Last Week
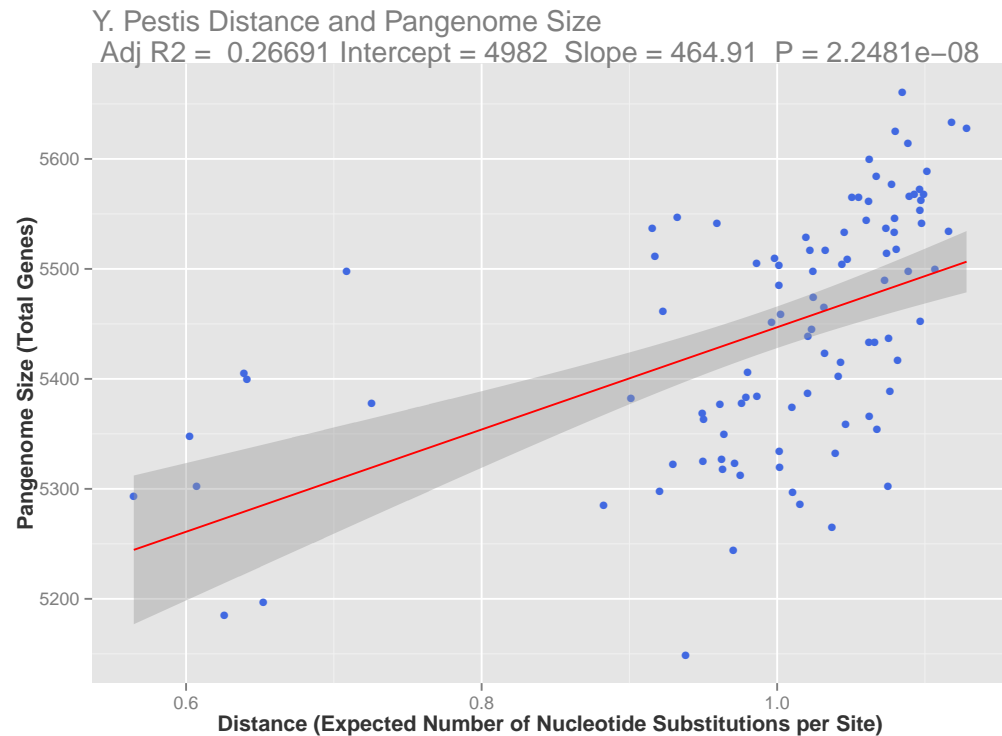
- Downloading files from Assembly

- Discovering and accounting for plasmid issue

This Week

Nothing new or interesting, mostly just doing repetitive things, but I manage to accumulate data for 8 bacterial species at last (with two more being put through Roary, so i should have them to work with by Monday, totalling up to 10), and though it's still a small sample size, it's still something to work with. I've sent the pangenome and distance correlations fitted to a linear regression on the separate file.

They mostly showed similar results, and even then right not I'm not sure if the linear model is the best model to fit them to (I simply did it because it seemed to be the most intuitive). I also played around with the plasmids a bit.

The one data set that ended up with the strangest appearing data was definitely *Y. pestis*, with strange anomalies towards the lower, bottom-left end of the graph:

Y. Pestis Distance and Pangenome Size
Adj R2 = 0.26691 Intercept = 4982 Slope = 464.91 P = 2.2481e−08

I checked the GFF files, and this was one of the ones where I ran Roary *with* the plasmids, so this may have been responsible for that.

Next Week

- **actual statistics?** but what to do...

- Consistency check: make sure they *all* don't have plasmids now, as I'd left some in in order to experiment about but now that all have to contain the same kind of data

- Begin looking at lateral gene transfer for methods for the present 10 bacteria data sample data

- Run rarefaction stuff on the background (takes a long time!)