Goldschmidt 2021 Abstract
https://doi.org/10.7185/gold2021.4886

Illuminating the Microbial Dark Matter Driving Energy Transformations in the Environment with a Universal Language of Life

ADRIENNE HOARFROST[1], ARIEL APTEKMANN[1], GONZALO FARFAÑUK[2], PAUL FALKOWSKI[1] AND YANA BROMBERG[1]

[1]Rutgers University

[2]Universidad de Buenos Aires

Presenting Author: adrienne.l.hoarfrost@gmail.com

The microbial communities that drive Earth's biogeochemical cycles are dominated by "microbial dark matter" of unknown identity and/or function. This unknown microbial diversity, coupled with the vast complexity of microbial interactions with their environment, limits our ability to accurately model the microbial systems responsible for energy transformation and biogeochemical regulation of the planet. Deep learning offers a promising toolbox for these complex systems, but typically requires infeasibly large datasets to perform well. Transfer learning provides a solution, applying domain knowledge learned in one model setting to a different but related problem, resulting in better models using less data than would otherwise be possible.

To facilitate deep transfer learning-based models in microbial systems, we developed LookingGlass, a deep model of the 'universal language of life'. LookingGlass is pretrained on millions of DNA reads from across the microbial tree of life, and captures the functionally, evolutionarily, and environmentally relevant features underlying microbial systems in general. Using LookingGlass as a starting point, we used transfer learning to create a model to classify novel oxidoreductases in the environment, identifying even those sequences with no close homology to known oxidoreductases. Oxidoreductases are the class of enzymes responsible for electron transfer, and underpin the metabolic processes that drive all biogeochemical cycles.

We applied this oxidoreductase model to a global dataset of marine metagenomes, and investigated global-scale patterns in oxidoreductase biogeography. 18-22% of reads in marine metagenomes were identified as oxidoreductases, and their relative abundances followed spatial patterns with depth and latitude. A higher proportion of oxidoreductases occupied mesopelagic depths relative to surface waters, and higher latitudes relative to lower latitudes. Furthermore, homology-based annotation tools could not annotate the majority of sequences (>60%), and did not capture the observed depth and latitudinal trends. This difference in observed geospatial trends highlights the importance of accounting for "microbial dark matter" for modeling energy transformations in the Earth system. Future work will leverage LookingGlass to target additional functional targets of biogeochemical importance, and to combine this functional information into a whole-metagenome embedding that can facilitate microbially explicit models of rates of biogeochemical processes across environmental gradients.