Appendix 3

Literature review

We conducted a qualitative literature overview of the bacterial diversity-ecosystem functioning literature that used natural bacterial communities and a dilution-to-extinction approach to create a diversity gradient. Studies that generated a diversity gradient through other means (e.g. fumigation or application of stressors) were not included as such a comprehensive overview was beyond the scope of this article.

We searched for relevant articles on Google scholar with the search string ["dilution to extinction" AND "bacterial diversity" OR "microbial diversity" AND "community function" OR "ecosystem function"]. This search resulted in 72 articles (accessed on Sept 11th 2015), 12 of which met the above-mentioned criteria. We then searched the literature cited by these articles for further relevant studies. This resulted in a total of 22 articles, all but one published between 2001 and 2015. Among the 22 initially selected articles, two had to be excluded. One did not present statistical evidence and the other conducted the final experiments on agar plates, which we defined as not comparable.

We grouped the response variables into 10 different categories. These were: (1) Abundance or biomass. If both were measured and contradicted each other, the result for biomass was taken. If the abundance was reported separately for subgroups of heterotrophic bacteria, we only looked at the results for all heterotrophs together. If the data for all heterotrophs were not provided, we summed up the reported groups; (2) Activity, measured either as respiration

or the uptake rate of isotope-labeled amino acids and nucleic acids; (3) Degradation of carbon sources, measured as the ability to metabolize a set of carbon substrates (usually using BioLog(R) plates) or the ability to degrade certain carbon sources such as phenol, chitin and lignin; (4) Resistance and (5) resilience, measured as the deviation of the perturbed community from a control, either immediately after the perturbation (resistance) or at the final time point (resilience). If the authors chose another definition we reanalyzed the data. Only deviation in levels of ecosystem functioning were taken into account; (6) Stability, measured as the temporal stability of either biomass or activity; (7) Nitrogen cycling. Includes denitrification, potential nitrification, nitrate accumulation, nitrite oxidation, and arginine ammonification; (8) Enzyme multifunctionality, measured as the capacity to sustain the simultaneous activity of a set of extracellular enzymes at certain threshold levels (following Gamfeldt et al 2008) If multifunctionality was measured, the single functions were not counted separately; (9) Invasion resistance, measured as the ability of an invader to survive in the host community; (10) Enhancing plant productivity, measured as the effect of soil bacterial diversity on plants.

For each article, we looked at the relationship between manipulated diversity and ecosystem functioning. As estimates of realized diversity were seldom available, or given in incomparable units (e.g. bands on DGGE gel, morphotypes of colony forming units or OTUs derived from pyrosequencing), diversity was usually reported on an ordinal scale which corresponded to the dilution factor. In general we tried to follow the authors' interpretation of the results. However, we counted the results only if statistical evidence was presented or the data were presented in an extractable form (i.e. either in a table

or graphically with sample mean, a measure of error, and sample size). If the authors used anova (analyzing diversity level as a categorical variable) we first tracked whether or not a significant effect was found between the lowest and highest diversity levels (as assessed from a post hoc test or from accompanying tabular or graphical representation of the data). If the significant anova effect was explained by intermediate diversity levels, the data were extracted and reanalyzed using regression on the full range of diversity levels. If data were presented from different time points, only the last data point was evaluated.

If data were presented graphically, we assembled them using GraphClick 3.0.3. We then used the function "mvrnom" from the MASS package in R (R Core Team 2015) to generate a sample with the same sample size, mean and variance as given by the extracted summary statistics. Then we regressed ecosystem function against diversity level and fitted a linear model to the data.

A relationship was counted as significant if the p-value was below 0.05. We categorized the relationships as either positive, negative, non significant or ambiguous. The last category was applied if two different response variables were presented that measured the same function according to our definition and the results did not agree. If a study presented several separate experiments or treatments we counted each experiment treatment separately, unless the authors made the choice to pool the data before the analysis in which case we took the results as presented by the authors. In total we counted 82 relationships.

76 References

77 R Core Team (2015). R: A language and environment for statistical
78 computing. R Foundation for Statistical Computing, Vienna, Austria.
79 Gamfeldt, L., H. Hillebrand, and P. R. Jonsson. 2008. Multiple functions increase
80 the importance of biodiversity for overall ecosystem functioning. *Ecology*
81 89:1223–1231.