**Reviewer #1**

*The research article entitled ‘Environmental factors driving spatial heterogeneity in desert halophile microbial community’ and submitted by Uritskiy and collaborators aimed at deciphering the drivers of prokaryotic communities of halite from the Atacama Desert at different scales. I have some mixed feelings regarding it and this is the main reason why it has taken me quite a long time to review it (I would like to apologize for this). It is indeed a very interesting study given the type of samples collected and the abiotic data recovered. However, I think that the analyses regarding the microbial communities c/should be improved prior to the publication. Therefore, I recommend that in this form the article cannot be published and will describe thereafter why.*

Given that only the 16S rRNA gene diversity was analyzed it is difficult to write that microbial communities were studied; particularly in the title. So this should be clarified. And to continue with the title, the concept of ‘halophile community’ is I think too vague as generally desert environments, and particularly their soils, are rich in salts. I would recommend to be more specific and clearly state that halites are studied here.

We strongly disagree with the statement that microbial communities cannot be studied using 16S rRNA gene analysis. As described in Knight et al., 2018 (<https://doi.org/10.1038/s41579-018-0029-9>) and Goodrich et al., 2014 (<http://dx.doi.org/10.1016/j.cell.2014.06.037>), marker gene analysis is indeed a valid method to investigate microbial communities. We would like to keep the title as is because “halite” is a rather specialize term that might not be recognized by non-specialists.

**Introduction**

L46-49: Nitrogen is generally described as the most important limiting factors, after water, for microbial communities in deserts. So this should be added.

This was added to the MS

L51-53: This is a very risky comment as, in deserts, Scola et al 2018 – Microbial Ecology has performed such analyses. Also this sentence made me expect analyses like ‘variation partitioning’ to actually clearly evaluate the importance of stochastic vs deterministic processes in determining the assembly of the halite communities. I hope in the next version to see such analyses.

We did not claim that there were no studies but rather “relatively few studies” and cited Scola et al. 2018, along with two other papers. Regarding the work presented in this manuscript, our analysis and conclusions reflect our data and the samples we were able to collect, given the limitation of doing fieldwork in the Atacama Desert. A comprehensive “stochastic vs deterministic” statistical analysis was not in the scope of this study, as explained below.

L62-65: I really do not understand how and why halite microbial communities represent compelling models for climate change studies. Indeed, both the halite niche and its community are so peculiar that making conclusions interesting at the global scale very difficult. Therefore, I suggest either the authors to remove this sentence or to explain better the rationale behind their thoughts here.

We clearly defended this position in the introduction by stating that halite nodules are 1) spatially isolated and thus provide greater resolution, and 2) are under a clear and acute limiting factor (water) and are thus very sensitive to minor changes. It is indeed essential to investigate the resilience of communities in hyper-arid deserts, as these are fragile ecosystems and their inhabitants at the front line of major changes in climate ahead of us.

L85: why the ‘and’ in “[…] photosynthesis and carbon fixation […]”?

We changed to photosynthesis in the text.

**Mat/Met**

As a general comment, it’s very difficult to find some of the information stated. I would strongly recommend to be specific on where exactly to find the data in the Supp Data 1. Could you please clearly state the exact number of samples processed?

The three sections in Supplementary Data1were re-labeled and we added the corresponding labels to the Mat/Met section text.

This information was added to the text.

It is written that some halites were collected in February 2016 and some in February 2018. But some were also collected in August 2017. Why is it not specified here too?

This information was added to the text.

All samples were collected in February and while sample names for the “Large scale” sampling effort were accurate (i.e. Feb7-1), we found mistakes in the sampling dates (column E) in Data S1. The metadata was corrected and we understand how these errors impaired the understanding of our sampling efforts and dates.

I was wondering if there is a chance or not that this wide temporal scale of sampling may have had an impact on the communities. Could you please test for this? Particularly as the August and February samples originate from two different seasons which can be drastically different in desert environments.

The scale of this study was such that the three major scales of diversity could not be sampled in one field season, but instead were divided into three distinct sampling efforts over three sampling seasons (i.e. three years), all in February (see above). It is important to note that these three study groups were processed and analyzed separately, without directly comparing them, i.e. we derive conclusions for each section independently. We did not discuss the three study groups together until the discussion, where we compare the conclusions from each distance scale to derive a more universal conclusion for the whole ecosystem. The only metric for which we felt justified to compare all the samples together was the community dissimilarly changes across increasing distances (i.e. Fig. 6). We added clarification on these caveats in the opening paragraph of the results.

All three sampling efforts were done in February – see above. We apologize for the mistakes in Data S1.

L120: the nodules were stored in the dark. Can this have an impact on the abundance of photosynthetic organisms?

All nodules were stored in a dark dry location to prevent nucleic acid degradation. There is no reason to believe that the DNA from photosynthetic organisms might be more liable to degradation under those conditions.

L136: It’s at the beginning of this section that it should be specified that light transmission measurements was performed in the lab. L151 is too late. On that note, in the supp material maybe, a picture of the setup could be a good addition. I’m just mentioning this as I found it a bit difficult to follow how it was performed (I’m obviously not a specialist).

This was added at the beginning of the section. A photo of the experimental set up was added in supplementary material.

**Results**

As a general remark, I find that some section/sentences could be moved to the mat/met (eg: L30-305). So this should be checked for.

This section was moved to Materials and Methods. However, because of the complexity of our sampling strategy and the multiple scales of diversity we are addressing, we think that reminding the reader of our experimental approach is beneficial to his/her understanding of the results.

Also, I generally found that the authors tried to “squeeze” their data to the maximum to show something. And to my point of view, this is the weakness of the paper. At one point, it made me doubt about the results. This is why after thinking about it for quite some time (sorry again for the delay), I have btw decided not to comment as yet on the discussion. So this is something to be taken into account in the next version. I would recommend therefore to only use one (and only one). For ex: what is the conclusion of L410-413?

Sorry, but we do not understand what the reviewer means by “I would recommend therefore to only use one (and only one)”and therefore cannot reply to that comment.

Regarding the data in L410-413, it is an important bit of data that is not overlooked, but adds to the conclusions we draw in the discussion.

L250-256: Please remind here the number of samples used for each. I must also here mention that I was expecting some distance-decay type of analyses and was surprised not to. That could be a very powerful approach given the sampling strategy that covers from cm to many km of distances. I would really advise the authors to perform such analyses.

Number of samples were added.

Because the three major sampling efforts were conducted in three different years, we cannot collectively analyze them in a conventional distance decay analysis as a certain amount of year-to-year variation is expected. Instead, we made sure to process, analyze, and compare the three groups of samples only within themselves, and derive independent conclusions. We added clarification on this at the beginning of the Results section, and also clearly outlined this in the Methods. The only analysis in which we felt justified comparing all three sample groups together was in Fig. 6, where we compared intra-site dissimilarity metrics between the different scales.

L300: subset ? n=?

Details added to the text.

L299-311: I’m not convinced at all by this section. Everything seems arbitrary to my point of view. To keep using this approach (and therefore the conclusions related to it) every step must be better justified. The general randomness shown on Fig 4 supports this view I think. However, I was wondering if the amount of mDNA extracted g-1of halite could not be used as a proxy rather here?

We are surprised that this reviewer qualifies as arbitrary the approach we used to estimate the biomass in halite nodules. It is the same approach that was used by Finstad et al., 2017 (doi: 10.3389/fmicb.2017.01435) to estimate the biomass of halite nodules from Salar Llamara, another salar in the Atacama Desert. There are only two differences between the protocols: (1) we sliced the nodules vertically in 3 sections, and (2) we counted all cells, stained with DAPI, rather than just counting autofluorescence cells. We also used a minimum of 5 replicates in our analysis to calculate reliable deviation and averages.

We found that DNA recovery was even less reliable for these samples due to the difficulty of extraction. The cell count results were reproducible when extracted from the same location in the same nodule.

We agree that there is a high level of heterogeneity in Fig. 4 between samples from the same location but there are also very significant trends between locations; we addressed both points in our manuscript. Regarding cell numbers, there are three levels that could introduce heterogeneity in this study: 1) different halite nodules and different slices within the nodule, 3) different positions (top, middle, bottom) within the same nodule, and 4) technical replicates within the same position in the same nodule. We included a reasonable number of replicates at each of these levels to insure a good statistical analysis. The large amount of work required for each replicate is why we could not count cells for every sample in this study.

L324-334: this should be reworked. What is the conclusion? The description of the PCA analyses could be shorten too. For ex, PC1 and PC2 will always explain better sample variation than PC3 and 4... L327-329: What is the point of this sentence?

We simplified this section to make it more understandable. While we agree that PC1 and PC2 better explain sample variable, PC3 and higher can reveal valuable information for example regarding lowly abundant keystone species, information that is masked in PC1 and PC2 by the highly abundant taxa (this being all relative abundance, of course). While we are not claiming that Cyanobacteria, Actinobacteria, Nanohaloarchaea, and Proteobacteria are keystone species in halite nodule, their relative abundance was much lower than that of Halobacteria and Bacteriodetes, indicating why PC3 and PC4 might be more informative for those taxa.

On a final positive note, I’d like to mention that I particularly enjoyed the second section of the results (Difference …). Such data are amazing.

Thank you for this.

**Reviewer #2**

*Uristskiy´s et al manuscript analyzes the microbial diversity in several halite nodules across different spatial scales inside the Salar Grande, and attempts to investigate the forces involved in driving spatial heterogeneity. They sequenced the 16S gene of the communities and measured the relative humidity and temperature in each site, and the relative humidity, temperature and light transmission inside the nodules to estimate the available water and photosynthetically active radiation. In addition, they measured the biomass by DAPI cell count. After that, they performed statistical comparisons at different diversity scales and at different spatial scales. They concluded that water, light, and community drift impact microbiome assembly differently at different distance scales. Water availability is more important in determining the microbiome assemblies from distant crusts as light availability is in determining the different assemblies inside the nodules. The impact of the architecture of the substrate was also inferred considering the distribution of the biomass and the light availability. Furthermore, they inferred that progressively smaller scales of diversity become less dependent on stochastic processes (L 547-8). The relative abundances of the major six phyla were analyzed in order to explain the reason for the differences in the microbial structures. The main strengths of the manuscript are the systematic sampling strategy, the homogeneity of the system selected for the study that allows us to infer that there are no other geochemical and mineralogical factors involved in shaping the microbial assemblies, and the statistical approaches used to answer the specific question addressed. The manuscript is very well organized and written and the conclusions are mostly supported by the results.*

I couldn´t find the statistical methods used to link environmental conditions to the observed community composition patterns. They compared the patterns of sites with different RH and temperature and they found that those are significantly different but they do not correlate those patterns with the environmental conditions. In that case, why don´t you consider that temperature is the determining factor, instead of water?

Because of the challenges in collecting the samples, we did not collect samples from a wide enough range of locations to conduct a statistically rigorous correlation analysis of microbiome composition and environmental variables. The sample collected for the three main distance scales (intra-nodule, landscape, regional) were also collected in three different years, so we had to analyze each set of samples independently. That being said, our observations and conclusions in each of scale of diversity we investigated pointed to water being the main driver, which the discussion outlines in detail. The only metric for which we felt justified to compare all samples together was the community dissimilarly changes across increasing distances (i.e. Fig. 6).

Besides, I wonder if you will reach the same conclusion by using the relative abundances of a lower taxonomic rank. In your previous paper, you mentioned that the results indicated that "the fine-scale composition" is more representative of the changes (Uritskiy 2019a). Wouldn´t you expect a higher degree of dependence on stochastic processes by using a lower taxonomic rank?

The main difference with the 2019 study is that we used WGS metagenomics and thus we had a much better resolution to study strain dispersal. The focus of the current study was the effects of environmental conditions on the taxonomic landscape as a whole. For this we collected and processed a large number of samples, making it prohibitive to apply WGS to all. This is a very good point that should be investigated in the future.

You used well-suited methods considering the research question. Considering that you conclude, "relatively small changes in water availability have significant implications for microbial community structure inside halite nodules," I expected to see the statistical results that link environmental conditions to the observed community composition patterns.

As outlined above, the fact that we collected data across multiple years for multiple locations did not allow for such statistics to be used. Instead, we derive independent conclusions from three independent sampling efforts at different distance scales, and link them together in the discussion, pointing out that all three studies point to the same conclusion. This relationship between atmospheric water availability and community structure in halite nodules was also reported in previous studies (Robinson et al., 2015; Finstad et al., 2017).

Besides, when you mentioned that you have installed Data Loggers at each sampling location, does it mean that you have data from 85 locations? or do you assume that the temperature and RH are the same in the 500 m2 and 20 m2 sections? Furthermore, you mentioned that "the internal nodule condition measurements were recorded one year after the intra-nodule sampling took place, however, the atmospheric temperature and RH conditions were very similar between the two years". In that case, I understand that you observed different patterns and because the different areas have different temperatures and RH you assume that those are the deterministic factors.

We unintentionally used misleading language in the methods – we used three loggers at the three major sampled regions. This has been fixed in the text.

This is correct and it was an assumption we unfortunately had to make on because of the multi-year scale of this project. We took care of validating this assumption in the results by comparing the atmospheric conditions between the years (see Fig. S3). We indeed relied on the conclusions from the three main independent sampling experiments (the major samples scales) to derive these conclusions, however we also outline a few alternate explanations in the discussion.

L56 there are not geological features analyzed in this article but RH and Temperature.

Thank you for pointing this out, it was corrected.

L59-60 "these diverse communities develop largely independently..." It does not seem to be the conclusion of those papers. Crist-Cristoph worked with a composed DNA and they do not analyze the turnover. Uritskiy 2019 looks for common features "can be recapitulated by two general modes of community shifts —a rapid Type 1 shift and a more gradual Type 2 adjustment".

Crits-Christoph was an early study of this system and at the time halite nodules were aggregated with the assumption that their microbiome were very similar. Uritskiy et al., 2019 showed the high level of heterogeneity of microbiomes across nodules and that the individual nature of the nodules allowed for independent strain-level evolution within each nodules. We replaced the Crits-Christoph citation with Wierzchos et. al, 2006, which provides a good overview of the isolated nature of these communities.

There is a methodology described for light transmission measurements in the M&M sections. The results should be the effective transmission and the approximate effective PAR available. In the results section, you mention the average PAR available. It is not clear if you are addressing the effective PAR available calculated as it was mentioned in the M&M section.

This is correct, we calculated the effective PAR available as described in the M&M section. We unintentionally forgot to mentioned that we repeated these measurements in three independent nodules and calculated averages. This has been corrected in the results sections.

L 355 Chlorophyta is not a class name.

L517-528. In comparing the results of the biomass estimation with previous results (Finstand et al 2017) you should consider that only autofluorescent cells were counted in that paper while in yours, total cells were quantified. Moreover, Finstad et al results found that the cell number was always higher in the middle and bottom sections when significant differences were observed. In the driest site, a random distribution of biomass was observed.

This is a good point and we added this distinction in the text. We also discussed that Salar Llamara, the site of study in Finstad et al., 2017, has a much higher relative humidity overall, even for the driest sites, than Salar Grande.

L568 at the functional level and, perhaps, at lower taxonomic rank.

This is an excellent good point that we added to the conclusions. Thank you!