Subs Paper Things to Do:

- ~~why does sinoC have omega lin reg = 0 near and far from the origin?~~
- create new graphs for selection analysis
- ~~find and example of high substitution bar in *Streptomyces* and put this into supplement as an example of really diverged taxa (and that subs are real!)~~
- ~~discuss removing omega outliers in methods~~
- ~~double check that the ter and ori and max genome pos are correct~~
- ~~make graphs proportional to length of respective cod/non-cod regions~~
- ~~test examples for genes near and far from terminus (robust log reg/results)~~
- ~~linear regression on 10kb regions for weighted and non-weighted substitutions~~
- ~~average number of substitutions in 20kb regions near and far from the origin~~
- ~~figure out why the data is weird for number of cod/non-cod sites~~
- why are the lin reg of dN, dS and ω NS but the subs graphs are...explain!
- ~~grey out outliers in subs graphs?~~
- mol clock for my analysis?
- GC content? COG? where do these fit?

Gene Expression Paper Things to Do:

- if necessary add a phylogenetic component to the analysis
- codon bias?
- ~~make corrections based on Brian's edits~~
- ~~create a clean copy of the paper (no strikeout) for re-submission~~

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

✓define a theme for the substitutions graphs (and selection graphs)

✓Added selection values summary plot (violin plots) to supplement with caption to explain all the intricacies of the plot

✓attempted to use the median instead of the mean for the *S. meliloti* chromosome plot (did not make a difference)

✓check into high *dS* values for *Streptomyces*

✓why does it look like some bac have missing data (i.e. *B. subtilis*, *S. meliloti* chrom ...etc.)

I added in the version of the violin plot that you liked to the supplementary material of the manuscript and made an appropriate caption for it. This might get moved to the main paper later, but we will see!

I tried to use the median of each 100Kbp region as points for the trend lines in the selection distribution graph for *S. meliloti* chromosome, but it did not make a difference in making the graph "look" nicer or more accurate. So I will be sticking with what I did previously.

I looked into the two sections of the *Streptomyces* selection distribution graph that had very high *dS* values and they are real! These are genes that just have a lot of synonymous changes and only one non-synonymous change!

For the portions of the selection distribution graphs that look like there are large chunks of data missing, it looks like this is actually missing data. This is what I spent most of my week trying to figure out (past me was not as good at keeping notes as current me). This "missing data" was discarded at the stage where any block with a tree that was significantly different from the overall tree (via SH-Test in RAxML) was chucked. I had to re-do the calculations for which block trees did not match the overall tree. Below is Table 1, showing the results from the SH-test. For *Streptomyces*, there are only 3 taxa and you therefore can not perform an SH-test (it needs minimum of 4). I am not sure what to do because in the case of *E. coli* and the chrom of *S. meliloti*, about 25% of the whole alignment is thrown out. I am not sure if we need to do this step since we are now being very conservative with which alignments we are keeping later on (trimal + my code). Since the test can not be done on *Streptomyces*, I am also not sure if we should use it for the other replicons. **Thoughts?**

Bacteria Replicon	# of Total LCBs with Identical Tree	# of Total LCBs with Non-identical Tree	% of Total Alignment Discarded
E. coli Chromosome	30	7	25.44%
B. subtilis Chromosome	10	2	21.62%
Streptomyces Chromosome	NA	NA	NA
S. meliloti Chromosome	9	2	25.06%
S. meliloti pSymA	35	0	0%
S. meliloti pSymB	8	0	0%

Table 1: Number of Locally Colinear Blocks that had identical topologies to the "super sequence" tree, not identical to the "super sequence" tree, and the proportion of the total alignment that was represented by the non-identical tree topologies. Topologies that were not identical were determined to be different at the 5% significant value using an SH-test in `RAxML` cite raxml here!. The SH-Test could not be preformed on *Streptomyces* as there were only three taxa present and `RAxML` needs a minimum of 4 taxa for the test cite raxml.

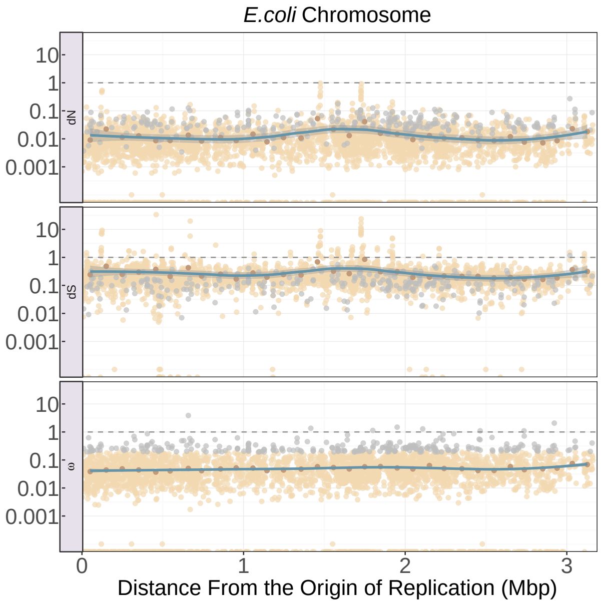
This Week

- find a block where mauve aligns non-homologous regions and put into supplement
- re-do selection and substitutions analysis (if necessary, because of not throwing out blocks with different trees than the overall tree)
- re-do substitutions and selection graphs (with new theme) and these and put in paper

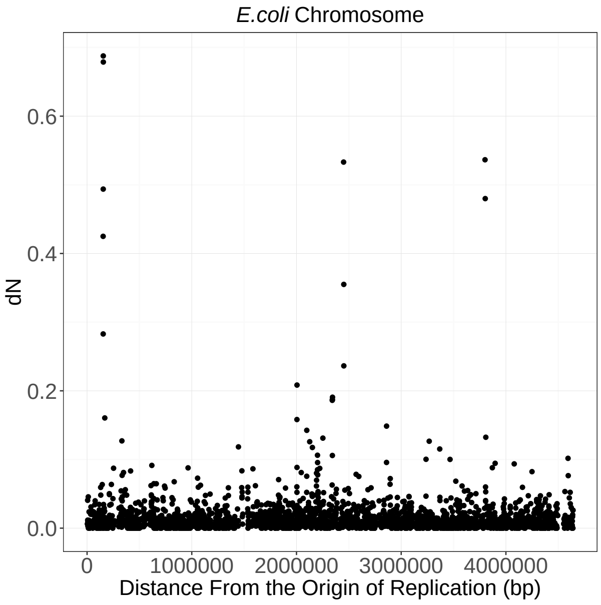
Next Week

- look into whats up with *S. meliloti* chrom bc it does not look right at all

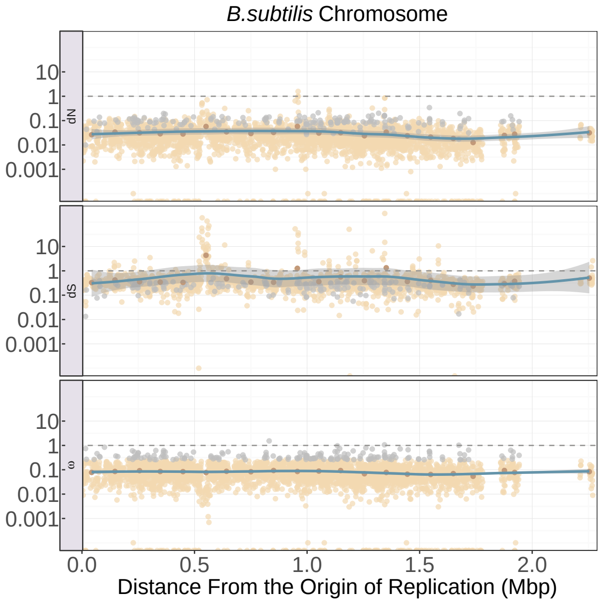
- update methods in paper draft (make sure they are the same as what I actually did bc things have changed a lot)
- add reviewers comments from gene expression paper into this one (most will apply)



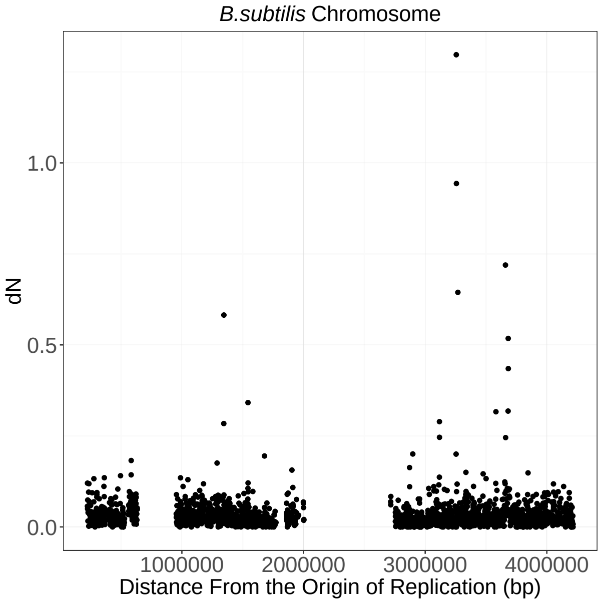
(a) Selection data **accounting** for bidirectional replication



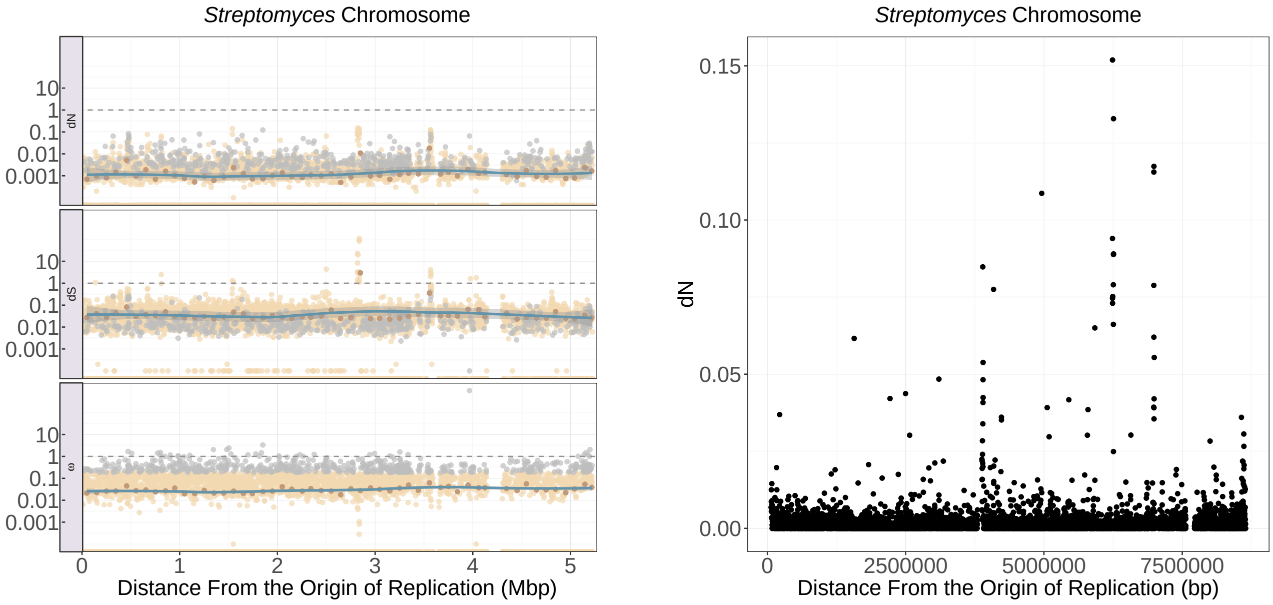
(b) Selection data with NO bidirectional replication accounted for. To show where missing data is



(a) Selection data **accounting** for bidirectional replication

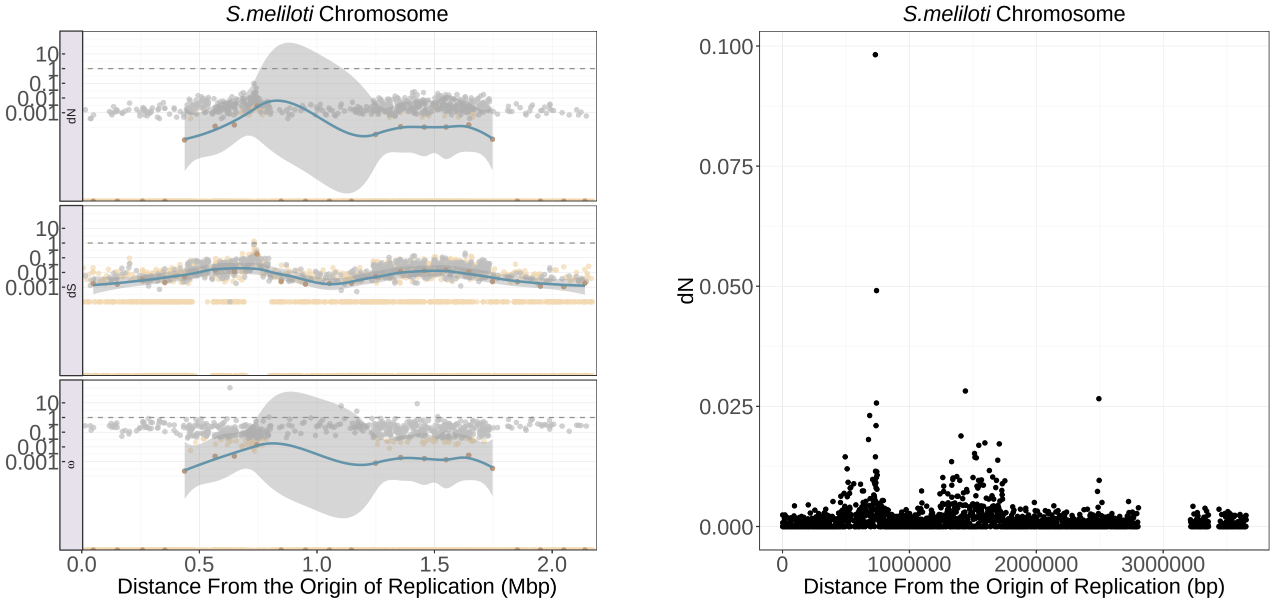


(b) Selection data with NO bidirectional replication accounted for. To show where missing data is



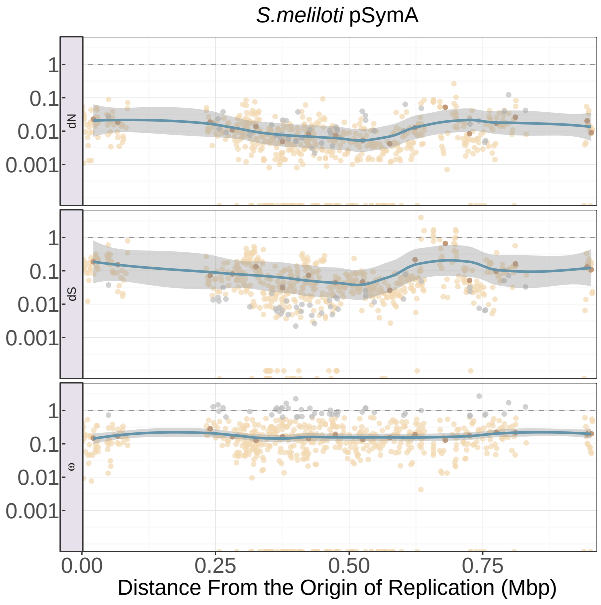
(a) Selection data **accounting** for bidirectional replication

(b) Selection data with NO bidirectional replication accounted for. To show where missing data is

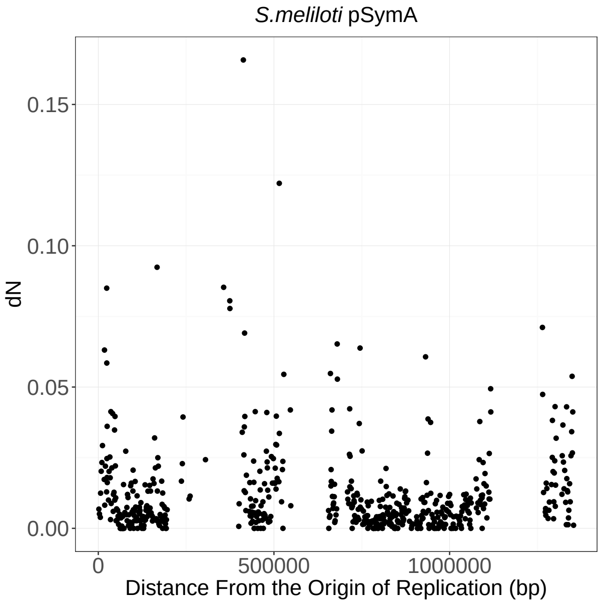


(a) Selection data **accounting** for bidirectional replication

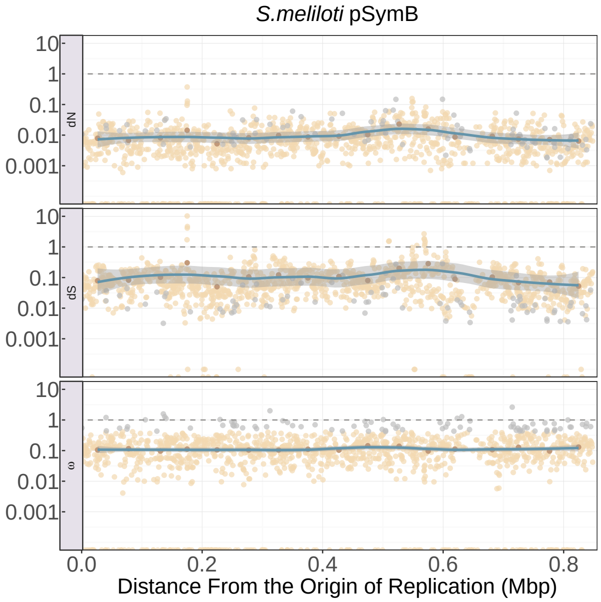
(b) Selection data with NO bidirectional replication accounted for. To show where missing data is



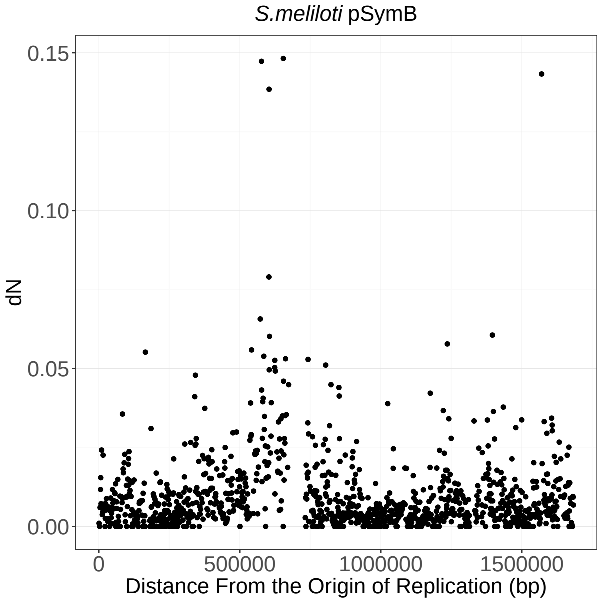
(a) Selection data **accounting** for bidirectional replication



(b) Selection data with NO bidirectional replication accounted for. To show where missing data is



(a) Selection data **accounting** for bidirectional replication



(b) Selection data with NO bidirectional replication accounted for. To show where missing data is