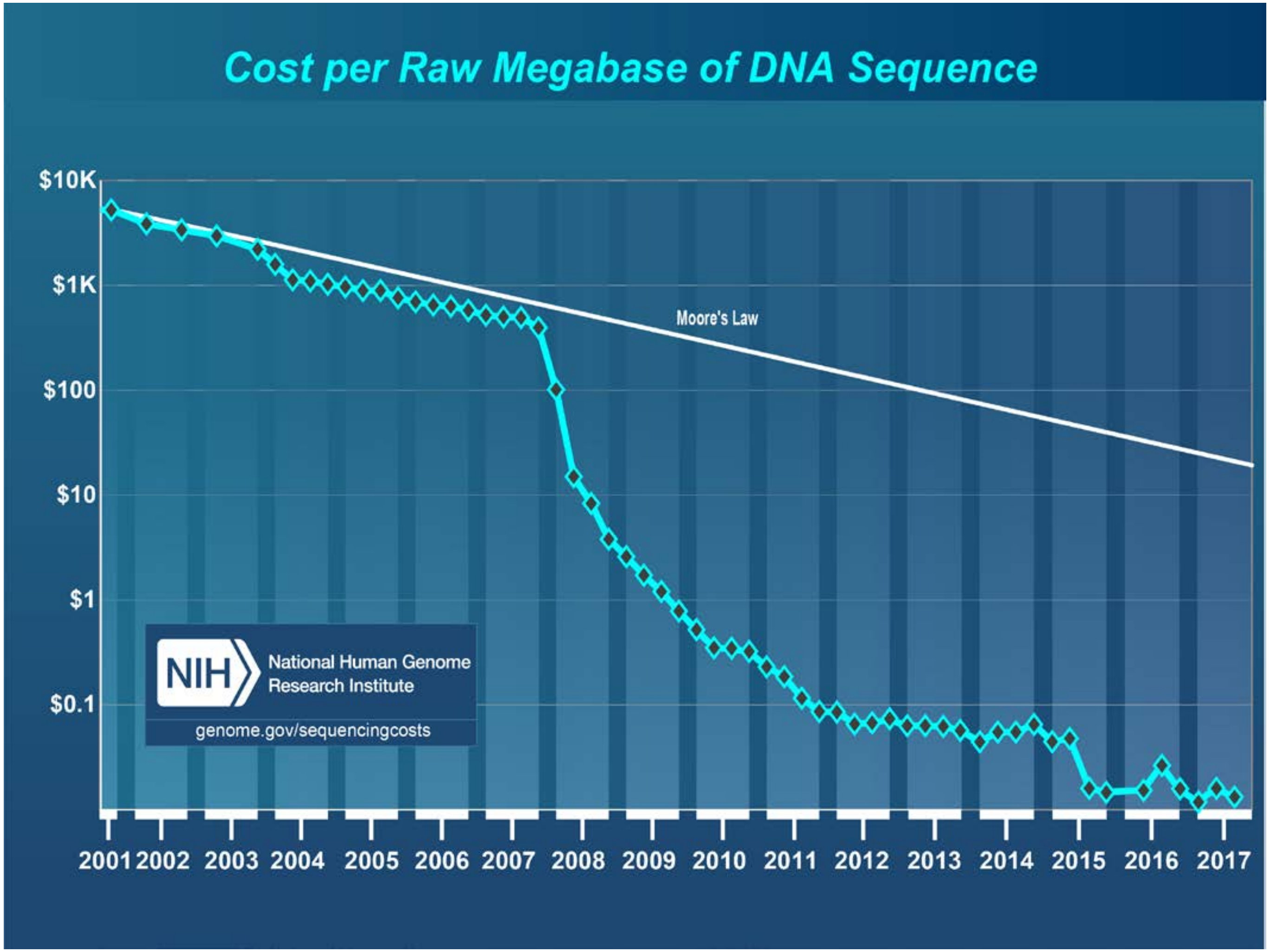
GenomeTrakr database and network: WGS network for real-time characterization and source tracking of foodborne pathogens

Marc Allard, Ph.D.

Center for Food Safety and Applied Nutrition,
US Food and Drug Administration

Almond Board Food Quality and Safety Symposium***: June 2019***

NGS Costs



genome.gov/sequencingcosts

Illumina Suite

- Massively parallel, short read, Sequencing-by-Synthesis
- Cluster generation, reversible dye-terminator dNTPs
- Instruments range from low to high throughput:
  iSeq, MiniSeq, MiSeq, NextSeq, HiSeq, NovaSeq
- Read Length:
  Paired end reads (forward and reverse)
  150-300 bp long (300-600 total)
- Max Output:
  4M to 20B Reads
  From 1.2 GB to 6 TB

https://www.illumina.com/systems/sequencing-platforms.html

Ion Torrent by ThermoFisher Scientific

- > Ion Torrent semiconductor sequencing
- > Measures hydrogen ions (pH) during sequencing-by-synthesis reactions
- > No modified bases required
- > Instruments: Ion S5, S5 Plus, S5 Prime
    - 200-600 bp read length
    - Up to 130M Reads
    - Scalable output from 15 to 25 GB

https://www.thermofisher.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-run-sequence/ion-s5-ngs-targeted-sequencing.html

SMRT Sequencing by PacBio

- Single Molecular, Real-Time (SMRT) Sequencing
- Zero-mode waveguides and phospholinked nucleotides (dyes)
- PacBio RS II – Original Long Read Sequencer
  - 60,000 bp max read length
  - 0.5-1 Gb per cell
  - Up to 16 cells per run
- PacBio Sequel – High throughput WGS
  - Targeting *de novo* assemblies
  - ~365,000 reads per cell
  - 5-8 Gb per cell

http://www.pacb.com/products-and-services/pacbio-systems/

Oxford Nanopore Technologies

> Single molecule, real-time, nanopore sequencing

> MinION Portable Sequencer

- 5-10 Gb per flow cell
- Extreme read lengths (record is > 2 MB)
- Real-time data analysis
- Compact size.

  Weighs <100g and plugs into PC via USB

> PromethION High-throughput Sequencer

- Up to 48 flowcells
- Simultaneous sample processing
- Increases data yield to TB

> SmidgION smartphone device in development

https://nanoporetech.com/

Why Develop a WGS based Network?

- Tracking and Tracing of food pathogens
  - Insufficient resolution of current tools
    -matching clinical to environmental
    -improve the environmental database
  - Faster identification of the food involved in the outbreak
  - Limited number of investigators vs. facilities and import lines
  - Global travel
  - Global food supply

Benefits of a WGS Approach

- More discriminatory and informative than PFGE
- Clues to geographic origin of pathogen

This means:

- Greater certainty when matching clinical, environmental, and product sample isolates
- Links between illnesses and the potential source of contamination can be made with fewer isolates
- Investigators can be deployed in a more targeted manner, saving resources

End Result:

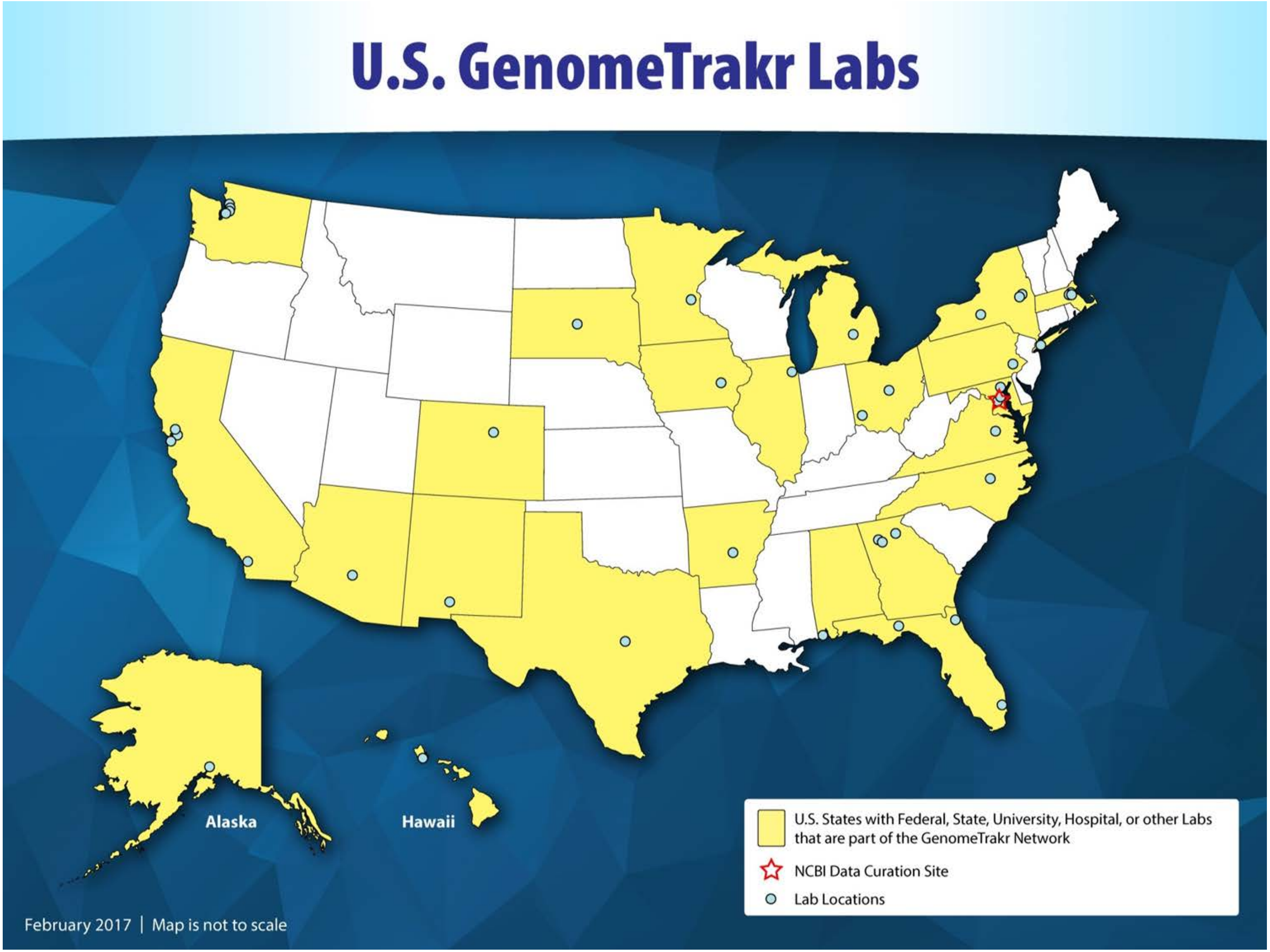
- Faster identification of the food involved in the outbreak
- Potential to help reduce the number of foodborne illnesses and deaths over the long term both in the U.S. and abroad.

Current Scope of GenomeTrakr Network

- Network includes labs at FDA, CDC, FSIS, 17 state health and university labs, 1 U.S. hospital lab, and 11 labs located outside the U.S.
  - Contributing labs are on 4 continents and in 10 countries
- The network provides high resolution genomic sequences of food pathogens, ex. *Salmonella, Listeria*, STEC's, others. Greater than 250,000 sequences in the database
- New GenomeTrakr labs are coming on-line
- Partnered with CDC in 2013 to study all clinical and environmental isolates of *Listeria monocytogenes, now E. coli, (Salmonella coming)*

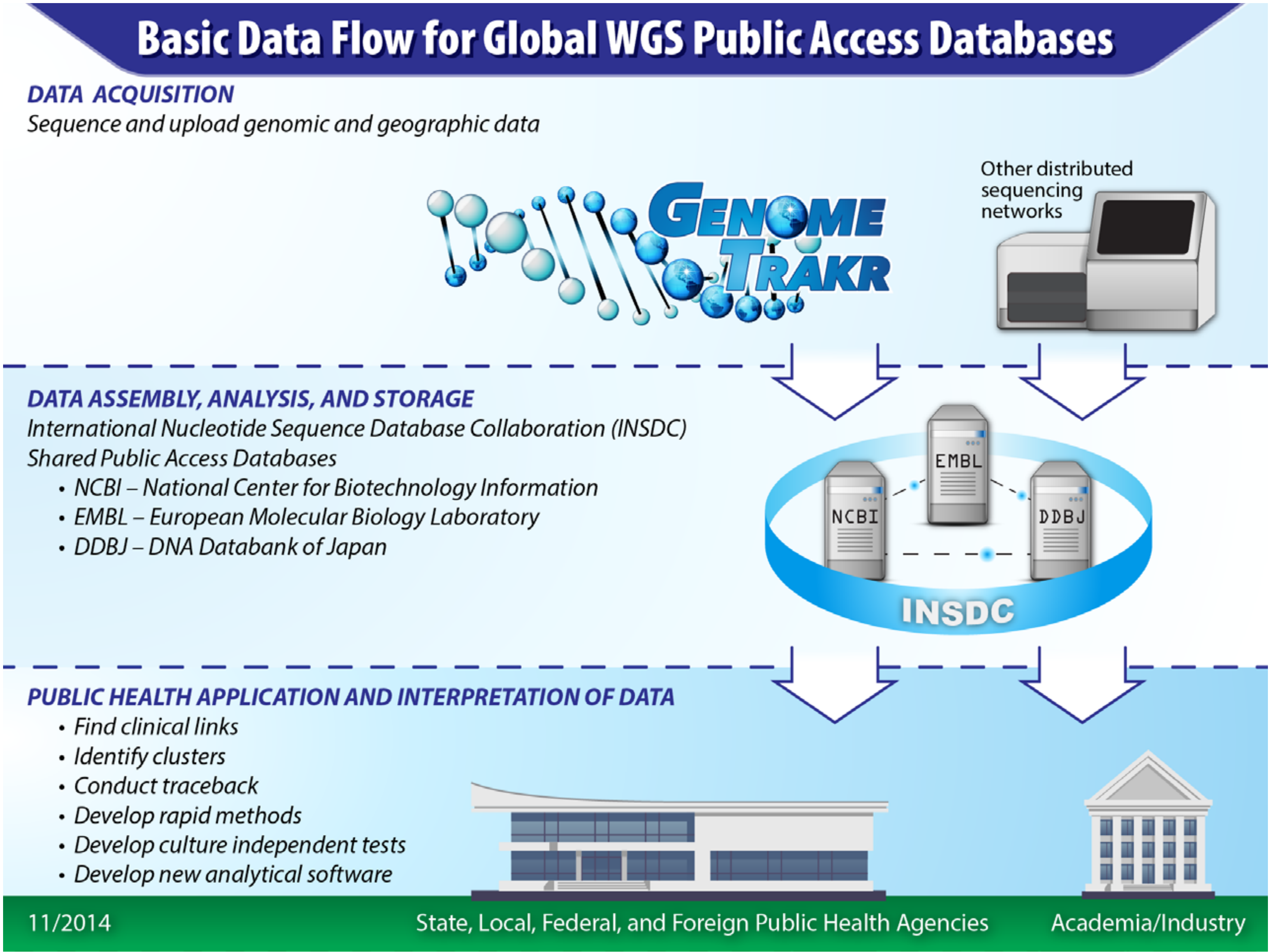
FDA GenomeTrakr website
http://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/ucm363134.htm

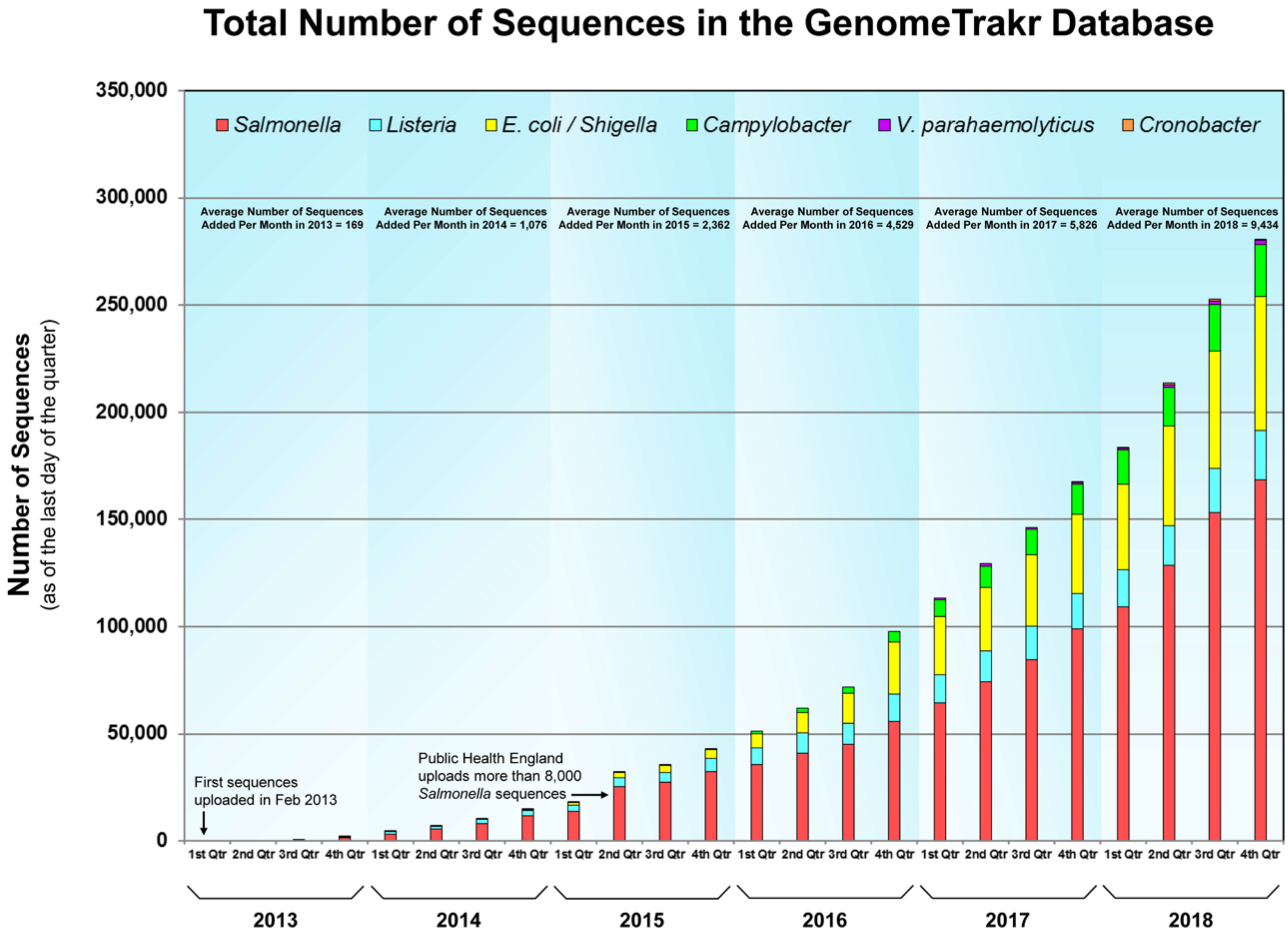
FDA U.S. FOOD & DRUG ADMINISTRATION



Labs Outside the U.S. Contributing to GenomeTrakr

U.S. FOOD & DRUG ADMINISTRATION





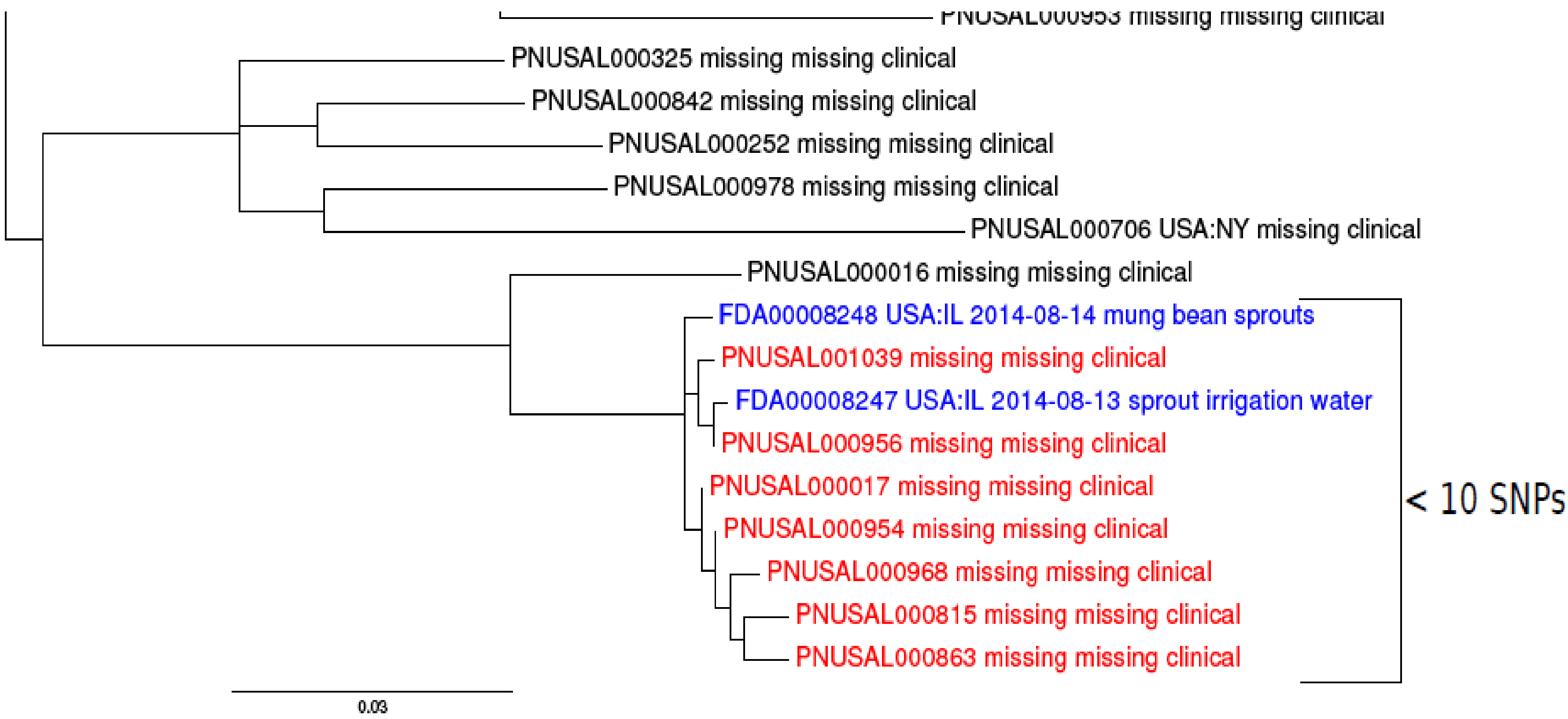
Total Number of Sequences in the GenomeTrakr Database
Salmonella
Listeria
E. coli / Shigella
Campylobacter
V. parahaemolyticus
Cronobacter
Number of Sequences
(as of the last day of the quarter)
350,000
300,000
250,000
200,000
150,000
100,000
50,000
0
Average Number of Sequences Added Per Month in 2013 = 169
Average Number of Sequences Added Per Month in 2014 = 1,076
Average Number of Sequences Added Per Month in 2015 = 2,362
Average Number of Sequences Added Per Month in 2016 = 4,529
Average Number of Sequences Added Per Month in 2017 = 5,826
Average Number of Sequences Added Per Month in 2018 = 9,434
First sequences uploaded in Feb 2013
Public Health England uploads more than 8,000 Salmonella sequences
1st Qtr
2nd Qtr
3rd Qtr
4th Qtr
2013
2014
2015
2016
2017
2018

May 09, 2019 GenomeTrakr Numbers

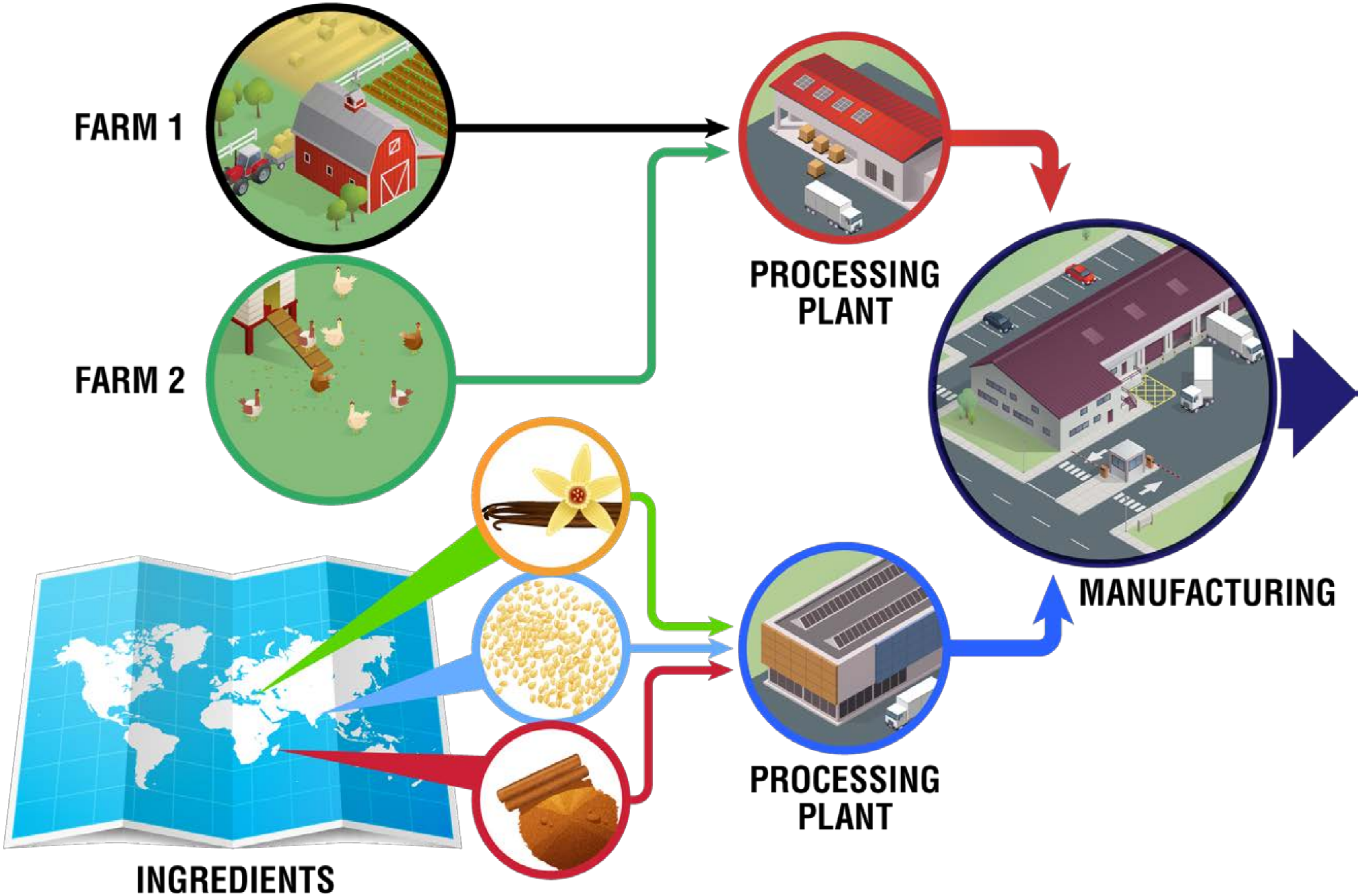
Species	Total Isolates
Salmonella enterica	196,325
E.coli and Shigella	71,760
Campylobacter jejuni	29,770
Listeria monocytogenes	26,923
Vibrio parahaemolyticus	1,859
Cronobacter	605
Total	**327,242**

FDA U.S. FOOD & DRUG ADMINISTRATION

How do we use the GenomeTrakr information? Example of Listeria in sprouts using a phylogenetic perspective.



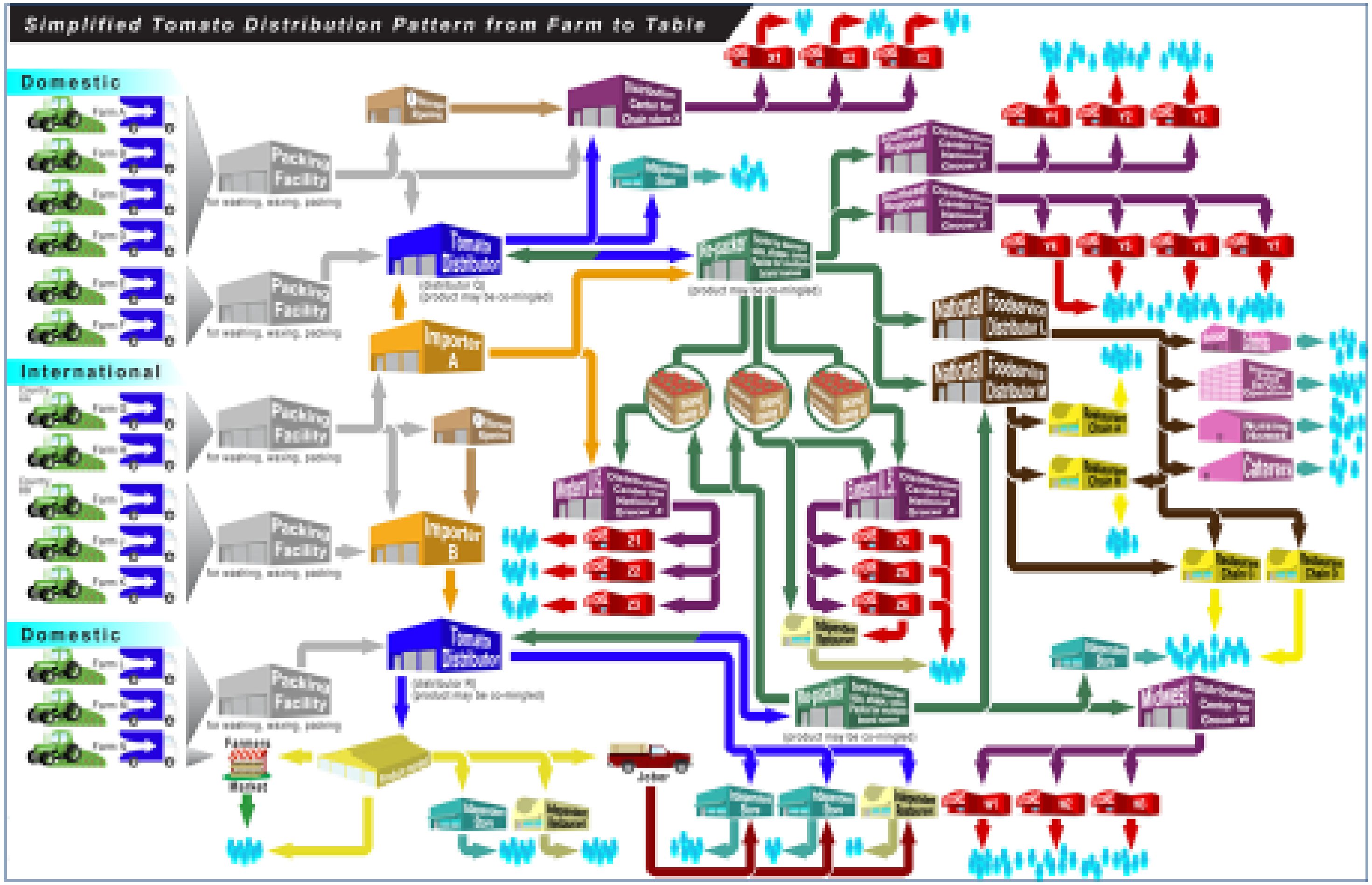
U.S. FOOD & DRUG ADMINISTRATION



Further perspectives on the food supply

U.S. FOOD & DRUG ADMINISTRATION
CENTER FOR FOOD SAFETY & APPLIED NUTRITION

Once tomatoes reach the supply chain, things really “*simplify*”.

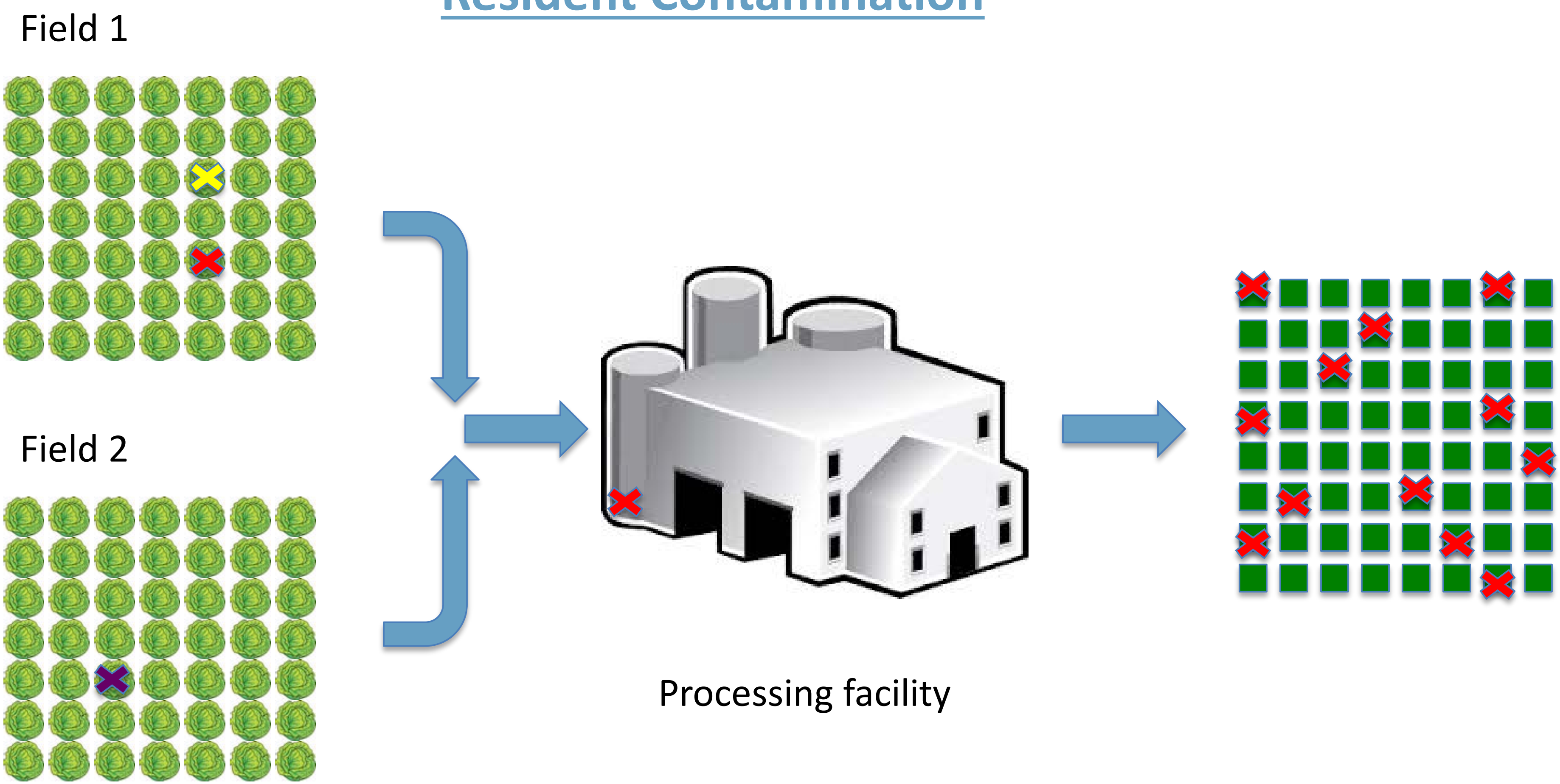


The Fresh-cut Tomato Supply Chain

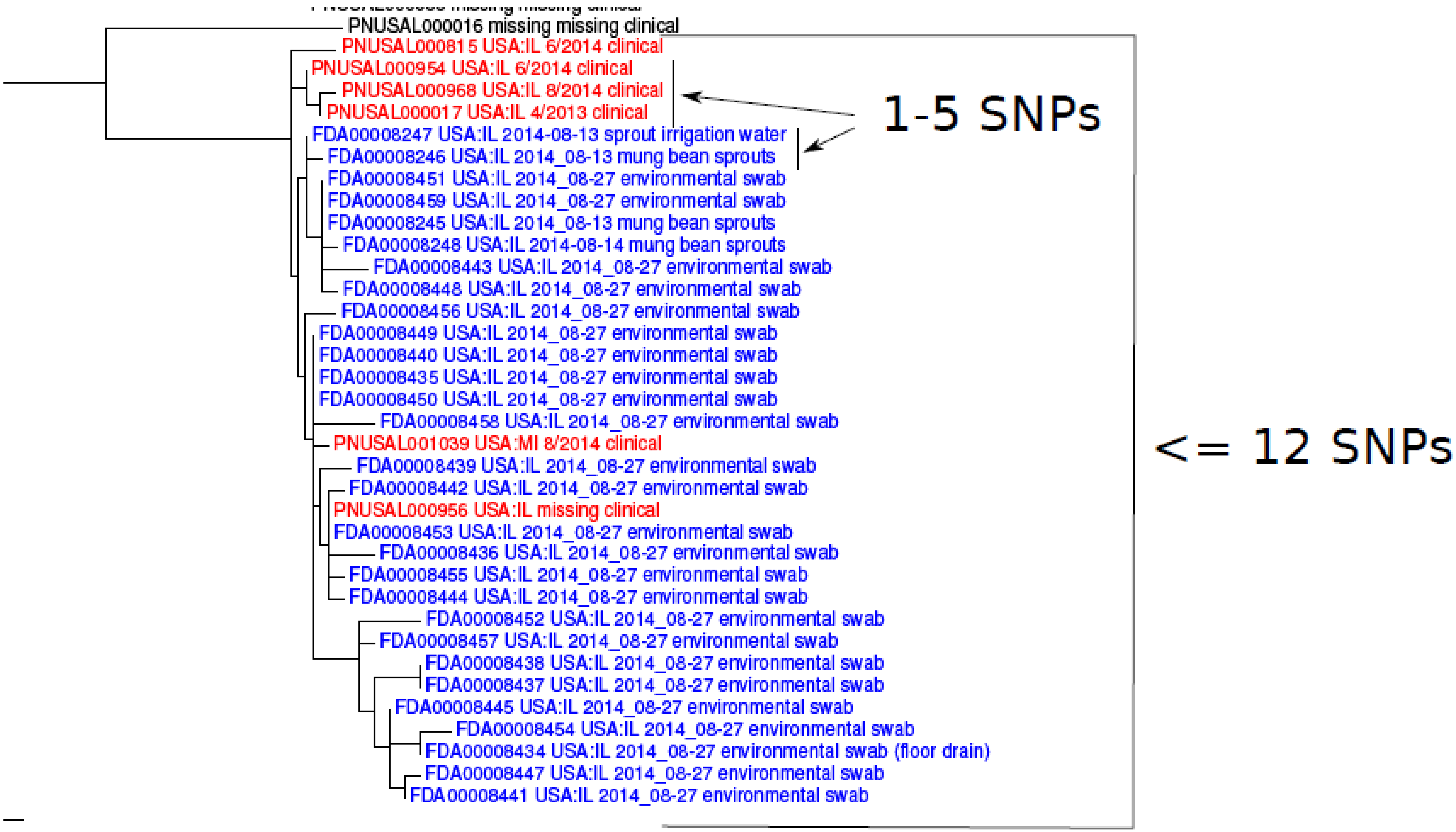
Office of Regulatory Science

FDA U.S. FOOD & DRUG ADMINISTRATION

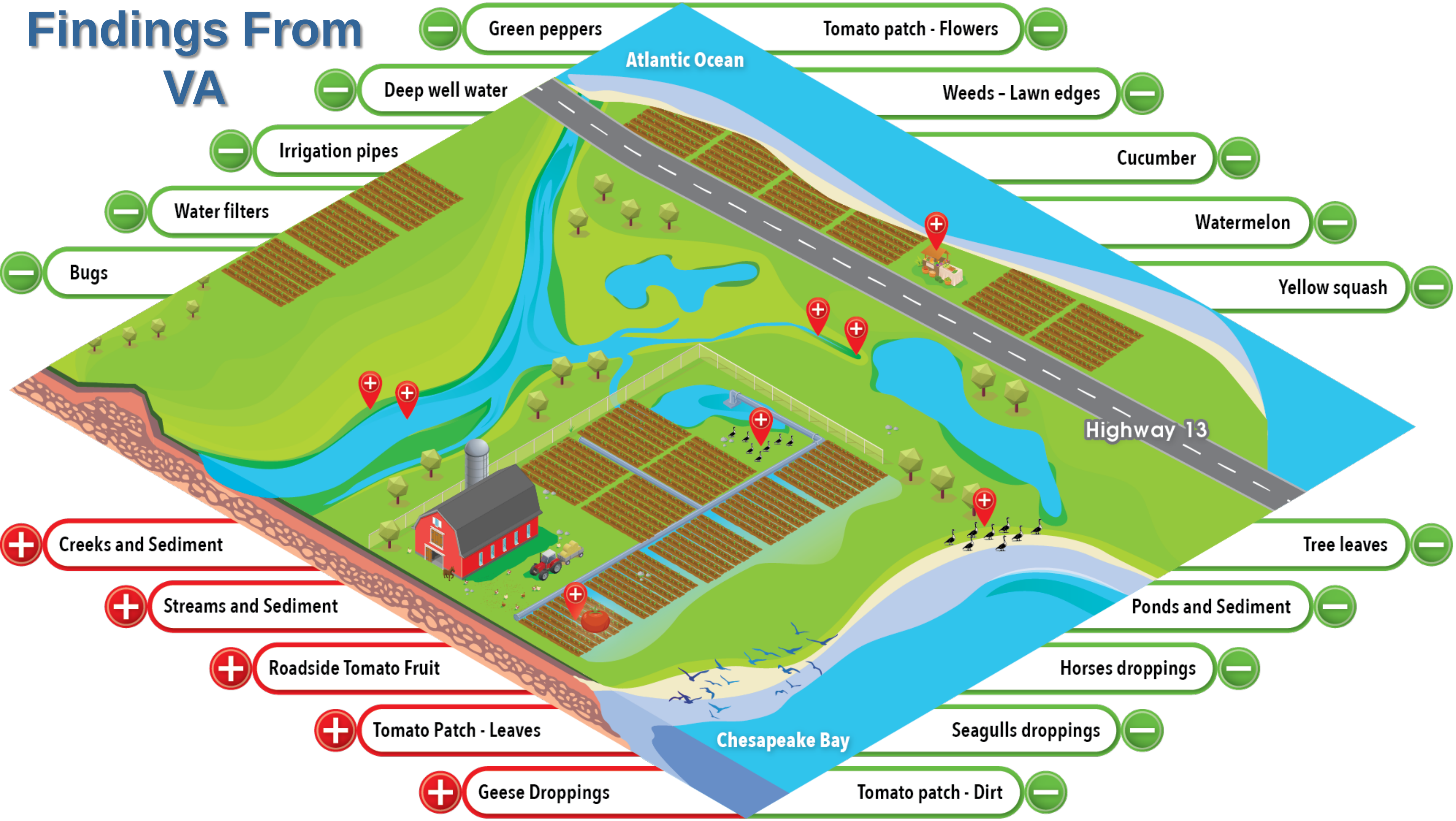
Scenario 2 and statistics: Resident Contamination



Post inspection determines root cause of contamination event.



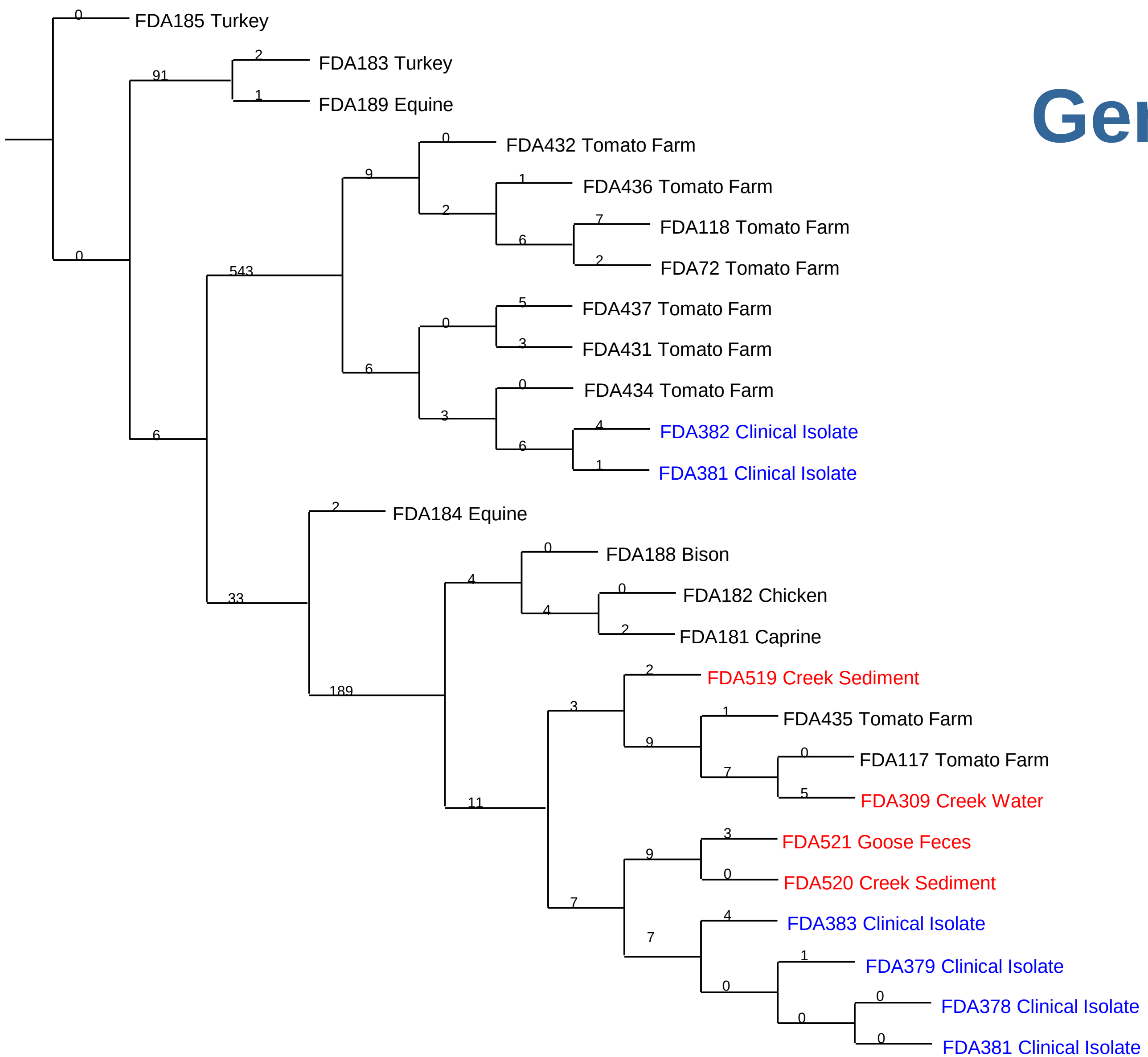
Findings From VA



Collected over 500 isolates:

2010-2011: **237** 2012: **137** 2014: **133+**

Whole Genome Tree



Key Conclusions from VA surveys Identify Water, Water, and Water

- Sampled Areas:
  - Virginia Tech Agricultural Research and Education Center (AREC): 2009-2011
  - 6 environmental waters and sediments: 2011, 2012, 2014-2015
- *Salmonella* Pattern 61 and other clinically relevant isolates found consistently in waters and sediments at AREC and the 6 other locations
- Environmental waters and sediments are potential reservoirs for *Salmonella*
- Enhanced agricultural practices related to the use of surface waters are important to mitigating *Salmonella* transfer on to crop plants


Ecological prevalence, genetic diversity, and epidemiological aspects of Salmonella isolated from tomato agricultural regions of the Virginia Eastern Shore.
Bell RL et al
Front Microbiol. 2015 May 7;6:415.

Current Status

- WGS is now routine in FDA's outbreak response and compliance/surveillance. Internally (across our agency), and in collaboration with FSIS and CDC, WGS has now been deployed and benefitted the traceability of numerous foodborne contamination events. **Weekly Regulatory Guidance from SME.**
- Numerous offshoot applications exist (i.e., supply chain management, quality assurance, process evaluation, etc.).
- Genome sequences are **portable, instantly cross-compatible and highly scalable**. One technology approach irrelevant of organism.
- Have to balance the need for increased number of well characterized **environmental** (food, water, facility, etc.) sequences with the need for extensive clinical isolates
- WGS, unlike PFGE, is more than a surveillance tool. It provides information on **AMR, Virulence, serotype**, and other critical factors in one assay, including historical reference to pathogen emergence.
- The volume of WGS and associated work continues to grow. The CDC currently sequences all ~800 **Listeria** clinical samples they receive annually, and we expect them to start sequencing all **STECs** and **Salmonella** clinical samples by 2017.
- As internal and external sequence data is integrated and analyzed in concert, we expect many more Illnesses that were previously attributed to **sporadic infections** to now be linked to specific food or environmental sources. This will result in the detection of a **large number of small outbreaks.**

U.S. FOOD & DRUG ADMINISTRATION

www.ncbi.nlm.nih.gov/pathogens/

Home - Pathogen Detect

https://www.ncbi.nlm.nih.gov/pathogens/

Pathogen Detection BETA

NCBI Pathogen Detection integrates bacterial pathogen genomic sequences originating in food, environmental sources, and patients. It quickly clusters and identifies related sequences to uncover potential food contamination sources, helping public health scientists investigate foodborne disease outbreaks.

Find isolates now!

Explore the Data

Species	New Isolates	Total Isolates
Salmonella enterica	124	53,003
E.coli and Shigella	1	19,861
Listeria monocytogenes	20	12,267
Campylobacter jejuni	0	4,309
Acinetobacter baumannii	0	2,651
Klebsiella pneumoniae	1	2,212
Enterobacter	0	1,151
Vibrio parahaemolyticus	0	817

Learn More

About

FAQ

Antimicrobial Resistance

Contributors

Data Resources

Isolates Browser

Antimicrobial resistance reference gene database

Isolates with antibiotic resistant phenotypes

Beta-lactamase resources

Download analysis results (FTP)

Submit

How to submit data

How to submit antibiotic resistance phenotypes

How to submit beta-lactamases

NCBI Submission Portal

New Isolate Check - Listeria

National Library of Medicine | NCBI National Center for Biotechnology Information

Health > Pathogen Detection

Pathogen Detection BETA

NCBI Pathogen Detection integrates bacterial pathogen genomic sequences originating in food, environmental sources, and patients. It quickly clusters and identifies related sequences to uncover potential food contamination sources, helping public health scientists investigate foodborne disease outbreaks.

Find isolates now!

Explore the Data

Species	New Isolates	Total Isolates
Salmonella enterica	368	63,000
E.coli and Shigella	1	26,603
Listeria monocytogenes	3	13,343
Campylobacter jejuni	11	7,857

Learn More

About

FAQ

Antimicrobial Resistance

Contributors

Data Resources

Isolates Browser

Antimicrobial resistance reference gene database

Isolates with antibiotic resistant phenotypes

Beta-lactamase resources

Download analysis results (FTP)

New Isolate Check - Listeria

Health > Pathogen Detection > Isolates Browser

taxgroup_name:"Listeria monocytogenes" AND new:1 | Search

Listeria monocytogenes PDS000000277.50 (1)
Listeria monocytogenes PDS000003244.7 (1)
Show all 3 clusters

Listeria monocytogenes

Filters

Page 1 of 1 | 20

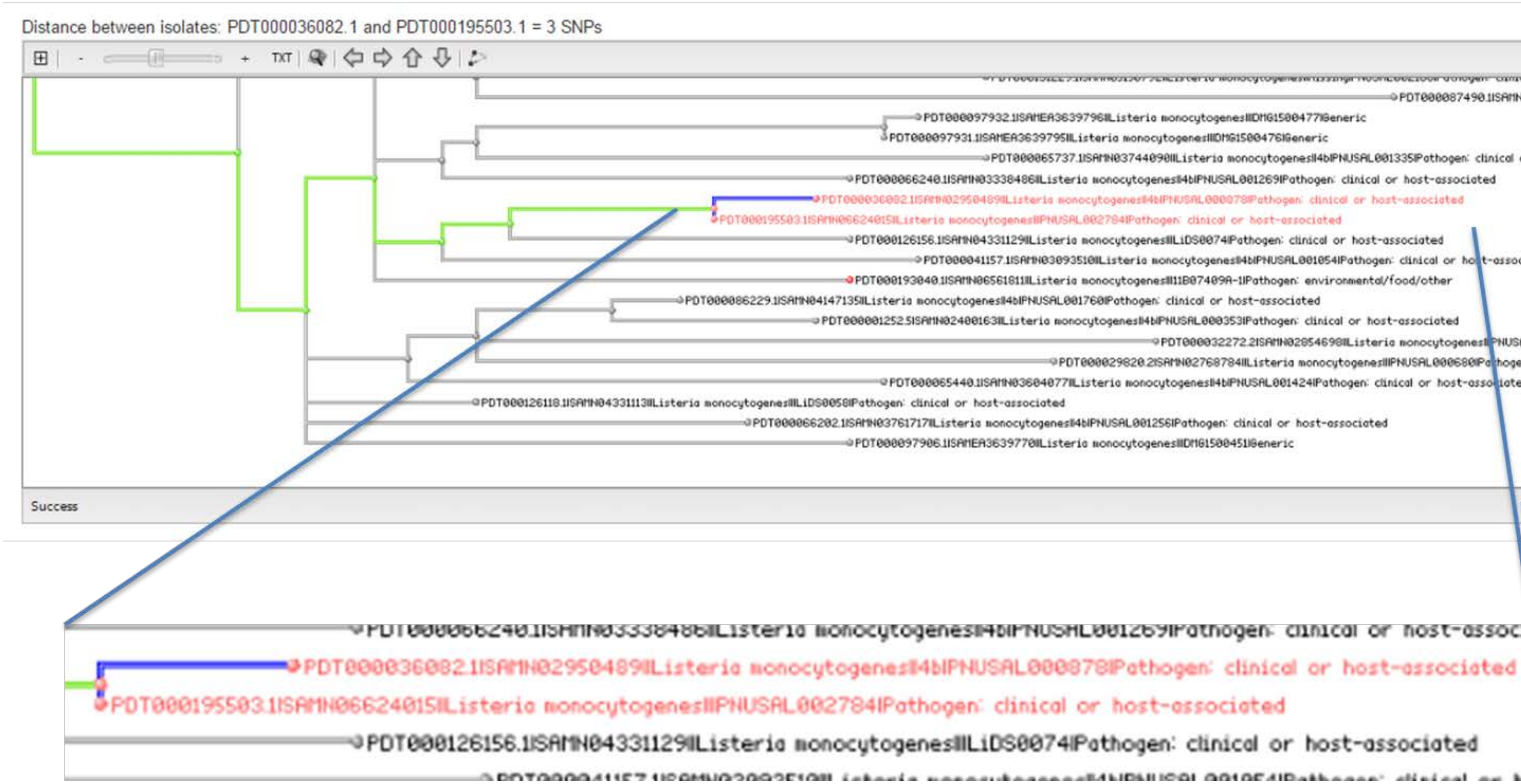
#	Organism Group	Strain	Serovar	Isolate	Create Date	Location	Isolation Sou	Isolation type	Host	SNP cluster	Min-sar	Min-diff	BioSample	Assembly	K-mer group
1	Listeria monocytogenes	PNUSAL00		PDT000195503.1	2017-03-21	USA		clinical		PDS000003244.7	3	24	SAMN06624015		PDG000000001.646
2	Listeria monocytogenes	PNUSAL00		PDT000195504.1	2017-03-21	USA		clinical		PDS000011887.1	n/a	21	SAMN06624016		PDG000000001.646
3	Listeria monocytogenes	ICDC-LM188		PDT000195835.1	2017-03-20	China: Beijing	blood	clinical	Homo sapiens	PDS000000277.50	27	22	SAMN04884549	GCA_002043045.1	PDG000000001.646

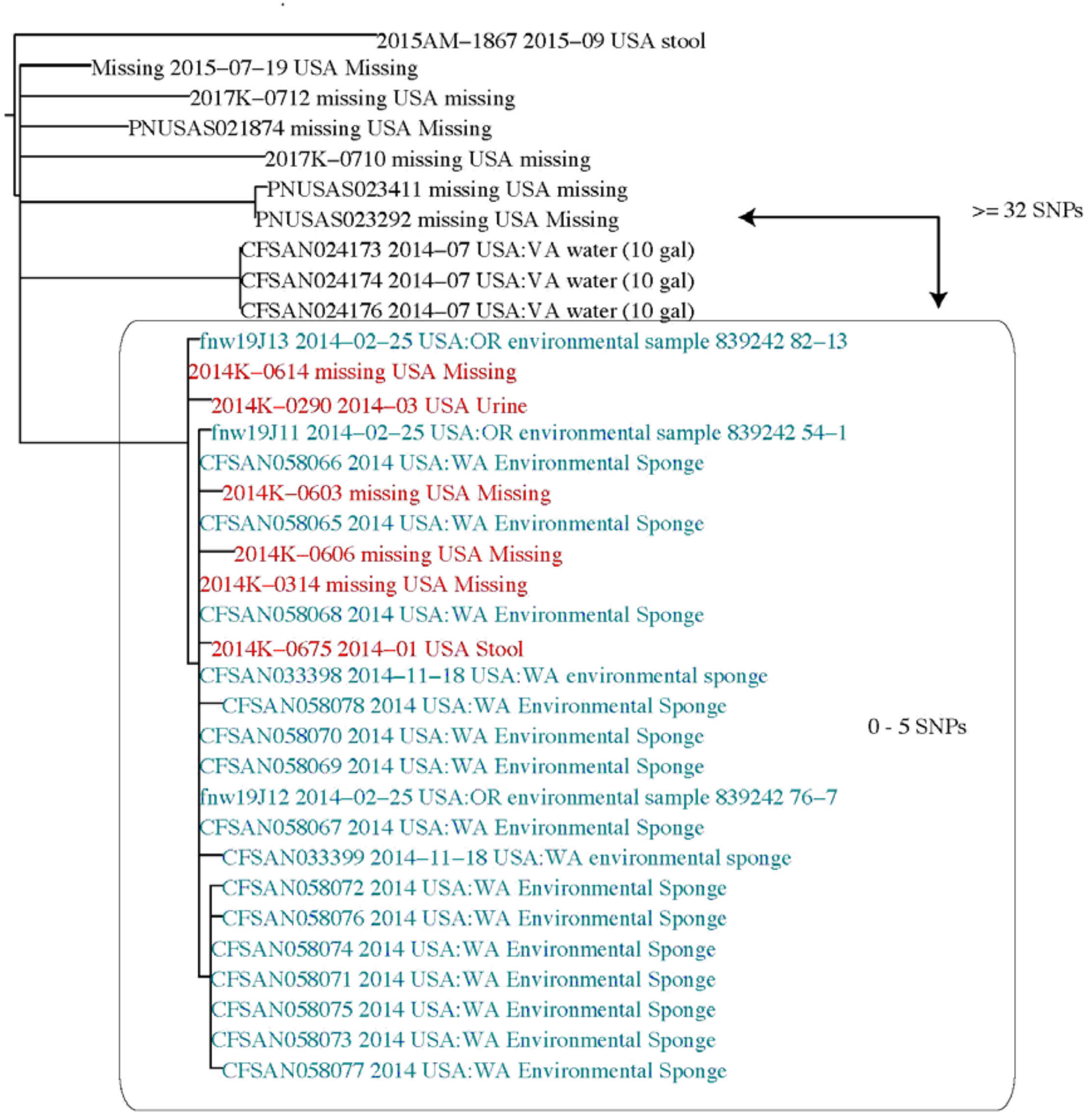
3 New Clinical Isolates
- Inconclusive food/env matches, 21-24 SNPs
- Clinical match, 3 SNPs

of 1 | 20

SNP cluster	Min-sar	Min-dif	Bi
PDS000003244.7	3	24	SAM
PDS000011887.1	n/a	21	SAM
PDS000000277.50	27	22	SAM

PDS000003244.7 Tree View



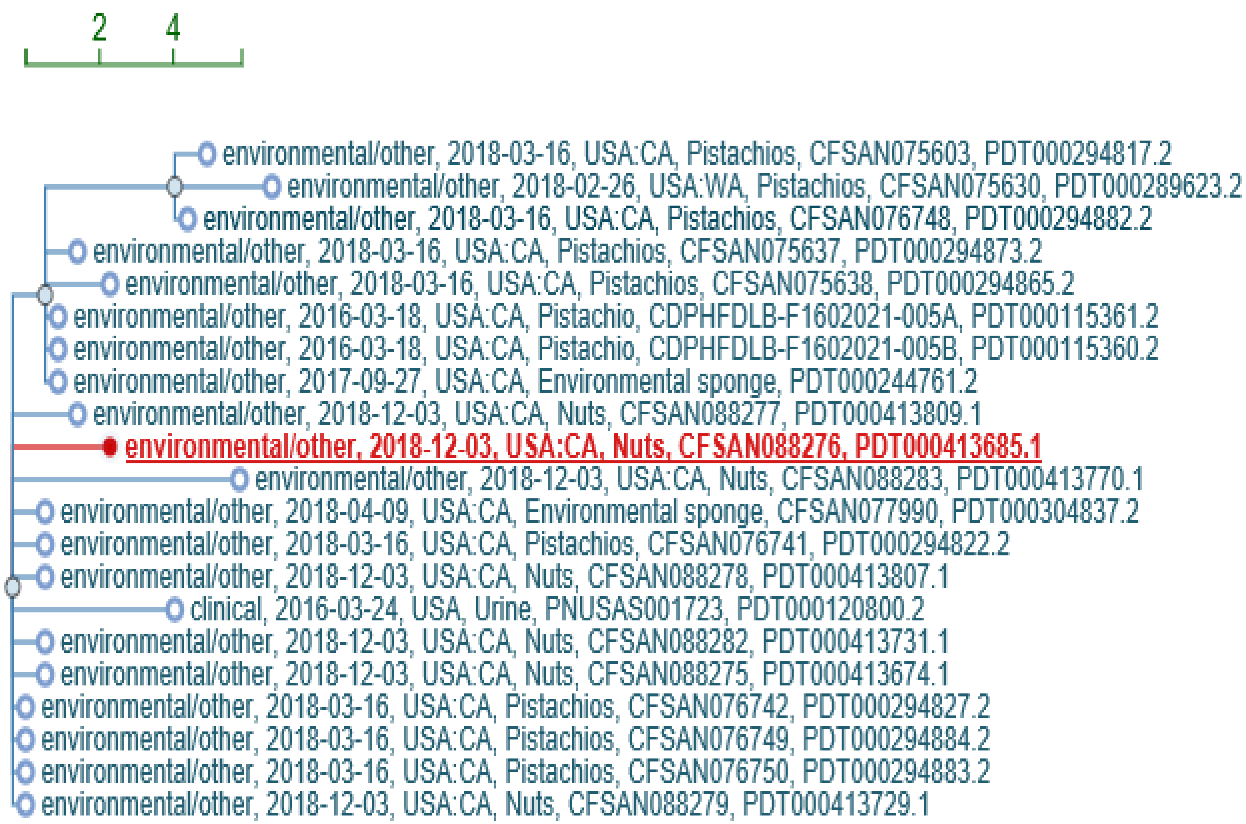

2015AM–1867 2015–09 USA stool
Missing 2015–07–19 USA Missing
2017K–0712 missing USA missing
PNUSAS021874 missing USA Missing
2017K–0710 missing USA missing
PNUSAS023411 missing USA missing
PNUSAS023292 missing USA Missing
CFSAN024173 2014–07 USA:VA water (10 gal)
CFSAN024174 2014–07 USA:VA water (10 gal)
CFSAN024176 2014–07 USA:VA water (10 gal)
>= 32 SNPs
fnw19J13 2014–02–25 USA:OR environmental sample 839242 82–13
2014K–0614 missing USA Missing
2014K–0290 2014–03 USA Urine
fnw19J11 2014–02–25 USA:OR environmental sample 839242 54–1
CFSAN058066 2014 USA:WA Environmental Sponge
2014K–0603 missing USA Missing
CFSAN058065 2014 USA:WA Environmental Sponge
2014K–0606 missing USA Missing
2014K–0314 missing USA Missing
CFSAN058068 2014 USA:WA Environmental Sponge
2014K–0675 2014–01 USA Stool
CFSAN033398 2014–11–18 USA:WA environmental sponge
CFSAN058078 2014 USA:WA Environmental Sponge
CFSAN058070 2014 USA:WA Environmental Sponge
CFSAN058069 2014 USA:WA Environmental Sponge
fnw19J12 2014–02–25 USA:OR environmental sample 839242 76–7
CFSAN058067 2014 USA:WA Environmental Sponge
CFSAN033399 2014–11–18 USA:WA environmental sponge
CFSAN058072 2014 USA:WA Environmental Sponge
CFSAN058076 2014 USA:WA Environmental Sponge
CFSAN058074 2014 USA:WA Environmental Sponge
CFSAN058071 2014 USA:WA Environmental Sponge
CFSAN058075 2014 USA:WA Environmental Sponge
CFSAN058073 2014 USA:WA Environmental Sponge
CFSAN058077 2014 USA:WA Environmental Sponge
0 - 5 SNPs

Phylogenetic results of Salmonella enterica clustering pine nuts and clinical isolates, NCBI Pathogen detection cluster PDS000032222.4. The number line corresponds to SNPs, where the branch length is proportional to the number of SNPs present.

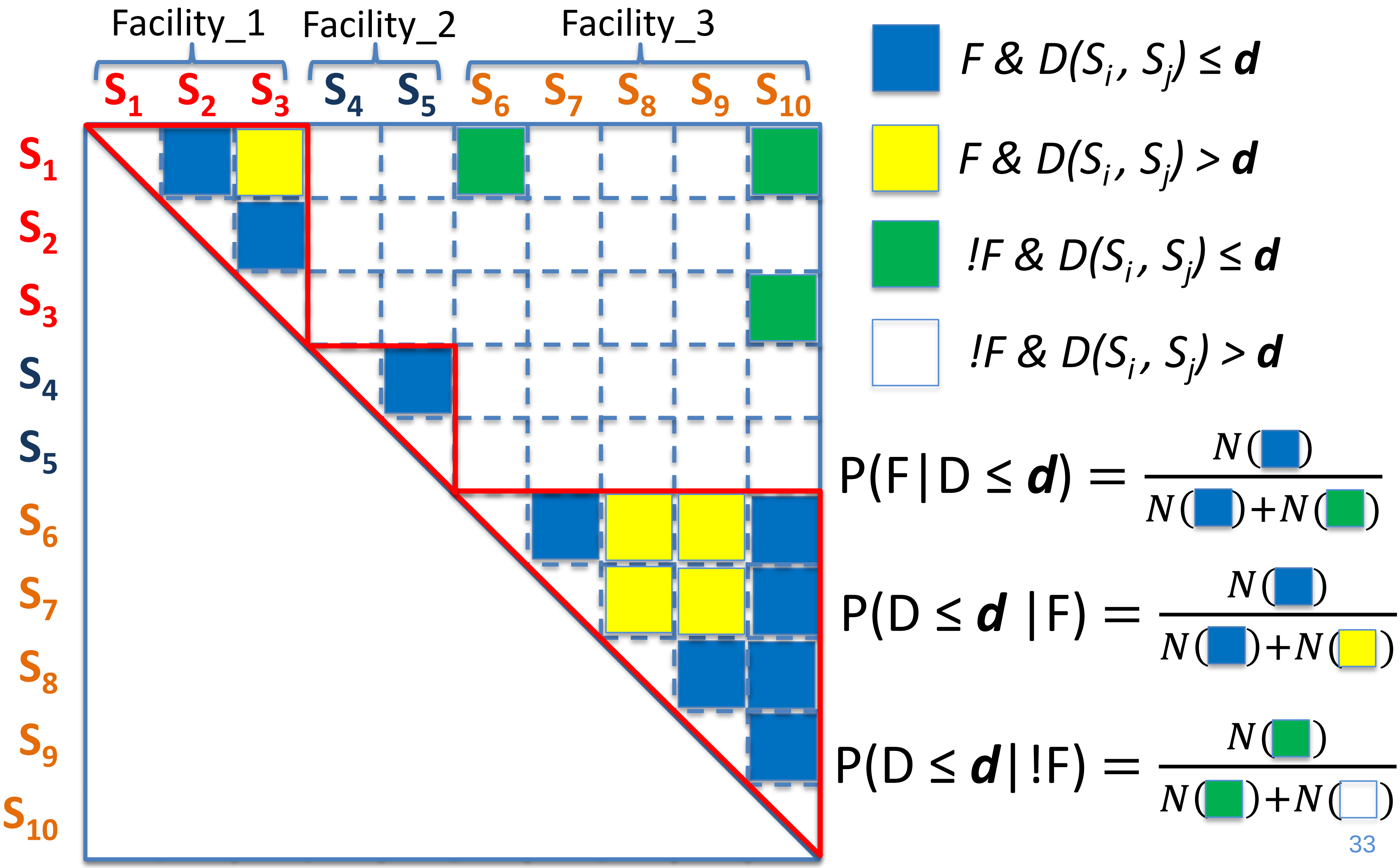
Phylogenetic results of Salmonella enterica clustering hazelnuts and clinical isolates, NCBI Pathogen detection cluster PDS000027740.2. The number line corresponds to SNPs, where the branch length is proportional to the number of SNPs present.

Phylogenetic results of Salmonella enterica clustering nuts and clinical isolates, NCBI Pathogen detection cluster PDS000005469.11. The number line corresponds to SNPs, where the branch length is proportional to the number of SNPs present.

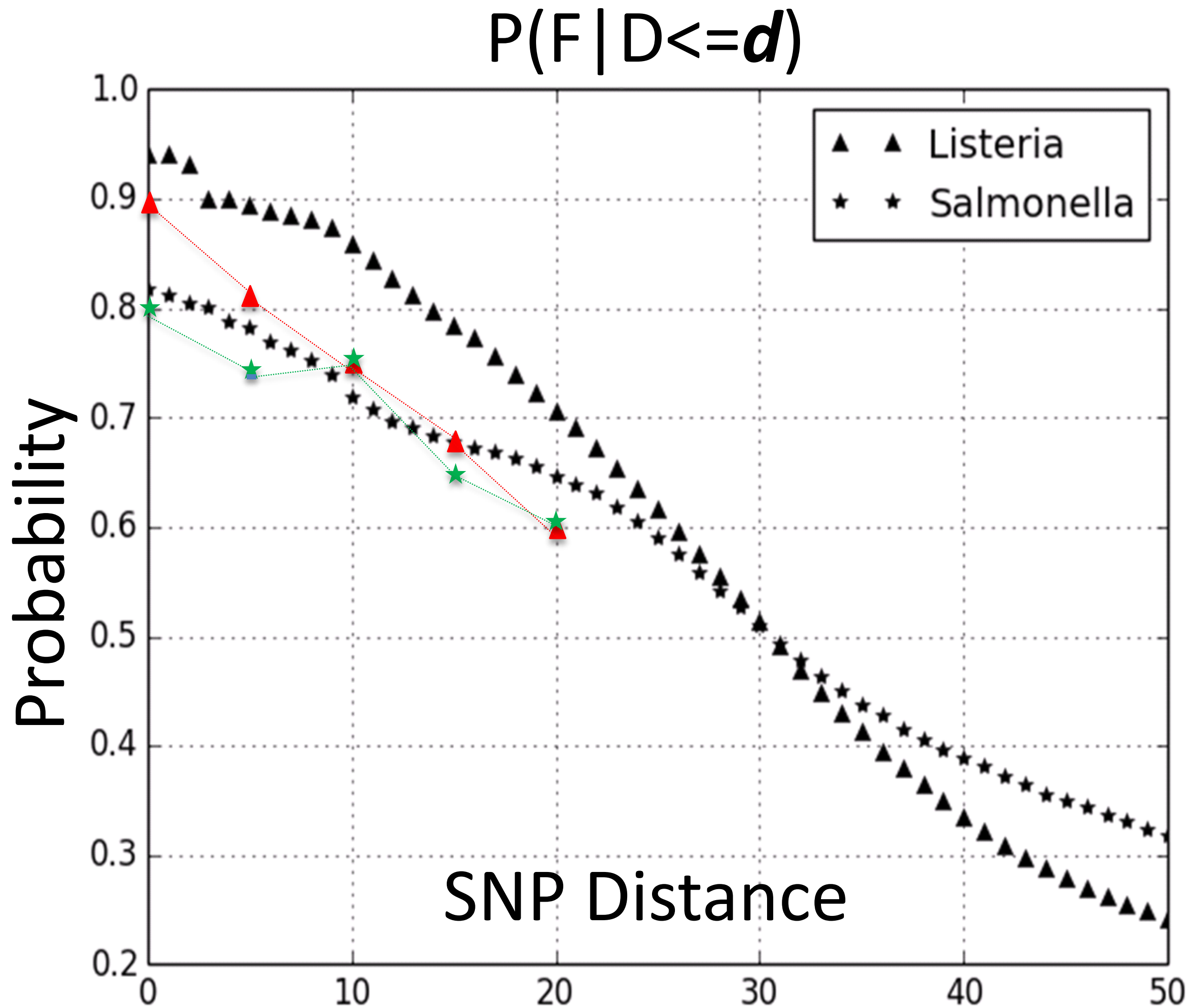
Phylogenetic results of Salmonella enterica clustering Pistachio and clinical isolates, NCBI Pathogen detection cluster PDS000027237.52. The number line corresponds to SNPs, where the branch length is proportional to the number of SNPs present.



Combining Genetic Distances and Facility Information

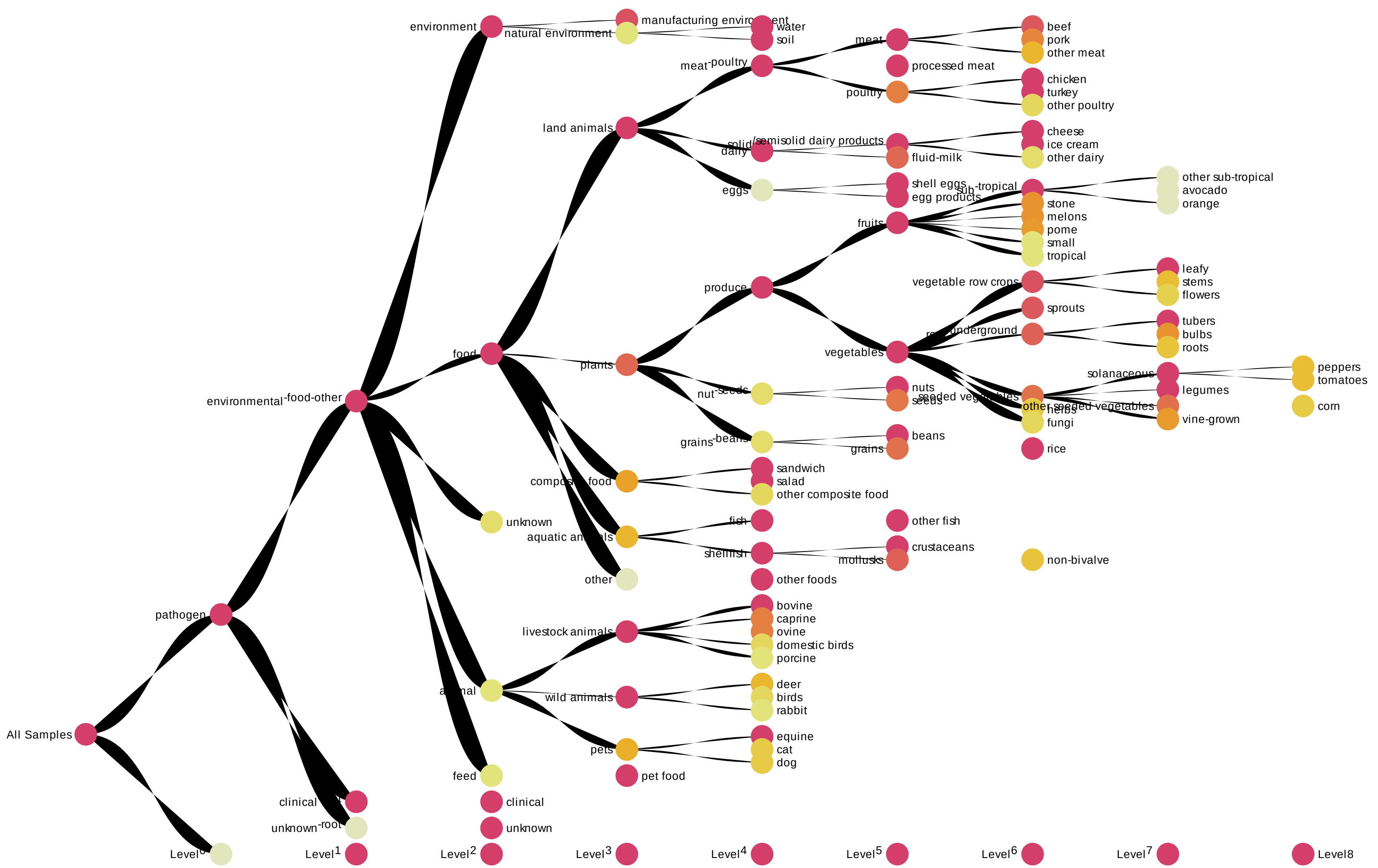


Facility Match Probability



SNP cutoff	*Sal.*	*Lis.*
0	0.82	0.94
5	0.78	0.89
10	0.72	0.86
15	0.68	0.79
20	0.65	0.70

Yu Wang, James B. Pettengill, Arthur Pightling, Ruth Timme, Marc Allard, Errol Strain, and Hugh Rand (2018) Genetic Diversity of Salmonella and Listeria Isolates from Food Facilities. Journal of Food Protection: December 2018, Vol. 81, No. 12, pp. 2082-2089.



GenomeGraphR: WGS data integration, analysis, and visualization for risk assessment and management: https://fda-riskmodels.foodrisk.org/genomegraphr/

Moez Sanaa, Régis Pouillot, Francisco J Garces-Vega, Errol Strain, Jane M Van Doren doi: https://doi.org/10.1101/495309 2018.

U.S. FOOD & DRUG ADMINISTRATION

Center for Veterinary Medicine — Protecting Human and Animal Health

NARMS — National Antimicrobial Resistance Monitoring System

Food and Drug Administration • Centers for Disease Control and Prevention • United States Department of Agriculture

FDA CDC USDA

From WGS to Antibiotic Resistance Genotype

DNA from Single colony

Sequencing With Illumina Miseq

CLC bio A QIAGEN Company

Assembly CLC Genomics Workbench

BLAST ®

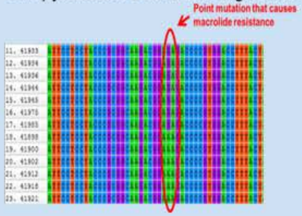
Local Blast AR Gene Database

aac(3)-IIa, aadA1, aph(3')-Ia catA1, tetO …

Acquired AR genes

Sequences alignment *gyrA* gene 23S rRNA gene

Campylobacter 23S rDNA alignment



Point mutations related to AR

AR Genotype

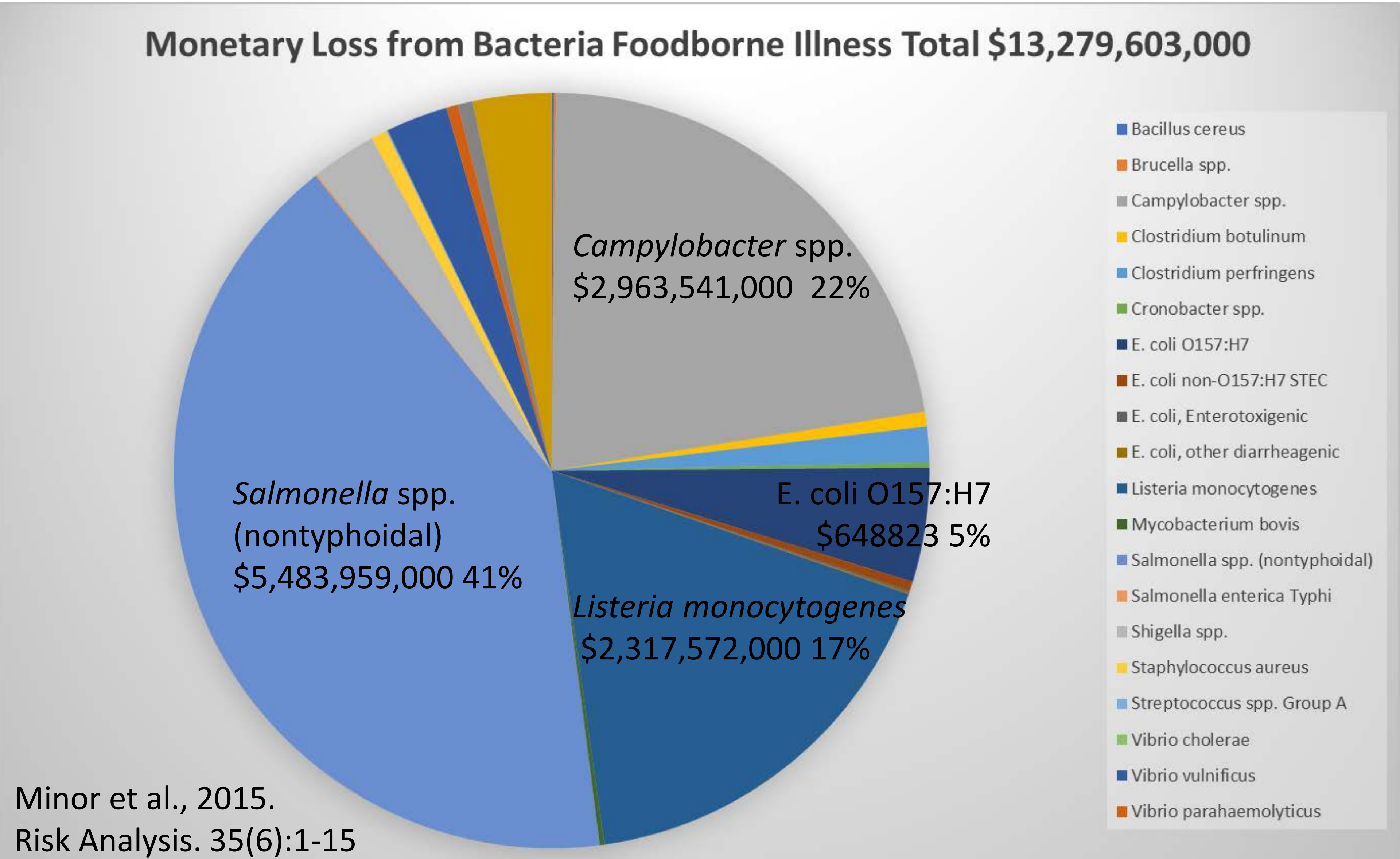

RESISTOME TRACKER
Salmonella
Select an icon or alert below to get started.
CUSTOMIZE
COMPARE
DISCOVER
EXPLORE
ALERTS
There are 11 records with 2 flagged genes uploaded in the last 30 days

Economic Analysis of *Salmonella* outbreaks reduction using WGS

	Canada	United States
Incidence of illness	47,028	1,200,000
Costs to adopt WGS	$158,340,000	$100,000,000
QUALY lost	469.75	16,782
Total Illness costs	$287,770,000	$3,300,000,000
Total net benefit of adopting WGS	$90,250,000	$1,000,000,000

Model assumes 70% reduction in numbers of illnesses due to WGS implementation.
Benefits gained due to earlier detection and decreased time to recall food items.
United States estimates are adjusted based on increase population size.
Additional analysis is needed to adjust to US illnesses and US health care costs.

Jain et al., 2019. Food Res. Int. 116: 802-809



Minor et al., 2015.
Risk Analysis. 35(6):1-15

U.S. FOOD & DRUG ADMINISTRATION

Acknowledgements

- **FDA**
  - Center for Food Safety and Applied Nutrition
  - Center for Veterinary Medicine
  - Office of Regulatory Affairs
- **National Institutes of Health**
  - National Center for Biotechnology Information
- **State Health and University Labs**
  - Alaska
  - Arizona
  - California
  - Florida
  - Hawaii
  - Maryland
  - Minnesota
  - New Mexico
  - New York
  - South Dakota
  - Texas
  - Virginia
  - Washington
- **USDA/FSIS and ARS**
- **CDC**
  - Enteric Diseases Laboratory
- **INEI-ANLIS "Carolos Malbran Institute," Argentina**
- **Centre for Food Safety, University College Dublin, Ireland and Irish FSA**
- **Melbourne (FSA). Australia**
- **Public Health England, UK**
- **Institute for Food Safety and Health (IFSH)**
- **WHO and FAO**
- **Illumina**
- **Pac Bio**
- **Other independent collaborators**

FDA U.S. FOOD & DRUG ADMINISTRATION