Last Week

- Completed rarefaction curves for sample of 10 bacteria
- Tried to run gene family codes on large salmonella files.
- Adjusted Athena's code as necessary as well as wrote my own in attempts to make it work

This Week

No results this week, but I believe I got everything to finally work. Instead of trying to run everything on a gigantic file with 98 species, I instead randomly chose five salmonella genomes and tested everything against that.

As Athena's code was highly specific, I had to do a lot of adjustments and such, but hopefully by next week, I can start running it on the actual bacteria files. I have yet to generate the final prsence-absence matrix but that remains the only step and everything else should hopefully work out. As I spent last week trying to put the files through genefamily11.R, all the blastp files should be functional so I won't have to rerun that.

At this point, I felt like I had a lot of steps I wrote everything down into a text file that I have attached as a sort of "draft-documentation". Everything I worked on up to now is located in that file, as well as the differences between the methods I used and Athena originally used.

Next Week

- Run the sequence of code on actual bacterial genomes and generate gene families
- Work on generating the absence-presence matrix