Response to Reviewer Comments on Manuscript bg-2019-144: "Microbial community composition and abundance after millennia of submarine permafrost warming"

Comments of Reviewer #2:

We would like to thank the reviewer for evaluating our manuscript positively and giving such useful and important advises and comments, especially those which helped to improve our knowledge on statistical analyses and the application of them in this manuscript.

general comments

The manuscript of Mitzcherling et al. describes a field survey in the arctic, which tested the hypothesis that the effect of permafrost warming can be examined already before the thawing starts. Intriguingly, the experimental design was to use frozen sediment cores of diverging base temperatures ranging between -12° and -1.4 °C, but from the approximately same age. This is a very clever setup, however, the implementation of this was limited by using only four sites (maybe because of the costs and logistics of such an expedition), which also limits the statistics and the final conclusions. The authors tried to compensate the limited sampling by taking 6-10 replicates (which could be statistically interpreted as pseudoreplicates) per site from the targeted frozen period, but in the end could only see a moderate, and even negative effect of temperature on the microbial parameters. Furthermore, other factors such as depth and potential differences in the palaoenvironmental origin obscured the temperature signal, which the authors discussed, accordingly. In general, the study is cleverly designed and provides new research concepts for studying permafrost changes over long time scales. Although the results leave room for discussions due to the limited sampling sites, this work is an interesting study, and a good basis for future studies within the same setting.

specific comments

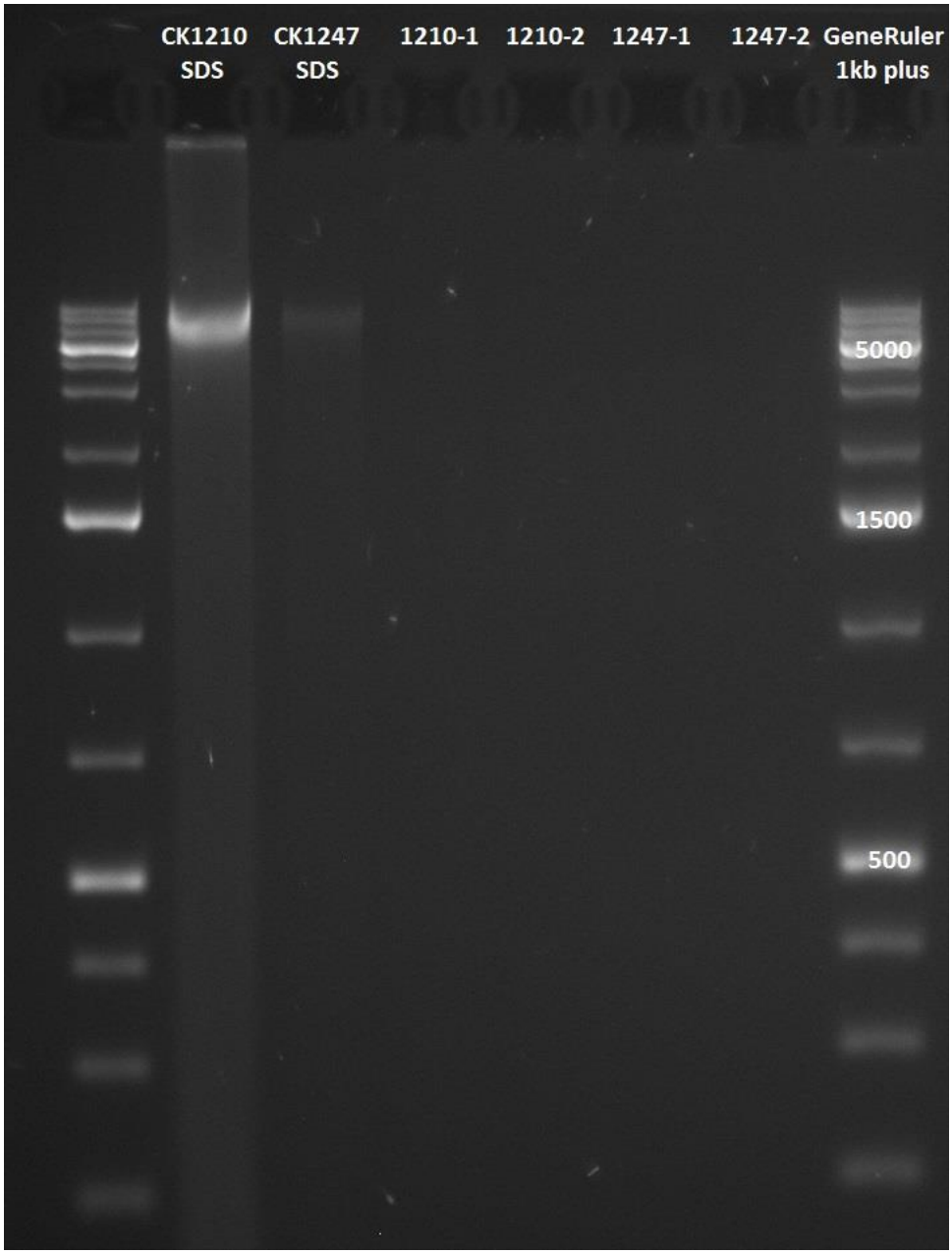
Page 4 – Age measurements: The authors estimated the age of the core profiles, but it wasn't included in the statistics (or at least I couldn't find it); I would assume that age could explain part of the variation in the microbial community composition.

We agree with this idea. Including the sediment ages in the statistics is, however, difficult as sediment age was determined in only a small number of sediment horizons which are different from the samples analyzed here. Furthermore, different dating methods were used across cores and the transect, and the age dates of the different methods differed strongly from one another (Winterfeld et al., 2011). However, sediment depth could be seen as an analogue for age and depth explains a large part of the variation in the microbial community.

Page 5 – DNA extraction: Could the authors maybe in the supplement provide a gel picture of the extracted DNA? I am asking this, since the fragmentation of the DNA can also be seen as an

indicator for the presence of non-cellular ancient DNA (aDNA). In addition, could the authors please specifiy which size fraction was extracted from gel?

We exemplarily provide the following gel picture of two samples from core C2 (CK1210 SDS: 0.05 m bsf, 265 ng/g and CK1247 SDS: 52.7 m bsf, 33.4 ng/g) and add this to the supplement. The two examples show that there is not much fragmentation likely due to constantly freeze-locked conditions. Hence, we used the DNA extracts without gel purification for downstream analyses.



Page 6 – HTS: Please keep in mind that 35 PCR cycles is an unusual high cycle number for an amplicon based microbiome analysis, which will probably cause larger shifts in relative abundance values of microbial groups (this may become relevant when you try to implement some of the RC1 comments).

We thank the reviewer for this reminder. In low biomass environments like permafrost it is, however, difficult to amplify a sufficient amount of DNA for sequencing. There are a number of other studies which implemented the same or even a higher number of PCR cycles in their work (Koebsch et al., 2019; Wen et al., 2018; Winkel et al., 2018, 2019).

Page 7 – Multivariate Statistics: To me the authors used a suboptimal set of statistical methods for analysing the microbial community composition. While Mantel tests are good for testing the correlation of two matrices (e.g. the community matrix with the environmental matrix or a subset of parameters (e.g. a matrix of depth, temp, stable isotopes), it is rather uncommon to use it for testing single parameters. (As a side mark, the Mantel test can be performed rank-based or parametric based, this should be specified). Current alternatives to Mantel tests are PERMANOVA variants that are suitable for continuous variables, or for a priori hypothesis such as the temperature hypothesis distance based redundancy analysis followed by an ANOVA test. However, since the authors start with an exploratory analysis, one of the most frequent used methods is to fit the variables into the ordination by regression/correlation. CCA may not be the best option in this case, but a PCoA or an NMDS would be the preferred method.

We thank the reviewer for giving these important advices. We changed the analyses as suggested and changed the text and figures accordingly:

The description of the statistical analysis in the methods part was changed as follows (page 6 line 29):

Variation in $OTU_{0.03}$ composition, 16S rRNA gene and total cell abundance between samples and among drill sites, as well as correlations of the abundance and $OTU_{0.03}$ composition with environmental parameters were assessed using the Past 3.14 software (Hammer et al., 2001) and **R, especially the vegan and MASS packages.** Principal component analyses (PCA) based on Euclidean distance were used to assess variation in environmental variables across the different sediment units and within Unit II. Prior to analysis, all environmental data were standardized by subtracting the mean and dividing by standard deviation. To assess the correlations of bacterial and microbial abundance with environmental parameters the rank-based Spearman correlation was calculated. **The Bray-Curtis dissimilarity was used to assess the beta diversity of the microbial communities in a non-metric multidimensional scaling (NMDS) plot. Environmental factors that might influence its composition were determined by an environmental fit into the ordination. The significance of the variance introduced by the identified environmental factors was tested using a permutational approach as implemented in the adonis function of the vegan package. Factors were tested for auto-correlation as implemented in the corrplot package. A linear model of the remaining factors was subject to a redundancy analysis which was tested for significance using the analysis of variance (ANOVA).**

The produced NMDS plot replaced the CCA in Figure 6.

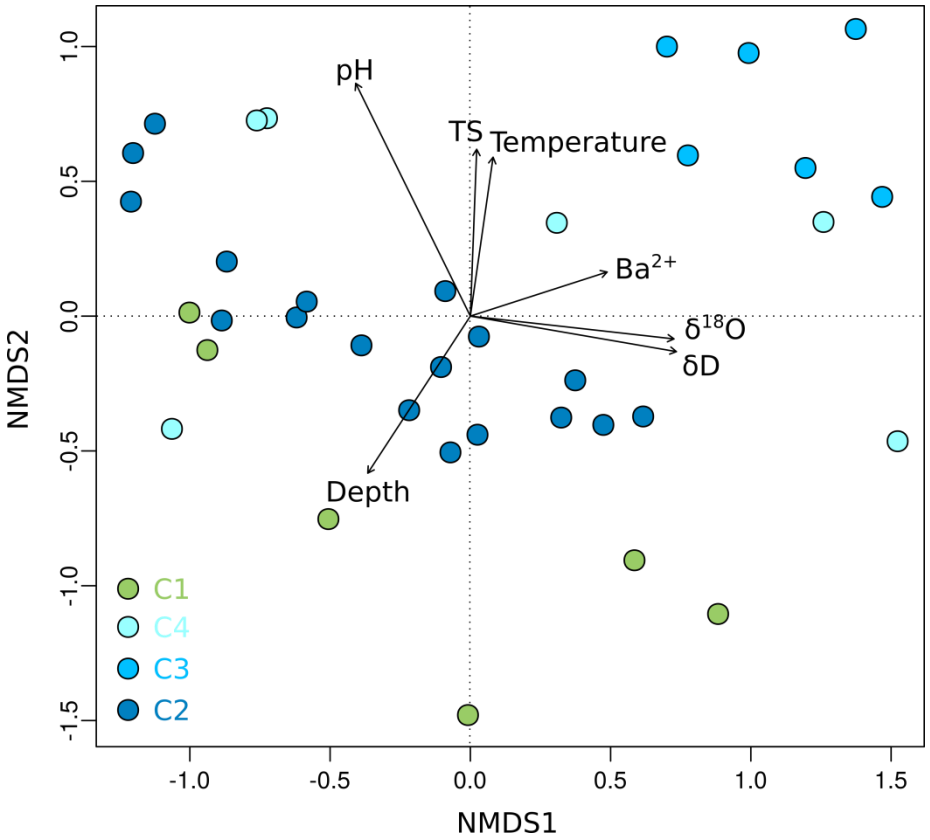


Figure caption: Non-metric multidimensional scaling (NMDS) plot of $OTU_{0.03}$ data from Unit II in dependence on environmental parameters. Shown are environmental factors that contribute significantly ($p < 0.05$) to the variance of the community data. The stress value of the NMDS plot is 0.13.

Environmental factors that might influence the microbial community composition were tested by Permutational MANOVA. This table replaced the Mantel tests in table S11:

	Dim1	Dim2	r^2	p-value
Depth [mbs/msbf]	**-0.53174**	**-0.84691**	**0.3322**	**0.006**
Temperature	**0.13632**	**0.99067**	**0.2487**	**0.015**
Ba	**0.94807**	**0.31805**	**0.1859**	**0.031**
Si	0.90304	-0.42956	0.1541	0.056
Ca	0.50032	0.86584	0.01	0.835
K	0.81761	0.57578	0.0612	0.341
Mg	0.80879	0.58809	0.0684	0.297
Na	0.99177	0.12804	0.0813	0.241
Nitrate	-0.8121	0.58351	0.028	0.637
Chloride	0.98966	0.14344	0.0527	0.391
Sulfate	-0.28689	0.95796	0.1014	0.161
Bromide	0.92727	0.37439	0.0629	0.326
Salinity	0.99532	0.0966	0.0459	0.443
δ18O	**0.99329**	**-0.11569**	**0.3753**	**0.001**
δD	**0.9843**	**-0.17648**	**0.3914**	**0.001**
pH	**-0.42785**	**0.90385**	**0.6412**	**0.001**
TC	0.41379	-0.91037	0.1053	0.149
TN	-0.38942	-0.92106	0.0268	0.640

TS	**0.03653**	**0.99933**	**0.2694**	**0.004**
TOC	0.40692	-0.91346	0.0974	0.170
Clay	0.47503	0.87997	0.1123	0.132
Silt	0.76336	0.64597	0.0532	0.405
Sand	-0.70792	-0.70629	0.063	0.330
Conductivity	0.98987	0.14199	0.0419	0.478

As the orientation and the distribution of the OTU data changed slightly, we adjusted the description of the results as follows (page 9 line 3-17):

Grouping patterns of the bacterial community based on the $OTU_{0.03}$ composition of the samples and the Bray-Curtis dissimilarity were visualized **using a non-metric multidimensional scaling (NMDS, Fig. 5)**. The **NMDS** showed a clustering of samples according to their borehole location for C2 and C3, while communities of C1 and C4 were more scattered. **We fitted environmental gradients with the NMDS ordination in order to test for correlation between the bacterial community compositions at each drill site with environmental parameters ($p < 0.05$). Samples located at the bottom left of the plot originated from a greater depth (C1 and C2) than samples to the top right (C3 and C4). Variance of samples from the bottom to the top was explained by rising pH, temperature and total sulphur content, while variance of samples from the left to the right side are likely explained by increasing values of the stable water isotopes $\delta^{18}O$ and δD, and Ba^{2+}. The bacterial community of C3 was most distinct and clustered furthest from communities of all other sites, and was linked with stable water isotopes, Ba^{2+} and sample depth. The variance between C1, C4 and C2 samples are explained by temperature differences. A subsequent permutational analysis of variance showed that depth, temperature, pH, TS, δD, $\delta^{18}O$, and Ba^{2+} contribute to the variance in the microbial community composition (Table S11), whereof $\delta^{18}O$ and δD show a high auto-correlation. A redundancy analysis showed that the explanatory variables depth, temperature, pH and $\delta^{18}O$ significantly explain parts of the variance in the microbial composition ($p = 0.001$).**

Finally, we added some discussion on the pH as one of major factors shaping the bacterial community in submarine permafrost samples (page 11 line 7).

The strongest correlation of the bacterial community composition was, however, found with pH. Soil pH is a major factor controlling the bacterial diversity, richness and community composition on a continental scale (Fierer and Jackson, 2006; Lauber et al., 2009; Rousk et al., 2010). On a global scale pH is also one of the major controls of archaeal communities (Wen et al., 2017). Fierer and Jackson (2006) showed that the richness and diversity of bacterial communities differed between ecosystem types, which could be explained by pH. This substantiates our suggestion that Unit II and the bacterial community therein was formed under different paleo-climatic conditions and varying landscape types during the last glacial cycle.

Page 7 – statistics: Isn't the Dunn's test the PostHoc test for non-parametric tests such as Kruskal-Wallice? For an ANOVA, I would have expected a Tukey-HSD. Please doublecheck.

We thank the reviewer for pointing to this mistake. It is absolutely right, that the post-hoc test of ANOVA should be the Tukey's test. We changed this accordingly in the supplementary:

Table S1: Analysis of variance (ANOVA) of DOC concentrations between all four cores and Tukey's pairwise post-hoc test with p-values according to Copenhaver-Holland above and the Tukey's Q below the diagonal.

	Sum of squares	df	Mean square	F	p (same)
Between groups:	24714.2	3	8238.06	4.814	0.003712
Within groups:	155731	91	1711.34		**Permutation p (n=99999)**
Total:	180446	94			0.02357

	C1	C4	C3	C2
C1		0.066	0.996	0.052
C4	3,540		0.299	**0.002**
C3	0.310	2.490		0.739
C2	3.676	5.209	1.441	

… and in the text page 10 line 27 :

In detail, TCC in C2 were higher than in the other submarine cores while DOC values were lower in C2**, significantly different from C4 and C1** (Table S13).

Page 7 – General statistics: The authors will need to think about corrections for multiple comparisons. In particular in Table 1 or for the Mantel tests presented in the supplemental, which will both require p-value corrections (e.g. using Bonferroni). If these corrections are not done, this has to be stated explicitly.

We thank the reviewer for this remark. As the Mantel tests were replaced by a PerMANOVA we performed a p-value correction only on table 1 according to Holm. The results roughly reflected the original results but with less significant p-values. The correlation of abundance measures to salinity became restricted to 16S rRNA gene copies normalized to DNA g^{-1}. We weakened corresponding statements about the significance in the text and rephrased the following sentences.

Abstract (page 1 line 29):
Based on correlation analysis TCC unlike bacterial gene abundance showed a significant rank-based negative correlation with increasing temperature while bacterial gene copy numbers showed a strong negative correlation with salinity

Discussion (page 10 line 8)
Besides permafrost warming changing pore-water salinity had an effect on the microbial abundance. Rising permafrost temperature strongly correlates with TCC whereas salinity correlates strongest with bacterial gene copy numbers (Table 1). Bacterial 16S rRNA gene copy numbers were lowest in core C4 (10^5 gene copies), where pore-water salinities were elevated (electrical conductivity values >2000 µS cm^{-1}, Table S3).

We furthermore indicated in the caption of table S10 that a p-value correction was not performed here.

Values in bold are significant (< 0.05) when omitting p-value corrections.

Page 8 – Line 16: Please state the correlation values between DNA, copy numbers, and cell counts. This may be important to interpret Table 1.

As also reviewer 1 asked for the correlation between the abundance measures we added detailed information about p-values and correlation coefficients to table S10 and the corresponding information to table 1 showing the results of the rank-based Spearman's correlation.

	16S Bacteria	16S/DNA	TCC	Temp	Salinity	Depth [mbs/ mbsf]	Ba^{2+}	Ca^{2+}	K^{+}	Mg^{2+}	Na^{+}	Cl^{-}	SO_4^{2-}	Br^{-}	$\delta^{18}O$	δD	pH	TC	TOC	Grav. Water Content
DNA	**0.87**	**0.47**	**0.68**	**-0.37**	**-0.35**	0.30	-0.08	-0.09	**-0.39**	-0.32	**-0.39**	**-0.43**	-0.14	**-0.41**	**-0.37**	-0.33	**-0.44**	**0.40**	**0.34**	**0.47**
16S Bacteria		**0.79**	**0.61**	-0.24	**-0.48**	**0.51**	0.03	-0.20	**-0.49**	**-0.46**	**-0.57**	**-0.56**	-0.26	**-0.54**	**-0.38**	-0.33	**-0.52**	**0.44**	**0.39**	**0.47**
16S / DNA			**0.36**	-0.12	**-0.63**	**0.47**	0.04	**-0.47**	**-0.55**	**-0.60**	**-0.71**	**-0.67**	**-0.40**	**-0.66**	-0.16	-0.11	**-0.54**	0.21	0.19	0.26
TCC				**-0.64**	**-0.44**	0.26	**-0.38**	-0.23	**-0.42**	-0.37	**-0.50**	**-0.52**	-0.09	**-0.50**	**-0.37**	**-0.37**	-0.28	0.06	0.14	0.16

Page 9 – curiosity comment: The authors took a vertical profile of each core, but this is not really implemented in the study. Out of curiosity: Do the points in e.g. the CCA or the PCA also structure according to the vertical profile? If so, this could be an interesting aspect that may also explain some of the variance observed, caused by differences in ages and/or the paleoenvironment.

For the CCA (now NMDS) this is unfortunately not the case. The sample points do not show any gradient according to depth location. In the PCA it is hardly visible if there is any structure according to the vertical profile, especially in the cluster of samples from Unit II.

Page 9 – Discussion: The discussion is rather comprehensive and understandable. I think, I can agree with the arguments of the authors. Please include a brief discussion on the limited sampling design of only 4 sites. With such a high variation between the cores, it may require > 30 cores to really answer the hypothesis.

We added a short discussion on the limited sampling size that followed the discussion on the correlation of bacterial community composition and pH.

However, the limited number of environmental samples and the inference of other correlating environmental factors might decrease the statistical powers to see a more significant effect of temperature on the microbial community.

Figures:

Optional comment: Since DOC became important for the discussion of the cell counts (and equivalents), maybe it could be worthwhile to include a figure on this in the main text. Please discuss this among yourselves.

We are happy about the reviewers suggestion but decided to leave the figure in the supplemental as we already have a sufficient number of figures in the main text. The supplemental is open and easily accessible for everybody who is interested in more details.

technical corrections

Page 6 – please specifiy which Illumina MiSeq chemistry was used (2x 250 or 2x300nt?)

Due to the wish of reviewer 1 for more detailed information on the sequence analysis and bioinformatical tools we decided to also add details on the PCR conditions, library preparation and sequencing instead of referencing them. We changed the paragraph as follows:

2.7 High throughput Illumina16S rRNA gene sequencing

Sequencing of each sample was performed in two technical replicates. Primers comprised different combinations of barcodes (Table S6). PCR amplification was carried out with a T100™ Thermal Cycler (Bio-Rad Laboratories, CA, USA). The PCR mixtures (25 µl) contained 1.25 U of OptiTaq DNA Polymerase (Roboklon), 10x concentrate buffer C (Roboklon), 0.5 µM of the sequencing primers S-D-Bact-0341-b-S-17 and S-D-Bact-0785-a-A-21 (Table S5), dNTP mix (0.2 mM each), additional 0.5 mM of $MgCl_2$ (Roboklon), PCR-grade water, and 2.5 µl of template DNA. **PCR conditions comprised an initial denaturation at 95°C for 5 min, followed by 35 cycles of denaturation (95°C for 30 s), annealing (56°C for 30 s) and elongation (72°C for 1 min), and a final extension step of 72°C for 10 min. The PCR products were purified from agarose gel with the HiYieldPCR Clean-Up and Gel-Extraction Kit (Südlabor, Gauting, Germany) and were quantified with the QBIT2 system (Invitrogen, HS-Quant DNA). They were mixed in equimolar amounts and sequenced from both directions (GATC Biotech, Konstanz) based on the Illumina MiSeq technology. The library was prepared with the MiSeq Reagent Kit V3 for 2× 300 bp paired-end reads. The 15% PhiX control v3 library was used for better performance due to different sequencing length.**

Page 6 – brackets are falsely set in line 13 for Llobet-Brossa et al. 1998

Preparation and quantification of the total cell abundance per g sediment were performed after **Llobet-Brossa et al. (1998)**.

Page 7 – line 19: Please indicate the PSU of the seawater in this area

The seawater in the Arctic Ocean is mostly stratified due to the freshwater inflow of large Arctic rivers. Thus, the bottom water salinity is the best to represent the Arctic Ocean waters (Guieu et al., 1996). Bottom-water salinity at the drill sites at Cape Mamontov Klyk were measured in the framework of the drilling campaign and were around 30 PSU. We added the information to the text as follows:

In C4, the drill site located closest to the coast, Unit II had the highest pore water salinity (mean = 5.6 PSU) ranging from 0.9 to 17.6 PSU (Table S2), which spans freshwater to mesohaline water but is much below seawater salinities. **In comparison, bottom-water salinities at the drill sites ranged between 29.2 and 32.2 PSU (Overduin et al., 2008)**.

<u>Figures:</u>

Figure 4: Please indicate the number of measurements for each box (n= : : :)

We are happy to provide the number of samples for each box. The number of technical replicates measured for each sample is provided in the figure caption.