Subs Paper Things to Do:

- why are the lin reg of dN, dS and ω NS but the subs graphs are...explain!
- mol clock for my analysis?
- GC content? COG? where do these fit?

Inversions and Gene Expression Letter Things to Do:

- ~~create latex template for paper~~
- confirm inversions with dot plot
- make dot plot of just gene presence and absence matrix (instead of each site) to see if this will go better
- look up inversions and small RNA's paper Marie was talking about at Committee meeting
- write outline for letter
- write Abstract
- ~~write intro~~
- write methods
- compile tables (supplementary)
- write results
- write discussion
- write conclusion
- do same ancestral/phylogenetic analysis that I did in the subs paper

General Things to Do:

- summarize references 40 and 56 from Committee meeting report (Brian was asking)

Last Week

Inversions + Gene Expression:

✓figured out mapping issue with BW and K-12

✓classify inversions based on whole block and individual sequences within a block

✓simple t-test and Wilcox test to compare inversions (all and within each block)

✓looked into overlapping regions in depth

✓started to look at DESeq

✓Queenie: brought to my attention the issue with BW and K12 mapping

Inversions + Gene Expression:

Queenie let me know that there was hardly any genes (300) that mapped between BW and K12 so I looked into this and realized there was an issue with my code and genes that overlapped. I fixed this and now there are 4000 genes that map between those two strains. Hooray!

I also included information about if an entire block had an inversion and if each individual sequence in that block was apart of that inversion or not.

I have been playing with the "toy" dataframe that Queenie made for me. I performed a simple t-test and Wilcox test comparing all inverted blocks to all non-inverted blocks. As well as comparing inverted and non-inverted sequences within each block. The results for this are irrelevant because this data set only has two of the 4 taxa. I mostly just wanted to get the code working so that when we do have the real dataset, the analysis will be speedy. I started looking into DESeq and how I could use inversions as a "treatment" to compare gene expression. But DESeq wants raw counts and the dataframe Queenie is working on is normalized CPM. So I will need to think about this more and figure out how to do this.

I also looked into some of the overlapping blocks (from PARSNP) and it appears as though they are relatively small (containing one gene or less), and are mostly intergenic. So again, I think it will not be a huge issue if we just remove data from these overlapped regions.

This Week

- Discuss next committee meeting and exit seminar
- t-test/Wilcox test comparing all individually inverted seqs with non-inverted ones
- think about (and execute) how to run DESeq to compare inverted and non-inverted regions
- get Queenie to remove overlapping regions from the dataframe
- get Queenie started on creating lists of which genes match (from all strains) in the parsnp alignment

Next Week

- edit another section of dissertation intro

- get Queenie to create a plot of the inversions
- think about (and execute) how to incorporate distance from the origin into the inversion analysis