Microbial diversity analysis of the bacterial and archaeal population in present day stromatolites

Maya Ortega

Mentor: Dr. Jamie Foster

Kennedy Space Center

Abstract and Introduction

Stromatolites are layered sedimentary structures resulting from microbial mat communities that remove carbon dioxide from their environment and biomineralize it as calcium carbonate. Although prevalent in the fossil record, stromatolites are rare in the modern world and are only found in a few locations including Highborne Cay in the Bahamas. The stromatolites found at this shallow marine site are analogs to ancient microbial mat ecosystems abundant in the Precambrian period on ancient Earth. To understand how stromatolites form and develop, it is important to identify what microorganisms are present in these mats, and how these microbes contribute to geological structure. These results will provide insight into the molecular and geochemical processes of microbial communities that prevailed on ancient Earth. Since stromatolites are formed by lithifiying microbial mats that are able to mineralize calcium carbonate, understanding the biological mechanisms involved may lead to the development of carbon sequestration technologies that will be applicable in human spaceflight, as well as improve our understanding of global climate and its sustainability.

The objective of my project was to analyze the archaeal and bacterial diversity in stromatolites from Highborn Cay in the Bahamas. The first step in studying the molecular processes that the microorganisms carry out is to ascertain the microbial complexity within the mats, which includes identifying and estimating the numbers of different microbes that comprise these mats. My main project was to analyze the archaeal population by generating a sequence library of the 16S ribosomal RNA genes (16S rRNA), which is used to classify and compare

Bacteria and Archaea, using the high-throughput 454-pyrosequencing. When the sequence data is returned, statistical analysis of the resulting 16S rRNA gene sequence data will be performed to assess the abundance of Archaea and the diversity of the population in each of the mat types. Following completion of preparing the archaeal library for sequencing, I shifted my focus to computer analysis of raw 16S rRNA bacterial sequence data from stromatolites that was previously obtained by the Foster lab. From the classification analysis, we found that sequencing picked up many previously undetected bacterial phyla in the stromatolites. This experimental approach builds upon previous studies by applying a novel deep-sequencing method to improve sequencing coverage of the mat populations and account for previously undetected microbes in the stromatolitic mats.

Goals and Purpose of the Project

This is my second summer internship at Kennedy Space Center, where I worked in the lab of Dr. Jamie Foster, an Assistant Professor at the University of Florida. Dr. Foster's research at the Space Life Science Lab focuses on communities of microbes and the mechanisms with which they interact with their environment. Lithifying microbial mats are used a model systems for understanding the molecular and metabolic processes that these mats undergo. As previously mentioned, these microbial mats serve as models for understanding ancient Earth ecosystems and bioregenerative life support.

This summer I returned to the Foster lab to research another type of microbial mat, the modern marine stromatolite. Unlike the thrombolites, which are mats in which organisms and sediment are arranged in clots, I worked with last summer, stromatolites are characterized by lamination and have a distinct, layered structure. The stromatolites from our study site at Highborne Cay in the Bahamas cycle through three types of microbial communities, types 1, 2, and 3, and each display a certain degree of microbial diversity and lithification. My project consisted of two objectives: to generate an archaeal 16S rRNA gene amplicon library from the stromatolites, and to analyze the bacterial community composition in an existing set of 16S rRNA gene sequence data.

The first portion focused on analyzing the archaeal population within each of the three mat types. The standard method for studying microbial community diversity is through sequencing and comparative analysis of regions of the 16S rRNA gene, which is used as a phylogenetic marker due to the high level of sequence conservation between different species. The 16S rRNA gene rate of mutation is constant over time within a species, thus small sequence variations can be attributed to evolution and used to distinguish between different microbial species. Statistical methods are applied to large sets of 16S rRNA gene sequences to identify the degree of variation between sequences, thereby establishing how many different organisms are represented in the microbial community and their phylogenetic relatedness, which reveals how similar or dissimilar organisms are from one another. For my project, 16S rRNA gene sequence

libraries generated from DNA extracted from the three types of stromatolitic mats from Highborne Cay Bahamas.

I dissected small samples of mat from type 1, 2, and 3 stromatolitic mats and extracted the DNA through a series of physical and chemical treatments that release DNA from the cells in the mat, followed by purification steps to isolate high quality community DNA. This purified DNA was used in a polymerase chain reaction (PCR) to amplify a specific portion of the target 16S rRNA gene. PCR works as a copy machine that is used to selectively amplify and produce millions of copies of a gene of interest (amplicon). One key element of PCR is the primer set, which are small pieces of DNA that bind to complementary regions of the DNA template strand. Primers act as markers to indicate where on the DNA template strand the polymerase (the machinery that builds new strands by attaching nucleotides) should begin and end. Choosing primers is a selective process because they must properly anneal at the correct position to the target DNA. For my project, I researched and selected two different primer sets that fit the requirements of the PCR conditions, and compared them to a database of known bacterial and archaeal sequences to ensure that the selected primer sequence would selectively anneal only to archaeal DNA.

Using DNA extracted from the three stromatolitic mat types, I performed a PCR for each primer set and confirmed proper amplification using a method called gel electrophoresis, which allows for visualization of the amplified DNA. Three replicates of PCR were run for each mat type to minimize bias from PCR, amounting to nine reactions per mat type. The resulting PCR

products were purified, pooled together into a single tube, and sent to the University of Florida in Gainesville for 454 pyrosequencing. A pyrosequencing run takes around three weeks to complete - such that my internship ended before I received my data. Therefore, for the remainder of my internship, I worked to analyze another data set of stromatolite sequences, this time focusing on bacterial population and diversity. This data set was generated from Illumina sequencing, another high throughput DNA sequencing method. The analysis of Illumina-sequenced 16S rRNA data is a relatively new field, thus my mentor and I had to come up with a new way of analyzing these data. We first trimmed the sequences by removing the primers and selectively removing sequences that were shorter or longer than our expected amplicon size. After removing repeated sequences to increase analysis efficiency, we aligned the unique sequences to a database of sequences belonging to known bacteria, and created a taxonomic report that classified each of the sequences by phyla, all the way to species when the sequence confidence interval would allow.

The classification analysis of the bacterial 16S rRNA gene libraries from the stromatolitic mats confirmed that the deep-sequencing approach generated greater coverage than that in previous studies in stromatolitic mat communities that used an older sequencing technology. Due to our ability to detect organisms found in very small amounts in the mats, at least seven newly accounted for phyla are represented in our dataset. Additionally, the high abundance of unclassified bacterial sequences suggests that all three stromatolitic mats contain novel organisms that are not similar to those in founds in the curated 16S rRNA gene database (Figure

1). These novel sequences will require further study and classification to identify what organisms they represent. This may be achieved through target study of specific sequences that resemble known organisms. The 16S rRNA gene data also suggests that mat types 1, 2, and 3 contain different bacterial communities. Type 1 mats contained the least amount of phyla richness, or number of different organisms represented in sequences, compared to types 2 and 3. Each mat also displayed a fairly different abundance of (or number of sequences classified as a) specific phyla, indicating that there is a developmental progression from type 1 to 3. This bacterial data, combined with the archaeal library findings will provide the most comprehensive sequencing of stromatolitic communities to date. By working to identify what species are composing these communities, we will lay the groundwork for further study of stromatolites including understand the role of microbes in such critical processes such as carbonate lithification.

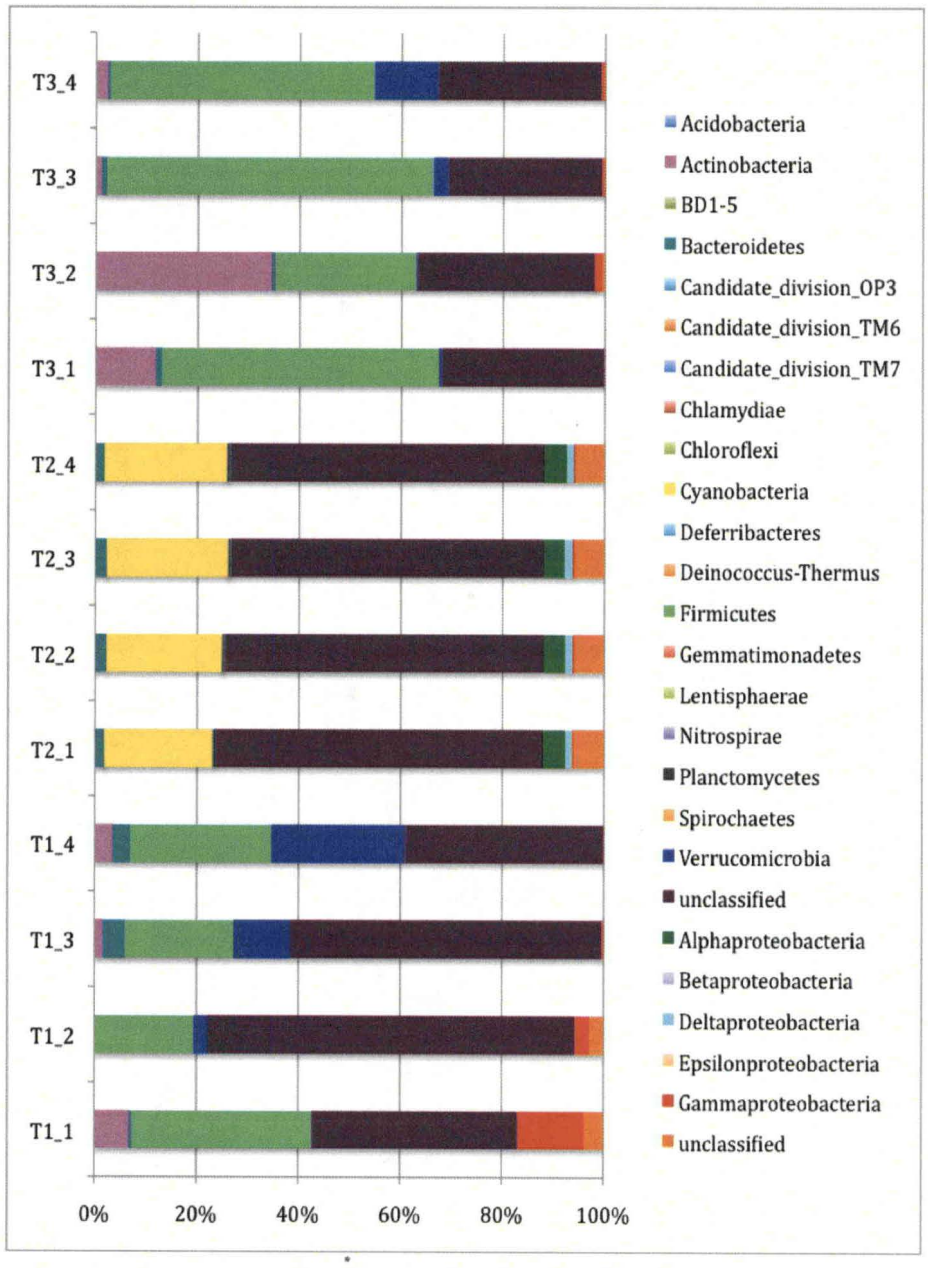

Figure 1 - Histogram representing the relative abundance of each phyla found in the stromatolites as a percentage of the total number of sequences classified into specific phyla. Unclassified sequences (dark purple) could not be assigned a phyla based on database sequences.

Impact of the MUST Internship on My Career Goals

This summer I had the opportunity to work on a mini-research project, which gave me a lot of insight into the environmental microbiology research, and what a career in academic research might be like. My mentor gave me a lot of space to work independently on my project, and as a result I gained a lot of experience with all aspects of working in a lab, from designing an experiment to working out procedural problems to using computer software to analyze data. I learned how to design primers, prepare DNA for sequencing, analyze a large set of 16S rRNA sequences, and troubleshoot PCR. I learned a lot about reading scientific literature, and how to glean important information from papers more efficiently. This is a very important skill to have both for gaining knowledge as well as eventually writing papers of my own. I also had an introduction to bioinformatics, and I learned about the general process used to analyze high-throughput sequence data. Computers are being incorporated more and more, not only in environmental microbiology, but in many life science fields as well, and having some basic knowledge of using computers to apply their computational capacity to a large data set is becoming increasing valuable in numerous fields of molecular and microbiology. Working with a type of data that has never been generated before, the Illumina sequences, allowed me to examine the process used for other data sets and work to optimize it for my specific data.

During the summer I had the chance to meet and work with people from many different stages in the field of research. In my lab I worked with a post-doctoral research assistant, a doctorate candidate, and a university professor. I also met graduate students and other principle

investigators, and each person had their own experiences, knowledge, and tips to tell me about. I appreciated the chance to hear about things such as applying to graduate school, picking research topics, and the process of obtaining funding from different sources and grants. I also benefited from the knowledge each one had in the field, and everyone was very willing to explain procedures and rationales to me so I could learn more about the research going on in my field. Overall my internship experience gave me a lot of valuable exposure to the world of research that is not often afforded to an undergraduate student. I was fortunate to be able to work directly on my own project and take part in a novel data analysis, and my experience is something that I can apply to my future career in science.