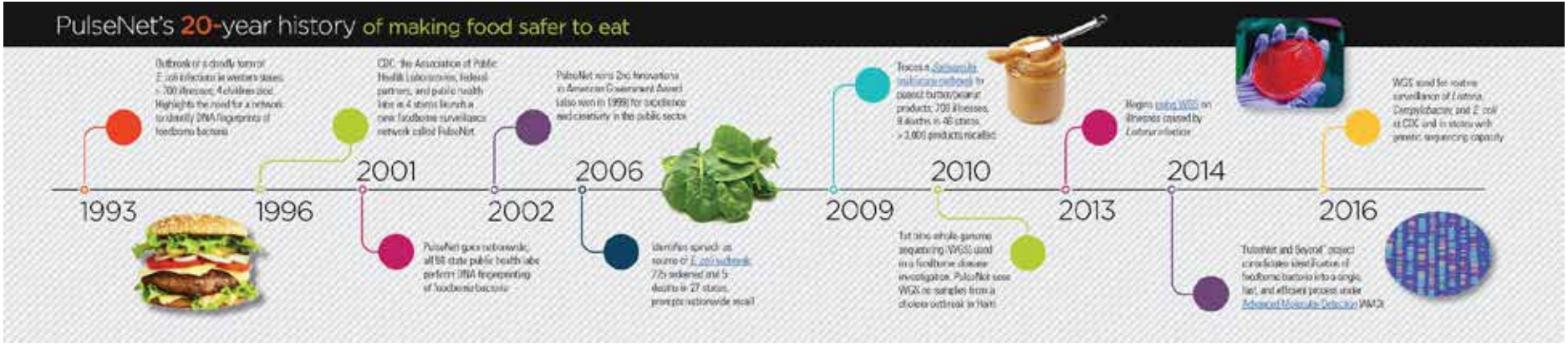
Whole Genome Sequencing (WGS) at CDC (PulseNet)

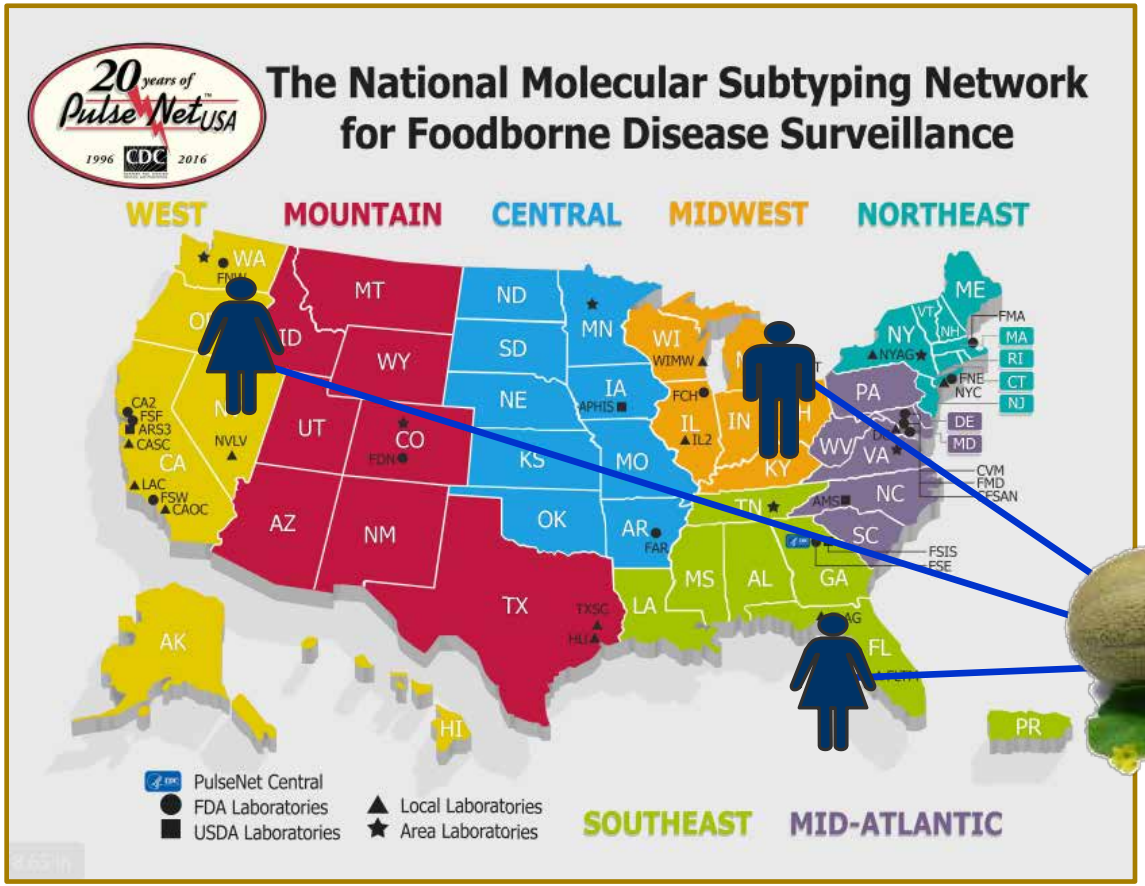
John Besser, PhD
Enteric Diseases Laboratory Branch
CDC/NCEZID/DFWED

Whole Genome Sequencing (WGS) at CDC (PulseNet)

- Background
- What has *not* changed from the PFGE era?
- The PulseNet WGS infrastructure
- How is WGS different?
- The road ahead

PulseNet's 20th Anniversary





Human disease surveillance

Food monitoring

Food animal surveillance

Cost-Benefit Study (Pre-WGS)

ARTICLE IN PRESS

An Economic Evaluation of PulseNet

A Network for Foodborne Disease Surveillance

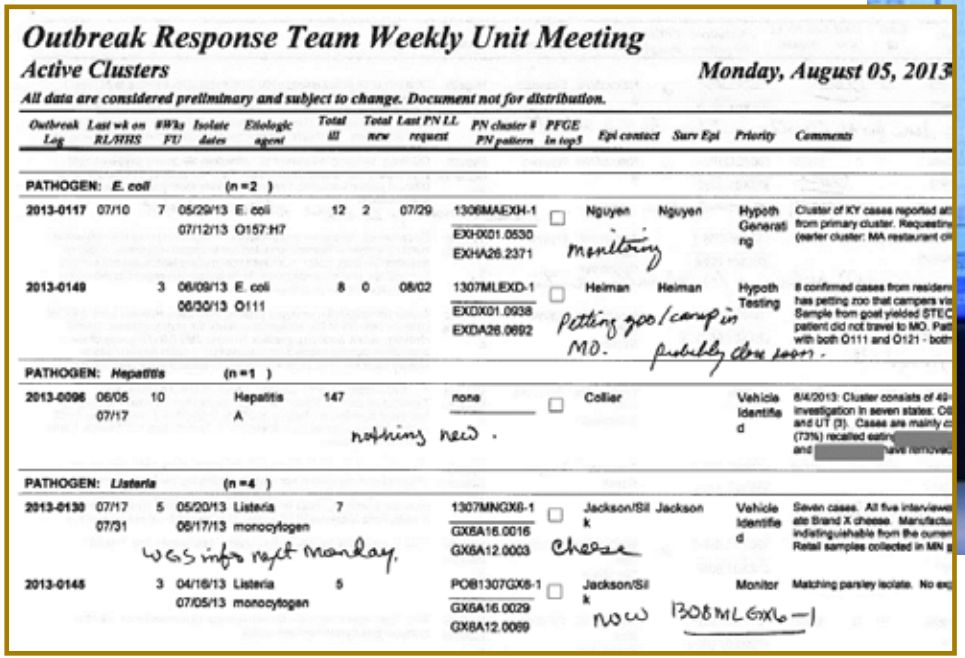
Robert L. Scharff, PhD, JD,[1] John Besser, PhD,[2] Donald J. Sharp, MD,[2] Timothy F. Jones, MD,[3]
Peter Gerner-Smidt DMS, MD,[2] Craig W. Hedberg, PhD[4]

The PulseNet surveillance system is a molecular subtyping network of public health and food regulatory agency laboratories designed to identify and facilitate investigation of foodborne illness outbreaks. This study estimates health an staggered adoption of PulseNet across the effectiveness, which is measured as reductio enhanced industry accountability, and mo

Each year, an estimated.....

- 270,000 cases are prevented
- $507,000,000 saved

National Cluster Investigations
(30 – 60 clusters monitored per week)



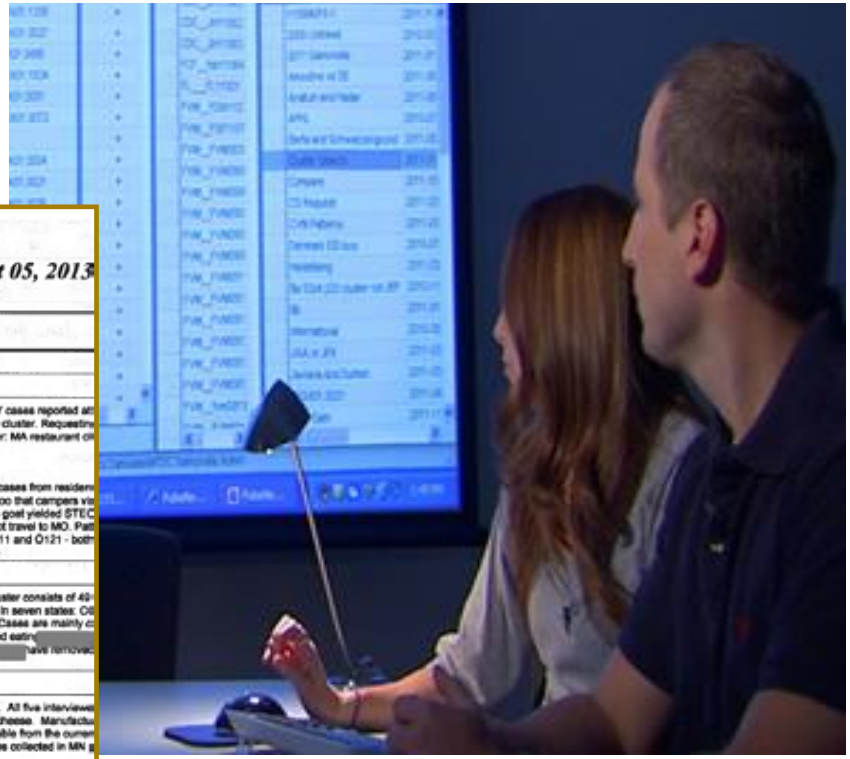
Outbreak Response Team Weekly Unit Meeting

Active Clusters — ***Monday, August 05, 2013***

All data are considered preliminary and subject to change. Document not for distribution.

Outbreak Log	Last wk on RL/HHS	#Wks FU	Isolate dates	Etiologic agent	Total ill	Total new	Last PN LL request	PN cluster # / PN pattern	PFGE in top5	Epi contact	Surv Epi	Priority	Comments
PATHOGEN: *E. coli*													
2013-0117	07/10	7	05/29/13 07/12/13	E. coli O157:H7	12	2	07/29	1306MAEXH-1 / EXHX01.0530 / EXHA26.2371	☐	Nguyen monitoring	Nguyen	Hypoth Generating	Cluster of KY cases reported at[illegible] from primary cluster. Requestin[illegible] (earlier cluster: MA restaurant [illegible]
2013-0149		3	06/09/13 06/30/13	E. coli O111	8	0	08/02	1307MLEXD-1 / EXDX01.0938 / EXDA26.0692	☐	Heiman petting zoo/camp in MO. probably close contact.	Heiman	Hypoth Testing	8 confirmed cases from residen[illegible] has petting zoo that campers vi[illegible] Sample from goat yielded STEC[illegible] patient did not travel to MO. Pati[illegible] with both O111 and O121 - both[illegible]
PATHOGEN: *Hepatitis*													
2013-0096	06/05 07/17	10		Hepatitis A	147			none	☐	Collier nothing new.		Vehicle Identified	6/4/2013: Cluster consists of 49[illegible] investigation in seven states: Co[illegible] and UT (3). Cases are mainly c[illegible] (73%) recalled eatin[illegible] and [redacted] have removac[illegible]
PATHOGEN: *Listeria*													
2013-0130	07/17 07/31	5	05/20/13 06/17/13	Listeria monocytogen	7			1307MNGX6-1 / GX6A16.0016 / GX6A12.0003	☐	Jackson/Silk cheese	Jackson	Vehicle Identified	Seven cases. All five interviewe[illegible] ate Brand X cheese. Manufactu[illegible] indistinguishable from the outbre[illegible] Retail samples collected in MN g[illegible]
2013-0145		3	04/16/13 07/05/13	Listeria monocytogen	5			POB1307GX6-1 / GX6A16.0029 / GX6A12.0069	☐	Jackson/Silk now B08MLGX6-1		Monitor	Matching parsley isolate. No exp[illegible]

was info rept Monday.



State and Local Cluster Investigations
~ 1,500 per year

PulseNet: A Distributed Laboratory Network

- Local testing
- Local control of analysis using standard computer
- Centralized quality assurance
- Centralized bioinformatics / high performance computing (WGS)

PulseNet: A Distributed Laboratory Network

Rationale…..

- Functionality for local, national, and international investigations
- Rapid turnaround time
- Large pool of resources
- Local control of patient and commercial confidential information

PFGE à WGS: What has *NOT* changed? (1)

	PFGE	WGS
Subtyping methods work by grouping together cases most likely to share an common exposure (such as a food)	Yes	Yes
“Matches” between cases and food/environmental isolates provide a hypothesis for investigators	Yes	Yes
The historical database is routinely examined for “matches” to current clusters	Yes	Yes

PFGE à WGS: What has *NOT* changed? (2)

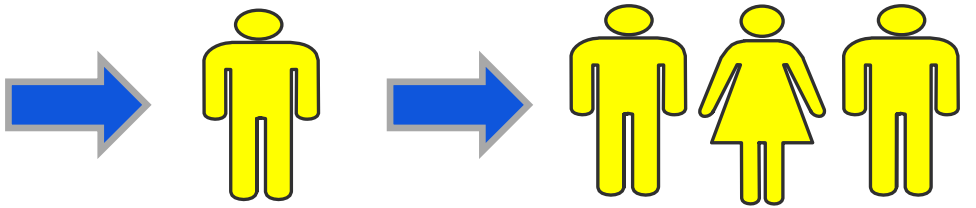
Comparisons between clinical cases and comparisons with food and environmental isolates

	PFGE	WGS
A "match" means that an association is *more likely* than if there is no match	Yes	Yes
A "miss-match" means an association is *less likely* (but not impossible)	Yes	Yes
Other types of data are needed to support conclusions*	Yes	Yes

* Such as epidemiological data, trace-back information

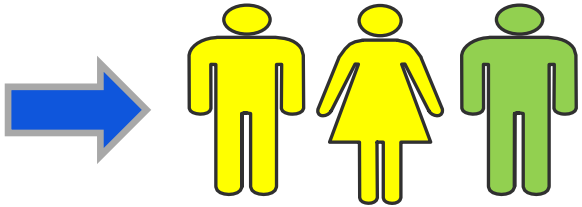
Strain Diversity and Outbreak Ecology: A Spectrum

Point source, single agent contamination



Closely related isolates

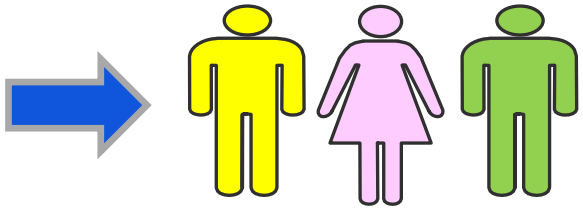
Point source, direct contact with environmentally contaminated commodities



Some diversity

Mass or chronic contamination

Contaminated irrigation water

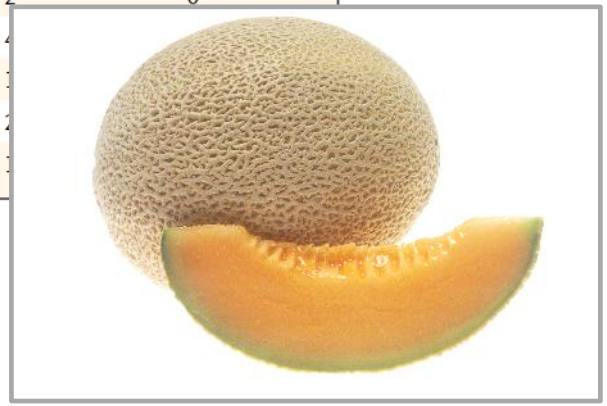


High diversity

Single Outbreaks May Involve Multiple Strains, Serotypes, Even Species

Table 3. Serotypes and Number of Illnesses, Deaths, Cantaloupe Samples, and Environmental Swabs Yielding Outbreak-Related *Listeria monocytogenes* Isolate Subtypes, According to PFGE Pattern.

PFGE Pattern No	PulseNet Pattern Name*		Serotype	Illnesses (N=147)	Deaths (N=33)	Source of Cantaloupe Samples†			Environmental Swabs from Processing Facility‡
	*Asc*I Restriction Enzyme	*Apa*I Restriction Enzyme				Farm A Cooler§	Retail Location	Patient's Home	
				no. (%)¶		*no. of samples*			*no.*
1	GX6A16.0029	GX6A12.0069	1/2a	48 (33)	12 (36)	0	3	2	0
2	GX6A16.0019	GX6A12.0227	1/2b	40 (27)	10 (30)	4	16	4	
3	GX6A16.0099	GX6A12.0001	1/2a	28 (19)	3 (9)	0	2	1	
4	GX6A16.0001	GX6A12.0001	1/2a	30 (20)	7 (21)	1	2	2	
5	GX6A16.1471	GX6A12.2154	1/2b	1 (<1)	1 (3)	0	0	1	



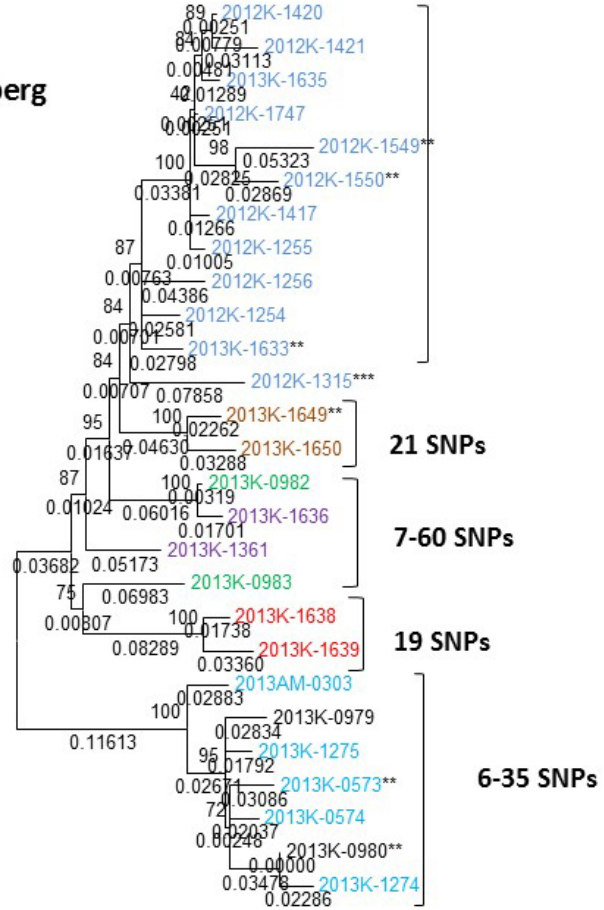
McCollum et al. NEJM. 2013 369(10):944-53

Single Outbreaks May Involve Multiple Strains, Serotypes, Even Species

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



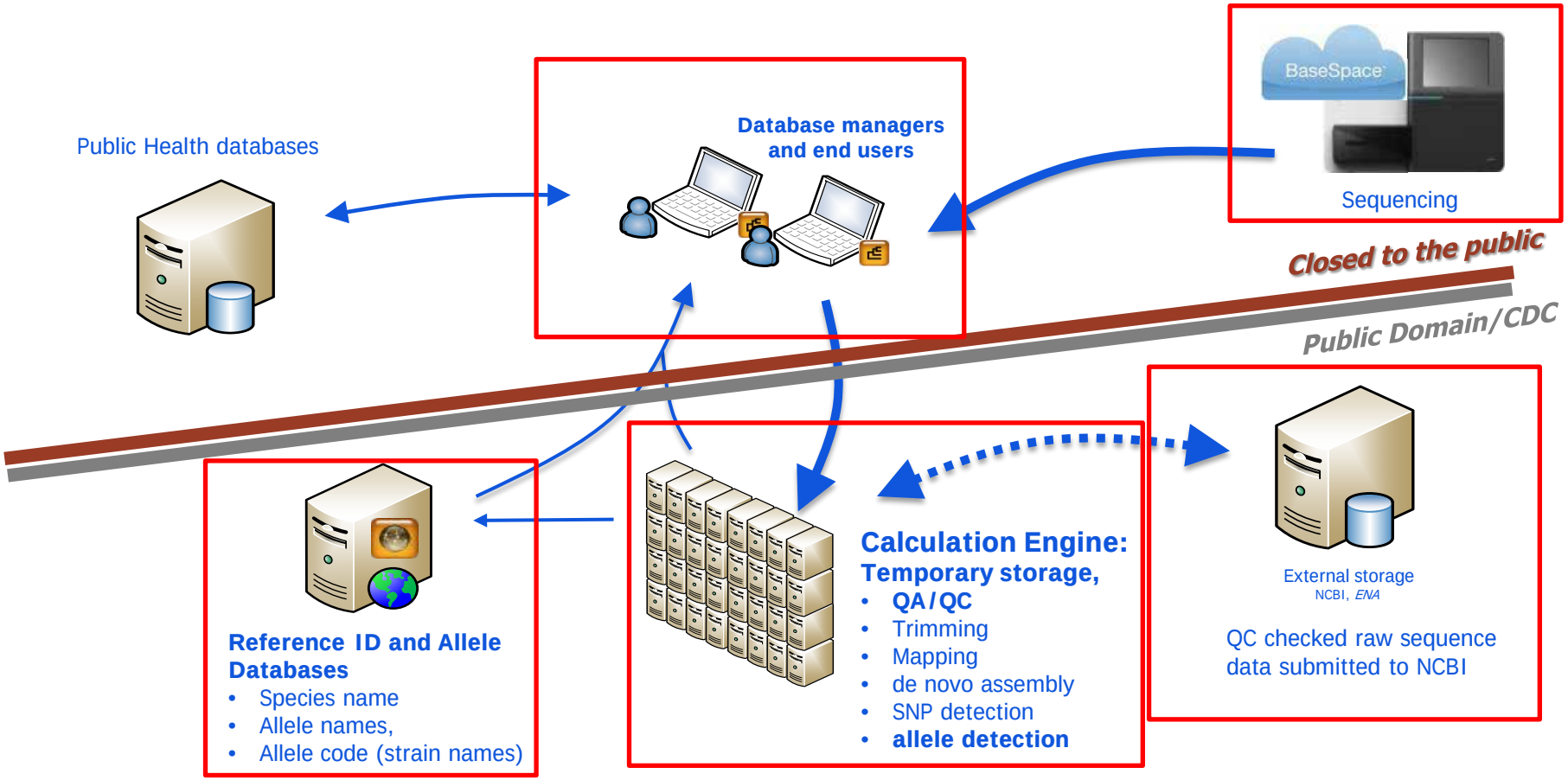
** Chicken from manufacturer X

*** Clinical isolate from 2011

Whole Genome Sequencing (WGS) at CDC (PulseNet)

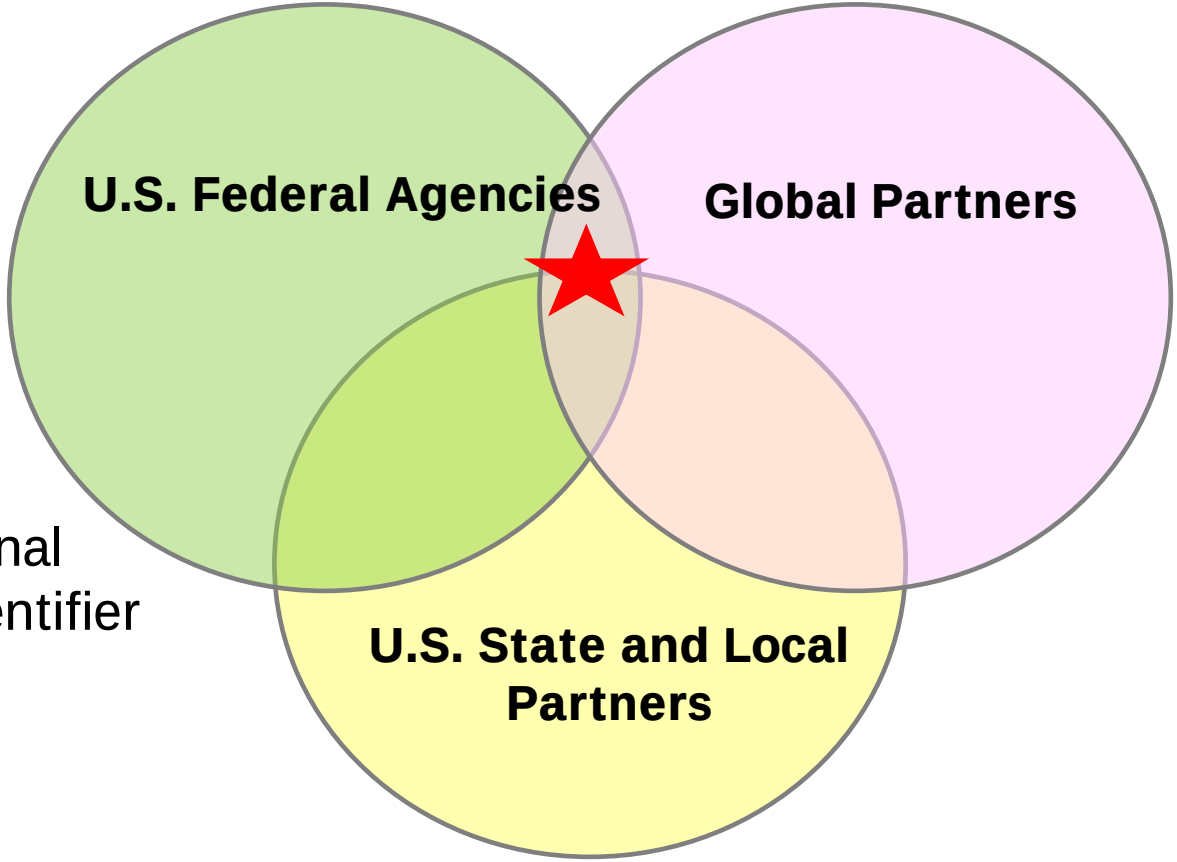
- Background
- What has *not* changed from the PFGE era?
- The PulseNet WGS infrastructure
- How is WGS different?
- The road ahead

WGS analysis workflow in PulseNet Labs: BioNumerics

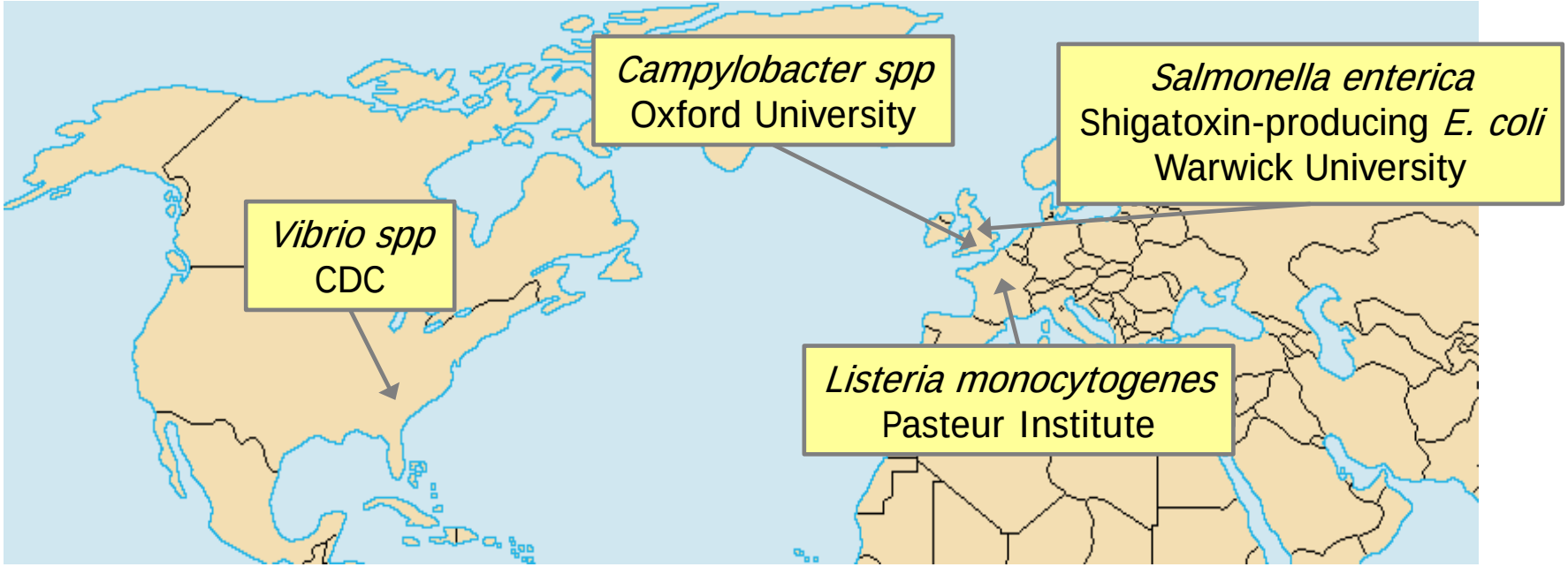


Coordination

- PulseNet USA
- Genome TrakR
- Gen-FS
- PulseNet International
- Global Microbial Identifier



Core Genome MLST (cgMLST) Scheme Development



PulseNet International: Global Vision

Eurosurveillance, Volume 22, Issue 23, 08 June 2017

Review

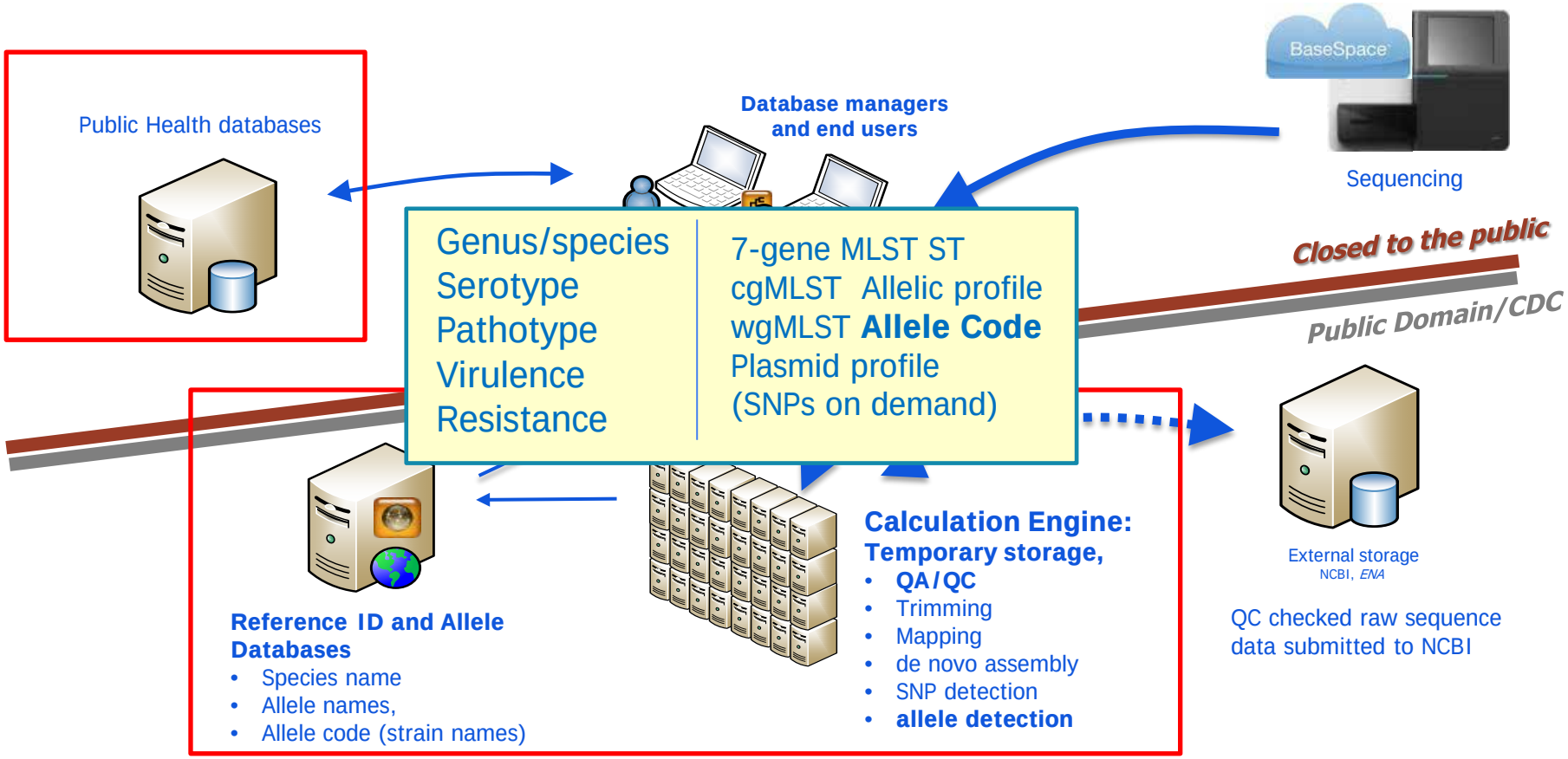
PULSENET INTERNATIONAL: VISION FOR THE IMPLEMENTATION OF WHOLE GENOME SEQUENCING (WGS) FOR GLOBAL FOOD-BORNE DISEASE SURVEILLANCE

C Nadon [1 2], I Van Walle [2 3], P Gerner-Smidt [4], J Campos [5], I Chinen [5], J Concepcion-Acevedo [4], B Gilpin [6], AM Smith [7], KM Kam [8], E Perez [9], E Trees [4], K Kubota [10], J Takkinen [3], EM Nielsen [11], H Carleton [4], FWD-NEXT Expert Panel [12]

+ Author affiliations

1. Public Health Agency of Canada, National Microbiology Laboratory, Canada
2. These authors contributed equally to this work
3. European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden
4. Centers for Disease Control and Prevention, United States
5. National Institute of Infectious Diseases "Dr Carlos G. Malbran", Argentina
6. Institute of Environmental Science and Research Limited; Christchurch, New Zealand
7. National Institute for Communicable Diseases, South Africa
8. Chinese University of Hong Kong, Hong Kong Special Adminstrative Region, China
9. Pan American Health Organization/World Health Organization, Washington, DC, United States
10. Association of Public Health Laboratories, United States
11. Statens Serum Institut, Denmark
12. The members of the FWD-NEXT Expert Panel are listed at the end of the article

WGS analysis workflow in PulseNet Labs: BioNumerics



Reference Characterization by WGS:

'One Shot' Characterization Of STEC

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service

Centers for Disease Control
and Prevention (CDC)
Atlanta GA 30333

Genus / Species: *Escherichia coli*

Serotype: O104:H4

Pathotype: Shiga toxin-producing and enteroaggregative *E. coli* (STEC/EAEC)

Virulence profile: *stx2a, aggR, aggA, sigA, sepA, pic, aatA, aaiC, aap*

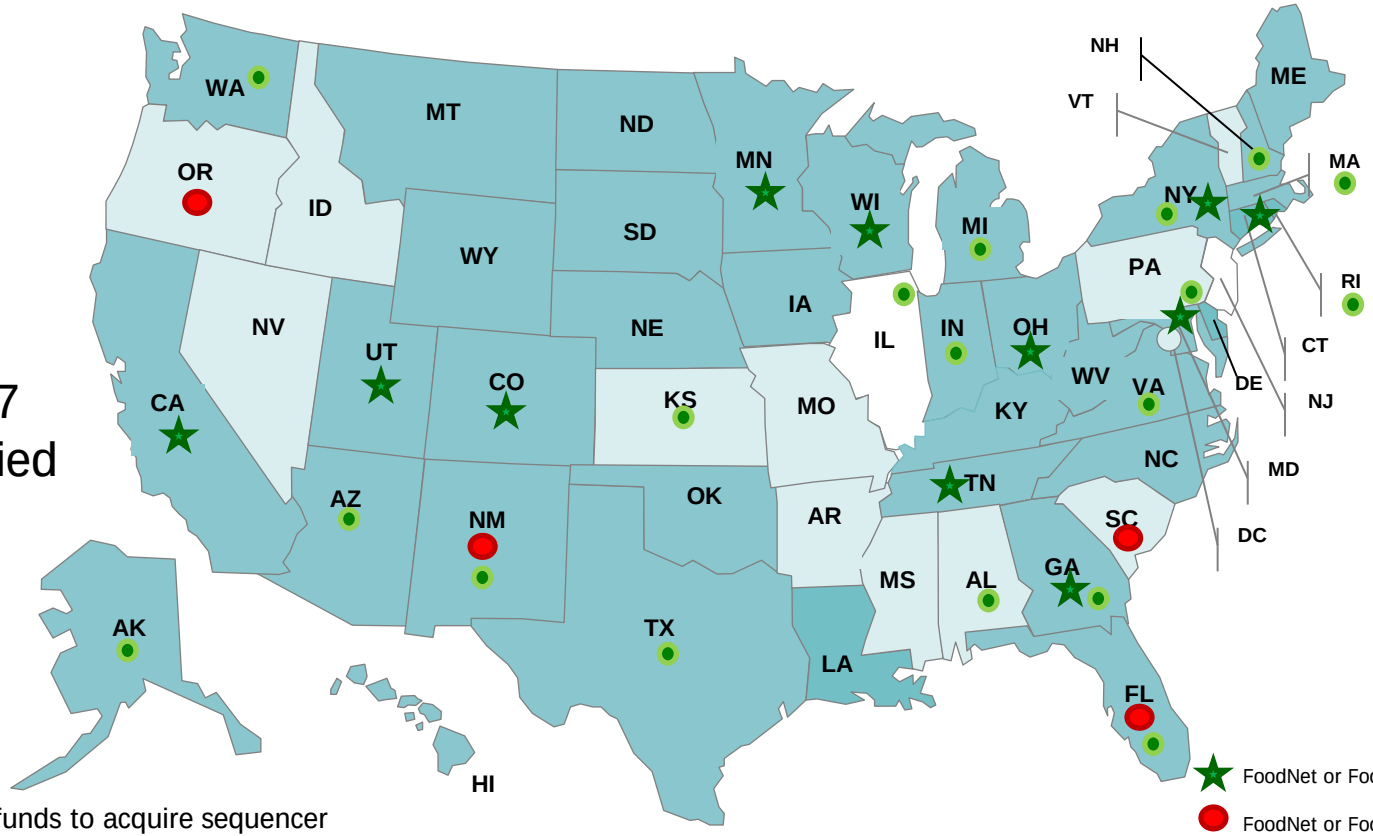
Sequence Type: ST678

wgMLST code: 102.45.26.35.3

Antimicrobial resistance genes: bla_{TEM-1}, $bla_{CTX-M-15}$, *strAB, sul2, tet(A)A, dfrA7*

PulseNet Sequencing Capacity – September 2017

43 labs in 37 states certified



Awarded ELC funds to acquire sequencer
Has sequencer but not certified
PulseNet certified

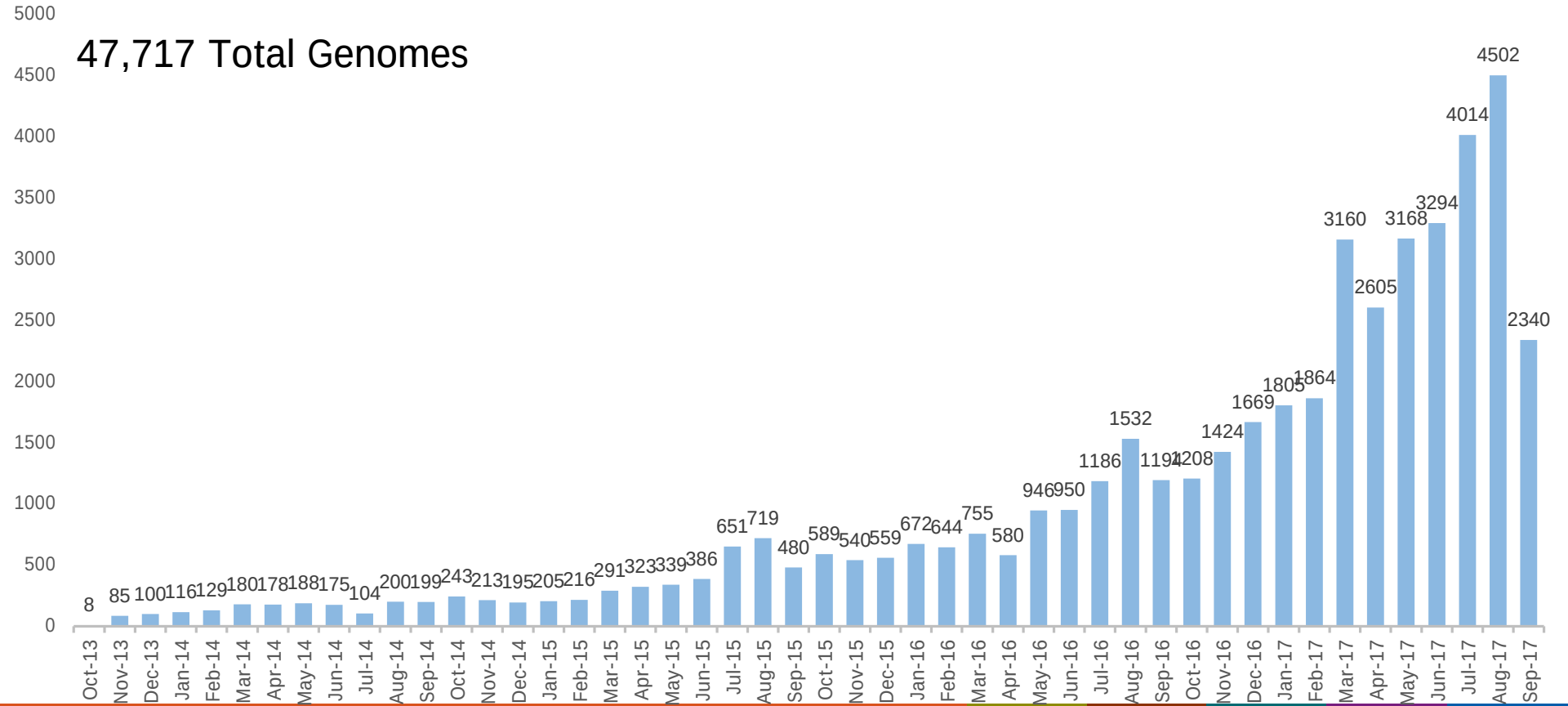
FoodNet or FoodCore with WGS capacity
FoodNet or FoodCore without WGS capacity
Outbreak Net Enhanced

Subtyping by WGS: whole genome MLST

Database	Current Version	Total Loci: Whole Genome (wgMLST)	Total Loci: Core Genome (cgMLST)	Current Status	Projected Release to Pilot
Campylobacter	v5	6,651	1,343 (*C. jejuni, C. coli*)	Validation	Early 2018
Escherichia	v4	34,453	2,513 (*E. coli*)	Validation	Early 2018
Listeria	Final (v4)	4804	1,748 (*L. monocytogenes*)	In Pilot	released
Salmonella	v4	19,384	3,002 (*S. enterica*)	Validation	Early 2018
Vibrio	TBD	TBD	TBD	In Development	TBD

Total Genomes Sequenced and Submitted to PulseNet

(Through September, 2017)



PulseNet Cases (Human): U.S.*

Disease	Cases/year
Listeriosis	800
Shiga toxin-producing *E coli* *disease*	5,000
Salmonellosis	42,000

*laboratory confirmed, approximate

PFGE à WGS: What is DIFFERENT? (1)

	PFGE	WGS
Specificity	High for some organisms, low for others	Very high
Our ability to evaluate the "closeness" of strains	Low	High
Nature of the data	Categorical*	Continuous

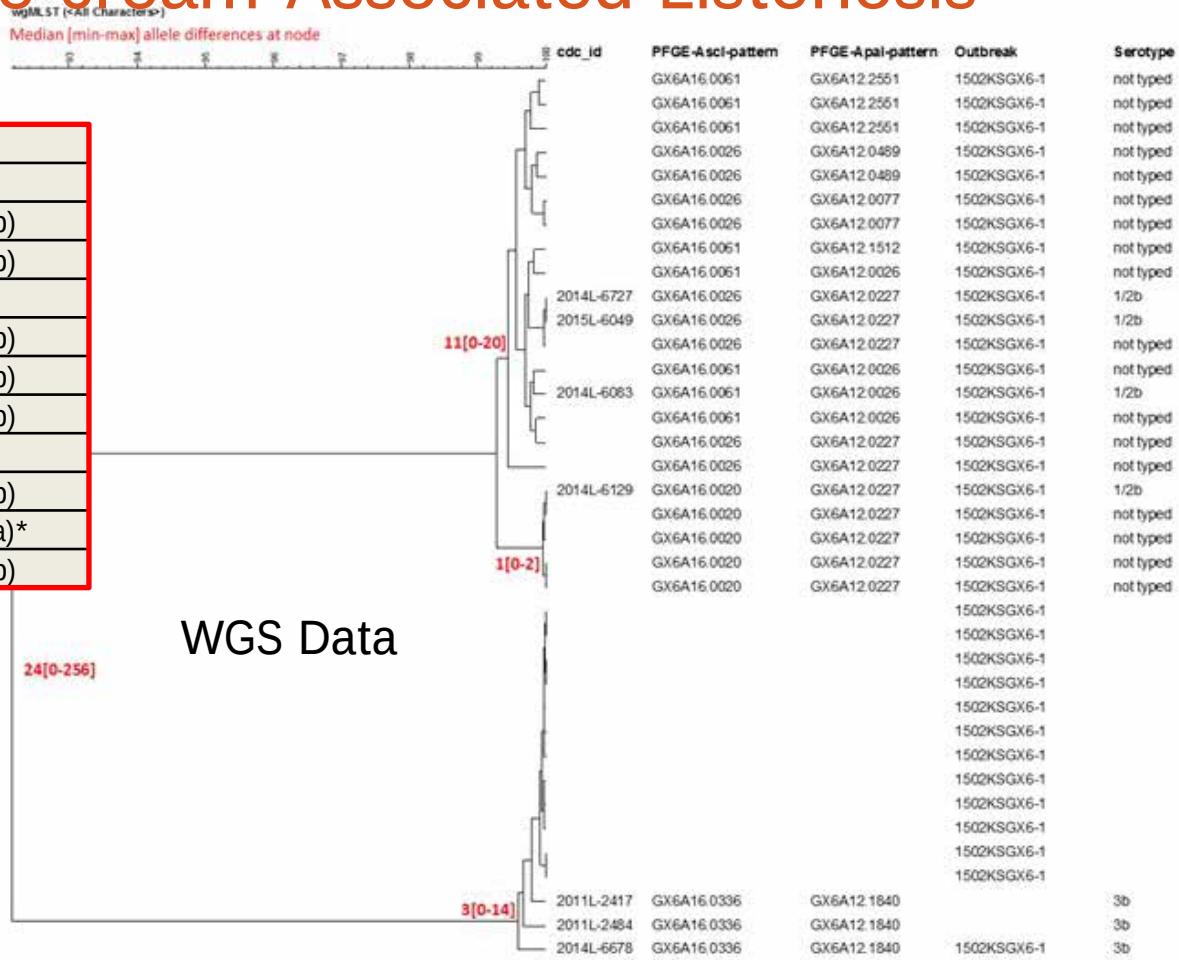
*i.e. patterns are either indistinguishable or not

Ice cream-Associated Listeriosis

Patterns in OB (11)
GX6A16.0336/GX6A12.1840 (3b)
GX6A16.0061/GX6A12.0026 (1/2b)
GX6A16.0026/GX6A12.0227 (1/2b)
GX6A16.0336/GX6A12.2255 (3b)
GX6A16.0020/GX6A12.0227 (1/2b)
GX6A16.0026/GX6A12.0489 (1/2b)
GX6A16.0061/GX6A12.2551 (1/2b)
GX6A16.0617/GX6A12.1840 (3b)
GX6A16.0026/GX6A12.0077 (1/2b)
GX6A16.0282/GX6A12.0355 (1/2a)*
GX6A16.0061/GX6A12.1512 (1/2b)

PFGE Data

WGS Data



Ice cream-Associated Listeriosis

wgMLST (<All Characters>)
Median [min-max] allele differences at node

Patterns in OB (11)
GX6A16.0336/GX6A12.1840 (3b)
GX6A16.0061/GX6A12.0026 (1/2b)
GX6A16.0026/GX6A12.0227 (1/2b)
GX6A16.0336/GX6A12.2255 (3b)
GX6A16.0020/GX6A12.0227 (1/2b)
GX6A16.0026/GX6A12.0489 (1/2b)
GX6A16.0061/GX6A12.2551 (1/2b)
GX6A16.0617/GX6A12.1840 (3b)
GX6A16.0026/GX6A12.0077 (1/2b)
GX6A16.0282/GX6A12.0355 (1/2a)*
GX6A16.0061/GX6A12.1512 (1/2b)

PFGE Data

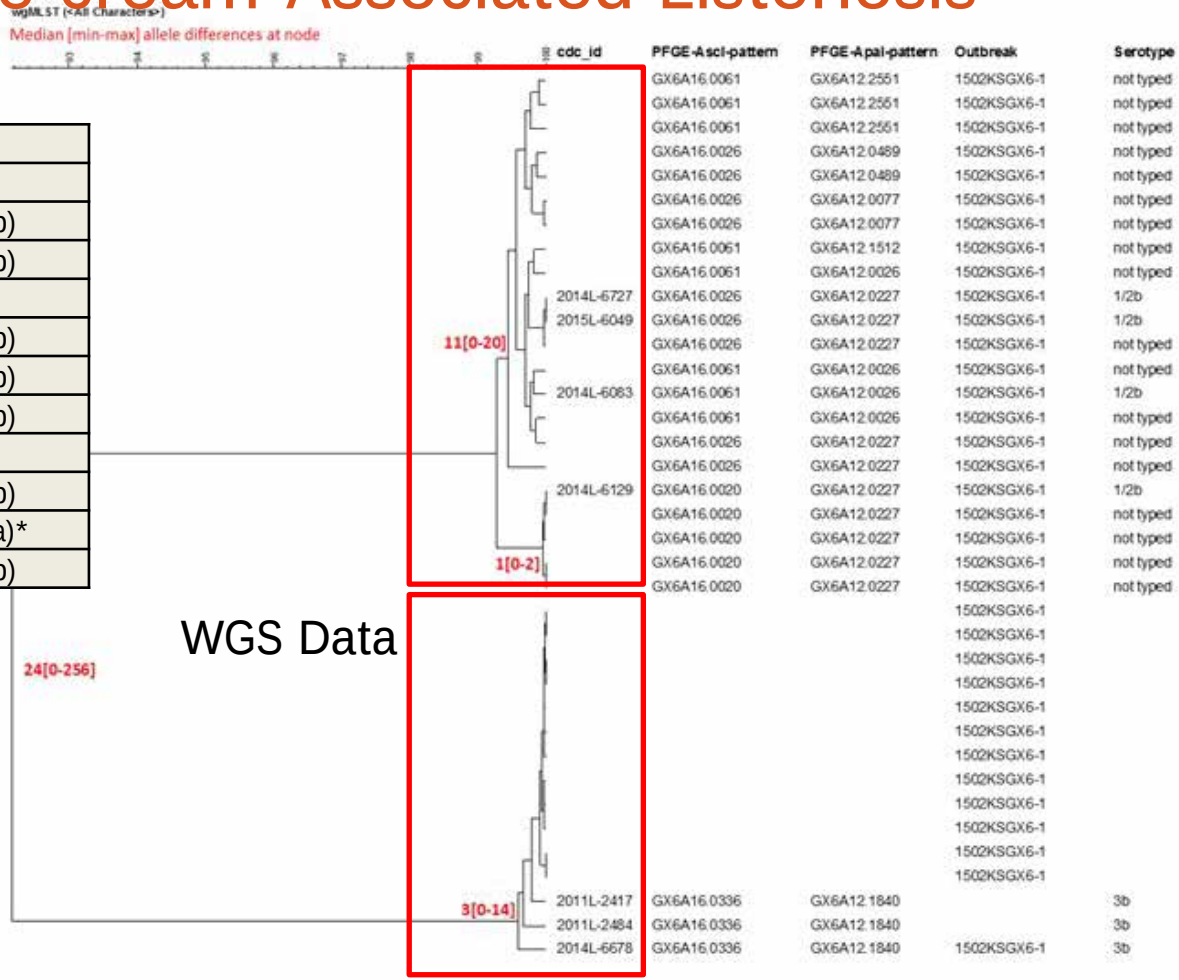
cdc_id	PFGE-AscI-pattern	PFGE-ApaI-pattern	Outbreak	Serotype
GX6A16.0026	GX6A12.0227	1502KSGX6-1		1/2b
GX6A16.0026	GX6A12.0227	1502KSGX6-1		1/2b
GX6A16.0026	GX6A12.0227	1502KSGX6-1		not typed
GX6A16.0026	GX6A12.0227	1502KSGX6-1		not typed
GX6A16.0026	GX6A12.0227	1502KSGX6-1		not typed
GX6A16.0020	GX6A12.0227	1502KSGX6-1		1/2b
GX6A16.0020	GX6A12.0227	1502KSGX6-1		not typed
GX6A16.0020	GX6A12.0227	1502KSGX6-1		not typed
GX6A16.0020	GX6A12.0227	1502KSGX6-1		not typed
GX6A16.0020	GX6A12.0227	1502KSGX6-1		not typed

Ice cream-Associated Listeriosis

Patterns in OB (11)
GX6A16.0336/GX6A12.1840 (3b)
GX6A16.0061/GX6A12.0026 (1/2b)
GX6A16.0026/GX6A12.0227 (1/2b)
GX6A16.0336/GX6A12.2255 (3b)
GX6A16.0020/GX6A12.0227 (1/2b)
GX6A16.0026/GX6A12.0489 (1/2b)
GX6A16.0061/GX6A12.2551 (1/2b)
GX6A16.0617/GX6A12.1840 (3b)
GX6A16.0026/GX6A12.0077 (1/2b)
GX6A16.0282/GX6A12.0355 (1/2a)*
GX6A16.0061/GX6A12.1512 (1/2b)

PFGE Data

WGS Data

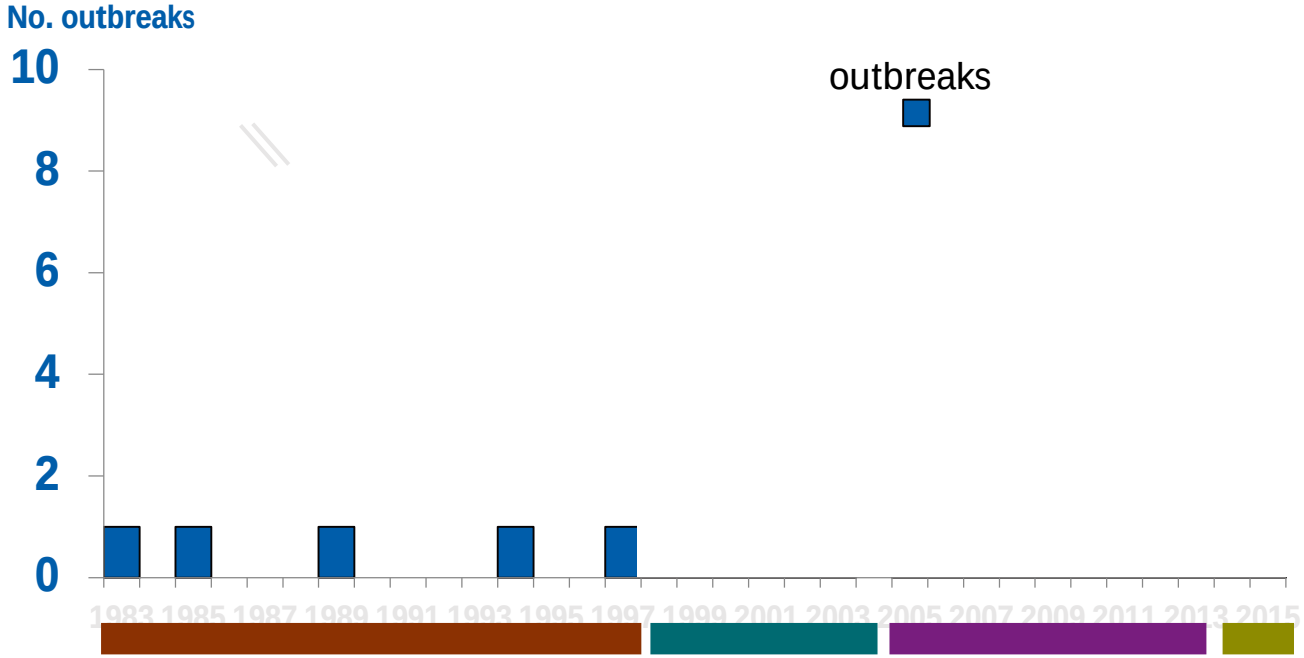


PFGE à WGS: What is DIFFERENT? (2)

	PFGE	WGS
Size of outbreaks (number of cases) when detected	Larger	Smaller
Number of outbreaks detected and solved	Less	More
Possible outbreaks *ruled out*	Less	More
Interpretation of case, food, and environmental "matches"	New or very rare PFGE pattern: Strong hypothesis* Common pattern: Weak hypothesis	Strongest hypothesis*

* Likely *some* association, although it may not be direct

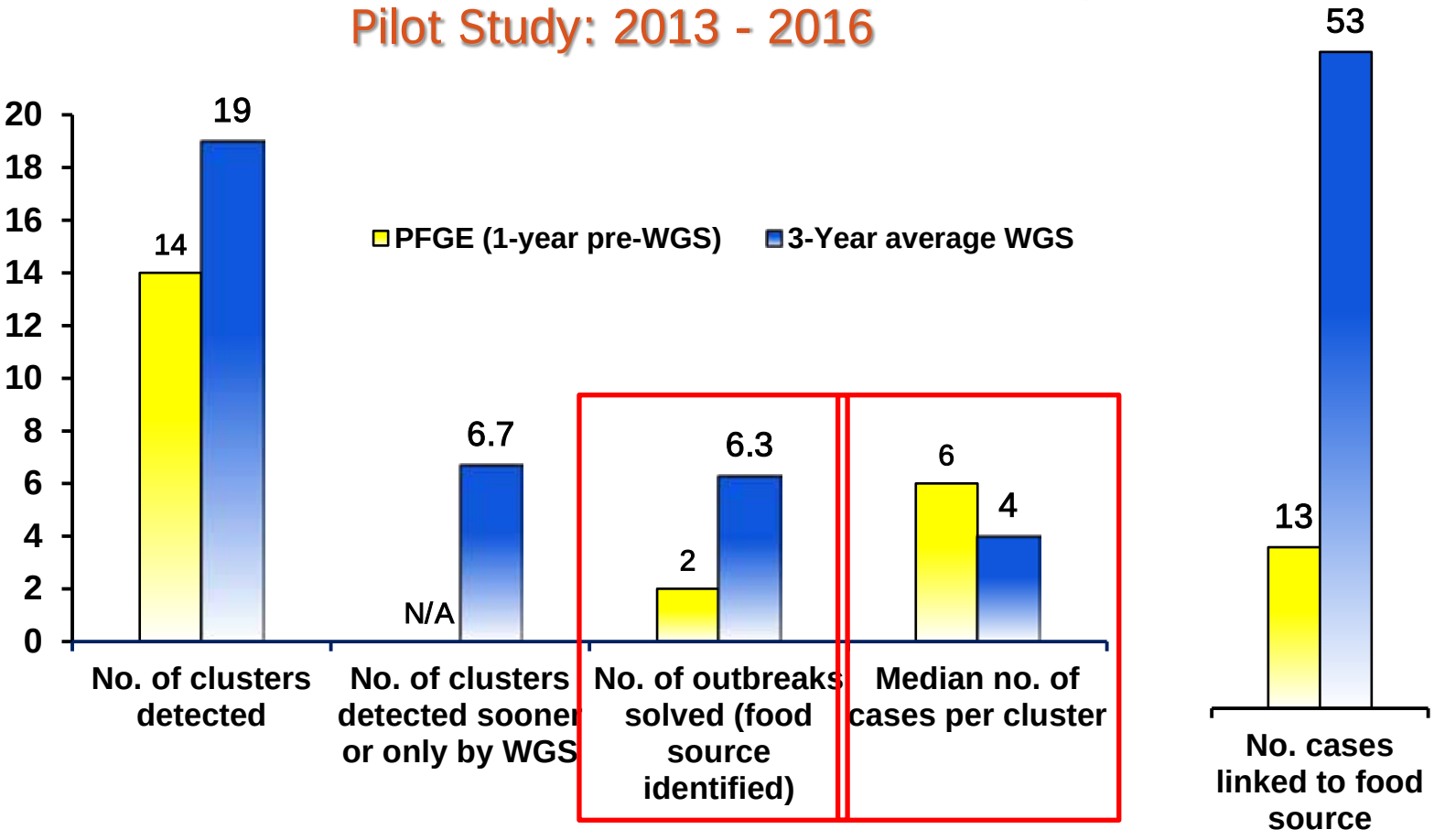
Listeriosis Outbreaks, 1983-2015



Era	Pre-PulseNet	Early PulseNet	*Listeria* Initiative	WGS
Outbreaks per year	0.3	2.3	2.6	7.5
Median cases per outbreak	69	11	5.5	4

*2015 incidence rate preliminary data from FoodNet

Collaborative Nation-Wide Real-time *Listeria monocytogenes* WGS Pilot Study: 2013 - 2016



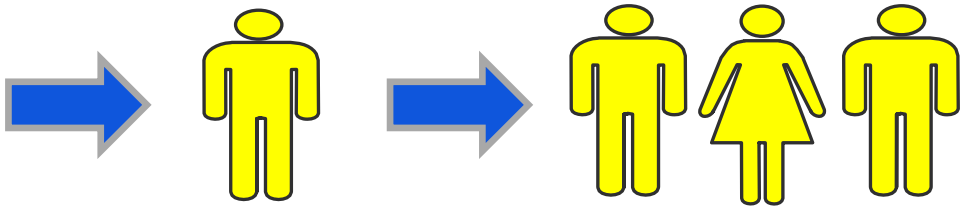
Is It Possible to have SNP/Allele Cutoff Values? (defining related / unrelated)

- Rule-of-thumb values used routinely

Current CDC “Rule-of-thumb” cutoffs	
Listeria monocytogenes	<20 alleles (<30 “watch”)
Salmonella / STEC; Suspected foodborne	<10 alleles or SNPs
Salmonella / STEC; Suspected zoonotic	< 20 alleles or SNPs

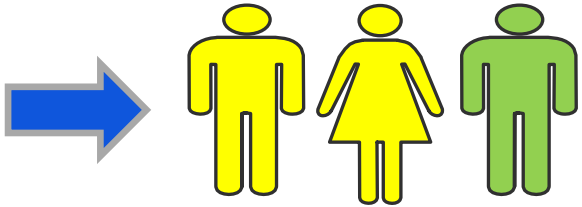
Strain Diversity and Outbreak Ecology: A Spectrum

Point source, single agent contamination



Closely related isolates

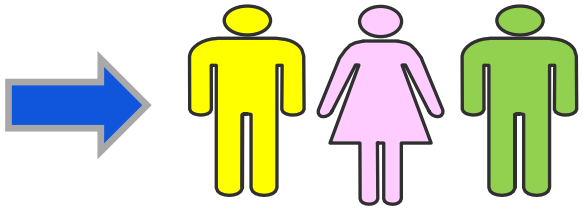
Point source, direct contact with environmentally contaminated commodities



Some diversity

Mass or chronic contamination

Contaminated irrigation water



High diversity

Is It Possible to have SNP/Allele Cutoff Values? (defining related / unrelated)

- Rule-of-thumb values used routinely

- “Absolute” cutoff values not possible

- Combining epidemiology, microbiology, and trace-back data is the best accurate way of “defining” clusters.

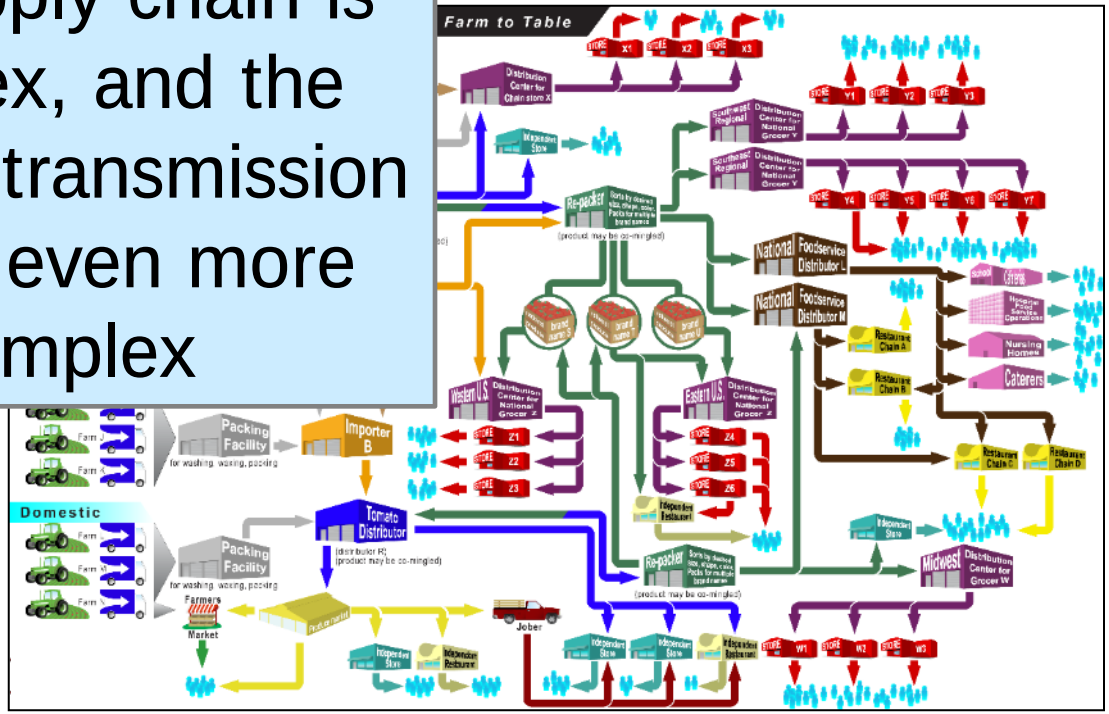
PFGE à WGS: What is DIFFERENT (2)?

	PFGE	WGS
Size of outbreaks (number of cases) when detected	Larger	Smaller
Number of outbreaks detected and solved	Less	More
Possible outbreaks *ruled out*	Less	More
Interpretation of case, food, and environmental "matches"	New or very rare PFGE pattern: Strong hypothesis* Common pattern: Weak hypothesis	Strongest hypothesis*

* Likely *some* association, although it may not be direct

…and that is just on the farm. Once it reaches the supply chain, things really "*simplify*".

The supply chain is complex, and the chain-of-transmission can be even more complex

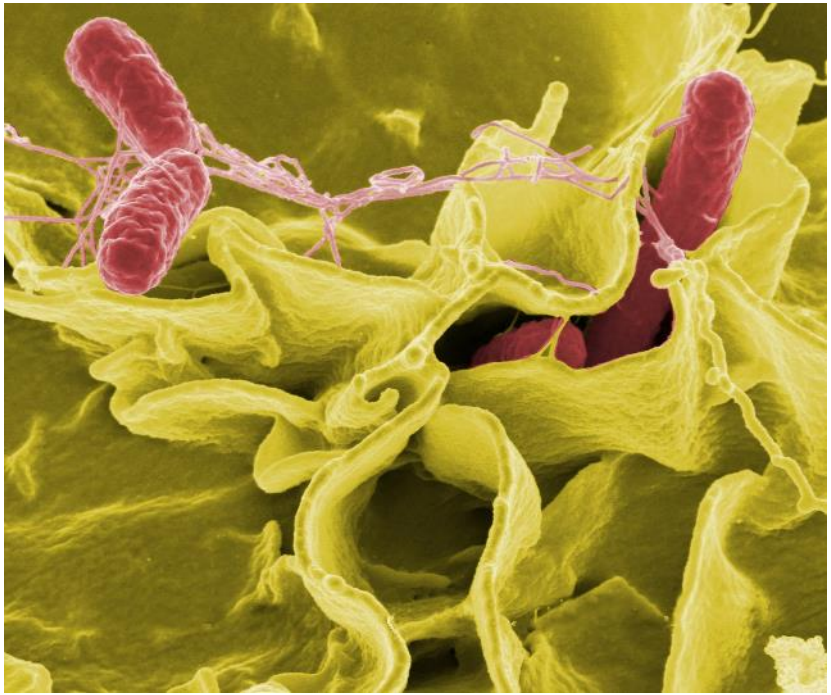


Schematic model of the Fresh-cut Tomato Supply Chain

Whole Genome Sequencing (WGS) at CDC (PulseNet)

- Background
- What has *not* changed from the PFGE era?
- The PulseNet WGS infrastructure
- How is WGS different?
- The road ahead

Using Big Data to Improve Food Safety

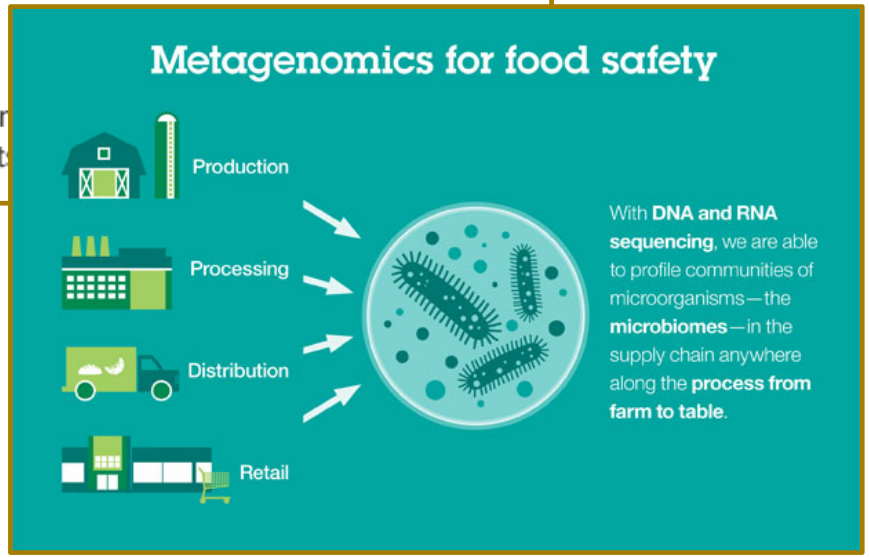


- Advanced analytics
- Metagenomics

Shotgun Metagenomics

IBM Research and Mars, Incorporated Launch Pioneering Effort to Drive Advances in Global Food Safety

New Sequencing the Food Supply Chain Consortiur
metagenomics study and unlock food safety insight

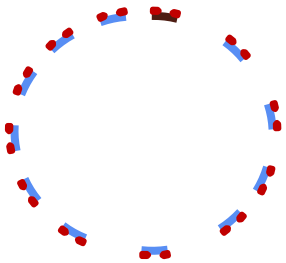


Selected Metagenomics Food Safety Applications

- Pathogen discovery
- *In situ* (direct-from-specimen) pathogen characterization
- Food identification (in and outside the body)
- Understanding (and controlling) spoilage, contamination
- Outbreak root cause analysis / environmental analysis
- Population biology (e.g. dysbiosis)
- Host factors

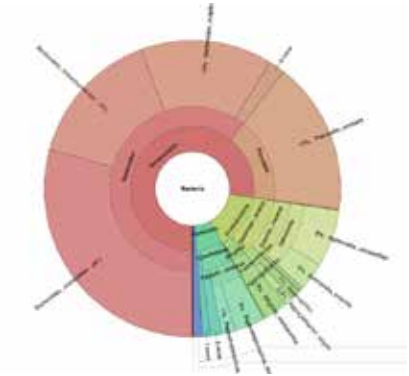
CDC Applied Research: Direct-from-Specimen Pathogen Characterization Development

Amplicon sequencing



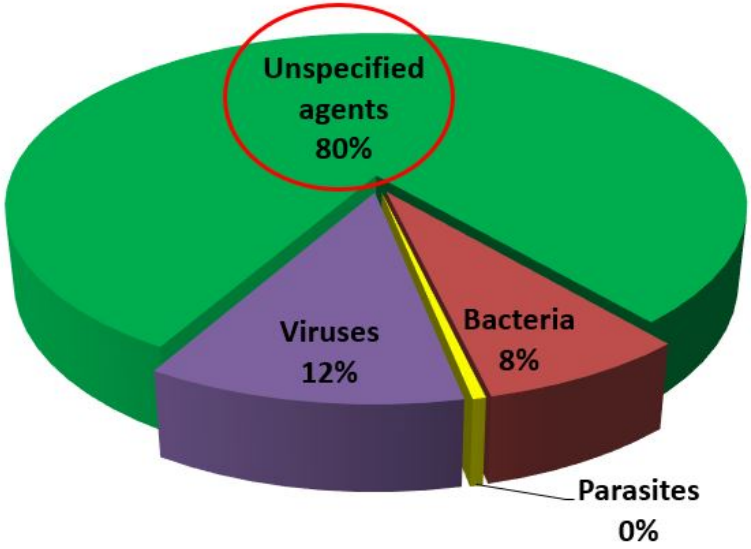
1,000's of MLST targets

Shotgun metagenomics



- Enrichment of pathogen targets
- Unbiased sequencing

Pathogen Discovery

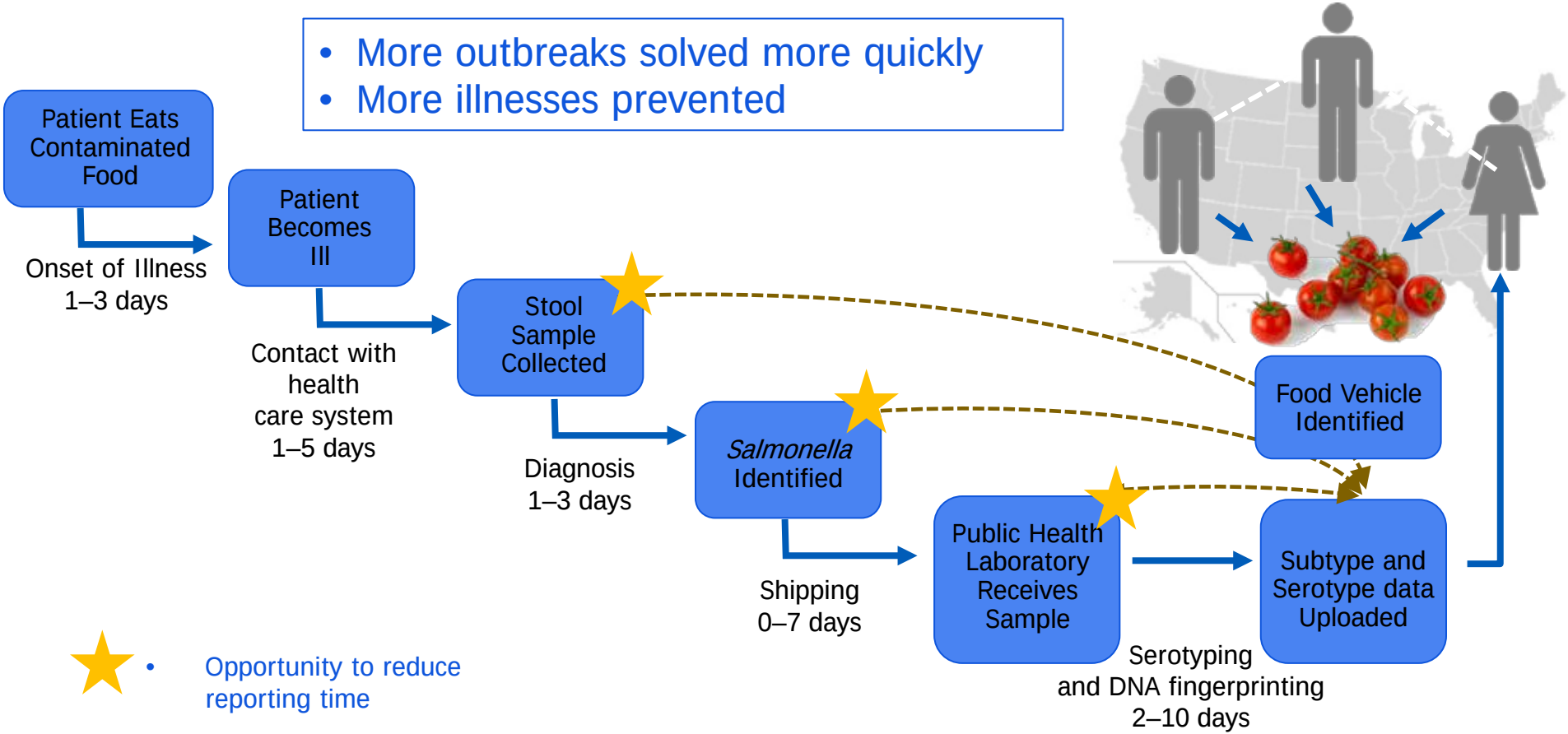


Etiology of Acute Gastroenteritis: U.S.[1,2]

[1]Scallan E, et al. Emerg Infect Dis. 2011 Jan;17(1):16-22
[2]Scallan E, et al. Emerg Infect Dis. 2011 Jan;17(1):7-15.

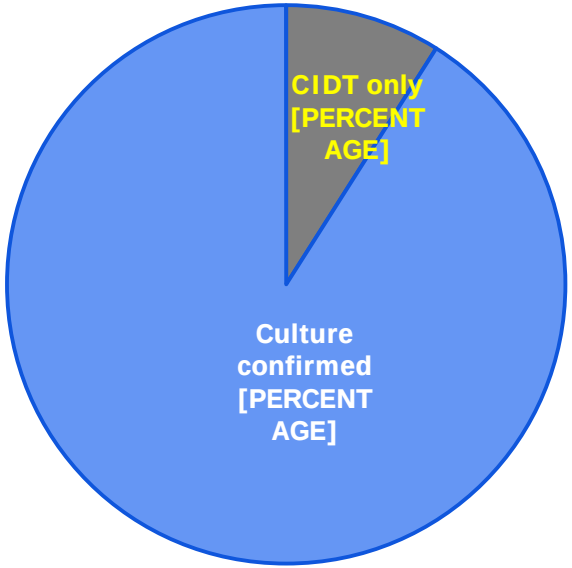
Direct-from-specimen Tests: Reduce Time to Actionable Results

- More outbreaks solved more quickly
- More illnesses prevented

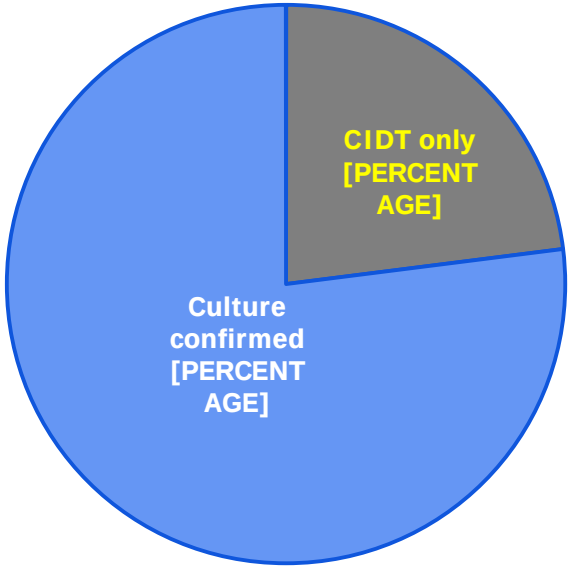


Use of Culture-Independent Diagnostic Tests (CIDTs) Are Increasing — FoodNet, 2012–2016

Annual percentage of bacterial infections diagnosed by CIDTs



2012–2015

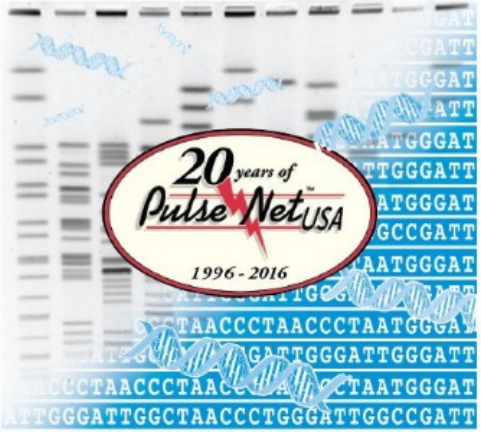


2016

Subtyping Methods: Isolate Dependency

Method	Isolates Required?
PFGE	Yes
WGS	Yes

Whole Genome Sequencing (WGS) at CDC (PulseNet)



The findings and conclusions in this presentation are those of the author and do not necessarily represent the views of the Centers for Disease Control and Prevention