Spatial Patterns of Molecular Trends in Bacterial Genomes

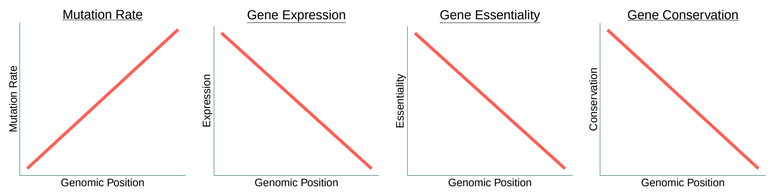
Daniella LATO* G. Brian GOLDING

*latodf@mcmaster.ca

Department of Biology, McMaster University, Hamilton, Ontario, Canada

Molecular Traits Vary with Genomic Position

Molecular traits such as mutation rate and gene expression change with distance from the origin of replication. Genes that are highly conserved, essential, and expressed at higher levels are typically found near the origin of replication[1;2;3;4;5]. The disparity in molecular features between the origin of replication and the terminus is thought to be due to decreased mutation rate near the origin of replication[1;3;4] and the requirement to maintain proper gene dosage[3] amount and replication timing[1;5]. These **spatial molecular trends have not been analyzed while accounting for genomic reorganization** such as rearrangements and inversion, which provide bacteria with the opportunity to gain new genetic information.



***Objective:* In depth analysis of genomic patterns of substitutions and gene expression in bacterial genomes.**

***Hypothesis:* The number of substitutions increases with increasing distance from the origin of replication.**

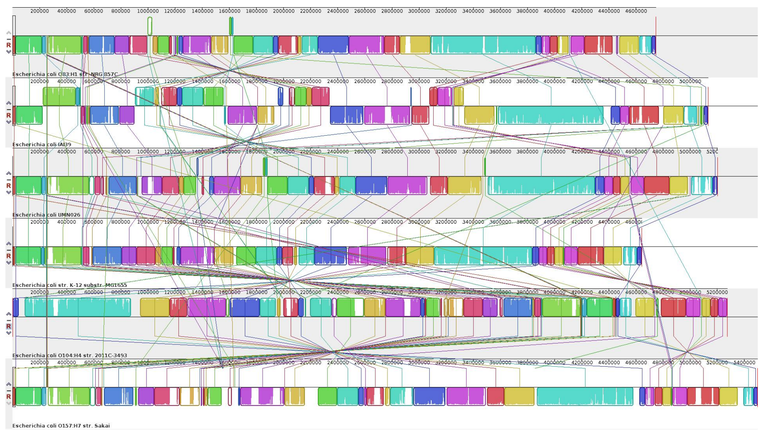
1. Methods

25 bacterial genomes and gene expression data from **9 bacterial species**.

Bacteria and Replicon	Number of Strains	Genome Structure
Escherichia coli Chromosome	6	○
Bacillus subtilis Chromosome	7	○
Streptomyces Chromosome	6	—
Sinorhizobium meliloti Chromosome	6	○ ○ ○
Sinorhizobium meliloti pSymA	6	○ ○ ○
Sinorhizobium meliloti pSymB	6	○ ○ ○

Alignment:

1. Each whole genome was **globally aligned** to obtain **L**ocally **C**o-linear **B**locks (**LCB**) which **allows for genome reorganization** such as inversions and rearrangements (`progressiveMauve`[6]).
2. **Each LCB** was **locally re-aligned** with `MAFFT`[7] for a more accurate alignment.



Example `progressiveMauve` alignment of *E. coli* genomes.

Phylogeny:

3. **Bootstrap phylogenetic trees** were constructed for each bacterial replicon (`PHYLIP`[8]).

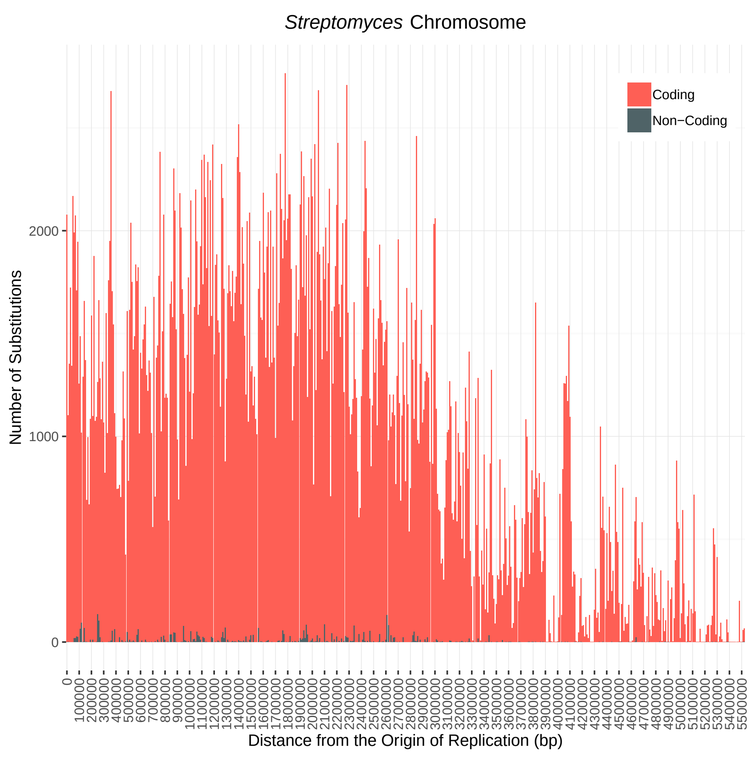
Statistical Analysis:

4. Each **genomic position** was **scaled** to the origin of replication and accounted for bidirectional replication (`R`[9]).
5. To track genome rearrangements, **nucleotide substitutions and genomic positions** were **reconstructed** in extinct taxa and **ancestors** on the given phylogenies (custom `Perl` and `Python` scripts, `PAML`[10]).
6. **Logistic regression and linear regression** were performed to compute changes in substitutions and gene expression across all coding and non-coding segments of the genome (`R`[9]).

See `GitHub` for more information.

www.github.com/dlato/SMBE_Mutation_Rate_Evolution_Poster

2. Substitutions ↘ with Genomic Position



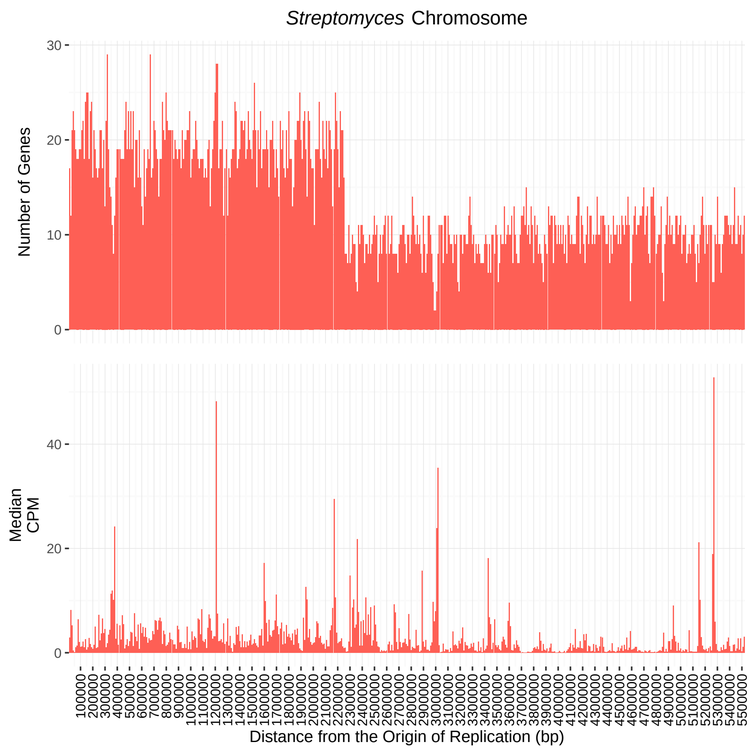
Bacteria and Replicon	Coefficient Estimate		Genome Structure
	Coding Sites	Non-Coding Sites	
E. coli Chromosome	-9.119×10^{-8}***	7.022×10^{-8}***	○
B. subtilis Chromosome	-1.273×10^{-7}***	-9.861×10^{8}***	○
Streptomyces Chromosome	-7.945×10^{-9}***	3.637×10^{-7}***	—
S. meliloti Chromosome	-1.550×10^{-7}***	-1.510×10^{-7}*	○ ○ ○
S. meliloti pSymA	-1.156×10^{-7}*	NS	○ ○ ○
S. meliloti pSymB	2.587×10^{-7}***	8.591×10^{-7}***	○ ○ ○

Significance Codes: p < 0.001 = '***', 0.001 < 0.01 = '**', 0.01 < 0.05 = '*', NS = Not Significant

Discussion: Increased transposon insertion events[11] and potential genomic and pathogenicity islands[12;13] near the origin of replication can cause asymmetry in nucleotide composition[14], GC content[15], and mutation rate[1;3;4]. This could be why we see an increase in the number of substitutions near the origin of replication.

Conclusion: The number of substitutions decreases with increasing distance from the origin of replication in most bacterial replicons at most coding sites.

3. Gene Expression ↗ with Genomic Position



Bacteria and Replicon	Coefficient Estimate	Genome Structure
E. coli Chromosome	-6.03×10^{-5}***	○
B. subtilis Chromosome	-9.7×10^{-5}***	○
Streptomyces Chromosome	-1.17×10^{-6}***	—
S. meliloti Chromosome	NS	○ ○ ○
S. meliloti pSymA	1.39×10^{-3}***	○ ○ ○
S. meliloti pSymB	NS	○ ○ ○

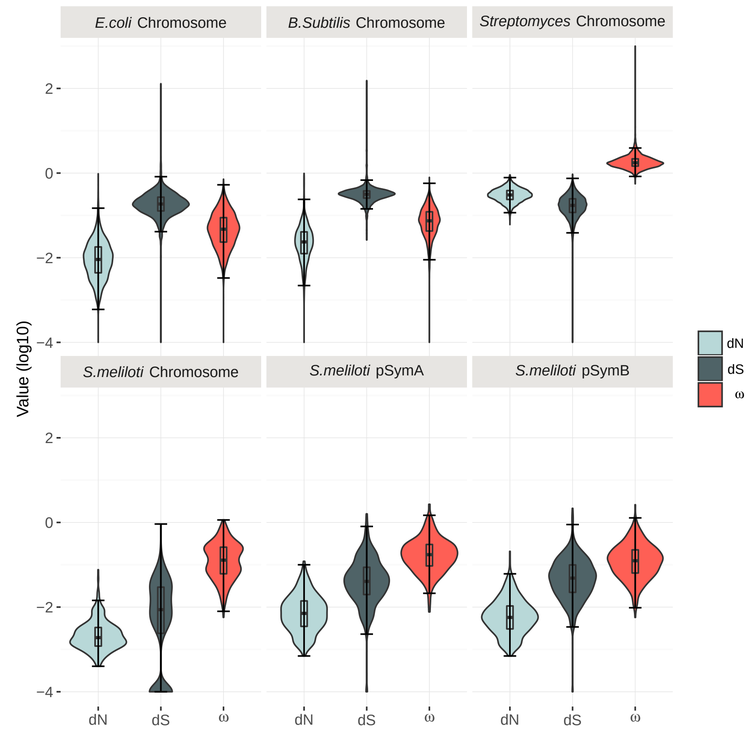
Significance Codes: p < 0.001 = '***', 0.001 < 0.01 = '**', 0.01 < 0.05 = '*', NS = Not Significant

Discussion: Replication error increases when moving away from the origin of replication, therefore, genes that are highly expressed are often located near the origin.[16;17;18]

Conclusion: Gene expression decreases with increasing distance from the origin of replication in most bacterial replicons.

4. Preliminary Selection Results

Distribution of *dN*, *dS*, and ω for the coding regions (`codeml` `PAML`[10]).



dN = # of non-synonymous substitutions per site, *dS* = # of synonymous substitutions per site,
ω = ratio *dN*/ *dS*

Most of the bacterial replicons are under **purifying selection** ($\omega < 1$).

Conclusions and Ongoing Research

Determining how the number of substitutions are distributed spatially throughout bacterial genomes broadens our understanding of their evolution. **Most coding regions of the replicons considered in this study have the number of substitutions and gene expression decrease with increasing distance from the origin of replication.** The exceptions to this were pSymB of *S. meliloti*, *E. coli* and *Streptomyces* which had the number of substitutions increase when moving away from the origin of replication. These spatial substitution and gene expression results can be used to determine if all bacteria possess the same evolutionary patterns.

Current work involves:

- Determining the *dN*/*dS* ratio for each coding section of the bacterial genomes and if this changes with genomic position.
- Identifying inversions within bacterial genomes and how they might alter gene expression
- Performing ancestral reconstruction of gene expression in several *E. coli* species

References

[1] J D Morrow and V S Cooper. Evolutionary effects of translocations in bacterial genomes. *Genome Biology and Evolution*, 4(12):1256–1262, 2012.
[2] P M Sharp, E Bailes, R J Grocock, J F Peden, and R E Sockett. Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Research*, 33(4):1141–1153, 2005.
[3] E Couturier and E P Rocha. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Molecular Microbiology*, 59(5):1506–1518, 2006.
[4] V S Cooper, S H Vhor, S C Wrocklage, and P J Hatcher. Why genes evolve faster on secondary chromosomes in bacteria. *PLoS Computational Biology*, 6(4):e1000732, 2010.
[5] K M Flynn, S H Vohr, P J Hatcher, and V S Cooper. Evolutionary rates and gene dispensability associate with replication timing in the archaeon Sulfolobus islandicus. *Genome Biology and Evolution*, 2:859–869, 2010.
[6] A E Darling, B Mau, and N T Perna. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS one*, 5(6):e11147, 2010.
[7] K Katoh, K Misawa, K Kuma, and T Miyata. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14):3059–3066, 2014.
[8] J Felsenstein. PHYLIP -Phylogeney Inference Package(Version 3.2). *Cladistis*, 5:164–166, 1989.
[9] R Development Core Team. R: a language and environment for statistical computing, 2014.
[10] Z Yang. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics*, 13(5):555–556, 1997.
[11] S Y Gerdes, M D Scholle, J W Campbell, G Balazsi, E Ravasz, M D Daugherty, A L Somera, N C Kyrpides, I Anderson, M S Gelfand, A Bhattacharya, V Kapatral, M D'Souza, M V Baev, Y Grechkin, F Mseeh, M Y Fonstein, R Overbeek, A L Barabasi, Z N Oltvai, and A L Osterman. Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. *Journal of Bacteriology*, 185(19):5673–5684, 2003.
[12] S Karlin. Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes. *Trends in Microbiology*, 9(7):335–343, 2001.
[13] A Mira, A B Martin-Cuadrado, G D'Auria, and F Rodriguez-Valera. The bacterial pan-genome:a new paradigm in microbiology. *International Microbiology*, 13(2):45–57, 2010.
[14] P Mackiewicz, A Gierlik, M Kowalczuk, M R Dudek, and S Cebrat. How does replication-associated mutational pressure influence amino acid composition of proteins? *Genome Research*, 9(5):409–416, 1999.
[15] H Ikeda, J Ishikawa, A Hanamoto, M Shinose, H Kikuchi, T Shiba, Y Sakaki, M Hattori, and S Omura. Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. *Nature Biotechnology*, 21(5):526–531, 2003.
[16] E P C Rocha. The replication-related organization of bacterial genomes. *Microbiology*, 150(6):1609–1627, 2004.
[17] R S Washburn and M E Gottesman. Transcription termination maintains chromosome integrity. *Proceedings of the National Academy of Sciences*, 108(2):792–797, 2011.
[18] D H S Block, R Hussein, L W Liang, and H N Lim. Regulatory consequences of gene translocation in bacteria. *Nucleic acids research*, 40(18):8979–8992, 2012.

McMaster University

Natural Sciences and Engineering Research Council of Canada
Conseil de recherches en sciences naturelles et en génie du Canada
Canada