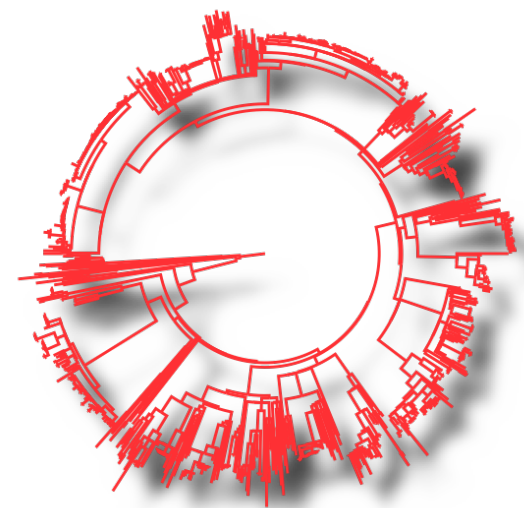
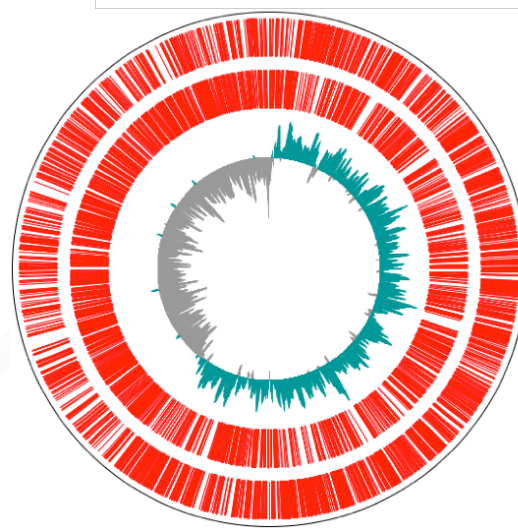
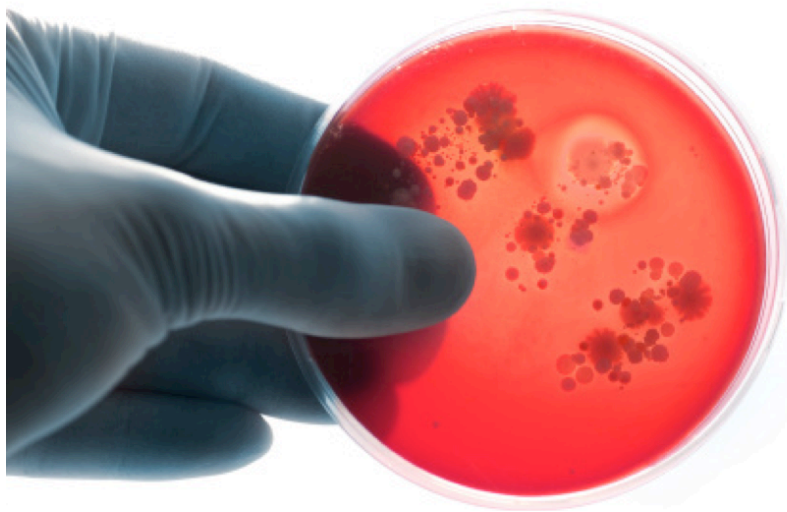


What Do We Need from Microbial Genomics Surveillance Software?

Kathryn Holt

www.holtlab.net - kholt@unimelb.edu.au - @DrKatHolt



Underlying question:

**What is the source?
(and how sure are we about that)**

- Assuming someone's already done the hard work of isolating and sequencing genomes for us to investigate...
- Compare to receiving a set of PFGE gel images & sample info
- Rule in/rule out are both important

<http://github.com/katholt/NGSchallenge/>

Methodological approaches

- Which bit of the genome are we looking at?
 - Whole genome? Genes? Core genes?
- SNPs? MLST? Kmers?
- Reads or assemblies? Do we need a reference?
- Output trees or distances?
- Confidence values?

#ASMNGS

**⇒ Not so important as long as it's
useable, reproducible, interpretable & accurate**

Useable

- **Do I have to be a...**

- Computer programmer?
- Bioinformatics grad student?
- Epidemiologist?
- ⇒ Aiming for intuitive automated execution

- **Do I have to have...**

- A laptop? Compute cluster? Cloud?
- Minutes? Hours? Days?

- **Will it work with my...**

- Input standards (sample IDs, metadata, LIMS)
- Reporting requirements, data management, certification standards

Reproducible & robust

- **Will I get the same answer...**
 - With different reference / comparator genomes?
 - With different sequencing runs?
 - With different sequencing tech?
 - In another lab?
 - If someone else runs the analysis? (users make choices?)
- **What happens when I...**
 - Add new strains tomorrow?
 - Need to compare to growing historical/background data?
 - Sequence a mixed culture by accident?
 - Have too much data or too little?
 - Have low or variable quality data?
 - Need to investigate a different species?

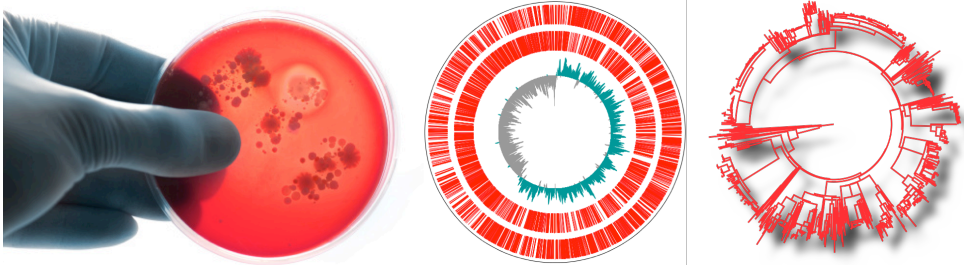
Interpretable

- **What's the output?** Do I need a degree in phylogenetics?
 - Trees? Distances? Rule in/out?
 - Visualization or reporting of results?
 - How is metadata / epi data incorporated?
 - Formats? How to store? How to share?
- **Decision points? Cut-offs?**
- **Confidence in result?**
 - Ruling out needs to be confident as well as ruling in
 - Known unknowns

Accurate?

- Will I make the right decision?
- How will I know?

<http://github.com/katholt/NGSChallenge/>



www.holtlab.net
kholt@unimelb.edu.au
[@DrKatHolt](https://twitter.com/DrKatHolt)

Pathogen Surveillance Software Demonstration: Listeria Monocytogenes & Salmonella Enteritidis Data

**FDA/CFSAN/ORS Division of Microbiology
George Kastanis, Dwayne Robertson, and Charles Wang**

**Errol Strain, Marc Allard, Kathryn Holt, Bill
Klimke**

9/26/2015

***Listeria monocytogenes* (n=16+2)**

Multistate Outbreak of Listeriosis Linked to Roos Foods Dairy Products (Final Update)

Posted April 18, 2014 9:30 AM ET

This outbreak appears to be over. *Listeria monocytogenes* infection (listeriosis) is an important cause of illness in the United States. More information about listeriosis, and steps people can take to reduce their risk of infection, can be found on the [CDC *Listeria* Web Page](#).

Highlights

- [Read the Advice to Consumers & Cheese Retailers»](#)
- A total of eight persons infected with the outbreak strain of *Listeria monocytogenes* were reported from two states.
 - The number of ill persons identified in each state was as follows: California (1) and Maryland (7)

At a Glance:

- Case Count: [8](#)
- States: [2](#)
- Deaths: **1**
- Hospitalizations: **7**
- Recall: [Yes](#)

LM Questions

1. Do the product isolates from facility #1 match the environmental swabs from facility #1? **(Yes)** Do the product isolates match any other food/environmental isolates currently in the NCBI/SRA database under BioProjects PRJNA212117 or PRJNA215355? **(Yes, multiple isolates from swabs, cheese, and cheese curd)**

2. Do the environmental/product isolates from either facility match clinical isolates in the BioProjects listed in question 1, or any other clinical *Listeria monocytogenes* isolates available from other public data sources? **(Yes, 4 clinical isolates)** Can you identify the clinical cases associated with this contamination event?

CFSAN007567 2013-11-13 USA:CA swab

ASM_52

CFSAN007549 2013-11-13 USA:CA swab

ASM_51

NYAG12B11863-1 2012 USA:NY spanish style cheese

PNUSAL000355 missing USA missing

CFSAN010072 2014-02-20 USA:MD cheese

CFSAN011016 2014-02-28 USA:MD fresh cheese curd

VA-WGS-00234 2014-02-11 USA:VA fresh cheese curd

CFSAN011018 2014-02-28 USA:MD fresh cheese curd

PNUSAL000571 2013-08-09 USA Blood

VA-WGS-00228 2014-02-11 USA:VA fresh cheese curd

CFSAN010075 2014-02-20 USA:MD cheese

PNUSAL000570 2013-08-14 USA Blood

VA-WGS-00241 2014-02-11 USA:VA fresh cheese curd

VA-WGS-00239 2014-02-11 USA:VA fresh cheese curd

VA-WGS-00235 2014-02-11 USA:VA fresh cheese curd

CFSAN010076 2014-02-20 USA:MD cheese

VA-WGS-00236 2014-02-11 USA:VA fresh cheese curd

FDA00004369 2014-05-27 USA:DE environmental swab

PNUSAL000517 2013-10-14 USA Stool

VA-WGS-00240 2014-02-11 USA:VA fresh cheese curd

PNUSAL000569 missing USA missing

VA-WGS-00222 2014-02-11 USA:VA fresh cheese curd

CFSAN010074 2014-02-20 USA:MD cheese

CFSAN010077 2014-02-20 USA:MD cheese

VA-WGS-00007 2014-02-11 USA:VA fresh cheese curd

***Salmonella* Enteritidis (n=50)**

Multistate Outbreak of Human *Salmonella* Enteritidis Infections Associated with Shell Eggs (Final Update)



NOTICE: This outbreak is over. The information on this page has been archived for historical purposes only and will not be updated.

Posted December 2, 2010

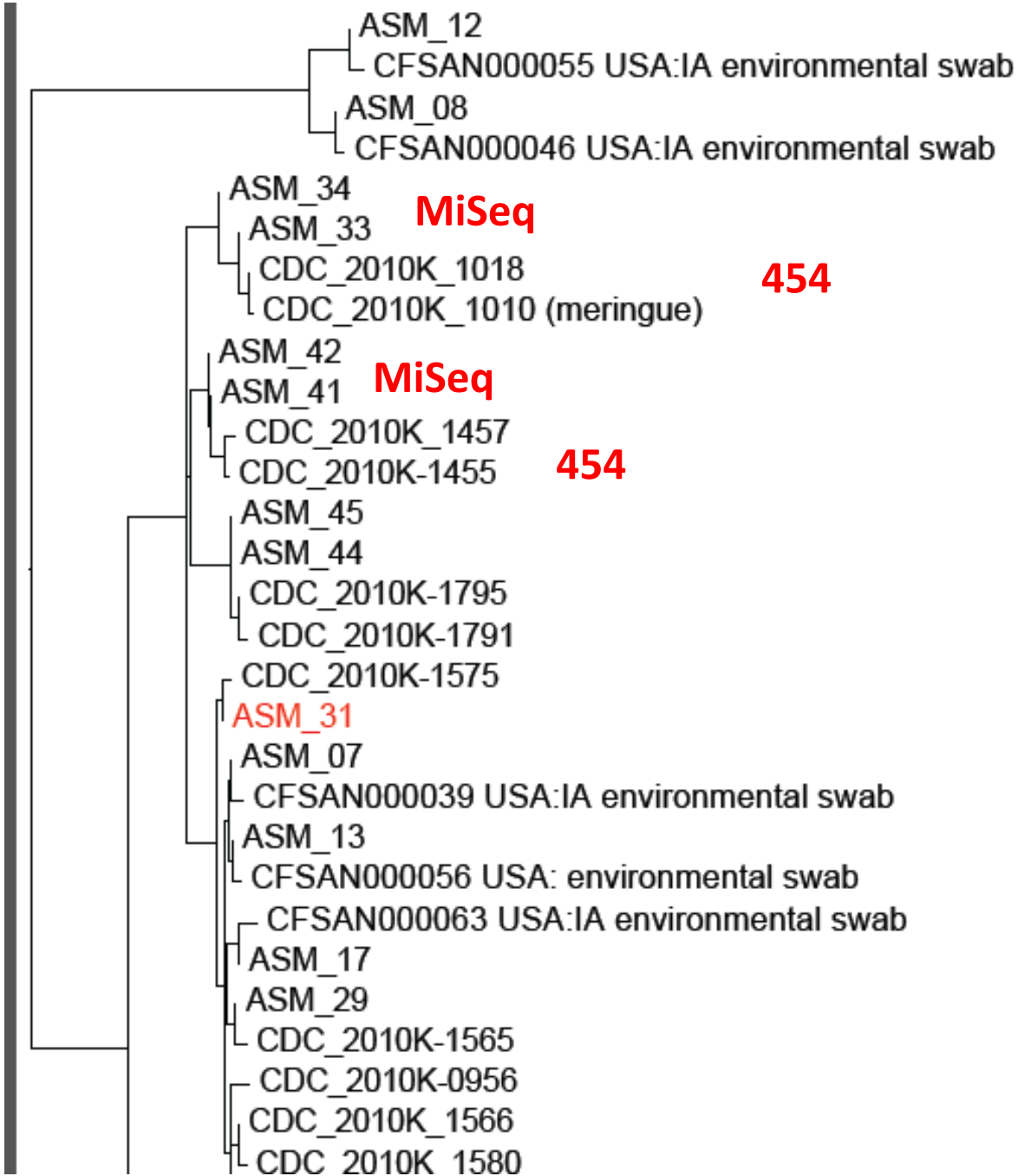
Outbreak Summary

CDC collaborated with public health officials in multiple states, the U.S. Department of Health and Human Services' Food and Drug Administration (FDA), and the U.S. Department of Agriculture's Food Safety and Inspection Service to investigate a nationwide increase of *Salmonella* Enteritidis (SE) infections with an indistinguishable pulsed-field gel electrophoresis (PFGE) pattern JEGX01.0004. This is the most common PFGE pattern for SE in the PulseNet database. Investigators used DNA analysis of SE bacteria obtained through diagnostic testing to identify cases of illness and restaurant or event clusters (where more than one ill person with the outbreak strain had eaten) that may have been part of this outbreak. Because the SE PFGE pattern commonly occurs in the United States, some of the cases identified may not have been related to this outbreak.

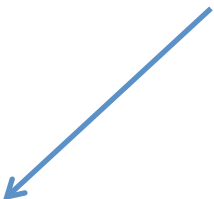
"... approximately 1,939 reported illnesses that are likely to be associated with this outbreak."

Questions

1. Do the 4 clinical isolates (ASM_26, ASM_31, ASM_49, and ASM_50) that are epidemiologically linked to eggs from state #2 (#1 OH, #2 IA) match any of the environmental or food swabs collected at those facilities? (Yes – ASM_31: IA, ASM_49 & 50 match IA & OH. ASM_26?)
2. Do any of the remaining 19 clinical isolates match clinical isolates from question #1? (Yes, but 3 different clusters) Do they match any of the food or environmental isolates? (Yes) Are there additional clinical clusters? (Yes, 5-6+)
3. Are there other clinical Salmonella isolates in public databases (including, but not limited to, samples from BioProjects PRJNA237212, PRJNA227458, PRJNA252015, and PRJNA230403) that match food or environmental isolates collected at these facilities in these two states? (Yes, maybe newer isolates?)



ASM_18
 ASM_37
 ASM_40
 ASM_26
 ASM_39
 ASM_25
 ASM_38
 CDC_2010K_0899 (mexican meal)
 CDC_2010K_1884 (rattlesnake cake)
 CDC_2010K_1882
 CDC_2010K-0968
 CDC_2010K_0895
 ASM_23
 ASM_06
 ASM_22
 ASM_24
 ASM_11
 CFSAN035272 2013 USA:TX product-eggs-raw-whole
 ASM_09
 FCC0177 2010-09-27 USA:OH swab
 ASM_03
 FCC0174 2010-09-23 USA:OH swab
 FCC0183 2010-10-05 USA:OH environmental swabs
 FCC0182 2010-10-05 USA:OH environmental swabs
 ASM_04
 FCC0173 2010-09-23 USA:OH swab
 ASM_16
 FCC0185 2010-10-05 USA:OH environmental swabs
 ASM_05



2013