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:

✓Queenie: complete dataframe for inversions and gene exp!

✓Queenie: working on getting summary statistics for inversions (how many, how big, in what taxa, where are they located, expression averages..etc)

✓started to create mapping file to show which genes match in all the strains in each block (based on subst paper "good" alignments)

✓started looking into genes that were missing from BW and K12MG mapping file

Subst Paper:

✓added more recent references (chromids and plasmids)

✓add program parameters to main text

✓better explanation for removing $dN = dS = 0$ and $dN = 0$

✓better explanation for normalization for number of sites

✓commented on multiple rep termini

✓addressed origin scaling figure

✓clearer figure captions

✓started running larger window size supplemental tests

✓progressiveMauve for *S. meliloti* chrom with 23 seqs (trying the analysis on this first)

✓research how many more complete genomes there are

Inversions + Gene Expression:

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.

Queenie is working on creating summary statistics and graphics for the inversions (how big are they, in what taxa..etc.). **If you have any things you would like to know regarding this, please let me know!**

I have been trying to create a mapping of which genes match genes in all taxa (based on the alignment) so this can be used to compare genes in inverted and non inverted regions.

Substitution Paper I have spent most of the week working on the simple revisions from the paper. Adding extra references, explanations, and small analysis (larger window size).

I am still trying to figure out how to add more species to the analysis. I am starting to re-do the analysis with all *S. meliloti* complete genomes available (23) and seeing what happens. I am also thinking about how we identify similar sequences and wondering if we should be changing the methods. I know we are doing everything we can to ensure that we are aligning homologous

sequences..but I worry. Although, I am not sure how we would re-do everything while still accounting for genomic reorganization. I am hoping that using blast to verify the alignment in the inversions project will be helpful here. Maybe I could use this as proof to say that we blasted things and still it looks good (or removed things that were not RBBHs), not sure. **let me know what you think about completely re-hauling the methods.**

This Week

- complete mapping of all genes in all strains
- get Queenie to use ↑ information in her summary statistics of inversions
- get Queenie to start creating visualizations on the summary statistics for inversions
- complete larger window size supplementary analysis
- keep working on scaling up the subst data
- add in new paper notes to subst intro

Next Week

- keep working on scaling up subst data
- comment on repeated genes (TEs) in subst analysis and not using core genome
- get Queenie to create a plot of the inversions
- think about (and execute) how to incorporate distance from the origin into the inversion analysis