Lab 7: Post Lab

OneCard: 1955791 October 31, 2017

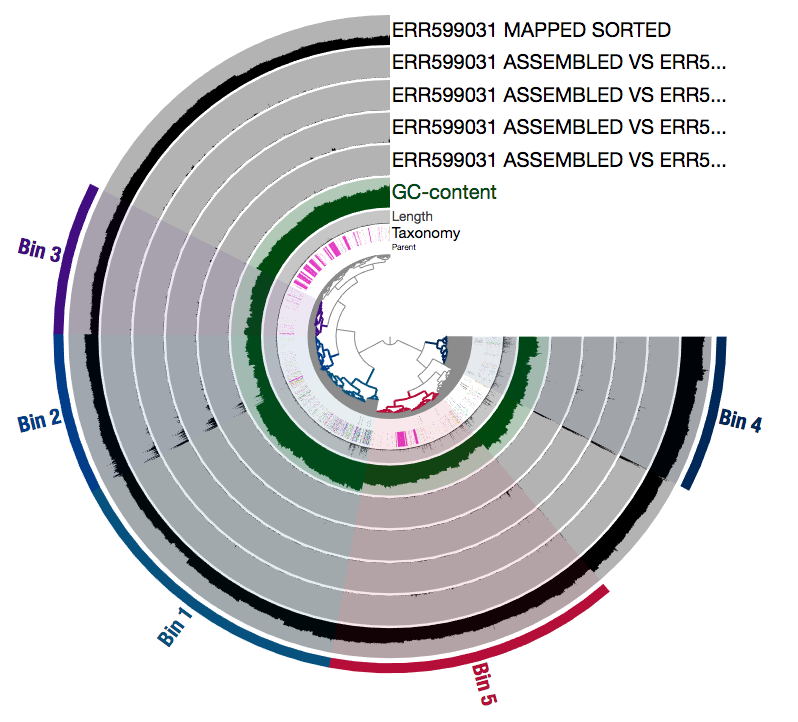


Figure . Anvi’o-interactive view of my own metagenomics sample (ERR599031) from mesopelagic zone of Arabian Sea, mapped and binned with other samples from mesopelagic zone (ERR599031, ERR599078, ERR599090, ERR599142).

# 1. Bin completeness

**1a. Which bin was the most complete?**

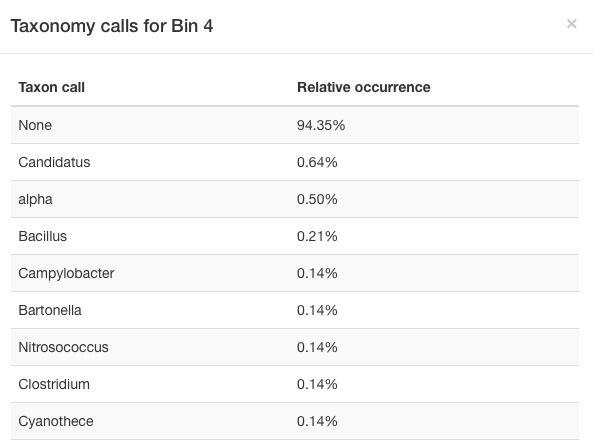
My Bin\_4 was most complete, at 98.56115108 percent.

**1b. What was the N50 of that bin?**

11734

**1c. What was the anvi'o-labeled taxonomy of this bin?**

None, with 94.35% of calls… Did I make a mistake?



**1d. Now find your bin folder…**

I did a blastp of two of the longest contigs. Both returned “no significant similarity.”

# 2. Coverage across all bins for one sample

**2a. Which bin had the highest coverage across all samples, when looking only at the mapping of your own dataset (self-to-self)?**

Bin 4 again! With a coverage of 25.54. (The other bins had coverage between 3-10).

**2b. What is the taxonomy of that bin, according to anvi'o and according to BLAST (see part 1 above)?**

Again, None…

**2c. Make a bar graph in which each bin is a different bar, and the bar height indicates the mean coverage. Call this 'Figure 2' and include a figure caption.**

Figure . Mean coverage for each bin, for mapping of my own dataset

**2d. What does it mean, biologically/ecologically speaking, for a bin to have high coverage?**

High coverage within a bin could indicate an abundance of organisms contained within that bin. In my particular sample, there were more organisms containing contigs that I binned into bin 4 than organisms containing contigs that got binned into the other four groups.

# 3. Coverage of one bin across samples

**3a. Make a bar graph in which each sample is a different bar, and the bar height indicates the coverage of your bin in that sample.**

Figure . Mean coverage for each sample in Bin 1.

Table . Mean coverage for each sample in bin1.

|  |  |
| --- | --- |
| Sample | Mean Coverage |
| ERR59903 vs ERR598992 | 0.000892548 |
| ERR599031 vs ERR599078 | 0.003170638 |
| ERR599031 vs ERR599090 | 0.000480909 |
| ERR599031 vs ERR599142 | 0.001571686 |
| ERR599031 vs Itself | 6.29560438 |

**3b. In which sample does your bin have the highest coverage? Second highest? What might this imply about the abundance of this microbe in different ecosystems?**

It seems that there is very low coverage for every sample except my own. Mine has 6.29, while the others all have less than 0.01. Something wrong?

# 4. Variability of all bins for one sample

Use the SAMPLES-MERGED/SUMMARY\_my\_bins/bins\_across\_samples/variability.txt file to answer these questions.

**4a. Which bins had the highest and lowest variability (single nucleotide variants per kilobase pair (SNVs/kbp)) across all samples, when looking only at the mapping of your own dataset (self-to-self)?**

Bin 4 has the highest with 25.537 SNVs/kbp.

Bin 3 had the lowest with 0.118468786.

**4b. What is the taxonomy of those bins, according to anvi'o and BLAST (see above)?**

Both were called as predominantly None.

**4c. What do you think the variability might imply about the microbial populations represented by those bins? What might it imply about the selection pressure on those populations? (We'll talk more about this in class on Wednesday, so you might want to wait until after then...)**

High variability could mean that a microbial population is similar enough to be grouped into a bin, but there is lots of variation within that bin (different alleles that are all allowed to thrive under these selective conditions, perhaps?)