GENOMIC ANALYSES OF CANDIDATE PHYLUM EM19 POPULATIONS IN GEOTHERMAL SPRINGS USING A NOVEL READ-FIRST BINNING METHOD.

E. D. Becraft[1], J. A. Dodsworth[2], S. K. Murugapiran[2], I. Ohlsson[1], B. P. Hedlund[2] and W. D. Swingley[1*], [1] Northern Illinois University, DeKalb, IL, [2] University of Nevada Las Vegas, Las Vegas, NV. [*] Corresponding author: wswingley@niu.edu. First author: ebecraft@gmail.com.

The vast majority of microbial life remains uncatalogued due to the inability to cultivate these organisms in the laboratory. This "microbial dark matter" represents a substantial portion of the tree of life and of the populations that contribute to chemical cycling in most ecosystems. The Great Basin is home to many alkaline high temperature geothermal springs that harbor thermophilic microbes that are of interest for understanding how life originally evolved to the extreme conditions of early Earth environments and the biological limits of habitability in our universe. Great Boiling Spring (GBS) has become a model system for Great Basin studies, with multiple ecological gradients targeted for dark matter studies [1]. In this work, we leveraged an existing single-cell genomic dataset sequenced from Great Boiling Spring representing the candidate bacterial phylum Calescamantes (EM19) to define distinct but related metagenomic bins in Great Boiling Spring in Nevada's Great Basin, as well as Octopus Spring and Bison Pool in Yellowstone National Park, and Gongxiaoshe Spring in Yunnan Province, China. Taxonomic binning was achieved by using a novel, read-based approach using machine learning algorithms trained on single-amplified genomes (SAGs) and published genomes, which yielded genomes with a higher predicted genome completeness than existing assembly-based binning methods. Metabolic reconstruction suggests that Calescamantes are heterotrophic, facultative anaerobes, which can utilize oxidized nitrogen sources as terminal electron acceptors for respiration in the absence of oxygen, and use proteins as the primary carbon source. Despite their phylogenetic divergence and the high fraction of proteins of unknown function, the geographically separate Calescamantes populations were highly similar in their predicted metabolic capabilities and core gene content. This work represents the first detailed exploration of the Calescamantes microbial dark matter phylum in different environments, and demonstrates that read-based binning approaches can be a valuable tool in the coordinated analysis of community metagenomes and targeted single-cell genome sequencing, and provides a framework for further testing the metabolic capabilities of Calescamantes.

[1] Rinke C. et al. (2013) Nature 499:431-437.