

- X Jan 6: Write up methods for COG paper
- ✓ May 7: Revise summer goals if accepted for Chicago Conference
- ✓ May 11: Find gene expression papers for bacteria and specific bacteria, printed
- ✓ May 25: Read above papers and make notes (one a day?)
- ✓ June 8: Have process down for testing position clustering
- ✓ June 12-29: Have first draft of ISMB presentation done (and present for the lab)/ prepare for conference questions
- ✓ July 5: Have final edits for ISMB presentation finished
- ✓ July 5-12: ISB Chicago Conference
- ✓ July 20: Have date booked for Comps
- ✓ July 20: Have all resources to read for comps topics
- July 21-Aug 10: Read all above resources
- ✓ July 21-Aug 4: Write committee report (first draft)
- July 30 - Aug 13: Work on first draft of presentation
- ✓ Aug 4: First draft of comps report due
- Aug 15: Finish reading all resources for comps
- Aug 13: Submit Comps Report to Committee
- Aug 13: First draft of presentation due
- Aug 18: Second draft of presentation due
- Aug 21: Comprehensive Exam 10:30am
- Aug 21: Have all clustering testing complete for all bacteria

Aug 26: make new list of dates for goals

Aug 26: Gather gene expression data for the above mentioned *E. coli* strains

Sep 9: Think about/compile list of inversions in *E. coli* for new paper

Last Week

Last week I finished the first draft of my comps report and sent it to you. Again, I would appreciate any feedback you have on it. I was mostly working on my presentation for my comps and I think I am almost done the first draft of it. Again, I have been reading 2 articles a day and making notes on them to prepare for the comps topics.

This Week

I will be working on my presentation for my comps, and continuing to read the papers I have found related to my comps topics.

I have all the steps for the clustering analysis separate so it is taking longer to run because I have to run each step. I want to work on condensing this so there is less “checking in” that I have to do.

Next Week

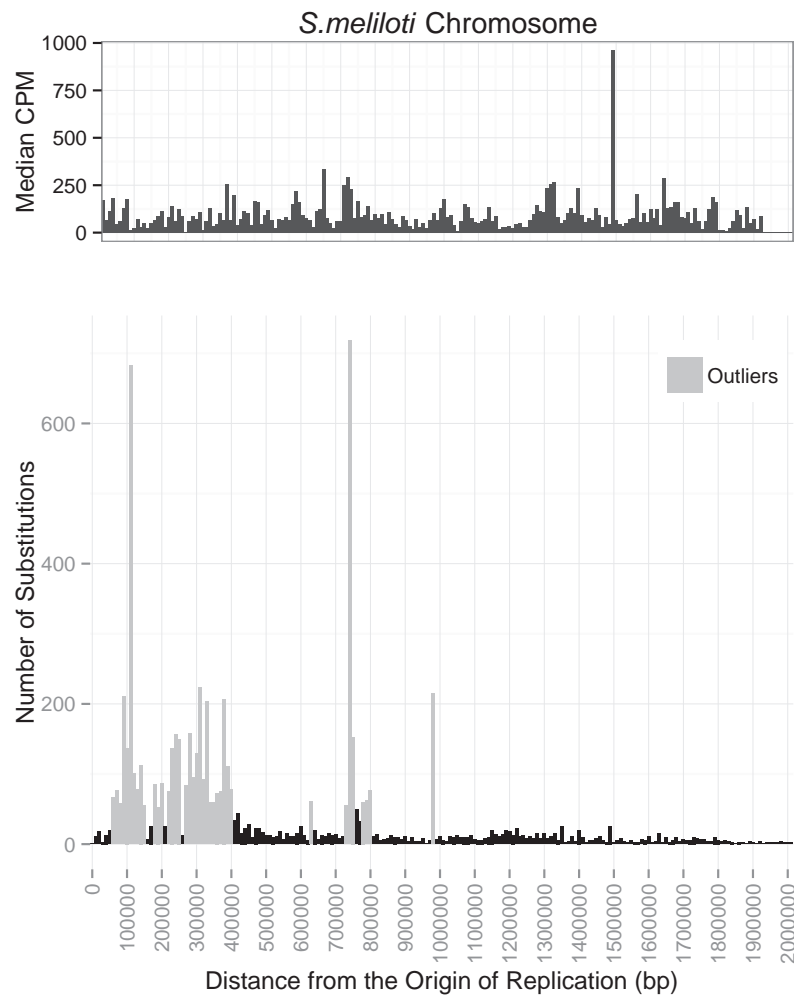
I will be continuing to read papers, edit and polish my comps report, and finish my presentation.

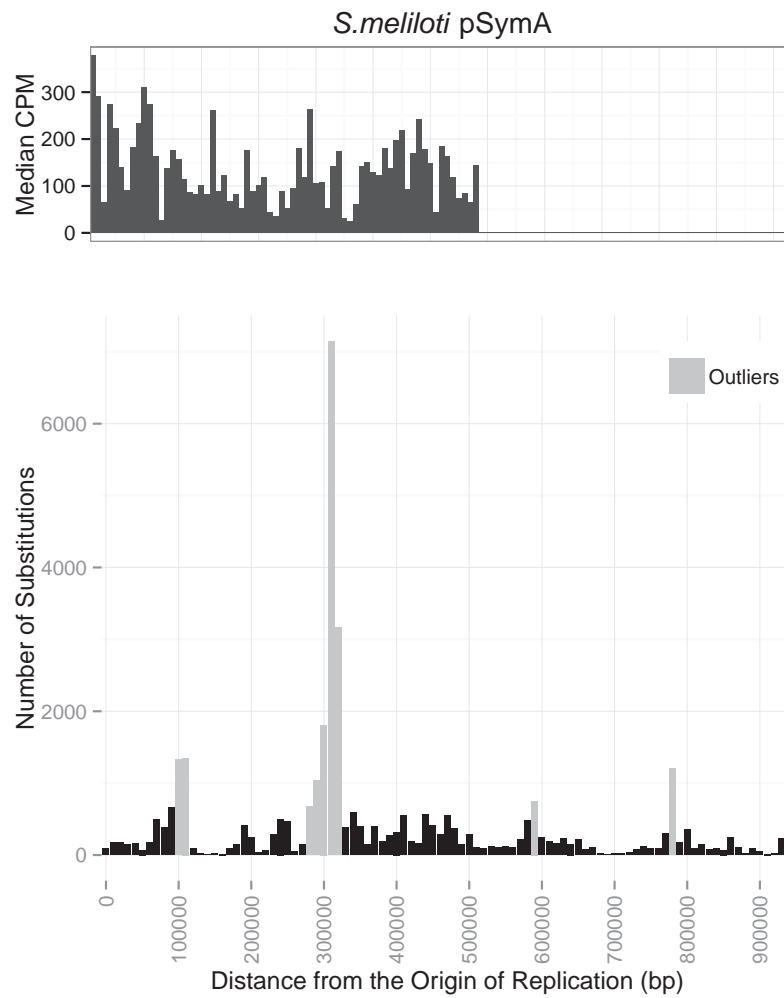
Bacteria and Replicon	Coefficient Estimate	Standard Error	P-value
<i>E. coli</i> Chromosome	-6.41×10^{-5}	1.65×10^{-5}	1.1×10^{-4}
<i>B. subtilis</i> Chromosome	-9.9×10^{-5}	2.18×10^{-5}	6×10^{-6}
<i>Streptomyces</i> Chromosome	-1.5×10^{-6}	1.4×10^{-7}	$< 2 \times 10^{-16}$
<i>S. meliloti</i> Chromosome	3.19×10^{-5}	3.57×10^{-5}	3.7×10^{-1}
<i>S. meliloti</i> pSymA	-5.36×10^{-5}	6.34×10^{-4}	9.33×10^{-1}
<i>S. meliloti</i> pSymB	5.05×10^{-4}	2.6×10^{-4}	5.3×10^{-2}

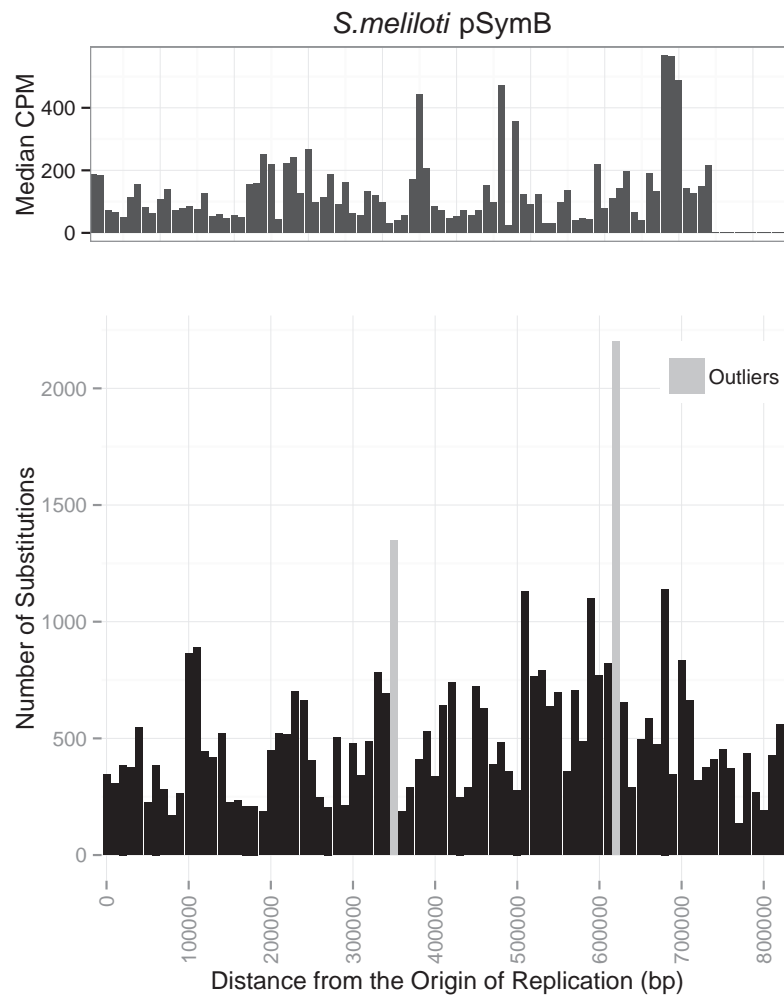
Table 1: Linear regression analysis of the median counts per million expression data along the genome of the respective bacteria replicons. Grey coloured boxes indicate statistically significant results at the 0.5 significance level. Linear regression was calculated after the origin of replication was moved to the beginning of the genome and all subsequent positions were scaled around the origin accounting for bidirectionality of replication.

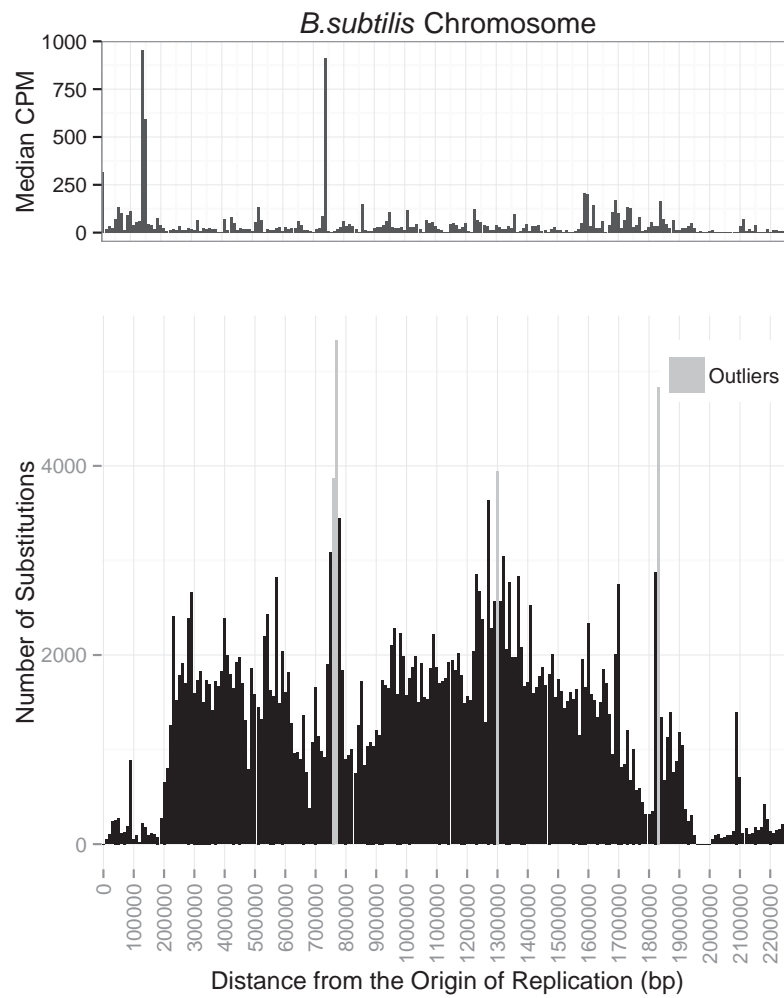
Bacteria and Replicon	Coefficient Estimate	Standard Error	P-value
<i>E. coli</i> Chromosome	-1.394×10^{-7}	2.425×10^{-9}	$< 2 \times 10^{-16}$
<i>B. subtilis</i> Chromosome	-2.538×10^{-8}	1.58×10^{-9}	$< 2 \times 10^{-16}$
<i>Streptomyces</i> Chromosome	1.736×10^{-8}	7.231×10^{-10}	$< 2 \times 10^{-16}$
<i>S. meliloti</i> Chromosome	-1.541×10^{-6}	3.042×10^{-8}	$< 2 \times 10^{-16}$
<i>S. meliloti</i> pSymA	-9.130×10^{-7}	1.975×10^{-8}	$< 2 \times 10^{-16}$
<i>S. meliloti</i> pSymB	2.488×10^{-7}	1.964×10^{-8}	$< 2 \times 10^{-16}$

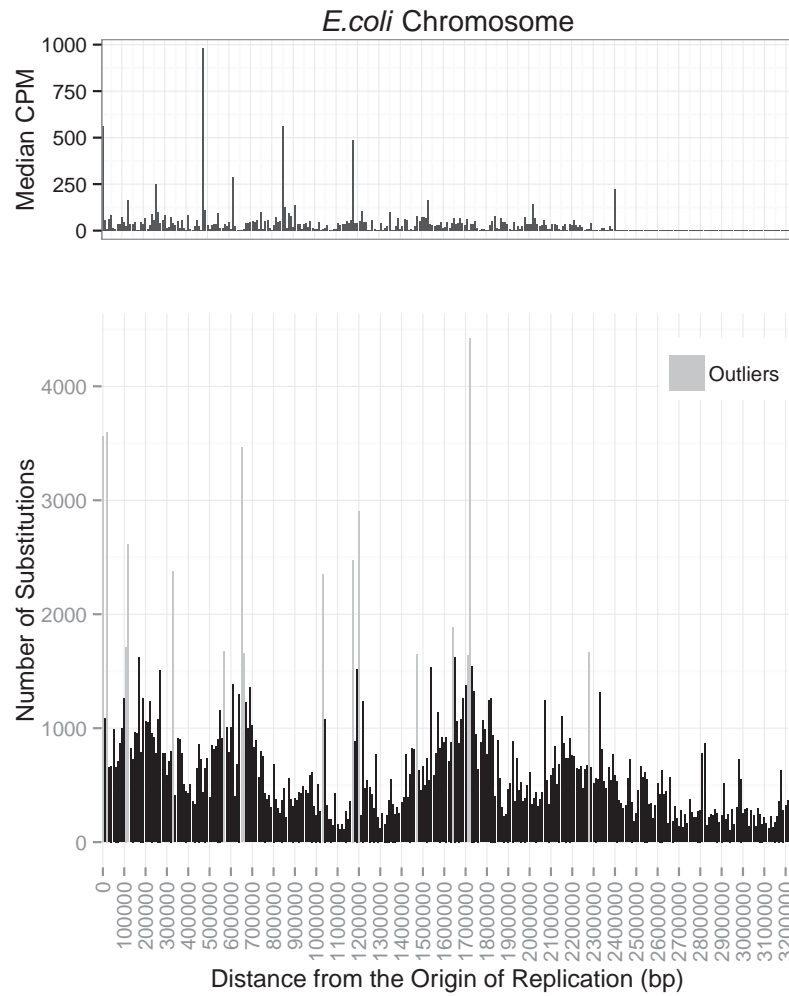
Table 2: Logistic regression analysis of the number of substitutions along the genome of the respective bacteria replicons. Grey coloured boxes indicate a negative logistic regression coefficient estimate. All results are statistically significant. Logistic regression was calculated after the origin of replication was moved to the beginning of the genome and all subsequent positions were scaled around the origin accounting for bidirectionality of replication.

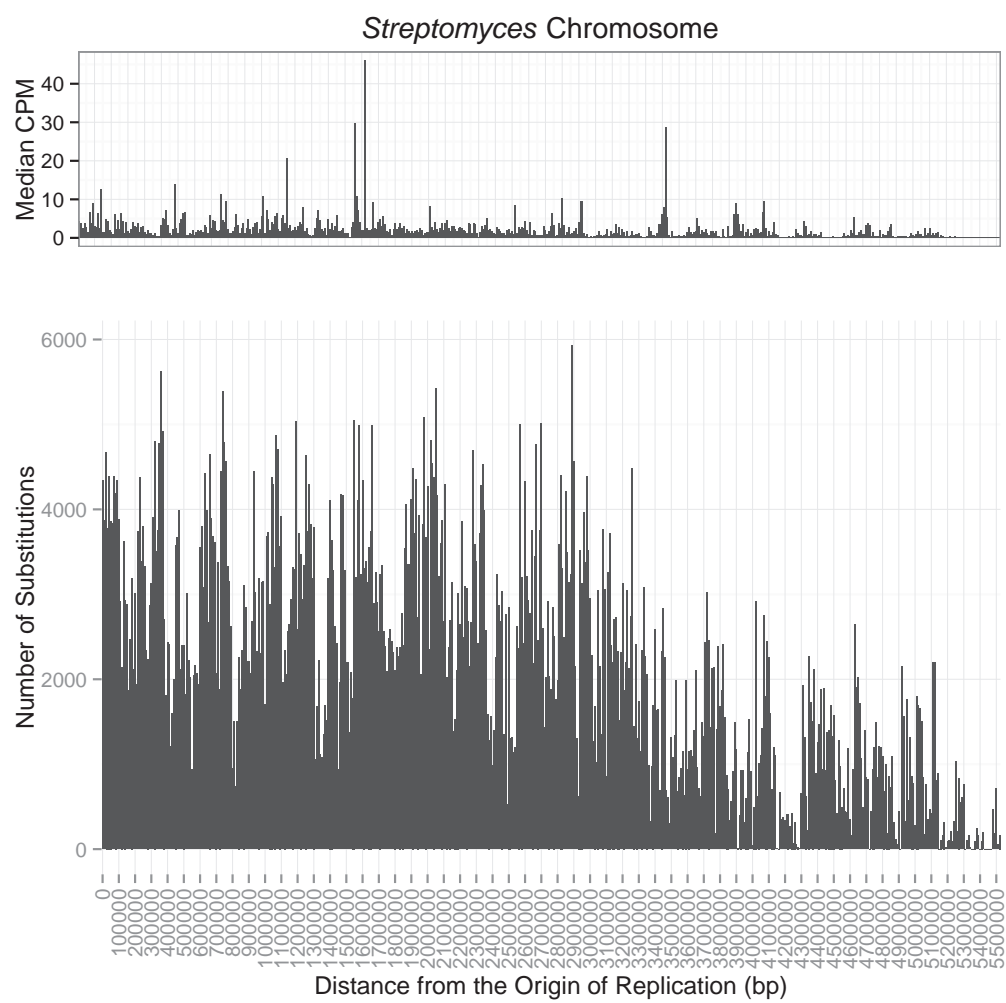












Origin Location	<i>E. coli</i> Chromosome	<i>B. subtilis</i> Chromosome	<i>Streptomyces</i> Chromosome	<i>S. meliloti</i> Chromosome	<i>S. meliloti</i> pSymA	<i>S. meliloti</i> pSymB
Moved 100kb Left	$-1.445 \times 10^{-7***}$	$4.374 \times 10^{-9*}$	$6.909 \times 10^{-9***}$	$-1.316 \times 10^{-6***}$	$-1.058 \times 10^{-6***}$	$-2.009 \times 10^{-7***}$
Moved 90kb Left	$-1.544 \times 10^{-7***}$	$-1.036 \times 10^{-7***}$	$5.677 \times 10^{-9***}$	$-1.32 \times 10^{-6***}$	$-1.246 \times 10^{-6***}$	$-1.357 \times 10^{-7***}$
Moved 80kb Left	$-1.65 \times 10^{-7***}$	$-1.072 \times 10^{-7***}$	$8.11 \times 10^{-9***}$	$-1.338 \times 10^{-6***}$	$-1.398 \times 10^{-6***}$	$-6.57 \times 10^{-8***}$
Moved 70kb Left	$-1.667 \times 10^{-7***}$	$-1.102 \times 10^{-7***}$	$6.716 \times 10^{-9***}$	$-1.363 \times 10^{-6***}$	$-1.405 \times 10^{-6***}$	9.83×10^{-8}
Moved 60kb Left	$-1.64 \times 10^{-7***}$	$-1.19 \times 10^{-7***}$	$8.7 \times 10^{-9***}$	$-1.324 \times 10^{-6***}$	$-1.394 \times 10^{-6***}$	$1.129 \times 10^{-7***}$
Moved 50kb Left	$-1.446 \times 10^{-7***}$	$-1.211 \times 10^{-7***}$	$1.045 \times 10^{-8***}$	$-1.36 \times 10^{-6***}$	$-1.403 \times 10^{-6***}$	$1.521 \times 10^{-7***}$
Moved 40kb Left	$-1.4 \times 10^{-7***}$	$-1.299 \times 10^{-7***}$	$1.214 \times 10^{-8***}$	$-1.255 \times 10^{-6***}$	$-1.422 \times 10^{-6***}$	$1.543 \times 10^{-7***}$
Moved 30kb Left	$-1.498 \times 10^{-7***}$	$-1.292 \times 10^{-7***}$	$1.24 \times 10^{-8***}$	$-1.26 \times 10^{-6***}$	$-1.392 \times 10^{-6***}$	$1.63 \times 10^{-7***}$
Moved 20kb Left	$-1.51 \times 10^{-7***}$	$-1.1 \times 10^{-7***}$	$1.395 \times 10^{-8***}$	$-1.525 \times 10^{-6***}$	$-1.412 \times 10^{-6***}$	$1.603 \times 10^{-7***}$
Moved 10kb Left	$-1.262 \times 10^{-7***}$	-2.602×10^{-9}	$1.563 \times 10^{-8***}$	$-1.599 \times 10^{-6***}$	$-9.499 \times 10^{-7***}$	$2.973 \times 10^{-7***}$
Moved 10kb Right	$-1.305 \times 10^{-7***}$	$-2.045 \times 10^{-8***}$	$1.578 \times 10^{-8***}$	$1.614 \times 10^{-6***}$	$-1.026 \times 10^{-6***}$	$3.505 \times 10^{-7***}$
Moved 20kb Right	$-1.454 \times 10^{-7***}$	$-1.006 \times 10^{-7***}$	$1.903 \times 10^{-8***}$	$-1.634 \times 10^{-6***}$	$-1.475 \times 10^{-6***}$	$1.649 \times 10^{-7***}$
Moved 30kb Right	$-1.548 \times 10^{-7***}$	$-8.596 \times 10^{-8***}$	$2.046 \times 10^{-8***}$	$-1.698 \times 10^{-6***}$	$-1.417 \times 10^{-6***}$	$1.526 \times 10^{-7***}$
Moved 40kb Right	$-1.632 \times 10^{-7***}$	$-8.378 \times 10^{-8***}$	$2.125 \times 10^{-8***}$	$-1.719 \times 10^{-6***}$	$-1.367 \times 10^{-6***}$	$1.589 \times 10^{-7***}$
Moved 50kb Right	$-1.856 \times 10^{-7***}$	$-7.879 \times 10^{-8***}$	$1.957 \times 10^{-8***}$	$-1.735 \times 10^{-6***}$	$-1.277 \times 10^{-6***}$	$1.654 \times 10^{-7***}$
Moved 60kb Right	$-1.91 \times 10^{-7***}$	$-6.98 \times 10^{-8***}$	$1.974 \times 10^{-8***}$	$-1.788 \times 10^{-6***}$	$-1.169 \times 10^{-6***}$	$1.645 \times 10^{-7***}$
Moved 70kb Right	$-1.892 \times 10^{-7***}$	$-6.634 \times 10^{-8***}$	$1.934 \times 10^{-8***}$	$-1.854 \times 10^{-6***}$	$-1.059 \times 10^{-6***}$	$1.843 \times 10^{-7***}$
Moved 80kb Right	$-1.879 \times 10^{-7**}$	$-5.814 \times 10^{-8***}$	$2.313 \times 10^{-8***}$	$-1.891 \times 10^{-6***}$	$-9.07 \times 10^{-7***}$	$1.90 \times 10^{-7***}$
Moved 90kb Right	$-1.862 \times 10^{-7***}$	$-4.314 \times 10^{-8***}$	$2.304 \times 10^{-8***}$	$-1.865 \times 10^{-6***}$	$-7.171 \times 10^{-7***}$	$2.415 \times 10^{-7***}$
Moved 100kb Right	$-1.799 \times 10^{-7***}$	$-2.597 \times 10^{-8***}$	$1.945 \times 10^{-8***}$	$-1.525 \times 10^{-6***}$	$-6.572 \times 10^{-7***}$	$3.095 \times 10^{-7***}$

Table 3: Logistic regression analysis of the number of substitutions along the genome of the respective bacteria replicons. All results are marked with significance codes as followed: $< 0.001 = '***'$, $0.001 < 0.01 = '**'$, $0.01 < 0.05 = '*'$, $0.05 < 0.1 = '.'$, $> 0.1 = ''$. 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 bidirectionality of replication.