# Bioinformatics:
# the hot interdisciplinary field

**Daniella Lato**
PhD Candidate
Biology Department
Golding Lab, McMaster

✉ latodf@mcmaster.ca
GitHub: https://github.com/dlato

```
cagccagatggggggagggggtgagcgctctcccgctcaaa
acctccagcactttgcgatgcgtttcgctcacttgccgct
tcctaatctaaaaataaaactgaatttataaggttttctat
ttcttagagtcggtgggtatcaaggattaaaatcaatcct
catcctatccagtccgcgtcaatctccggcaaagaggcgg
gagagattccggccgaattgcgcggtaagcggggaaaccc
ggtaaaacactgcagagtcagccccttgccggcgatcggg
agtttgcggttcctgtggatgaggagtcgatctgcgtgac
aatttgtcgtggccatgagtcttgtccacatgcccgggca
agagatttccggttaggtgtcagatgtgaacaagtcgccg
cttttccacttgcctgaagccaggcgccgccgttagctgt
ttttgtcccgccttatcaacagaccgcggagaattgcgtg
gagaaccgcaaaaactacttccatctgcatttgatctccg
attcgaccggcgaaactctgatcgccgccggccgagcggc
tgctgcgcaattccagtcctcccatgcgctggaacacgtc
tatccgctgatccgcaaccggaagcagctgatgcaggtca
```
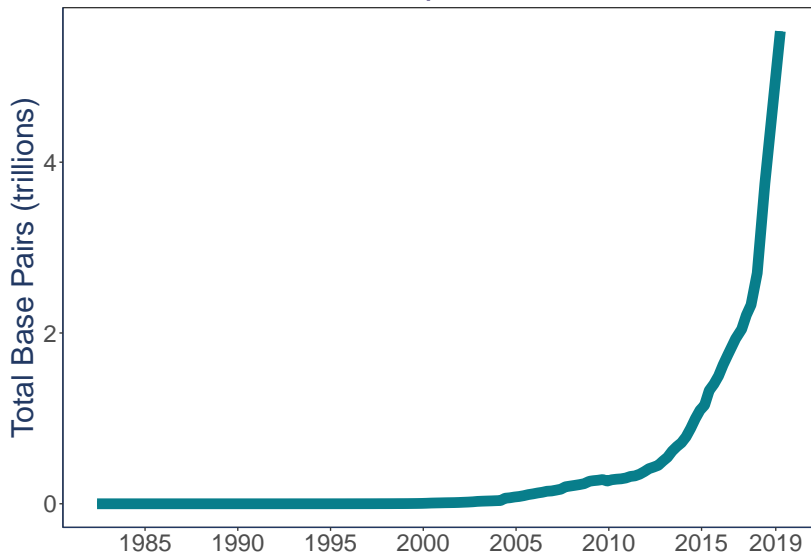
Sequencing Cost per Human Genome
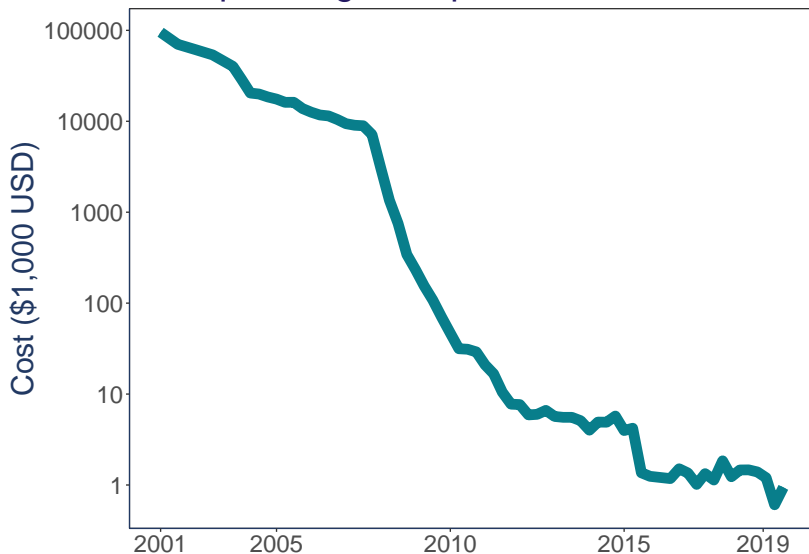
adapted from: Kris Wetterstrand, National Human Genome Research Institute

# EMBL Sequence Data

Data from: EMBL

All that sequence data would equal about

# 11,066,077,343,692 GB

$$\approx \textbf{512GB}$$

22 BILLION LAPTOPS

Bioinformatics to the rescue!

Bioinformatics is taking **biological data, processes and theories** and **applying** "informatics" techniques (derived from disciplines like **math, computer science, and statistics**) to understand, organize, and predict biological processes.

1. Data Analysis
2. Software Development
3. Modeling
4. Combination

1. Data Analysis
2. Software Development
3. Modeling
4. Combination

# Generating, interpreting, and explaining any biological data

# 1. Data Analysis

## Building the Human Genome



1989: The Banbury meeting at Cold Spring Harbor Laboratory in New York before the launch of the Human Genome Project.

# The Human Genome is Sequenced!

International Human Genome Sequencing Consortium Announces
"Working Draft" of Human Genome, June 2000

# 1. Data Analysis

The Human Genome is Sequenced...Again!

The Human Genome is Sequenced...Again!



February 2001

# 1. Data Analysis

The Human Genome is Sequenced...yet AGAIN!

The Human Genome is Sequenced...yet AGAIN!

# **The complete Human Genome is announced by NHGRI**

# But is the human genome really "Complete"?

1. Data Analysis
2. Software Development
3. Modeling
4. Combination

# Creating bioinformatics tools for other people to use

# 2. Software Development

`progressiveMauve`: Whole genome alignment program

# 2. Software Development

`progressiveMauve`: Whole genome alignment program



Darling et al. 2010

# 2. Software Development

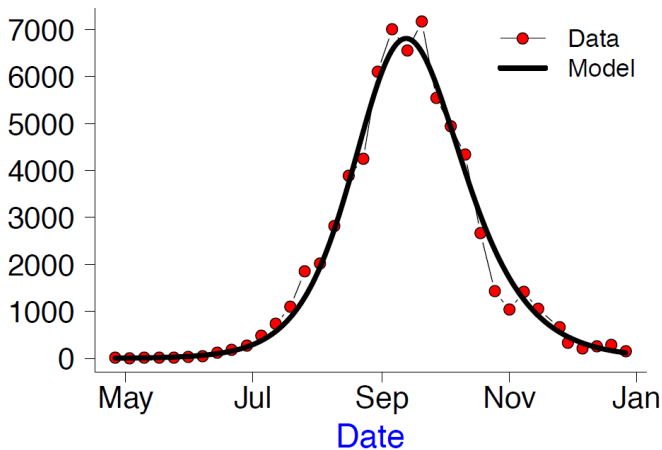`progressiveMauve`: Whole genome alignment program



Darling et al. 2010

1. Data Analysis
2. Software Development
3. Modeling
4. Combination

# Using mathematical and statistical principals to represent and predict biological systems or data
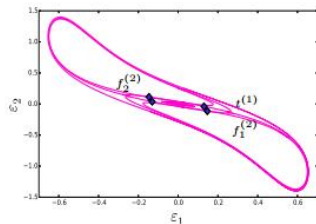
# 3. Modelling

SIR model and the Great Plague of London
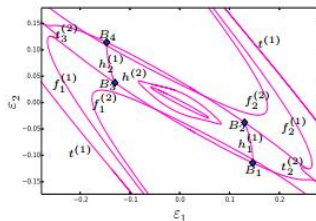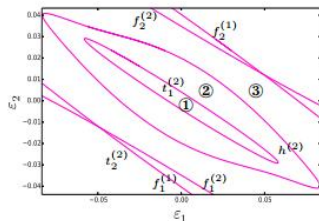


Weekly Deaths from Plague

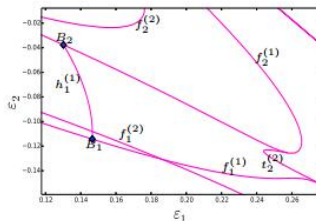# 3. Modelling

## Bifurcation theory and predator-prey relationships



Li et al. 2018, from Gail Wolkowicz's lab at McMaster University

1. Data Analysis
2. Software Development
3. Modeling
4. Combination

# 4. Combination

- Modelling + Software Development
- Data Analysis + Modelling
- Data Analysis + Software Development
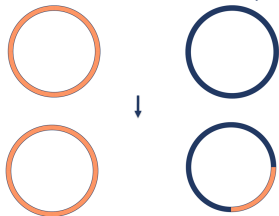- Data Analysis + Modelling + Software Development

# Spatial Patterns of Molecular Trends in Bacterial Genomes

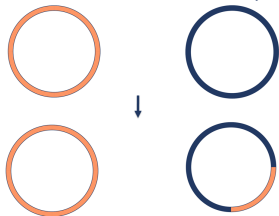(How do molecular trends change with position in the genome?)

# Bacteria are bizarre!
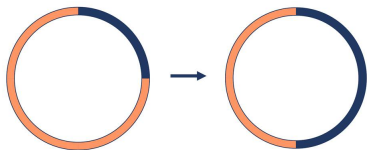
# Bacteria are bizarre!

**H**orizontal **G**ene **T**ransfer (**HGT**)

# Bacteria are bizarre!

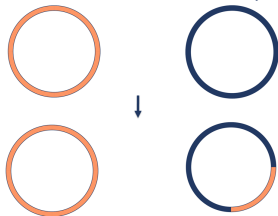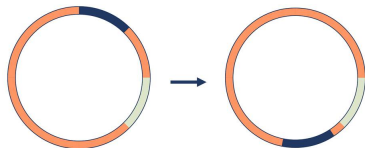**H**orizontal **G**ene **T**ransfer (**HGT**)

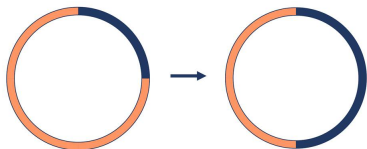

Duplication

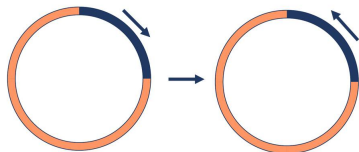# Bacteria are bizarre!



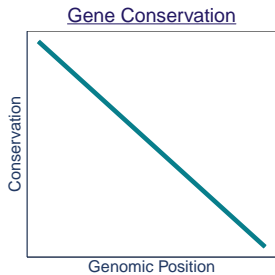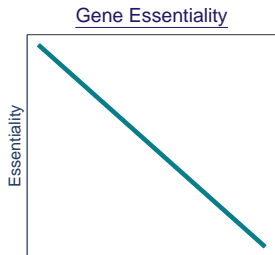Horizontal Gene Transfer (HGT)

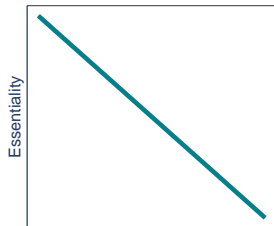Rearrangement and Translocation

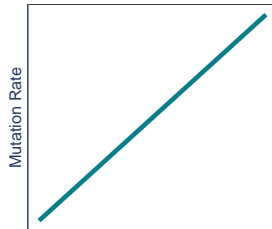Duplication

Inversion

# My Research: Spatial molecular trends

# My Research: Spatial molecular trends



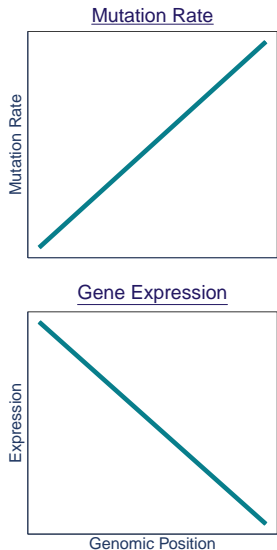Couturier et al. 2006, Cooper et al. 2010, Sharp et al. 2005, Morrow et al. 2012, Cooper and Rocha 2006

# My Research: Spatial molecular trends



### Mutation Rate

### Gene Expression

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

Bacteria:

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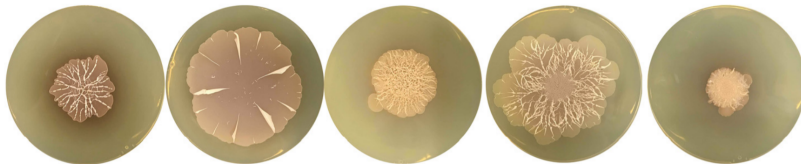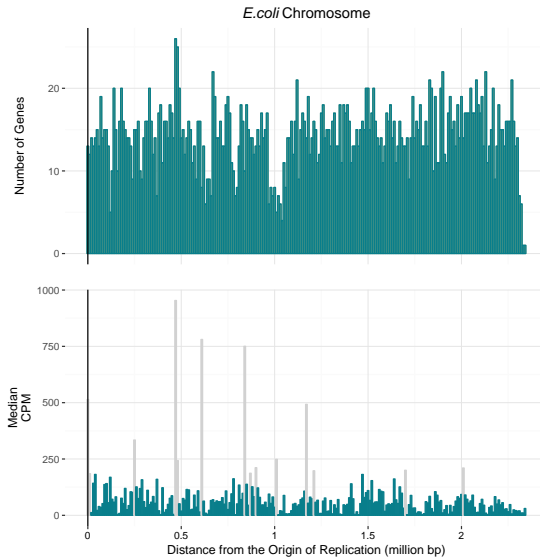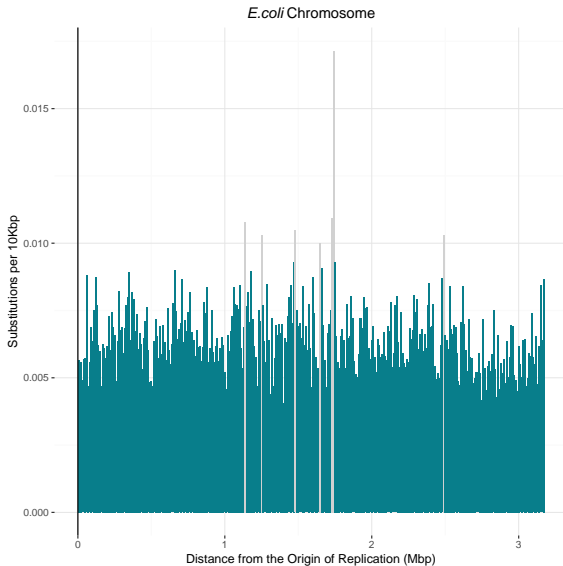- *Streptomyces*
- *Sinorhizobium meliloti*



Photo: *Streptomyces* by Stephanie Jones, Marie Elliot's Lab at McMaster University

# My Research: Gene Expression



*E.coli* Chromosome

33

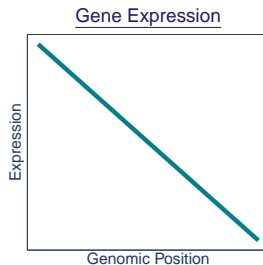# My Research: Substitutions



*E.coli* Chromosome

**Previous Studies:**

Mutation Rate

Gene Expression

**My Research:**

Substitutions

Gene Expression
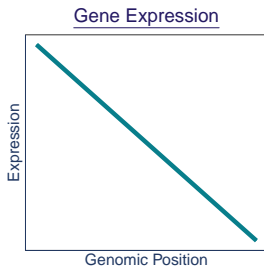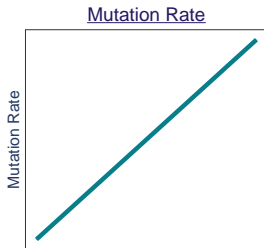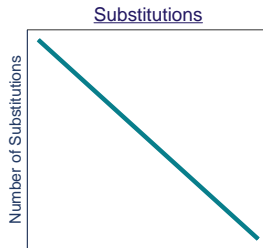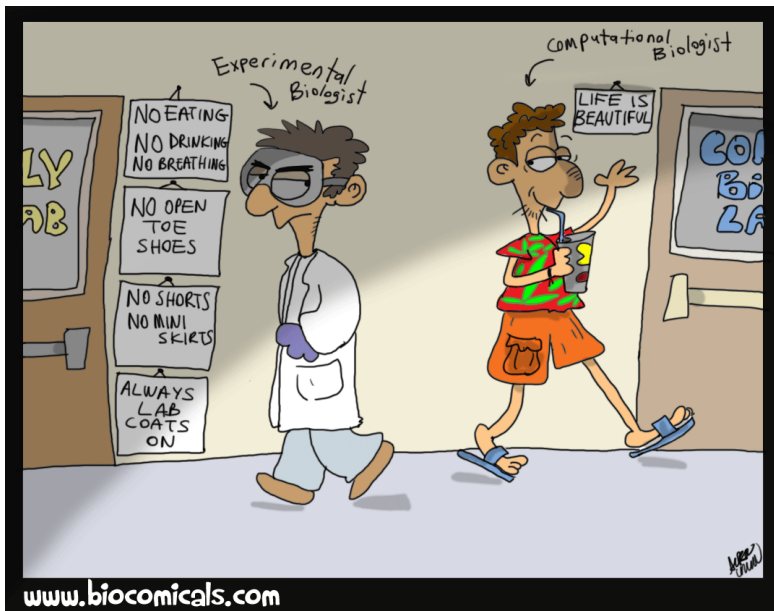
- **Bio 3S03:** Intro to Bioinformatics
- **Bio 3SS3:** Population Ecology
- **Math 4MB3:** Mathematical Biology
- **Math 3MB3:** Introduction to Modelling
- suggestions??

# Questions?

latodf@mcmaster.ca