Last Week

- Generating and completing the correlation between distance and pangenome for ten separate bacterial species

- Fitted graphs to a linear regression model and concluded that there is vague correlation between pangenome size and distance (i.e. the smaller pangenomes do appear to have less distance from the tips to the roots)

- Compared plasmid data vs non-plasmid data. Made note that differences are not too significant.

- Finalized pipeline for generating distance and pangenome plots

This Week

This week was the beginning of looking into the lateral gene transfer at long last, beginning with *Salmonella*. I took the liberty of going through one tutorial so far on perl, so I can adjust the codes as necessarif I have to. I considered using the Genbank files generated from prokka (that I already had) to put into the getfeatures profile but they didn't seem to correlate to the same proteins as the files downloaded directly from Genbank (identified through their ascension number)

Running the blastp ended up taking a lot longer than expected... I ran Athena's code exactly as she wrote it but it ended up giving me the "segmentation error"

This is the code I originally ran:

```
blastp −db parsed10mycogenomeswotransposase_seqs −query
10mycogenomeswotransposase_seqs.fna −out
allvsall_wotransposase_softmask_sw −evalue 5e−2 −num_threads 10
−outfmt '7 qseqid qlen qstart qend length sseqid slen qcovs score
bitscore evalue' −soft_masking true −use_sw_tback &
```

However I still get results, but because I wanted to make sure that it was accurate, I ran the blastp code again without the -num_thread arguements. It is still running...

Next Week

- Run rarefaction stuff on the background for each bacteria to show Roary's consistency.

- Finish off the gene family pipeline for *Salmonella*