

Subs Paper Things to Do:

- why does ~~sineC~~ 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  $\omega$  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). I am working to figure out where this data was “thrown out”. I think it might have been at the stage where any block with a tree that was significantly different (via SH-Test in RAxML) was chucked. I am still working this out.

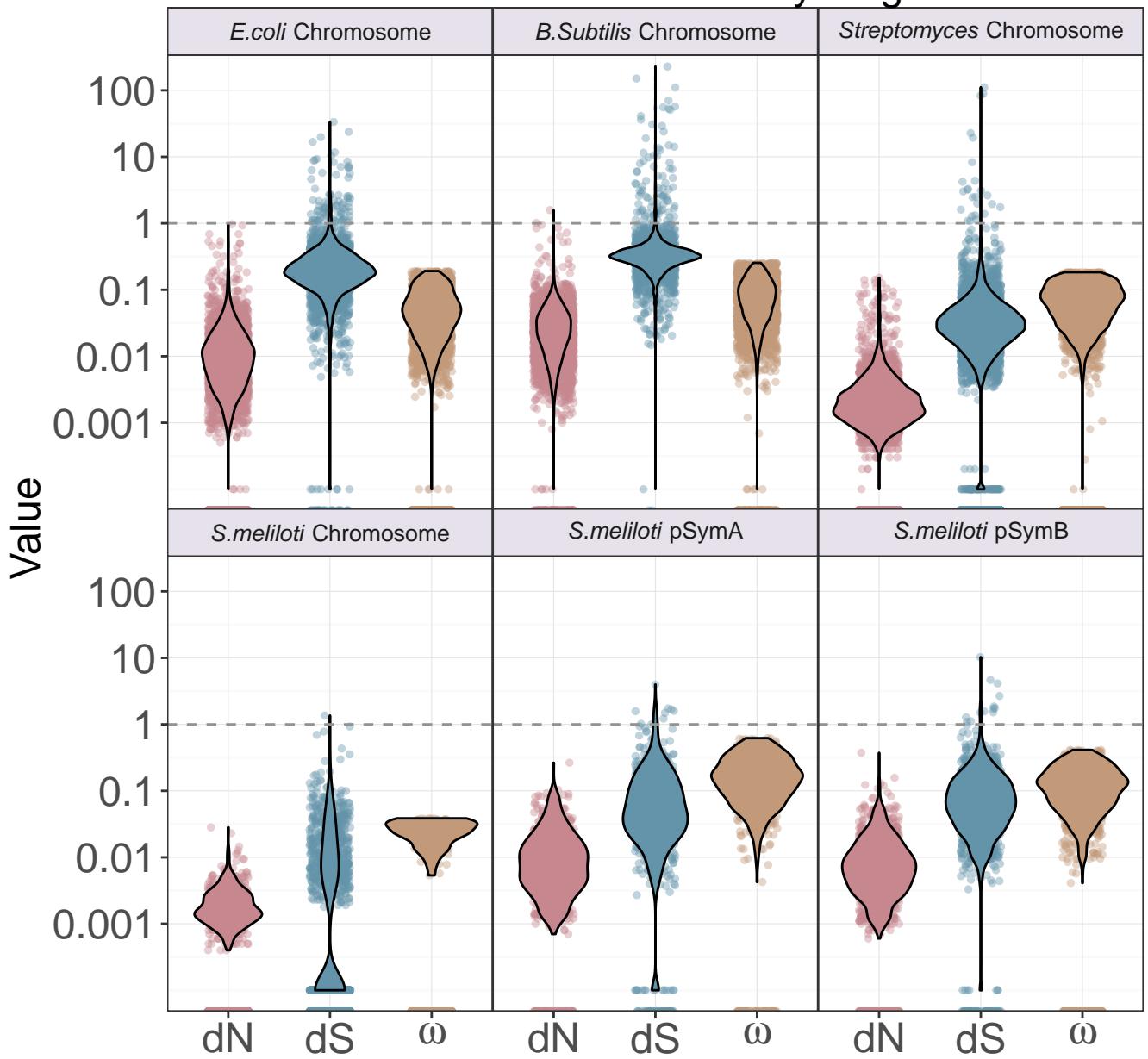
## This Week

- check why there are gaps in the selection data (across the genome) in some of the bac (*B. subtilis*, *S. meliloti* chrom..etc)
- check 2 spots where  $dN$  and  $dS$  are high in *Streptomyces*
- try computing median instead of mean for *S. meliloti* chrom to see if that makes the graphs look different

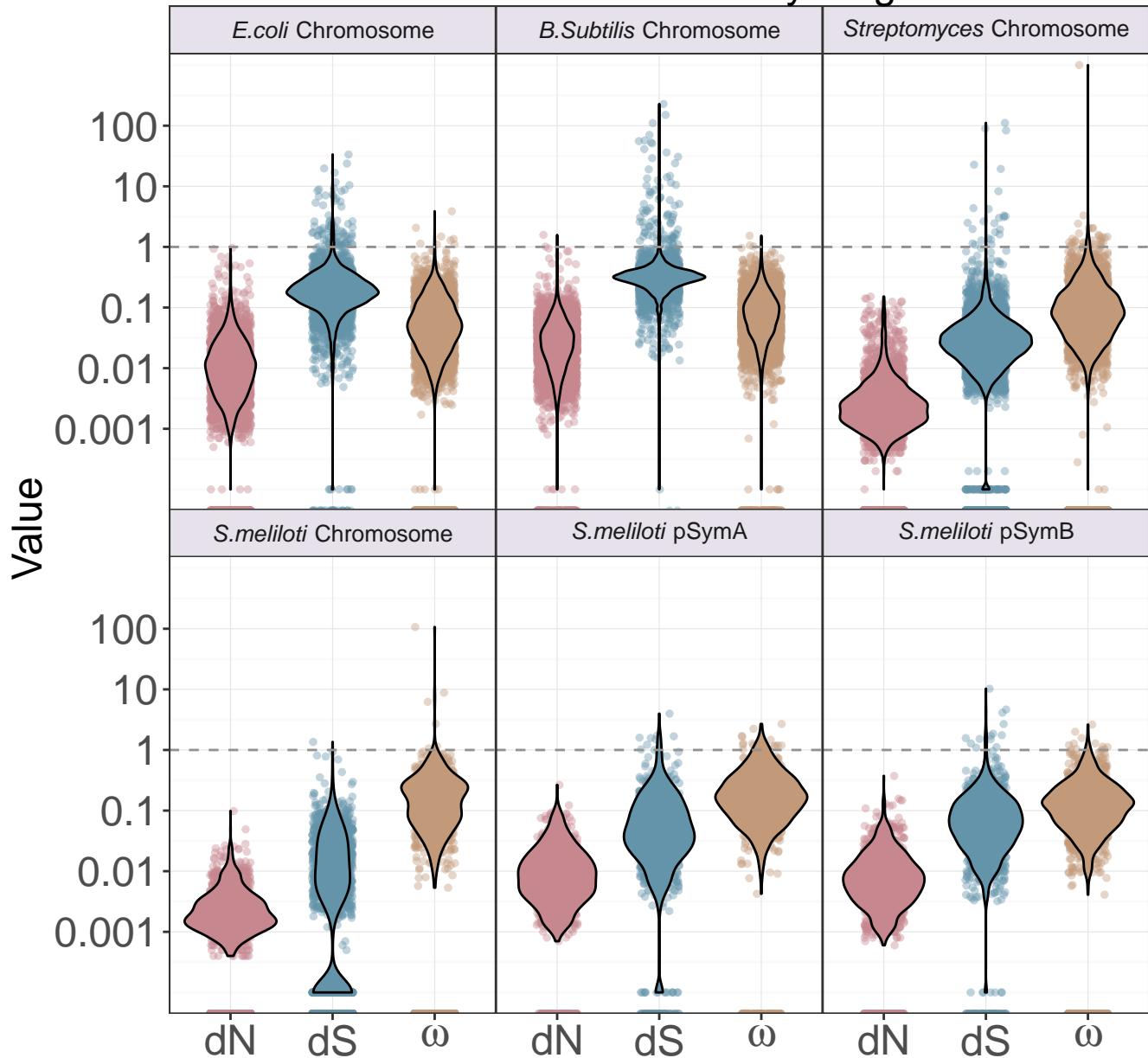
## Next Week

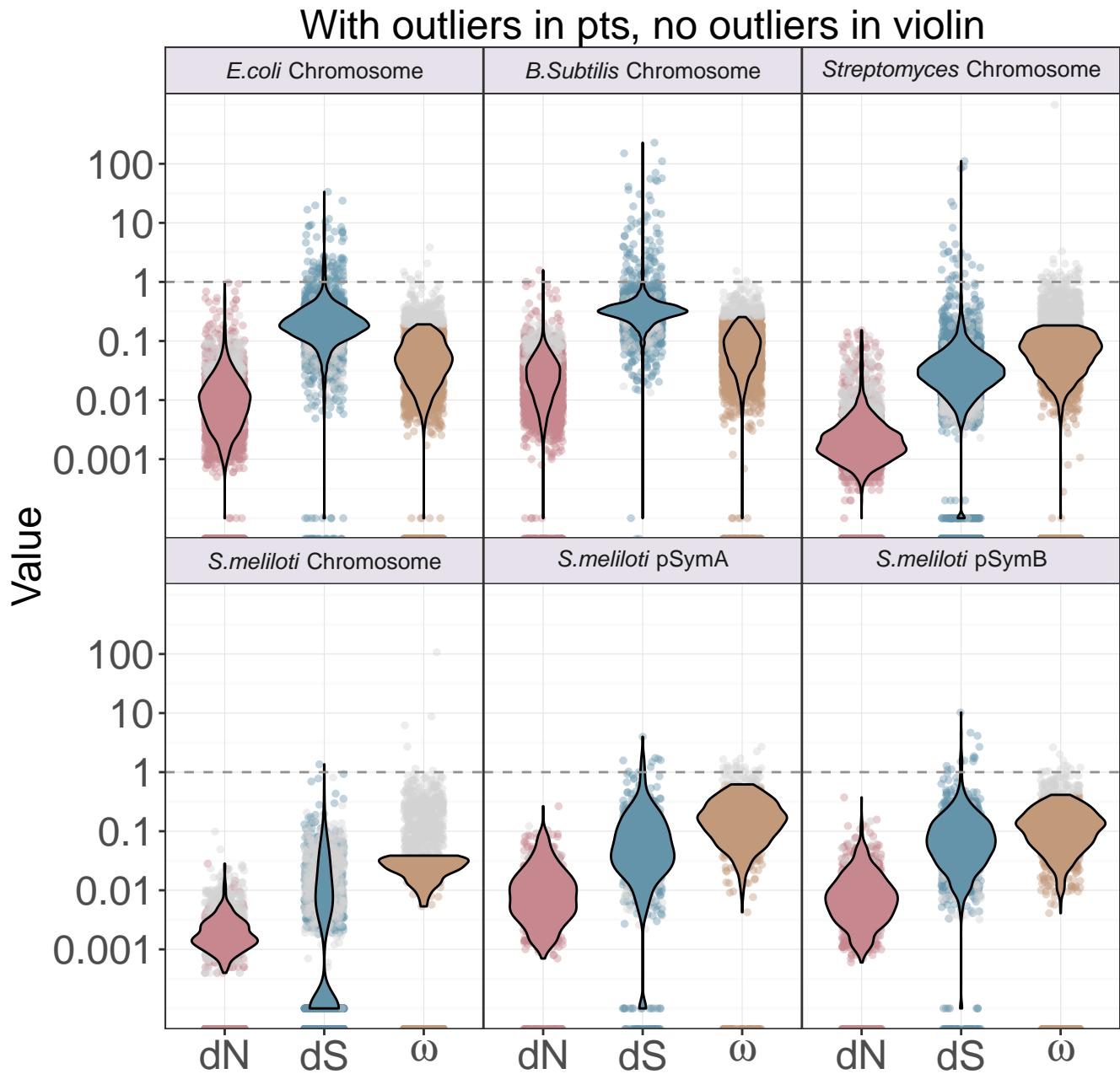
- find a block where mauve aligns non-homologous regions and put into supplement
- re-do substitutions and selection graphs (with new theme) and these and put in paper
- look into whats up with *S. meliloti* chrom bc it does not look right at all

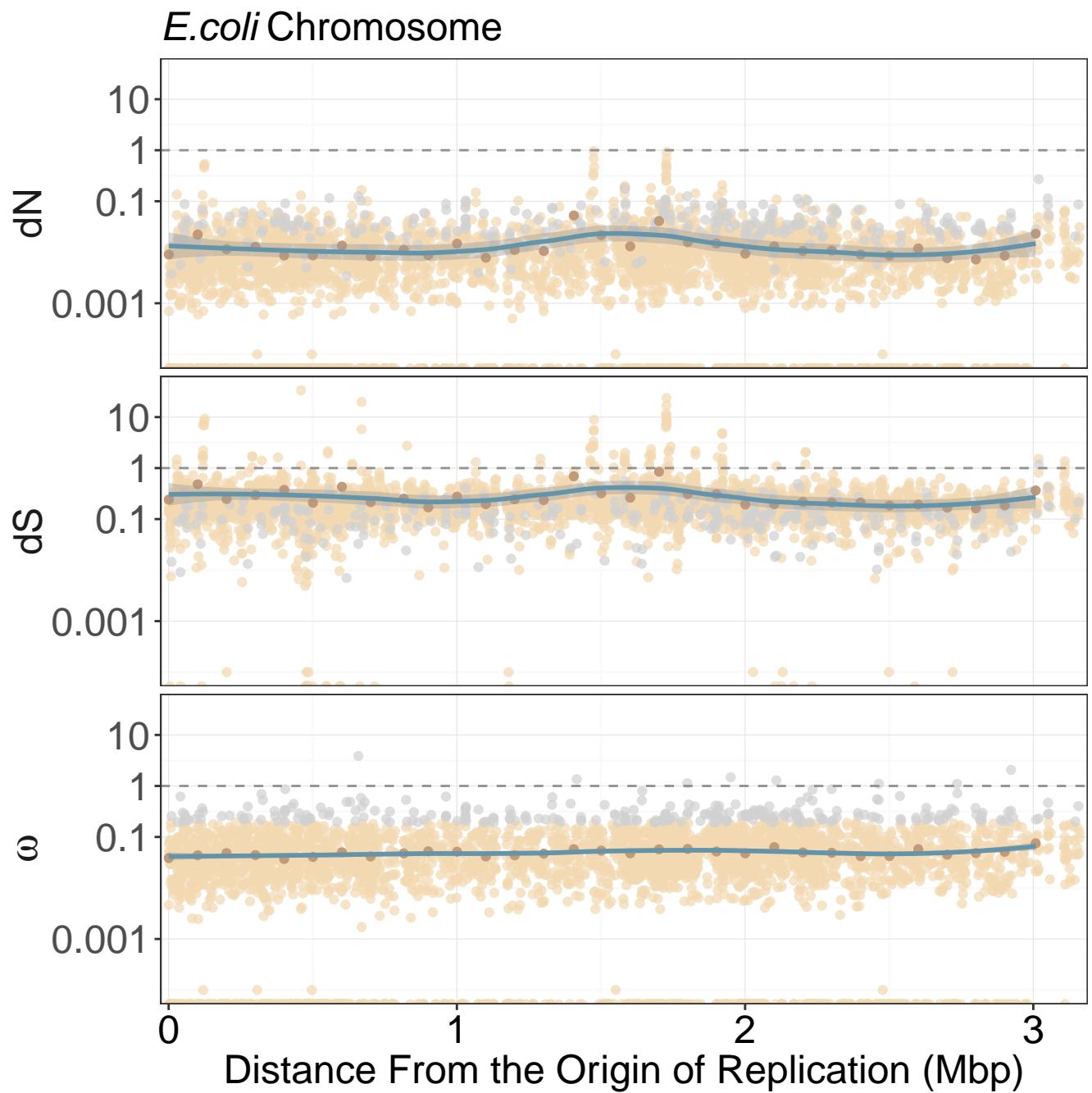
## Without outliers in anything

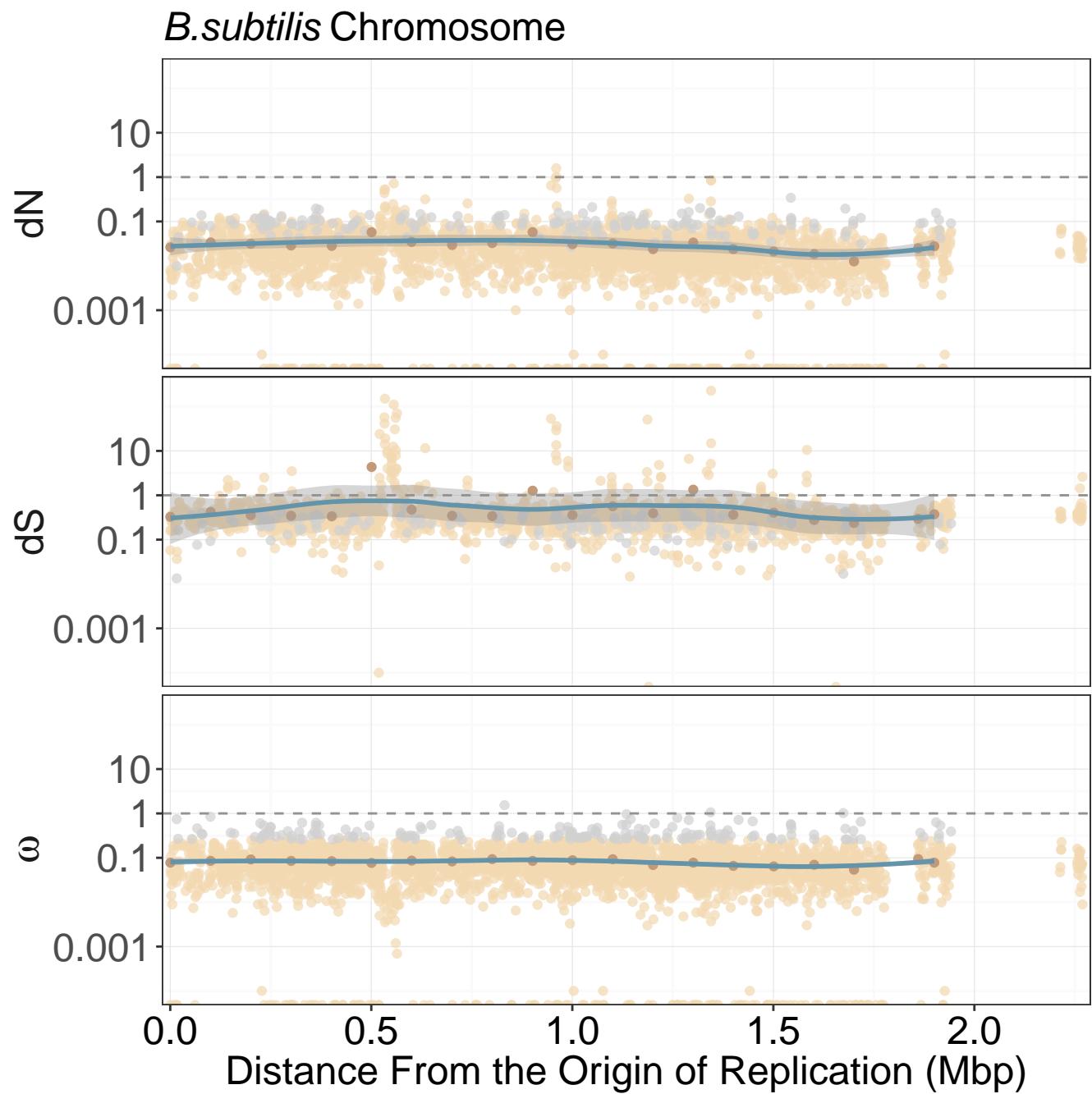


## With outliers in everything

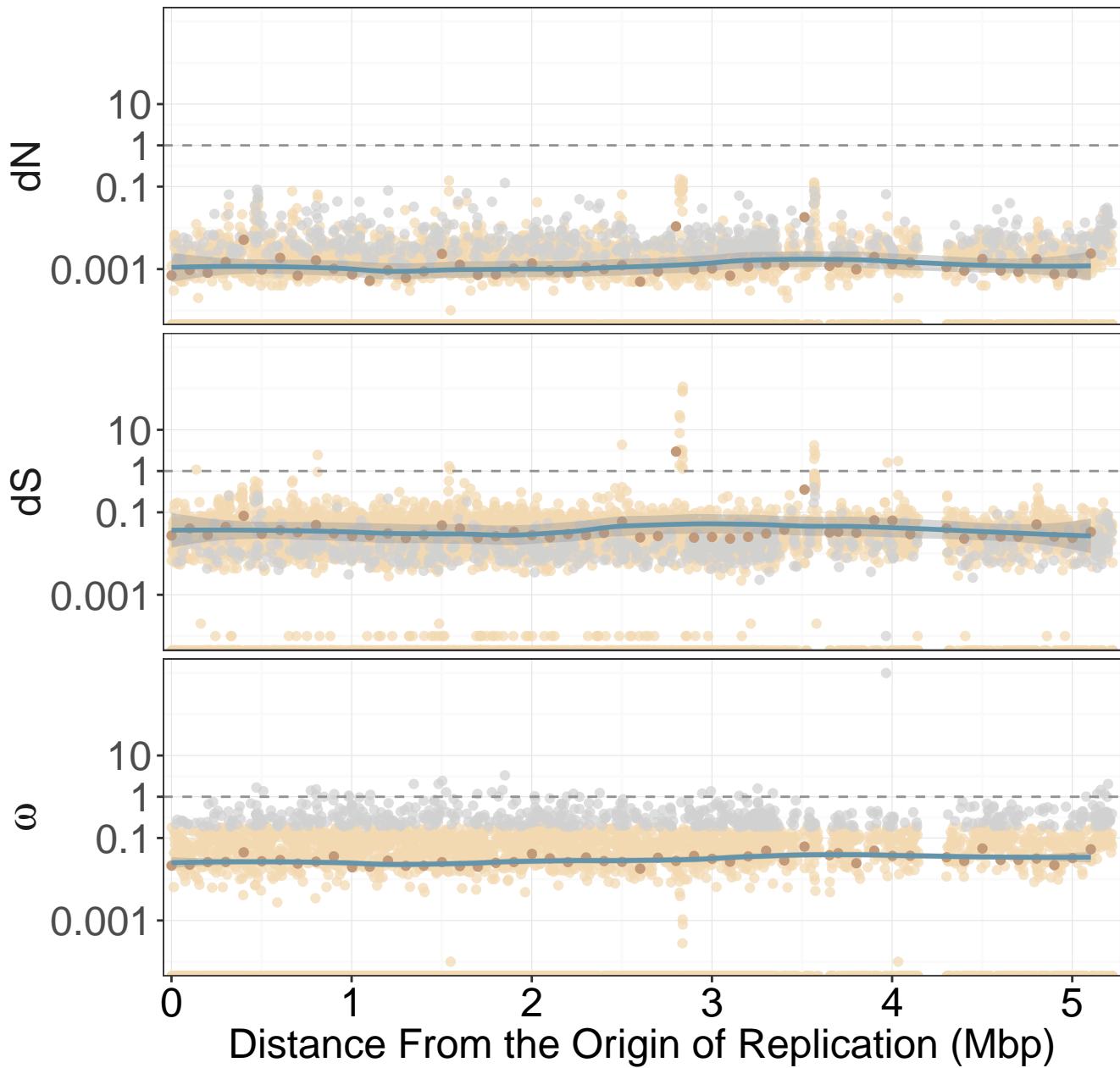


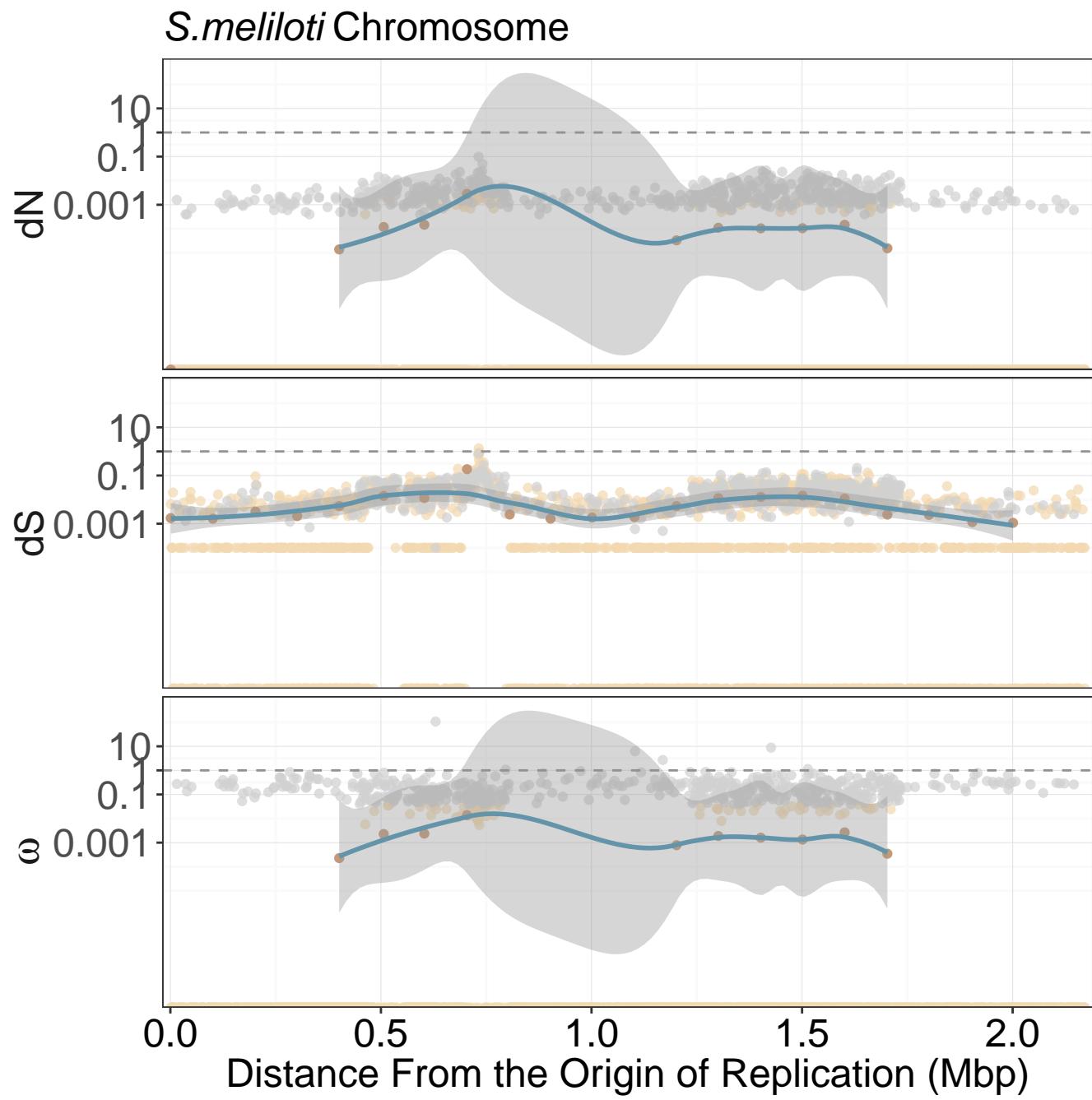






## *Streptomyces* Chromosome





*S.meliloti* pSymA