

A model for assessing the concordance between genetic and epidemiological similarity of *Campylobacter* isolates: Towards improved application of genomic epidemiology in public health

**Benjamin Hetman, MSc. Candidate (University of Lethbridge)
Supervisor: Dr. Eduardo Taboada**



PROTECTING CANADIANS FROM ILLNESS

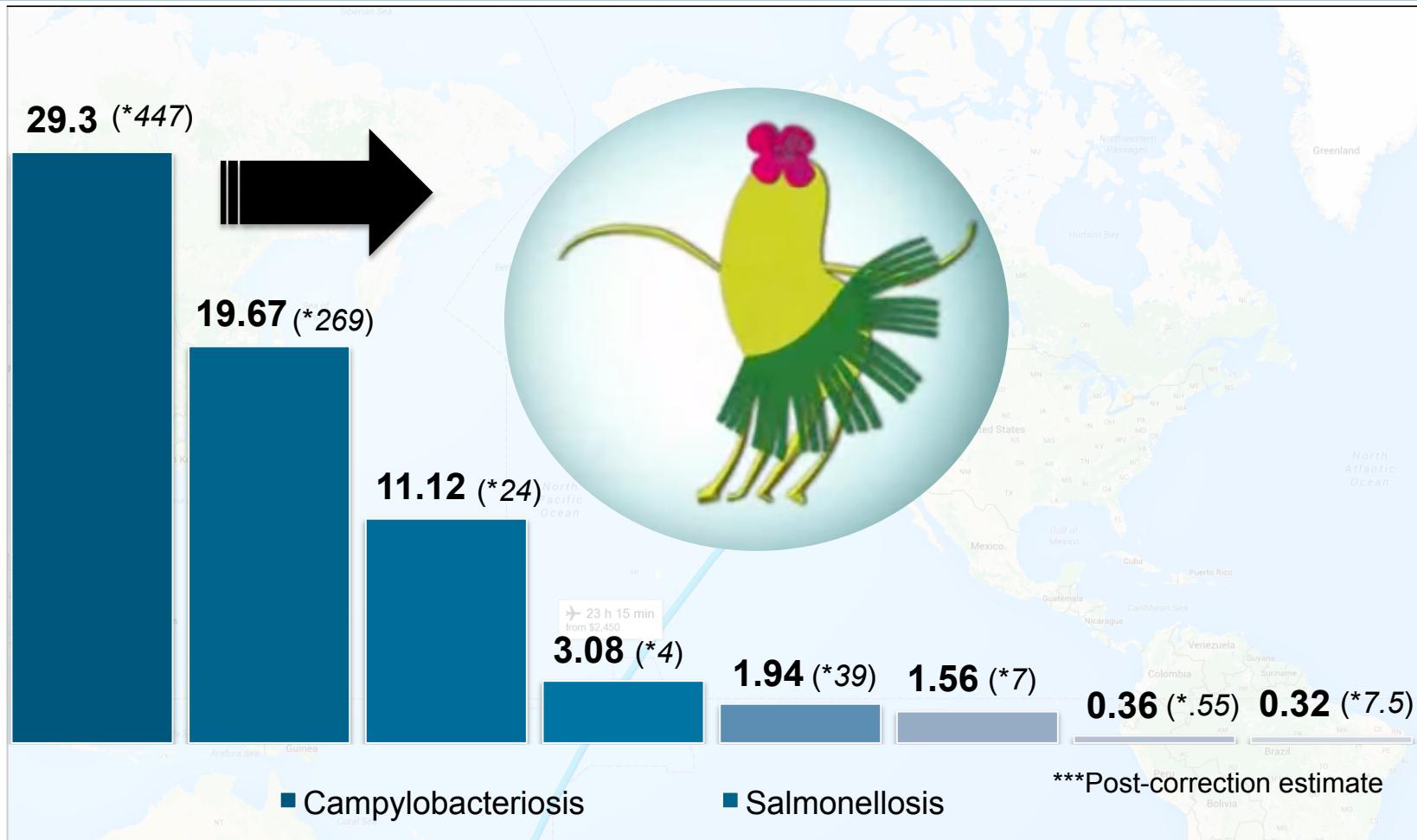


Public Health
Agency of Canada

Agence de la santé
publique du Canada

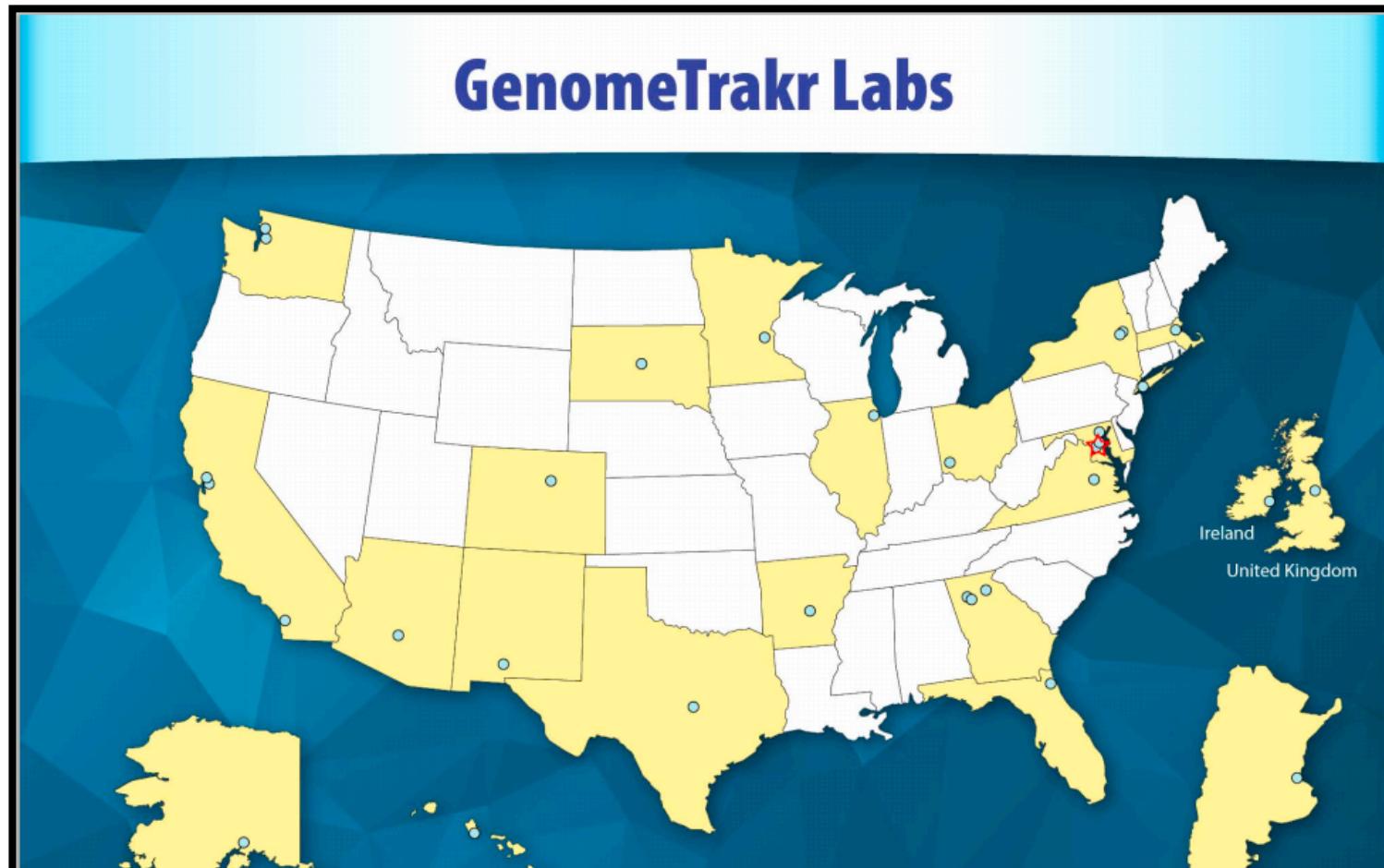
Canada

Campylobacteriosis in Canada: Cases per 100,000



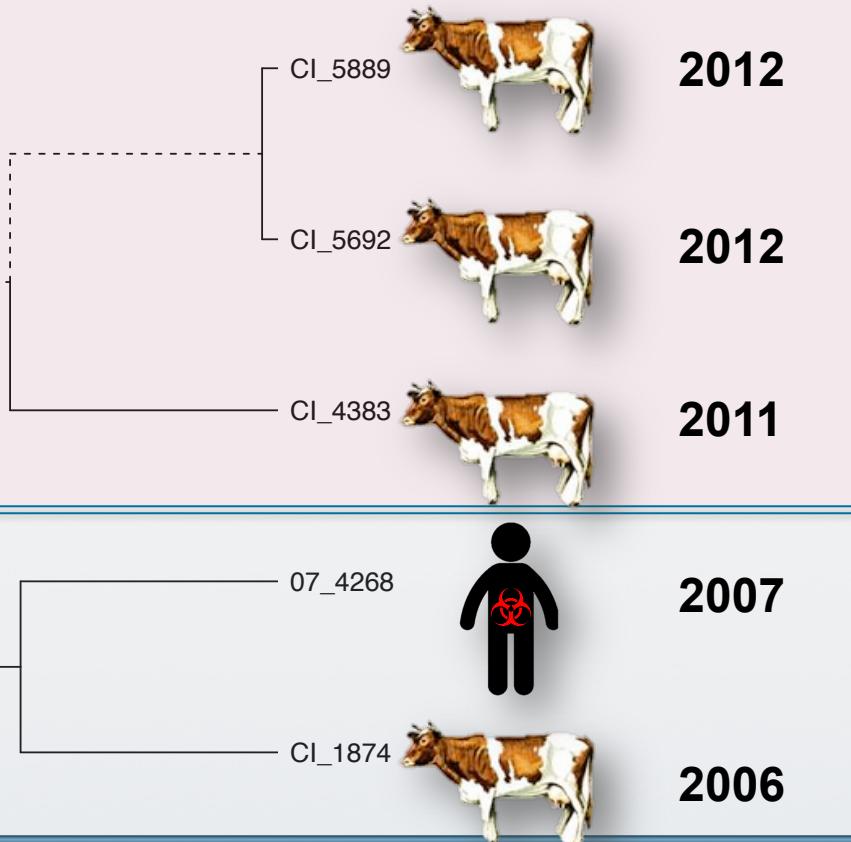
- #1 bacterial gastrointestinal disease in Canada
- Highly underreported (~\$350M annual economic burden)
- Sporadic disease: difficult to identify routes of transmission

WGS for Routine Pathogen Surveillance:



- Public Health Laboratories worldwide are transitioning to a WGS-based approach to routine surveillance
 - WGS has replaced many older, molecular based techniques
 - **WGS can be adapted to low – or – high resolution analyses**

A tale of two resolutions: Scenario 1



- Five CJ genomes that are *indistinguishable* by MLST
- Start examining the epidemiology:
 - **Strong human/cattle link!**
 - Temporal components don't exactly add up...
- Assess at level of cgMLST:
 - **Groups split at temporal divide**
- Increasing the resolution provided enhanced understanding of the relationship between isolates

- Using increased resolution can help match up the genomic signal with the underlying epidemiology

A tale of two resolutions: Scenario 2

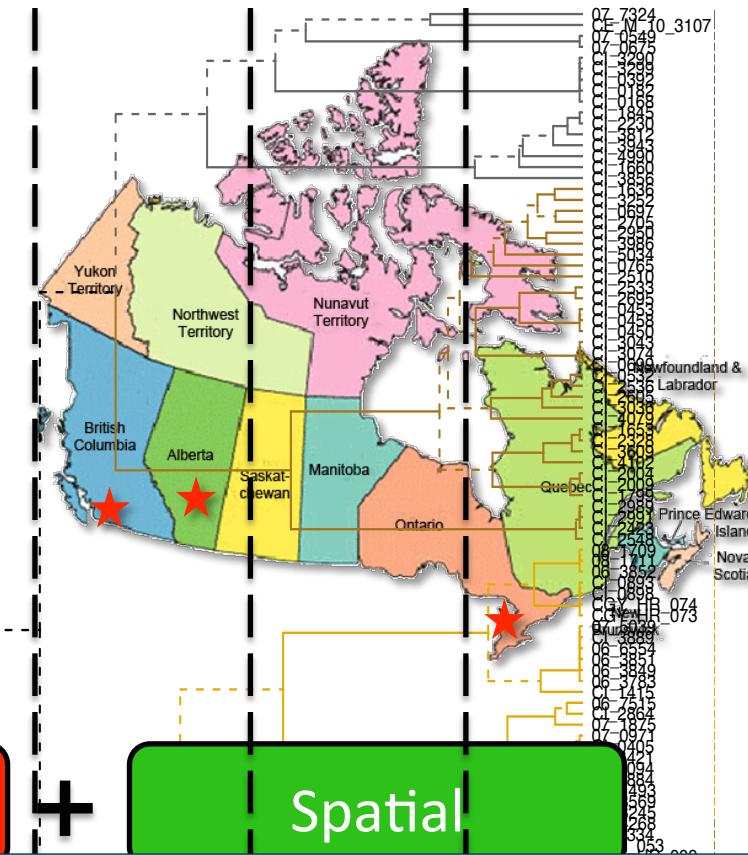


08_4461
08_4460
08_4466
-
08_4468
08_4472
08_4474

- Eight genomes split individually using cgMLST
- Examine the Epidemiology:
 - Same **source**
 - Same **location**
 - Same **date**
- Isolates from confirmed outbreak: ON 2008

- WGS based analyses are better but we still require knowledge of the epidemiology to cluster genomic data into relevant groupings
- What if we could calibrate our analysis of WGS data that ensured our results were high resolution and epidemiologically relevant?

Summarizing CJ Epidemiology



Source

+

Temporal

+

Spatial

- Establish a metric that summarizes the relationships between isolates of CJ based on their epidemiologic metadata
- Calculate an epidemiologic similarity network between CJ isolates
- Directly compare the epidemiology of CJ isolates with their genomic signal

Workflow Overview:

WGS Workflow



Canadian CGF DB

Isolate Selection



Source



Location



Date

Epi Workflow



Sequencing



Assembly



Annotation



Core Analysis

MIST

In-Silico Typing

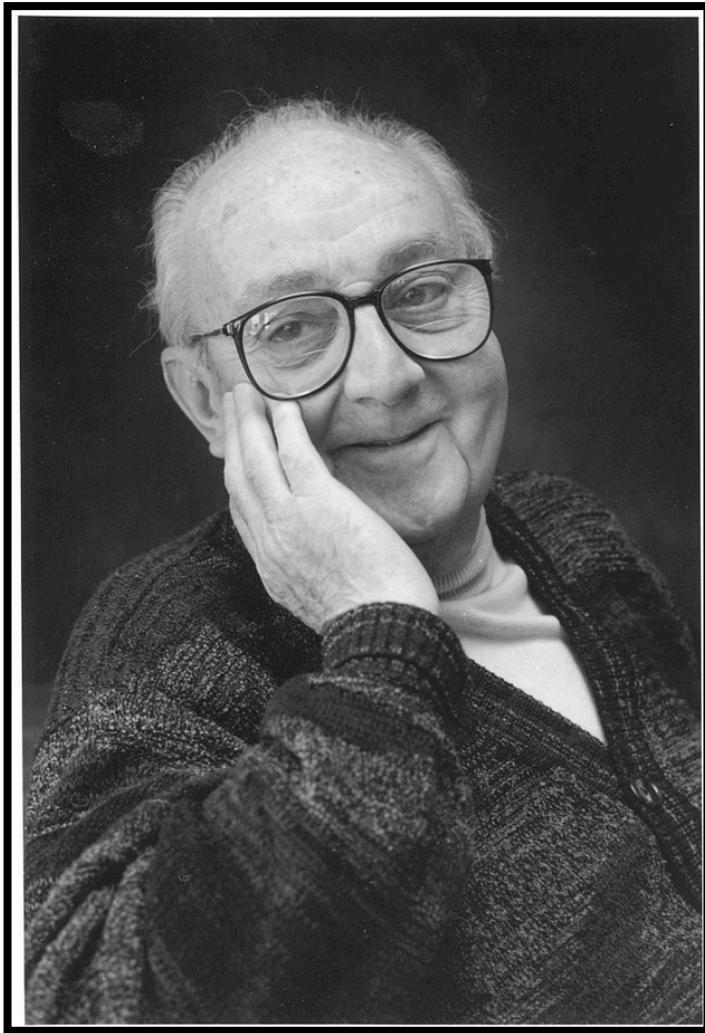


Cluster Analysis



- ❑ Need a model to assess relationships based on isolate epidemiology so we can directly compare them against the WGS data

Building a model for quantifying metadata similarity



***“Essentially, all models are wrong,
but some are useful.”***

**George E.P. Box
(1919-2013)**

Quantifying epidemiological relationships:

Given any two *Campylobacter* isolates with epidemiological metadata, the epidemiologic similarity (Epi_{sym}) between the two can be quantified by:

Source_{A→B}

+

Spatial_{A→B}

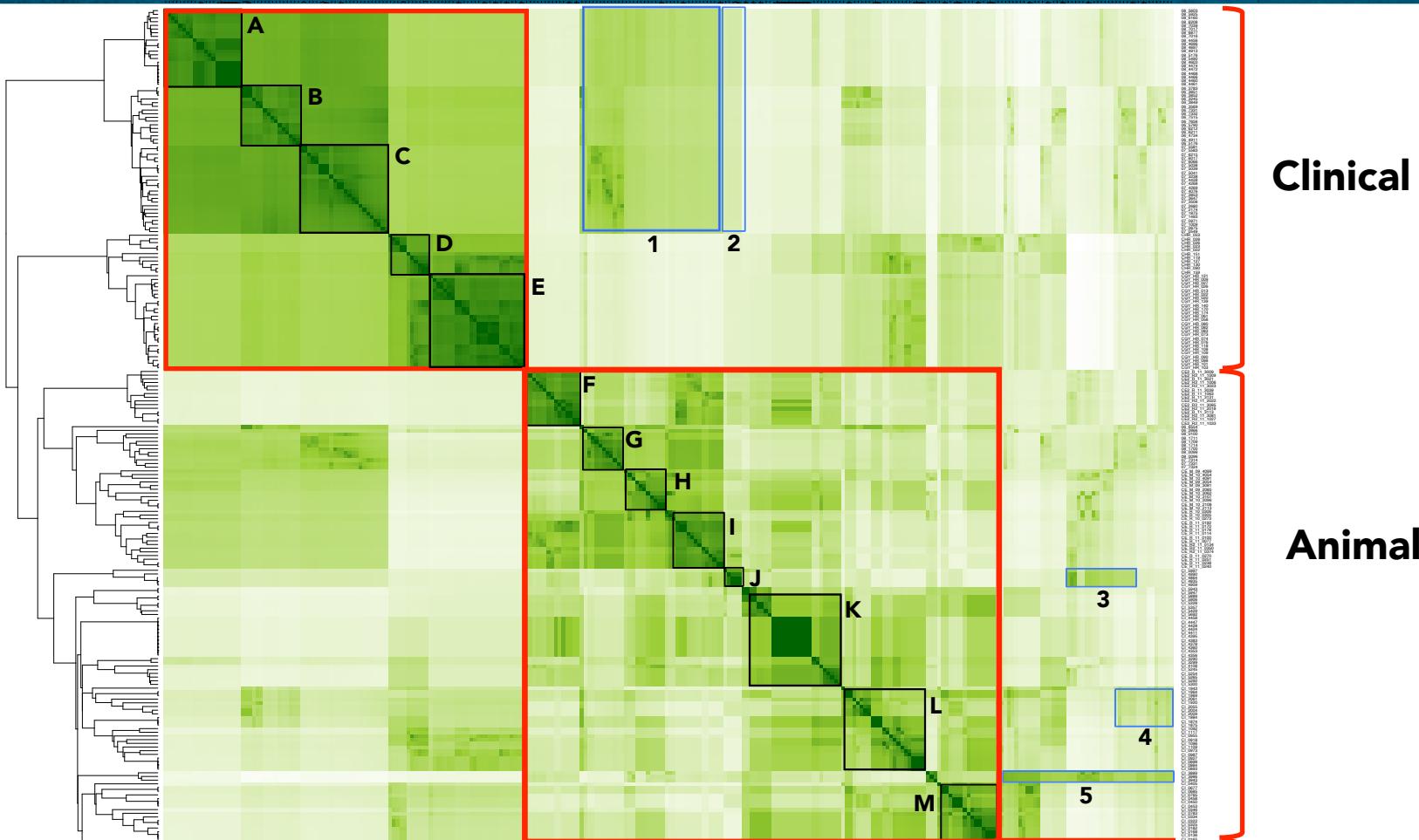
+

Temporal_{A→B}



- Putting all 3 components together, we can now quantifiably cluster *Campylobacter* isolates based on their epidemiological relationships

Results: Epidemiological Clustering of *C. jejuni*

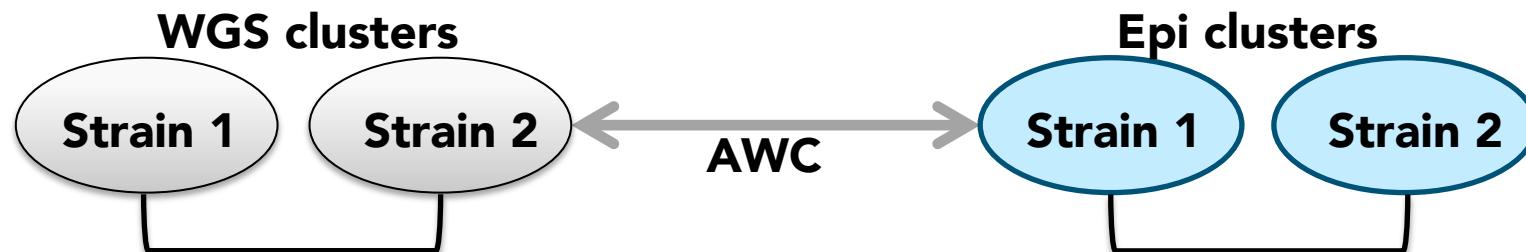


- ☐ Clusters of secondary heat correspond to isolates with similar geography and temporal data, but different sources

How to relate epidemiologic and genomic clustering:

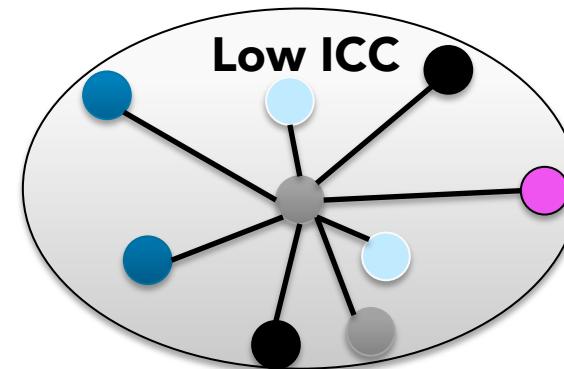
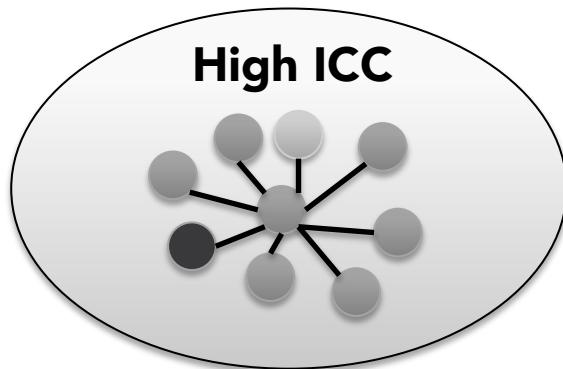
1. Adjusted Wallace Coefficient: (AWC)

The directional likelihood that two isolates clustered together using one method will be grouped together in the second method

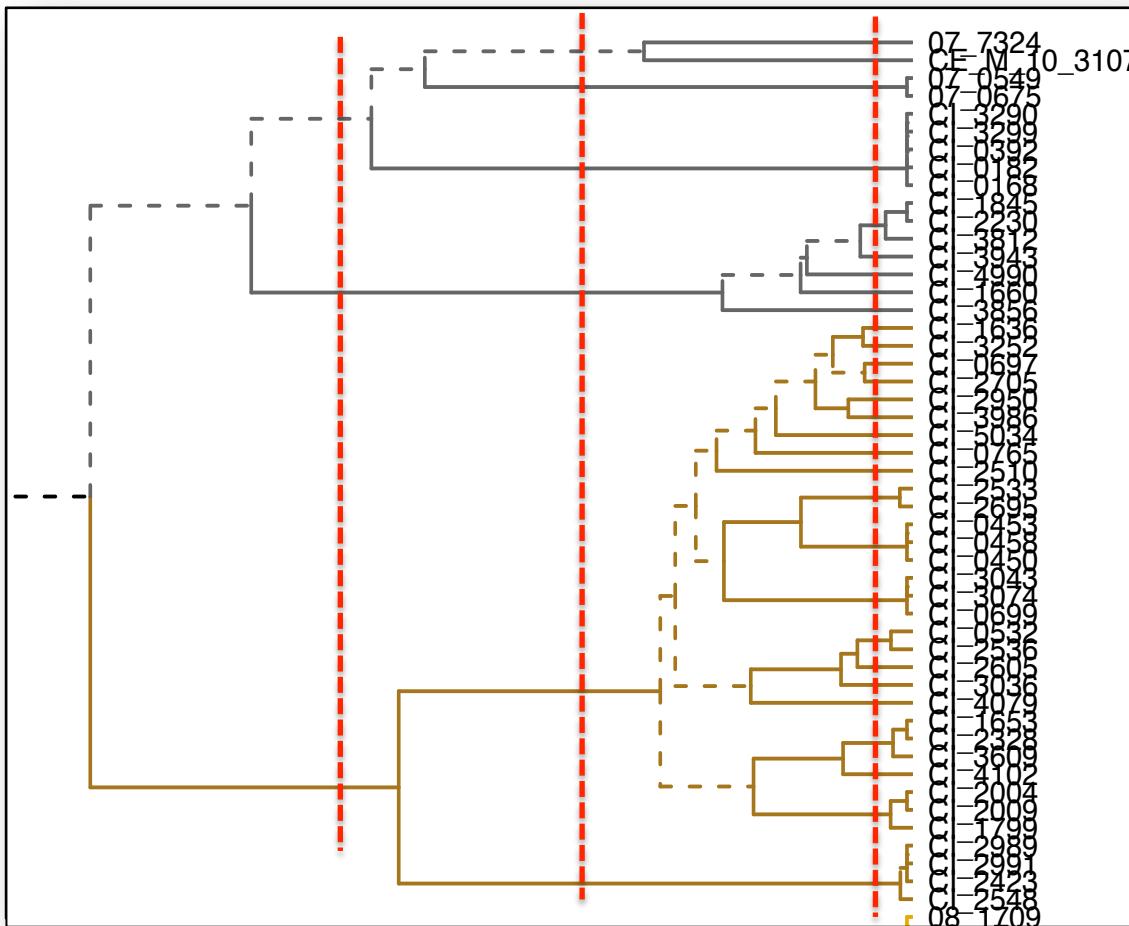


2. Intra-cluster cohesion: (ICC)

- A measure of the genomic and epidemiologic homogeneity of the bacterial isolates within a cluster

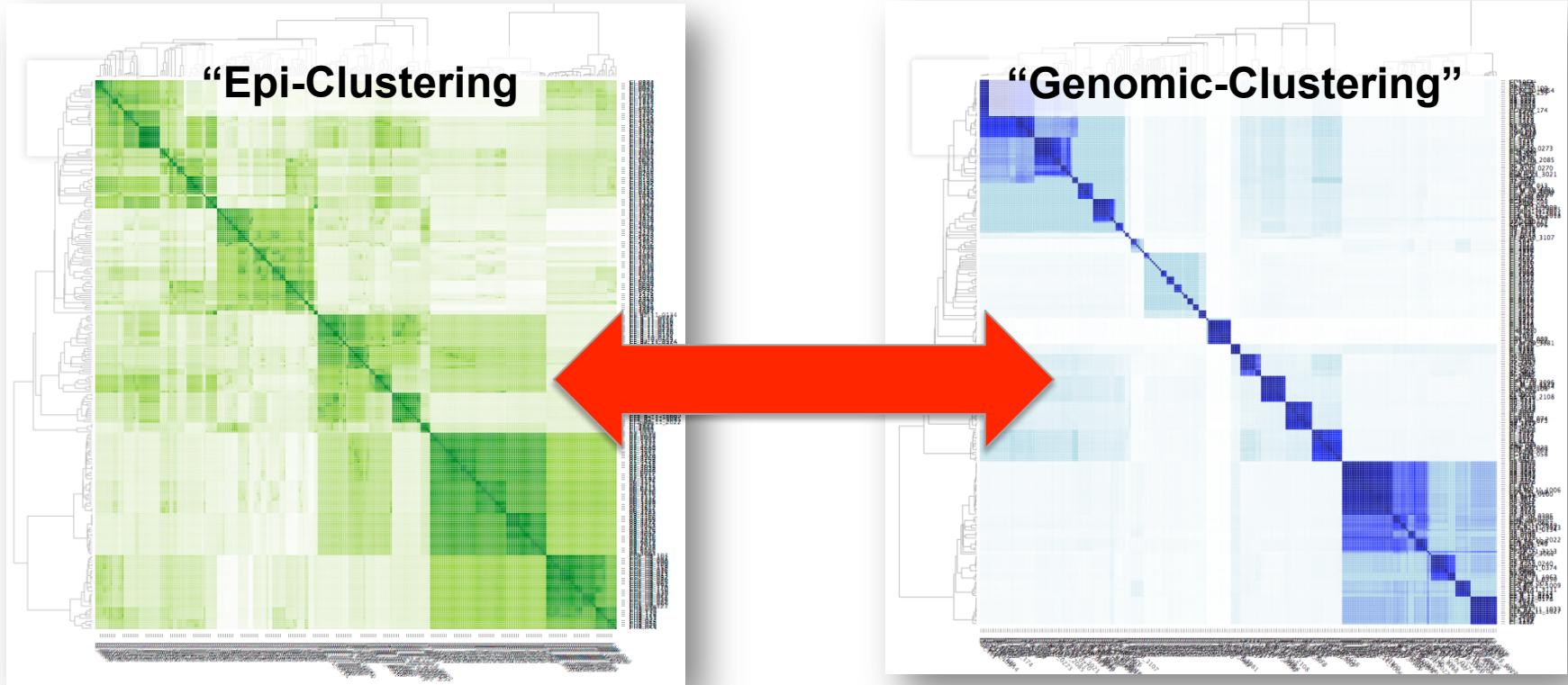


Calibrating WGS methods for epidemiologic investigations



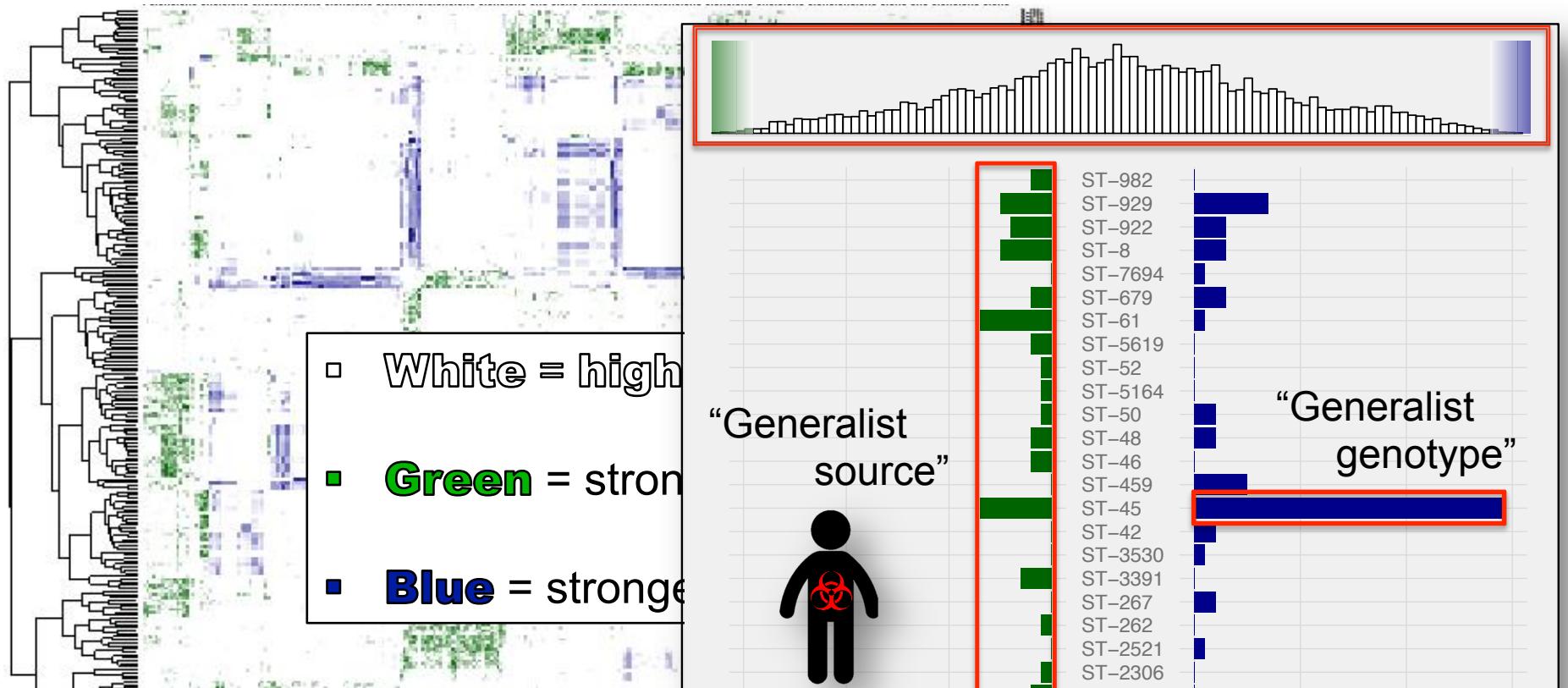
- Genomic cluster homogeneity
- Epidemiologic cluster homogeneity
- Calculate point of highest genomic-cohesion while maintaining
 - a) Multi-isolate clusters
 - b) High epidemiologic validity

Outlier clustering analysis: Epi vs. WGS clustering



- Strains with similar epidemiology don't necessarily have to be similar genetically (and vice-versa!)
- By overlaying the two methods, we can identify clusters that group together significantly stronger via genomic or epidemiologic relationships

Identifying Epidemiologic and Genomic Outliers



- ‘Generalist’ genotypes → CJ that persist across many source, temporal and spatial ranges
- ‘Generalist’ epidemiologic reservoirs → support the persistence of a broad range of genotypes

Summary:

- Developed a model to help **guide our analysis of *Campylobacter* WGS data for practical public health purposes**
- Epidemiologically relevant clustering thresholds for genomic data allow for targeted research approaches
- Interactive web application under development (Check it out! <https://hetmanb.shinyapps.io/EpiQuant/>)



Acknowledgements:



People

- **Supervisors:**
Ed Taboada + Jim Thomas
- **Lab members:**
Steven Mutschall (PHAC)
Peter Kruczkieicz (PHAC)
Dillon Barker (PHAC/ULeth)

Funding

- **University of Lethbridge**
- **Public Health Agency of Canada A-base**
- **Gov't of Canada: Genomics Research and Development Initiative**



Public Health
Agency of Canada

Agence de la santé
publique du Canada