Dear reviewer,

Thank you for your thorough and constructive review. We appreciate the careful attention you have given to our manuscript, especially in highlighting limitations that deserve clarification.

Based on your comments, we recognize that some of our conclusions may appear somewhat ambitious. We have revised the manuscript to present our findings more cautiously, providing more adequate background on prior MICP experiments that led to our methodological choices.

Accordingly, we have refined our conclusions to better reflect the experiment's scope, adding appropriate reservations. We believe that the main findings from this study should be considered in future MICP research due to the potential impacts on microbial diversity. In light of your feedback, we have adjusted the manuscript as follows:

- *The title, however, gave me a wrong expectation of the scientific approach.*

  We understand that the original title may have led to some misinterpretation of our approach. Therefore, we have revised it to: **Differential responses of Aridisol microbiomes from different depths to ureolytic biostimulation.**

  Additionally, to clarify the experimental nature of our study as an incubation experiment, we have emphasized this in the Introduction, within the paragraph outlining the research approach and objectives.

1. *As expected, exposing the microbial community from arid, nutrient-poor soil to water and high concentrations of urea (and other nutrients contained in the yeast extract) resulted in a completely different community. This outcome can, however, no longer be ascribed to single effect (urea), but is instead very likely the result of overall drastically changed environmental conditions. This flaw in the experimental design is not mentioned in the manuscript.*
2. *This experimental setup leads to several flawed conclusions. As the native microbiome is subjected to drastically different incubation conditions, the alpha diversity of the microbial community declines. Even though it cannot be ascribed solely to urea addition, the authors clearly relate this decline to biostimulation and make it an important point both in the discussion and in the conclusions.*
4. *Throughout the manuscript, I found no mention of a control treatment. If the authors choose to drastically change the environment of the soil microbiome by generating a soil slurry, I suggest adding a treatment in form of a soil slurry where urea is omitted from the nutrient medium. By comparing the control treatment to the native soil, the authors could disentangle the influence of incubation conditions compared to the influence of urea on the native community.*

   To address points 1, 2, and 4, we understand that our definition of ureolytic biostimulation and our choice of experimental setup may have lacked clarity. The experiment was not aimed to isolate the microbial response to a single factor (urea addition) but to assess the response to the overall change in conditions typical of MICP experiments — specifically, the addition of urea and an organic carbon source. We have thus included the following clarifications to: (1) clarify the choice of experimental design and background experiments; and (2) emphasize that the microbial response studied is due to the combined standard conditions in biostimulated MICP research.

To clarify, we have added the following text to line 77 of the Introduction:
"Experiments using various medium compositions have been conducted across a range of setups, including incubation experiments, soil column rinsing, and field-scale studies (Gat et al., 2016; Gomez et al., 2018; Gomez et al., 2019; Ghasemi and Montoya, 2022; Ghasemi et al., 2022; Graddy et al., 2021; Lee et al., 2019; Ohan et al., 2020). Hereafter, we refer to the application of urea and an organic carbon source to induce ureolysis as ureolytic biostimulation."

Additionally, we elaborated on our experimental design and its basis in prior research (line 141):
"We aimed to examine the response of edaphic microbiomes to the chemical solution composition that is the common ground to MICP experiments at various scales, since considerable shifts in microbial diversity and medium properties reported in previous studies stemmed from varied setups and media compositions (Gomez et al., 2018; Gomez et al., 2019; Ghasemi and Montoya, 2022; Ghasemi et al., 2022; Graddy et al., 2021; Lee et al., 2019; Ohan et al., 2020). A prior study has shown that the characteristic biostimulation response—a marked pH increase, ureolysis, and dominance of Firmicutes—requires both urea and an organic carbon source (Gat et al., 2016). Accordingly, our study employed a 'before and after' design, where biostimulation was performed by…"

3. *Another flawed conclusion concerns the characterization of ureolysis-related environmental changes. The outcome of the biostimulation experiment is a drastic increase of pH. In the discussion, the authors suggest a similarity between the measured increase in pH in soil slurry and potential pH increase in MICP-treated soil pore fluids. I find the conditions in the described soil slurry incomparable to conditions, geochemical properties and natural buffering capacities of soil. Due to this reason, I consider drawing parallels between the two systems inappropriate.*
While the two systems are not directly comparable, and pH levels were monitored to track the reaction rather than for environmental considerations (see Methods), we believe that our findings hold relevance for future studies with potential environmental implications, such as percolation. In this research, the experimental design was based on previous data attained by our group (Gat et al. 2017) in long-term experiments (exceeding 6 months), and was specifically aimed to follow the biostimulation phase (for typicaly two weeks). In the long-term biostimulation experiment showed that the initial increase in pH was followed by a gradual decline, ultimately converging to control levels of untreated (water only) samples. We address this point in the Results section, lines 221-223.

Additionally, we removed lines 215–217, which stated: "Our results add further concerns to the previously raised question regarding the potential pollution of deeper soil layers and aquifers by ammonium, a prominent MICP byproduct, considering its potential hazard to human and environmental health (Lee et al., 2019)."

We have reiterated this point in line 335 of the Discussion.

Minor remarks:

Introduction

59 – *Is the citation on the importance of cyanobacteria a bit too general? The reference is a book titled "Biological soil crusts: structure, function and management" from 2003. Wouldn't it be better to find a source which directly claims that cyanobacteria (and not for example lichens or algae) constitute a key group in arid biocrusts? According to descriptions in "What is a biocrust? …", in hyperarid regions, biocrusts consist of cyanobacteria and / or algae, while in arid regions, they are generally dominated by cyanobacteria or lichens, with patches of bryophytes commonly found in wetter microsites. In the manuscript, there is a strong accent on cyanobacteria – why is the significance of algae or lichens not discussed? Is it because they cannot be characterized by 16S sequencing? A photograph of sampling area, where studied biocrust are visible, would be helpful as part of the Supplementary Data.*

We focused on cyanobacteria as they are the primary microorganisms involved in biocrust formation in the Negev Desert (as detailed in the cited reference; we added this information in the description of the study site). In this region, biocrusts are typically subtle in appearance. Algae, lichens, and other organisms fall outside the scope of this manuscript.

67 – *typographic error; I assume the authors meant "drought" (a shortage of rainfall) and not "draught" (a cold burst of wind).*

Thank you, typographic error corrected.

86 – *one of the references for archaea becoming more abundant in deeper soil horizons may not apply; from my understanding, the paper by Sokol et al. 2022 is nowhere stating that archaea are more abundant in deeper soil horizons.*

Thank you, reference deleted.

Materials and methods

122 – *from this sentence, it seems like biostimulation experiments were only performed on soils from 3rd site. The 3rd, disturbed site is not clearly described – how was it disturbed? Were upper soil layers placed on the bottom and vice-versa? A photo in Supplementary Data would also be helpful.*

Thank you for pointing out this unclarity. We corrected the sentence regarding the soil origin in lines 126-130.

123 – *"disturbance approximately 20 before this study" – I guess 20 years?*

Thank you, "years" indeed added.

125 - *I am guessing that overall, 12 samples representing Negev soil mean only Site 1 and Site 2, because the math otherwise does not add up (3 sites x 3 depths x 2 replicates = 18 sites; 2 sites x 3 depths x 2 replicates = 12 sites).*

You are correct, we referred to the sequenced samples, and therefore moved this part of the sentence to the suitable place in line 134 for clarification.

138 – *here it seems again like only samples from Site 1 and Site 2 were biostimulated, as the number of biostimulated samples is 12? Then how come the biostimulation effect is later also described for the 3rd, disturbed site?*

As we addressed the samples intended for sequencing, we added this point in line 150 following your comment.

150 – *If possible, I would advise not to use NanoDrop spectrophotometers for DNA extracted from environmental samples; a fluorimetry-based assay, such as Qubit, is more reliable for measuring DNA concentration. Spectrophotometry-based quantification is often reported to overestimate DNA concentrations and is strongly influenced by other proteins and contaminants, which would in case of DNA extracted from soil include humic acids.*

Thank you for the comment. We would like to clarify that NanoDrop spectrophotometry was performed for QC purposes only and not for the quantification of the DNA concentration. We have added clarification for this point on line 161.

152 – *I would expect more details for the library preparation: which exact region of the 16S rRNA gene was amplified, which primers were used (including references where primer design is described), how long was the expected PCR product, details about the PCR program (steps, temperatures, number of cycles).*

Library preparation and 16S amplicon sequencing were designed and performed by Qiagen Genomic Services (Hilden, Germany). The details of the primers and process are the intellectual property of Qiagen. We added available information in line 163.

Results and discussion

213 (figure 1) – *the green and blue line look very similar; it's sometimes hard to distinguish between them.*

As the data points in many cases overlap, it is hard to distinguish between the line even when the colors are different. Therefore, we adjusted the legend to stress the different shapes representing the different depths.

227 (figure 2) – *in this figure, there are samples which strongly separate on the PC1 axis from the rest of the samples (these are on the left side) – I think a reader would like to know what are these samples. The colors are very similar, especially in case of the PCA graph, it's hard to spot the difference under certain light conditions.*

We enhanced the figure following your comment to make the sites more easily distinguished.

305 (figure 5) – *the colors close on the spectrum are very similar; it's very hard to see on the graphs which colors correspond to which taxa.*

We enhanced the figure following your comment to make the taxa more easily distinguished.

451 – *there is an error in the reference; ;Asce, S.M and Asce, M. are not author names.*

Absolutely, corrected.