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

Andrew H. Morris and Brendan J. M. Bohannan

# Abstract

# Introduction

Soil microbial communities are central to the functioning of ecosystems (Crowther et al. 2019). And yet, microbes have historically been excluded from ecosystems models in favor of abiotic variables (D. S. Schimel 1995). This is in part because attempts to establish quantitative relationships between microbial marker genes and ecosystem functions have not been fruitful Rocca et al. (2015). Here we use a novel approach to estimate how much of the variation in ecosystem function can be attributed to variation in the microbial community. We then identify microbial markers of ecosystem function. These experiments will expand our ability to model microbial ecosystem-function relationships and allow us to manage ecosystems for particular ecosystem functions J. P. Schimel and Gulledge (1998).

In macroorganismal communities, there appears to be a relationship between the diversity of a community and the resilience of ecosystem functions to environmental perturbations (Hooper et al. 2005). Typically, diversity is represented as the richness of a particular functional group, such as photosynthetic plants. Similar studies in microbial communities rarely find a relationship between diversity and function (Graham et al. 2016). However, these studies define diversity much more broadly by looking at the richness of the entire bacterial and archaeal community using 16S rRNA gene variants. Studies that focus more narrowly on a particular functional group still rarely find a significant correlation between the abundance of a functional group and the rate of the corresponding process (Rocca et al. 2015). For certain functions in certain ecosystems, these relationships can be informative. For example, the abundance of a single methanogenic species (Candidatus ‘*Methanoflorens stordalenmirensis*’) is correlated with the CH4:CO2 ratio in a permafrost ecosystem (McCalley et al. 2014). In addition, bacterial and archaeal community composition is correlated with methane emissions in forests and pastures of the Brazilian Amazon (Meyer et al. 2020). Therefore, there appears to be some indication that microbial community structure drives variation in methane cycling in soils.

Why have we been unsuccessful at identifying relationships between microbial community structure and ecosystem functions? Microbial communities are extremely diverse, particularly in soil environments (Whitman, Coleman, and Wiebe 1998). Characterizing these communities using broad phylogenetic markers, such as 16S, reveals thousands of mostly uninformative taxa which makes it difficult to identify markers of ecosystem function. Some studies constrain this diversity by characterizing individual protein-coding genes. However, this approach assumes investigators already know what limits the rate of a process in a particular ecosystem. Typically, this is assumed to be the gene that encodes the final step in a pathway–for example, the *mcrA* gene, which encodes a subunit of the enzyme that performs the final step in methanogenesis. However, regulation may happen further upstream (or downstream) within the pathway and therefore the abundance of the final gene would not correlate with the rate of the process. In the case of methane emissions, the limitation may be in generating the substrates for methanogenesis or may be in the consumption of methane by methanotrophs.

To further investigate the relationship between microbial community structure and ecosystem function, we want to determine how much of the variation in soil methane cycling is attributable to variation in the microbial community. This ecological question is analogous to the problem of estimating heritability in quantitative genetics (Morris, Meyer, and Bohannan 2020). Variation in an organismal trait is determined by the sum of genetic variation, environmental variation, and the interaction between the two (Falconer and MacKay 1996). One goal of quantitative genetics is to determine the proportion of the variation attributable to genetics. This is commonly estimated as the narrow-sense heritability (h2) defined as the additive genetic variance. This goal can be achieved through artificial selection experiments with selection on the character of interest (Falconer and MacKay 1996). We applied artificial ecosystem selection to whole microbial communities by selecting on the ecosystem-scale methane oxidation rate in order to estimate the additive genetic variance of soil methane oxidation (Swenson, Wilson, and Elias 2000).

To identify which organisms are important for regulating methane cycling, we compared the composition of the artificially selected microbial communities and a control set of communities without selection. Artificial ecosystem selection has a similar effect to enrichment culturing by amplifying the population of interest - in this case methanotrophs (Swenson, Wilson, and Elias 2000). This will reduce the diversity of soil microbes and allow for greater power in detecting significant markers of methane oxidation rate. To do this, we extracted DNA from several of the ecosystems within the selection experiment and sequenced the 16S rRNA gene in order to perform a differential abundance test on methane oxidation rate.

# Materials & Methods

## Selection experiment

We performed an artificial ecosystem selection experiment (*sensu* Swenson, Wilson, and Elias 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 positive jars. The other line was a neutral selection line where an equal number of jars as the positive line were chosen at random to inoculate the next set of neutral jars. 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 experiment was carried out over five passages.

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 in Eugene, OR, USA. Incubations were performed in 500 mL mason jars with rubber septa installed in the lids. Each jar was sterilized with 70% ethanol to which was added 45 g of autoclaved potting mix, 5 g of living 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. To create the two treatment groups, twenty-four jars were created in an identical manner and then randomly assigned to either the positive or neutral selection line. Jars were flushed and respiked twice per week to maintain aerobic conditions and elevated CH4 concentrations and were incubated at ambient temperature for approximately three weeks. Methane oxidation rates were determined at the end of the incubation period. These values were then used to determine the jars selected to inoculate the next set of jars.

## Methane oxidation rate measurements

Methane oxidation rates were determined after flushing and spiking jars to approximately 1000 ppm CH4. Headspace 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 SRI model 8610C gas chromatograph (Torrance, CA, USA) to determine the headspace CH4 concentration. Fluxes were calculated from a first-order exponential decay function as decay constant k with units day-1. Oxidation rates are presented as the additive inverse of k (i.e., -k) so that a more positive value represents a greater oxidation rate. The jars selected for the positive treatment in passage 2 had the lowest methane oxidation rate of the twelve jars due to a calculation error in the methane oxidation rate. All other passages correctly used the top three jars.

## Soil DNA extraction and sequencing

A subsample of soil from the starting inoculum and from every jar in passages 2 and 5 was collected and stored at -80C. Soil DNA was extracted from 0.25 g soil. Negative controls were extracted from autoclaved potting mix and DNase-free water. Extractions were performed using the DNeasy PowerSoil kit (QIAGEN, Germany) and quantified using Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, USA). In order to estimate the diversity and relative abundance of the bacterial and archaeal taxa in our soil ecosystems, we sequenced the V4 region of the 16S rRNA gene using the using the 515F − 806R primer combination (Caporaso et al. 2011). PCR mixtures were: 10 l NEBNext Q5 Hot Start HiFi PCR master mix, 9.2 l primer mixture (1.09 M concentration), and 0.8 l of DNA template. Reaction conditions were: 98C for 30 s (initialization), 35 cycles of 98C for 10 s (denaturation), 61C for 20 s (annealing), and 72C for 20 s (extension), and 72C for 2 m (final extension). Amplicons were purified twice using 0.8x ratio Mag-Bind RxnPure Plus isolation beads (Omega Bio-Tek, USA). Sequencing libraries were prepared using a dual-indexing approach (Fadrosh et al., 2014; Kozich et al., 2013). Amplicon concentrations were quantified using Qubit (Thermo Fisher Scientific Technologies, USA) multiplexed at equimolar concentration. Sequencing was performed at the University of Oregon Genomics Core Facility on the Illumina NovaSeq 6000 with paired-end 150 bp reads.

## Bioinformatics

Bioinformatics processing was performed in R (R Core Team 2021). Demultiplexed sequencing reads were denoised using DADA2 (Callahan et al. 2016). Taxonomic assignment for the 16S reads was performed using the RDP naive Bayesian classifier (Wang et al. 2007).

## 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 a difference in slopes between the positive and neutral selection lines. The CH4 oxidation rates were strongly right skewed with most values close to zero and few large, positive values. This resulted in residuals that did not meet the assumptions of constant variance and normal distribution. Therefore, CH4 oxidation rates were log10 transformed to better meet the assumptions of a linear model and to make figures easier to interpret. First, we tested the effect of treatment by fitting two nested models with and without treatment using lm in R. The model formulation for the full model was flux ~ passage \* treatment and the model formulation for the reduced model was flux ~ passage. We compared these models using the likelihood ratio test with the anova function in R. We then fit the full model to estimate the slope of the positive line, which represents the change in CH4 oxidation rate per passage as a response to selection.

### Ecosystem heritability

To estimate the proportion of variance in ecosystem function due to variation in the microbial community, we estimated narrow-sense heritability () as the regression of offspring on mid-parent (Falconer and MacKay 1996). The mid-parent was the mean for all three selected jars and the offspring was the mean for all jars produced by those parents. First, we tested if there was an effect of treatment on the heritability estiamte. The model formulation for the full model was offspring ~ parent \* treatment and the model formulation for the reduced model was offspring ~ parent. We compared these models using the likelihood ratio test with the anova function R. We then fit the full model to estimate the heritability of the Positive and Neutral treatments.

### Community response to selection

Diversity estimates and tests were performed using DivNet (Willis and Martin 2020). We tested a difference in Shannon diversity between the Positive and Neutral treatment and between the beginning and the end of the experiment with model formulation diversity ~ treatment \* passage. We also tested a difference in beta-diversity as both Bray-Curtis distance and Jaccard distance between the Positive and Neutral treatment and the beginning and end of the experiment with model formulation diversity ~ treatment \* passage.

To identify taxa that responded to selection on CH4 oxidation rate, we tested differential abundance using a beta-binomial model (Martin, Witten, and Willis 2020). We compared taxa in passage 5 between the positive and neutral treatment to identify taxa that were enriched or depleted as a response to selection. We also tested differential abundance with differences in CH4 oxidation rate to identify taxa correlated with CH4 flux.

Possible tests:

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

Test if random forest or gradient boosting can predict whether this is the positive line or the neutral line based solely on the microbial community?

# Results

## Response to selection

We observed a response to selection on soil methane oxidation rate at the whole ecosystem level (Figure 1). In addition, the response to selection varied with treatment (LRT of nested models with and without treatment: df = 2, ss = 1.86, p = 0.02). At the start of the experiment, the Positive treatment had a mean methane oxidation rate that was 24% lower than the Neutral treatment (difference of y-intercepts = -0.34, SE = 0.16, t = -2.14, p = 0.03). There was no change in methane oxidation rate of the Neutral treatment over the five passages (slope = -0.01, SE = 0.05, t = -0.26, p = 0.80). By contrast, the Positive treatment had a 51% increase in methane oxidation rate per passage (slope = 0.18, SE = 0.06, t = 2.76, p = 0.01).

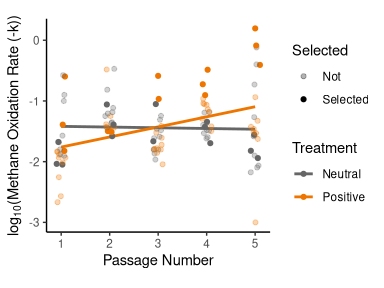


Figure 1: Response to selection on soil CH4 oxidation rate fit by ordinary least squares regression. The y-axis is the log10 of the additive inverse of the first-order exponential decay constant k (i.e., -k) with units day-1. The x-axis is the passage number. The orange line and points are the Positive treatment and the gray line and points are the Neutral treatment. The dark points are the selected jars in each passage and the light points are the non-selected jars. Regression lines aIn passage 2, the jars with the small -k were selected due to an error in the flux calculations.

## Ecosystem heritability

We estimated heritability as the regression of mid-offspring on mid-parent. Offspring CH4 oxidation rates were correlated with parental CH4 oxidation rates in both the Positive treatment (slope = 1.08, SE = 0.24, t = 4.43, p = 0.01) and the Neutral treatment (slope = -0.58, SE = 0.20, t = -2.89, p = 0.04). Notably, the sign of the effect was reversed with a positive heritability for the Positive treatment and a negative heritability for the Neutral treatment.

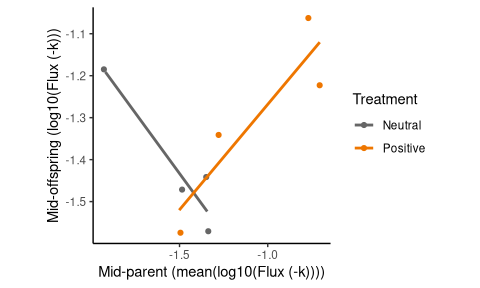


Figure 2: Ordinary least squares regression of mid-offspring CH4 flux on mid-parent CH4 flux. Mid-parent is the mean of the jars selected to inoculate the next passage. Mid-offspring are the mean of all twelve jars produced in one passage. The orange lines and points are the Positive treatment and the gray lines and points are the Neutral treatment.

## Community Response

Alpha diversity - treatment - flux

Beta diversity - treatment - flux

Differential abundance - treatment - flux

# Discussion

We observed a response to selection on soil CH4 oxidation rate of the whole soil microbial community. This type of response has been observed for other functions, such as chloroaniline degradation in water (Swenson, Wilson, and Elias 2000), but is the first, to our knowledge, of selection being performed on soil biogeochemical cycling.

If this experiment had been run for longer, we might be expected to see crashes in methane oxidation as the process of random sampling at each generation

Ecosystem heritability sounds confusing because it is difficult to imagine heritability without discrete generations. And how could a soil have “generations” (in the absence of coalescence events? (cite that one paper…)). However, this definition of heritability is conflated with the colloquial use of the term as referring to *inheritance*. But looking at the definition of narrow-sense heritability, it is simply the additive genetic variation associated with a trait. In this case, additive genetic effect would refer to the identity and relative abundance of microbial taxa in the metagenome. This notably excludes any interactions among taxa and interactions with the environment (i.e., GxE effects or context-dependent effects).

The heritability we observed was quite large at h2 = 1.08. However, given the small sample size and and wide confidence intervals (0.40, 1.75

The negative heritability of methane oxidation rate in the Neutral selection treatment may result from selection on alternative characters in those populations. The strong, positive heritability in the Positive treatment … . It seems strange that the 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 = (t = , p = ). With an inbreeding coefficient of ,

## 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.

# Supplement

First, we found an effect of treatment by comparing the full model to the reduced model (Likelihood ratio test: df = 2, ss = 1.86, p = 0.02).

# Junk

To identify communities of bacteria and archaea that collectively perform a high rate of methane oxidation, we performed an artificial ecosystem selection experiment (Swenson et al. 2000). We generated twenty-four soil ecosystems in the lab using a sterilized potting mix as the substrate. We then inoculated these ecosystems with a small amount of living soil to generate variation in community structure across the twenty-four ecosystems. These ecosystems were then assigned to one of two selection treatments: neutral selection or positive selection. Each ecosystem was maintained at 1000 ppm CH4 over several weeks to allow for colonization and growth of the microbial community and to enrich for methane oxidizers. We then determined the methane oxidation rate for each ecosystem. To impose selection, we selected three jars from the positive treatment with the highest rate of methane oxidation, homogenized them, and used that soil to inoculate the next set of jars. For the neutral treatment, we selected three jars at random to inoculate the next set of jars. This process continued for five passages. At the end of the experiment, we extracted DNA from soils in the second and fifth passage for both the neutral and positive treatments. By comparing the response to selection within the community between the two treatments, we can identify microbial species that contribute to variation in ecosystem methane emissions.

This approach is powerful because by using a common soil substrate and maintaining a constant headspace concentration of methane, we can eliminate much of the environmental variation that would be present in an observational study. In addition, by applying selection on methane oxidation rate in the positive treatment, we can enrich for taxa involved in methane cycling making it easier to perform association mapping between microbial taxa and ecosystem methane emissions. In this paper, we address whether there is a response to selection on methane oxidation rate at the whole ecosystem level. We then quantify the amount of variation in methane oxidation attributable to variation in community composition within our laboratory environment. Finally, we identify microbial markers of soil methane cycling using association mapping in order to deduce the mapping between microbial community structure and ecosystem function.

(Goodnight et al. 1997, Williams and Lenton 2007)

# References

Callahan, Benjamin J., Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes. 2016. “Dada2: High-Resolution Sample Inference from Illumina Amplicon Data.” *Nature Methods* 13 (7): 581–83. <https://doi.org/10.1038/nmeth.3869>.

Caporaso, J. Gregory, Christian L. Lauber, William A. Walters, Donna Berg-Lyons, Catherine A. Lozupone, Peter J. Turnbaugh, Noah Fierer, and Rob Knight. 2011. “Global Patterns of 16S rRNA Diversity at a Depth of Millions of Sequences Per Sample.” *Proceedings of the National Academy of Sciences* 108 (Supplement 1): 4516–22. <https://doi.org/10.1073/pnas.1000080107>.

Crowther, T. W., J. van den Hoogen, J. Wan, M. A. Mayes, A. D. Keiser, L. Mo, C. Averill, and D. S. Maynard. 2019. “The Global Soil Community and Its Influence on Biogeochemistry.” *Science* 365 (6455). <https://doi.org/10.1126/science.aav0550>.

Falconer, D. S. (Douglas Scott), and Trudy F. C. MacKay. 1996. *Introduction to Quantitative Genetics*. Harlow : Prentice Hall.

Fuhrman, Jed A. 2009. “Microbial Community Structure and Its Functional Implications.” *Nature* 459 (7244): 193–99. <https://doi.org/10.1038/nature08058>.

Graham, Emily B., Joseph E. Knelman, Andreas Schindlbacher, Steven Siciliano, Marc Breulmann, Anthony Yannarell, J. M. Beman, et al. 2016. “Microbes as Engines of Ecosystem Function: When Does Community Structure Enhance Predictions of Ecosystem Processes?” *Frontiers in Microbiology* 0. <https://doi.org/10.3389/fmicb.2016.00214>.

Hooper, D. U., F. S. Chapin, J. J. Ewel, A. Hector, P. Inchausti, S. Lavorel, J. H. Lawton, et al. 2005. “Effects of Biodiversity on Ecosystem Functioning: A Consensus of Current Knowledge.” *Ecological Monographs* 75 (1): 3–35. <https://doi.org/10.1890/04-0922>.

Martin, Bryan D., Daniela Witten, and Amy D. Willis. 2020. “Modeling Microbial Abundances and Dysbiosis with Beta-Binomial Regression.” *The Annals of Applied Statistics* 14 (1): 94–115. <https://doi.org/10.1214/19-AOAS1283>.

McCalley, Carmody K., Ben J. Woodcroft, Suzanne B. Hodgkins, Richard A. Wehr, Eun-Hae Kim, Rhiannon Mondav, Patrick M. Crill, et al. 2014. “Methane Dynamics Regulated by Microbial Community Response to Permafrost Thaw.” *Nature* 514 (7523): 478–81. <https://doi.org/10.1038/nature13798>.

Meyer, Kyle M., Andrew H. Morris, Kevin Webster, Ann M. Klein, Marie E. Kroeger, Laura K. Meredith, Andreas Brændholt, et al. 2020. “Belowground Changes to Community Structure Alter Methane-Cycling Dynamics in Amazonia.” *Environment International* 145 (December): 106131. <https://doi.org/10.1016/j.envint.2020.106131>.

Morris, Andrew, Kyle Meyer, and Brendan Bohannan. 2020. “Linking Microbial Communities to Ecosystem Functions: What We Can Learn from Genotypephenotype Mapping in Organisms.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (1798): 20190244. <https://doi.org/10.1098/rstb.2019.0244>.

Rocca, Jennifer D., Edward K. Hall, Jay T. Lennon, Sarah E. Evans, Mark P. Waldrop, James B. Cotner, Diana R. Nemergut, Emily B. Graham, and Matthew D. Wallenstein. 2015. “Relationships Between Protein-Encoding Gene Abundance and Corresponding Process Are Commonly Assumed yet Rarely Observed.” *The ISME Journal* 9 (8): 1693–99. <https://doi.org/10.1038/ismej.2014.252>.

Schimel, David S. 1995. “Terrestrial Ecosystems and the Carbon Cycle.” *Global Change Biology* 1 (1): 77–91. <https://doi.org/10.1111/j.1365-2486.1995.tb00008.x>.

Schimel, Joshua P., and Jay Gulledge. 1998. “Microbial Community Structure and Global Trace Gases.” *Global Change Biology* 4 (7): 745–58. <https://doi.org/10.1046/j.1365-2486.1998.00195.x>.

Swenson, William, David Sloan Wilson, and Roberta Elias. 2000. “Artificial Ecosystem Selection.” *Proceedings of the National Academy of Sciences* 97 (16): 9110–14. <https://doi.org/10.1073/pnas.150237597>.

Wang, Qiong, George M. Garrity, James M. Tiedje, and James R. Cole. 2007. “Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy.” *Applied and Environmental Microbiology* 73 (16): 5261–67. <https://doi.org/10.1128/AEM.00062-07>.

Whitman, William B., David C. Coleman, and William J. Wiebe. 1998. “Prokaryotes: The Unseen Majority.” *Proceedings of the National Academy of Sciences* 95 (12): 6578–83. <https://doi.org/10.1073/pnas.95.12.6578>.

Willis, Amy D, and Bryan D Martin. 2020. “Estimating Diversity in Networked Ecological Communities.” *Biostatistics*, no. kxaa015 (May). <https://doi.org/10.1093/biostatistics/kxaa015>.