Further supplemental information and code are available on GitHub at www.github.com/dlato/Spatial_Patterns_of_Substitutions.

Sequences

Bacteria Strain/Species	Accession Number	Date Accessed
Escherichia coli		
E. coli 0104H4	CP003289	September 29, 2016
E. coli 0157H7	BA000007	September 29, 2016
E. coli 083H1	CP001855	September 29, 2016
E. coli IAI39	CU928164	September 26, 2016
E. coli K12	U00096	September 26, 2016
E. coli UMN026	CU928163	September 26, 2016
Outgroup: *Salmonella enterica* LT2	AE006468	September 29, 2016
Bacillus subtilis		
B. subtilis 168	NC_000964	November 10, 2016
B. subtilis BS38	NZ_CP017314	November 11, 2016
B. subtilis BSn5	NC_014976	November 11, 2016
B. subtilis PY79	NC_022898	November 11, 2016
B. subtilis QB928	NC_018520	November 11, 2016
B. subtilis RONN1	NC_017195	November 11, 2016
B. subtilis W23	NC_014479	November 11, 2016
Outgroup: *Listeria monocytogenes* EDGe	NC_003210	November 11, 2016
Streptomyces		
S. venezuelae B65442	NC_CP018074	April 28, 2017
S. venezuelae 15439	NC_CP013129	April 28, 2017
S. venezuelae 10712	NC_018750	April 28, 2017
S. lividans TK24	NZ_GG657756	April 28, 2017
S. lividans 1362	NZ_CM001889	April 28, 2017
S. coelicolor A3	AL645882	November 30, 2016
Outgroup: *Mycobacterium tuberculosis* H37Rv	NC_000962	November 30, 2016
S. meliloti Chromosome		
S. meliloti 2011	NC_020528	April 24, 2017
S. meliloti 1021	NC_003047	June 3, 2014
S. meliloti AK83	NC_015590	June 3, 2014
S. meliloti BL225C	NC_017322	June 3, 2014
S. meliloti SM11	NC_017325	June 3, 2014
S. meliloti RMO17	NC_CP009144	April 24, 2017
Outgroup: *Agrobacterium tumefaciens* C58 chromosome	AE007869	December 19, 2015
S. meliloti pSymA		
S. meliloti 2011	NC_020527	April 24, 2017
S. meliloti 1021	NC_003037	June 3, 2014
S. meliloti AK83	NC_015591	June 3, 2014
S. meliloti BL225C	NC_017324	June 3, 2014
S. meliloti SM11	NC_017327	June 3, 2014
S. meliloti RMO17	NC_CP009145	April 24, 2017
Outgroup: *Agrobacterium tumefaciens* C58 plasmid	AE007872	Jan 11, 2016
S. meliloti pSymB		
S. meliloti 2011	NC_020560	April 24, 2017
S. meliloti 1021	NC_003078	June 3, 2014
S. meliloti AK83	NC_015596	June 3, 2014
S. meliloti BL225C	NC_017323	June 3, 2014
S. meliloti SM11	NC_017326	June 3, 2014
S. meliloti RMO17	NC_CP009146	April 24, 2017
Outgroup: *Agrobacterium tumefaciens* C58 chromid	NC_003063	May 1, 2017

Table S1: Strains and species used for each replicon analysis. Accession numbers, date accessed, and outgroups for each replicon are provided.

progressiveMauve Alignment

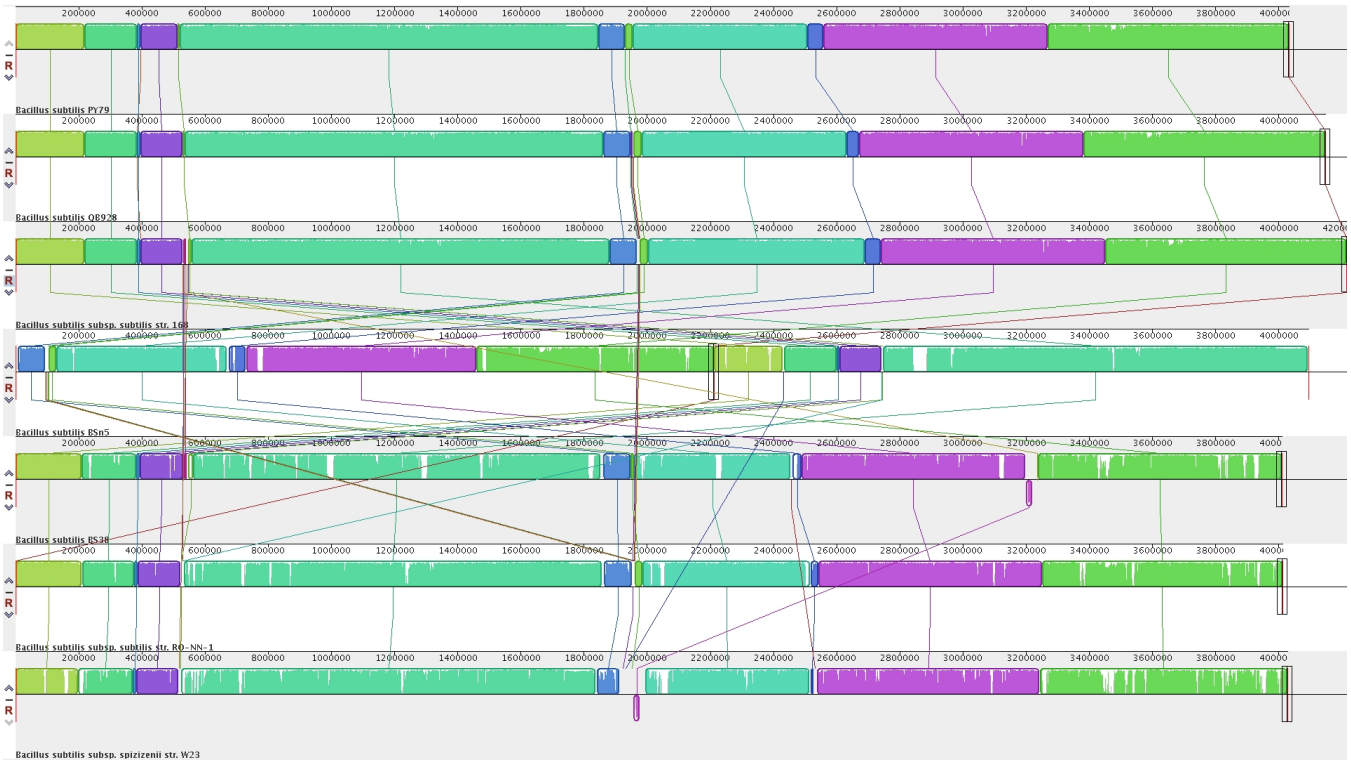


Figure S1: Visualization of the progressiveMauve alignment of the *B. subtilis* genomes. Each coloured block represents a different locally colinear block (LCB). Coloured lines connect LCBs that are similar btween taxa. The black lines underneath each LCB represent the whole genome sequence of each of the *B. subtilis* taxa. From top to bottom the taxa are: *B. subtilis* PY79, *B. subtilis* QB928, *B. subtilis* 168, *B. subtilis* BSn5, *B. subtilis* BS38, *B. subtilis* RONN1, *B. subtilis* W23. Each LCB can be treated as a rearrangement, there have therefore been 12 rearrangements between these *B. subtilis* genomes.

Phylogenetic Trees

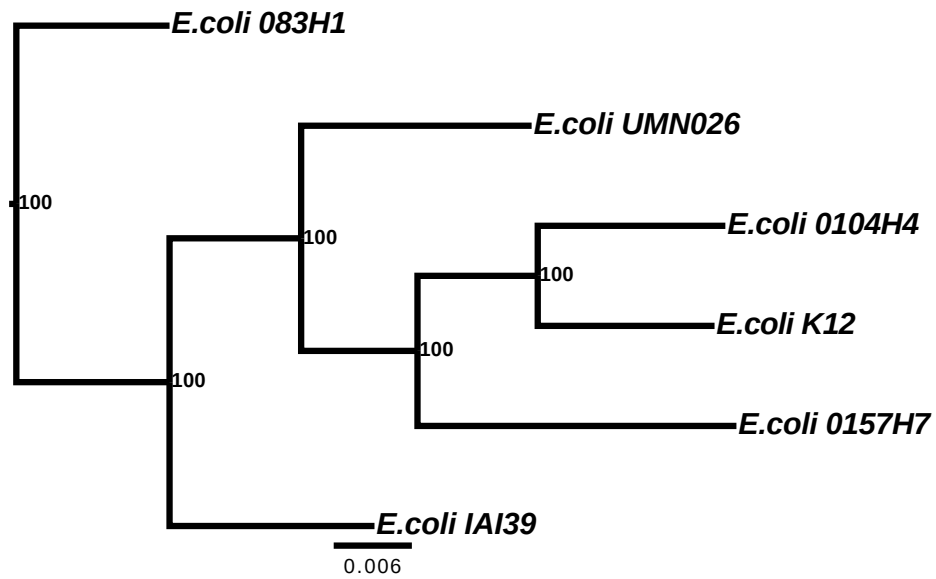


Figure S2: Phylogenetic tree of *E. coli* genomes. *Salmonella enterica* was used as an outgroup to root the tree. Branch lengths are to scale. The numbers at each node indicate the bootstrap value as a percentage. The number of bootstrapped trees was 100.

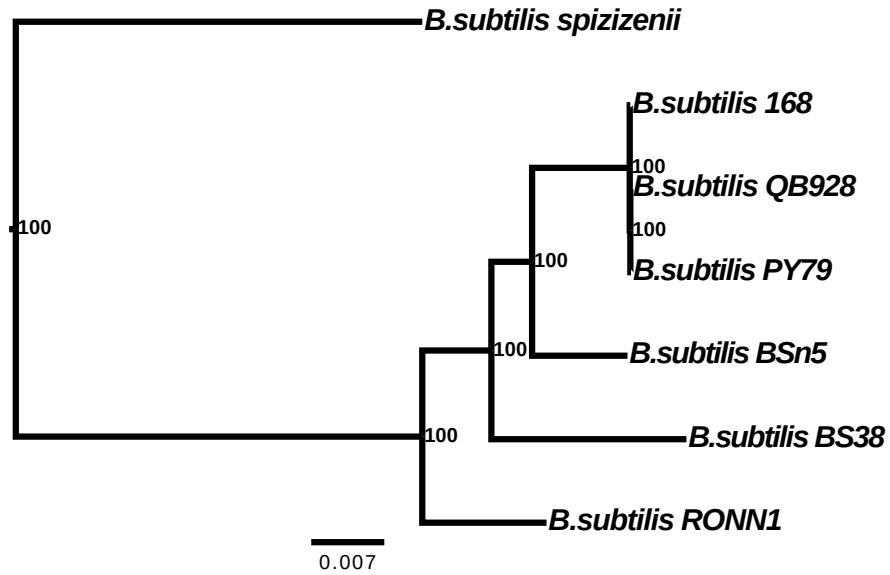


Figure S3: Phylogenetic tree of *B. subtilis* genomes. *Listeria monocytogenes* was used as an outgroup to root the tree. Branch lengths are to scale. The numbers at each node indicate the bootstrap value as a percentage. The number of bootstrapped trees was 100.

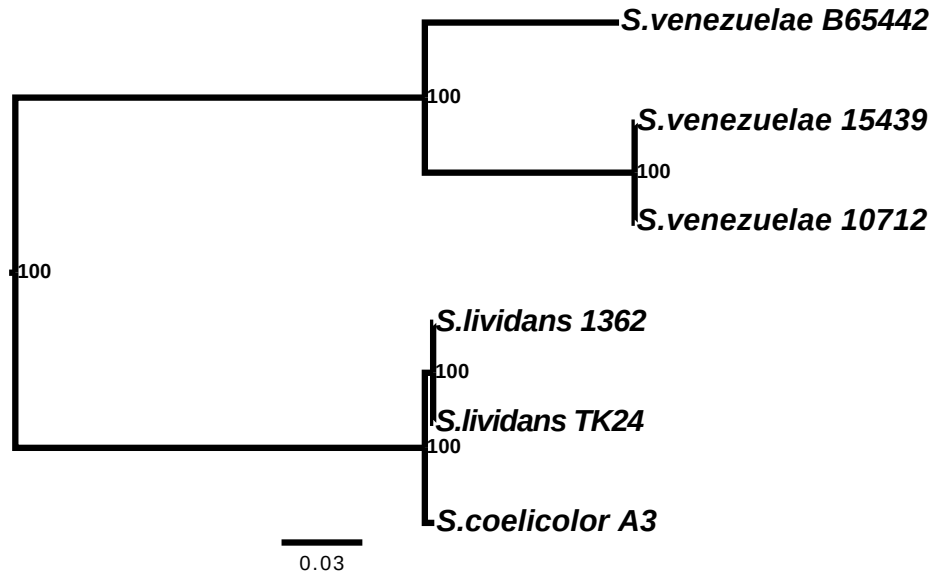


Figure S4: Phylogenetic tree of *Streptomyces* genomes. *Mycobacterium tuberculosis* was used as an outgroup to root the tree. Branch lengths are to scale. The numbers at each node indicate the bootstrap value as a percentage. The number of bootstrapped trees was 100.

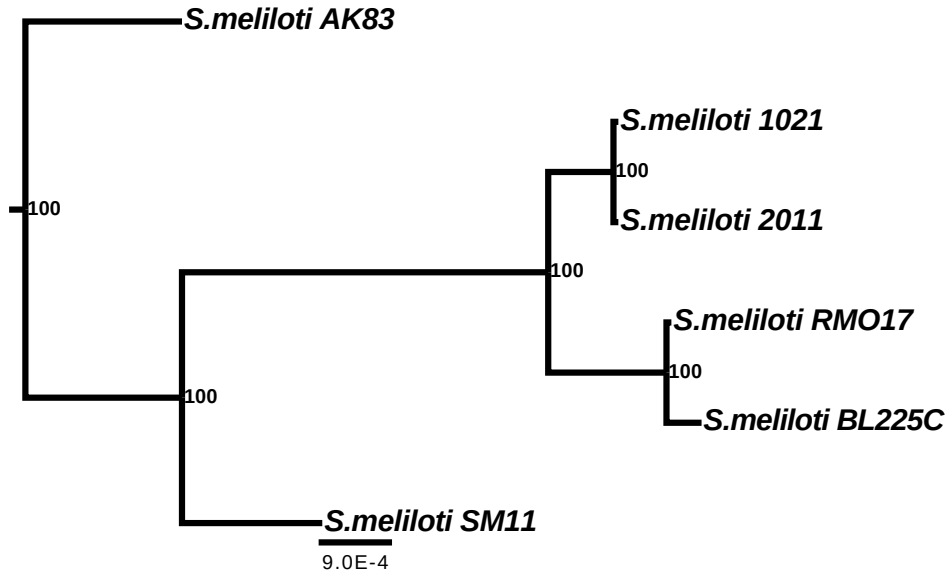


Figure S5: Phylogenetic tree using only the chromosomes of *S. meliloti*. *A. tumefaciens* circular chromosome was used as an outgroup to root the tree. Branch lengths are to scale. The numbers at each node indicate the bootstrap value as a percentage. The number of bootstrapped trees was 100.

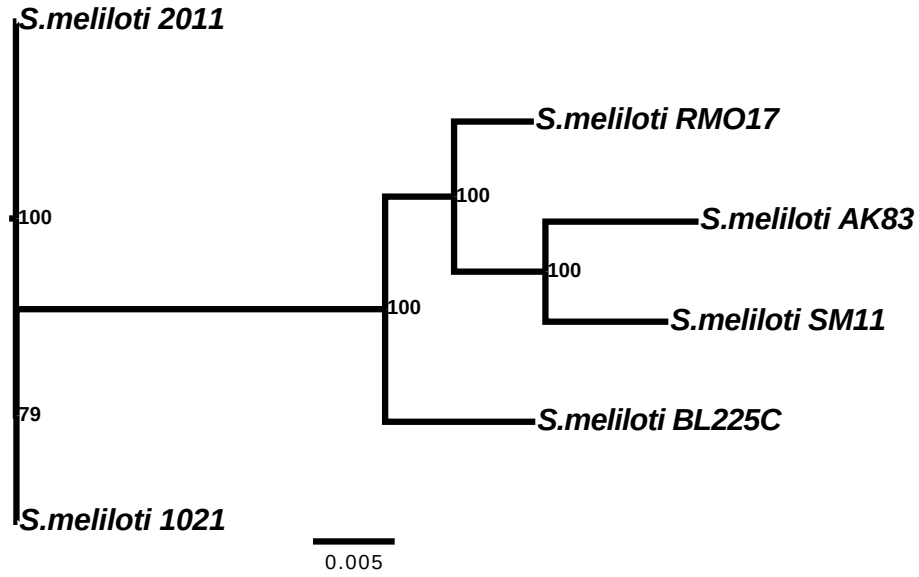


Figure S6: Phylogenetic tree using only pSymA of *S. meliloti*. *A. tumefaciens* circular plasmid was used as an outgroup to root the tree. Branch lengths are to scale. The numbers at each node indicate the bootstrap value as a percentage. The number of bootstrapped trees was 100.

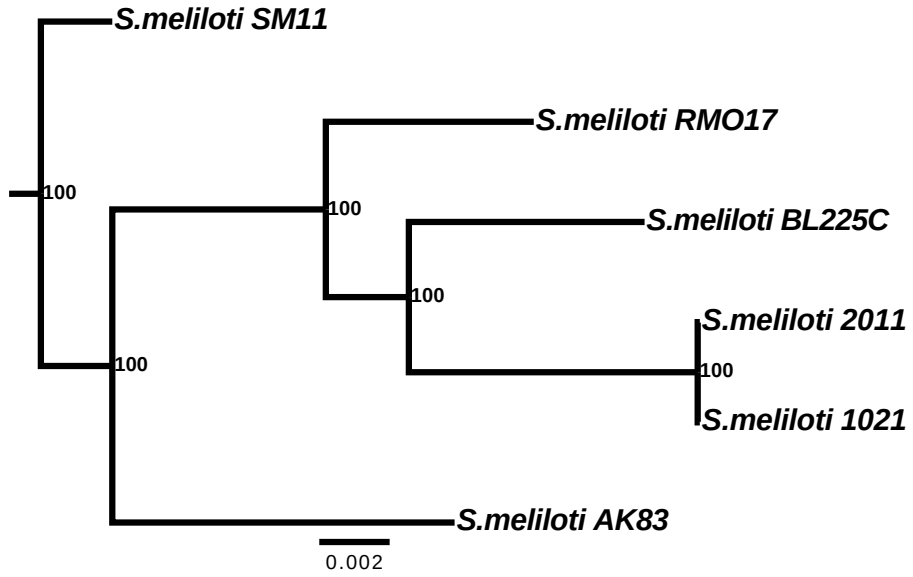


Figure S7: Phylogenetic tree using only pSymB of *S. meliloti*. *A. tumefaciens* circular chromid was used as an outgroup to root the tree. Branch lengths are to scale. The numbers at each node indicate the bootstrap value as a percentage. The number of bootstrapped trees was 100.

Bacteria Replicon	% of Total LCBs with Identical Tree	% of Total LCBs with Not Identical Tree	% of Total Alignment Discarded
E. coli Chromosome	81.58%	18.42%	25.18%
B. subtilis Chromosome	83.33%	16.67%	19.37%
Streptomyces Chromosome	96.53%	3.47%	12.42%
S. meliloti Chromosome	81.82%	18.18%	25.42%
S. meliloti pSymA	100%	0%	0%
S. meliloti pSymB	100%	0%	0%

Table S2: Proportion 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.

Origin and Terminus Locations

Bacteria	Origin of Replication	Terminus of Replication
E. coli	3925744	1678398
B. subtilis	1	1942542
Streptomyces	3419363	1 & 9054831
S. meliloti Chromosome	1	1735626
S. meliloti pSymA	1350001	672888
S. meliloti pSymB	55090	896756

Table S3: Origin of replication and terminus of replication positions in replicons of *E. coli*, *B. subtilis*, *Streptomyces*, and *S. meliloti*. The linear nature of *Streptomyces* chromosome gives it two termini, one at each end of the chromosome.

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

Table S4: Logistic regression analysis of the number of substitutions along the genome of the respective bacterial replicons after the origin location was moved by the specified increments from the original origin of replication position (listed in Table S3). 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 new location in the genome and all subsequent positions were scaled around the origin accounting for bidirectionality of replication.

Bacteria Strain	Accession Number	Date Accessed
E. coli K12 Chromosome	U00096	September 26, 2016
B. subtilis 168 Chromosome	NC_000964	November 10, 2016
S. coelicolor A3 Chromosome	AL645882	November 30, 2016
S. meliloti Chromosome 1021	NC_003047	June 3, 2014
S. meliloti pSymA 1021	NC_003037	June 3, 2014
S. meliloti pSymB 1021	NC_003078	June 3, 2014

Table S5: Strains and species used for determining the protein coding and non-protein coding regions of each bacterial replicon. GenBank reference annotation was used to determine all protein coding and non-protein coding sections of the replicons. NCBI accession numbers and date accessed are provided.

Genomic Position Clustering

A custom R script was used to cluster genomic positions together based on a user specified genetic distance using single-link clustering. For example if the user specified the genetic distance to be 10, all genomic positions within 10 base pairs would be clustered together. Consider the example where we are looking at 4 taxa with genomic positions 10, 12, 14, and 20 and the genetic clustering distance was chosen to be 2bp. Based on the clustering algorithm, positions 10, 12 and 14 would be grouped into cluster A and position 20 would be grouped into cluster B. Once the clusters were determined, a new genomic position for each of the clusters was calculated using the average of all positions within that cluster. Referring to the above mentioned example, cluster A would have a new genomic position of 12 (the average between those three positions) and cluster B would have the same genomic position of 20. The new list of genomic positions for the 4 taxa would be: 12, 12, 12 and 20. This clustering was done for genomic distances beginning at 1bp and increasing by one order of magnitude until 1,000,000bp difference exists between the taxa genomic positions. These newly clustered genomic positions were then put into the same substitution analysis as mentioned previously to determine the impact of this position clustering on the spatial substitution trends through a linear regression. A complete table of the statistical results from the clustering assessment are found in Table S6.

The results from this analysis indicate that genomic positions up to 1,000,000bp apart can be considered a singular genomic position without altering the overall spatial substitution analysis.

Position Difference	*E. coli* Chromosome	*B. subtilis* Chromosome	*Streptomyces* Chromosome	*S. meliloti* Chromosome	*S. meliloti* pSymA	*S. meliloti* pSymB
1bp	-1.394×10^{-7}**	-2.538×10^{-8}**	1.736×10^{-8}**	-1.541×10^{-6}**	-9.130×10^{-7}**	2.488×10^{-7}***
10bp	-1.394×10^{-7}***	-2.518×10^{-8}***	-4.484×10^{-9}***	-1.627×10^{-6}***	-9.13×10^{-7}***	3.487×10^{-7}***
100bp	-1.764×10^{-7}***	-1.417×10^{-8}***	1.448×10^{-8}***	-1.605×10^{-6}***	-1.166×10^{-6}***	4.021×10^{-7}***
1000bp	-1.784×10^{-7}***	-1.417×10^{-8}***	1.505×10^{-8}***	-1.605×10^{-6}***	-1.153×10^{-6}***	4.021×10^{-7}***
10000bp	-1.712×10^{-7}***	-3.496×10^{-8}***	4.790×10^{-8}***	-1.605×10^{-6}***	-3.570×10^{-8}*	3.784×10^{-7}***
100000bp	-2.061×10^{-7}***	-3.561×10^{-8}***	4.167×10^{-9}***	-1.605×10^{-6}***	-4.676×10^{-7}***	3.784×10^{-7}***
1000000bp	4.229×10^{-8}***	-7.710×10^{-9}***	6.083×10^{-8}***	-1.605×10^{-6}***	4.285×10^{-6}***	-8.888×10^{-7}***

Table S6: Results from the position clustering analysis. Logistic regression analysis of the number of substitutions along the genome of the respective bacterial replicons to test position differences. The "Position Difference" column denotes different base pair distances that the positions in the genome were clustered together as. 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 positions in the genome were determined to be the same at each position difference listed in the first column.

Bacteria and Replicon	Average Replicon Length	Number of Sites	Number of Substitutions
E. coli Chromosome	5082529	3314712	215136
B. subtilis Chromosome	4077077	4409372	206644
Streptomyces Chromosome	8497577	6657766	28891
S. meliloti Chromosome	3426881	2869482	7439
S. meliloti pSymA	1455940	627779	13666
S. meliloti pSymB	1664597	2650130	30409

Table S7: Total number of protein coding sites in each replicon for this analysis and the number of those sites that have a substitution (multiple substitutions at one site are counted as two substitutions).

	Protein Coding			
	Correlation Coefficient 20kb Near		Number of Substitutions per 20kb Near	
Bacteria and Replicon	Origin	Terminus	Origin	Terminus
E. coli Chromosome	-1.17×10^{-5}**	NS	5.91×10^{-3}	6.92×10^{-3}
B. subtilis Chromosome	NS	-8.96×10^{-5}***	1.97×10^{-3}	9.04×10^{-3}
Streptomyces Chromosome	7.45×10^{-5}***	-1.32×10^{-4}***	6.48×10^{-4}	6.73×10^{-3}
S. meliloti Chromosome	8.26×10^{-5}*	NS	9.79×10^{-5}	5.07×10^{-5}
S. meliloti pSymA	NS	NS	9.80×10^{-4}	3.24×10^{-3}
S. meliloti pSymB	-1.42×10^{-5}*	-6.32×10^{-5}***	1.97×10^{-3}	1.24×10^{-3}

Table S8: Logistic regression on 20kb closest and farthest from the origin of replication after accounting for bidirectional replication and outliers. All results are marked with significance codes as followed: < 0.001 = '***', $0.001 < 0.01$ = '**', $0.01 < 0.05$ = '*', > 0.05 = 'NS'.

	Protein Coding	
Bacteria and Replicon	Weighted	Non-Weighted
E. coli Chromosome	-2.25×10^{-10}**	-1.89×10^{-4}***
B. subtilis Chromosome	-8.35×10^{-10}**	-1.91×10^{-4}**
Streptomyces Chromosome	5.45×10^{-11}***	NS
S. meliloti Chromosome	-1.11×10^{-10}***	-1.28×10^{-5}***
S. meliloti pSymA	NS	NS
S. meliloti pSymB	NS	NS

Table S9: Linear regression on 10kb sections of the genome with increasing distance from the origin of replication after accounting for bidirectional replication. Weighted columns have the total number of substitutions in each 10kb section of the genome divided by the total number of protein coding and non-protein coding sites in the genome. Non-weighted columns are performing a linear regression on the total number of substitutions in each 10kb section of the genome. All results are marked with significance codes as followed: < 0.001 = '***', $0.001 < 0.01$ = '**', $0.01 < 0.05$ = '*', > 0.05 = 'NS'.

Bacteria and Replicon	Coefficient Estimate
E. coli Chromosome	-2.51×10^{-2}***
B. subtilis Chromosome	-2.00×10^{-2}**
Streptomyces Chromosome	-1.74×10^{-3}***
S. meliloti Chromosome	-1.88×10^{-2}***
S. meliloti pSymA	-2.50×10^{-2}**
S. meliloti pSymB	NS

Table S10: Linear regression analysis of the total number of protein coding sites per 10kb along the genome of the respective bacteria replicons. 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. All results are marked with significance codes as followed: $<$ 0.001 = '***', 0.001 $<$ 0.01 = '**', 0.01 $<$ 0.05 = '*', $>$ 0.05 = 'NS'.

Distribution of *dN*, *dS*, and ω

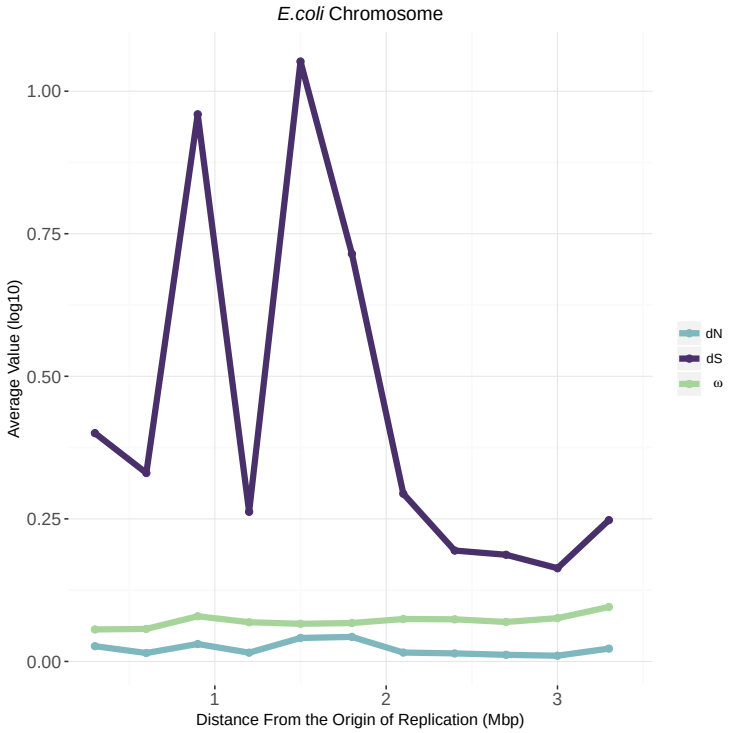
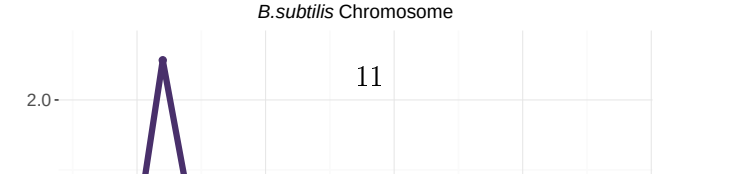


Figure S8: Distribution of all gene *dN*, *dS*, and ω values across the genome of *E. coli*. The x-axis denotes the genomic position while account for bidirectional replication. The y-axis denotes the log10 value for *dN*, *dS*, or ω.



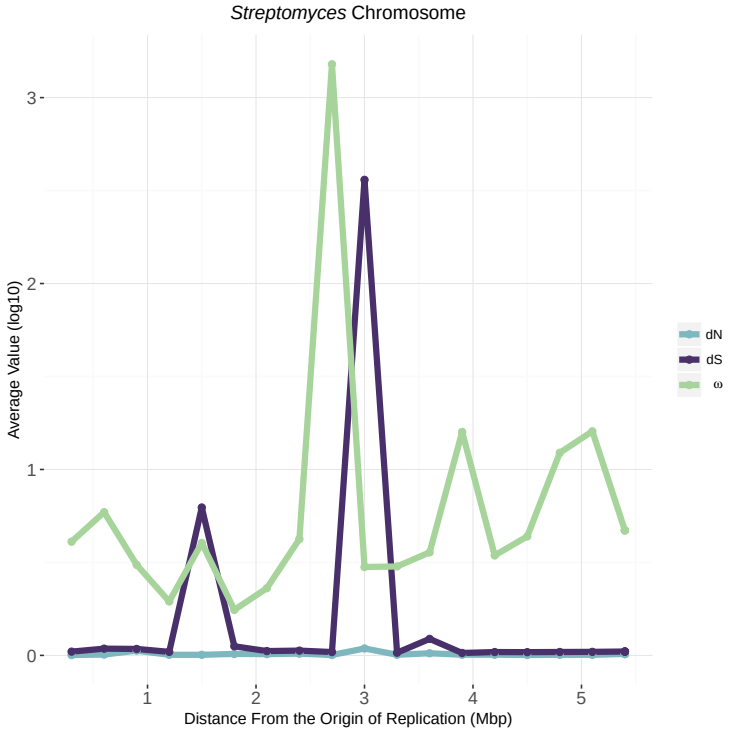


Figure S10: Distribution of all gene dN, dS, and ω values across the genome of *Streptomyces*. The x-axis denotes the genomic position while account for bidirectional replication. The y-axis denotes the log10 value for dN, dS, or ω.

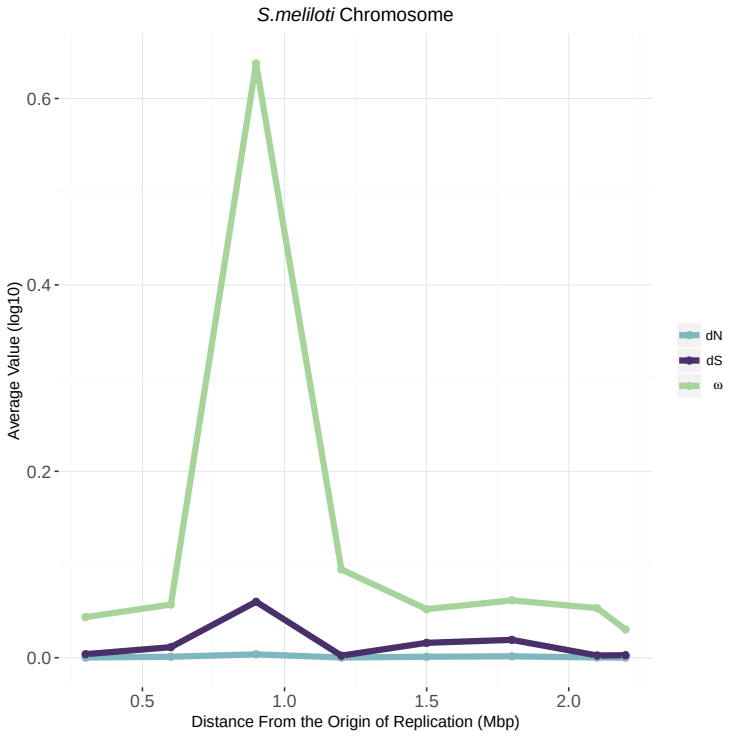


Figure S11: Distribution of all gene dN, dS, and ω values across the chromosome of *S. meliloti*. The x-axis denotes the genomic position while account for bidirectional replication. The y-axis denotes the log10 value for dN, dS, or ω.

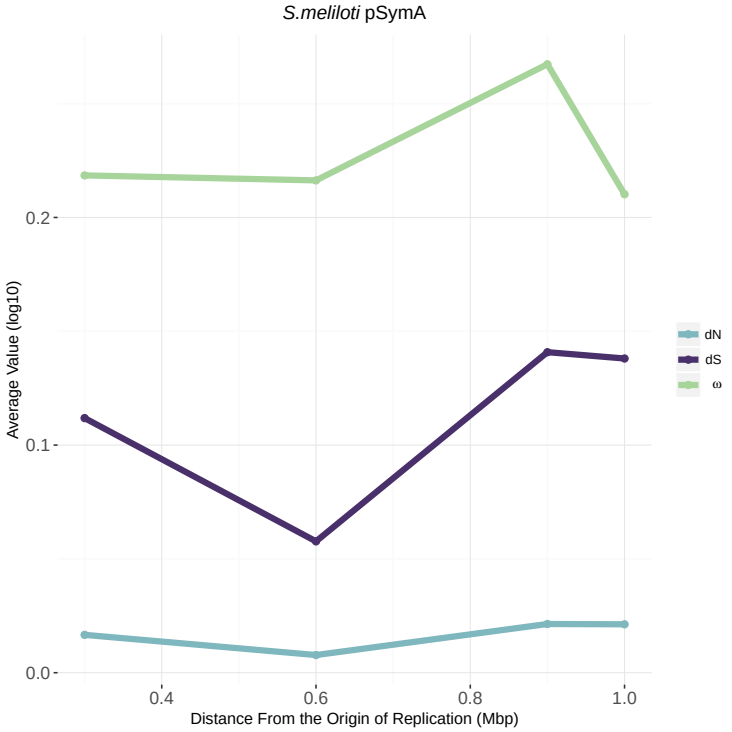


Figure S12: Distribution of all gene dN, dS, and ω values across pSymA of *S. meliloti*. The x-axis denotes the genomic position while account for bidirectional replication. The y-axis denotes the log10 value for dN, dS, or ω.

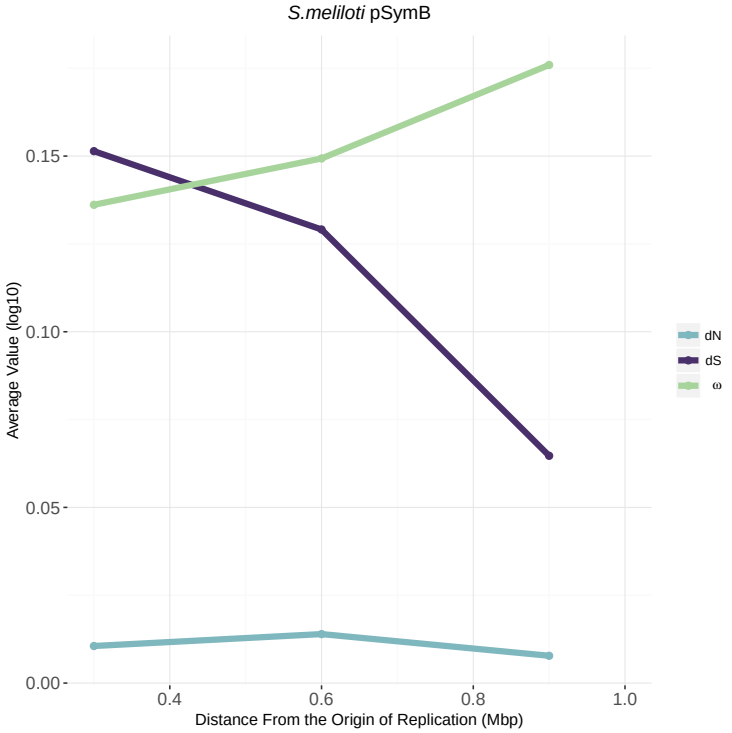


Figure S13: Distribution of all gene dN, dS, and ω values across pSymB of *S. meliloti*. The x-axis denotes the genomic position while account for bidirectional replication. The y-axis denotes the log10 value for dN, dS, or ω.

Bacteria and Replicon	*dN*	*dS*	ω
E. coli Chromosome	NS	NS	6.62×10^{-9}**
B. subtilis Chromosome	-1.31×10^{-8}**	-5.92×10^{-7}*	-9.71×10^{-9}**
Streptomyces Chromosome	NS	NS	NS
S. meliloti Chromoeom	NS	NS	NS
S. meliloti pSymA	NS	NS	NS
S. meliloti pSymB	NS	NS	6.36×10^{-8}**

Table S11: Linear regression for *dN*, *dS*, and ω calculated for each bacterial replicon on a per genome basis. All results are marked with significance codes as followed: p: < 0.001 = '***', $0.001 < 0.01$ = '**', $0.01 < 0.05$ = '*', > 0.05 = 'NS'.

	Near Origin			Near Terminus		
Bacteria and Replicon	*dN*	*dS*	ω	*dN*	*dS*	ω
E. coli Chromosome	NS	-1.08×10^{-5}*	NS	NS	NS	NS
B. subtilis Chromosome	NS	NS	NS	NS	NS	NS
Streptomyces Chromosome	NS	NS	NS	-3.11×10^{-7}*	NS	NS
S. meliloti Chromosome	NS	NS	NS	NS	NS	NS
S. meliloti pSymA	NS	NS	NS	NS	NS	-2.89×10^{-6}*
S. meliloti pSymB	NS	NS	-1.49×10^{-5}*	NS	NS	NS

Table S12: Linear regression for *dN*, *dS*, and ω calculated for each bacterial replicon for the 20 genes closest and 20 genes farthest from the origin of replication. All results are marked with significance codes as followed: p: < 0.001 = '***', $0.001 < 0.01$ = '**', $0.01 < 0.05$ = '*', > 0.05 = 'NS'.