SOIL Discuss., referee comment RC1
https://doi.org/10.5194/soil-2021-14-RC1, 2021

Comment on soil-2021-14

Grace Pold (Referee)

Referee comment on "Whole-soil warming decreases abundance and modifies the community structure of microorganisms in the subsoil but not in surface soil" by Cyrill U. Zosso et al., SOIL Discuss., https://doi.org/10.5194/soil-2021-14-RC1, 2021

Cyrill Zosso and colleagues present an interesting look into how deep soil warming alters microbial biomass and high-level community abundance in a forest ecosystem. Overall, I think this is a very interesting paper and applaud the authors for their use of lipids rather than getting stuck in the mud with sequencing, as it provides a different perspective on warming impacts on microbial communities than is typically used. The paper is also well-written and flows well, and was enjoyable to read. There are a few areas I think the manuscript could benefit from clarification on. In particular, how and/or why certain comparisons for depth*warming interactions were made, and a better integration and justification for the brGDTs in hypothesis testing.

Main comments:

The scale on which the data are presented does not seem to match the scale on which the analyses were completed. It would be very helpful to post the R scripts with the manuscript. I found it hard to follow what was being included as a random effect (or whether random effects were nested) in the different results. Was a model of form "lipid ~ warming *depth + (1|block/warm)" or "lipid ~ warming *depth + (1|block/depth)" fit? Or something else?

Related, there seem to be a lot of post-hoc tests, but it is not always clear how these post-hoc tests were selected for completion. Why is the top 10cm sometimes compared to the bottom 10cm, and other times some intermediate depth compared to the deep? Or why is the cutoff for depth 30cm in some instances, and 50cm for others (ex. total PLFA vs. Actinobacteria)? Why is there only one p-value reported (ex. L229-230) when looking at all the depths below a certain point, and not one for each of the depths analyzed as show in the figure? Was the total lipid below a certain depth summed for each core to complete this analysis? [I think this would make sense from a statistics standpoint]

Second, please add more justification for why brGDGTs were measured (in terms of specific hypotheses) and how they should be interpreted. The authors mention that they turnover slowly…is this why they were chosen? If so, what does this mean for interpreting the data from a 4-year warming study? Since these are predominantly necromass, is the idea that microbes under heated and control conditions might be preferentially consuming these or warming might accelerate their turnover? Or is the idea to try and see if there is a signal in the brGDGTs that might indicate how microbial communities have overall changed in the past 4 years, which is not visible in the more rapidly cycling PLFAs? Or is the idea to capture predominantly the archaea community, which would not be captured by PLFAs?

Minor comments:

Since the soil is warmed by a smaller degree in the top 20cm compared to below that, why not just discard the shallow soil data since it cannot fairly be compared with the deeper samples? Also, since lipids were extracted from less soil in shallow compared to deep samples, the deep samples could just be more representative of deep soil and therefore easier to detect a difference in.

L135: why add the standard in after lipid extraction, and not directly to the soil so that the authors could get a better idea of extraction efficiency for the different soil depths?

Could figure 4 be presented as an ordination instead? There is a lot to digest here.

Grammar/style

The "gram" in Gram positive/negative should be capitalized, as it refers to someone's name (Hans Christian Gram)

L177: please mention what kind of post-hoc test was used.

L281: correlations are generally reported as R (for Pearson) or rho (for Spearman); R^2 is the coefficient of determination.

L300: there are also a lot of unknowns with respect to extraction efficiency of chloroform fumigation extraction. It almost certainly underestimates the C content of high surface area:volume small cells, as it predominantly captures cytoplasm.

Please also make sure to report the F/T/Z statistic and degrees of freedom, preferably in the text, or otherwise in a supplementary table.

Whenever someone says something increased/decreased with depth, it sounds like depth has been treated as a continuous rather than categorical variable. So please try to avoid this.