Metagenomic Analysis of Dr. Bruns' Cyanobacteria Samples

Eric Marti

Introduction

Cyanobacteria share many of their characteristics with algae, which is why a common name for cyanobacteria is "blue-green algae." However, these organisms are not algae; cyanobacteria are a phylum of prokaryotic photosynthetic bacteria. Unlike other photosynthetic bacteria, cyanobacteria have the same type of pigment that is in the chloroplasts of algae and plants: chlorophyll a. Combined with phycocyanin, another photosynthetic pigment, colonies of cyanobacteria exhibit a blue-green appearance that gives them their misleading name of blue-green algae [1].

Typically colonies of cyanobacteria do not exhibit much differentiation. This is because these organisms reproduce via fission, which limits the genetic variation. However, it is possible for cyanobacteria to exchange DNA. Adjacent cyanobacteria can exchange DNA through the conjugation pilus that connects two cells. In addition, a cell can "transform" by absorbing free DNA. A virus can also replace a cyanobacteria's DNA with its own. Lastly, genetic mutations are common in cyanobacteria. These methods help cyanobacteria overcome their inability to reproduce sexually.

Cyanobacteria can survive in a wide range of environments. Although commonly found in basic aquatic habitats, cyanobacteria are also found in soil. These organisms can often be the primary colonizers of a new environment due to their ability to survive in harsh conditions. This means that they are responsible for fertilizing the habitat with organic matter for other organisms to thrive. This is why Dr. Mary Ann Bruns of Pennsylvania State University is interested in using cyanobacteria to revitalize poor soil and reduce nutrient run-off to the Chesapeake Bay. She says,

“If [cyanobacteria] can be cultivated and more widely introduced, they could substitute for some of the fertilizer now being applied by farmers” [2]. This research can extend globally if Dr. Bruns can use cyanobacteria to regulate the richness of soil.

Dr. Bruns’ Pennsylvania Soil Samples

Each of the three metagenomic samples came from a replicate enrichment culture grown in a bioreactor. This means that there should not be any significant variation in any of the samples. The original sample came from "photosynthetic crust" from the surface of soil in a greenhouse. Dr. Bruns believes that there are at least two closely related species of cyanobacteria belonging to the Nostocaceae family. There may also be strains of alpha and gamma proteobacteria. The objective is to identify the species of cyanobacteria in the samples.

Pipeline

To determine the taxonomy of the metagenome, the DNA samples must first be assembled de-novo using IDBA-UD. Then, the assembled contigs will be plugged into MaxBin2 to be binned into individual microbial species. MaxBin is a package that automates the binning process by running through a script that uses FragGeneScan, HMMer3, Bowtie2, and Velvet. The package utilizes the Expectation-Maximization algorithm, which uses nucleotide composition as well as the contig abundance to bin metagenomic contigs. By default, MaxBin will look for the 107 marker genes found in over 95% of bacteria, but it also has an option to search for the 40 marker gene sets common to bacteria and archaea. Next, each bin must be run through BLAST to determine its taxonomic classification against the latest NCBI Bacteria database. Finally, MEGAN will be used to visualize the taxonomy of each bin.

Results

MaxBin identified six microbial species. Using the plotmarker flag of MaxBin, it was possible to visualize the abundance of marker genes in each bin of the data.

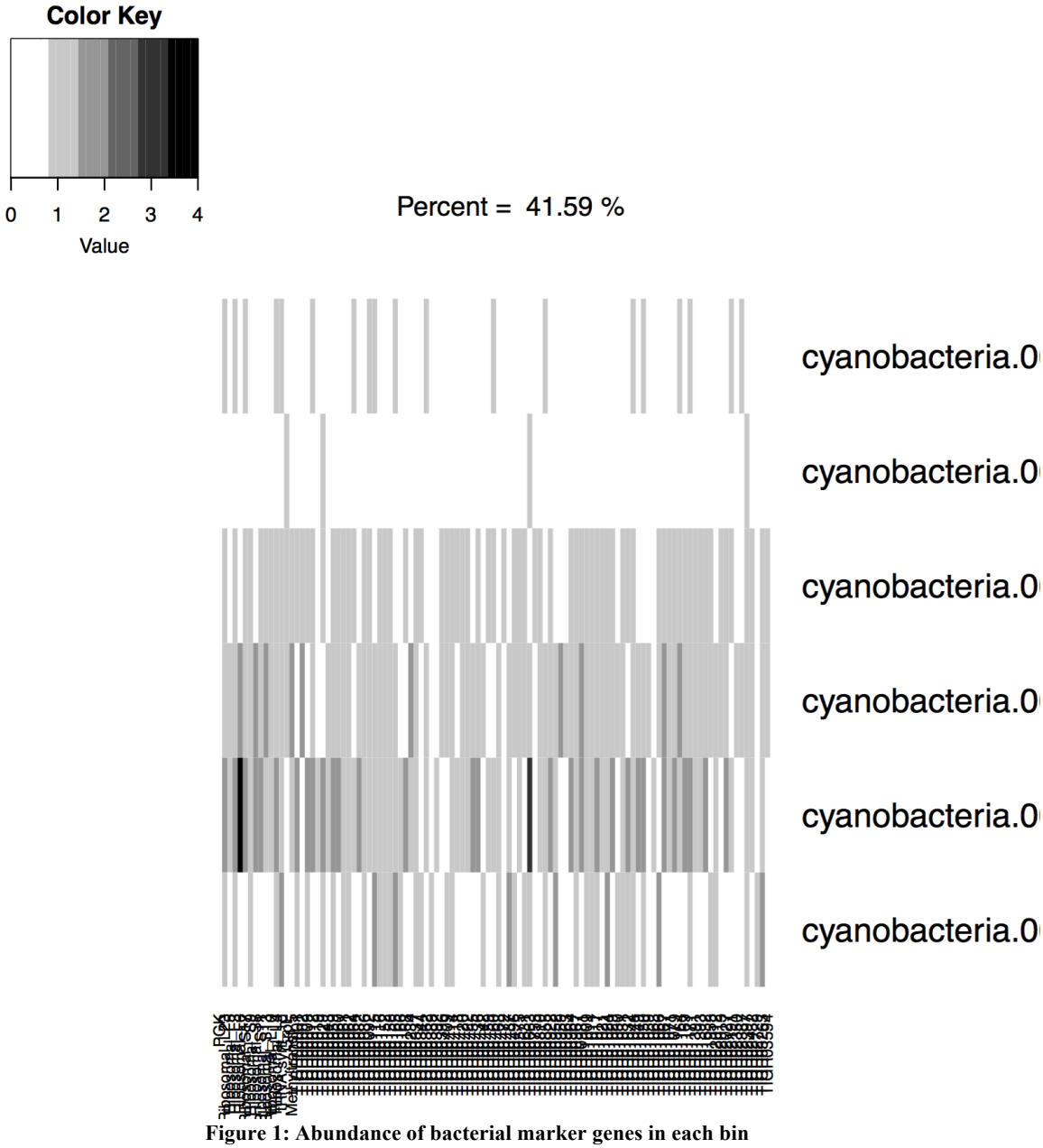

Figure 1: Abundance of bacterial marker genes in each bin

Here it is clear that the species of bacteria in the first two bins is much different than the bacteria in the rest of the bins. Below is a table with information about the abundance of each bin inside each metagenomic sample.

Bin name	1_S1_L001.fasta	2_S2_L001.fasta	3_S3_L001.fasta
cyano.001.fasta	55.76%	55.32%	55.55%
cyano.002.fasta	38.60%	38.29%	38.42%
cyano.003.fasta	3.78%	4.72%	4.55%
cyano.004.fasta	0.73%	0.82%	0.77%
cyano.005.fasta	0.70%	0.60%	0.52%
cyano.006.fasta	0.43%	0.25%	0.19%

Figure 2: Abundance of each bin inside each sample

There are no significant differences in abundance between samples. Each sample came from a replicate enrichment culture, so this is expected. Bins 1 and 2 seem to be the majority of each of the samples, so the species of cyanobacteria must exist in these two bins. Since cyanobacteria are closely related to algae, it makes sense that there would be less bacterial marker gene hits than other bacteria.

The results from MEGAN clearly confirm the suspicion that Bins 1 and 2 are species of cyanobacteria. However, it is not clear what the two species are. The next page features a visualization of the two bins corresponding to cyanobacteria.

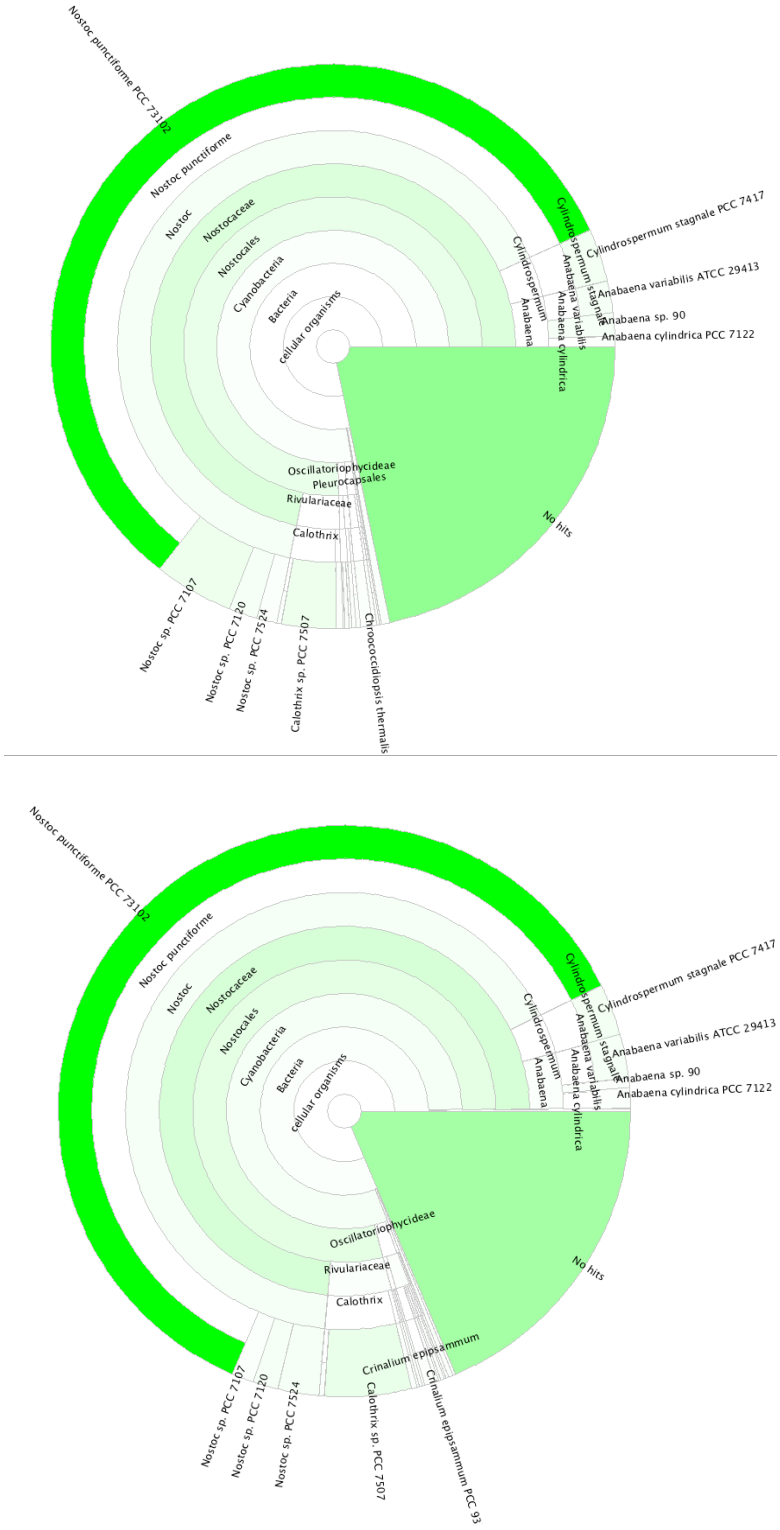

Figure 3: Bins 1 & 2

In both bins, a majority of the reads were assigned to one species of cyanobacteria. In the first bin, 60.3% of the reads were assigned to Nostoc punctiforme PCC 73102, and in the second bin, 59.3% of the reads were assigned to the same species. This is odd because there is a clear difference in the abundance of the different bacterial marker genes across both bins. However, 27.8% and 22.9% of the reads from Bin 1 and Bin 2 respectively did not have any hits with the NCBI Bacteria database. Below is a figure showing the distribution of species read hits across the entire Cyanobacteria phylum.

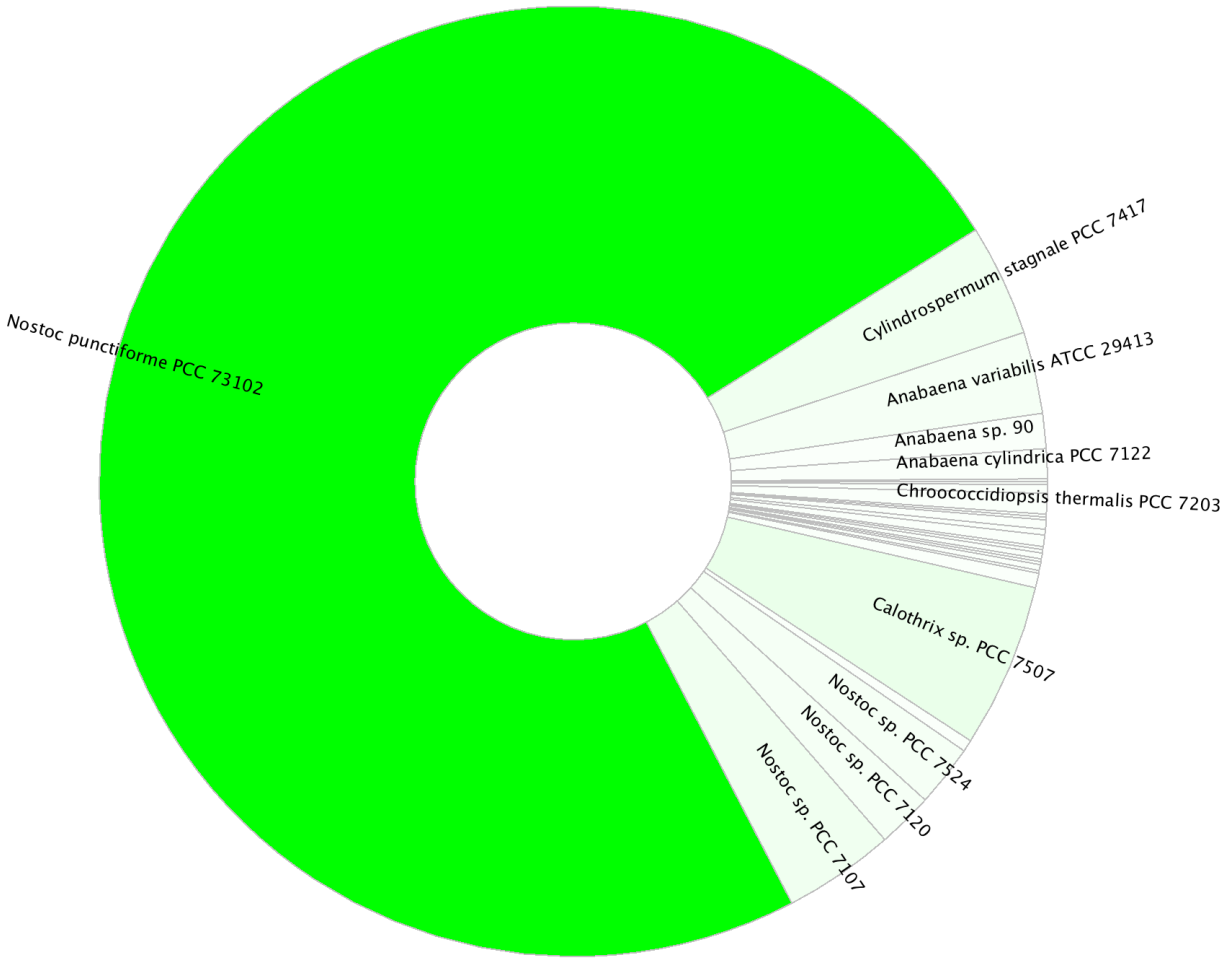

Figure 4: Reads assigned in Cyanobacteria Phylum

The other bins were associated with species of bacteria in the Proteobacteria phylum. Most notably, there was a high hit rate for Sphingopyxis alaskensis RB2256, which is an alphaproteobacteria that showed up in Bin 6, which represents less than 1% of each sample. Below is the distribution between all of the proteobacteria found in the samples.

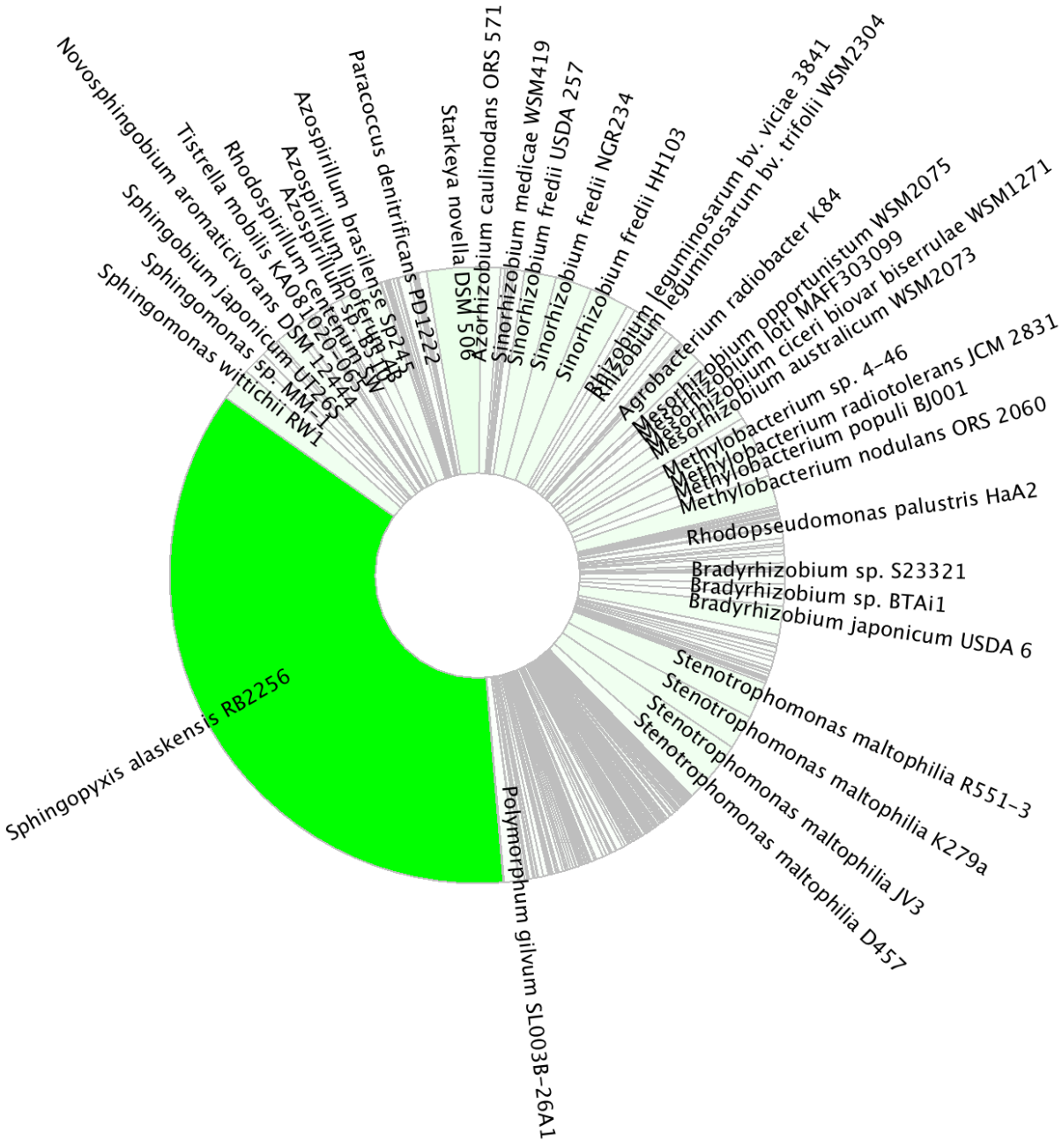

Figure 5: Reads assigned in Proteobacteria Phylum

Below are the distributions for each of the remaining bins.

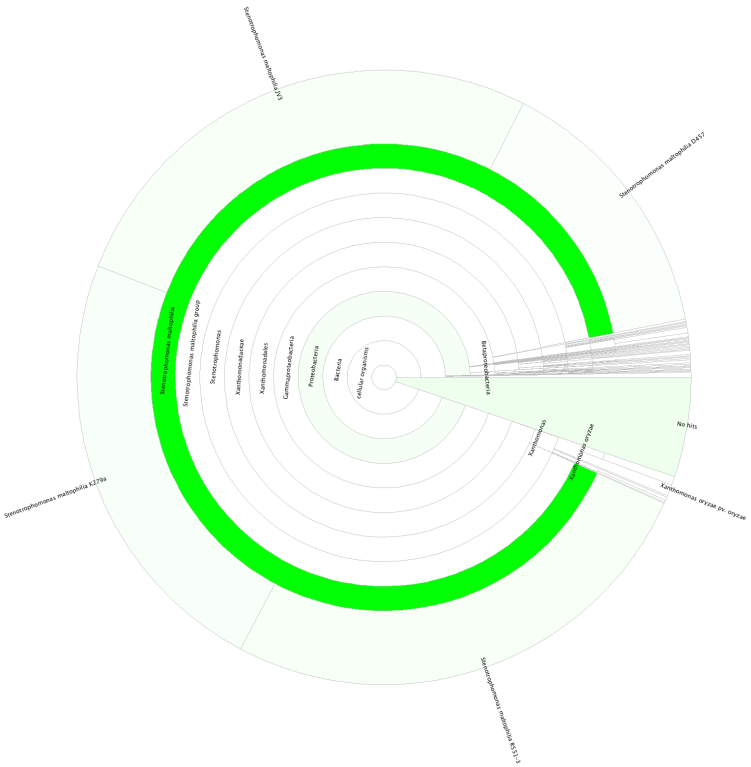

Figure 6: Bin 3

Figure 7: Bin 4

Figure 8: Bin 5

Figure 9: Bin 6

Below is the taxonomic tree associated with the cyanobacteria hits.

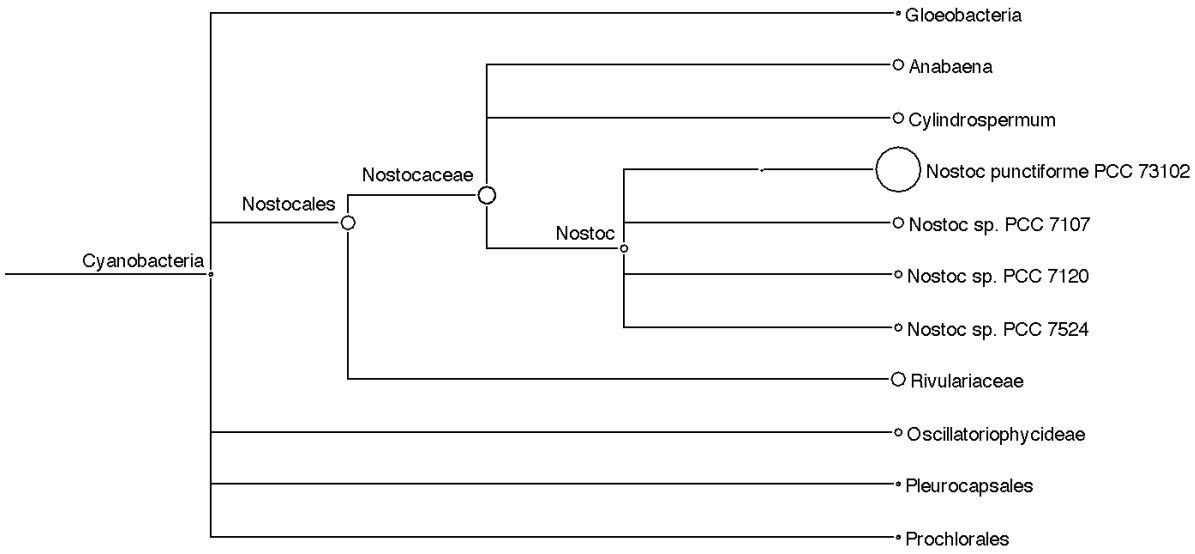

Figure 10: Cyanobacteria branch with read hits

Conclusion

The only species of cyanobacteria successfully identified is Nostoc punctiforme PCC 73102. However, there seem to be some differences between the bins that both exhibited read hits with that species. One of these differences was in marker gene abundance. The two bins had different patterns, as seen in Figure 1. In addition, there was a high percentage of unassigned reads in both bins. A future work may be to use a different binning package such as GroopM. MaxBin marked a lot of reads as "no classification." The marker gene method may not sufficient enough to identify cyanobacteria.

References

[1] B. Gow. *Cyanobacteria*. Available: http://academics.smcvt.edu/dfacey/AquaticBiology/Freshwater%20Pages/Cyanobacteria.html

[2] J. Mulhollem. (2015). *Cyanobacteria could help manage nitrogen to benefit Chesapeake Bay*. Available: http://news.psu.edu/story/347590/2015/03/06/research/cyanobacteria-could-help-manage-nitrogen-benefit-chesapeake-bay