**Supplementary Information**

**Title:** What drives study-dependent differences in distance-decay relationships of microbial communities?

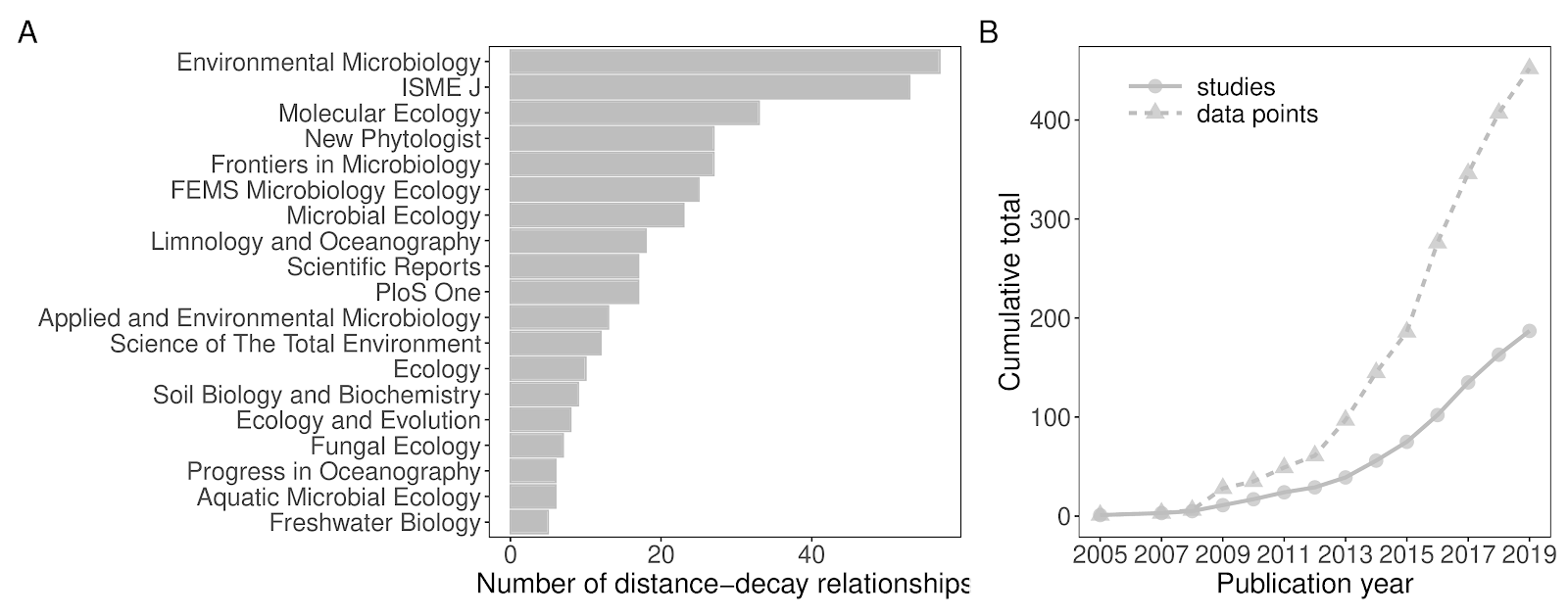
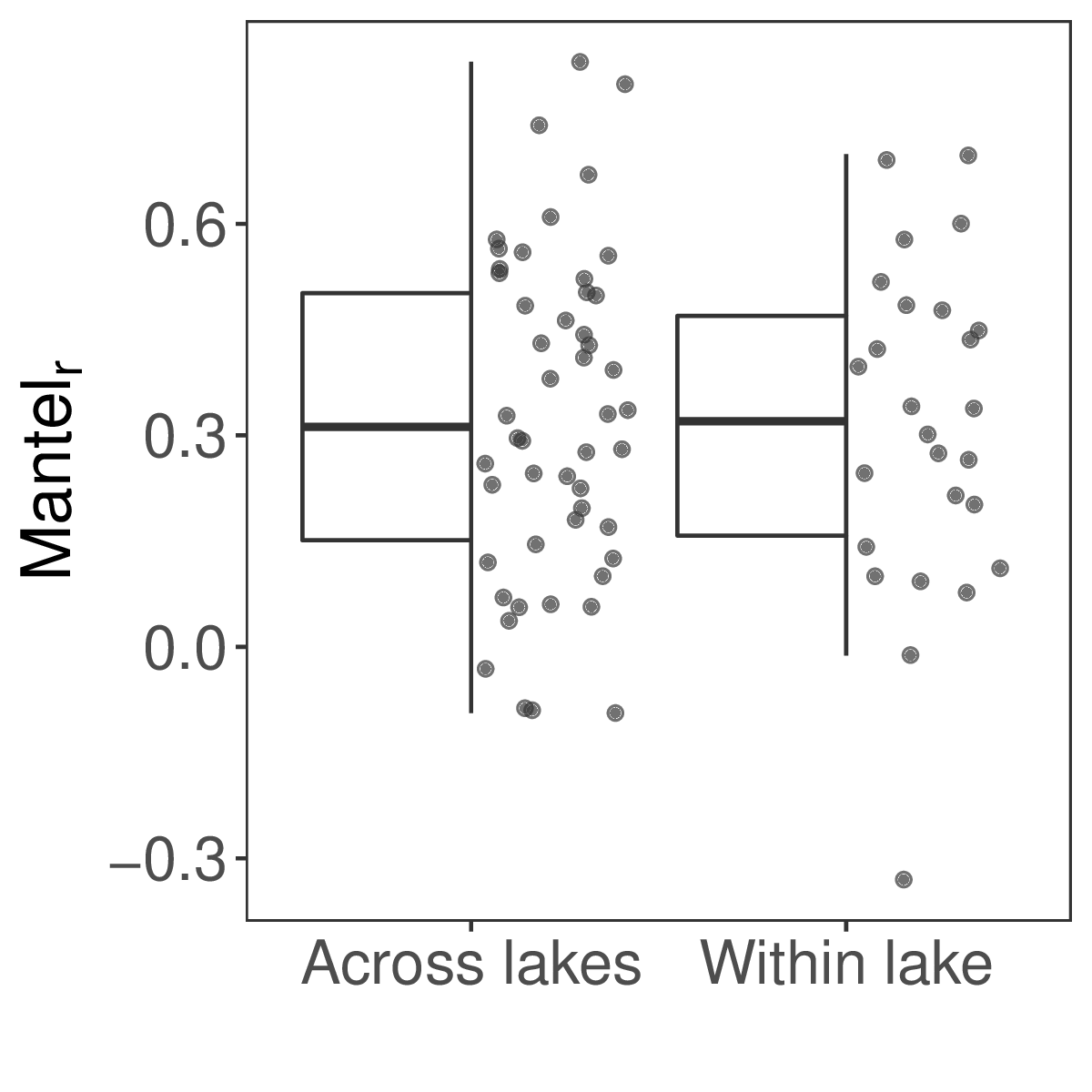


Figure S1. (A) The number of distance-decay relationships obtained from each journal. Only journals with five or more distance-decay relationships are shown for clarity. (B) The cumulative total of suitable studies (circles) and distance-decay relationships (triangles) according to publication year.

Figure S2. A comparison of the strength of microbial distance-decay relationships from studies conducted within single lakes compared to those across multiple lakes.

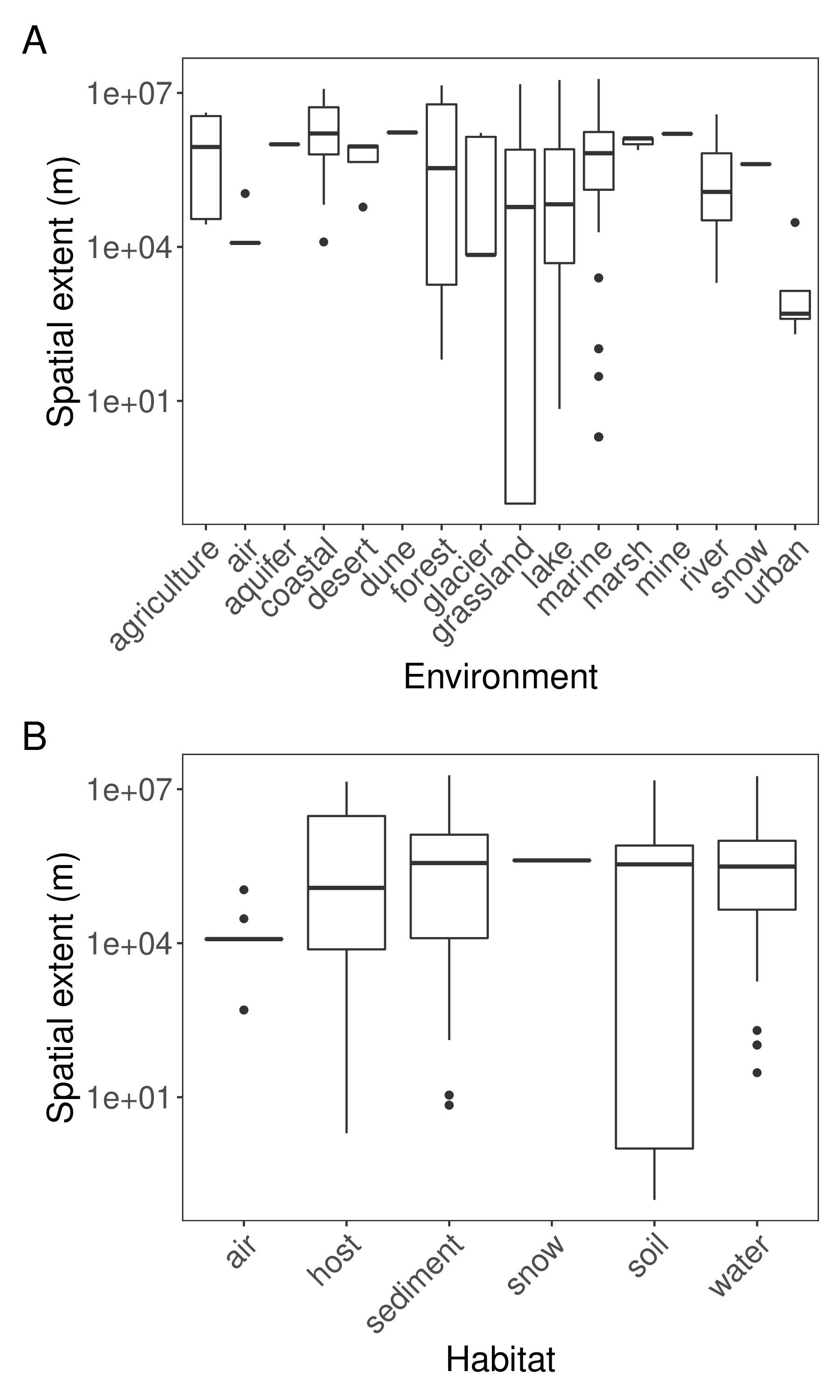


Figure S3. A comparison of the spatial extents of distance-decay relationships across environments and habitats. Note that the y-axis is plotted on a log10 scale for clarity.

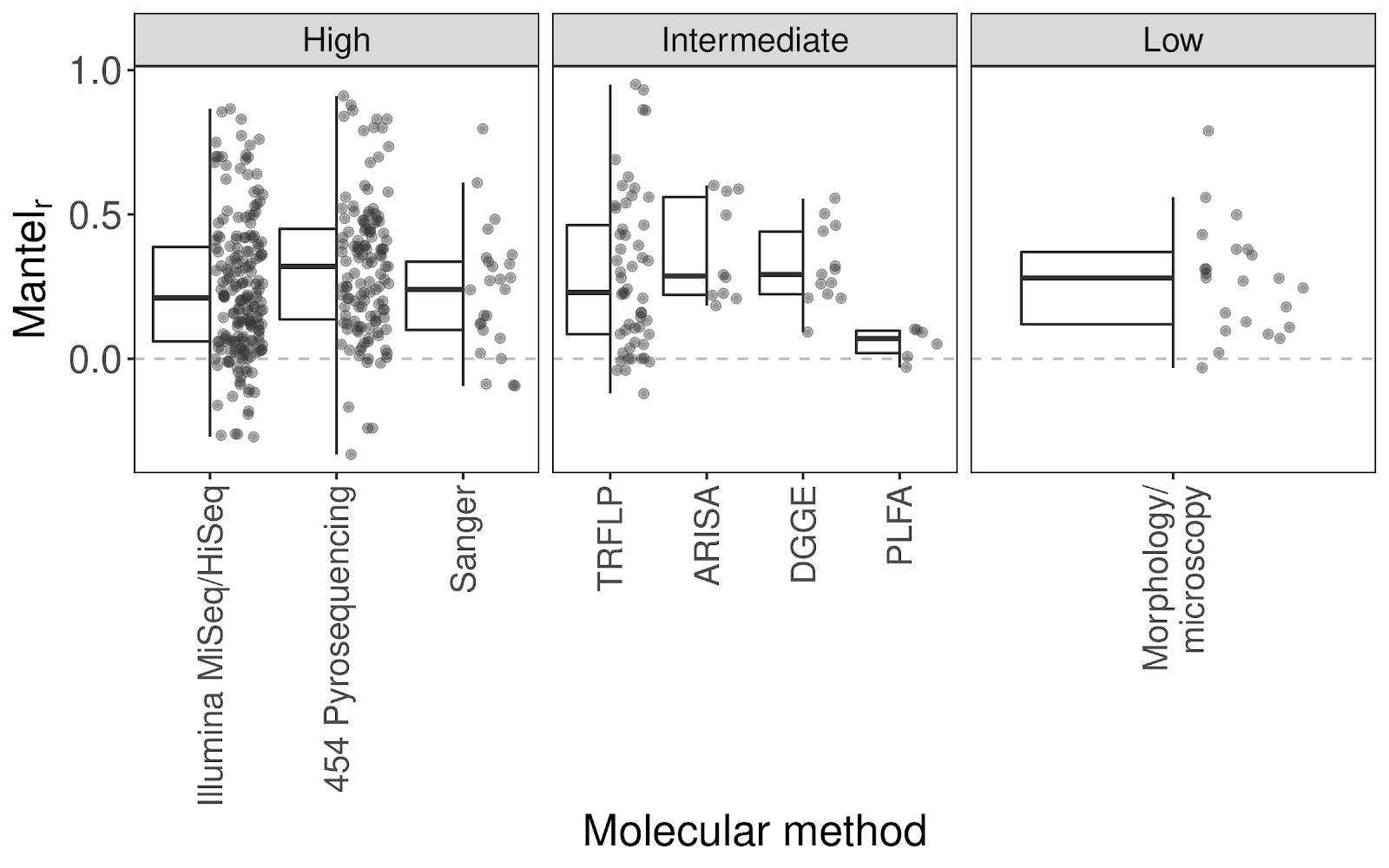


Figure S4. Mantel correlation coefficients of distance-decay relationships where the microbial community was characterised using molecular or morphological methods of varying resolution. High Mantel correlation coefficients indicate a stronger distance-decay relationship. Abbreviated molecular techniques are defined as follows: (TRFLP = Terminal Restriction Fragment Length Polymorphism; ARISA = Automated Ribosomal Intergenic Spacer Analysis; DGGE = Denaturing Gradient Gel Electrophoresis; PLFA = PhosphoLipid Fatty Acid analysis).

Table S1. Metadata extracted from each study confirmed to be suitable for inclusion in this study. Data will be deposited to the Dryad data repository upon acceptance of this manuscript.

Table S2. Full details of all statistical results obtained. For categoric variables, likelihood ratio tests were used to assess the statistical significance of variables, and post-hoc Tukey HSD tests to identify significantly different groups. For continuous variables, Wald tests were used to assess the statistical significance of variables.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Covariate | Coefficient | Test-statistic | Degrees of freedom  (used, residual) | Explained variation | P-value |
| ~ taxon | Taxon | NA | F = 0.99 | 5, 441 | R2 = 0.01 | 0.43 |
| ~ environment | Environment | NA | *F* = 3.1872 | 10, 432 | *R*2 = 0.07 | < 0.001 |
| ~ environment \* habitat | Environment | NA | *F* = 3.29 | 9, 420 | *R*2 = 0.12 | < 0.001 |
| Habitat | NA | *F* = 6.65 | 3, 420 | < 0.001 |
| Environment \* Habitat | NA | *F* = 1.93 | 4, 420 | 0.10 |
| ~ within/between lakes | within\_lake | NA | *F* = 0.11 | 1, 74 | *R*2 = < 0.01 | 0.743 |
| ~ (random intercept | study) + log10(scale) | log10(scale) | 0.29 | *T* = 8.47 | 1, 435 | *Marginal R*2 = 0.02  Conditional R2 = 0.48 | < 0.0001 |
| Intercept | 0.02 | *T* = 1.58 | 0.11 |
| ~ resolution | Resolution | NA | *F* = 0.56 | 2, 449 | <0.01 | 0.57 |
| ~ method | Method | NA | *F* = 1.97 | 7, 437 | 0.03 | 0.06 |
| ~ (random intercept | study) + log10(community coverage) | Community coverage | 0.06 | *T = 2.73* | 1, 337 | *Marginal R*2 = 0.04  Conditional R2 = 0.57 | < 0.01 |
| Intercept | 0.13 | *T* = 1.55 | 0.12 |
| ~ similarity index | Similarity index | NA | *F* = 7.24 | 9, 424 | *R2 =* 0.13 | < 0.001 |
| ~ correlation type | Correlation type | NA | *F* = 2.47 | 1, 146 | *R*2 = 0.02 | 0.12 |