Simulating Nitrogen Interactions Within Microbial Communities From Metagenome Assembled Genomes (MAGs) and Genome-Scale Metabolic Models (GEMs)

Jorge Lopez-Nava[1, 2], Zoey Werbin[3], Jennifer Bhatnager[3]

[1]Swarthmore College, [2]Boston University Bioinformatics BRITE REU Program, Summer 2021,
[3]Department of Biology, Boston University

Abstract

Metagenomic analysis is a common method of studying unculturable bacteria within microbial communities, but it is difficult to retrieve complete genomes from hyperdiverse environments such as forest soils. Recently, metagenomic sequences have been published for thousands of soils from across the U.S. by the National Ecological Observatory Network (NEON). With 47 sites and 4 years of annual sequencing, NEON data can be used to explore how microbial genes and species, active in carbon and nutrient cycling processes, vary in space and time. Sample analysis was performed using an existing tool named metaGEM, a bioinformatics pipeline developed by Francisco Zorrilla. Interactions between the resulting GEMs were simulated with different nitrogen inputs.

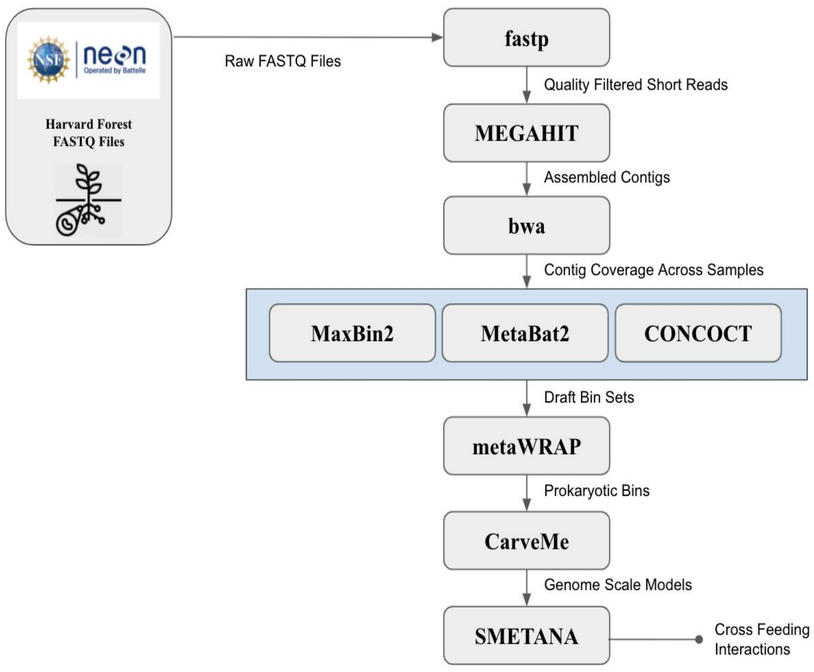
Background and Motivation

The Bhatnagar lab studies ecological communities and is using metagenomic samples from NEON to provide insight into the relationships responsible for metabolic processes in ecosystems. The goal of this project was to investigate how soil microbial communities, responsible for transforming nitrogen, can be represented using metabolic modeling of novel genomes. These metabolic processes drive nutrient recycling that the larger ecosystem depends on.

Methods

Metagenomic sequencing data from the Harvard Forest, in the form of FASTQ files, were run through the metaGEM pipeline using the steps below.

1. Quality Filter with **fastp**
2. Assembly with **MEGAHIT**
3. Draft bins with **MaxBin2, MetaBat2,** and **CONCOCT**
4. Refine and reassemble bins with **metaWRAP**
5. Reconstruct and evaluate GEMs with **CarveMe** and **memote**



Data and Results

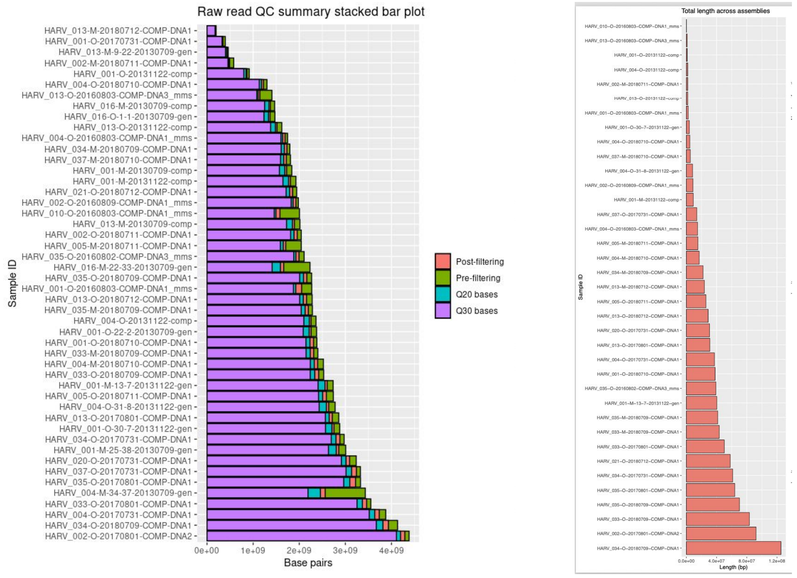


Figure 1: Quality Filtering across soil samples from Harvard Forest, MA. Quality Filtering summary is calculated from raw read length (base pairs) for each sample after fastp.
Figure 2: Assembly of each sample after MEGAHIT assembly step.

A metaGEM report (Figure 1) showed that NEON FASTQ files had substantial variation depth between samples. High-quality raw sequences improve the accuracy of gene annotation and assemblies. Similarly, another report (Figure 2) shows substantial variation in assembly length across samples. MetaGEM was able to successfully produce bins for samples with higher quality and more base pairs.

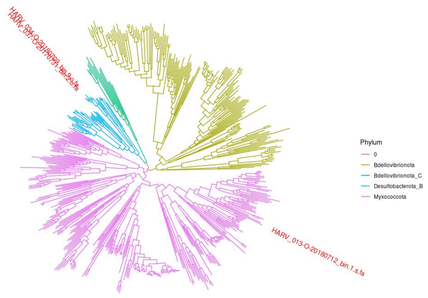


Figure 3: Phylogenetic tree showing the placement of X metagenome-assembled genomes (MAGs) within the reference tree provided by the Genome Taxonomy Database (GTDB).

Genomes were dereplicated using dRep before phylogenetic analysis using GTDB-tk. Newly-assembled MAGs are colored green, while reference genomes are gray. Asterisks indicate genomes that represent a possibly novel lineage.

Conclusion and Future Work

From 48 raw soil metagenome sequencing samples from the Harvard Forest NEON site, we assembled 41 medium-quality metagenome assembled genomes (MAGs)and genome-scale metabolic models (GEMs). Most genomes were not identifiable to the species level, due to either incompleteness or genetic novelty. The simulated metabolic exchanges were qualitatively consistent with soil nitrogen fluxes. The GEMs of these genomes revealed that interactions within microbial communities were dependent on different nitrogen inputs. These newfound conditions at the microbial level provide an important understanding of the local and global nitrogen cycling processes. Our approach therefore has potential for uncovering microbial controls on local and global nitrogen cycling processes.

While MAGs can be assembled directly from metagenomes, MAGs from soil microbes tend to be low quality and newer assembly techniques (long read assembly) might be necessary for obtaining good soil genomes. Most metabolic models came from just a few high-quality samples, usually these were from the organic layer. From these advanced techniques, we can improve overall scores shown in Figure 4.

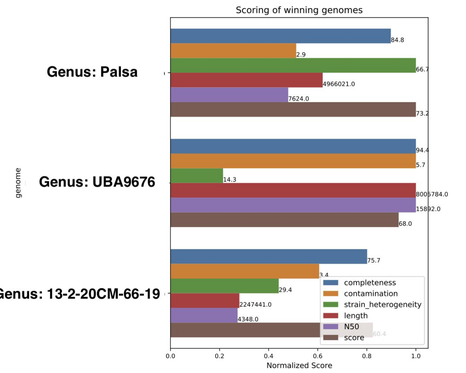


Figure 4: Genome scores showing measures of completeness, contamination, strain heterogeneity, length, N50, and overall score

Future models will use spatially-explicit flux-balance analysis, providing more realistic representations of soil microbial interactions and growth that influence global biogeochemistry. Though our work is ongoing, we anticipate learning about metabolic dependencies between microbes participating in nitrogen cycling, which could allow us to generate testable hypotheses about the function of soil bacteria in the wake of the climate crisis.

References and Acknowledgments

This work was funded, in part, by NSF grant DBI-1949968, awarded to the Boston University Bioinformatics BRITE REU program

- Werbin, Z. R., Hackos, B., Dietze, M. C., & Bhatnagar, J. M. (2021). The National Ecological Observatory Network's soil metagenomes: Assembly and basic analysis. F1000Research, 10, 299. https://doi.org/10.12688/f1000research.51494.1
- Chen, L.-X., Anantharaman, K., Shaiber, A., Eren, A. M., & Banfield, J. F. (2020). Accurate and complete genomes from metagenomes. Genome Research, 30(3), 315–333. https://doi.org/10.1101/gr.258640.119
- Cardona, C., Weisenhorn, P., Henry, C., Gilbert, J. (2016). Network-based metabolic analysis and microbial community modeling. Current Opinion in Microbiology, 31, 124-131. https://dx.doi.org/10.1016/j.mib.2016.03.008
- Francisco Zorrilla, Kiran R. Patil, Aleksej Zelezniak (2020). metaGEM: reconstruction of genome scale metabolic models directly from metagenomes bioRxiv 2020.12.31.424982; doi: https://doi.org/10.1101/2020.12.31.424982