollen after mitosis 1, depending on inheritance. So Pollen contains cpDNA as well as mtDNA in addition to the nuclear of course. This has been demonstrated in a number of different experimental studies in the lab conducted on fresh pollen (e.g. Nagata et al 1999) as well as in sito in ancient sediments: cpDNA has been amplified from pollen of plants with maternal inheritance and vice versa for mtDNA in conifers for ex. It is now know, but it still require some studies, that plant and in sediments comes

mainly from plant tissues like macros and that likely it mainly present in its extracellular forms in sediments (linked and protected by solid particles). However this is a field that requires lots of studies in the future: taphonomy of plant ancient DNA in sediments.

Thank you very much for this statement and the additional information which you provided in the discussion forum. We agree and are excited to read your paper when it is published.

It is not clear in the results how many samples are analysed for DNA. The authors write of extraction batches of 11 samples. But how many in total? In that paragraph this should be stated more clearly. Were all 32 samples analysed for pollen also analysed for DNA?

We agree, and included a sentence on Page 6, Lines 24-26:

"All 54 samples were analysed for sedaDNA while only 32 were processed for pollen analysis."

The authors write that they run up to 4 PCR per samples. This is not very good if it means that some samples perhaps have been amplified only one for example. State in a table how many PCR per sample or state the minimum number amplified. Why not amplify the same number for all samples?

We are sorry for this confusion. We pooled exactly two PCR-products for each sample. However, in a few cases up to 4 PCRs were necessary to get two positive PCR-products with clean PCR negative controls. We changed the corresponding sentences from: "All samples were subjected to up to four PCR runs. Once two positive amplifications were obtained, while the associated NTC and extraction blank remained negative, they were used for sequencing."

To:

"For each sample we pooled two positive amplifications for sequencing, under the condition that the associated NTCs and extraction blank were negative."

Did the authors sequence the positive blanks?

The extraction blanks as well as the PCR negative controls were all negative, indicated by the gel-electrophoreses after the PCRs. Nonetheless, we sequenced all controls. On page 8 line 5-6 of the first version we wrote: "All extraction blanks and NTCs were included in the sequencing run, using a standardized volume of 10 µl, even though they were negative in the PCRs." Furthermore, the (unfiltered) sequencing results of extraction blanks and PCR controls are reported in Supplement S6. We clarified this by adding the sentence:

"The sequencing results of extraction blanks and PCR controls are reported in the supplementary file S4."

Since we will refer to this supplement in an earlier section now, the supplement order will be changed accordingly to S4 instead of S6.

Paragraph 6-14 on page 8 should be moved earlier in the paragraph.

We will move the paragraph to the beginning of part 3.3.

I did not understand the sentence on page 18 line 18: The sedDNA record does not contain conifer-derived DNA (with reference from Birks and van der Knaap). Why are these two reference are used here? To confirm what, that DNA of conifer is not detectable by sedDNA. this is not the case. Not clear.

We agree. This is a remnant from an older version of this sentence and we rephrased the sentence:

"In most of the samples we did not detect conifer-derived sedaDNA, although they are present in the pollen record. The conifer pollen presumably originated from extra-regional stands through long-distance dispersal (Birks, 2001; van der Knaap, 1987)."

Page 19, line 17-18. the authors should consider to use MapDamage 2.0 to test for patterns and rates of DNA damage and assure for authenticity of their reads.

We considered using mapDamage 2.0 but came to realize that the program is not adequate for metabarcoding. mapDamage 2.0 requires a (genomic) reference sequence to which the reads of interest are aligned. From this alignment it uses a Bayesian approach to estimate nucleotide misincorporation rates and fragmentation patterns (Jónsson et al., 2013).

The first problem we encountered is that the amplicons we use are highly variable and do not allow for an overall alignment, even if we use closely related taxonomic subsets. The second problem is that the numbers of sequences without an exact match to their corresponding reference sequence are too low to produce an error distribution. The third problem is, that in our study we analysed only sequences with an exact match to a reference sequence. Therefore, we cannot really "prove" that these sequences are authentic. mapDamage would (theoretically) only prove that the sequences, which we discarded are probably authentic. However, if you have any suggestions on how we could apply mapDamage appropriately, we are very interested.

The manuscript and especially discussion is long and can be shortened in my opinion.

We will shorten the manuscript accordingly.

References

Birks, H. H.: Plant macrofossils, in Tracking Environmental Change Using Lake Sediments, vol. 3 Terrestrial, Algal, and Siliceous Indicators, edited by J. P. Smol, H. J. B. Birks, W. M. Last, R. S. Bradley, and K. Alverson, Springer Netherlands, Dordrecht. [online] Available from: http://link.springer.com/10.1007/0-306-47668-1 (Accessed 29 July 2016), 2001.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics, 29(13):1682-1684. doi:10.1093/bioinformatics/btt193, 2013

van der Knaap, W. O.: Transported pollen and spores on Spitsbergen and Jan Mayen, Pollen Spores, 24, 449–453, 1987.