Dear Dr Soininen,

We would like to thank the two reviewers and yourself, for considering our manuscript "What drives study-dependent differences in distance-decay relationships of microbial communities?" for publication in Global Ecology and Biogeography. We are delighted to see that both reviewers were very positive about our manuscript, and we are grateful for the insightful and constructive comments received. We believe our revisions in response to these comments have considerably improved our manuscript, and we hope they are well received. We provide our point-by-point response to all reviewers' comments below. Please also find a "clean" version of our manuscript attached, with all modifications and track-changes accepted. I confirm that all authors have seen and approved this revised version of our manuscript.

We thank you for considering a revised version of our manuscript for publication in Global Ecology and Biogeography.

Dr Dave Clark (on behalf of all authors)

Point-by-point response to reviewer comments.
Editor.

"While both of them are in principle positive, they raise a number of points you need to carefully address in the major revision. I think the points they raise are clear and rather straightforward. I want to emphasize the scale dependence of distance decay. In theory, distance decay follows power law just like species-area relationship and thus slope should become milder and halving distance longer with extent in lin-lin regression from (Soininen et al. 2007). Here you have used Mantel r which measures strength of the pattern, not steepness (slope) as also pointed out by Referee 2. You need to very carefully think about how to formulate your hypothesis about Mantel r and spatial extent and discuss the difference between slope and Mantel r in Discussion. Some good advice about strength (r) and steepness (slope) of the macroecological pattern can be found in Hillebrand (2004) Am Nat."

- We appreciate the suggestion to reframe our scale-hypothesis, and to add some discussion about the difference between slope and strength of distance-decay relationships. We have reframed our hypothesis as suggested (lines 150-153), and have also added some text to the discussion (405-424). Furthermore, we have revised our language throughout to ensure that it is clear that we are analysing distance-decay strength rather than slope.

Reviewer 1.
"*The study is well conducted and the paper well written. I have only a few comments and suggestions of improvements.*"

- We thank the reviewer for the constructive and encouraging comments.

"Regarding spatial scales, I wonder how many tried to investigate really small spatial scales (millimeters or centimeters) and did that vary among environments? This could potentially have an effect on the results of the comparison among environments."

- We agree with the Reviewer's suggestion that differences in the minimum spatial scales between environments could alter the relationship between scale and the strength of microbial distance-decay relationships. We have clarified the range of spatial scales investigated within each environment in a new supplementary figure (Fig. S3). We have also added some text to the results (lines 269-272) and the discussion (lines 426-438) to speculate on some of the reasons why minimal spatial scales could differ across environments.

"When talking about environmental context I thought it would be more about environmental heterogeneity, and aspect that perhaps could be expanded."

- We agree that our discussion of the role of environmental heterogeneity in distance-decay relationships may not have come across clearly. Therefore, we have expanded the sentences on environmental context to include further discussion of heterogeneity (lines 88-92).

"Did the authors check if the original mantel tests used Spearman or Pearson correlations? Should this not be affecting the results?"

- We did not initially collect these data as we suspected (from reading much of the literature analysed here) that few studies described which correlation coefficient was used, and because the choice of which correlation coefficient to use lacks any underlying ecological rationale, unlike the methodological aspects that we analysed. However, we appreciate that this choice could have an effect and so have now updated our dataset with this information. We searched for explicit mentions of "Spearman" or "Pearson" in relation to the distance-decay analyses within each study. As expected, the vast majority of studies did not explicitly state which correlation type they used (n = 304), whilst Spearman's correlation coefficient was more frequently used (n = 86) than Pearson's (n = 62). Furthermore, we found no significant difference in the strength of microbial distance-decay relationships between Spearman's and Pearson's correlation coefficients (R^2 = 0.01, P = 0.118). Following comments from both reviewers to check potential differences between correlation types, we have added this extra analysis to the methods (Box 1) and results sections (lines 323-327).

"Regarding lakes, maybe clarify that distances were within lakes, if they were."

- We appreciate this point, and also a similar point from the Reviewer 2 regarding whether studies from lakes were conducted within a single lake or across multiple lakes, and how this might affect the strength of the distance-decay relationship. Of the 76 distance-decay relationships from lakes, 50 were conducted across multiple lakes, whilst 26 were conducted within a single lake. Despite this, Mantel correlation

coefficients were not significantly different between these groups, suggesting that within-lake environmental heterogeneity may balance with increasing dispersal limitation or environmental heterogeneity between lakes. We have added some text to the results (lines 254-258) and discussion (lines 388-398) to clarify this, and a new supplementary figure (S2) to demonstrate this result.

"I like the discussion about the dissimilarity indices (lines 471-499). What I am missing is the discussion about sequencing data being compositional, which also affects which indices are suitable. See Gloor et al. 2017 Front. Microbiol. 8:2224"

- We are glad the reviewer enjoyed this part of the manuscript and appreciate the suggestion to discuss the important issue of compositionality in sequencing data. We have now added some discussion of this issue (lines 549-560) as suggested by the reviewer.

"I think the final conclusions could be more edgy, it could maybe point forward in some research direction or provide some recommendations."

- We appreciate the suggestion to make our final conclusions bolder. We have now added some text describing our recommendations and vision for how the field should proceed (lines 602-611).

Reviewer 2.
"I see this is a very great study with clear conclusions."

- We are grateful for the reviewer's insightful suggestions and encouraging comments.

"Lines 51-52: Please consider add some references"

- We have added a small selection of relevant references as suggested, to highlight the array of taxa for which distance-decay relationships have been found (lines 52-53).

"Line 72 - Here starts my confusion, the authors mention steepness for the first time here, but many other parts of the text. Studies that mention steepness do it because they use regressions, studies with mantel tests are not able to provide steepness, instead, they provide a correlation value. This is very important to be clear as correlation and steepness are not interchangeable."

- We apologise for the confusion caused by mixed terminology. We have now clarified throughout that our analysis refers to the strength of distance-decay relationships, as quantified by correlation coefficients, rather than the slope. Furthermore, we have explicitly mentioned the differences in strength and slope in the discussion (see lines 376-379, and 407-414) which should draw additional attention to the important distinction between these terms.

"Line 98 - I don't know if the authors refer to strength (ability to predict) or steepness (rate of the decay) here, and If the latter, I would disagree. Distance-decay relationships are in theory similar to the species-area relationships (https://www.nature.com/articles/nature03034#ref-CR21). Because the regional species pool is limited by an upper boundary for regional diversity (https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2699.2011.02652.x). Therefore, the slope of the distance decay should saturate at some spatial extent, and increasing the sampled area would not make it any steeper, but milder. On the other hand, I would agree that it becomes stronger (higher correlation value) as the discovering rate of species decreases with increasing sampling effort, if that' the case."

- We appreciate that the mixed terminology highlighted by this reviewer's previous comment could have resulted in confusion regarding our point that increasing a spatial extent should increase the strength of distance-decay relationships. We did indeed mean strength in this instance, and thus are in agreement with the referee. As described in our response to the previous comment, we have revised our language throughout so as not to cause confusion between "steepness" and "strength" of relationships.

"Line 145 - "steep physicochemical gradients"? is it measured?"

- We thank the reviewer for pointing out this ambiguous sentence. We intended this sentence to mean environments such as sediments and soils, which are generally heterogeneous over very small spatial scales, will have stronger distance-decay relationships than environments such as seawater or air, which have more diffuse environmental gradients due to mixing, as described in the introduction. We have now clarified this sentence (lines 147-149).

"Lines 182-183 - Did the authors check whether the mantel test used "Pearson" or "Spearman" correlation rank? this may affect the results as the distance decay is not linear and the Pearson coefficient in untransformed data would lead to falsely low values especially at a large spatial extent."

- We thank the Reviewer for this suggestion. Reviewer 1 also raised this suggestion, and so our response to this point is provided in detail above.

"Box 1. ENVIRONMENT: was the choice arbitrary or there was some GIS methodology involved?"

- Our categorisation of each distance-decay relationships environment was based on the site description within each study. In the vast majority of cases, the site description provided a clear categorisation of the study site, including keywords such as "forest", "grassland", or "river". Whilst a GIS-based approach could be more systematic, it would not have been possible for some studies which did not give spatial coordinates, and may not reflect accurate assignments at the time each site was sampled. For instance, some sites may since have transitioned from grassland

to agricultural land or vice versa, and this may not be reflected in a GIS-based classification. Furthermore, the resolution at which landscape classification data is available may not be high enough to reflect local habitat pockets such as streams, small ponds, or forest fragments, and thus the site description would likely provide a more accurate assessment of the type of environment in which each study was undertaken.

"*The authors should consider giving a figure or table with the summary statistics of all datasets, I struggled to find how many studies per category, the gradient of spatial extent, community coverage, etc.*"

- Whilst we provided the raw data for readers to examine, we appreciate the suggestion that some summary statistics would help give context to the study. We have therefore added a new table (Table 2 in revised manuscript) to display the number of data points in categorical variables, as well as summary statistics of continuous variables.

"*A big note of caution for this section. Please homogenize the terminology! Once the authors wrote "weak distance decay", but after "low effect size". If I understood well, the two sentences are equivalent and should have been written consistently. I could also spot "larger" to antagonize "low"... should be "higher" as this is the real opposite.*"

- We agree with this comment and have thoroughly revised the manuscript to enure terminology is consistent throughout. We have therefore dropped any mention of effect sizes (apart from in the methods section) as this is a more statistical term and is less easily interpretable than talking in terms of the "strength" of relationships.

"*Also, avoid the use of adjectives. A statistical test is significant or not significant, the threshold for significance is debatable, but writing whether the differences are clear or not doesn't help the reader to understand precisely. For example, two individuals may have clear (noticeable) differences in height but this won't be testable if the population mean is not available.*"

- We apologise for the confusion caused. We were following recommendations described in Dushoff *et al*., (2019, Methods in Ecol & Evol), which makes the (philosophical) suggestion to frame the results of statistical tests in terms of clarity rather than significance. We have now explicitly stated this in the methods (lines 215-217) and have added the relevant reference. We appreciate that this change in terminology is a philosophical point and as such, we are happy to revert back to the more classic paradigm of describing results in terms of "significance" upon the Editor's recommendation.

"*Line 255: I am quite intrigued by this low R^2 value. Only 5% of the whole variation is accommodated by your model. Did the author check for model heteroscedasticity and residuals? Linear models tend to generate significant p values at higher n.*"

- We appreciate the Reviewer's suggestion to perform diagnostic checks on the model of Mantel statistics as a function of scale. We did indeed check model residuals and found evidence of heteroscedasticity in the residuals. Upon exploring this, we found that the heteroscedasticity was best accounted for by including a random effects structure in the model, specified as a study-dependent random intercept. Here, the total model explained 47.8% of the variation, although the marginal effect of scale was still very weak (marginal R^2 = 0.02) and was not significant. We also checked other regression models and, where necessary, accounted for heteroscedasticity in the same way. We have added these details to the methodology (lines 207-210) and have added some text to the discussion regarding the lack of relationship with scale (lines 407-426).

"I am aware that studies with microorganisms usually have low R² because of high stochasticity. Still, I can't help but thinking that this coefficient has low ecological meaning. Would be good to have somewhere in the discussion what the authors think about this low R². Furthermore, did the authors used the Z-score as a response in this model? Wouldn't it be better to provide the whole statistical results in table and also include the intercept? Slopes alone do not explain much."

- We agree that the R^2 is low, and have added some text to the discussion regarding the lack of relationship between distance-decay strength and spatial scale, as detailed in our response to the previous point.
- To answer the reviewer's other query, we used the transformed coefficients (Z-scores) as the response in our models. We have acted on the suggestion to add details of model intercepts and have included these in Table S2.

*"I can understand from the figure that the coefficient comes from Mantel values against log10 spatial distances in meters (because 0.1 + 7*0.03 = 0.31), but the authors wrote in methods (line 193-195) that all analyses would be carried using Z-scores, is there a reason to use actual values in here?"*

- To answer the reviewer's question, we used original Mantel correlation coefficients in our plots as we felt this would be more interpretable than presenting Z-scores. The transformation of Mantel coefficients to Z-scores is not straightforward (see Rosenberg *et al.*, (2013) for original formula), it would be challenging for readers to mentally perform the inverse transformation to derive the original correlation coefficients. Consequently, we chose to perform statistical analyses on the Z scores, but present figures using the original correlation coefficients for ease of interpretation, as we stated on lines 193-195 (in the original manuscript).

"Lines 265-269 belongs to methods"

- We appreciate that this introductory sentence may be repetitive of text in the methods, where our analyses are described in greater detail and have therefore deleted these lines.

"Lines 272-273: Among studies or within the same study?"

- Here, we describe how we filtered out molecular methods for which there were not a sufficient number of distance-decay relationships across the entire dataset. We have clarified this in the text (line 288).

"275-276 - here Log10, before in the text the authors wrote Log, the same for figure 3. guess they are all log10, please revise. Furthermore, log transformations should be given in the axis title."

- We appreciate the reviewer pointing out this inconsistency. We have revised the text throughout to ensure that all log transformations are stated as $\log_{10}$.

"Line 277: Here there is a good example of the confusion generated by the wording. Strength is being measured by R², or mantel r, or slopes? Confusing."

- We apologise for the confusion caused by our wording. We have ensured that throughout we have clarified our wording around measures of effect sizes. In this case, the strength of the linear regression is measured by the R^2.

"Also, R² 0.01 is weak, but not 0.05 ???"

- We agree with the reviewer and have amended our results to describe this R^2 as weak too.

"Line 278: Avoid adjectives in academic writing, outliers can be found e.g., using is_outlier() in R. However, extreme values have another methodology to be found and this was not mentioned in the methods.

L. de Haan, A. Ferreira (2006). Extreme Value Theory: An Introduction , Springer-Verlag."

- We appreciate the suggestion from the Reviewer on procedures to identify outliers. After adjusting our statistical approaches to account for heteroscedasticity (in response to an earlier Reviewer comment), we conducted a Rosner test to identify outliers with extreme community coverage values. Both of the points we removed previously were flagged as outliers and removed. However, after re-running our model, the results were qualitatively and quantitatively near-identical to the results with the outliers present. Therefore, we chose to leave these points in and present only the model with outliers present.

"Lines 361 - But lakes are not connected and work as islands within the terrestrial realm. Do the authors also expect weak distance decay for lakes?"

- We appreciate this important point about potential differences in distance-decay relationships within and between lakes. Reviewer 1 also raised a similar point, and we have responded to this comment in detail above.

"Lines 471-499: Note that the Raup-Crick method is less sensitive to richness than the Jaccard index. This could be a reason for weaker distance decay of similarities using such an index. Please see :

https://esajournals.onlinelibrary.wiley.com/doi/full/10.1890/ES10-00117.1%4010.1002/%28ISSN%292150-8925%28CAT%29VirtualIssue%28VI%29ECS2

https://onlinelibrary.wiley.com/doi/full/10.1111/j.1461-0248.2010.01552.x"

- We thank the reviewer for drawing our attention to this important point. We have now added some text to the discussion to acknowledge the referenced studies (lines 500-508).