University of Utah
UNDERGRADUATE RESEARCH JOURNAL

NOVEL LOST CITY METHANOGENIC ARCHAEA BELONGING TO THE ORDER METHANOSARCINALES

Briggs C. Miller (Faculty Mentor: William Brazelton)
School of Biological Sciences

The Lost City, which is located in the middle of the Atlantic Ocean, is a unique ecosystem that mimics the conditions found in Earth's oceans billions of years ago. Understanding the microbial diversity located here thus is one step toward understanding the complexity of life this environment supports and the strategies that the earliest lifeforms may have employed to survive such conditions. Methanogens, or organisms that are able to use methane in their metabolism, are one class of organism that can be found at the Lost City. Metagenomic analysis of samples from the Lost City revealed a novel genus and species of methanogenic archaeon belonging to the order *Methanosarcinales*. The Metagenome-Assembled Genomes (MAGs) representing this organism appear to be distinctly abundant at the Poseidon North Spire chimney which is also true of general methanogenesis genes. The presence of key methanogenesis genes, hydrogenases, and formate dehydrogenases provide evidence that this organism has the capacity to use methane, H_2, and formate in some aspect of its metabolism

Introduction:

The Lost City hydrothermal field is a series of hydrothermal vents situated upon the Atlantis Massif located in the North Atlantic Ocean. This notably rare ecosystem is fueled by hydrothermal circulation within the seafloor that transports compounds from deep within Earth to the organisms living on and within the chimneys. The Lost City is a unique ecosystem that mimics the conditions found in Earth's oceans billions of years ago. The chimneys vent fluids that are hot (up to 95 ºC), extremely alkaline (pH 9-11), and rich in methane and hydrogen gas. Due to these unique conditions, the Lost City provides a window into the history of life on Earth and provides possible clues into how life might have first emerged.

The hydrothermal vents of Lost City are home to countless microorganisms that rely on the unique chemistry of Lost City's venting fluids to obtain food and nutrients. Within this environment there is a great diversity of bacteria and archaea that join together to form biofilms and perform a wide variety of metabolic processes. Of the organisms present at Lost City, a select few have been investigated as potentially representative of this bizarre ecosystem and perhaps of the most ancient ecosystems on Earth.

Methanogens were the primary group of organisms of interest due to their ability to produce, and sometimes consume, methane. Bioinformatic methods were used to explore metagenomic sequences from the Lost City chimneys to search for potential genomes of methanogens.

Analysis of these hypothetical methanogens included genomic classification, calculation of normalized abundance, genome comparison, metabolic analysis, and phylogenetic assembly. The goal of this project was to investigate the diversity and genomic content of methanogens of the Lost City hydrothermal field in order to learn more about how they make a living in such extreme environmental conditions.

Methods:

Sample Collection and MAG Assembly

Sample collection from the Atlantis Massif and Lost City was carried out during the 2018 Lost City expedition aboard R/V *Atlantis* using the ROV *Jason*. DNA and RNA sequencing followed this collection and was accompanied by metagenomic assembly of collected data into Metagenome-Assembled Genomes (MAGs) for each sampling location and in a pooled assembly.

MAG Selection

Various methods were employed to select metagenome assembled genomes (MAGs) that had the highest possibility of belonging to methanogen microorganisms. From the pooled assembly, which combined samples from all chimneys into a single assembly, MAGs were chosen based upon whether or not there was a sufficient amount of methanogenesis genes present in the MAG. Of the total 156 MAGs in the pooled assembly only 26 were selected. Three additional MAGs were selected from the individual assemblies SOM1 (Sombrero 1), SOM2 (Sombrero 2), and PoCH (Poseidon Camel Humps) based upon their genomic classification using the GTDB-Tk application on KBase (The Department of Energy Systems Biology Knowledgebase).[4,3] The initial MAG selection contained a total of 29 MAGs and was further refined to include only four MAGs that had the most methanogenesis genes: BinSanityLC-kmean-bin_76-bin_2-refined_6, SOM1 Bin 62, SOM2 Bin 61, and PoCH Bin 22.

Calculation of Normalized Abundance

The general importance or abundance of the various selected MAGs was determined by calculating the overall proportion of the MAG sequences for each of the selected MAGs. Normalized abundance was calculated by dividing the sequence coverage for each of the MAGs by the total coverage for each of the specific chimneys. This value was then converted to FP or fragment proportions per million by multiplying by a factor of one million. The results of this calculation can be visualized in Figure 1.

MAG Comparison

Multiple techniques were used to determine the similarity between the selected MAGs and a MAG that was assembled from an earlier expedition to the Lost City Hydrothermal Field and presented by McGonigle, et al..[1] Genomic classification of the selected MAGs was done using the GTDB-Tk application on KBase.[4,3] Further comparisons were drawn between the MAGs using a Clustal Omega alignment[6] of the protein sequences for the mcrA (methyl coenzyme M reductase) gene. The results can be visualized in Figure 3.

Metabolic Gene Analysis
Genes selected for further analysis in methanogen MAGs included all genes utilized in the various methanogenesis pathways as well as genes that code for general hydrogenases. The genes selected were chosen based on their designation as utilized in methanogenesis by the GhostKOALA (KEGG Orthology And Links Annotation) database.[2] This database was used for detection of all genes in MAGs that can be used in any capacity in the various forms of metabolisms using methane.

Phylogenetic Analysis
Phylogenetic trees were assembled using sequences for the 16S rRNA section of the genome and for the mcrA gene. 16S sequences were compiled from the SILVA database[9] and selected based upon their measured relatedness to the MAG assembled by McGonigle, et al..[1] The mcrA sequences were found and selected based upon the similarity to the mcrA sequence of the provided MAG using the BLAST database.[10] In both cases, organisms were selected by genus to provide a broader perspective on the relatedness of the selected MAG to a more diverse set of organisms.

The 16S sequences obtained from the SILVA website were aligned by default. The mcrA sequences obtained from the BLAST database had to first be aligned to the mcrA gene from the extracted MAG.. This alignment was processed by Clustal Omega[6]. The 16S and mcrA trees were constructed using RAxML rapid bootstrapping[7] and visualized using Dendroscope.[8] These tree can be seen in Figure 4.

Results:

Each of the four selected MAGs originated from a different assembly. The BinSanityLC-kmean-bin_76-bin_2-refined_6 MAG was selected from the pooled assembly with an estimated completeness and contamination of 85.78% and 1.99% respectively. Bin 62 from the SOM1 assembly was estimated to have a completeness of 56.87% , Bin 61 from the SOM2 assembly was estimated to have a completeness of 73.86%, and Bin 22 from the PoCH assembly was estimated to have a completeness of 79.69%. All three of these MAGs had an estimated contamination of 0%. Similarity of the three chimney specific MAGs (SOM1 Bin 62, SOM2 Bin 61, and PoCH Bin 22) was done using Integrative Genomics Viewer (IGV)[5] and it appears that these three bins have very similar contig compositions.

The abundances of SOM1 Bin 62, SOM2 Bin 61, and PoCH Bin 22 were all much greater than the abundance of BinSanityLC-kmean-bin_76-bin_2-refined_6 as can be seen in Figure 1. Bin 22 form the PoCH assembly was also consistently the most abundant across all chimneys. Second, the abundances of all four MAGs is much greater in the Poseidon North Spire (PoNS) chimney than across the other chimneys. This may indicate that the environment of the Poseidon North Spire may be more hospitable to methanogens, or that methane may play a more important role within this location.

In addition to the abundances of the various MAGs, the abundances of specific methanogenesis genes was calculated and summarized within Figure 2. Overall abundances of the various methanogenesis genes appear to be greater at the Poseidon North Spire chimney than at the other

locations. This piece of data matches the observation made in Figure 1 that the selected methanogenesis bins had a greater abundance at the Poseidon North Spire chimney.

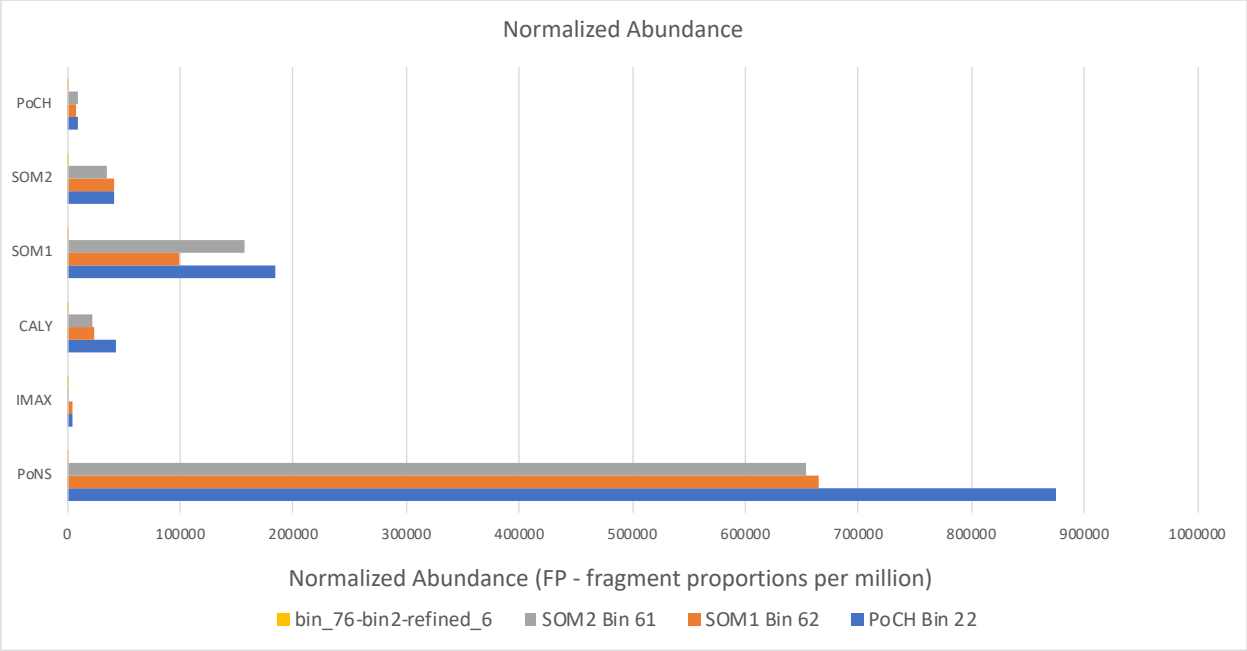

Figure 1. Graph depicting the normalized abundance for each bin and reported by chimney. Note that BinSanityLC-kmean-bin_76-bin_2-refined_6 values are all very small and are difficult to visualize.

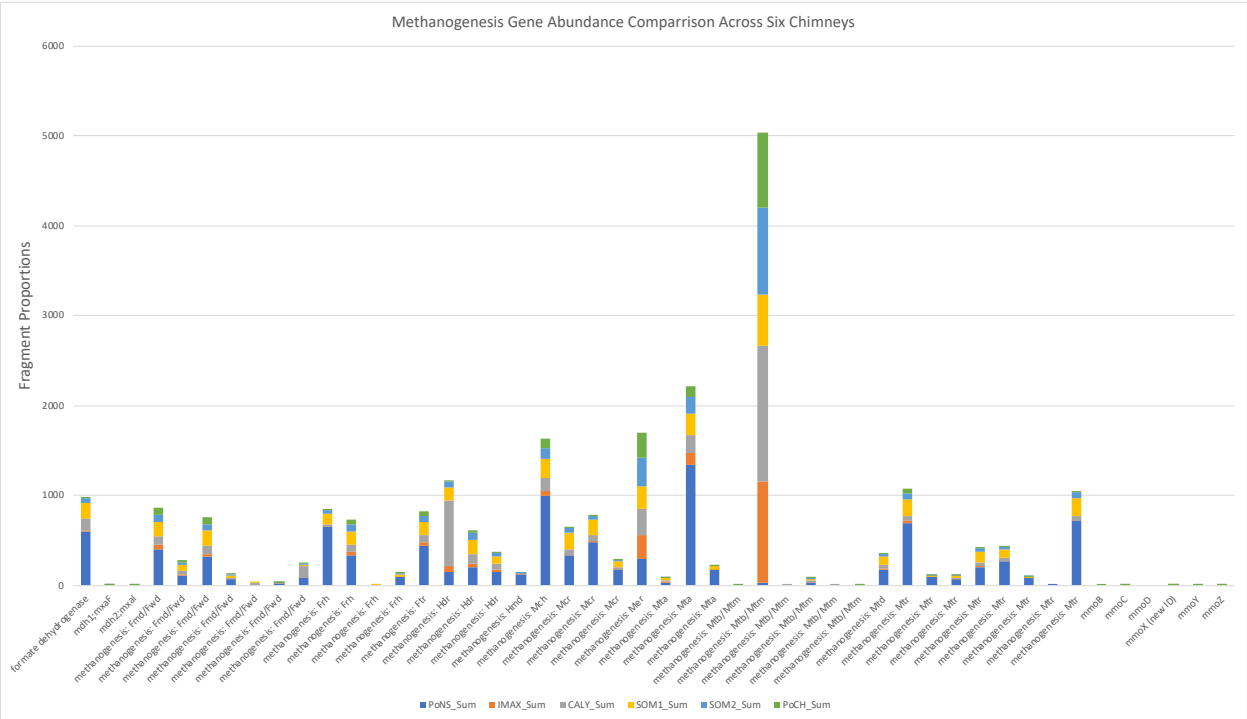

Figure 2. Graph depicting both the proportions of various genes and comparing the abundances across all chimneys.

Of the four MAGs that were being examined, there were three (SOM1 Bin 62, SOM2 Bin 61, and PoCH Bin 22) that were defined as *Methanosarcinales* through genomic classification. Similarly, the MAG discussed by McGonigle, et al.[1] had been given a "taxonomic assignment of *Methanosarcinales*". Figure 3 shows that the mcrA gene from PoCH Bin 22 was identical to the bin presented by McGonigle, et al.; these two bins only differed by a single amino acid in the mcrA genes of SOM1 Bin 62 and SOM2 Bin 61. The BinSanityLC-kmean-bin_76-bin_2-refined_6 bin was not included in the comparison because it lacked the mcrA gene. It appears that the bin assembled by McGonigle, et al. and SOM1 Bin 62, SOM2 Bin 61, and PoCH Bin 22 were all very similar and could possibly belong to the same organism. Thus these four MAGs were assembled into a single Lost City Methanosarcinales MAG for further observation.

The newly assembled Methanosarcinales MAG was used for phylogenetic analysis. From the assembled mcrA gene and the 16S sequence used by McGonigle, et al.,[1] two individual phylogenetic trees were created and can be visualized in Figure 4. Importantly, the Lost City Methanosarcinales MAG appears to represent a novel genus and species and belongs to the order *Methanosarcinales*. Thus this mag appears to indicate the presence of methanogenic archaeon characterized only at the Lost City.

Analysis of the genes found in Methanosarcinales MAG indicated the presence of the necessary components to utilize methane in its metabolism. A summary of the provided genes and corresponding KEGG IDs can be found in the Additional Information. Additionally, the presence of formate dehydrogenase (FdhA) and membrane bound hydrogenase (Mbh) were present in the Methanosarcinales MAG. A divergent form of formate dehydrogenase can also be found in this MAG. Its function has yet to be understood in great detail.

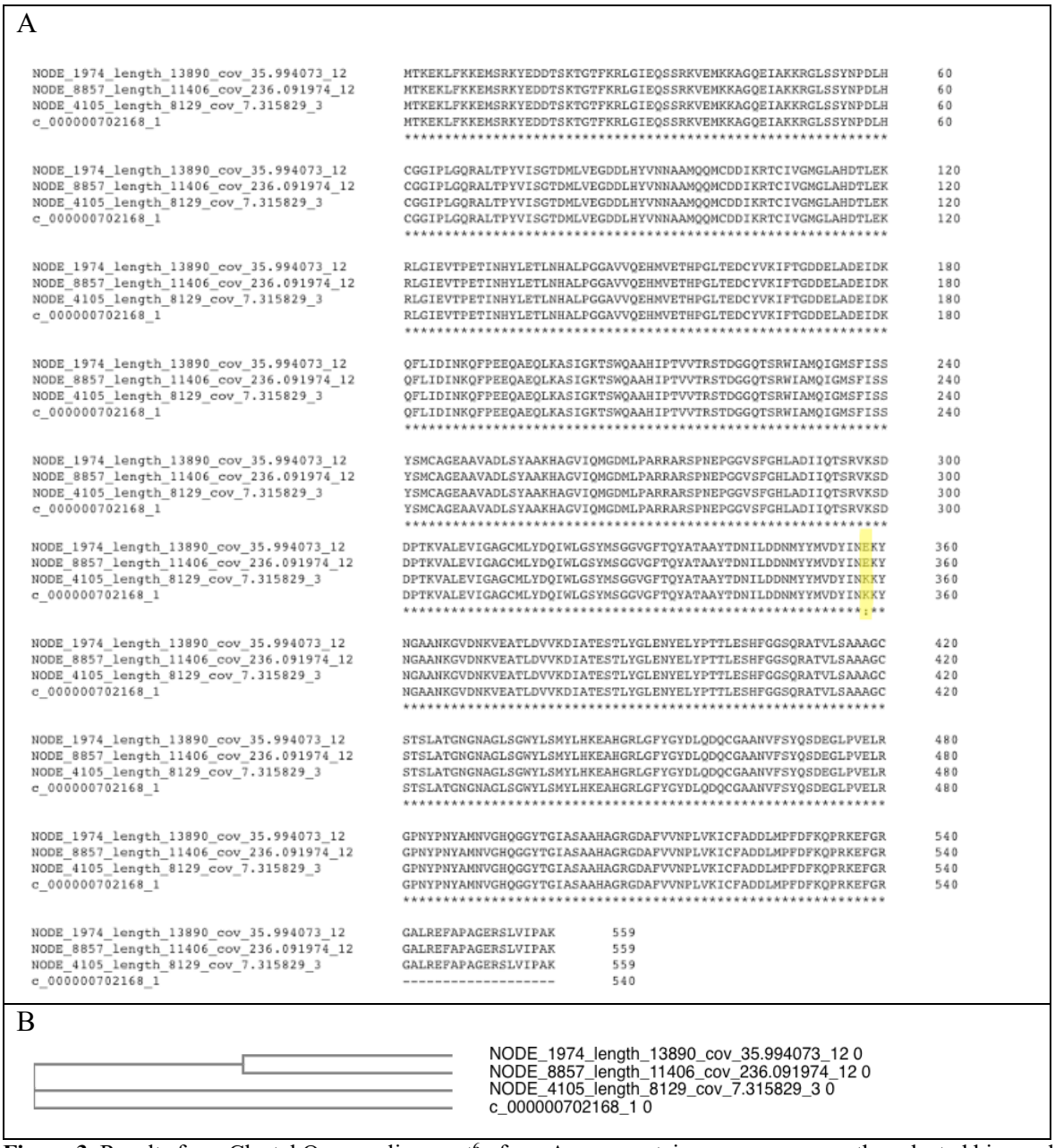

Figure 3. Results from Clustal Omega alignment[6] of mcrA gene protein sequence across the selected bins and the bin from McGonigle, et al..[1] SOM2 Bin 61 corresponds to NODE_1974_length_13890_cov_35.994073_12, SOM1 Bin 62 corresponds to NODE_8857_length_11406_cov_236.091974_12, PoCH Bin 22 corresponds to NODE_4105_length_8129_cov_7.315829_3, and the from McGonigle, et al. corresponds to c_000000702168_1. Image **A** shows the complete alignment of the amino acid sequences. Image **B** shows a hypothetical phylogenetic tree from the alignment.

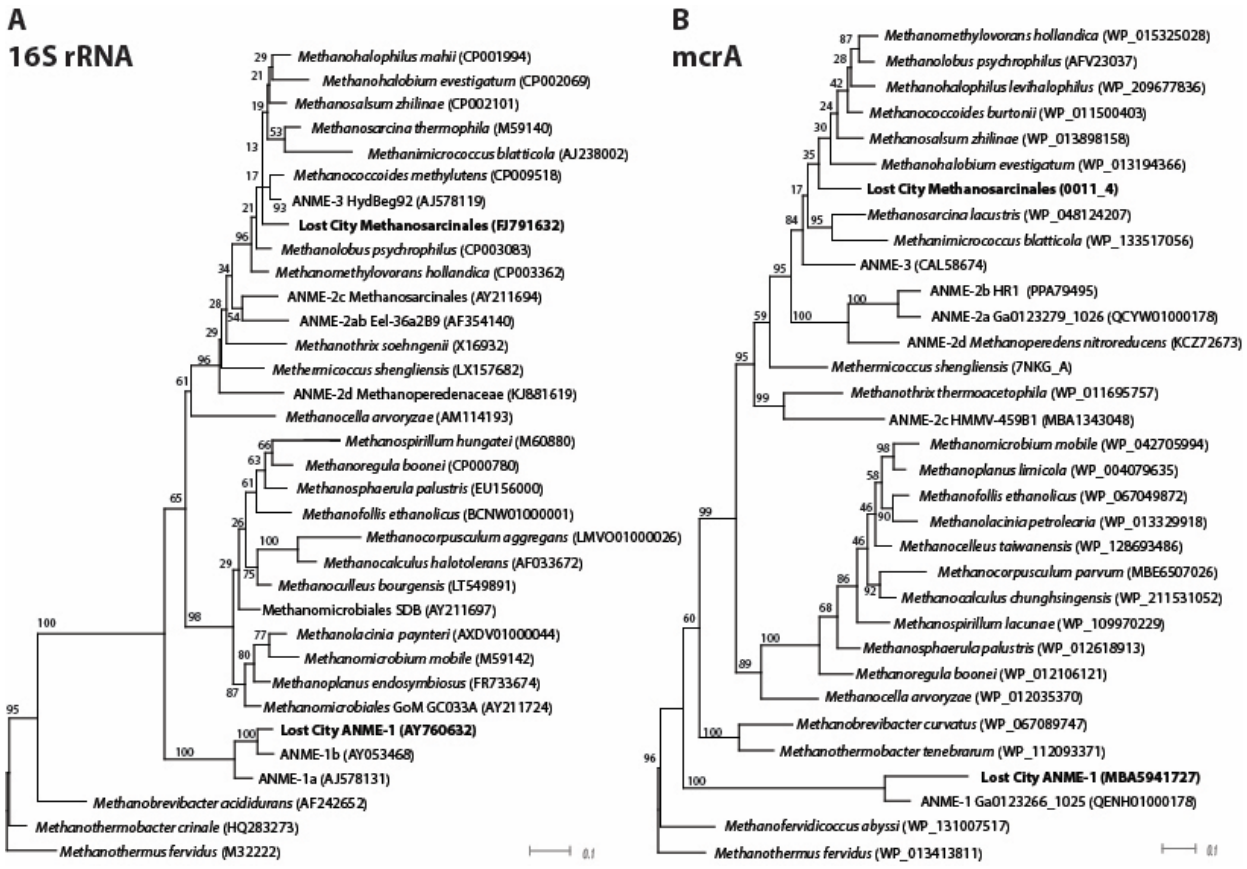

Figure 4. Phylogenetic trees of the Lost City Methanosarcinales MAG with other known methanogens developed by using **A** 16S rRNA and **B** the mcrA gene (necessary for methanogenesis). Accession IDs for the corresponding organisms can be found in the phylogenetic tree.

Discussion:

Thorough examination of the samples collected from the Lost City through metagenomics methods revealed the presence of four MAGs that are representative of methanogenic organisms. These MAGs appear to be more abundant at the Poseidon North Spire chimney and this is also true of the accompanying methanogenesis genes. It appears that three of the four bins that were selected (SOM1 Bin 62, SOM2 Bin 61, and PoCH Bin 22) were all very similar to a bin presented by McGonigle, et al..[1] For this reason, these bins were combined into a single Lost City Methanosarcinales MAG.

The combined Lost City Methanosarcinales MAG, through phylogenetic analysis, was shown to be a methanogenic archaeon that belonged to a novel genus and species. Formal naming of this organism is yet to be carried out, but it is known to belong to the order *Methanosarcinales*.

The presence of methanogenesis genes in this organism indicate that it is able to use methane, but it is not known if this organism is producing or consuming it as part of its metabolism. Additionally, this MAG possesses a set of membrane bound hydrogenases that could potentially allow it to use H_2. This MAG also has both a standard formate dehydrogenase and divergent

formate dehydrogenase that could provide this organism the capacity to incorporate formate in its metabolism. The divergent formate dehydrogenase may also confer some metabolic capacity unique only to organisms found at the Lost City.

The methanogenic archaeon that was discovered at the Lost City hydrothermal field appears to possess unique characteristics that allow it to survive this extreme environment. Further research into the dynamics of its metabolic strategies and interactions with other organisms should illuminate the dynamics of the Lost City. Understanding the complexity of Lost City methanogens and accompanying microorganisms may further shed light on how ancient ecosystems supported early life on Earth. This research contributes to a larger international effort to investigate how hydrothermal processes within the seafloor fuel the metabolic activity and the evolution of microscopic life.

Additional Information:

Any files and other information can be found at https://github.com/Brazelton-Lab/Briggs_Methanosarcinales .

Acknowledgements:

I would like to extend my gratitude to mentor Dr. William Brazelton for his guidance regarding my research and suggestions surrounding my writing.

References:

1 McGonigle, J. M., Lang, S. Q., & Brazelton, W. J. (2020). Genomic evidence for formate metabolism by Chloroflexi as the key to unlocking deep carbon in Lost City microbial ecosystems. *Applied and environmental microbiology*, *86*(8), e02583-19.

2 Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of molecular biology*, *428*(4), 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

3 Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., ... & Yu, D. (2018). KBase: the United States department of energy systems biology knowledgebase. *Nature biotechnology*, *36*(7), 566-569.

4 Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database.

5 Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature biotechnology*, *29*(1), 24-26.

6 Sievers, F., & Higgins, D. G. (2014). Clustal omega. *Current protocols in bioinformatics*, *48*(1), 3-13.

7 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312-1313.

8 Huson, D. H., & Scornavacca, C. (2012). Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic biology*, *61*(6), 1061-1067.

9 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., ... & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research*, *41*(D1), D590-D596.

10 Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410.