Responses to comments from Anonymous Referee #1

This review addresses storage, a very important methodological aspect in soil biology which has not received enough attention. As this work shows, we should take more account of the artefacts that storage could introduce to data. This is clearly a relevant topic and falls within the scope of SOIL. However, while this assembly of data is a strong starting point, in my view the analysis suffers from some crucial flaws, such that this contribution falls short of its potential and does not generate well-supported conclusions.

Thank you for underlining the importance of the subject addressed in this manuscript. I hear and understand your criticisms about the production of this synthesis, and I'm going to try to respond to them.

I propose to prepare a new version of the manuscript in which I will try to take into account the comments below.

#R1.1- The manuscript places considerable emphasis on the significance of the storage impacts, not just in the discussion of the results but fundamentally in the way the data were scored and collected from the underlying literature (line 140-150). However, the more important question is how big the effects are, and whether this can be so large as to seriously bias the conclusions of a study. In other words, the analysis would be much more relevant if it focused more on effect sizes and less on significance.

I completely agree that the size effect is a crucial issue while assessing the significance of storage impacts. I agree that a quantitative study based on effect sizes rather than on the presence / absence / inconsistency of storage effects would have been more informative. My question is: How to quantify these effect sizes in a comparable way across a wide range of metrics (and sometimes at different time steps for the same metric?).

In older articles, the authors frequently show data (microbial parameters) without/with storage as figures, or do not provide the raw data (*e.g.* for all replicates) (the sharing of raw data was not required for the publication of articles). Also, authors frequently only report the significance of the storage impacts using ANOVA-based analyses, only mentioning “higher” or “lower values”, which makes it impossible to calculate the size effect.

I felt it was more important to see the consistency of the storage effects (microbial parameter that increases/decrease or no impact in all soils, or inconsistent (“variable”) effect between the different soils tested) rather than to quantify the storage effects (especially as, as mentioned above, it wasn't always possible). With this approach, I obtained a single data (storage effect) per microbial variable and per storage modality for each study, avoiding to fragment information for easier synthesis.

#R1.2- Another problem with the focus on significance is that the analysis essentially follows a vote-counting approach (counting up the number of studies that found a significant effect in one direction or another). This sort of approach has been long recognized as statistically flawed and potentially misleading (e.g. Koricheva et al. 2013 Handbook of Meta-analysis in Ecology and Evolution). There are stronger statistical methods for meta-analysis that could be better applied in this situation.

As you must have realized, it was not my intention to carry out a meta-analysis (this term does not appear in the manuscript), but rather to provide a ‘state of the art’ of the knowledge about storage impacts on soil microbial properties and to demonstrate the small number of studies and fragmented data available on this subject. It's not interesting to know whether soil storage has an impact on the microbial parameters measured (it usually has an impact), but rather which storage method has the least impact on the analysis of a given microbial parameter, and perhaps dispel some preconceived ideas about best storage practices.

I felt that the information was too fragmented and studies were too dissimilar and then inadequate for a meta-analysis (The average number of studies for a given question, *i.e.* effect of a given storage practice on a given microbial parameter, is low: median = 3.0, mean = 3.78).

As mentioned above, the corpus of studies frequently failed to provide adequate data for inclusion in a meta-analysis (effect size or raw data needed to calculate it, standard errors or confidence intervals), especially in less recent papers. In these conditions, vote counting is the simplest method for synthetizing multiple independent studies.
Finally, because the methods used to characterize microbial parameters have evolved considerably over time, it may be problematic to compare the results of older and more recent studies (*e.g.* for sequence-based variables: sampling effort and sequencing depth are not comparable).

In the forthcoming revised version, I will try to clarify the objectives of the study in the introduction.

#R1.3- A review in my view should not only summarize results from the literature (which this manuscript has done) but also add something new to the literature through the analysis or interpretation of these results to create new knowledge. I don't think the manuscript has done enough to achieve this. The Discussion concludes that the effects are concerning and highly variable, but I would really like to hear something deeper: what might be causing the variation? What parameters of storage or removal of storage might be influential? When does it matter and when not? The Conlusions state that storage practices need to be adapted to the microbial parameters and soils being studied, so what guidance does the assembled data provide on how to do this?
Here again, I agree with this comment! The conclusion on this review is that the data are too scarce, the context of analyses is too variable (type of soil, condition of storage, duration of storage…) to produce recommendations. I think the points you raise above are addressed in the discussion, but not sufficiently developed because of the scarcity of data available. I propose, in the next version, to restructure the discussion by clearly mentioning these questions and trying to provide answers (or possible answers).

#R1.4- The analysis is not up to date, not including literature from the last three years (line 159) and excluding relevant recent work (l. 228). Such a meta-analysis should really be as current as possible.
Indeed, the analysis includes the literature up to 2021. This bibliographical synthesis required a huge amount of time and effort. I recognize that several relevant studies (*i.e.* explicitly comparing microbial parameters in stored *vs* unstored soil samples) have been published since 2021 (most of these studies were not available when I performed the analysis) and must be included in the corpus.
Querying the WOS database again, using the same keywords as those used for the study, gave 204 articles, including six relevant to the synthesis (indicated with an asterisk) (two of these papers, Finn et al. (2023) and Lane et al. (2022), are already mentioned in the manuscript):

* Lane et al. (2022) Soil sample storage conditions impact extracellular enzyme activity and bacterial amplicon diversity metrics in a semi-arid ecosystem. Soil Biology and Biochemistry 175, 108858, https://doi.org/10.1016/j.soilbio.2022.108858
Already mentioned in the manuscript (as additional reference): it will be included in the corpus of the synthesis in the next version.

* Finn et al. (2023) Importance of sample pre-treatments for the DNA-based characterization of microbiomes in cropland and forest soils. Soil Biology and Biochemistry, 184, 109077, https://doi.org/10.1016/j.soilbio.2023.109077
Already mentioned in the manuscript as additional reference. Although the authors present their results as mean relative change in microbial parameters for pre-treated (dried or fresh) soil in comparison to deep-frozen soil (their reference).

*Kushwaha et al. (2024) Field to greenhouse: How stable is the soil microbiome after removal from the field? Microorganisms, 12, 110, https://doi.org/10.3390/microorganisms12010110.

* Lee et al. (2021) Revisiting soil bacterial counting methods: Optimal soil storage and pretreatment methods and comparison of culture-dependent and -independent methods. PLoS ONE, 16, e0246142. https://doi.org/10.1371/journal.pone.0246142.
The authors tested the impact of storage condition on epifluorescence bacterial counts in soil samples stored at -20, 4, or 24 °C in one soil type (pre-study to determine the best storage temperature). This reference will also be cited as an example of pilot study.

* Moy & Nkongolo (2023) Variation in microbial biomass and enzymatic activities in metal contaminated soils during storage at low temperature (4°C). Chemistry and Ecology, 39, 688-709. https://doi.org/10.1080/02757540.2023.2253222
The authors showed that microbial PLFA-based biomass decreased during the first two weeks of COLD storage and remained unchanged thereafter, and that most enzymes inconsistently increased or decreased over time during storage at 4°C.

* Smenderovac et al. (2024) Drying as an effective method to store soil samples for DNA-based microbial community analyses: a comparative study. Scientific Reports, 14, 1725, https://doi.org/10.1038/s41598-023-50541-2
The authors tested several methods for DNA preservation strategy, including freezing at -20°C. Additionally, the authors tested some commercial preservatives, showing that storage at room temperature with silica gel packs gave results compatible to frozen samples. They acknowledge that the preservation methods should be studied on a grater range of soil samples and with more barecodes.

The following 6 references are not retained (I explain why):

Reardon et al. (2022) Enzyme activities distinguish long-term fertilizer effects under different soil storage methods. Applied Soil Ecology, 177, 104518, https://doi.org/10.1016/j.apsoil.2022.104518
Not included because the authors used drying at +40 °C (which is unusual).

Hu et al. (2023) The preservation of bacterial community legacies in archived agricultural soils. Soil & Tillage Research, 231, 105739, https://doi.org/10.1016/j.still.2023.105739
Compared soil samples stored at -80°C or air-dried samples (no comparison with fresh, unstored soils).

Brock et al (2024) Impacts of sample handling and storage conditions on archiving physiologically active soil microbial communities. FEMS Microbiology Letters, 371, fnae044. https://doi.org/10.1093/femsle/fnae044
The 'control' modality was soil samples immediately snap frozen in liquid nitrogen and stored at –80°C. The authors did not compare to non-stored, unfrozen soil.

Pavlovska M., Prekrasna I., Parnikoza I., Dykyi E. (2021) Soil sample preservation strategy affects the microbial community structure. Microbes and Environments, 36, ME20134, https://doi.org/10.1264/jsme2.ME20134
Their control was immediately frozen in liquid nitrogen (no fresh control)

Ouyang et al. (2021) Direct cell extraction from fresh and stored soil samples: Impact on microbial viability and community compositions. Soil Biology and Biochemistry, 155, 108178, https://doi.org/10.1016/j.soilbio.2021.108178
The paper focuses on complex cell extraction procedures before analysis of microbial parameters: this study cannot be compared to other DNA-based studies. The authors showed that cell viability changed and microbial community composition changed in all stored samples, but that the least changes were observed at +4°C. This result can be added in the discussion but not included in the corpus.

Edwards et al. (2024) Long- and short-term soil storage methods other than freezing can be useful for DNA-based microbial community analysis. Soil Biology and Biochemistry, 191, 109329, https://doi.org/10.1016/j.soilbio.2024.109329

The authors compared soils stored under different conditions (e.g; refrigeration versus storage) but not with the same soils immediately analyzed before storage.

#R1.5- The manuscript touches on the bias of effects (i.e. does storage affect all samples the same, so that experimental conclusions might still be valid even if absolute values are shifted?) but it doesn't provide a robust and transparent analysis on this point. This would be very valuable, because many workers defend even large experimental artefacts as acceptable if all samples are treated the same. Line 461 in the Conclusions supports this argument, but it is not clear what data this conclusion is based on.

Few authors address the question of effect bias, i.e. whether a given storage practice affects all samples in the same way. On way to evaluate whether storage impacts bias the conclusion of a study is to estimate if the ranking between fresh (unstored) and stored samples was conserved. However, as mentioned LL 153, this information is not always (and even rarely) reported.

Following your suggestion: I will add a quantitative analysis on this issue (for studies including several soil types / land use / agronomic practices: how many found consistent impact of storage for a given storage practice?).

#R1.6- - The presentation of results provides a raw summary of results from numerous papers, but is very dense reading without a clear line of argument. This would benefit greatly from summarizing the various findings in a figure and highlighting only the important points in the text.

The aim of this review was firstly to take stock of the current state of knowledge on the impact of different storage methods on microbial parameters measured in soils. To achieve this objective, it was necessary to carry out an exhaustive analysis of the existing literature. The fact that the content of the review is dense and difficult to synthesize is directly linked to the fact that the available data are scattered and carried out in very different contexts. It should also be noted that we have not considered all the contextual elements (soil type, storage time in particular), otherwise the available information would have been even more scattered and fragmented.

#R1.7- The Discussion section does provide a few conclusions, but to a large extent introduces new observations from the literature (e.g. lines 371-403) rather than discussing deeper trends or conclusions available from what was presented in the Results section.

For this review, I have limited the corpus of studies that explicitly compared soil microbial parameters in fresh, unstored and stored samples. In the discussion, I called on other studies that didn't necessarily meet these conditions to provide additional information.

One possible response could be to mix references that explicitly compare microbial parameters in stored and non-stored soils (references indicated by numbers) with additional references that provide additional information (references indicated by names), but the quantitative aspect of the synthesis would be lost.

- Other comments

It is not clear what the conceptual difference is between measurements of community composition and structure (l. 133-4)

"Composition" refers to the list of species, taxa or OTUs that are present while "structure" additionally includes the relative abundance of these species / taxa / OTUs. This will be clarified.

The abbreviation "qPCR" is preferable to "RT-PCR", since the latter is ambiguous ("real time" or "reverse transcription"). In line 215 they are both given, increasing the confusion.

The term "RT-PCR" will be replaced by "qPCR" in the revised version.

Line 244 not clear what "failed to report" means: didn't investigate; didn't show results; or found no significant effect?

Thank you for this comment. Actually, the mentioned studies found no significant impact of storage on DNA-based structure of soil microbial communities (as shown in Table 2). This sentence will be clarified in the revised version.

Would Table 1 not be better presented as a figure (e.g bar or line graph)?

Thank you for this suggestion. It will be applied in the next version.