

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

Gene Expression: ✓ published online! Woo!

✓ added program version table to GitHub

Substitutions:

✓ finished final edits to the manuscript!

Inversions + Gene Expression:

✓ started combining gene expression and PARSNP info

✓ finished running BLAST with various parameters

✓ Queenie: created graph to compare gene expression between datasets from the same strain

Inversions + Gene Expression: Queenie created a graph comparing the expression values between datasets from the same strain (similar to what we did for the gene expression paper) and this is in Figure 1. The values look like they are all mixing well together so I think it is fine that we take the average expression value of all the datasets for each gene.

Queenie is slowly working away at the tasks I gave her: verifying that gene expression is consistent across datasets for each strain and combining the parsnp inversions information with the gene expression + genome position data frame. I have been working on the blast portion of this project and running blast with multiple parameters to see what we get.

For the BW1255 strain, since one of the data sets is mapped to the K-12 genome and the other is unclear of where it is mapped, I think that using the K-12 genome annotation for this strain is best. Additionally, since the proteomes are redundant, I think we should have minimal issues using the K-12 annotation. I have asked Queenie to see how often the genome positions for the same gene (in the gbk file) differ between BW and K-12, so we can get a sense/justify if this is the correct decision.

This Week

- check on Queenie's progress and double check her normalization code
- Queenie should be done graphs to check that expression between samples is comparable
- Get Queenie started on combining information about Parsnp inversions and gene expression data
- continue with blast (extract info from blast results)

Next Week

- help queenie with anything else she might need
- continue to work on blast (reciprocal part of blast hit and extract info from blast)
- edit dissertation intro
- submit substitutions paper

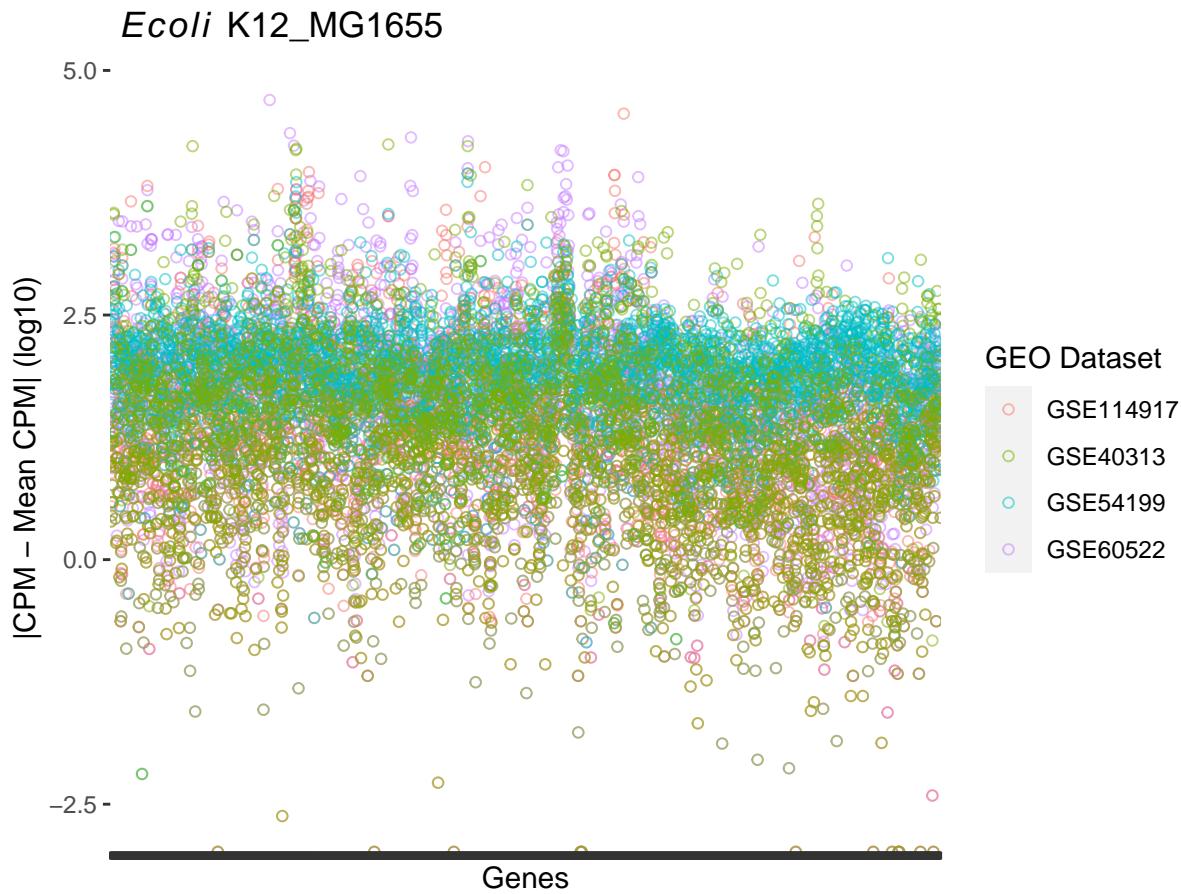


Figure 1

Bacteria and Replicon	Genome Average		
	dS	dN	ω
<i>S. meliloti</i> Chrom + <i>A. tumefaciens</i>	12.5529	0.0553	0.0265
<i>E. coli</i> Chromosome	0.2387	0.0101	0.0441
<i>B. subtilis</i> Chromosome	0.4201	0.0243	0.0714
<i>Streptomyces</i> Chromosome	0.0458	0.0011	0.0335
<i>S. meliloti</i> Chromosome	0.0029	0	0
<i>S. meliloti</i> pSymA	0.0835	0.0099	0.1645
<i>S. meliloti</i> pSymB	0.0940	0.0084	0.1142

Table 1: Weighted averages calculated for each bacterial replicon on a per genome basis using the gene length as the weight. Arithmetic mean calculated for the per gene averages for each bacterial replicon.

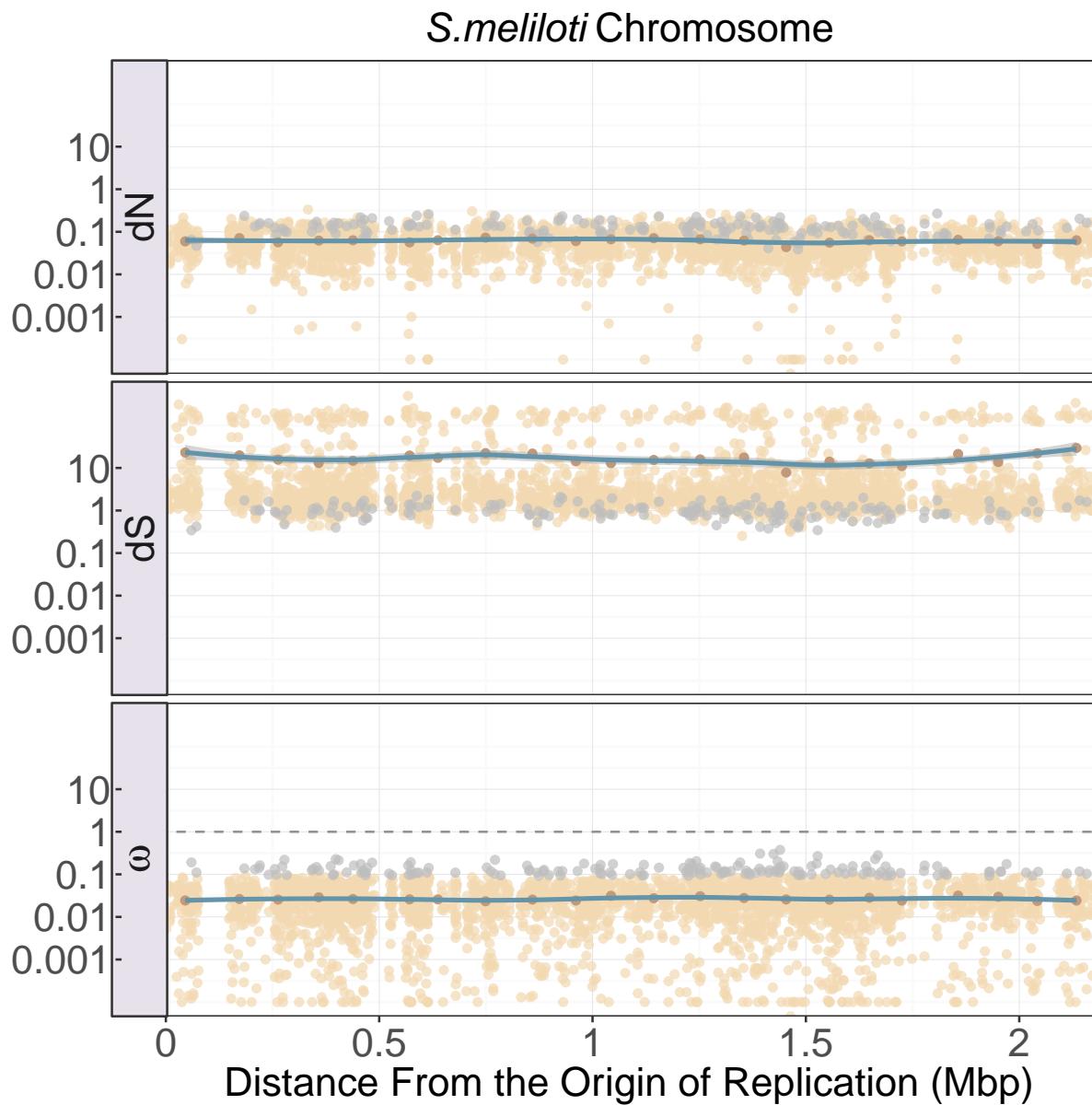
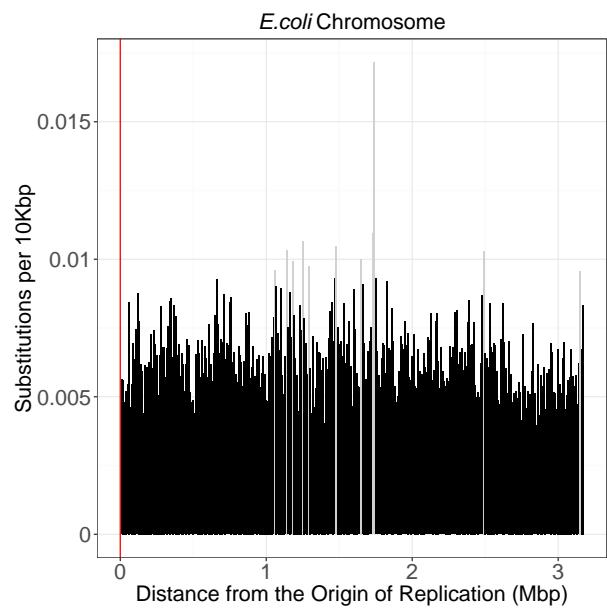


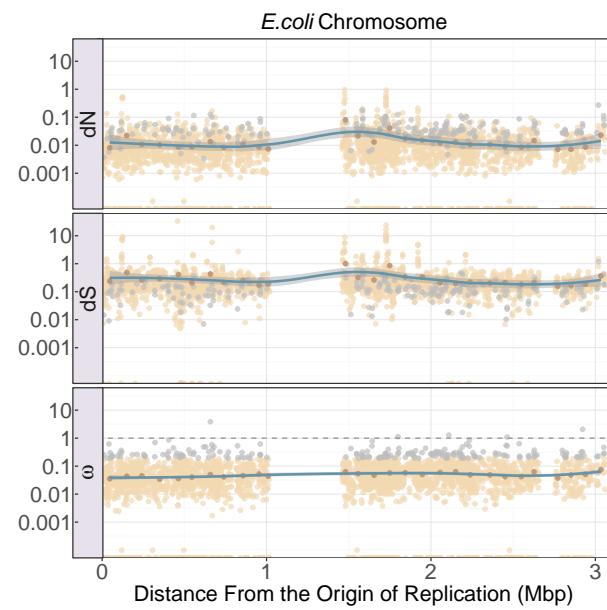
Figure 2: dN , dS , and ω values for *S. meliloti* chromosomes and *A. tumefaciens*.

Bacteria and Replicon	Average Number of Substitutions per bp
<i>E. coli</i> Chromosome	1.97×10^{-4}
<i>B. subtilis</i> Chromosome	1.93×10^{-4}
<i>Streptomyces</i> Chromosome	2.74×10^{-6}
<i>S. meliloti</i> Chromosome	9.72×10^{-5}
<i>S. meliloti</i> pSymA	6.54×10^{-5}
<i>S. meliloti</i> pSymB	1.99×10^{-4}

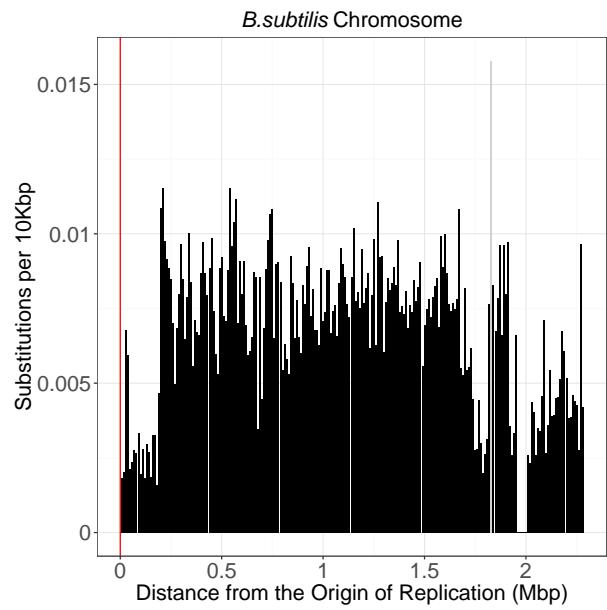
Table 2: Average number of protein coding substitutions calculated per base across all bacterial replicons. Outliers and missing data was not included in the calculation.



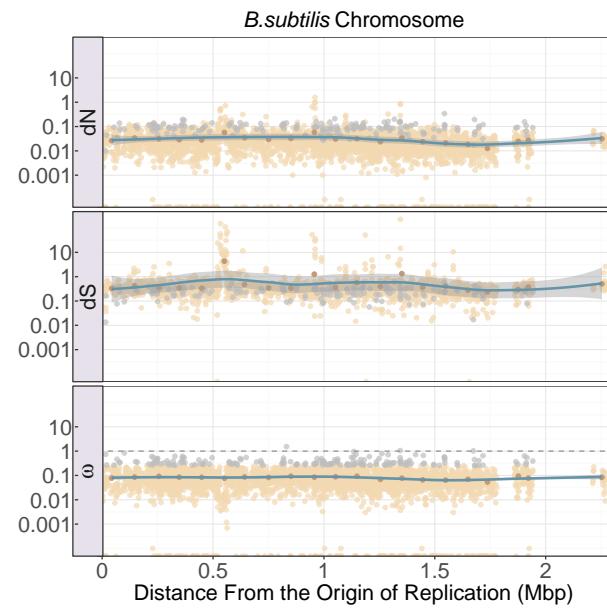
(a)



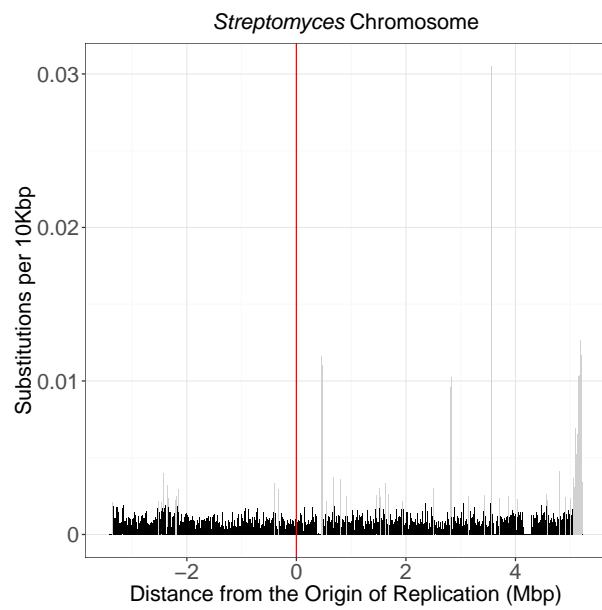
(b)



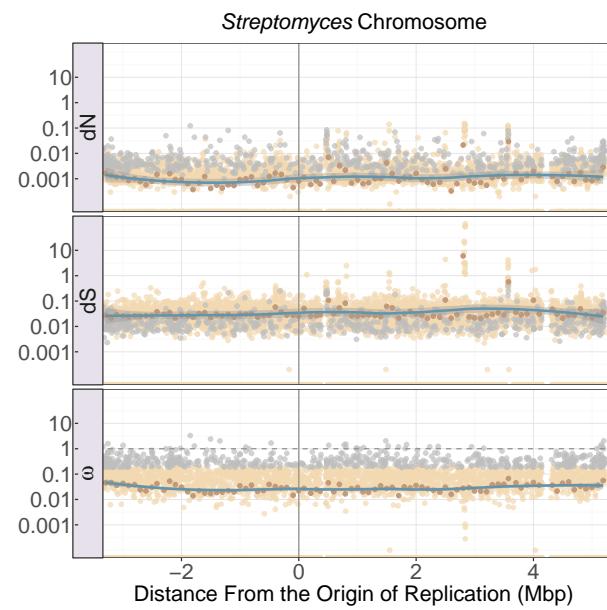
(a)



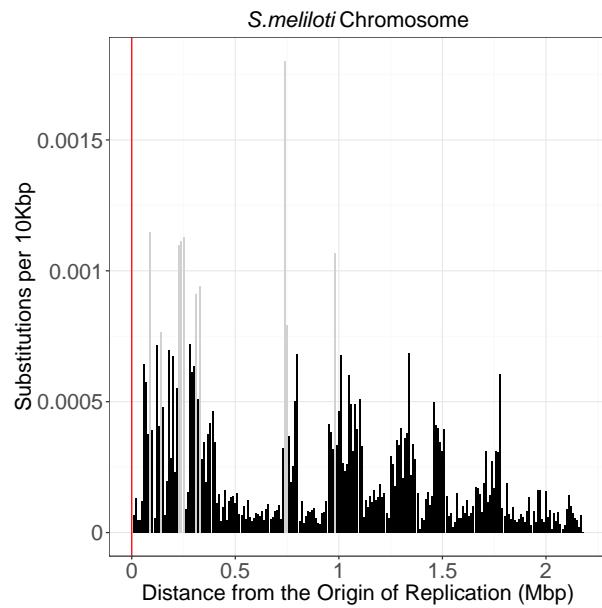
(b)



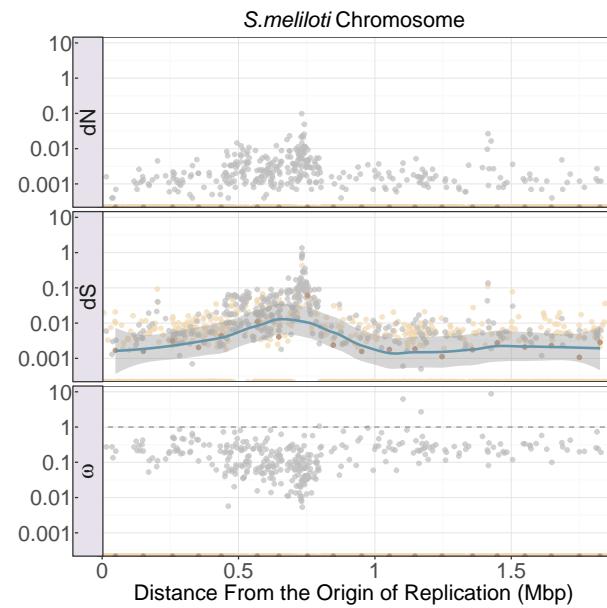
(a)



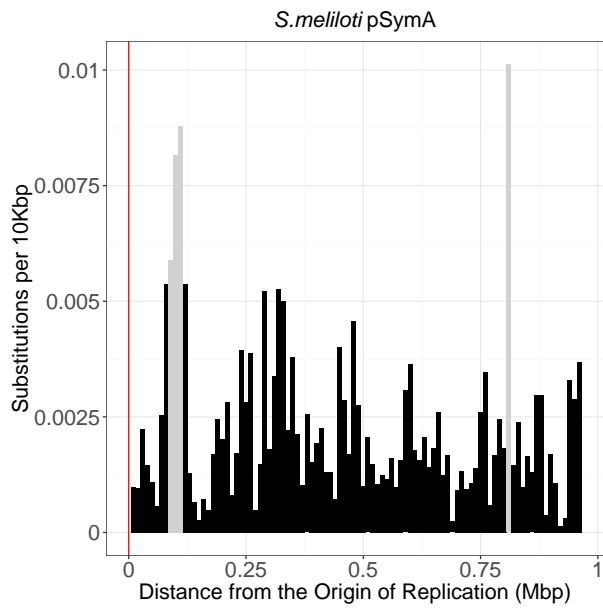
(b)



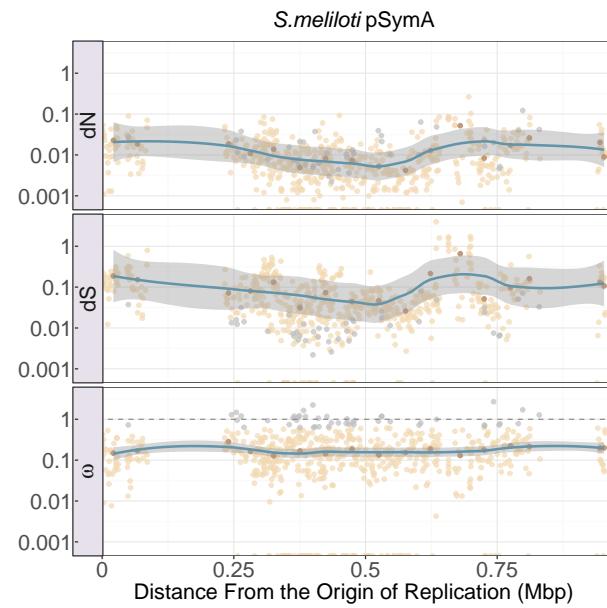
(a)



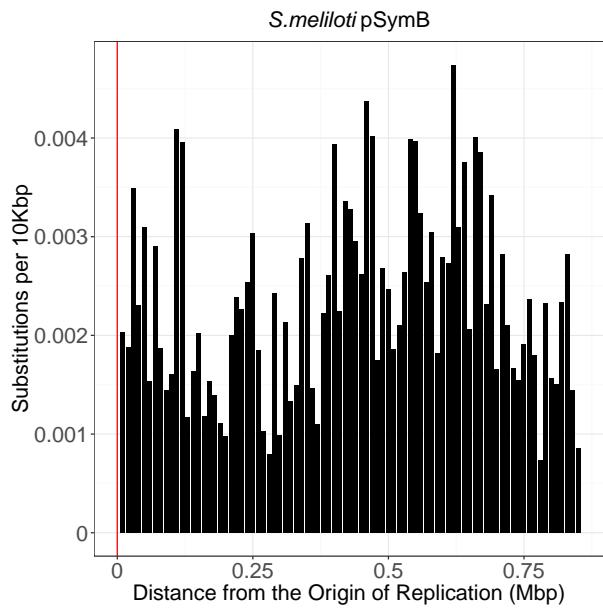
(b)



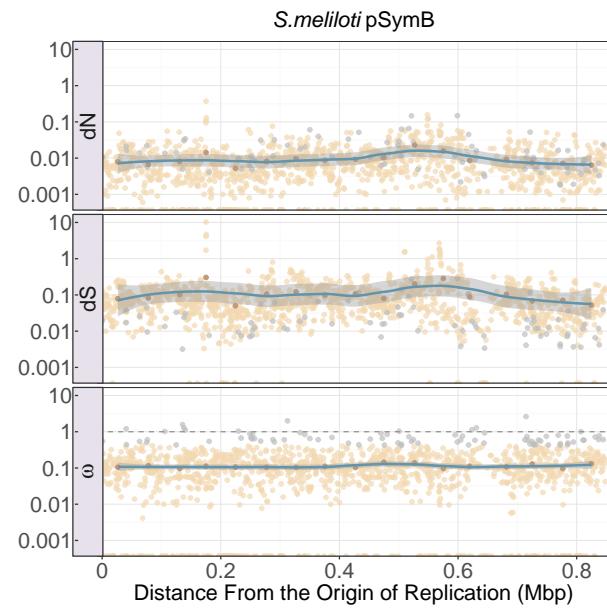
(a)



(b)



(a)



(b)

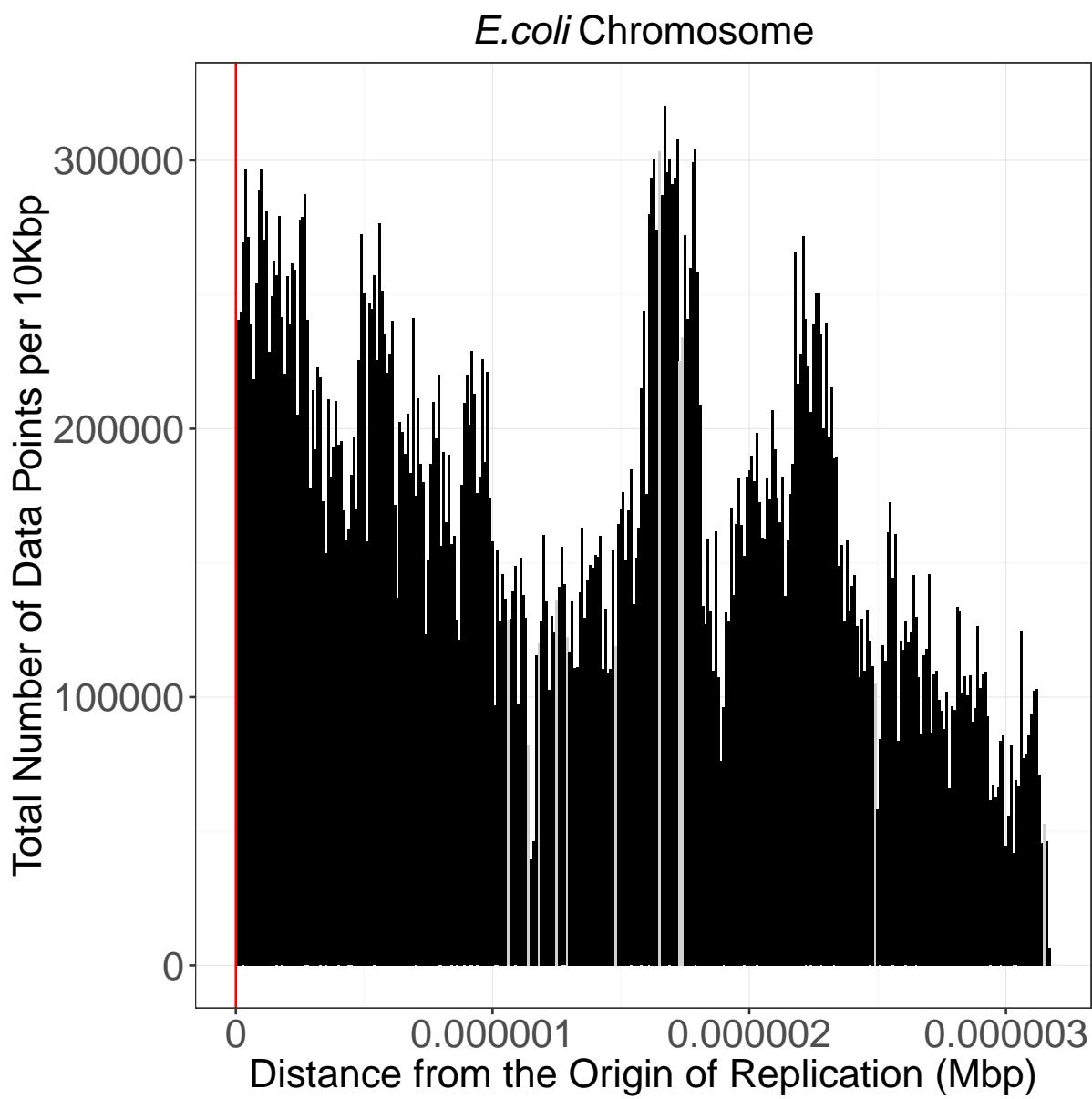


Figure 9: Distribution of total number of substitution data points per 10Kbp in genome.

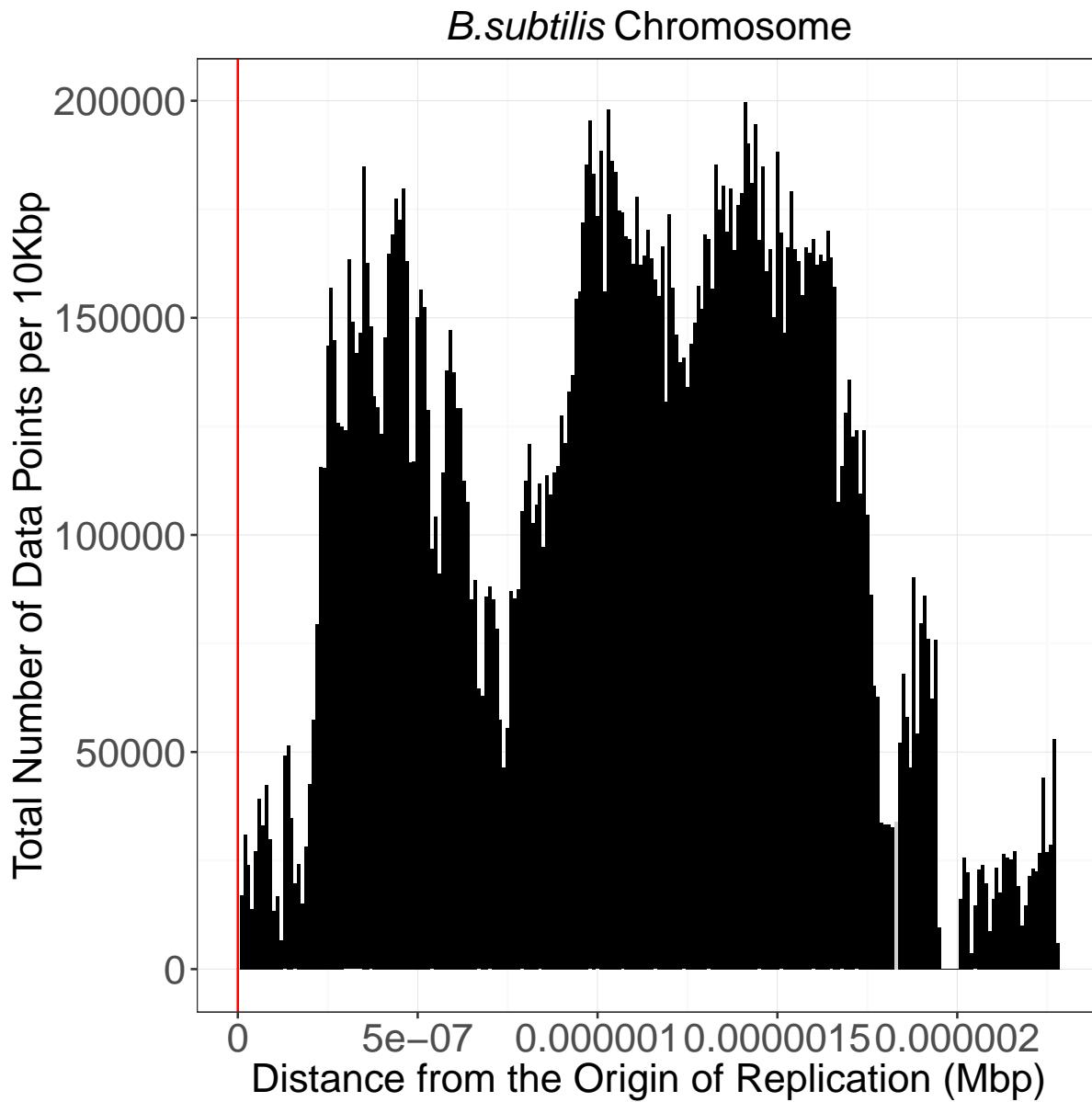


Figure 10: Distribution of total number of substitution data points per 10Kbp in genome.

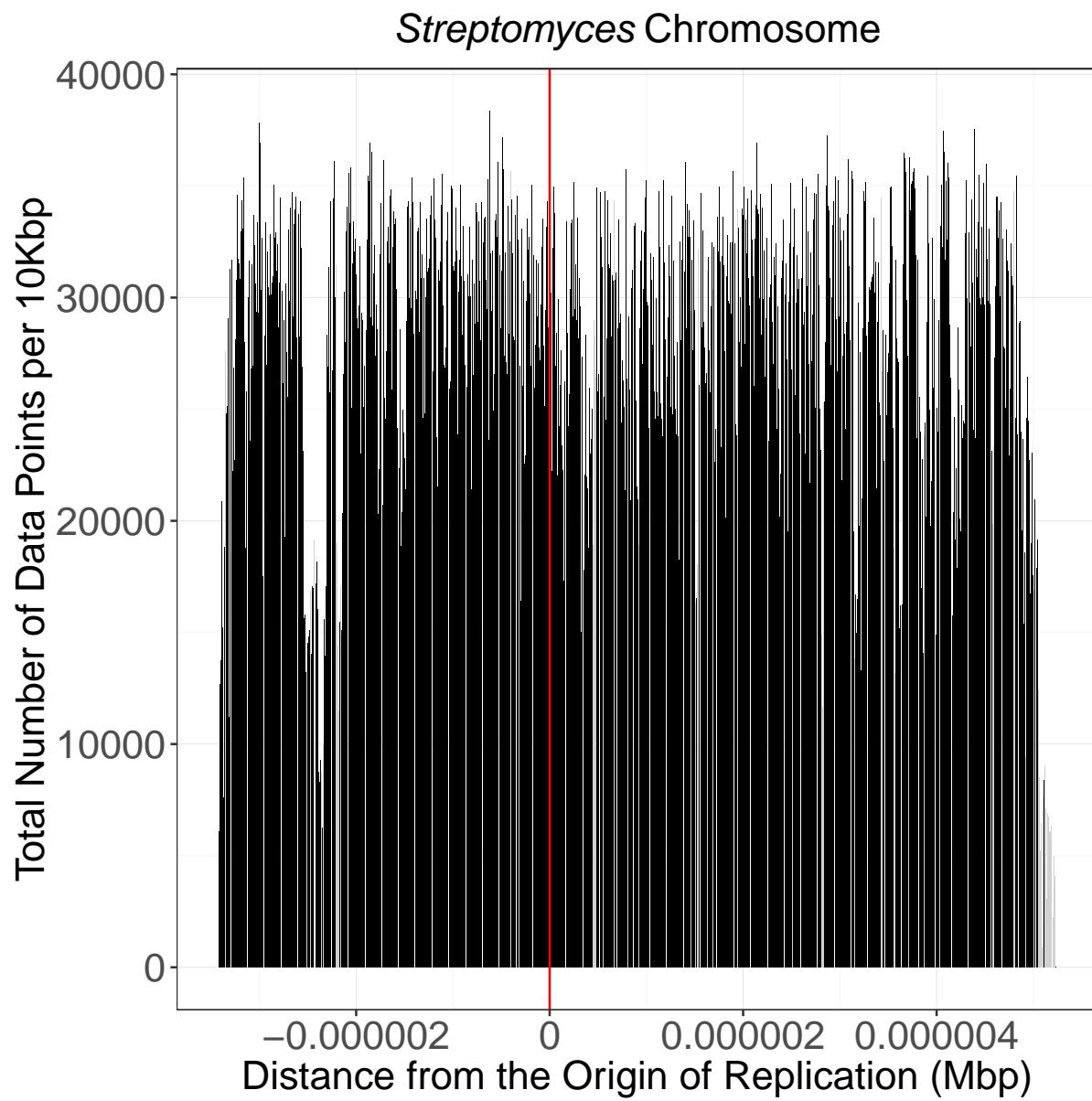


Figure 11: Distribution of total number of substitution data points per 10Kbp in genome.

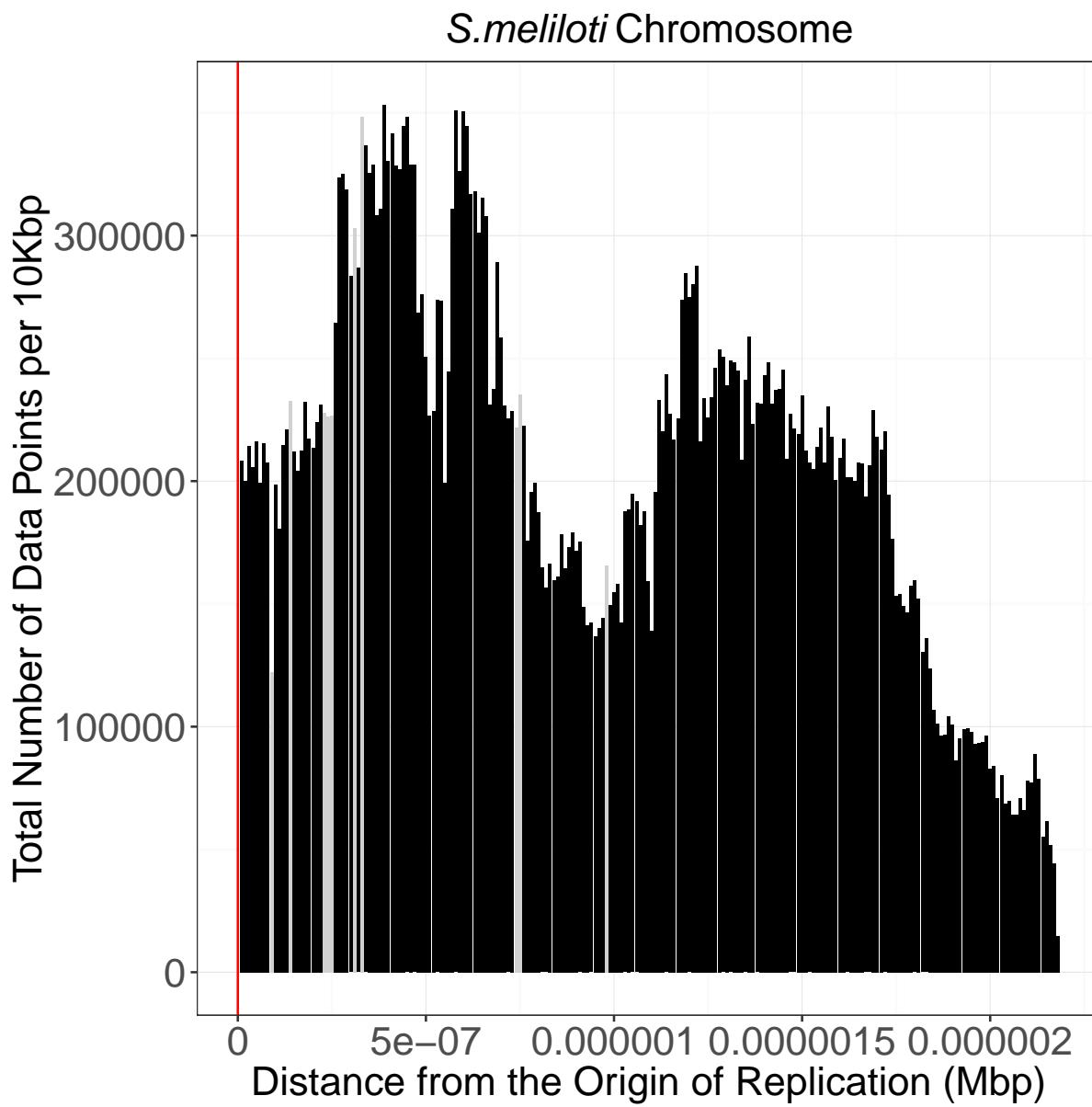


Figure 12: Distribution of total number of substitution data points per 10Kbp in genome.

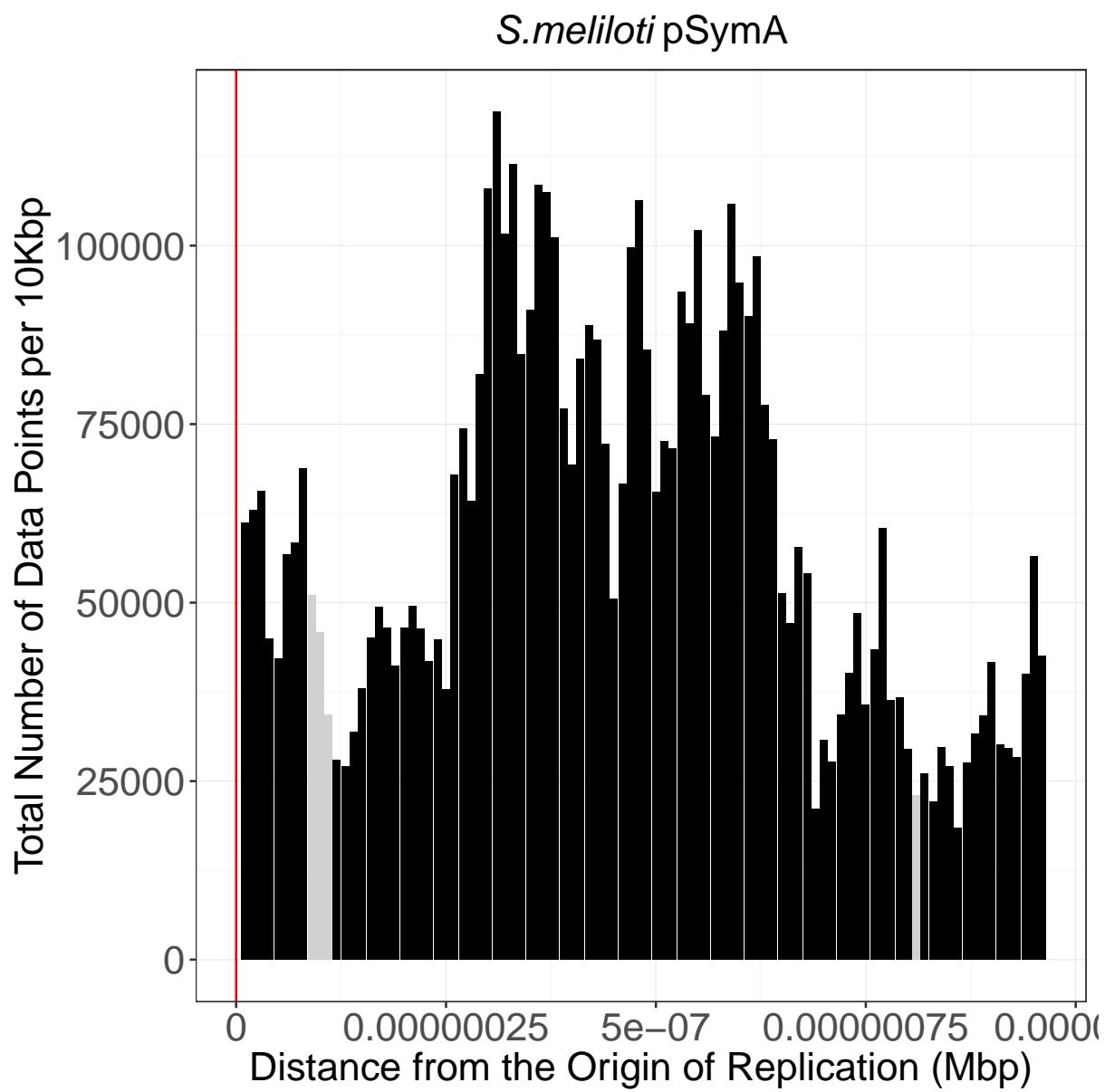


Figure 13: Distribution of total number of substitution data points per 10Kbp in genome.

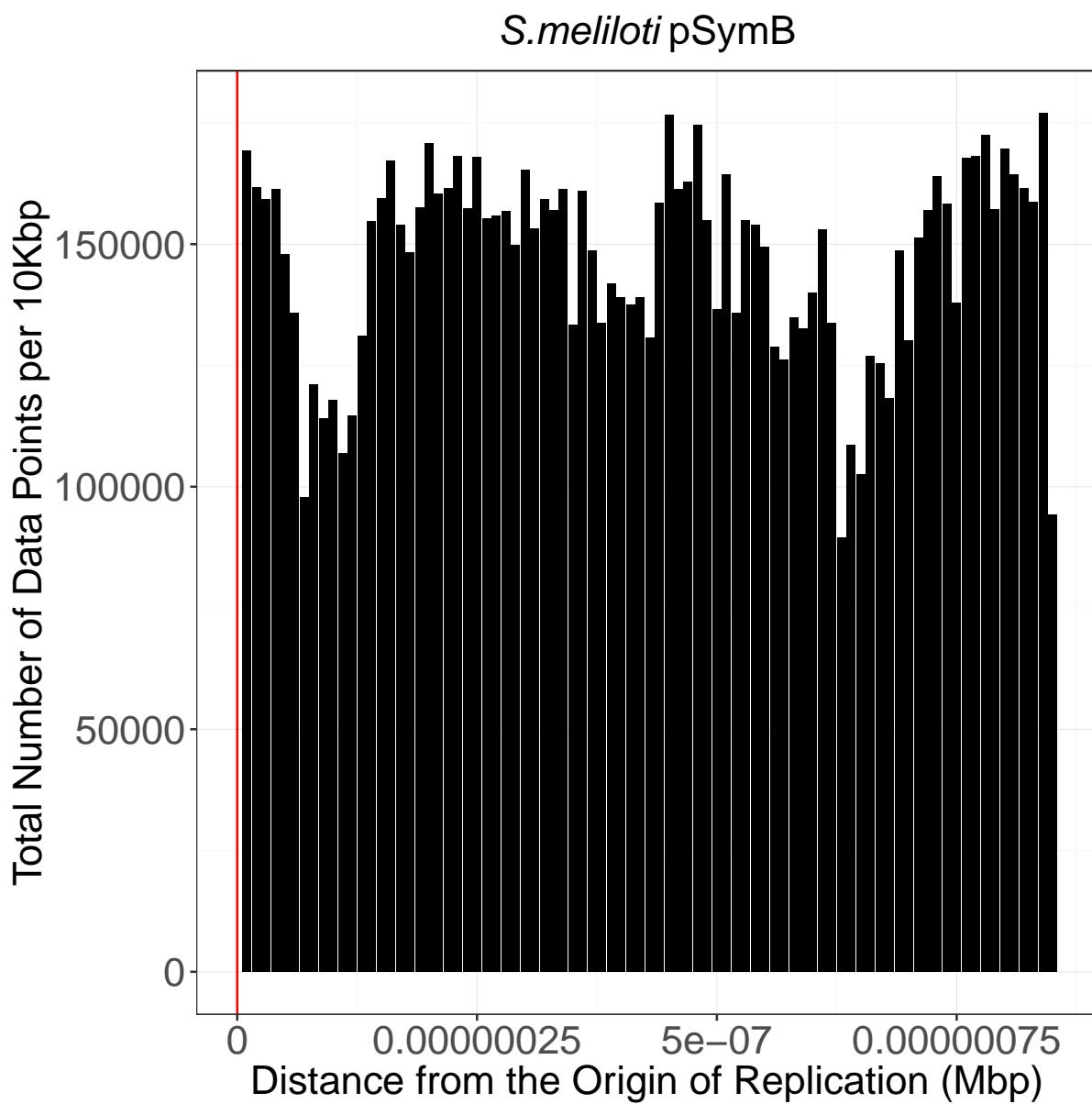


Figure 14: Distribution of total number of substitution data points per 10Kbp in genome.

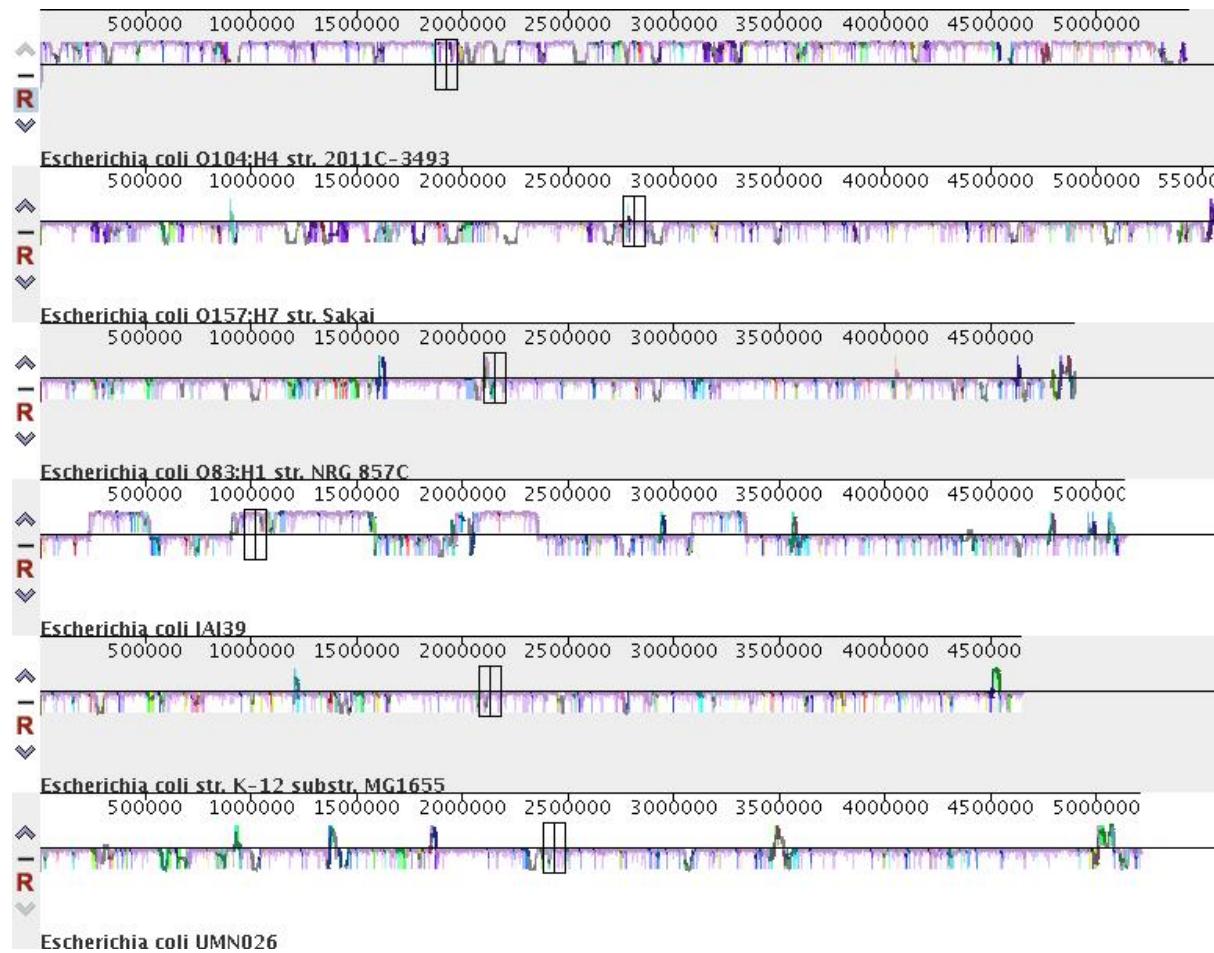


Figure 15: progressiveMauve alignment of *Escherichia coli* genomes highlighting the “backbone” of the alignment (matching regions).



Figure 16: progressiveMauve alignment of *S. meliloti* Chromosomes highlighting the “backbone” of the alignment (matching regions).

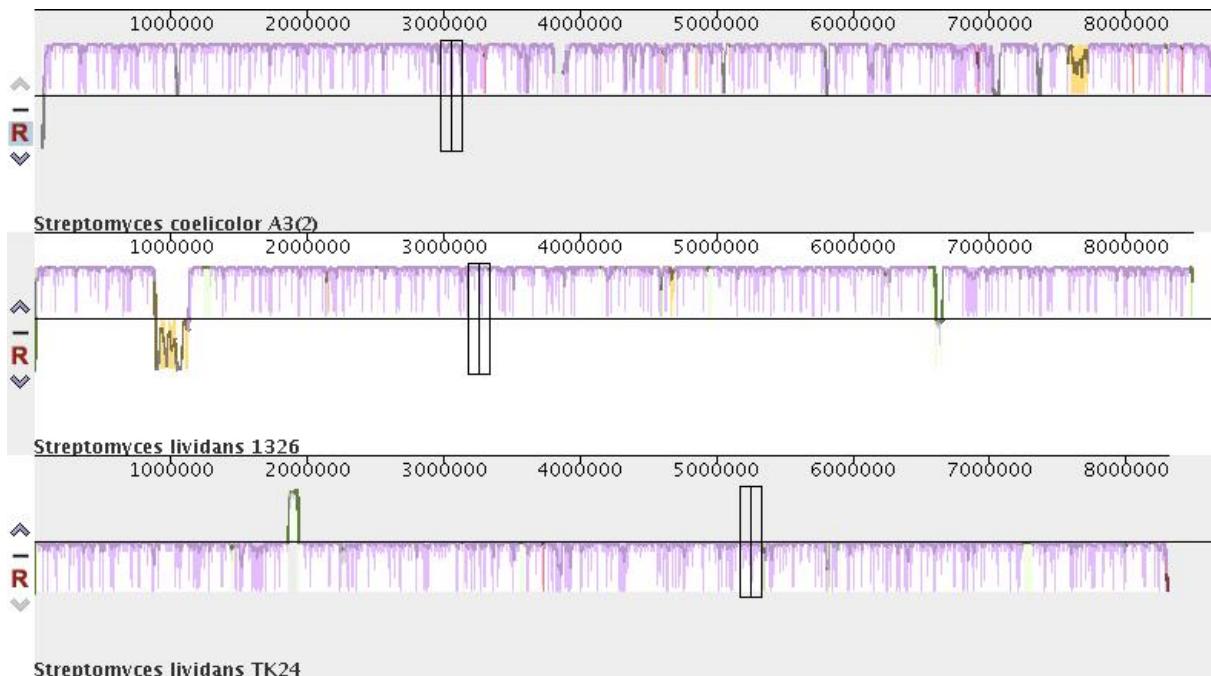


Figure 17: progressiveMauve alignment of *Streptomyces* genomes highlighting the “backbone” of the alignment (matching regions).