

Visualizing Molecular Trends in Bacterial Genomes

Daniella Lato and Jana Taha

Stats 744: Final Project

December 3, 2019

The Organisms

Bacteria:

- *Escherichia coli*
- *Bacillus subtilis*
- *Streptomyces*
- *Sinorhizobium meliloti*

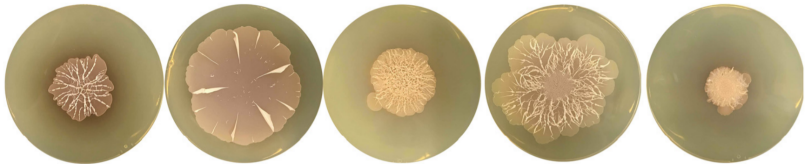
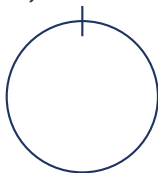


Photo: *Streptomyces* by Stephanie Jones, McMaster University

Biological Background

1) Circular



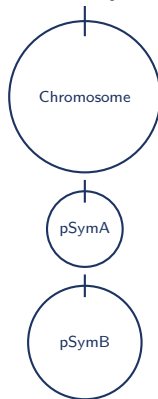
Escherichia coli
Bacillus subtilis

2) Linear



Streptomyces

3) Multi-repliconic



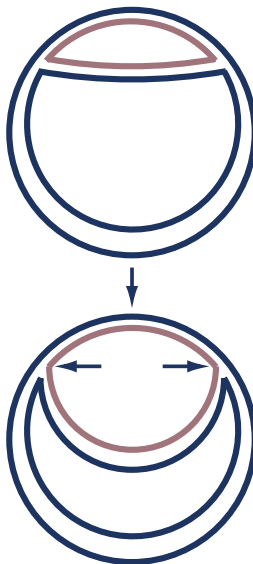
Sinorhizobium meliloti

Bacterial Genome Properties

Bidirectional Replication

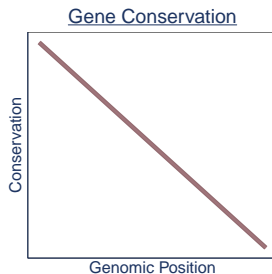
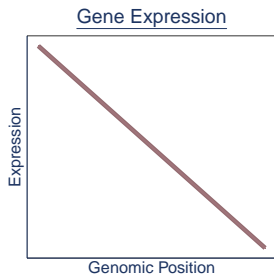
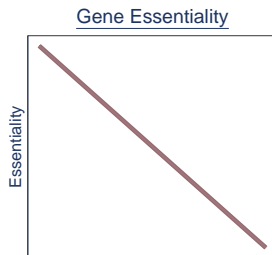
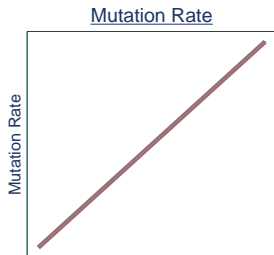
- Gene dosage
- Sequence composition
- Codon bias
- Replication errors
- Strand differences

– new DNA
– old DNA



Different **replicons** and **positions** of bacterial genomes can have **varying molecular trends**

Spatial Molecular Trends



Couturier et al. 2006, Cooper et al. 2010, Sharp et al. 2005, Morrow et al. 2012, Cooper and Rocha 2006

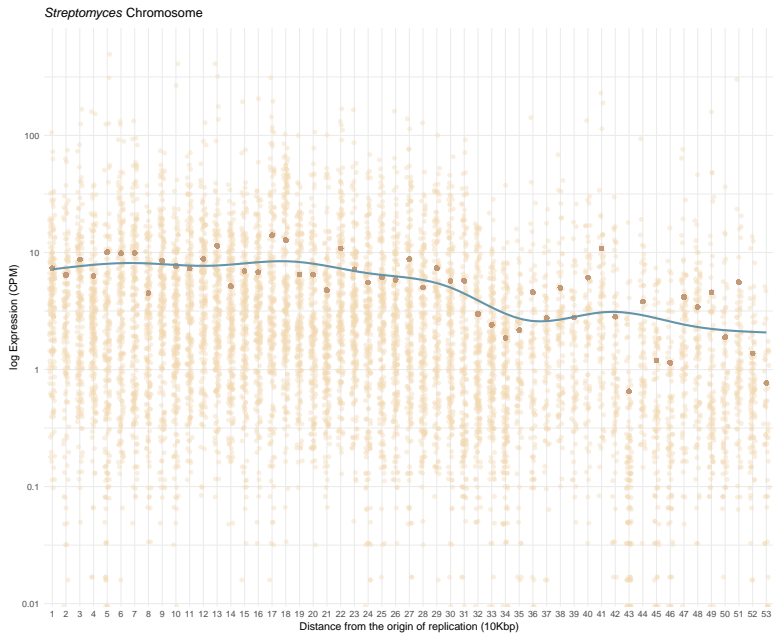
The Data

Gene Expression

- *E. coli*, *B. subtilis*, *Streptomyces*, and *S. meliloti*
- Gene expression + distance from the origin of replication
- **Expectation:** Near the origin or replication = Higher Expression

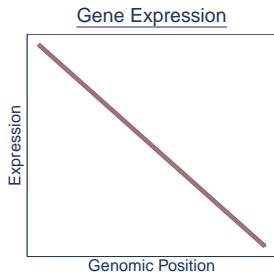
We will focus on *Streptomyces*

Gene Expression



Spatial Molecular Trends

Expectation:



Observation:

Gene expression decreases with increasing distance from the origin of replication in a **non-linear** manner

Selection

Measuring what evolutionary forces are acting on genes

Selection

Measuring what evolutionary forces are acting on genes

- **dN** = non-synonymous substitution rate
 - change in amino acid sequence
 - can alter function of the protein

Selection

Measuring what evolutionary forces are acting on genes

- **dN** = non-synonymous substitution rate
 - change in amino acid sequence
 - can alter function of the protein
- **dS** = synonymous substitution rate
 - no change in amino acid sequence
 - will not alter function of protein

Selection

Measuring what evolutionary forces are acting on genes

- **dN** = non-synonymous substitution rate
 - change in amino acid sequence
 - can alter function of the protein
- **dS** = synonymous substitution rate
 - no change in amino acid sequence
 - will not alter function of protein
- $\omega = (\text{dN}/\text{dS})$

Selection

Measuring what evolutionary forces are acting on genes

- **dN** = non-synonymous substitution rate
 - change in amino acid sequence
 - can alter function of the protein
- **dS** = synonymous substitution rate
 - no change in amino acid sequence
 - will not alter function of protein
- $\omega = (\text{dN}/\text{dS})$
 - $\omega > 1 \rightarrow$ positive selection
 - beneficial to the organism and will likely be maintained over time

Selection

Measuring what evolutionary forces are acting on genes

- **dN** = non-synonymous substitution rate
 - change in amino acid sequence
 - can alter function of the protein
- **dS** = synonymous substitution rate
 - no change in amino acid sequence
 - will not alter function of protein
- $\omega = (dN/dS)$
 - $\omega > 1 \rightarrow$ positive selection
 - beneficial to the organism and will likely be maintained over time
 - $\omega = 1 \rightarrow$ neutral selection
 - neither beneficial nor deleterious to the organism

Selection

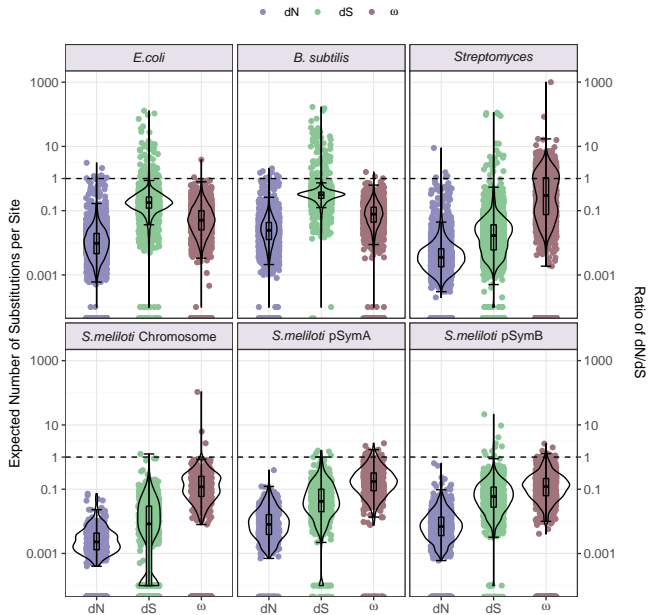
Measuring what evolutionary forces are acting on genes

- **dN** = non-synonymous substitution rate
 - change in amino acid sequence
 - can alter function of the protein
- **dS** = synonymous substitution rate
 - no change in amino acid sequence
 - will not alter function of protein
- $\omega = (dN/dS)$
 - $\omega > 1 \rightarrow$ positive selection
 - beneficial to the organism and will likely be maintained over time
 - $\omega = 1 \rightarrow$ neutral selection
 - neither beneficial nor deleterious to the organism
 - $\omega < 1 \rightarrow$ negative selection
 - deleterious to the organism and will likely not be maintained over time

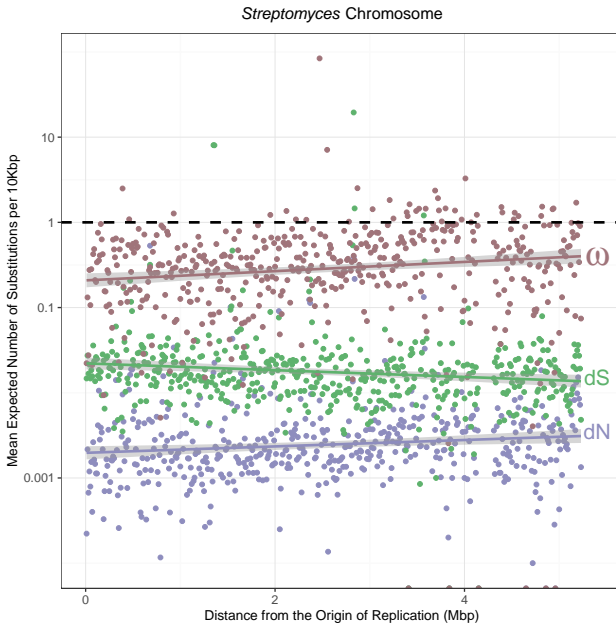
Selection

- *E. coli*, *B. subtilis*, *Streptomyces*, and *S. meliloti*
- dN, dS, and ω + distance from the origin of replication
- $\omega > 1$ (positive selection), $\omega = 1$ (neutral), $\omega < 1$ (negative selection)
- **Expectations:**
 1. dS > dN
 2. Most genes are under neutral or negative selection
 3. Genes that are under positive selection should most likely be accessory genes
 - Core genes are near the origin of replication
 - Accessory genes are usually near the terminus

Selection Summary



Streptomyces Selection



Summary

- ✓ Gene expression decreases (non-linearly) with increasing distance from the origin of replication
 - Wave-like pattern at the end of the graph?
- ✓ $dS > dN$
- ✓ Most genes are under neutral or negative selection ($\omega \leq 1$)
- ✓ In *Streptomyces*, genes under positive selection ($\omega > 1$) were near the terminus and most likely accessory genome