Identification and Characterization of Foodborne Pathogens by Whole Genome Sequencing: A Shift in Paradigm

Peter Gerner-Smidt, MD, ScD

Enteric Diseases Laboratory Branch

EFSA Scientific Colloquium Nº 20
Use of Whole Genome Sequencing (WGS) of foodborne pathogens for public health protection
Parma, Italy, 16-17 July 2014

National Center for Emerging and Zoonotic Infectious Diseases
Division of Foodborne, Waterborne, and Environmental Diseases

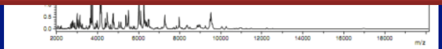
Current Methods of Characterizing Foodborne Pathogens in a Public Health Laboratory

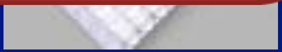
- Growth characteristics
- Phenotypic panels

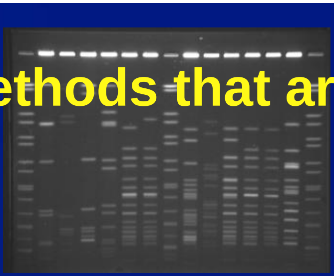
WGS could be used to characterize foodborne pathogens in a public health laboratory in one single efficient workflow!

TAT: (2-) 3- 4 days

- Electrophoresis (PFGE, capillary)







- **Each pathogen is characterized by methods that are specific to that pathogen in multiple workflows**
  - Separate workflows for each pathogen
  - TAT: 5 min – weeks (months)

Why Move Public Health Microbiology to WGS?

- Consolidation of workflows in the lab
- More efficient outbreak detection, investigation & control
    - Precise and flexible case definition
        - More outbreaks will be detected and solved when they are small
        - Scarce epi-resources may be focused
- More efficient surveillance of sporadic infections
    - Source attribution analysis of sporadic disease
    - Focus on pathogens of particular public health importance:
        - Virulence – Resistance - Emerging pathogens - Rapidly spreading clones/ traits- Vaccine preventable diseases

WGS in Public Health:

The tools must be

- Simple
  - Public health microbiologists are NOT bioinformaticians
  - Standard desktop software
- Comprehensive
  - All characterization in one workflow
- Work in a network of laboratories
  - Free sharing and comparison of data between labs
  - Central and local databases

Genomics for Diagnostics and Surveillance is a Global Issue

- **How can we implement genomics fast and efficiently at the global level?**
- **When to share data?**
    - Surveillance WGS should be shared in real-time
- **What is the minimal IT-infrastructure?**
- **What technical gaps need to be filled?**
- **QA/QC**
- **Political and ethical barriers**

GMI

Global Microbial Identifier

Generating succes stories:

Proof-of-Concept Study on the Use of Real-Time Whole Genome Sequencing in Conjunction with Enhanced Surveillance for Listeriosis

- Collaboration among the public health departments in the states, CDC, FDA, USDA, and NCBI
- International component: Developing and refining bioinformatics ‘pipelines’ with partners
  in Belgium. Canada, Denmark, England, and France

Why *Listeria monocytogenes?*

- **Illness is rare but serious, costly, and commonly outbreak associated**
- **Controllable**
- **Current subtyping methods are not ideal**
  - Not highly discriminatory
  - No evolutionary relationships
- ***Listeria* genome is fairly small and relatively easy to sequence and analyze**
- **Strong epidemiology surveillance (*Listeria* Initiative)**
- **Strong regulatory component**

Approach:

- **Sequence all clinical isolates in the U.S. during one year as close to real-time as possible <u>in parallel</u> with current surveillance**
  - PulseNet PFGE, strain characterization at CDC, interview of case-patients
- **Evaluate data on a weekly basis**
- **Follow up on clusters detected**
  - Both PFGE and WGS defined clusters
- **Upload sequences to NCBI (Genbank), a public database, as the sequences are generated**
  - With metadata that do NOT identify state or isolation date but with link to the PulseNet database

Listeria Whole Genome Sequencing Surveillance in Numbers

(8 months into the project, June 2014)

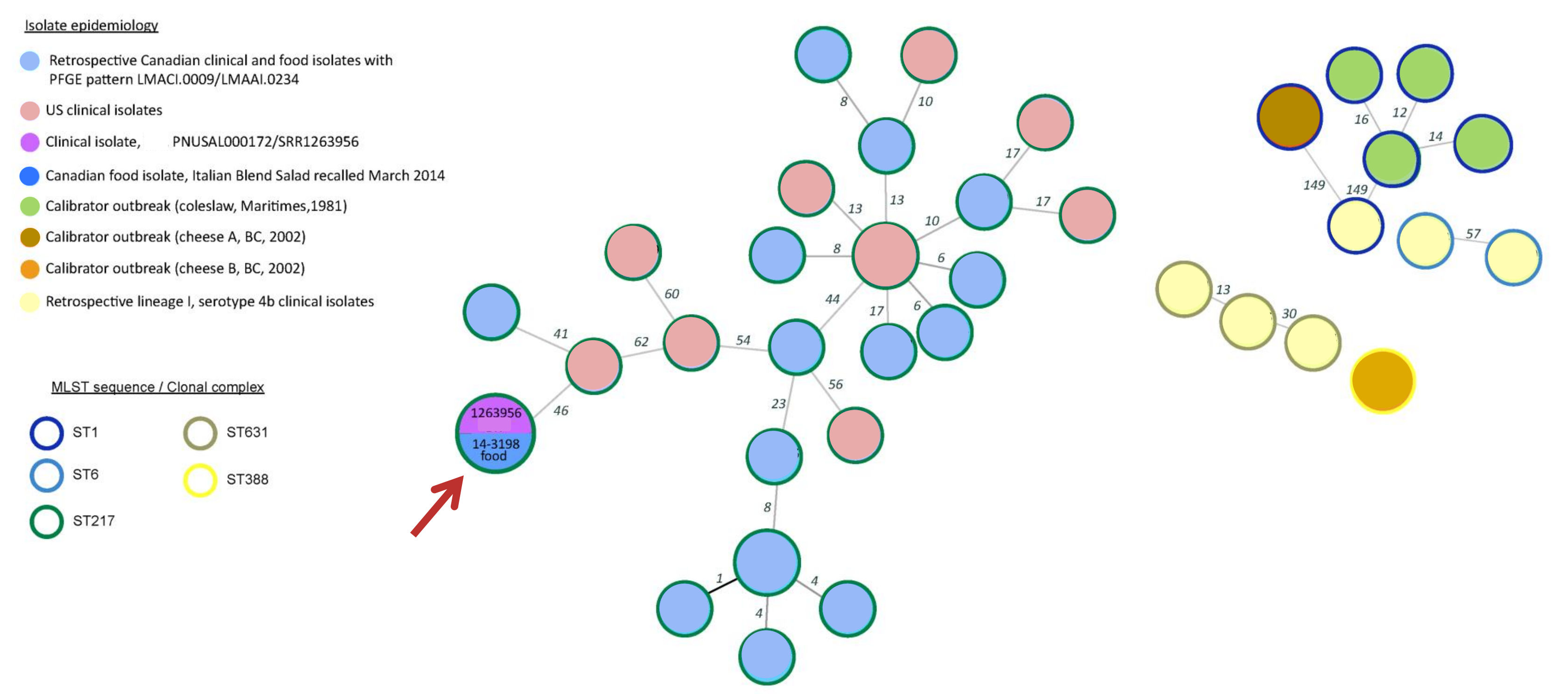
- ~ 690 clinical isolates sequenced in real-time
- ~ 75 environmental and historical clinical isolates sequenced
- 470 food isolates sequenced by FDA/GenomeTrakr
- 13 clusters identified by PFGE & WGS
- 3 clusters identified only by WGS

Listeria Whole Genome Sequencing Surveillance Lessons Learned:

- **Possible to perform WGS in real-time**
- **Historical data are critical for cluster determination**
- **WGS provides more resolution than PFGE**
  - One PFGE cluster could not be confirmed by WGS
    - Saving resources by interviewing only patients relevant to the cluster
  - Suspected source disproved by WGS in one cluster
- **A common source identified in one cluster until now**
  - Difficult to obtain relevant exposure information from patients
- **A sporadic case was linked to a lettuce recall by WGS**
- **Many good analytical approaches yielding (almost) the same information**

WGS Minimum Spanning Tree of *Listeria monocytogenes*

Connection between lettuce recall and a sporadic case?

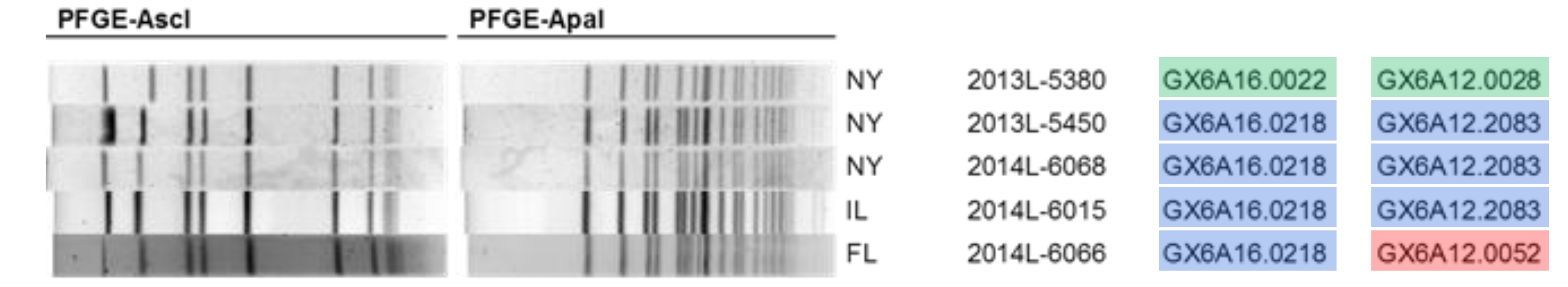


Courtesy Public Health Agency of Canada & the Canadian Food Inspection Agency

Lessons learned:
WGS may identify clusters <u>not</u> evident by PFGE

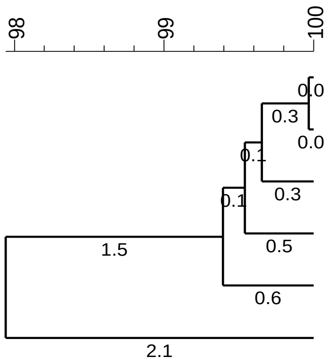
Cluster 1403MLGX6-1

- 5 cases, 3 PFGE patterns, 3 states
- All patients originally from former USSR or Poland



NY	2013L-5380	GX6A16.0022	GX6A12.0028
NY	2013L-5450	GX6A16.0218	GX6A12.2083
NY	2014L-6068	GX6A16.0218	GX6A12.2083
IL	2014L-6015	GX6A16.0218	GX6A12.2083
FL	2014L-6066	GX6A16.0218	GX6A12.0052

*wg*MLST



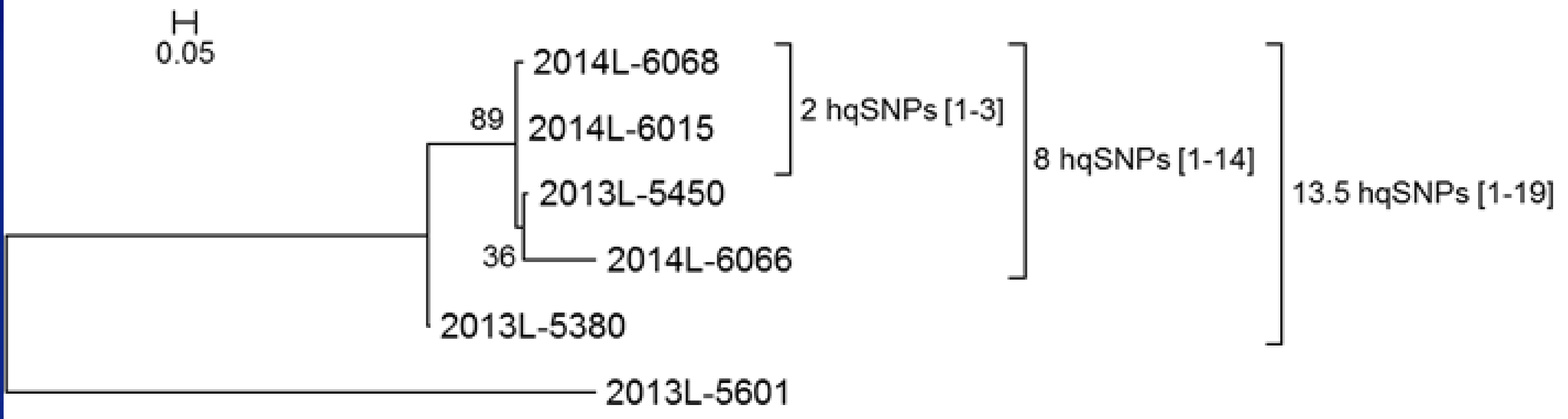
IL isolate
NYC isolate 1
FL isolate
NYC isolate 2
NYC isolate 3
2013 isolate – Nearest Neighbor

Lessons Learned:

WGS May Identify Clusters Not Evident by PFGE

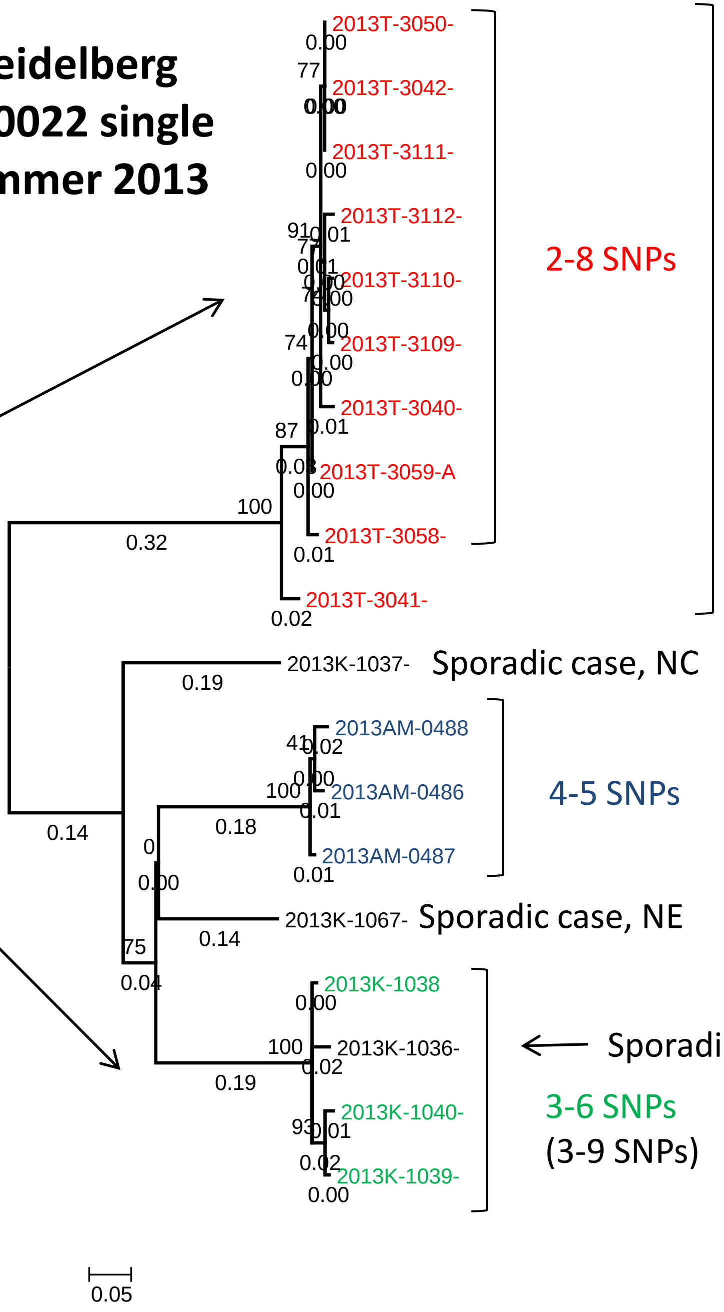
1403MLGX6-1WGS

hqSNP tree



How different are WGS profiles of isolates belonging to the same outbreak?

***Salmonella* serovar Heidelberg PFGE pattern JF6X01.0022 single state outbreaks in summer 2013**



105-113 SNPs

2-8 SNPs

2-16 SNPs

AL funeral

Sporadic case, NC

4-5 SNPs

NY daycare

Sporadic case, NE

Sporadic case, OH

3-6 SNPs

(3-9 SNPs)

CO family gathering

2012-2013 *Salmonella* serovar Heidelberg multistate outbreak associated with chicken from manufacturer X

PFGE patterns

🟦 Pattern 122

🟫 Pattern 672

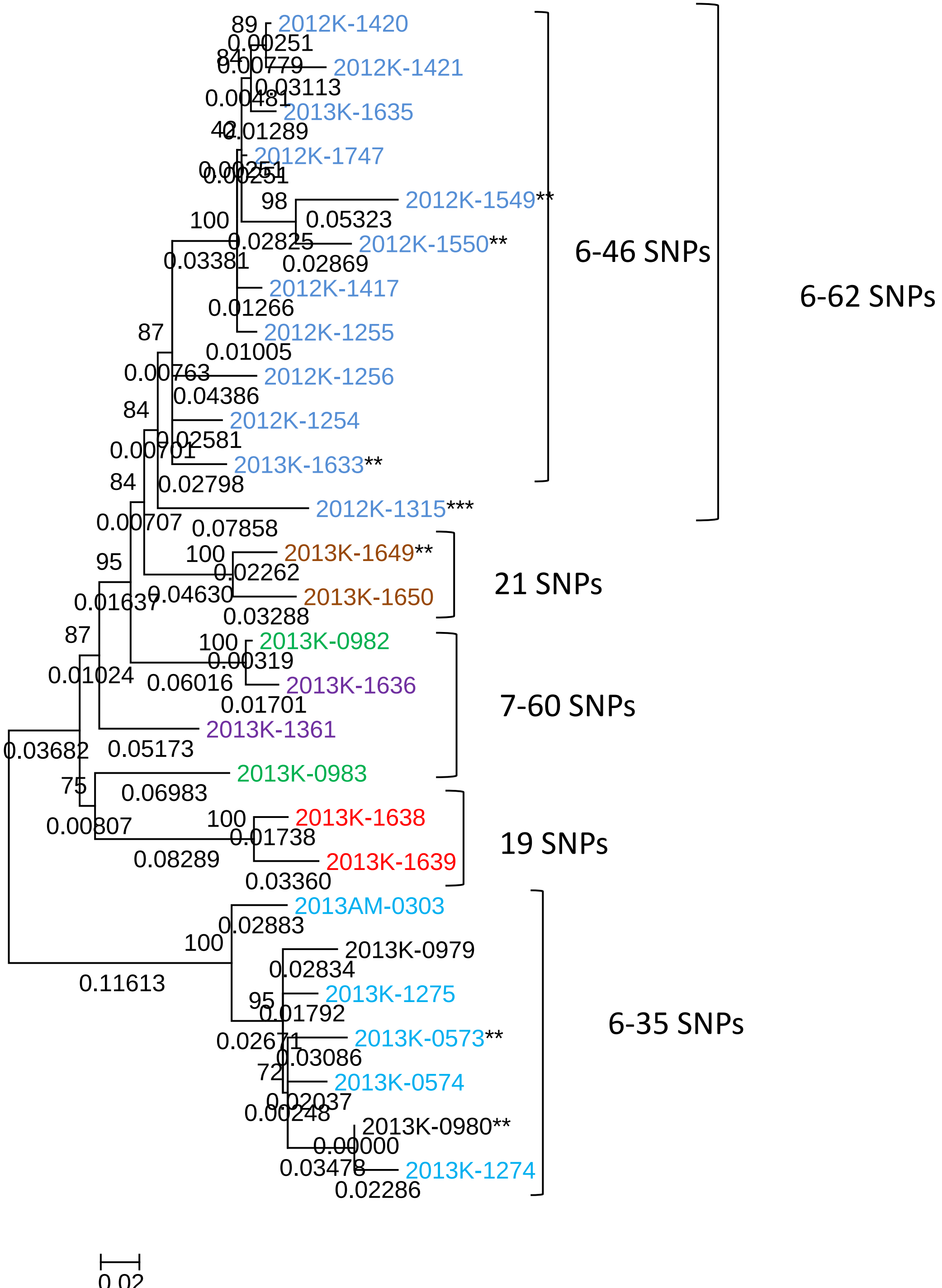
🟩 Pattern 45

🟪 Pattern 22

🟥 Pattern 326

🟦 Pattern 41

⬛ Pattern 258



** Chicken from manufacturer X

*** Clinical isolate from 2011

To SNP or Not to SNP?

in public health

- **Single Nucleotide Polymorphism (SNP) approaches**
    - Default for phylogenetic analyses of sequence data
    - Comparative subtyping by nature
    - Results difficult to communicate
    - Computationally intensive = SLOW
- **Gene- gene approach (wgMLST)**
    - Definitive subtyping
        - Leads to naming, tracking over time, easy communication
    - Computationally more simple = FAST but…
    - Sufficiently discriminatory?

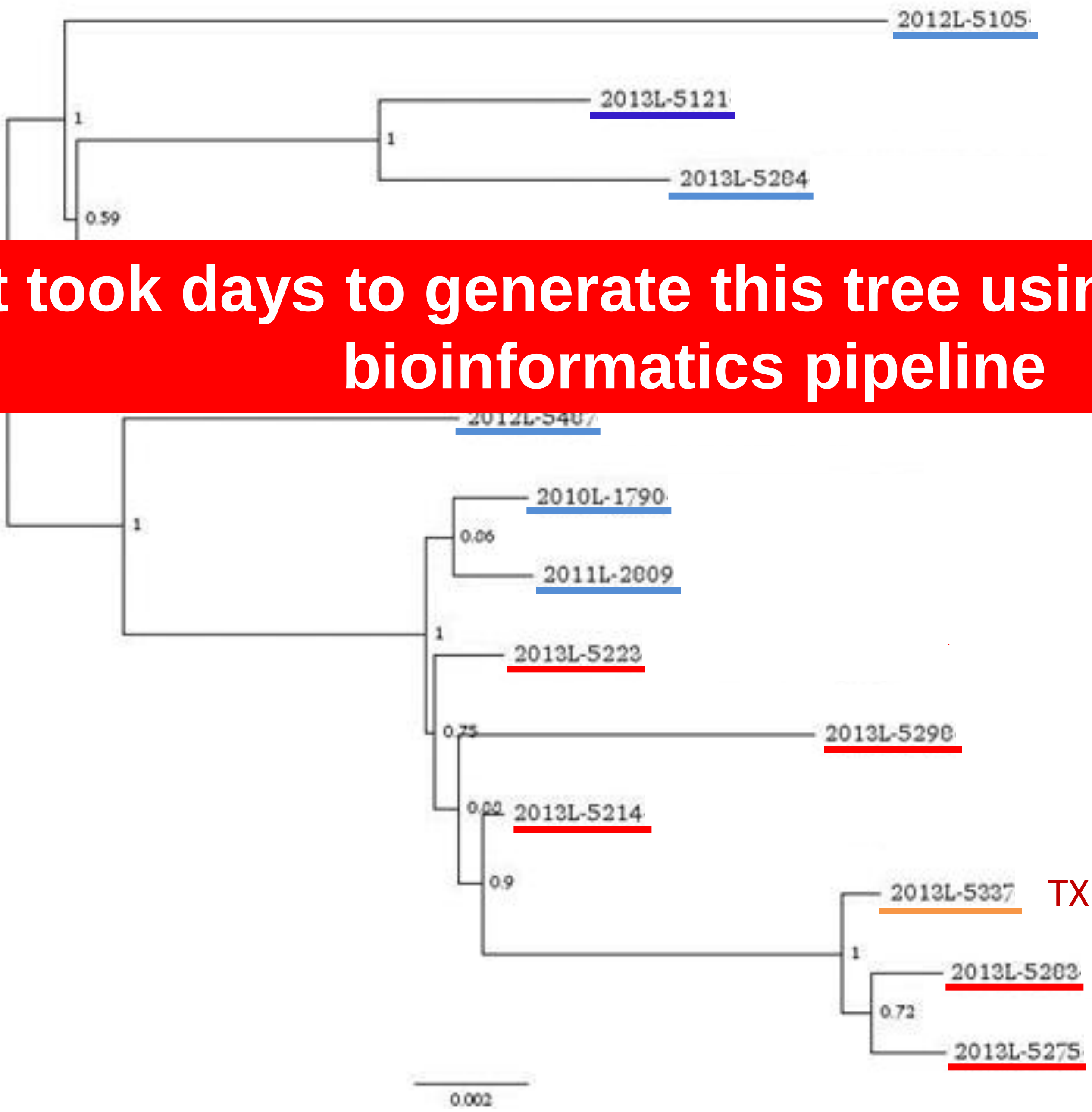
PulseNet™ International

The International Molecular Subtyping Network for Foodborne Disease Surveillance

USA, Canada, Europe, Asia Pacific, Latin America, Middle East Africa

Whole Genome Sequencing of *Listeria monocytogenes* during the Crave Brothers Cheese outbreak

High confidence core SNP



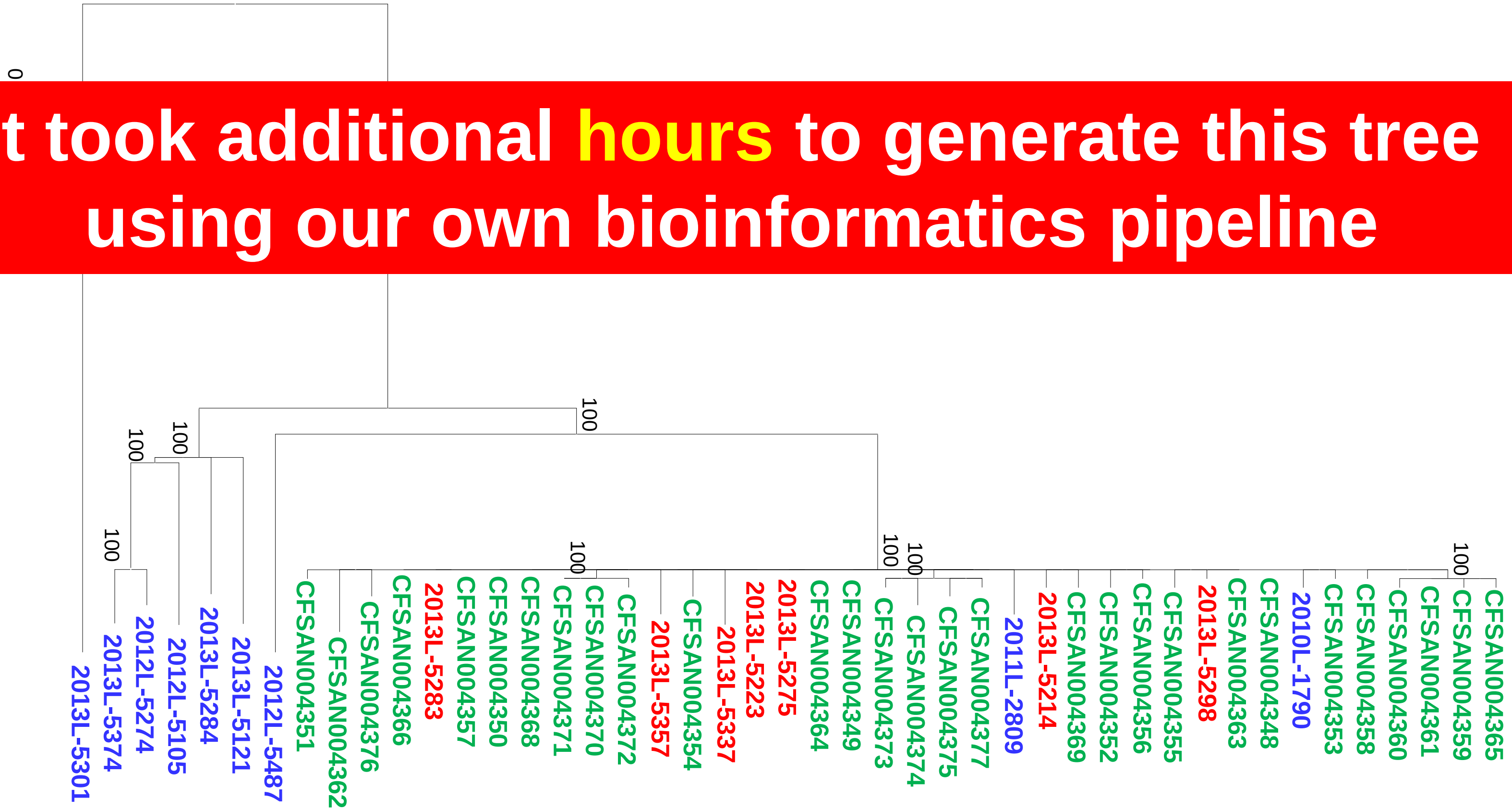
It took days to generate this tree using our own bioinformatics pipeline

TX - no exposure info

Whole Genome Sequencing of *Listeria monocytogenes* during the Crave Brothers Cheese outbreak

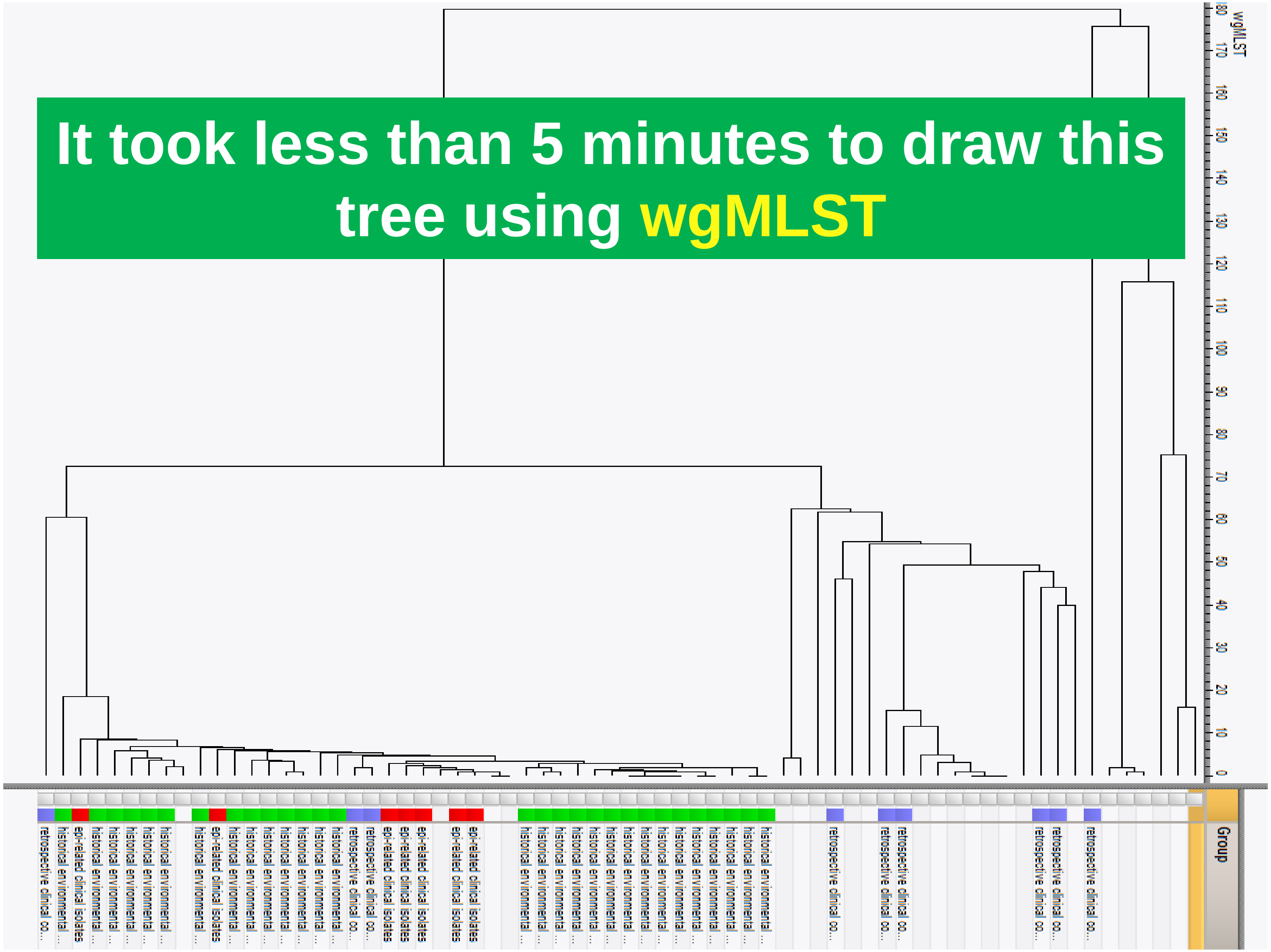
High confidence core SNP

Historical isolates from the plant environment added to the comparison (courtesy FDA/CFSAN)



It took additional hours to generate this tree using our own bioinformatics pipeline

It took less than 5 minutes to draw this tree using **wgMLST**



Gene – Gene Approach:

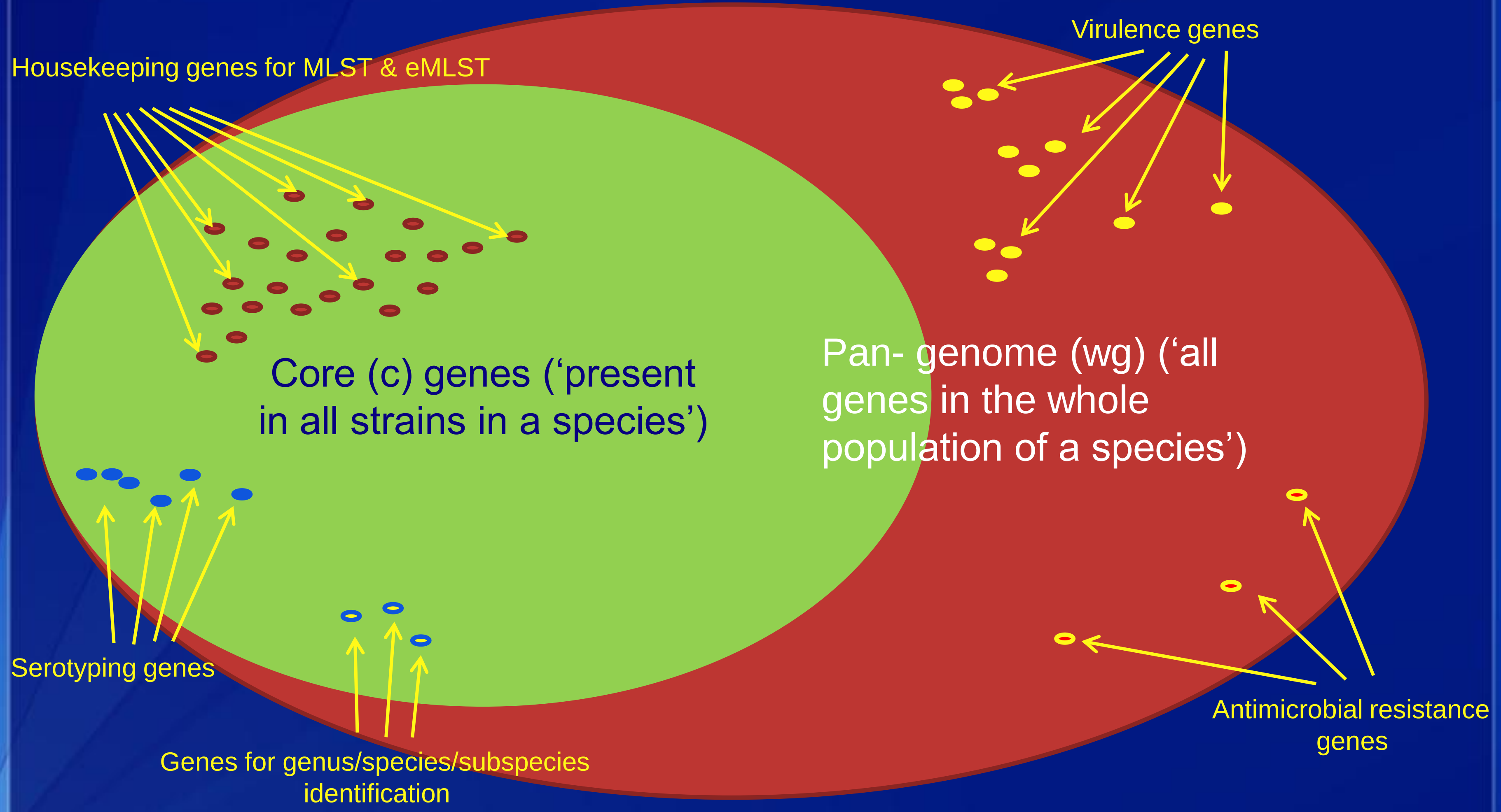
- Fixed set of genes ('loci') leading to typing schemes on different levels

MLST
Genus/Species
Serotype
AR

eMLST
cMLST
wgMLST

- Concept of allelic variation, not only point mutations
  - Evolutionary distance for events such as recombination and simultaneous close-range mutations are counted as one event
- Definitive subtyping
  - Leads to nomenclature
  - *Requires curation*

Genes That May Be Targeted In a Gene-Gene Analytical Approach



Gene – Gene Approach for Naming Subtyping in Keep with Phylogeny

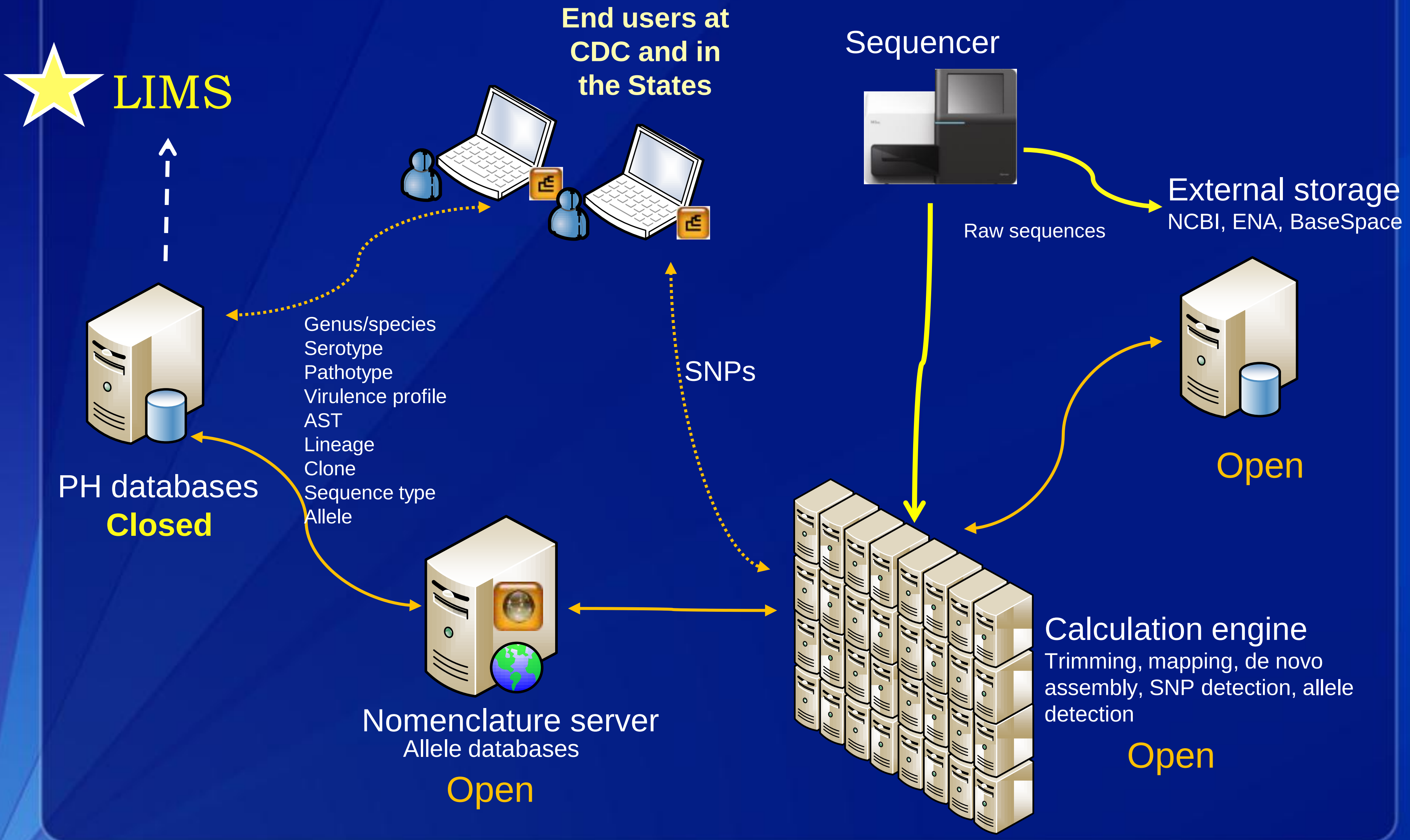
(concept to be developed)

Isolate A	**ST24 -**	**e12 -**	**c48** -	w214
Isolate B	**ST24 -**	**e12 -**	**c48** -	w352
Isolate C	**ST24 -**	**e12** -	c45 -	w132
Isolate D	ST31 -	e15 -	c60 -	w582

Isolate A and B **closely related**
Isolate C related to A and B but not as closely as A is to B
Isolate D unrelated to all the other isolates

Providing phylogenetic information in the name is important because isolates from the same source are more likely to be related than isolates from different sources

Public Health WGS Workflow



LIMS

End users at
CDC and in
the States

Sequencer

Raw sequences

External storage
NCBI, ENA, BaseSpace

Open

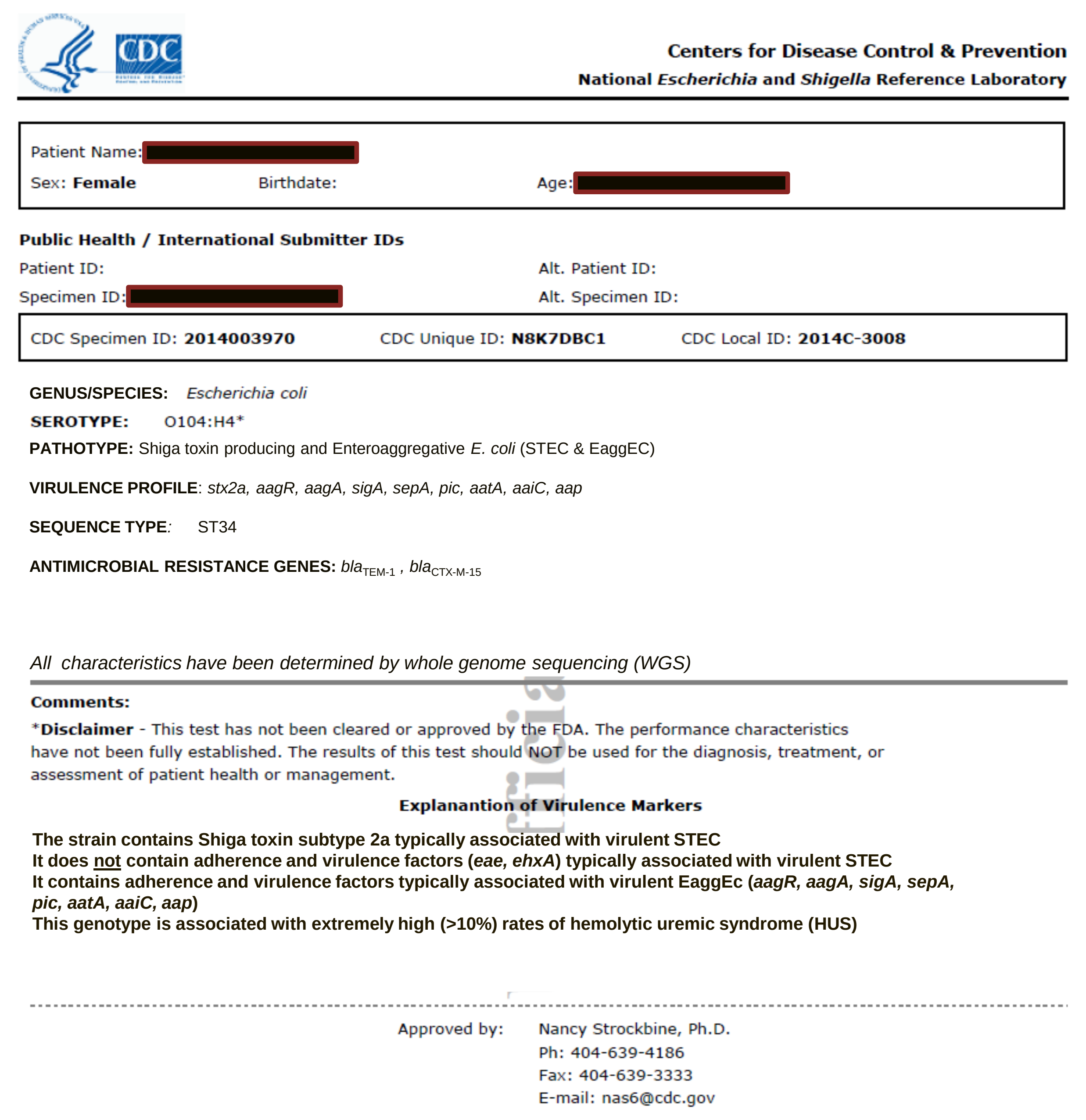
PH databases
Closed

Genus/species
Serotype
Pathotype
Virulence profile
AST
Lineage
Clone
Sequence type
Allele

SNPs

Nomenclature server
Allele databases
Open

Calculation engine
Trimming, mapping, de novo assembly, SNP detection, allele detection
Open



Centers for Disease Control & Prevention
National *Escherichia* and *Shigella* Reference Laboratory

Patient Name: [redacted]
Sex: **Female** Birthdate: Age: [redacted]

Public Health / International Submitter IDs

Patient ID: Alt. Patient ID:
Specimen ID: [redacted] Alt. Specimen ID:

CDC Specimen ID: **2014003970** CDC Unique ID: **N8K7DBC1** CDC Local ID: **2014C-3008**

GENUS/SPECIES: *Escherichia coli*

SEROTYPE: O104:H4*

PATHOTYPE: Shiga toxin producing and Enteroaggregative *E. coli* (STEC & EaggEC)

VIRULENCE PROFILE: *stx2a, aagR, aagA, sigA, sepA, pic, aatA, aaiC, aap*

SEQUENCE TYPE: ST34

ANTIMICROBIAL RESISTANCE GENES: $\mathit{bla}_{\text{TEM-1}}$, $\mathit{bla}_{\text{CTX-M-15}}$

All characteristics have been determined by whole genome sequencing (WGS)

Comments:

***Disclaimer** - This test has not been cleared or approved by the FDA. The performance characteristics have not been fully established. The results of this test should NOT be used for the diagnosis, treatment, or assessment of patient health or management.

Explanantion of Virulence Markers

The strain contains Shiga toxin subtype 2a typically associated with virulent STEC
It does <u>not</u> contain adherence and virulence factors (*eae, ehxA*) typically associated with virulent STEC
It contains adherence and virulence factors typically associated with virulent EaggEc (*aagR, aagA, sigA, sepA, pic, aatA, aaiC, aap*)
This genotype is associated with extremely high (>10%) rates of hemolytic uremic syndrome (HUS)

Approved by: Nancy Strockbine, Ph.D.
Ph: 404-639-4186
Fax: 404-639-3333
E-mail: nas6@cdc.gov

Real-time Sharing Of Sequence Data Is The Key To Successful Surveillance !

Acknowledgements

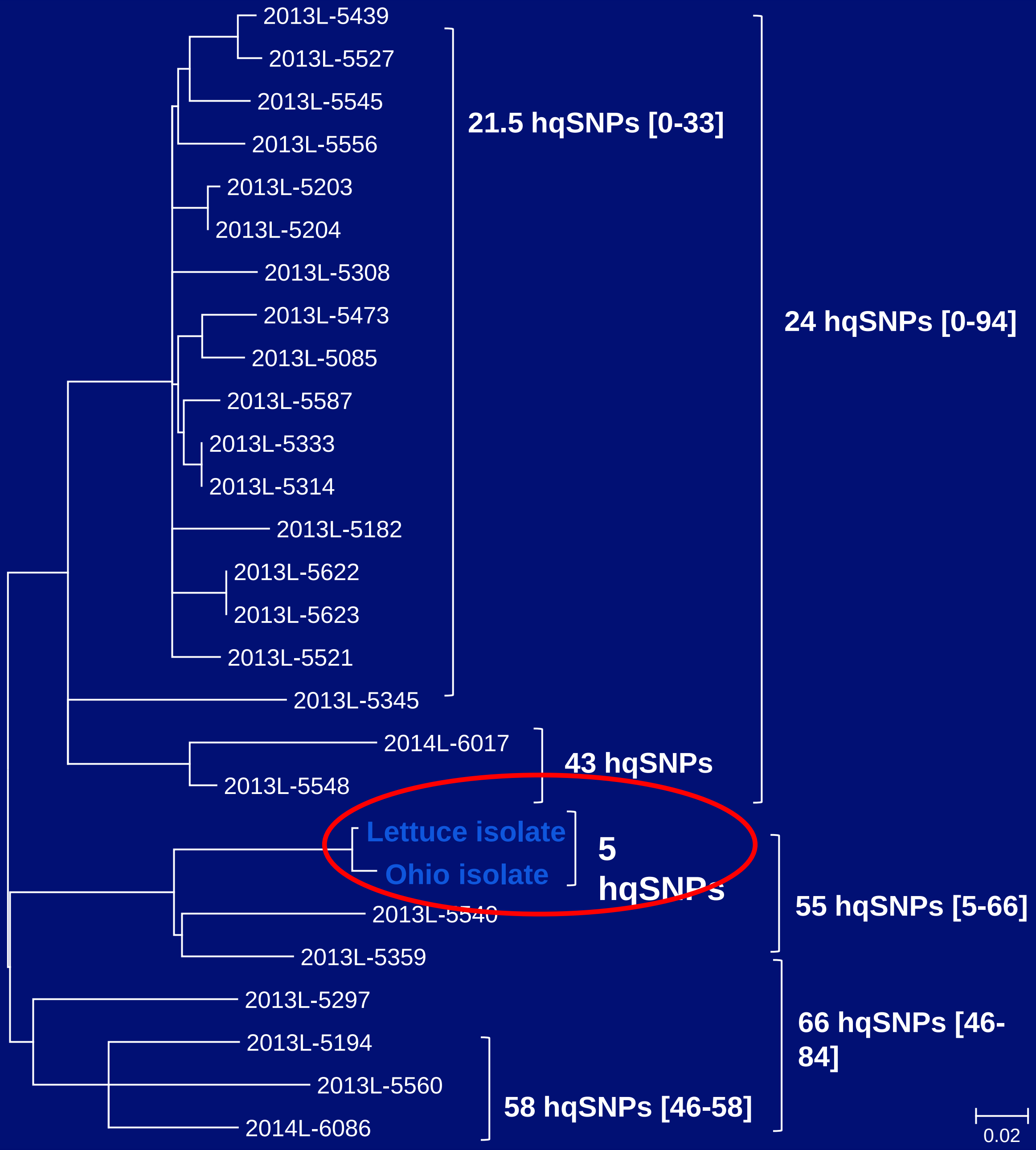
CDC: John Besser, Kelly Jackson, Amanda Conrad, Lee Katz, Steven Stroika, Heather Carleton, Zuzana Kucerova, Eija Trees, Katie Roache, Ashley Sabol, Andrew Huang, Lori Gladney, Maryann Turnsek, Ben Silk, Katie Fullerton, Matthew Wise, Ian Williams, Duncan MacCannell, Efrain Ribot, Patricia Griffin , Rob Tauxe, Chris Braden, Dana Pitts, Collette Leaumont, Patti Fields

Public Health Agency of Canada

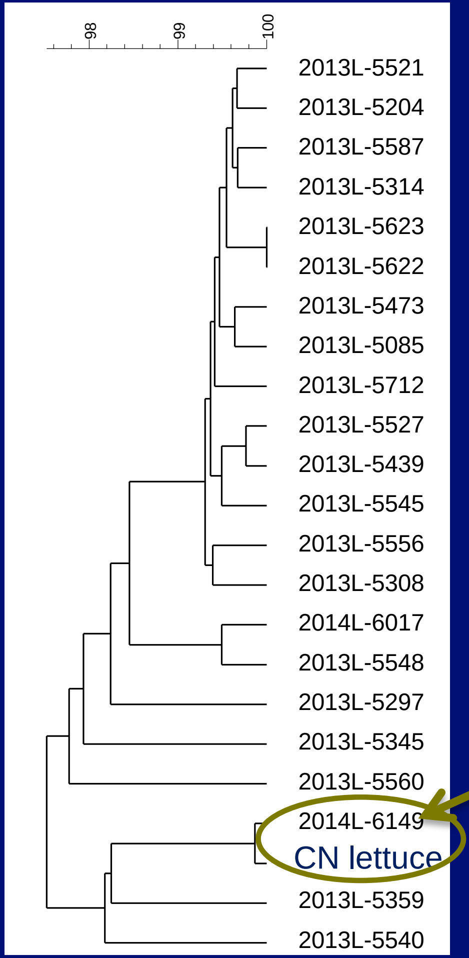
Disclaimers:

WGS Analysis at CDC

- ❑ Similar results
- ❑ Circumstantial evidence suggesting pre-packaged lettuce is the likely food vehicle in a sporadic case of listeriosis



Canadian Lettuce/OH case patient wgMLST



OH case

*UPGMA tree containing PFGE matches to Ohio and Canadian lettuce isolates

*wgMLST groups these isolates of interest as well as supports the general tree layout seen by hqSNP analysis