Implementing Whole Genome Cluster Analysis to Aid in *Salmonella* Outbreak Investigations

William Wolfgang
Wadsworth Center NYSDOH
APHL
06/01/14

CDC statistics on Foodborne Illness in America

Each year

- 1 in 6 (48 million) get sick.
- 128,000 are hospitalized
- 3,000 die
- For about 60% the cause is unknown.

In 2011.

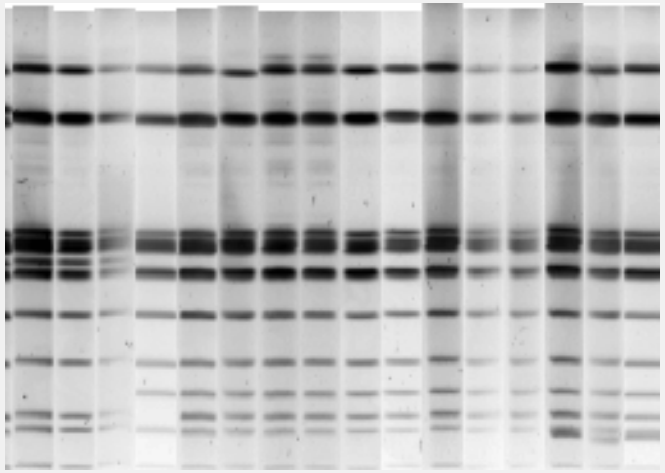
- *Salmonella* accounted for 1 million cases.
- 19,000 hospitalizations and 378 deaths.
- Rates of *Salmonella* infections are increasing.

Surveillance of *Salmonella* in New York State

In New York all **positive patient specimens** are submitted to the Wadsworth Bacteriology Laboratory.

Bacteriology receives ~1,800 *Salmonella* patient specimens each year.

- Serotyped.
- DNA is fingerprinted by PFGE.



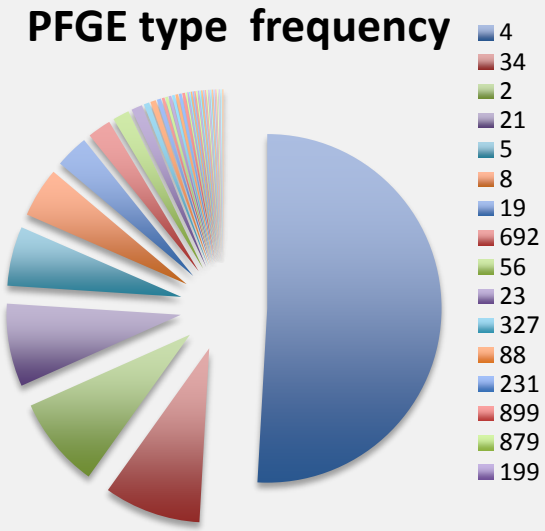
All data is sent to the CDC.

PFGE data bases are monitored:

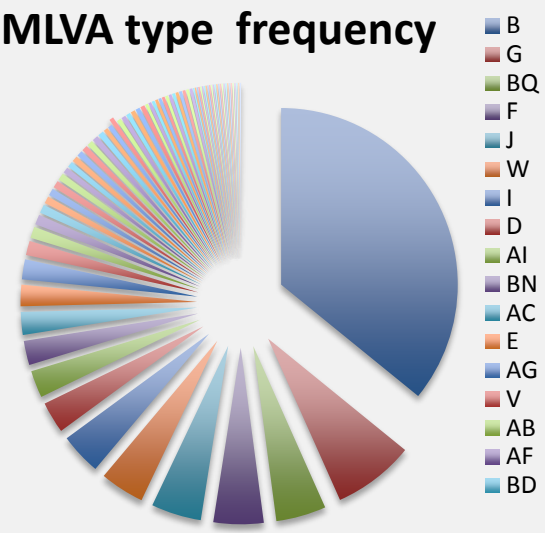
- To detect outbreaks in the patient population.
- To find source of the outbreak.

PulseNet International
The International Molecular Subtyping Network for Foodborne Disease Surveillance
USA, Canada, Europe, Asia Pacific, Latin America, Middle East
CDC

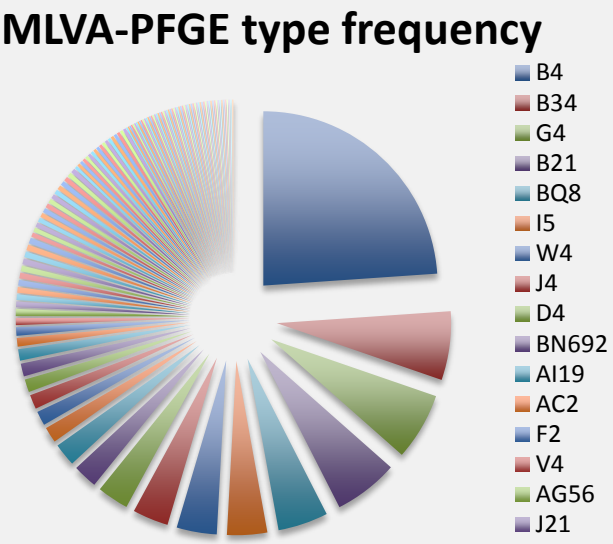
For *Salmonella* Enteritidis PFGE typing methods are poor at resolving clusters



52 PFGE types



98 MLVA types



163 combined MLVA-PFGE types

Thus *Salmonella* Enteritidis outbreaks are rarely detected

2010 Outbreaks

- Montevideo - salami.
- Typhimurium – long term care facility.
- Javiana - tomatoes.
- Saintpaul - restaurant.
- **Enteritidis** - long term care facility (4).

2011 Outbreaks

- Typhimurium - aquatic frogs.
- Heidelberg - ground turkey.
- **Enteritidis** - food worker at a deli (692).
- Typhimurium - ground beef.
- **Enteritidis** -Turkish pine nuts (8).

2012 Outbreaks

- Bareilly - sushi.
- Nchanga – sushi.
- Hartford - sub shop worker.
- Newport - chick and duck exposure.
- SanDiego & Poona - small turtles.
- Javiana - Mothers Day fruit baskets.
- **Enteritidis** - Cargill ground beef (9).

Retrospective studies have shown that WGS can improve pathogen tracking and surveillance.

- MRSA
- Tb
- Drug resistant *Klebsiella pneumoniae*
- Cholera outbreak in Haiti.
- *Salmonella* Montevideo outbreak from pepper.

Applied and Environmental Microbiology, Dec. 2011, p. 8648–8655 Vol. 77, No. 24
0099-2240/11/$12.00 doi:10.1128/AEM.06538-11
Copyright © 2011, American Society for Microbiology. All Rights Reserved.

A Whole-Genome Single Nucleotide Polymorphism-Based Approach To Trace and Identify Outbreaks Linked to a Common *Salmonella enterica* subsp. *enterica* Serovar Montevideo Pulsed-Field Gel Electrophoresis Type †

Henk C. den Bakker,[1]* Andrea I. Moreno Switt,[1] Craig A. Cummings,[2] Karin Hoelzer,[1] Lovorka Degoricija,[2] Lorraine D. Rodriguez-Rivera,[1] Emily M. Wright,[1] Rixun Fang,[2] Margaret Davis,[3] Tim Root,[4] Dianna Schoonmaker-Bopp,[4] Kimberlee A. Musser,[4] Elizabeth Villamil,[4] HaeNa Waechter,[5] Laura Kornstein,[5] Manohar R. Furtado,[2] and Martin Wiedmann[1]

Identification of a Salmonellosis Outbreak by Means of Molecular Sequencing

E. Kurt Lienau, Ph.D.
Errol Strain, Ph.D.
Charles Wang, B.S.
Jie Zheng, D.V.M., Ph.D.
Andrea R. Ottesen, Ph.D.
Christine E. Keys, M.S.
Thomas S. Hammack, M.S.
Steven M. Musser, Ph.D.
Eric W. Brown, Ph.D.
Marc W. Allard, Ph.D.
Food and Drug Administration
College Park, MD
marc.allard@fda.hhs.gov
Guojie Cao, M.S.
Jianghong Meng, D.V.M., Ph.D.
University of Maryland
College Park, MD
Robert Stones, M.S.
Food and Environment Research Agency
York, United Kingdom

Acknowledgments

- Cornell University
  Martin Wiedmann
  Henk den Bakker

- FDA
  Eric Brown
  Peter Evans
  Marc Allard
  Errol Strain
  Ruth Timme

- Connecticut DOH
  Stacey Kinney
  John Fontana

- Minnesota DOH
  David Boxrud
  Angie Jones

- NCBI
  Bill Klimke
  Martin Shumway

- Wadsworth Bacteriology Laboratory
  Kara Mitchell

- Wadsworth Genomics Core
  Matt Shudt
  Zhen Zhang
  Charles MacGowan
  Mark Rosenthal
  Melissa Leisner
  Danielle Loranger
  Mike Palumbo
  Pascal LaPierre

- Wadsworth PulseNet Lab
  Dianna Bopp
  Deb Baker
  Lisa Thompson

Cannoli outbreak

- Sept. 2010 Connecticut Dept. of Health identifies a *Salmonella* outbreak in a long term care facility.
- Outbreak was linked to cannoli from a Westchester bakery.
- Both NY and CT cases consumed cannoli's.

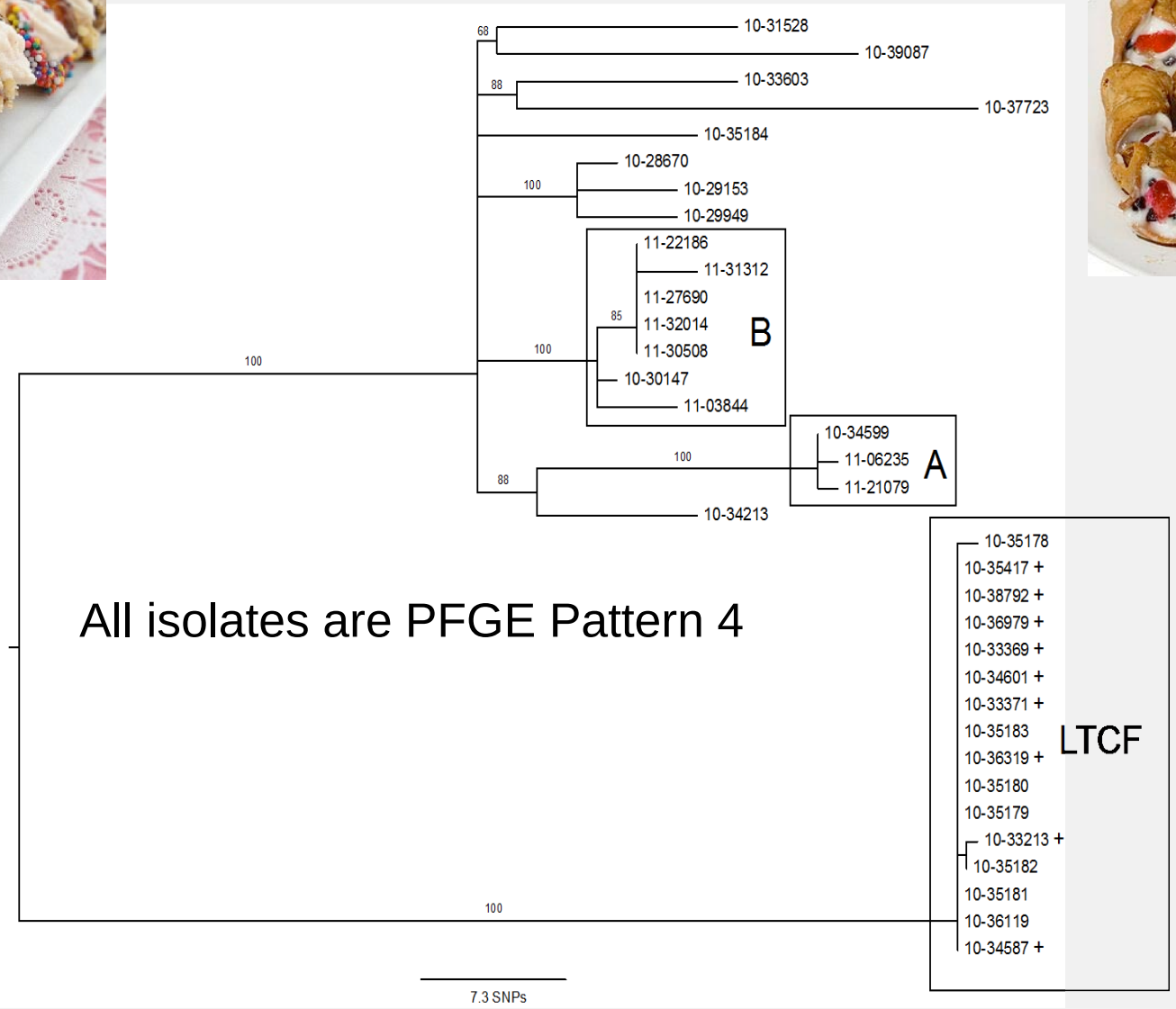
- Isolates had the most common PFGE pattern, JEGX01.0004.




Retrospective cohort

Key	County	Date	PFGE-MLVA Combined
IDR1000029153	Cattaraugus	8/10/10	JEGX01.0004W
IDR1000031528	Rockland	8/26/10	JEGX01.0004W
IDR1000033213	Putnam	9/10/10	JEGX01.0004W
IDR1000033369	Putnam	9/10/10	JEGX01.0004W
IDR1000033371	Putnam	9/11/10	JEGX01.0004W
IDR1000034601	Washington	9/13/10	JEGX01.0004W
IDR1000034587	Westchester	9/20/10	JEGX01.0004W
IDR1000035417	Putnam	9/22/10	JEGX01.0004W
IDR1000035178	Westchester	9/13/10	JEGX01.0004W
IDR1000035179	Greenwich CT	9/12/10	JEGX01.0004W
IDR1000035180	Westchester	9/12/10	JEGX01.0004W
IDR1000035181	Westchester	9/13/10	JEGX01.0004W
IDR1000035182	Westchester	9/12/10	JEGX01.0004W
IDR1000035183	Greenwich CT	9/16/10	JEGX01.0004W
IDR1000036119		9/17/10	JEGX01.0004W
IDR1100035184	**Westchester**	**9/16/10**	**JEGX01.0004AE**
IDR1000036319	Putnam	9/28/10	JEGX01.0004W
IDR1000036979	Putnam	10/8/10	JEGX01.0004W
IDR1000038792	Nassau	10/29/10	JEGX01.0004W
IDR1000034599	Orange	9/15/10	JEGX01.0004W
IDR1100006235	Westchester	2/21/11	JEGX01.0004W
IDR1100021079	Rockland	7/13/11	JEGX01.0004W
IDR1000030147	Out-Of-State	8/22/10	JEGX01.0004W
IDR1100003844	Onondaga	2/1/11	JEGX01.0004W
IDR1100022186	Yates	7/22/11	JEGX01.0004W
IDR1100027690	Erie	9/6/11	JEGX01.0004W
IDR1100030508	Madison	10/9/11	JEGX01.0004W
IDR1100031312	Suffolk	10/5/11	JEGX01.0004W
IDR1100032014	Onondaga	10/22/11	JEGX01.0004W
IDR1000028670	Nassau	8/8/10	JEGX01.0004B
IDR1000029949	Suffolk	8/16/10	JEGX01.0004B
IDR1000033603	Erie	9/14/10	JEGX01.0004B
IDR1000034213	Erie	9/13/10	JEGX01.0004B
IDR1000037723	Westchester	10/4/10	JEGX01.0004B
IDR1000039087	Westchester	10/27/10	JEGX01.0004B

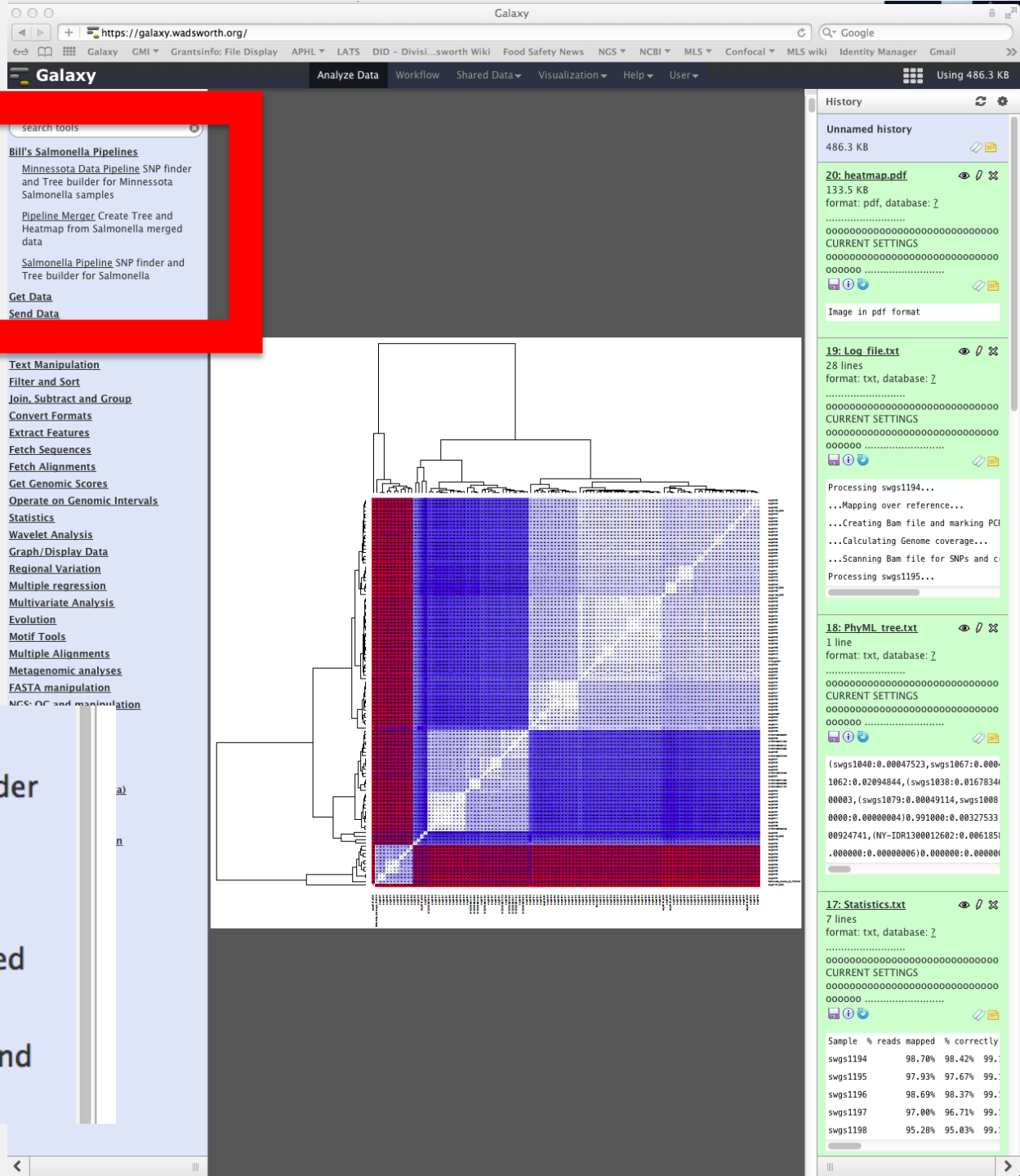
WGCA can identify an outbreak cluster not detected by PFGE



Implementing WGCA in real-time.

- Evaluate WGCA efficacy compared to PFGE.
  - Speed
  - Actionable Clusters
  - Cost
- Develop in house bioinformatic pipeline.
- Develop communication pipeline to epidemiologists.
- Determine cluster parameters that represent an outbreak from a single source.
- Acquire a real data set to evaluate evolving informatic methods.

Data is analyzed using a portal developed by Informatics core



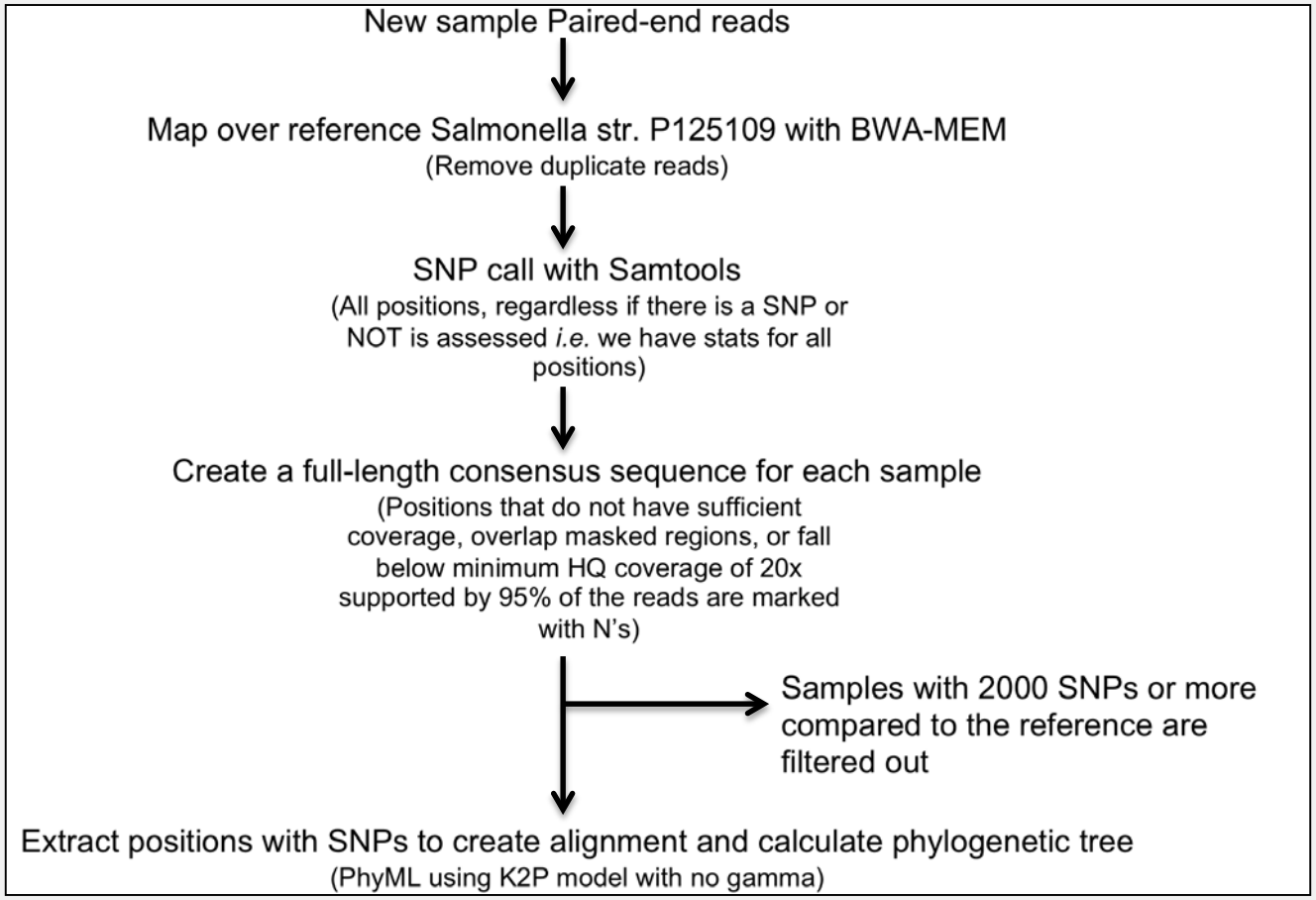
Bill's Salmonella Pipelines

Minnesota Data Pipeline SNP finder and Tree builder for Minnesota Salmonella samples

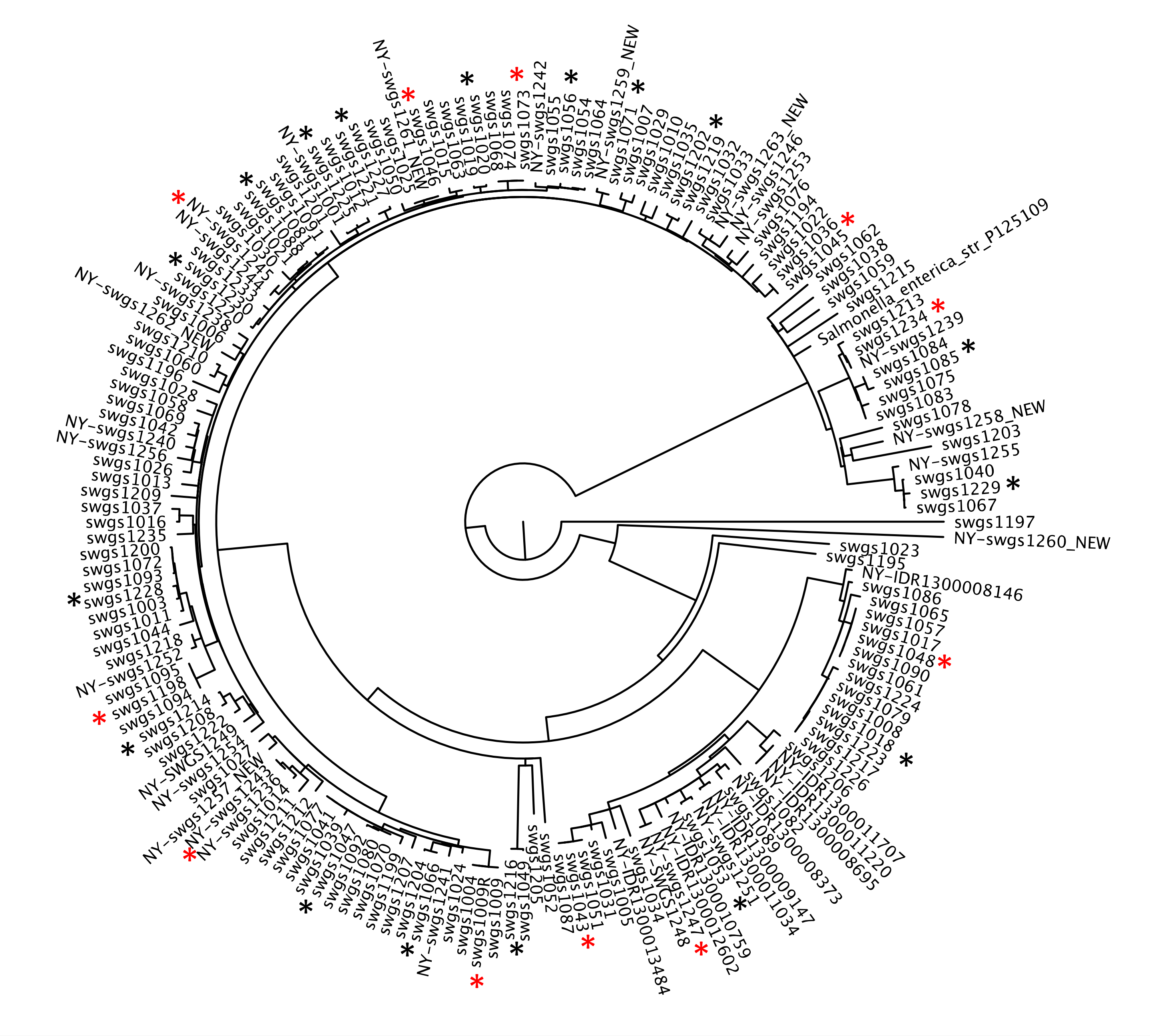
Pipeline Merger Create Tree and Heatmap from Salmonella merged data

Salmonella Pipeline SNP finder and Tree builder for Salmonella

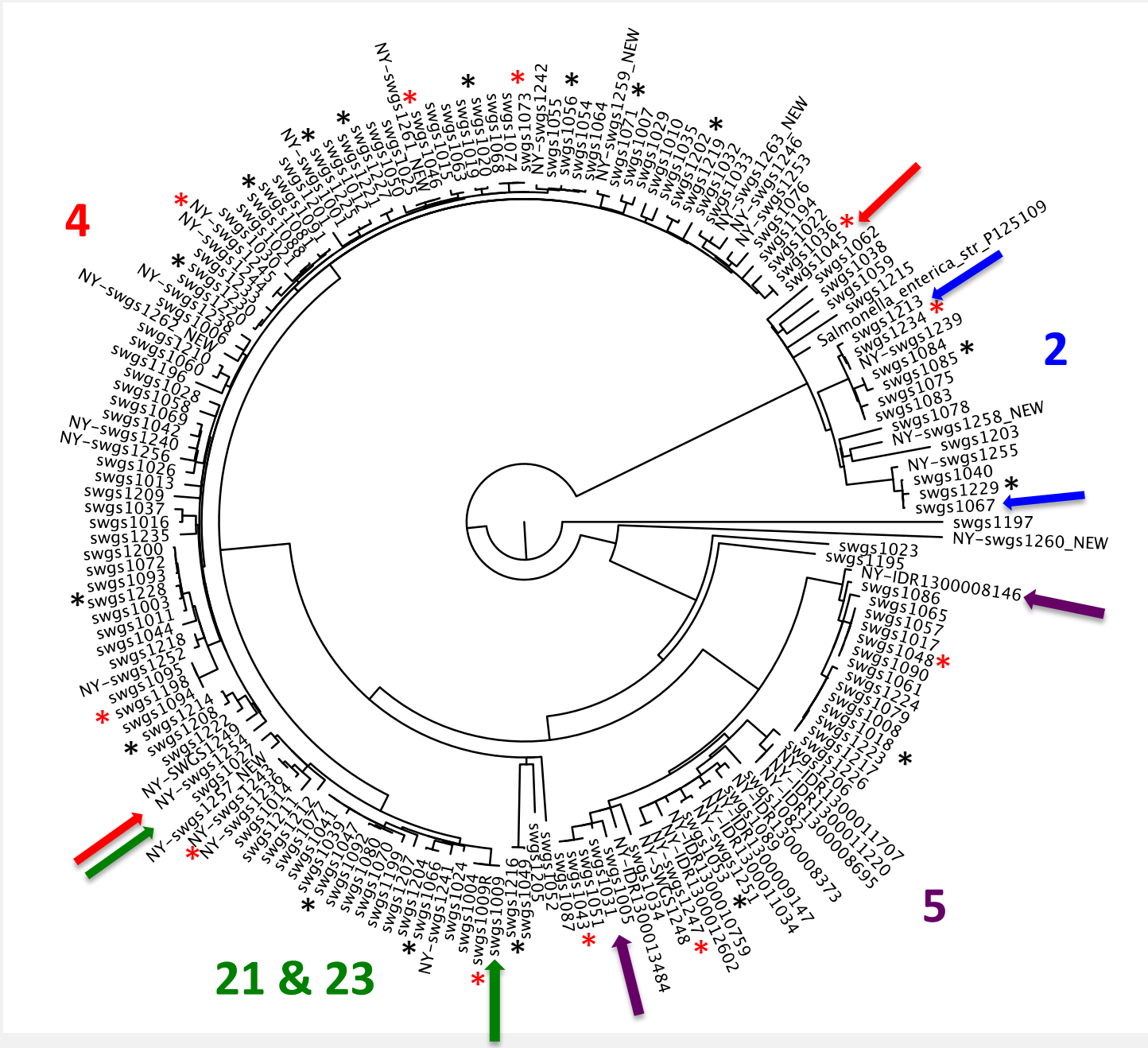
Pipeline


New sample Paired-end reads
Map over reference Salmonella str. P125109 with BWA-MEM
(Remove duplicate reads)
SNP call with Samtools
(All positions, regardless if there is a SNP or NOT is assessed *i.e.* we have stats for all positions)
Create a full-length consensus sequence for each sample
(Positions that do not have sufficient coverage, overlap masked regions, or fall below minimum HQ coverage of 20x supported by 95% of the reads are marked with N's)
Samples with 2000 SNPs or more compared to the reference are filtered out
Extract positions with SNPs to create alignment and calculate phylogenetic tree
(PhyML using K2P model with no gamma)

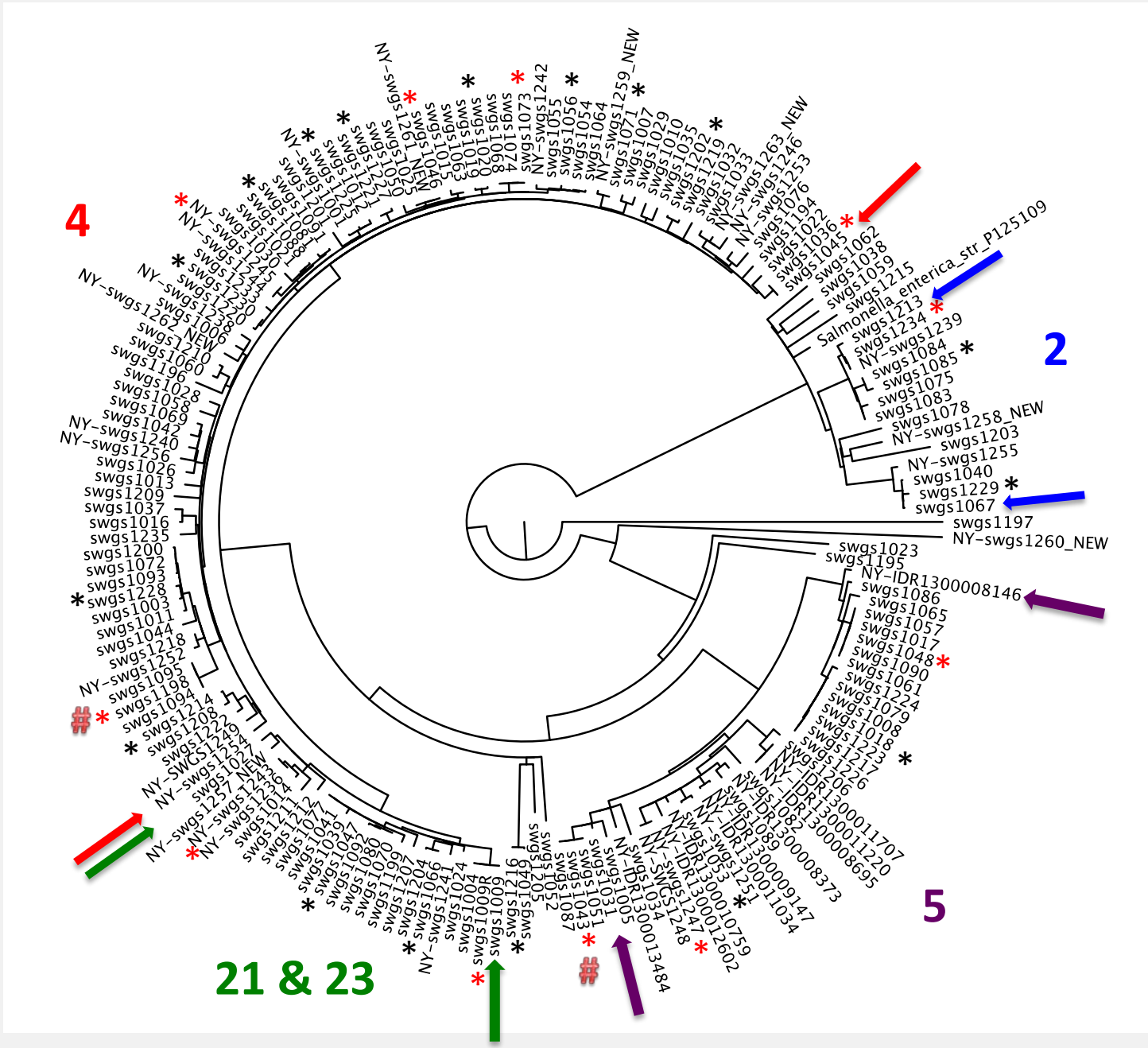
We see lots of clusters



Clusters vs. PFGE



Non-endemic clusters



In 6 months:

156 isolates have been sequenced and analyzed.

- .ca 6 isolates / week

28 clusters were reported to epidemiologists.

- 1 cluster every week

12 clusters have **zero** snp differences.

- Collected up to 6 months apart.

10 clusters acquired one or more new isolates.

- 7 isolates in the largest cluster

Clusters with identical SNPs

cluster	# of isolates	temporal distance	spatial distance	epidemiological report
GC-13	3	2d	metro	family thanksgiving meal
GC-02	2	4d	same county	shopped at same grocery store
GC-27	2	6d	same county	sibs
GC-09	4	42d	2 in one county; 2 distant	husband wife pair; other two not related
GC-07	3	6mo	metro	no common source identified

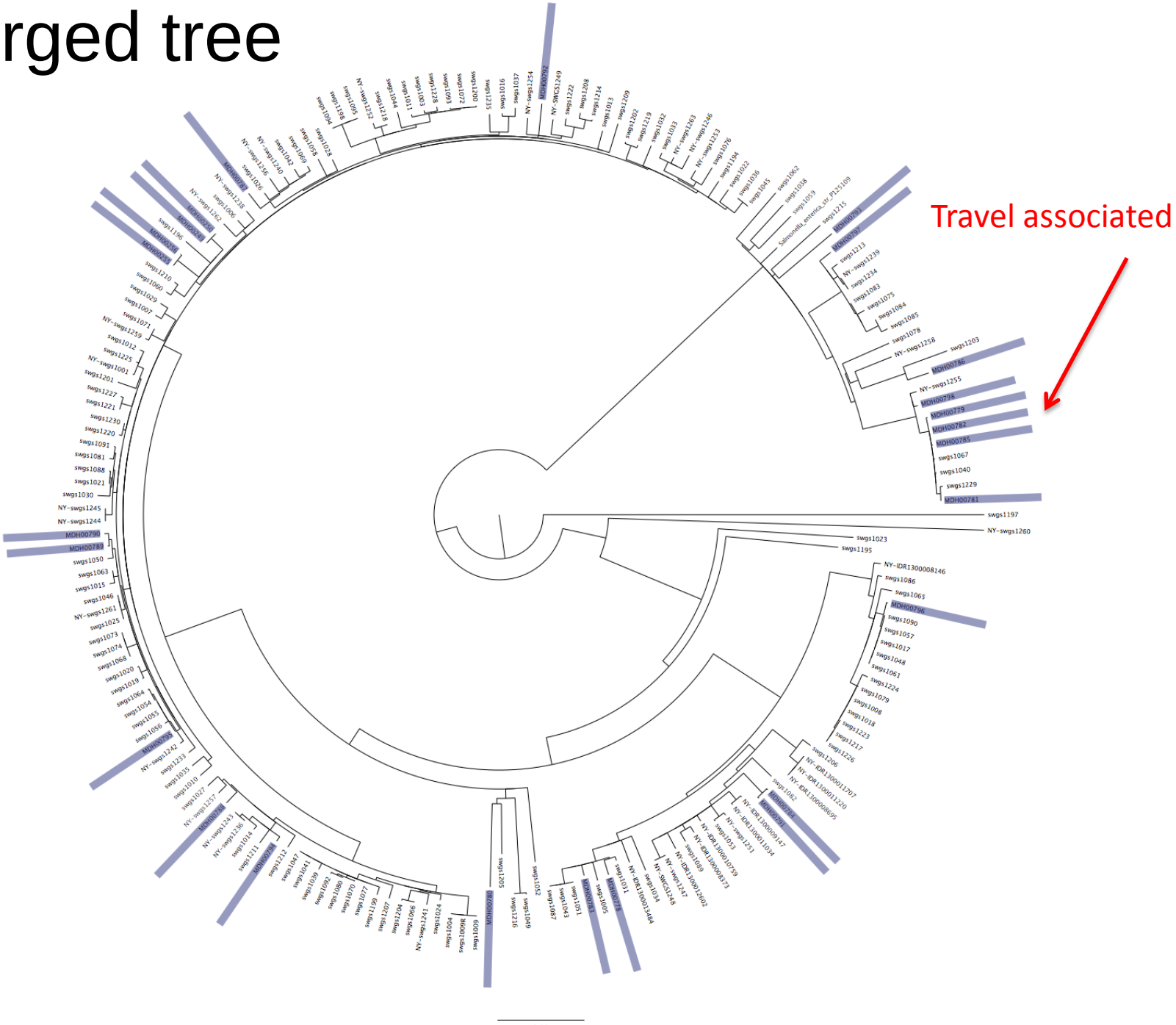
PFGE vs. WGCA for surveillance

metric	PFGE	WGCA
TAT from isolation	2 days	7 days
Cost	$69	$294
Technician time	8h	10h
Actionable clusters	2	28

Two State Network

- Collaborating with Minnesota.
  - Currently no informatics in house.
- We pull their sequences of Basespace.
- Run through our pipeline.

Merged tree



FDA Genome Trackr network

State Health labs

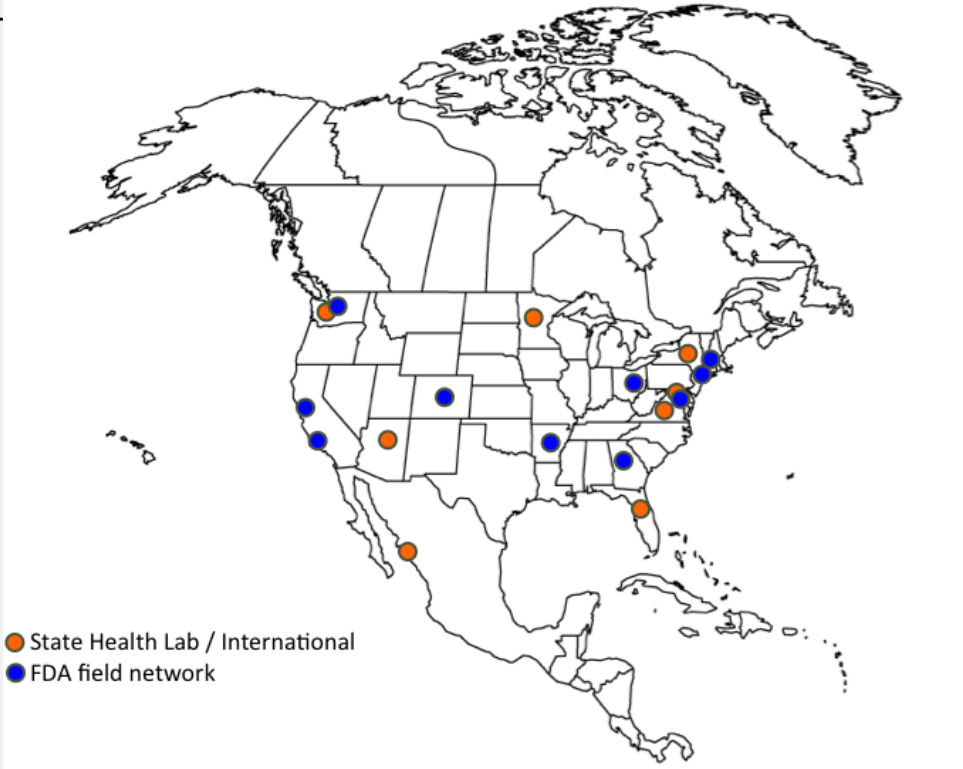
- New York
- Florida
- Arizona
- Washington
- Minnesota
- Virginia
- Maryland

FDA labs

- 9 FDA field labs
- CFSAN - MOD1
- CFSAN - Wiley
- IEH (contracting lab)

International labs

- Mexico
- Ireland
- UK (FERA)
- Columbia



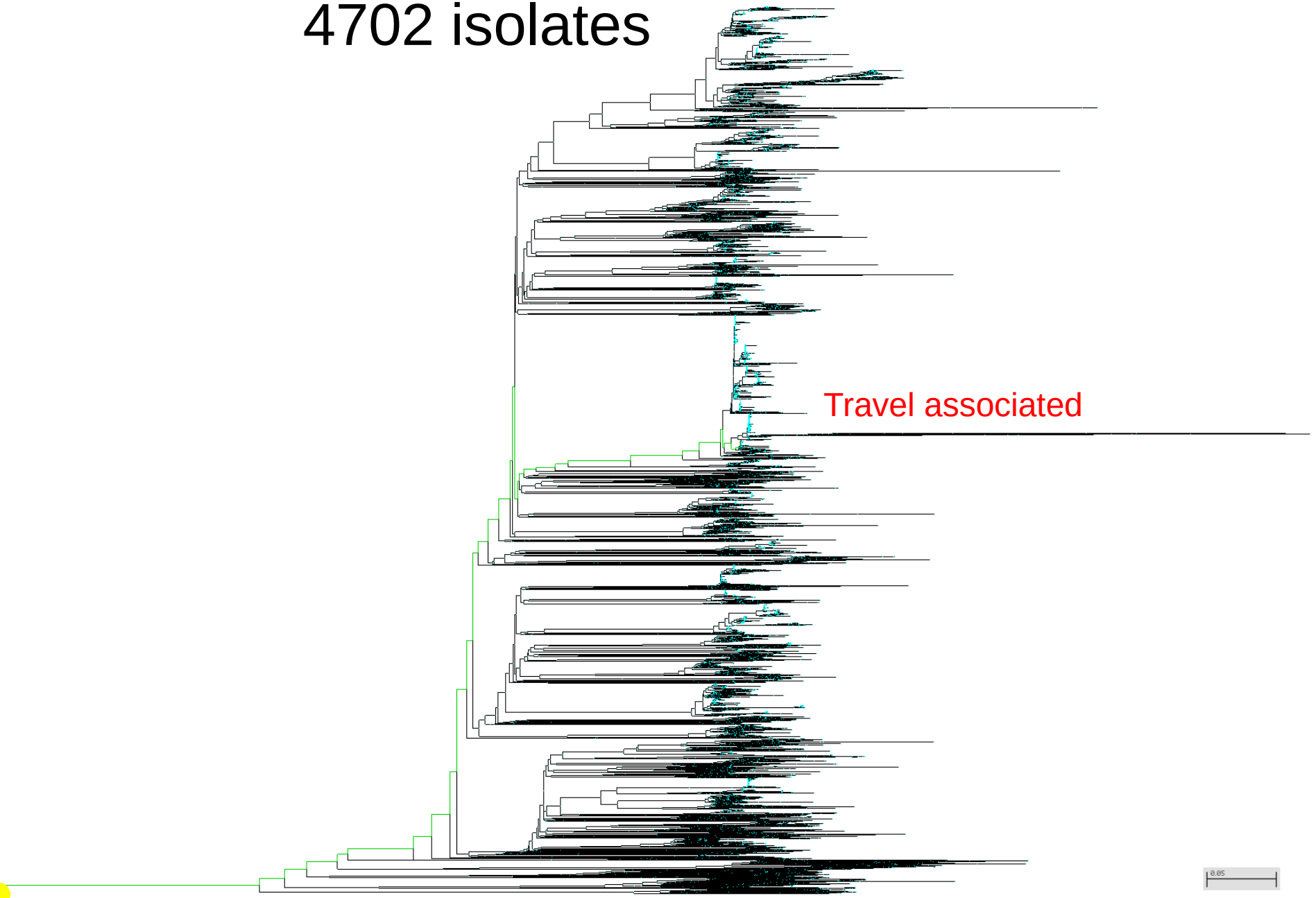
Contributors

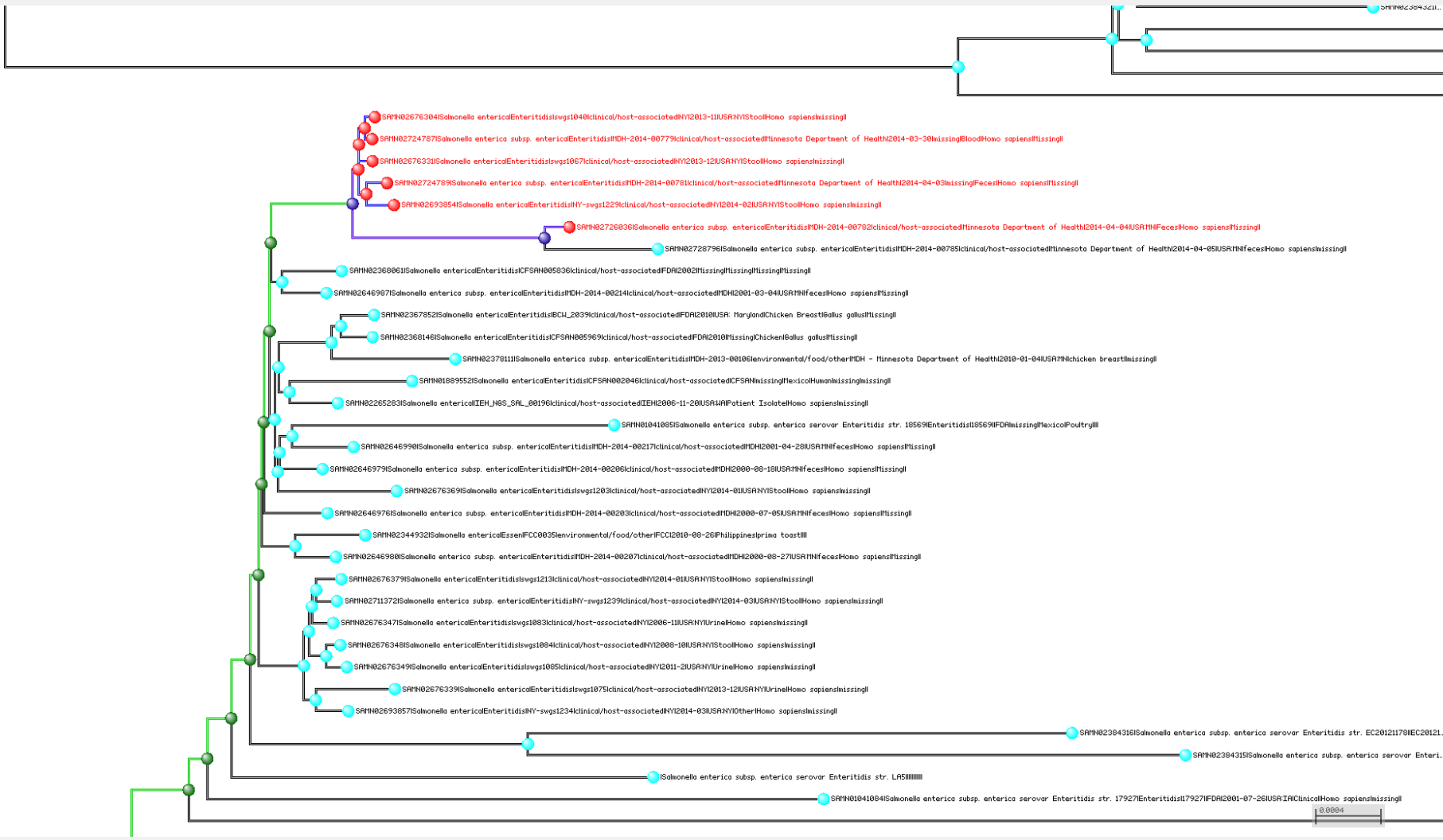
- Turkey
- Brazil
- Italy

Genomic Surveillance Machine

- State labs feed the machine by uploading sequences from isolates received through surveillance.
- Federal and other support for reagents and equipment.
- NCBI to analyze the products of this machine and reports results to state and federal agencies.

NCBI *Salmonella* tree
4702 isolates




SAMN02676304|Salmonella enterica|Enteritidis|swgs1040|clinical/host-associated|NY|2013-11|USA:NY|Stool|Homo sapiens|missing||
SAMN02724787|Salmonella enterica subsp. enterica|Enteritidis|MDH-2014-00779|clinical/host-associated|Minnesota Department of Health|2014-03-30|missing|Blood|Homo sapiens|Missing||
SAMN02676331|Salmonella enterica|Enteritidis|swgs1067|clinical/host-associated|NY|2013-12|USA:NY|Stool|Homo sapiens|missing||
SAMN02724789|Salmonella enterica subsp. enterica|Enteritidis|MDH-2014-00781|clinical/host-associated|Minnesota Department of Health|2014-04-03|missing|Feces|Homo sapiens|Missing||
SAMN02693854|Salmonella enterica|Enteritidis|NY-swgs1229|clinical/host-associated|NY|2014-02|USA:NY|Stool|Homo sapiens|missing||
SAMN02726036|Salmonella enterica subsp. enterica|Enteritidis|MDH-2014-00782|clinical/host-associated|Minnesota Department of Health|2014-04-04|USA:MN|Feces|Homo sapiens|Missing||
SAMN02728796|Salmonella enterica subsp. enterica|Enteritidis|MDH-2014-00785|clinical/host-associated|Minnesota Department of Health|2014-04-05|USA:MN|Feces|Homo sapiens|missing||
SAMN02368061|Salmonella enterica|Enteritidis|CFSAN005836|clinical/host-associated|FDA|2002|Missing|Missing|Missing|Missing||
SAMN02646987|Salmonella enterica subsp. enterica|Enteritidis|MDH-2014-00214|clinical/host-associated|MDH|2001-03-04|USA:MN|Feces|Homo sapiens|Missing||
SAMN02367852|Salmonella enterica|Enteritidis|BCW_2039|clinical/host-associated|FDA|2010|USA: Maryland|Chicken Breast|Gallus gallus|Missing||
SAMN02368146|Salmonella enterica|Enteritidis|CFSAN005969|clinical/host-associated|FDA|2010|Missing|Chicken|Gallus gallus|Missing||
SAMN02378111|Salmonella enterica subsp. enterica|Enteritidis|MDH-2013-00106|environmental/food/other|MDH - Minnesota Department of Health|2010-01-04|USA:MN|chicken breast|Missing||
SAMN01889552|Salmonella enterica|Enteritidis|CFSAN002046|clinical/host-associated|CFSAN|Missing|Mexico|Human|Missing|Missing||
SAMN02265283|Salmonella enterica|IIEH_NGS_SAL_00196|clinical/host-associated|IIEH|2006-11-20|USA:IA|Patient Isolate|Homo sapiens|Missing||
SAMN01041085|Salmonella enterica subsp. enterica serovar Enteritidis str. 18569|Enteritidis|18569|FDA|Missing|Mexico|Poultry|||
SAMN02646990|Salmonella enterica subsp. enterica|Enteritidis|MDH-2014-00217|clinical/host-associated|MDH|2001-04-28|USA:MN|Feces|Homo sapiens|Missing||
SAMN02646979|Salmonella enterica subsp. enterica|Enteritidis|MDH-2014-00206|clinical/host-associated|MDH|2000-08-18|USA:MN|Feces|Homo sapiens|Missing||
SAMN02676369|Salmonella enterica|Enteritidis|swgs1203|clinical/host-associated|NY|2014-01|USA:NY|Stool|Homo sapiens|missing||
SAMN02646976|Salmonella enterica subsp. enterica|Enteritidis|MDH-2014-00203|clinical/host-associated|MDH|2000-07-05|USA:MN|Feces|Homo sapiens|Missing||
SAMN02344932|Salmonella enterica|Essen|FCC003|environmental/food/other|FCC|2010-08-26|Philippines|prima toast|||
SAMN02646980|Salmonella enterica subsp. enterica|Enteritidis|MDH-2014-00207|clinical/host-associated|MDH|2000-08-27|USA:MN|Feces|Homo sapiens|Missing||
SAMN02676379|Salmonella enterica|Enteritidis|swgs1213|clinical/host-associated|NY|2014-01|USA:NY|Stool|Homo sapiens|missing||
SAMN02711372|Salmonella enterica subsp. enterica|Enteritidis|NY-swgs1239|clinical/host-associated|NY|2014-03|USA:NY|Stool|Homo sapiens|missing||
SAMN02676347|Salmonella enterica|Enteritidis|swgs1083|clinical/host-associated|NY|2006-11|USA:NY|Urine|Homo sapiens|missing||
SAMN02676348|Salmonella enterica|Enteritidis|swgs1084|clinical/host-associated|NY|2008-10|USA:NY|Stool|Homo sapiens|missing||
SAMN02676349|Salmonella enterica|Enteritidis|swgs1085|clinical/host-associated|NY|2011-2|USA:NY|Urine|Homo sapiens|missing||
SAMN02676339|Salmonella enterica|Enteritidis|swgs1075|clinical/host-associated|NY|2013-12|USA:NY|Urine|Homo sapiens|missing||
SAMN02693857|Salmonella enterica|Enteritidis|NY-swgs1234|clinical/host-associated|NY|2014-03|USA:NY|Other|Homo sapiens|missing||
SAMN02384316|Salmonella enterica subsp. enterica serovar Enteritidis str. EC20121178|EC20121...
SAMN02384315|Salmonella enterica subsp. enterica serovar Enteri...
|Salmonella enterica subsp. enterica serovar Enteritidis str. LA5|||||||||
SAMN01041084|Salmonella enterica subsp. enterica serovar Enteritidis str. 17927|Enteritidis|17927|FDA|2001-07-26|USA:IA|Clinical|Homo sapiens|missing||
0.0004

Expected Outcomes for WGS surveillance

Laboratory

- Improve outbreak cluster detection.
- Clusters will be detected more rapidly and from fewer isolates.

Epidemiologists

- Allow identification of clusters within **endemic** patterns.
- Allow more efficient use of resources by focusing on highly genetically related clusters.
- Solve more clusters.

Public Health

- More efficient identification and removal of pathogen sources.

Challenges exist in Creating a Network

- Increasing amounts of data.
- Metadata: how much should be public?
  - In real time?
  - What elements?

- Transitioning:
  - Integration with PFGE typing.
  - Integrating surveillance at a national level.
  - Paying.

- As sequencing technology and bioinformatics evolve:
  - Need to maintain backward compatibility

THE FUTURE

Near term

- Universal *Salmonella* tree
- Hands off data submission to NCBI
- Hands off data analysis

Further out

- Entire process automated
  - Sample preparation
  - Sequencing
  - Identification of clusters
  - Reporting

Summary

- WGS can improve surveillance activities and outbreak traceback.
- It is practical to develop network.
- It is likely the transition will be gradual – first to go live.
  - Pathogens with more stable genomes.
  - Pathogens with greatest Public Health Impact.