Microbes in Arctic Lakes

Biological Question

Do relative abundances of specific phyla, in microbial communities from sediments in Arctic lakes, correlate with classes of biolipids?

Introduction

Bacteria produce biolipids, also known as brGDGTs (branched glycerol dialkyl glycerol tetraethers), with varying abundances and distributions given environmental conditions such as temperature and pH (Pei et al., 2021). These biolipid can be classed based on their biochemical lipid compositions. Early biolipid research demonstrated that GDGTs were strongly with the survival bacterial and archaeal taxa found in extreme environments, such as hyperthermophilic microbes surviving in temperatures above 60°C (Schouten et al., 2007). Beyond, biochemical membranes in extreme environments, biolipid distribution of classes are often used to reconstruct paleolithic conditions of the ancient earth, which provide insight into early earth environmental conditions (Pei et al, 2021). What is less known about biolipid classes is the type of taxa that may be driving the lipid distributions in a microbial community. This is especially evident in soils and sediments in high altitude arctic lakes where early taxa may be correlated to a specific distribution of biolipid classes, which are linked to the abundance of taxa found in these arctic environments. My goal for this project was to illuminate community composition and distribution of microbes among different sample types from arctic sediments and compare relative abundances of specific taxa correlated with biolipids classes. The relationship between biolipids and environmental variables is well known, but the taxa related driving these biolipid trends are unknown.

Methods

Source of the data:

This is an original microbial dataset. However, the samples and biolipid information were provided by collaborators with the Institute of Arctic and Alpine Research (Raberg et al., 2021). The sequencing data was processed within the Fierer lab from DNA extracted from soils and sediments in arctic regions. The DNA sequences were processed using next-generation sequencing for the 16S rRNA gene the Illumina MiSeq. I prepared these sequences using the DADA2 bioinformatics pipeline in R (Callahan et al., 2016; Oliverio and Holland-Moritz, 2019) to create an ASV (amplicon sequence variants) table for further exploratory bacterial analysis and relative abundances, and regional and global trends associated with these types of samples.

Processed data:

I utilized the mctoolsr package in R (Leff, 2016), to combine both taxonomic and sequence ASV tables with metadata tables. I focused on only 74 of the original 150 samples from Iceland and Canadian soils and sediments for analysis because they contained the bulk of the metadata variables for comparison and to preserve privacy for this dataset.

The ASV data files were prepared to be combined with the microbiome r package mctoolsr (Leff, 2016) for microbiome processing. I processed all ASV files within that package, and the metadata file that contained a subset of the biolipid fractional abundances and environmental sample type, region, and pH. Then I combined the metadata and taxonomic sequence data to create an aggregated relative abundance of taxa in each sample at the phylum taxonomic level, and stored that table for later merging with metadata to conduct correlation and linear modeling using tidyverse (v. 1.3.1; Wickham et al., 2019). All graph visualizations were prepared with ggplot2 (v. 3.3.3.; Wickham, 2011) and cowplot (v.1.1.1; Wilke et al., 2020). Once relative abundances were aggregated, the metadata and relative abundance tables were merged using tidyverse for correlation and linear regression modeling. The size of the data tables varied with the sequences data 16.7 MB, the metadata file 7KB, and the aggregated relative abundance of the phylum level table was 53 KB in size.

In order to perform correlations, I merged the phyla relative abundance table and the metadata table with different fractional abundances of a small subset of biolipid classes. After initial correlation with all sample types, I focused my analysis further by subsetting the sediments in arctic lakes and selecting five phyla and five different bolipid for linear modeling.

I sought to look for correlations with some of the taxa at the phylum level and specific biolipid. However, extensive analysis to analyze every phylum with every biolipid class proved to be incredibly challenging. Therefore, I changed my approach to focusing on top phylum with only a couple of biolipids using linear modeling and correlations through tidyverse R packages. Any correlation with a p-value < 0.05 was considered significant. I also included pH in the analysis because pH has been associated with an influence on biolipid classes.

Results and Conclusions

Upon initial visualization, Proteobacteria was tested in all sample types. Its association with biolipid fractional abundances revealed distinct differences in the distribution of relative abundance to biolipid abundance dependent on sample type, with surface sediments showing high biolipid to a high relative abundance of phyla (Figure 1).

Chart, scatter chart

Description automatically generated

Figure 1: Scatter plot of Proteobacteria Relative Abundance vs. Biolipid Fractional Abundance. Color by type of sample.

The distinct sample type led to focusing analysis on only one sample type, surface sediments, with five different phyla Acidobacteria, Atribacteria, Cholorflexi, Proteobacteria, and Verrucomicobia. All common phyla which are abundant in soils and sediments. Out of the five taxa tested with linear modeling, Acidobacteria significantly correlated with three biolipid classes (Figure 2).

Chart, scatter chart

Description automatically generated

Figure 2: Scatter plots of Acidobacteria Relative Abundance vs. Biolipid Fractional Abundance. Color by pH and shape by region in the Arctic.

The visual depiction did not show strong linear association, but there was evidence from linear modeling with Proteobacteria significantly correlated with three classes fIIIa, fIIIb and fIIa\_methylation (R2 = 0.77; p = 0.00169, p = 0.02825, p = 0.01868, respectively). Both pH and the biolipid classes explained a high amount of variance in those models, supporting the overall idea that pH, biolipids and some taxa ae linked. Still, not all taxa demonstrate enough to show these correlations strongly and this could be explained by the variation that tend to occur from spatial differences given different sample sites in high latitude lakes.

Overall, this project was a proof of concept that we can potentially correlate biolipid classes based on their abundances to some taxonomic relative abundances in community samples. However, there are limitations to this approach because the analysis would need to be conducted on all classes and all taxa in the given sample types. If we are able to identify biolipid classes with specific taxa under varying environmental conditions, then we can link information about biolipids to specific taxa in low diversity sites or ancient sites.

References Cited

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods, 13(7), 581-583. https://doi.org/10.1038/nmeth.3869

Leff, J. W. (2016). mctoolsr: microbial community data analysis tools. available at github. com/leffj/mctoolsr.

Oliverio, A., Holland-Moritz, H. (2019). Dada2 tutorial with MiSeq dataset for Fierer Lab. Retrieved from https://github.com/amoliverio/dada2\_fiererlab

Pei, H., Zhao, S., Yang, H., & Xie, S. (2021). Variation of branched tetraethers with soil depth in relation to non-temperature factors: Implications for paleoclimate reconstruction. Chemical Geology, 572https://doi.org/10.1016/j.chemgeo.2021.120211

Raberg, J. H., Blumm, A., Harding, D. J., Crump, S. E., de Wet, G., Dildar, N., Kopf, S., Geirsdóttir, Á., Miller, G. H., and Sepúlveda, J.: Revised brGDGT fractional abundances and warm-season temperatures strengthen relationships between brGDGTs and environmental variables, EGU General Assembly 2021, online, 19–30 Apr 2021, EGU21-8362, https://doi.org/10.5194/egusphere-egu21-8362, 2021.

Schouten, S., Marcel T. J. van der Meer, Hopmans, E. C., W. Irene C. Rijpstra, Reysenbach, A., Ward, D. M., & Jaap S. Sinninghe Damsté. (2007). Archaeal and bacterial glycerol dialkyl glycerol tetraether lipids in hot springs of yellowstone national park. Applied and Environmental Microbiology, 73(19), 6181-6191. https://doi.org/10.1128/AEM.00630-07

Wickham, H. (2011). ggplot2. Wiley Interdisciplinary Reviews: Computational Statistics, 3(2), 180-185.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., ... & Yutani, H. (2019). Welcome to the Tidyverse. Journal of Open Source Software, 4(43), 1686

Wickham, H., Chang, W., & Wickham, M. H. (2016). Package 'ggplot2'. Create Elegant Data Visualisations Using the Grammar of Graphics. Version, 2(1), 1-189

Wilke, C. O. (2020). cowplot: streamlined plot theme and plot annotations for ‘ggplot2’. R package version 0.9. 2; 2017. URL https://CRAN. R-project. org/package= cowplot.