Artificial ecosystem selection to deduce the mapping between microbial community structure and ecosystem function

Andrew H. Morris and Brendan J. M. Bohannan

# Abstract

# Introduction

## Determining the mapping between microbial community structure and ecosystem function

Microbial communities mediate ecosystem nutrient cycling via decomposition, greenhouse gas cycling, carbon storage, and other processes. Despite the central role of bacteria in performing these processes, we have not found a consistent relationship between the composition of a microbial community and the rate of an ecosystem process across habitats. Typically, research in this area has followed the model of macroorganismal biodiversity-ecosystem function studies such as grassland productivity studies that measure or manipulate the diversity of plant species and correlate that with the rate of primary productivity. By analogy, microbial ecologists use microbial sequencing data to create taxonomic relative abundance tables and then infer diversity or relative abundance from the whole archaeal/bacterial community and correlate that with the rate of a microbial process. Alternatively, they estimate the abundance of a putative functional gene (such as *pmoA* for methane oxidation) and correlate abundance with function. These approaches only rarely show significant, positive relationships between biodiversity and ecosystem function. In my first chapter published in Phil Trans I lay out philosophically why this approach shouldn’t work – namely that marker gene sequences are not analogous to plant species counts, but are more analogous to measurements of gene markers as in evolutionary biology studies.

In evolutionary biology, genotype-phenotype mapping provides a way to understand the fitness effects of a genotype. While selection typically acts on phenotype, the genotypic changes are the fundamental driver of phenotypic differences and thus fitness differences (Stadler and Stephes 2005). A similar analogy can be drawn for biodiversity-ecosystem function relationships. Traits of a community are what drive community assembly, not the taxa or genotypes themselves (Inkpen et al. 2017). However, as ecologists, we wish to understand the effects of species on ecosystem function. Therefore, in a similar way that evolutionary biologists want to deduce the genotype-phenoype map to understand the genetic basis of fitness, community ecologists must describe the “mapping” between community genotype and ecoystem function to understand what drives communty assembly, and ultimately, ecosystem function.

The relationship that ecologists have described between plant diversity and productivity represents a relatively simple map between taxonomy and function. The number of unique members of a functional group (i.e., plant species) is correlated with the rate of the function performed by that group. However, for microbial communities that mapping is not quite as clear. For many microbial traits, there is not a clear coupling between taxonomy and function, except for in the case of highly conserved traits involving many genes (Martiny et al. 2013). In addition, most microbial taxa have not been sampled and so we don’t have a full picture of which taxa are members of a particular functional group (citation). Put another way, it’s much harder to identify methanotrophic prokaryotes in soil than it is to identify plants in a prairie. Finally, even for highly conserved traits with a putative functional marker, for example methane oxidation and the gene *pmoA*, functional group abundance is not correlated with the rate of the function performed by that group in most ecosystems (Rocca et al. 2016). Because of these differences between macroorganisms and microorganisms, it may be more useful to approach microbial biodiversity-ecosystem function in a way analogus to genotype-phenotype mapping in organisms (Morris et al. 2020).

One approach to deducing the genotype-phenotype map in organisms is through artificial selection (Fuller et al. 2005). In this approach, selection for a trait is imposed in a contrived laboratory environment and compared to laboratory environment without selection on that trait. By comparing the genetic respones between these two treatments, you can remove the environment-fitness connection that would be present in a comparative study and identify genes that are evolving in response to selection only on the trait of interest. It has been shown that whole communities can respond to artifical selection on traits of the ecosystem (Goodnight et al. 1997, Williams and Lenton 2007). And it has been proposed that this could be a suitable approach for identify communities of organisms that are capable of performing a particular function at a high rate (Swenson et al. 2000). Therefore, artificial ecosystem selection is a potentially fruitful approach to identifying the map between biodiversity and ecosystem function in microbial communities.

To identify the map between microbial community composition and ecosystem function, we imposed artificial ecosystem selection on methane oxidation rate of whole soil ecosystems. We prepared two lines of replicate ecosystems: one in which we imposed positive selection on methane oxidation by only using the top performing ecosystems to generate the next set of ecosystems and one in which we imposed neutral selection where a random set of ecosystems were chosen to generate the next set of ecosystems. We repated this process over multiple passages and then compared the resultant metacommunities in terms of their rate of ecosystem function and the composition of their communities.

With this experiment, we asked several questions. First, is there a response to selection on methane oxidation by soil ecosystems? And how strong is the response to selection? (i.e., how “heritable” is methane oxidation as a community trait in terms of the phenotypic resemblence of ecosystems to the ecosystems that generated them?) If methane oxidation is heritable, then what is the nature of the map between microbial community composition and methane oxidation? Is it the diversity of methane oxidizers, as it is for plants and productivity? Or is there a more complex map involving upstream metabolic regulation?

Things this Intro doesn’t yet discuss:

1. What a more complex map might look like. For example, maybe pmoA and methanotrophy aren’t correlated, but maybe understand methanogenesis or nitrogen cycling is helpful.
2. The idea of “heritability” for communities/ecosystems and why that relates to environmental vs. genetic control over ecosystem function.

# Materials & Methods

## Selection experiment

We performed an artificial ecosystem selection experiment similar to Swenson et al. (2000) by passaging replicate soil microbial communities. The trait we selected on was CH4 oxidation rate. Our experiment had two selection lines with twelve jars each for a total of twenty-four jars per passage. One line was a positive selection line where the two or three jars with the highest CH4 oxidation rate were chosen to inoculate the next set of jars. The other line was a neutral selection line where an equal number of jars as the positive line were chosen at random among the twelve. The number of jars chosen was based on the distribution of fluxes among the positive jars, i.e., we chose the top three jars unless only two jars had considerably greater oxidation rates based on visual inspection of histograms.

The initial soil microbial community was sampled from the top 10 cm of an upland mineral soil under a deciduous forest ecosystem near the University of Oregon campus. Incubations were performed in 500 mL mason jars (Ball Corporation, Broomfield, Colorado, USA) with rubber septa installed in the lids. Jars were incubated at ambient temperature on the benchtop. To create the first set of jars, twenty-four jars were created in an identical manner and then randomly assigned to either the positive or negative selection line. Each jar was sterilized with 70% ethanol to which was added 45 g of autoclaved potting mix, 5 g of live soil, and 3.5 mL of sterile deionized water to bring them to 60%??? of field capacity . The jars were then homogenized, capped, and injected with 4.3 mL of 99% CH4 to bring the headspace concentration to approximately 1000 ppm CH4. Jars were flushed and respiked twice per week to maintain elevated CH4 concentrations until travel, holidays, and COVID-19 got in the way and then we just did it as often as we could. Jars were incubated for approximately three weeks and then flux measurements were taken.

Once CH4 oxidation rates were determined, the top two or three jars were chosen for selection in the positive line and an equal number of jars were randomly chosen from the neutral line. The selected jars for each line were homogenized and used as the 5 g live soil for the subsequent passage. Twelve new jars were created for each line in the same manner as above using the inoculum for that line. This procedure continued until a divergence was observed in the mean flux of each line for two consecutive generations. Jars in passage number 2 and 5 were homogenized and a sample was collected and stored at -80C for later DNA extraction.

## Methane flux measurements Methane oxidation rates were determined after

flushing and spiking jars to approximately 1000 ppm CH4. Head-space samples of 1 mL were collected from each jar immediately after spiking and then at time points 3, 6, 24, and 48 hours for a 5-point curve. Samples were immediately injected into a ??? gas chromatograph fitted with an electron capture detector (company, city, state, USA) to determine the headspace CH4 concentration. Fluxes were calculated from a first-order exponential decay function as decay constant k. Oxidation rates are presented as the additive inverse of k (i.e., -k) so that a more positive value represents a faster oxidation rate.

## Soil DNA extraction and sequencing

Soil DNA was extracted from a 0.25 g subsample of soil from the starting inoculum and from each jar in passage 2 and 5. Negative controls were extracted from autoclaved potting mix and DNase-free water. Extractions were performed using the DNeasy PowerSoil kit (QIAGEN, Hilden, Germany) and quantified using Qubit (company, city, state). In order to estimate the diversity and relative abundance of the bacterial and archaeal taxa in our soil ecosystems, we sequenced the V3-V4 region of the 16S rRNA gene using the Earth Microbiome Project protocol and primers???. In addition, to estimate the diversity and relative abundance of CH4 oxidizing taxa, we sequenced the particulate methane monooxygenase subunit A gene (*pmoA*) using primers???. Sequencing was performed on the Illumina MiSeq in v3 mode with 300 bp paired-end reads.

## Bioinformatics Bioinformatic analyses were conducted using the QIIME2

bioinformatics platform (Bolyen et al. 2019). Raw sequencing reads were demultiplexed using ??? (citation), denoised using DADA2 (citation). Taxonomic assignment for the 16S reads was performed using ??? based on the Silva database. All subsequent tatistical analyses were performed in the R statistical programming environment (R Core Team 2020). Richness estimates were made using the breakaway package and diversity indices using the DivNet package (Willis et al. ???).

## Statistical Analysis

### Response to selection

To test whether there was a significant change in CH4 oxidation rate (k) as a response to selection, we tested the deviation between the positive and neutral selection lines over multiple passages. We calculated deviation as the difference in the mean oxidation rate between the positive and neutral treatment within each passage. The standard errors of the difference were calculated as the square root of the sum of the squared errors of the positive and neutral lines. We then fit an ordinary least squares regression model taking into account the standard errors of the deviations. The model was fit using the rma function from the metafor package in r (Viechtbauer 2010).

### Ecosystem heritability

To estimate the proportion of variance in ecosystem function due to additive genetic variance of the microbial communtiy, we applied the Breeder’s equation to the results of our selection experiment. the heritability of CH4 oxidation rate as an ecosystem property, we calculated narrow-sense heritability () as the regression of offspring on mid-parent (Goodnight 2000, Falconer 1997). The mid-parent was the mean for all three selected jars and the offspring was the mean for all jars produced by those parents.

In addition, we calculated the inbreeding coefficient, , as:

where is the population size and is the number of generations (Goodnight 2000 or Falconer 1997).

### Taxa differentially abundant in the positive selection treatment

To identify taxa that responded to selection on CH4 oxidation rate, we tested differential abundance using a beta-binomial model (Martin et al. 2020). We compared taxa in passage 5 between the positive and neutral treatment to identify taxa that

Possible tests:

* Random forest or gradient boosting to predict Positive vs. Neutral.

# Results

## Response to selection

Methane oxidation rate increased as a response to selection in the Positive selection treatment (Figure 1). This can be seen in the increase in deviation between the Positive and Neutral selection lines over the course of the experiment (z = 2.60, p = 0.009). Mean CH4 oxidation rate for the Positive treatment increased by 0.032 0.012 per passage relative to the Neutral treatment. The mean flux across all jars in all treatments in the first passage was 0.036 0.015 so this represents a doubling of CH4 consumption in every passage due to selection.

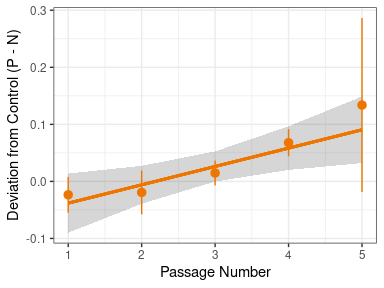


Figure 1: Deviance of CH4 oxidation rate for the Positive selection treatment relative to the Neutral selection treatment. Deviance is calculated as the mean of P within each passage minus the mean of N within each passage. Errorbars are the standard error of the difference between P and N for each passage. The best fit line is an ordinary least squares fit taking into account the standard errors of the deviations. The gray ribbon is the 95% confidence interval for the regression.

## Ecosystem heritability

Methane oxidation rate was heritable between consecutive passages (Figure 2). Specifically, the narrow-sense heritability, calculated as the slope of the mid-parent and mid-offspring, was = 1.46 0.97 (t = 1.51, p = 0.183). With an inbreeding coefficient of 0.19,

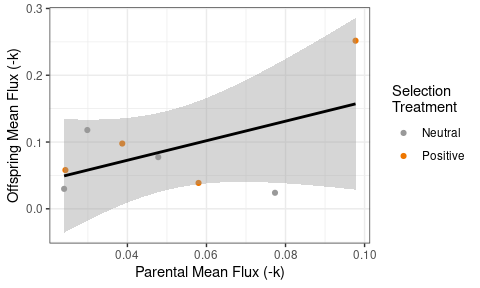


Figure 2: Regression of offspring CH4 flux on mid-parent CH4 flux. Mid-parent is the mean of jars selected to inoculate the next passage. The mid-offsprings are the mean of all twelve jars produced in one passage. The best fit line is an ordinary least squares regression line and the gray ribbon is the 95% confidence interval for the regression coefficient.

# Discussion

## Future Directions

This study focused on the practical implications of using AES as a tool to deduce how microbes map onto functions at the whole-ecosystem level. If one were interested in the evolutionary dynamics underlying this process, one could design a study to distinguish between individual selection on microbial strains and group selection or multilevel selection. That is, selection that cannot be recapitulated through selection at the individual level (Williams and Lenton 2007). This could be accomplished by culturing differentially abundant taxa in the positive selection treatment and then inoculating them in isolation (or in a non-selected background community) to see if the same level of ecosystem function could be recapitulated as in the final generation of the positive selection treatment.

Another avenue for research could be to look at correlated responses of ecosystem-level traits to selection. This would address questions such as, “Is methane oxidation rate correlated with other ecosystem processes due to pleiotropy or linkage disequilibrium?” At the community level, pleiotropy might be defined as “the production by a single taxon of two or more unrelated effects.” Further, linkage disequilbrium would represent correlated traits due to taxa who tend to co-associate due to shared environmental preferences or coupled metabolic pathways.

These experiments would further help lus understand the nature of the “map” between community structure and ecosystem function. This would not only reveal how best to predict ecosystem function in earth system models, but also demonstrate the mechanism underlying that relationship. This is akin to deducing the genetic control of trait variation in molecular biology.

Community epistasis is analogous to functional redundancy.

# Acknowledgments

This project was supported by the National Science Foundation Graduate Research Fellowship Program (grant no. DGE 1255832) and the ARCS Foundation Florence and Mike Nudelman Scholarship.

# Competing Interests

We declare we have no competing interests.

# References