

# SPATIAL PATTERNS OF SUBSTITUTIONS IN BACTERIAL GENOMES

DANIELLA LATO  
PAPER DRAFT

August 3, 2017

## **Abstract**

## **Materials and Methods**

## **Results**

## **Discussion**

## **Acknowledgements**

## Figures

## Supplementary Figures

### Logistic Regression

COG Abbreviation	COG Category	<i>E. coli</i> Chromosome
C	Energy Production and Conversion	$2.40 \times 10^{-7}$
E	Amino Acid Transport and Metabolism	$4.53 \times 10^{-7}$
F	Nucleotide Transport and Metabolism	NA
G	Carbohydrate Transport and Metabolism	NA
H	Coenzyme Transport and Metabolism	NA
I	Lipid Transport and Metabolism	$4.51 \times 10^{-7}$
J	Translation, Ribosomal Structure and Biogenesis	NA
K	Transcription	$2.22 \times 10^{-7}$
L	Replication, Recombination and Repair	$2.95 \times 10^{-7}$
M	Cell Wall/Membrane/Envelope Biogenesis	NA
N	Cell Motility	$-7.74 \times 10^{-7}$
O	Posttranslational Modification, Protein Turnover, Chaperones	$3.37 \times 10^{-7}$
P	Inorganic Ion Transport and Metabolism	NA
Q	Secondary Metabolites Biosynthesis, Transport and Catabolism	NA
T	Signal Transduction Mechanisms	NA
U	Intracellular Trafficking, Secretion, and Vesicular Transport	NA
V	Defense Mechanisms	$3.75 \times 10^{-7}$
W	Extracellular Structures	$-3.23 \times 10^{-6}$
X	Mobilome: Prophages, Transposons	$-1.09 \times 10^{-6}$
R	General Function Prediction Only	$2.61 \times 10^{-7}$
S	Function Unknown	$-1.43 \times 10^{-6}$

### Logistic Regression Stats

COG Abbreviation	COG Category	Coefficient Estimate	Standard Error	P-value
C	Energy Production and Conversion	$2.396 \times 10^{-7}$	$7.202 \times 10^{-8}$	$8.77 \times 10^{-4}$
E	Amino Acid Transport and Metabolism	$4.53 \times 10^{-7}$	$6.464 \times 10^{-8}$	$2.42 \times 10^{-12}$
I	Lipid Transport and Metabolism	$4.506 \times 10^{-7}$	$1.065 \times 10^{-7}$	$2.32 \times 10^{-5}$
K	Transcription	$2.221 \times 10^{-7}$	$7.031 \times 10^{-8}$	$1.58 \times 10^{-3}$
L	Replication, Recombination and Repair	$2.95 \times 10^{-7}$	$9.903 \times 10^{-8}$	$2.89 \times 10^{-3}$
N	Cell Motility	$-7.743 \times 10^{-7}$	$1.453 \times 10^{-7}$	$9.8 \times 10^{-8}$
O	Posttranslational Modification, Protein Turnover, Chaperones	$3.371 \times 10^{-7}$	$9.418 \times 10^{-8}$	$3.45 \times 10^{-4}$
R	General Function Prediction Only	$2.606 \times 10^{-7}$	$7.056 \times 10^{-8}$	$2.21 \times 10^{-4}$
S	Function Unknown	$-1.427 \times 10^{-6}$	$1.323 \times 10^{-7}$	$< 2 \times 10^{-16}$
V	Defense Mechanisms	$3.754 \times 10^{-7}$	$1.203 \times 10^{-7}$	$1.8 \times 10^{-3}$
W	Extracellular Structures	$-3.234 \times 10^{-6}$	$5.718 \times 10^{-7}$	$1.55 \times 10^{-8}$
X	Mobilome: Prophages, Transposons	$-1.085 \times 10^{-6}$	$1.962 \times 10^{-7}$	$3.23 \times 10^{-8}$

<i>B. subtilis</i>				
COG Abbreviation	COG Category	Coefficient Estimate	Standard Error	P-value
C	Energy Production and Conversion	$4.095 \times 10^{-7}$	$1.284 \times 10^{-7}$	$1.42 \times 10^{-3}$
F	Nucleotide Transport and Metabolism	$-7.49 \times 10^{-7}$	$1.771 \times 10^{-7}$	$2.35 \times 10^{-5}$
H	Coenzyme Transport and Metabolism	$-4.068 \times 10^{-7}$	$1.223 \times 10^{-7}$	$8.77 \times 10^{-4}$
I	Lipid Transport and Metabolism	$3.744 \times 10^{-7}$	$1.358 \times 10^{-7}$	$5.84 \times 10^{-3}$
J	Translation, Ribosomal Structure and Biogenesis	$-7.131 \times 10^{-7}$	$1.171 \times 10^{-7}$	$1.12 \times 10^{-9}$
K	Transcription	$7.615 \times 10^{-7}$	$1.001 \times 10^{-7}$	$2.83 \times 10^{-14}$
M	Cell Wall/Membrane/Envelope Biogenesis	$5.182 \times 10^{-7}$	$1.091 \times 10^{-7}$	$2.02 \times 10^{-6}$
N	Cell Motility	$1.009 \times 10^{-6}$	$2.191 \times 10^{-7}$	$4.11 \times 10^{-6}$
O	Posttranslational Modification, Protein Turnover, Chaperones	$3.508 \times 10^{-7}$	$1.451 \times 10^{-8}$	$1.56 \times 10^{-2}$
R	General Function Prediction Only	$3.197 \times 10^{-7}$	$8.967 \times 10^{-8}$	$3.63 \times 10^{-4}$
S	Function Unknown	$-1.185 \times 10^{-6}$	$1.286 \times 10^{-7}$	$< 2 \times 10^{-16}$
V	Defense Mechanisms	$7.148 \times 10^{-7}$	$1.711 \times 10^{-7}$	$2.94 \times 10^{-5}$
X	Mobilome: Prophages, Transposons	$-1.814 \times 10^{-6}$	$4.027 \times 10^{-7}$	$6.65 \times 10^{-6}$

<i>Streptomyces</i>				
COG Abbreviation	COG Category	Coefficient Estimate	Standard Error	P-value
C	Energy Production and Conversion	$-2.939 \times 10^{-7}$	$2.536 \times 10^{-8}$	$< 2 \times 10^{-16}$
E	Amino Acid Transport and Metabolism	$-1.967 \times 10^{-7}$	$1.849 \times 10^{-8}$	$< 2 \times 10^{-12}$
F	Nucleotide Transport and Metabolism	$-1.864 \times 10^{-7}$	$4.142 \times 10^{-8}$	$6.78 \times 10^{-6}$
H	Coenzyme Transport and Metabolism	$-1.112 \times 10^{-7}$	$2.172 \times 10^{-8}$	$3.06 \times 10^{-7}$
I	Lipid Transport and Metabolism	$-2.011 \times 10^{-7}$	$1.952 \times 10^{-8}$	$< 2 \times 10^{-3}$
J	Translation, Ribosomal Structure and Biogenesis	$-1.355 \times 10^{-7}$	$2.59 \times 10^{-8}$	$1.68 \times 10^{-7}$
L	Replication, Recombination and Repair	$-1.174 \times 10^{-7}$	$3.205 \times 10^{-8}$	$2.5 \times 10^{-4}$
M	Cell Wall/Membrane/Envelope Biogenesis	$-8.046 \times 10^{-8}$	$2.326 \times 10^{-8}$	$5.43 \times 10^{-4}$
N	Cell Motility	$-2.04 \times 10^{-7}$	$9.227 \times 10^{-8}$	$2.71 \times 10^{-2}$
O	Posttranslational Modification, Protein Turnover, Chaperones	$-7.751 \times 10^{-8}$	$2.777 \times 10^{-8}$	$5.25 \times 10^{-3}$
P	Inorganic Ion Transport and Metabolism	$-1.682 \times 10^{-7}$	$2.346 \times 10^{-8}$	$7.52 \times 10^{-13}$
R	General Function Prediction Only	$-6.914 \times 10^{-8}$	$1.285 \times 10^{-8}$	$7.38 \times 10^{-8}$
S	Function Unknown	$4.615 \times 10^{-7}$	$2.671 \times 10^{-8}$	$< 2 \times 10^{-16}$
T	Signal Transduction Mechanisms	$1.517 \times 10^{-7}$	$1.531 \times 10^{-8}$	$< 2 \times 10^{-16}$
V	Defense Mechanisms	$-1.205 \times 10^{-7}$	$2.739 \times 10^{-8}$	$1.09 \times 10^{-5}$
W	Extracellular Structures	$9.063 \times 10^{-7}$	$2.336 \times 10^{-7}$	$1.05 \times 10^{-4}$
X	Mobilome: Prophages, Transposons	$4.323 \times 10^{-7}$	$3.599 \times 10^{-8}$	$< 2 \times 10^{-16}$

<i>S. meliloti</i> Chromosome				
COG Abbreviation	COG Category	Coefficient Estimate	Standard Error	P-value
E	Amino Acid Transport and Metabolism	$2.662 \times 10^{-7}$	$8.996 \times 10^{-8}$	$3.08 \times 10^{-3}$
F	Nucleotide Transport and Metabolism	$-6.682 \times 10^{-7}$	$2.02 \times 10^{-7}$	$9.43 \times 10^{-4}$
G	Carbohydrate Transport and Metabolism	$-2.533 \times 10^{-7}$	$1.067 \times 10^{-7}$	$1.76 \times 10^{-2}$
H	Coenzyme Transport and Metabolism	$-1.095 \times 10^{-6}$	$1.574 \times 10^{-7}$	$3.5 \times 10^{-12}$
J	Translation, Ribosomal Structure and Biogenesis	$-1.233 \times 10^{-6}$	$1.544 \times 10^{-7}$	$1.42 \times 10^{-15}$
M	Cell Wall/Membrane/Envelope Biogenesis	$4.594 \times 10^{-7}$	$1.38 \times 10^{-7}$	$8.73 \times 10^{-4}$
O	Posttranslational Modification, Protein Turnover, Chaperones	$3.474 \times 10^{-7}$	$1.623 \times 10^{-7}$	$3.23 \times 10^{-2}$
P	Inorganic Ion Transport and Metabolism	$5.361 \times 10^{-7}$	$1.367 \times 10^{-7}$	$8.77 \times 10^{-5}$
R	General Function Prediction Only	$9.493 \times 10^{-7}$	$1.07 \times 10^{-7}$	$< 2 \times 10^{-16}$
S	Function Unknown	$-7.436 \times 10^{-7}$	$3.329 \times 10^{-7}$	$2.55 \times 10^{-2}$
T	Signal Transduction Mechanisms	$1.851 \times 10^{-6}$	$1.948 \times 10^{-7}$	$< 2 \times 10^{-16}$
U	Intracellular Trafficking, Secretion, and Vesicular Transport	$8.624 \times 10^{-7}$	$4.29 \times 10^{-7}$	$4.44 \times 10^{-2}$
V	Defense Mechanisms	$4.235 \times 10^{-7}$	$2.036 \times 10^{-7}$	$3.75 \times 10^{-2}$
X	Mobilome: Prophages, Transposons	$1.673 \times 10^{-6}$	$6.938 \times 10^{-7}$	$1.59 \times 10^{-2}$

<i>S. meliloti</i> pSymA				
COG Abbreviation	COG Category	Coefficient Estimate	Standard Error	P-value
E	Amino Acid Transport and Metabolism	$-2.077 \times 10^{-6}$	$3.619 \times 10^{-7}$	$9.45 \times 10^{-9}$
F	Nucleotide Transport and Metabolism	$-1.147 \times 10^{-6}$	$5.704 \times 10^{-7}$	$4.43 \times 10^{-2}$
G	Carbohydrate Transport and Metabolism	$9.779 \times 10^{-7}$	$4.287 \times 10^{-7}$	$2.25 \times 10^{-2}$
H	Coenzyme Transport and Metabolism	$-9.828 \times 10^{-7}$	$4.699 \times 10^{-7}$	$3.65 \times 10^{-2}$
I	Lipid Transport and Metabolism	$2.006 \times 10^{-6}$	$6.299 \times 10^{-7}$	$1.45 \times 10^{-3}$
J	Translation, Ribosomal Structure and Biogenesis	$-1.506 \times 10^{-6}$	$4.205 \times 10^{-7}$	$3.42 \times 10^{-4}$
L	Replication, Recombination and Repair	$1.416 \times 10^{-6}$	$6.18 \times 10^{-7}$	$2.19 \times 10^{-2}$
M	Cell Wall/Membrane/Envelope Biogenesis	$1.633 \times 10^{-6}$	$5.42 \times 10^{-7}$	$2.58 \times 10^{-3}$
Q	Secondary Metabolites Biosynthesis, Transport and Catabolism	$4.283 \times 10^{-6}$	$6.738 \times 10^{-7}$	$2.07 \times 10^{-10}$
S	Function Unknown	$2.53 \times 10^{-5}$	$9.966 \times 10^{-6}$	$1.111 \times 10^{-2}$

<i>S. meliloti</i> pSymB				
COG Abbreviation	COG Category	Coefficient Estimate	Standard Error	P-value
E	Amino Acid Transport and Metabolism	$-9.452 \times 10^{-7}$	$2.901 \times 10^{-7}$	$1.12 \times 10^{-3}$
G	Carbohydrate Transport and Metabolism	$2.048 \times 10^{-6}$	$4.07 \times 10^{-7}$	$4.87 \times 10^{-7}$
H	Coenzyme Transport and Metabolism	$-1.452 \times 10^{-6}$	$4.089 \times 10^{-7}$	$3.84 \times 10^{-4}$
I	Lipid Transport and Metabolism	$1.839 \times 10^{-6}$	$5.832 \times 10^{-7}$	$1.61 \times 10^{-3}$
J	Translation, Ribosomal Structure and Biogenesis	$-1.151 \times 10^{-6}$	$3.589 \times 10^{-7}$	$1.34 \times 10^{-3}$
K	Transcription	$-4.166 \times 10^{-6}$	$5.168 \times 10^{-7}$	$7.59 \times 10^{-16}$
M	Cell Wall/Membrane/Envelope Biogenesis	$5.408 \times 10^{-6}$	$7.131 \times 10^{-7}$	$3.37 \times 10^{-14}$
O	Posttranslational Modification, Protein Turnover, Chaperones	$1.078 \times 10^{-6}$	$5.23 \times 10^{-7}$	$3.93 \times 10^{-2}$
P	Inorganic Ion Transport and Metabolism	$-2.045 \times 10^{-6}$	$4.404 \times 10^{-7}$	$3.41 \times 10^{-6}$
Q	Secondary Metabolites Biosynthesis, Transport and Catabolism	$3.806 \times 10^{-6}$	$6.679 \times 10^{-7}$	$1.21 \times 10^{-8}$
S	Function Unknown	$4.141 \times 10^{-5}$	$1.595 \times 10^{-5}$	$9.44 \times 10^{-3}$
X	Mobilome: Prophages, Transposons	$-3.347 \times 10^{-6}$	$1.514 \times 10^{-6}$	$2.7 \times 10^{-2}$