

Further supplemental information and code are available on [GitHub](#) at [INSERT PROPER GIT LINK HERE](#).

Gene Expression Data

Strain	GEO Accession Number	Date Accessed	NCBI Accession Genome Used For Gene Position
<i>E. coli</i> K12 MG1655	GSE60522	December 20, 2017	U00096
	GSE114917	November 26, 2018	
	GSE54199	December 18, 2019	
	GSE40313	November 21, 2018	
<i>E. coli</i> K12 DH10B	GSE98890	March 13, 2018	NC_010473
<i>E. coli</i> BW25113	GSE73673	December 19, 2017	NZ_CP009273
	GSE85914	December 19, 2017	
<i>E. coli</i> ATCC 25922	GSE94978	November 23, 2018	NZ_CP009072 BA000007

Table S1: Strains and species used for each gene expression analysis. Gene Expression Omnibus accession numbers and date accessed are provided. NCBI genome accession numbers are listed for which genome was used to determine the gene position. Strains with multiple NCBI genome accession numbers had multiple genome versions/builds used to determine the genomic position.

Sequences

Strain	Accession Number	Date(s) Accessed
<i>E. coli</i> K-12 MG1655 *	U00096	September 26, 2016
<i>E. coli</i> K-12 DH10B	NC_010473	February 13, 2020
<i>E. coli</i> BW25113	NZ_CP009273	October 3, 2018
<i>E. coli</i> ATCC 25922	NZ_CP009072	December 18, 2018
Outgroup: <i>Salmonella enterica</i> LT2	AE006468	September 29, 2016

Table S2: *E. coli* strains used for the analysis. Accession numbers, date accessed, and outgroups for each replicon are provided. Multiple dates and accession numbers for one strain denote updated versions of the genome. An asterix (*) insicates the strain that was used as the representative strain.

Proteomes

Strain	UniProt Accession Number	NCBI Accession Number	Date(s) Accessed
<i>E. coli</i> K-12 MG1655	UP000000625	U00096	May 4, 2020
<i>E. coli</i> K-12 DH10B	UP000001689	NC_010473	May 4, 2020
<i>E. coli</i> BW25113	UP000029103	NZ_CP009273	May 4, 2020
<i>E. coli</i> ATCC 25922	UP000001410	NZ_CP009072	May 4, 2020

Table S3: Proteomes used for the *E. coli* analysis were downloaded from UniProt (**Unkn:19**). Accession numbers for both UniProt and NCBI as well as date accessed are provided.

Correlation of Gene Expression Over Datasets

To assess uniform expression over *E. coli* strains with multiple data sets we looked at the mean normalized expression values. Multiple replicates from a data set were combined by finding the median normalized CPM expression value for each gene. This was done for any data sets that had multiple replicates. For each gene (x_i) the mean normalized expression value was calculated across all data sets (\bar{x}_{ij}). Then the normalized 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 **INSER FIGURES HERE**. All data sets are well mixed, implying that the expression levels are consistent across all data sets. Only *E. coli* K-12 MG1655 and *E. coli* BW25113 strains had multiple expression datasets available so they are the only ones that were analyzed. *E. coli* ATCC 25922 and *E. coli* K-12 DH10B had only one data set each and therefore were not analyzed.