

All supplemental information including interactive graphs of the expression data can be found on GitHub at https://github.com/dlato/Spatial_Patterns_of_Gene_Expression.git.

Gene Expression Data

Bacteria Strain/Species	GEO Accession Number	Date Accessed
<i>E. coli</i> K12 MG1655	GSE60522	December 20, 2017
<i>E. coli</i> K12 MG1655	GSE73673	December 19, 2017
<i>E. coli</i> K12 MG1655	GSE85914	December 19, 2017
<i>E. coli</i> K12 DH10B	GSE98890	December 19, 2017
<i>B. subtilis</i> 168	GSE104816	December 14, 2017
<i>B. subtilis</i> 168	GSE67058	December 16, 2017
<i>B. subtilis</i> 168	GSE93894	December 15, 2017
<i>S. coelicolor</i> A3	GSE57268	March 16, 2018
<i>S. meliloti</i> RM2010 Chromosome	GSE69880	December 12, 2017
<i>S. meliloti</i> RM2010 pSymA	GSE69880	December 12, 2017
<i>S. meliloti</i> RM2010 pSymB	GSE69880	December 12, 2017

Table S1: Strains and species used for each gene expression analysis. Gene Expression Omnibus accession numbers and date accessed are provided.

Origin and Terminus Locations

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

Table S2: 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.

Correlation of Gene Expression Over Datasets

To assess uniform expression over bacteria with multiple data sets we looked at the mean normalised expression values. Multiple replicates from a data set were combined by finding the median normalised CPM expression value for each gene. This was done for any data sets that had multiple replicates. For each gene (x_i) the mean normalised expression value was calculated across all data sets (\bar{x}_{ij}). Then the normalised median expression value for each data set was subtracted from the mean across all expression

values ($|x_{ij} - \bar{x}_{ij}|$). The distribution of these $|x_{ij} - \bar{x}_{ij}|$ across all genes are found in Figures S1 and S2. All data sets are well mixed, implying that the expression levels are consistent across all data sets. Only *E. coli* and *B. subtilis* had multiple expression datasets available so they are the only ones that were analysed. *Streptomyces* and all replicons of *S. meliloti* had only one dataset each and therefore were not analysed.

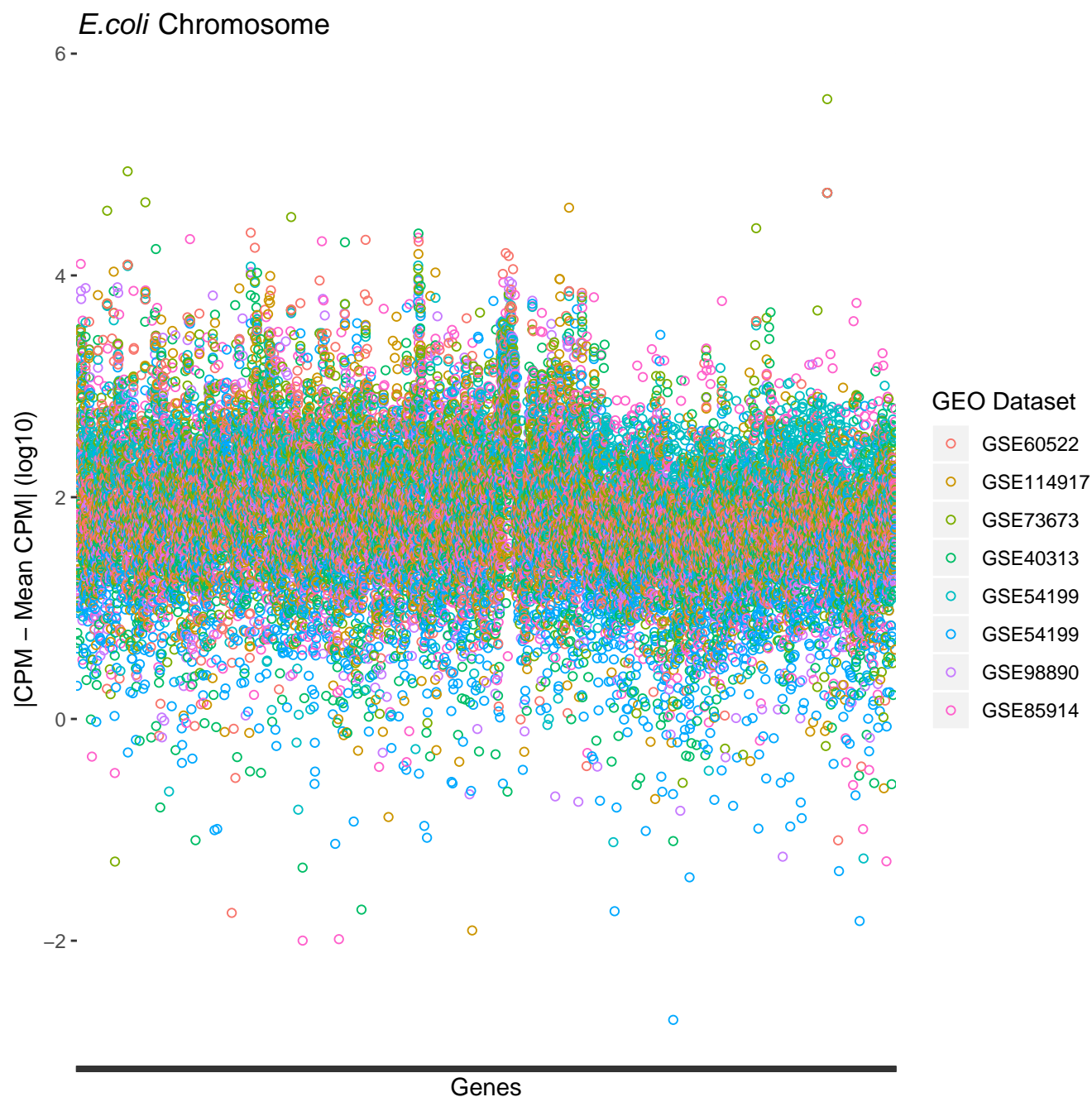


Figure S1: Distribution of the median expression value for each *E. coli* data set minus the mean expression value for that gene across all data sets. Each gene is shown on the x-axis and the log base 10 values are on the y-axis.

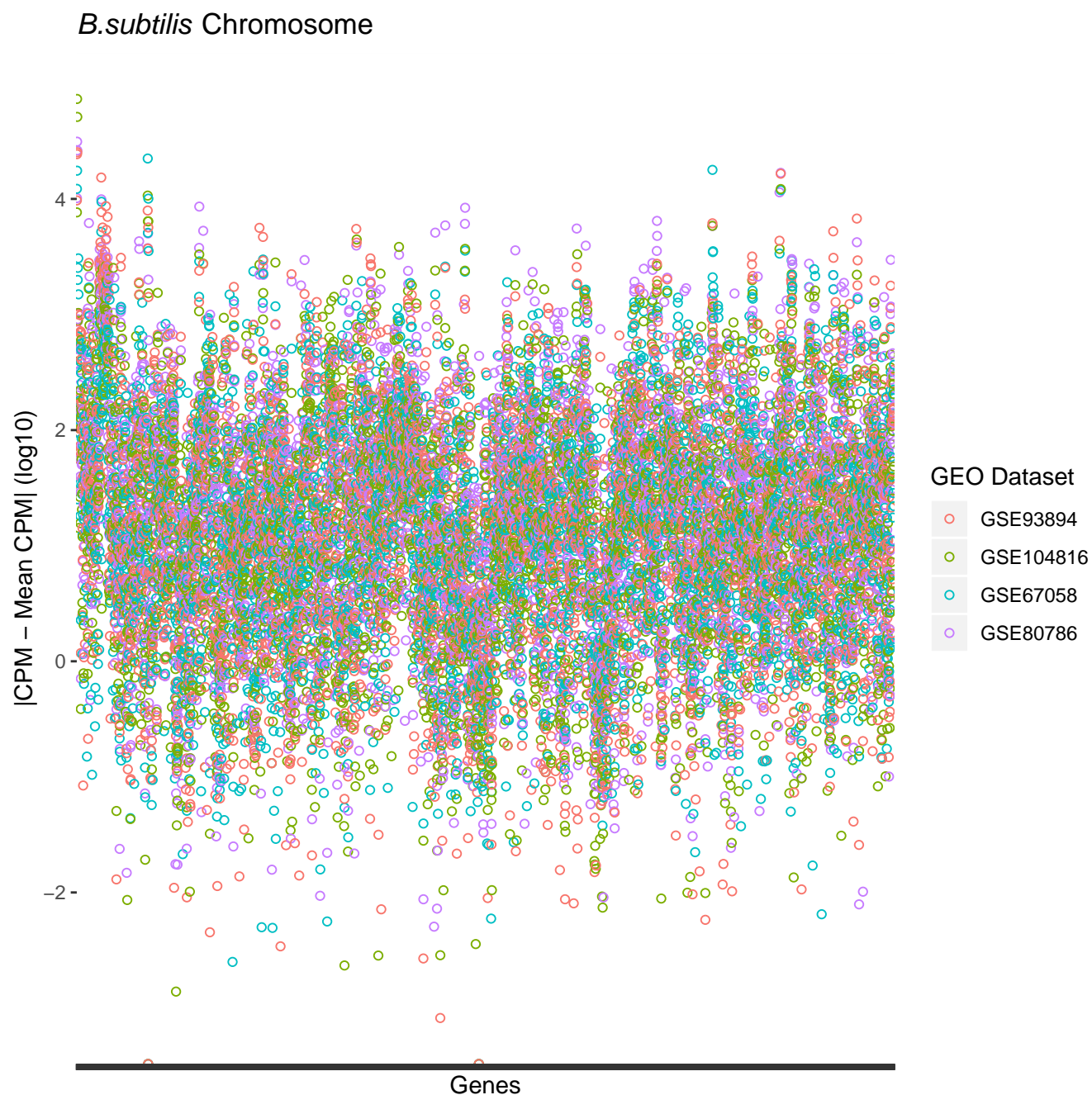


Figure S2: Distribution of the median expression value for each *B. subtilis* data set minus the mean expression value for that gene across all data sets. Each gene is shown on the x-axis and the log base 10 values are on the y-axis.

Linear Regression

Multiple more detailed linear regressions are performed to determine if there is any correlation between gene expression per gene and distance from the origin of replication. A linear regression to determine how the median CPM expression values per gene changes with genomic position was performed. Additionally, a linear regression to determine how the median CPM expression value for each 10Kbp section of the genome changes with genomic position was performed. Finally, a linear regression to determine how the total added expression over each 10Kbp region of the genome changes with genomic position was performed. All linear regression results mirror the results from the linear regression on the median gene expression CPM value per gene. Most bacteria have a negative correlation, implying that gene expression tends to decrease with distance from the origin of replication.

Bacteria and Replicon	Coefficient Estimate	Standard Error	P-value
<i>E. coli</i> Chromosome	-3.68×10^{-5}	9.30×10^{-6}	7.58×10^{-5}
<i>B. subtilis</i> Chromosome	-9.7×10^{-5}	2.0×10^{-5}	1.2×10^{-6}
<i>Streptomyces</i> Chromosome	-1.15×10^{-6}	8.12×10^{-8}	NS
<i>S. meliloti</i> Chromosome	9.57×10^{-6}	4.04×10^{-5}	NS
<i>S. meliloti</i> pSymA	1.39×10^{-3}	2.53×10^{-4}	4.9×10^{-8}
<i>S. meliloti</i> pSymB	1.55×10^{-4}	2.03×10^{-4}	NS

Table S3: Linear regression analysis of the median counts per million expression values per gene 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 bidirectional replication. NS indicates Not Significant at $P \leq 0.05$. A grey row indicates a significant negative trend.

Bacteria and Replicon	Coefficient Estimate	Standard Error	P-value
<i>E. coli</i> Chromosome	-6.38×10^{-6}	3.22×10^{-6}	4.85×10^{-2}
<i>B. subtilis</i> Chromosome	-4.04×10^{-6}	2.82×10^{-6}	NS
<i>Streptomyces</i> Chromosome	-6.29×10^{-7}	3.27×10^{-8}	NS
<i>S. meliloti</i> Chromosome	-6.84×10^{-6}	7.42×10^{-6}	NS
<i>S. meliloti</i> pSymA	-1.02×10^{-4}	6.75×10^{-5}	NS
<i>S. meliloti</i> pSymB	2.50×10^{-5}	5.86×10^{-5}	NS

Table S4: Linear regression analysis of the median counts per million expression data for 10Kbp segments of 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 bidirectional replication. Statistical outliers were removed from this linear regression calculation. NS indicates Not Significant at $P \leq 0.05$. A grey row indicates a significant negative trend.

Bacteria and Replicon	Coefficient Estimate	Standard Error	P-value
<i>E. coli</i> Chromosome	-3.26×10^{-4}	1.77×10^{-4}	1.30×10^{-8}
<i>B. subtilis</i> Chromosome	-5.63×10^{-4}	1.79×10^{-4}	1.87×10^{-3}
<i>Streptomyces</i> Chromosome	-1.37×10^{-6}	4.59×10^{-7}	2.88×10^{-3}
<i>S. meliloti</i> Chromosome	-5.39×10^{-4}	2.47×10^{-4}	3.02×10^{-2}
<i>S. meliloti</i> pSymA	3.88×10^{-3}	2.65×10^{-3}	NS
<i>S. meliloti</i> pSymB	-1.46×10^{-3}	1.95×10^{-3}	NS

Table S5: Linear regression analysis of total added expression and distance from the origin of replication. The total added expression values were calculated by summing the total counts per million expression value per 10Kbp section of the genome. 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 bidirectional replication. NS indicates Not Significant at $P \leq 0.05$. A grey row indicates a significant negative trend.

High Gene Expression Distribution

Bacteria and Replicon	Bidirectional Genomic Position (bp)	Protein/Gene Examples
<i>E. coli</i> Chromosome	0 - 10000	DNA replication and repair ATP-proton motive force ATP biosynthesis transport
	470000 - 480000	DNA replication and repair tRNA synthesis Ribosomal proteins Putative transport
	610000 - 620000	Ribosomal protein Translation modification tRNA modification RNA synthesis
	1520000 - 1530000	Energy metabolism
	2330000 - 2340000	Energy metabolism
	2770000 - 2780000	Energy metabolism
	2870000 - 2880000	Putative transport Transport
	3250000 - 3260000	Metabolism Putative transport
	0 - 10000	tRNA modification Ribosomal proteins DNA gyrase
	130000 - 150000	rRNA small subunit methylation Ribosomal proteins Elongation factor
	730000 - 740000	tRNA subunit Transcription regulation Glycolysis
<i>S. meliloti</i> Chromosome	30000 - 40000	Small molecule metabolism

		Macromolecule metabolism
		Hypothetical proteins
	1480000 - 1490000	Ribosomal proteins
		Structural elements
		Transmembrane proteins
	1550000 - 1560000	Small molecule metabolism
		Structural element
		Hypothetical proteins
	1930000 - 1940000	Hypothetical proteins
		Small molecule metabolism
<i>S. meliloti</i> pSymA	890000 - 900000	Cell processes
		Hypothetical proteins
		Macromolecule metabolism
	910000 - 920000	Hypothetical proteins
		Unknown protein
	950000 - 960000	Miscellaneous proteins
		Small molecule metabolism
<i>S. meliloti</i> pSymB	210000 - 220000	Unknown proteins
		Cell processes
		Hypothetical proteins
	290000 - 300000	Cell Division
		Small molecule metabolism
		Cell processes
	820000 - 830000	Small molecule metabolism
		Cell processes

Table S6: Table of high median CPM (Counts per Million) gene expression over 10kb genomic regions for each bacterial replicon and the associated proteins/gene functions found in that region. The genomic position begins at the origin of replication and continues in both directions until the terminus of replication (bidirectional replication).