

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 datasets we looked at the mean normalized expression values. Multiple replicates from a dataset were combined by finding the median normalized CPM expression value for each gene. This was done for any datasets that had multiple replicates. For each gene (x_i) the mean normalized expression value was calculated across all datasets (\bar{x}_{ij}). Then the normalized median expression value for each dataset 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 datasets are well mixed, implying that the expression levels are consistent across all datasets. Only *E. coli* and *B. subtilis* had multiple expression datasets available so they are the only ones that were analyzed. *Streptomyces* and all replicons of *S. meliloti* had only one dataset each and therefore were not analyzed.

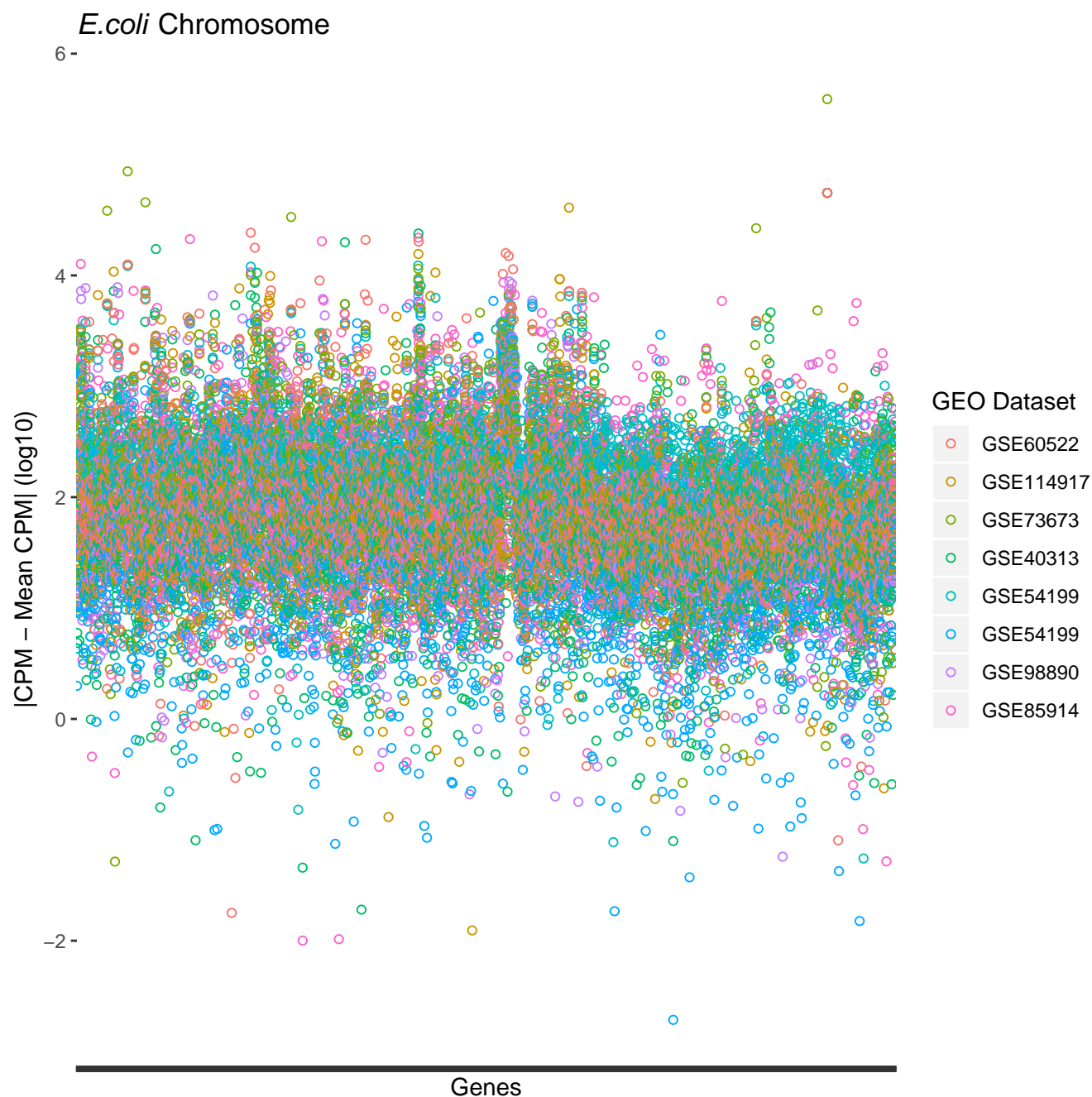


Figure S1: Distribution of the median expression value for each *E. coli* dataset minus the mean expression value for that gene across all datasets. Each gene is shown on the x-axis and the log base 10 values are on the y-axis.

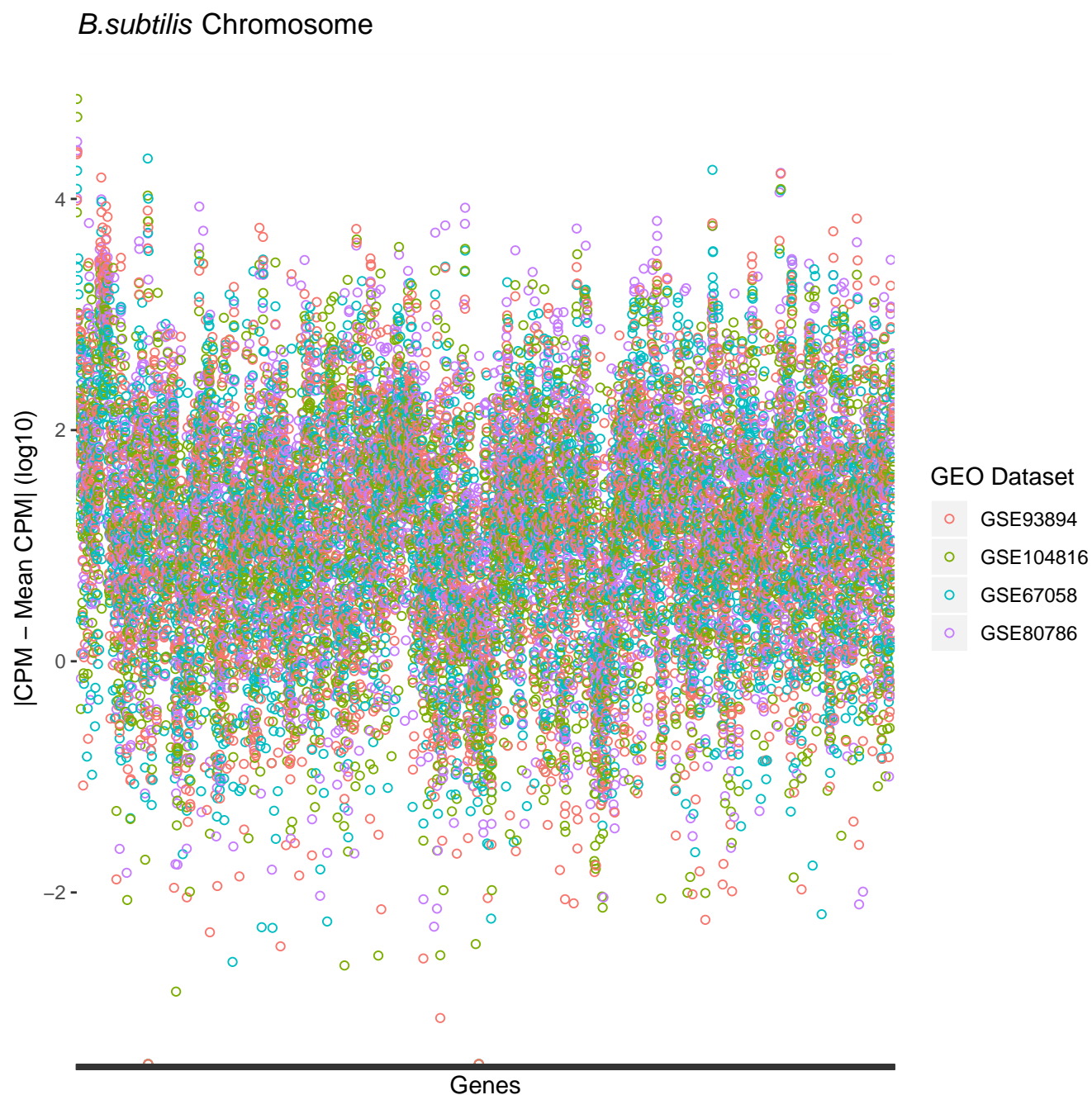


Figure S2: Distribution of the median expression value for each *B. subtilis* dataset minus the mean expression value for that gene across all datasets. Each gene is shown on the x-axis and the log base 10 values are on the y-axis.

Linear Regression

A more detailed linear regression was performed to determine if there is any correlation between gene expression per gene and distance from the origin of replication. This was done on a per gene basis where the median CPM expression value for each gene was considered a separate data point. These results mirror the results from the linear regression on the median gene expression value per 10kb genomic region. 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	-6.03×10^{-5}	1.28×10^{-5}	2.8×10^{-6}
<i>B. subtilis</i> Chromosome	-9.7×10^{-5}	2.0×10^{-5}	1.2×10^{-6}
<i>Streptomyces</i> Chromosome	-1.17×10^{-6}	1.04×10^{-7}	$< 2 \times 10^{-16}$
<i>S. meliloti</i> Chromosome	3.97×10^{-5}	4.25×10^{-5}	NS
<i>S. meliloti</i> pSymA	1.39×10^{-3}	2.53×10^{-4}	4.9×10^{-8}
<i>S. meliloti</i> pSymB	1.46×10^{-4}	2.03×10^{-4}	NS

Table S3: Linear regression analysis of the median counts per million expression data 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. A grey row indicates a significant negative trend.

Bacteria and Replicon	Coefficient Estimate	Standard Error	P-value
<i>E. coli</i> Chromosome	-1.59×10^{-5}	1.51×10^{-5}	NS
<i>B. subtilis</i> Chromosome	-1.01×10^{-4}	2.59×10^{-5}	1.31×10^{-4}
<i>Streptomyces</i> Chromosome	-1.26×10^{-7}	1.20×10^{-7}	NS
<i>S. meliloti</i> Chromosome	3.32×10^{-6}	3.48×10^{-5}	NS
<i>S. meliloti</i> pSymA	1.32×10^{-3}	3.67×10^{-4}	5.30×10^{-4}
<i>S. meliloti</i> pSymB	2.66×10^{-4}	3.64×10^{-4}	NS

Table S4: Linear regression analysis of normalized expression and distance from the origin of replication. The normalized expression values were calculated by dividing the total counts per million expression value per 10kb section of the genome by the total number of genes in the respective 10kb section. 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. NS indicates Not Significant at $P \leq 0.05$. A grey row indicates a significant negative trend.

Gene Distribution

Bacteria and Replicon	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
<i>B. subtilis</i> Chromosome	0 - 10000	tRNA modification Ribosomal proteins DNA gyrase rRNA small subunit methylation
	130000 - 150000	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