

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?

Gene Expression Paper Things to Do:

- if necessary add a phylogenetic component to the analysis
- codon bias?

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

✓supplemental selection analysis including outliers for *S. meliloti* chromosome

Edits to the Substitution Paper:

✓discuss selection linear regression results

✓alter main focus of paper in discussion and conclusion

✓interpret all other results

✓edit discussion and conclusion for flow

✓add in genome clustering figure to supplement to help explain that analysis

✓put new supplementary tables on git

✓bacteria names are italic in references

✓check all references work

✓edit abstract and intro

✓edit methods

✓edit results

✓check black and white look of figures (and alter colours/shapes as necessary)

I made lots of little edits and wrote more for the substitution paper (see above checklist). **The substitutions paper draft is almost complete so if you have any additional tests/analysis for this paper that you think I should do please tell me now so I can get to work on that. Otherwise, I think the analysis is done and the paper is ready for editing.**

I would like to get the formatting ready for submission of the paper (so it is a quick process when we are ready to submit). Last time we discussed this we talked about submitting to GBE, **please let me know if this is still the plan or if you think another journal would be better.**

The section below are the results from INCLUDING outliers (non-zero dN and ω values) in the *S. meliloti* chromosome selection analysis. The results are basically the same which is good! Because it means that even with these values, our point still stands that there is no evidence of a correlation between dN , dS and ω and distance from the origin of replication. **I would really appreciate it if you could read it over and let me know of any edits I should make.**

This Week

- send Brian results to edit
- incorporate Caitlin's comments on paper draft
- clean up code and add to substitutions paper git hub
- go over online checklist for substitutions paper and make sure it is all done

Next Week

- do Brian's edits on results
- send Brian discussion + conclusion to edit
- check on Queenie's progress and double check her normalization code
- gather all notes for inversions and gene expression paper and start to think about an outline for that paper

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.

Bacteria and Replicon	Protein Coding Sequences
<i>E. coli</i> Chromosome	$-1.43 \times 10^{-8}***$
<i>B. subtilis</i> Chromosome	$-5.55 \times 10^{-8}***$
<i>Streptomyces</i> Chromosome	$7.49 \times 10^{-8}***$
<i>S. meliloti</i> Chromosome	$-5.99 \times 10^{-7}***$
<i>S. meliloti</i> pSymA	$-5.18 \times 10^{-7}***$
<i>S. meliloti</i> pSymB	$1.67 \times 10^{-7}***$

Table 2: Logistic regression analysis of the number of substitutions along all protein coding positions of the genome of the respective bacteria replicons. Grey coloured boxes indicate a negative logistic regression coefficient estimate. All results are statistically significant. Logistic regression was calculated after the origin of replication was moved to the beginning of the genome and all subsequent positions were scaled around the origin accounting for bidirectional replication. All results are marked with significance codes as followed: $< 0.001 = '***'$, $0.001 < 0.01 = '**'$, $0.01 < 0.05 = '*'$, $> 0.05 = 'NS'$.

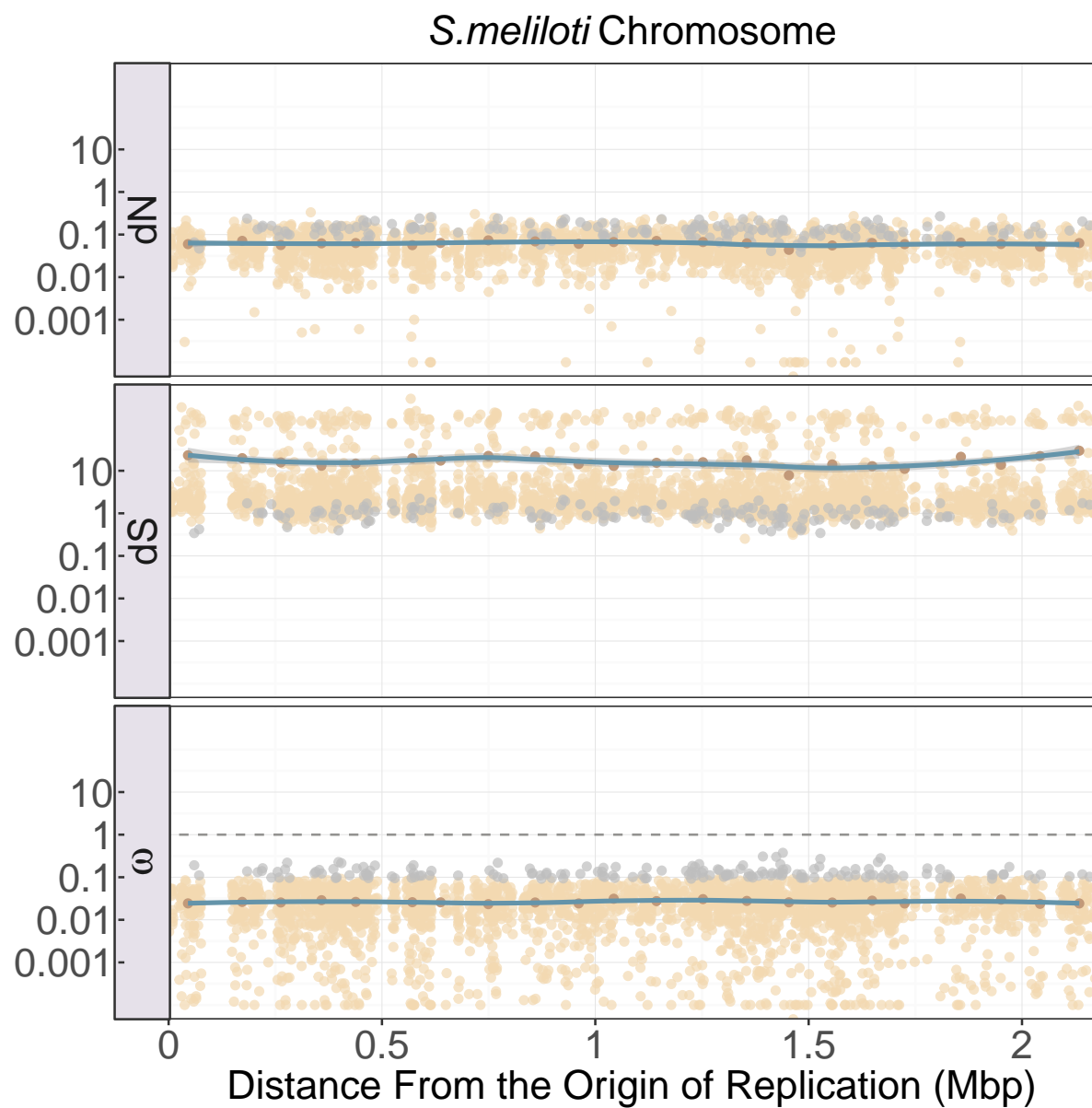
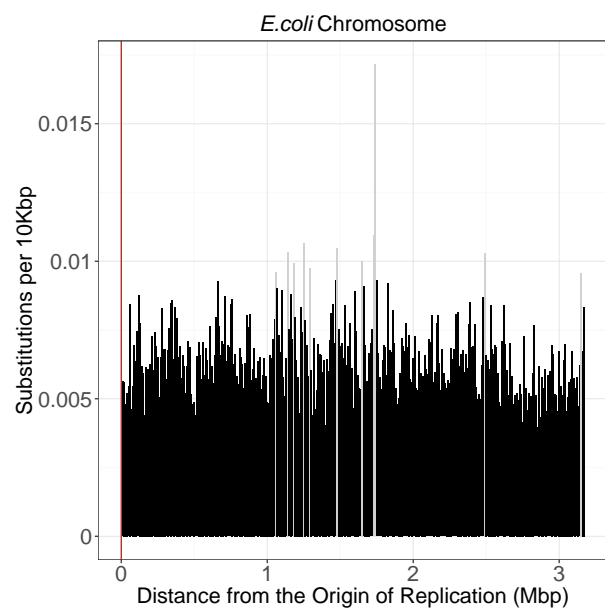
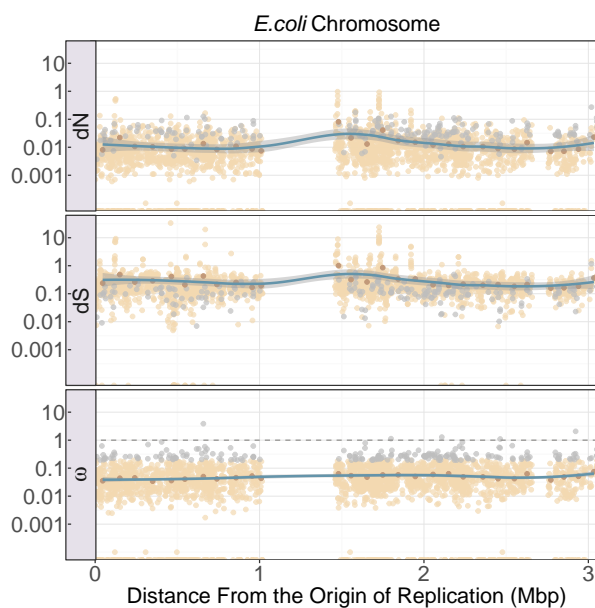


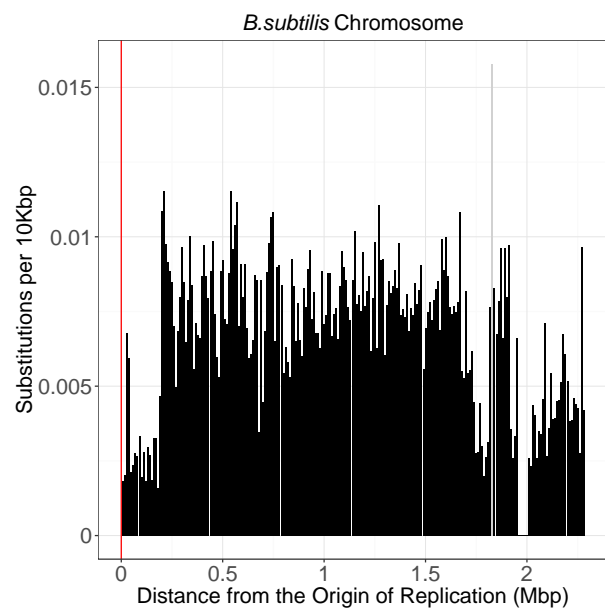
Figure 1: dN , dS , and ω values for *S. meliloti* chromosomes and *A. tumefaciens*.



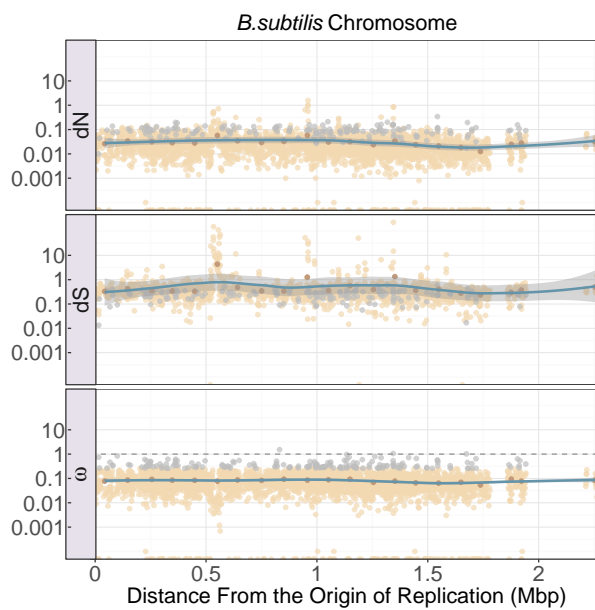
(a)



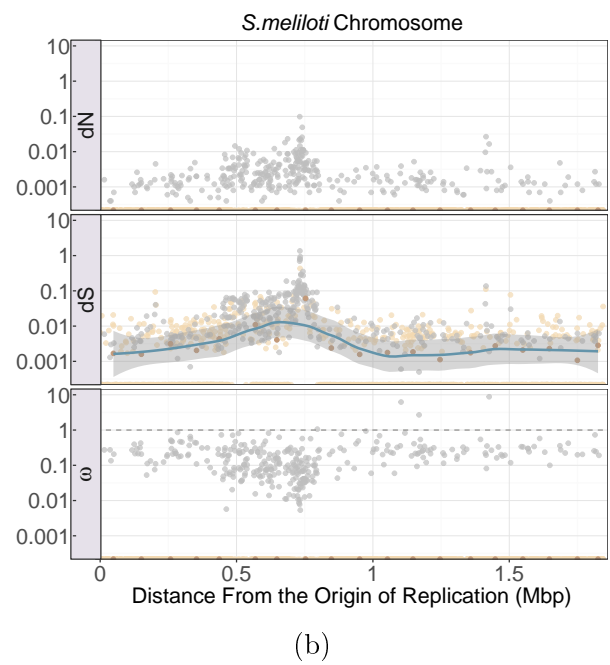
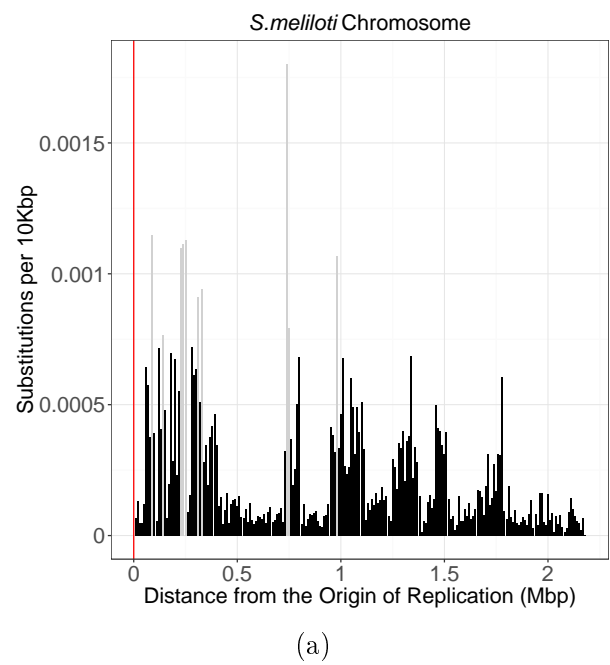
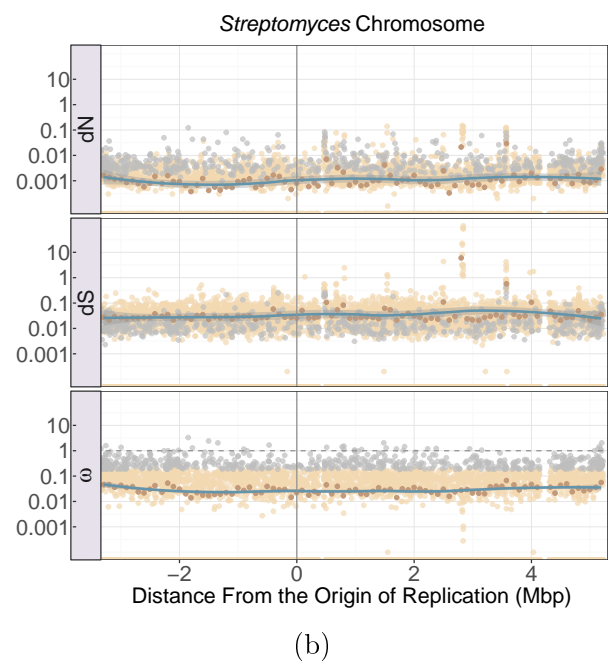
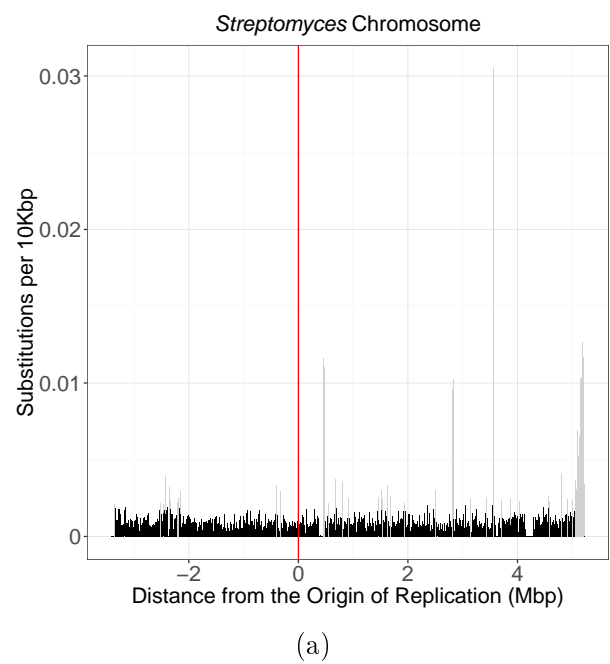
(b)

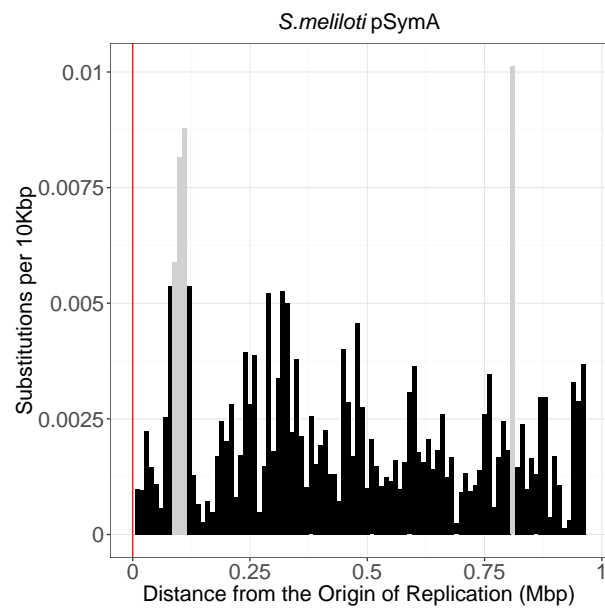


(a)

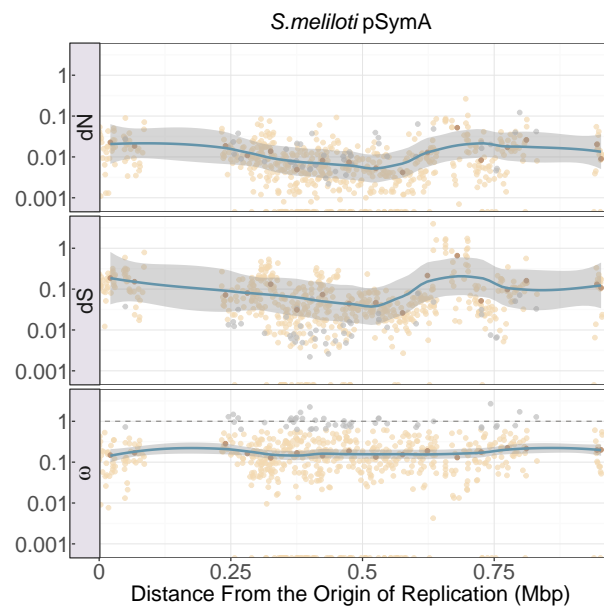


(b)

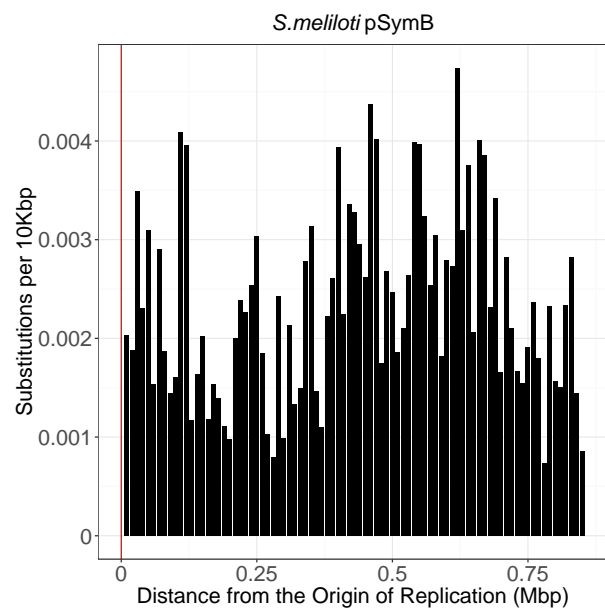




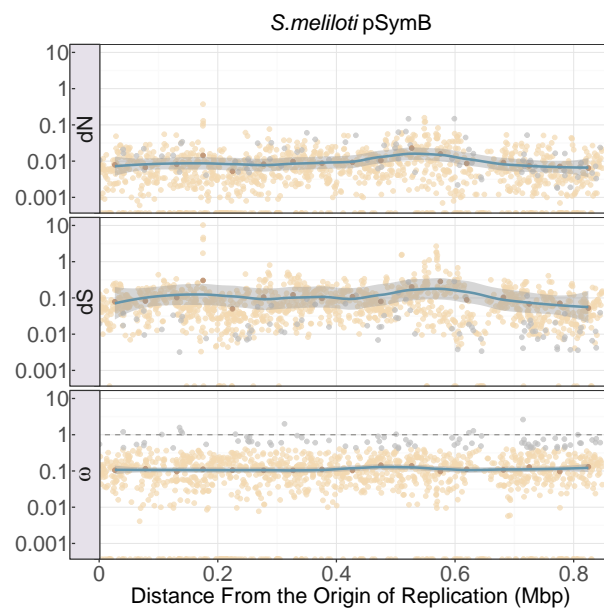
(a)



(b)



(a)



(b)

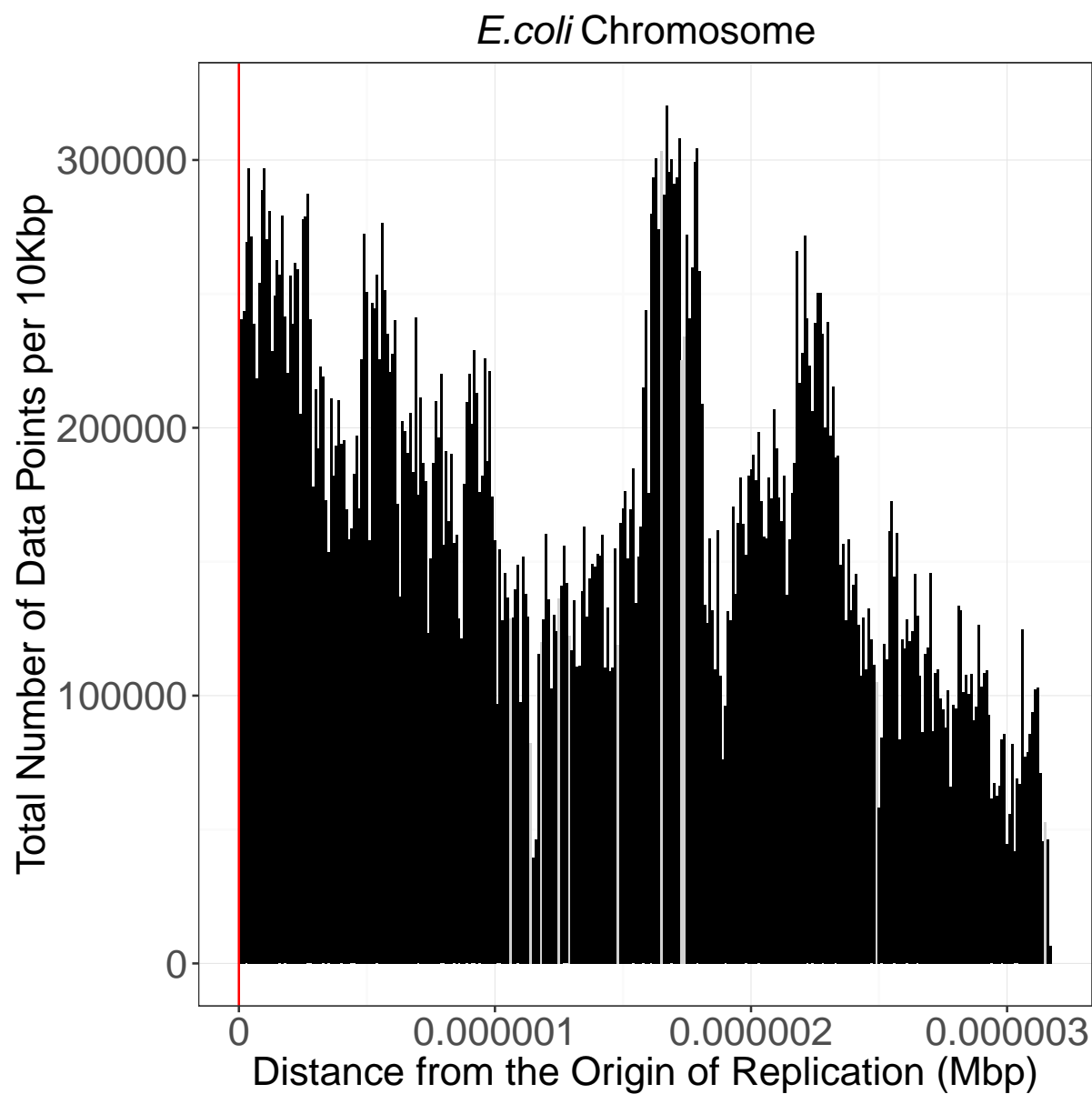


Figure 8: Distribution of total number of substitution data points per 10Kbp in genome.

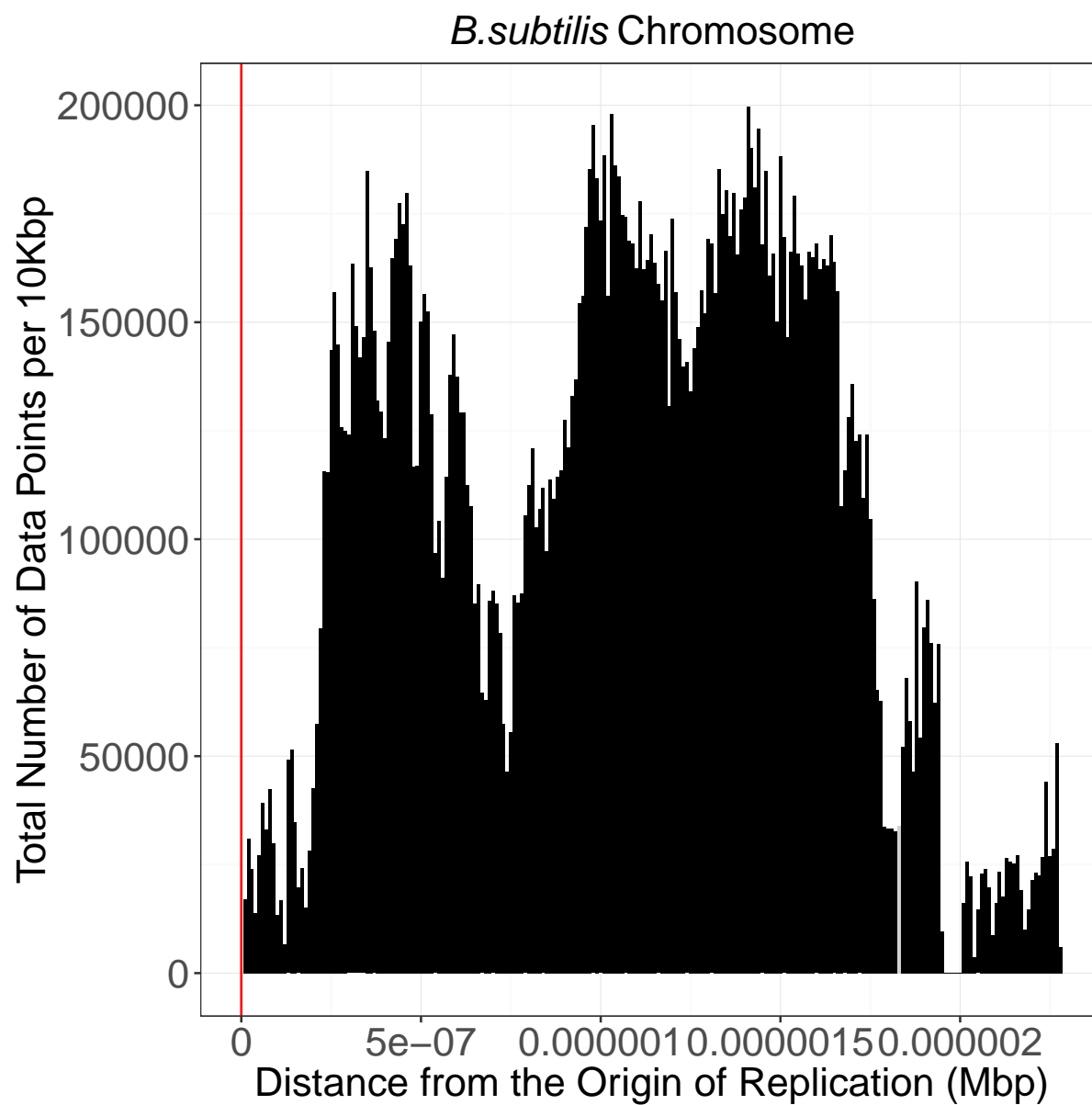


Figure 9: Distribution of total number of substitution data points per 10Kbp in genome.

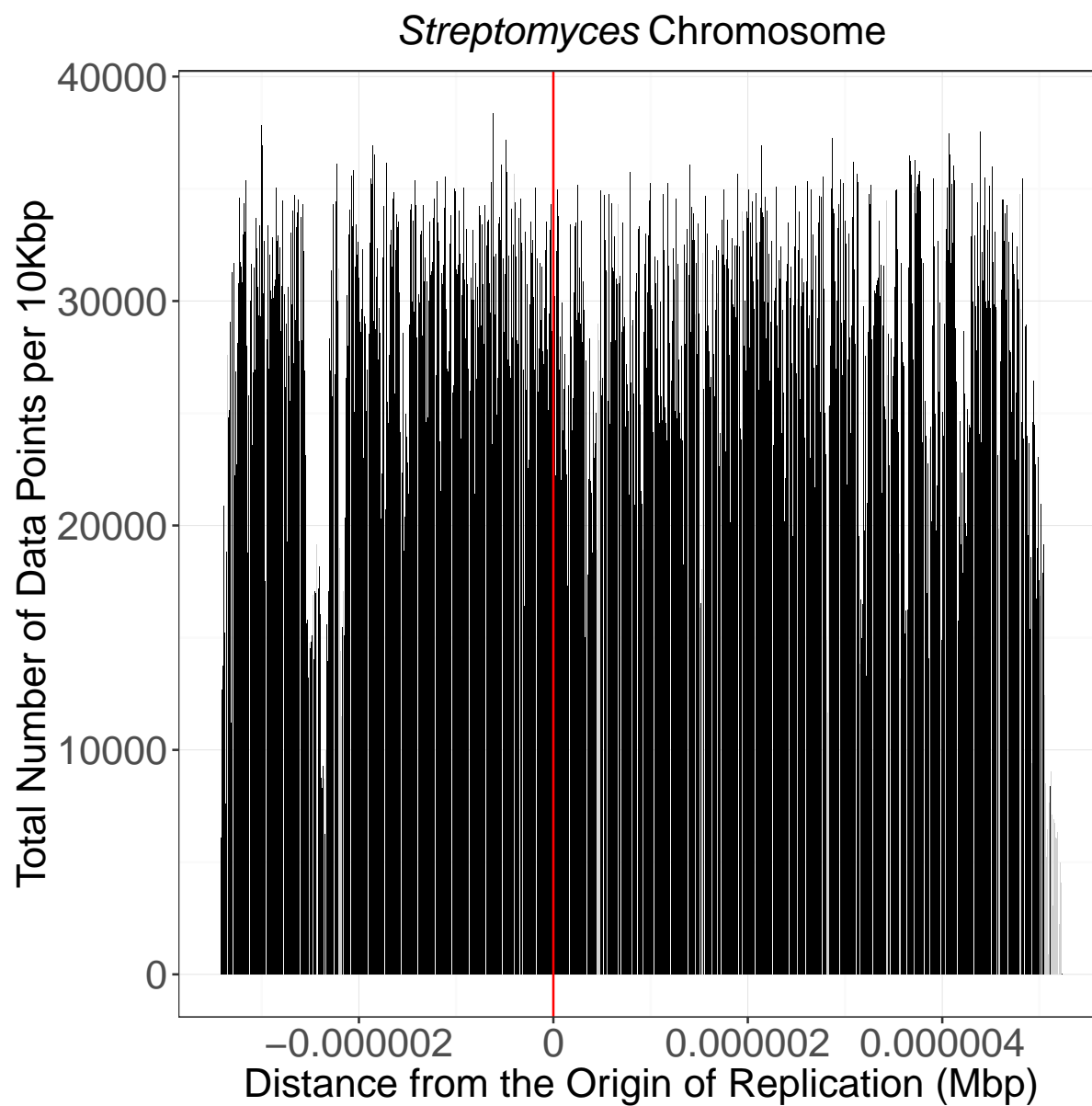


Figure 10: Distribution of total number of substitution data points per 10Kbp in genome.

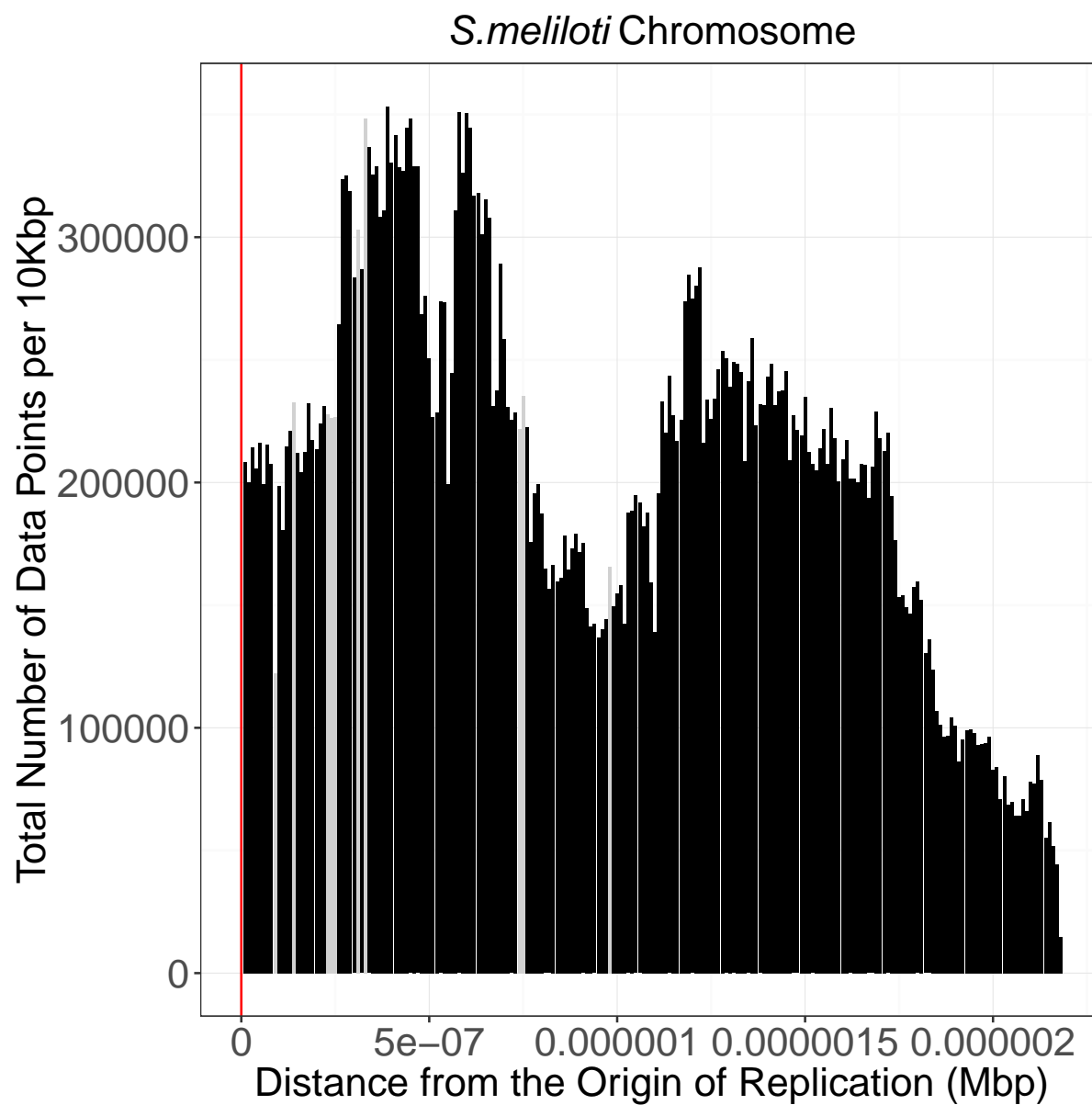


Figure 11: Distribution of total number of substitution data points per 10Kbp in genome.

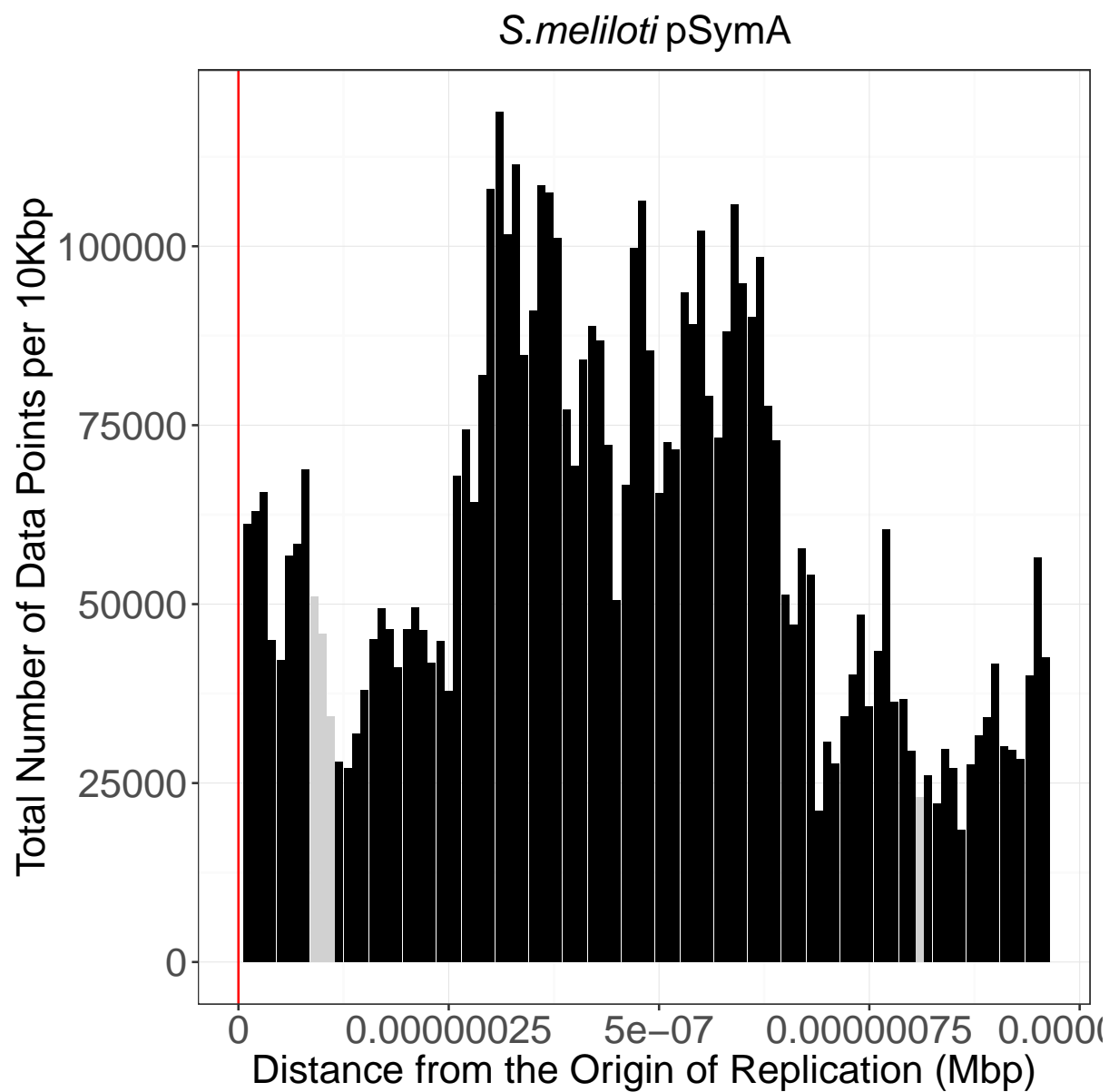


Figure 12: Distribution of total number of substitution data points per 10Kbp in genome.

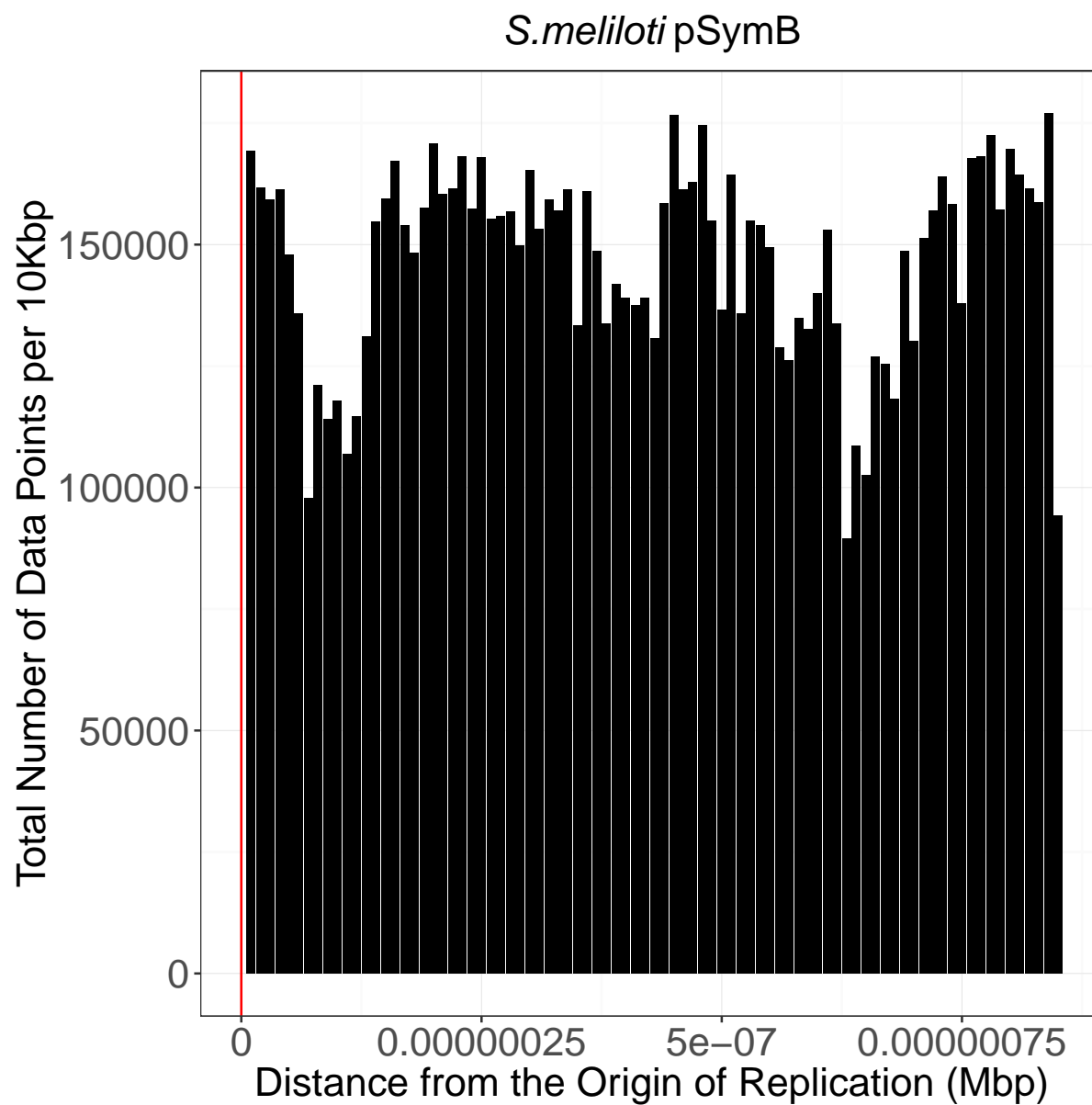


Figure 13: Distribution of total number of substitution data points per 10Kbp in genome.

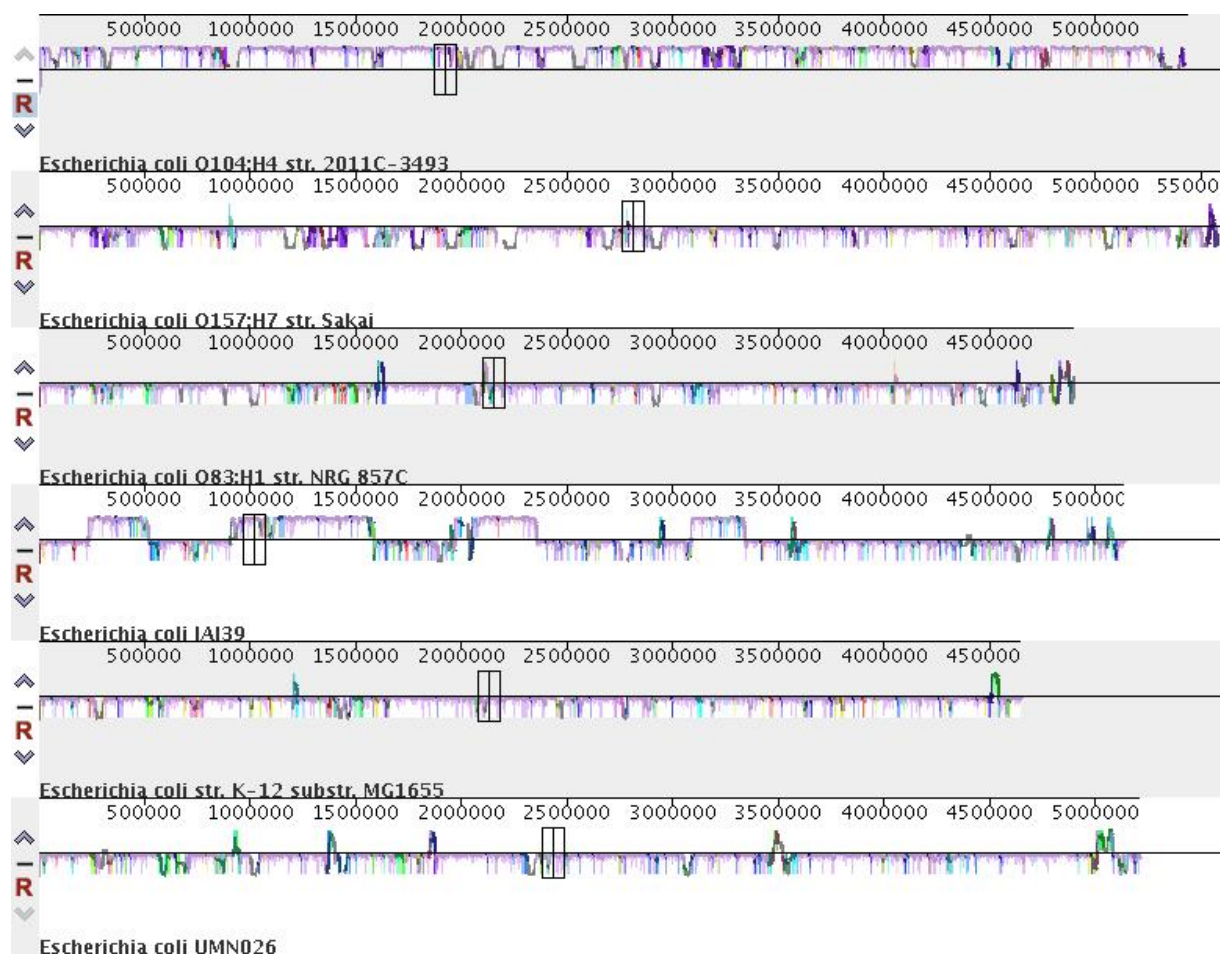


Figure 14: progressiveMauve alignment of *Escherichia coli* genomes highlighting the “backbone” of the alignment (matching regions).



Figure 15: progressiveMauve alignment of *S. meliloti* Chromosomes highlighting the “backbone” of the alignment (matching regions).

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

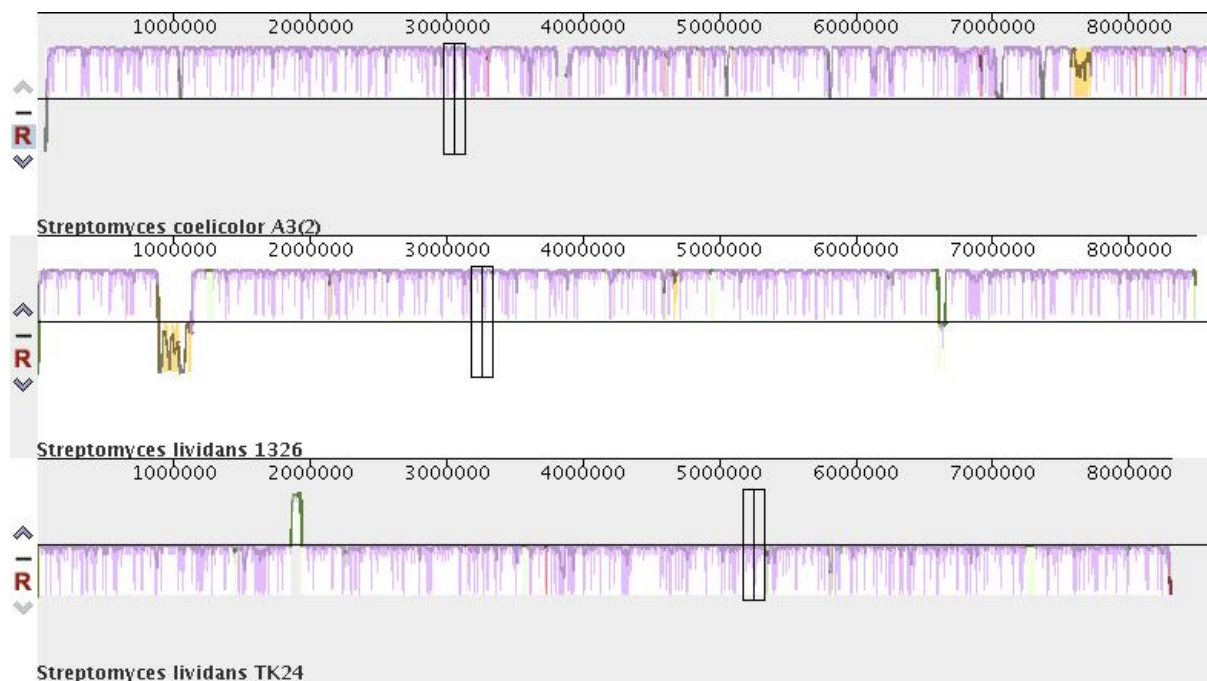


Figure 16: progressiveMauve alignment of *Streptomyces* genomes highlighting the “backbone” of the alignment (matching regions).