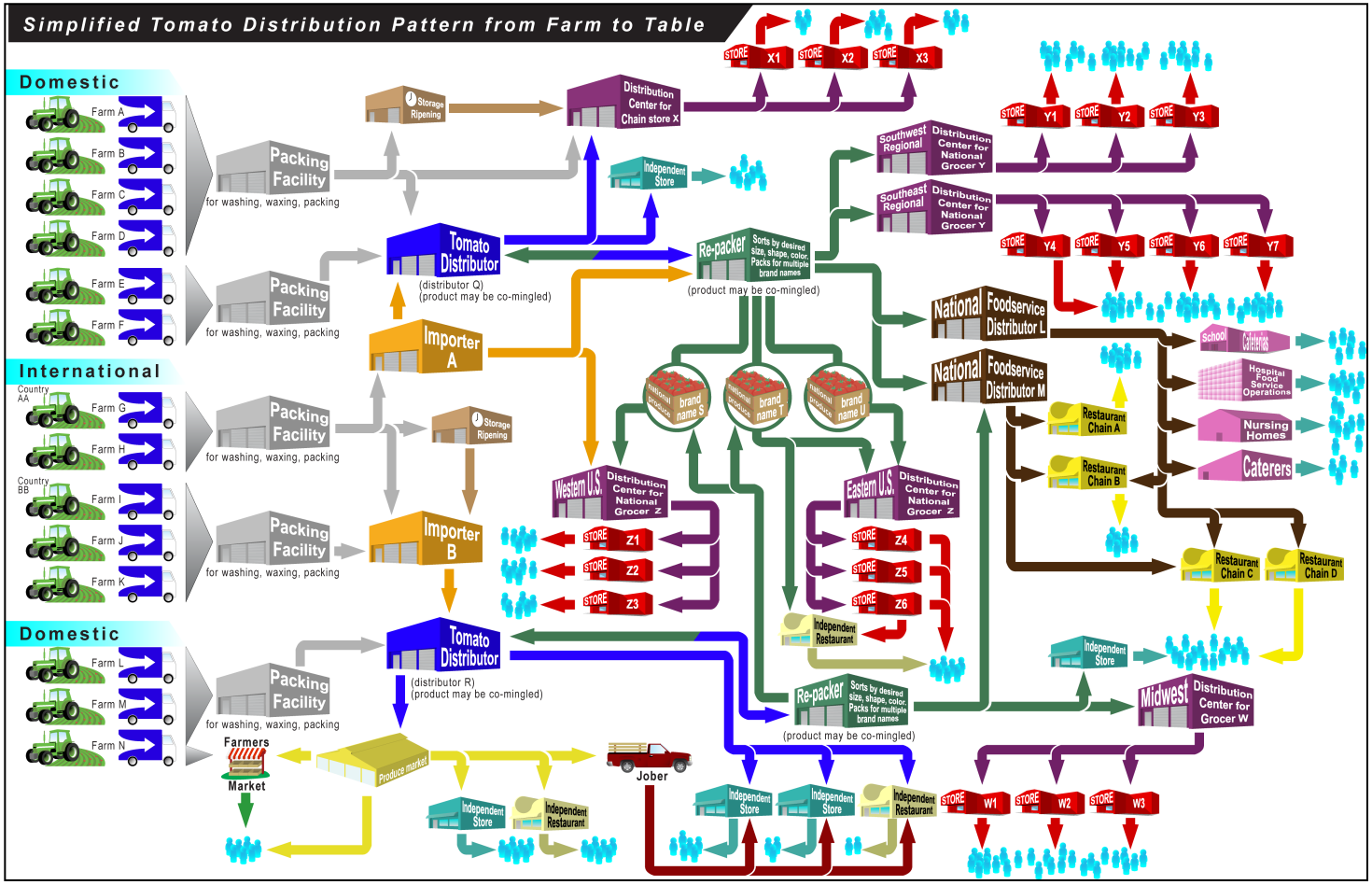
How Whole Genome Sequencing is Augmenting FDA's Role in Foodborne Outbreak Investigation

Eric L. Stevens, Ph.D.
FDA Commissioner's Fellow
U.S. Food & Drug Administration
Center for Food Safety and Applied Nutrition
Office of Regulatory Science
Division of Microbiology
College Park, MD 20740

Safe, Wholesome, Sanitary Foods

The fresh-cut tomato supply chain

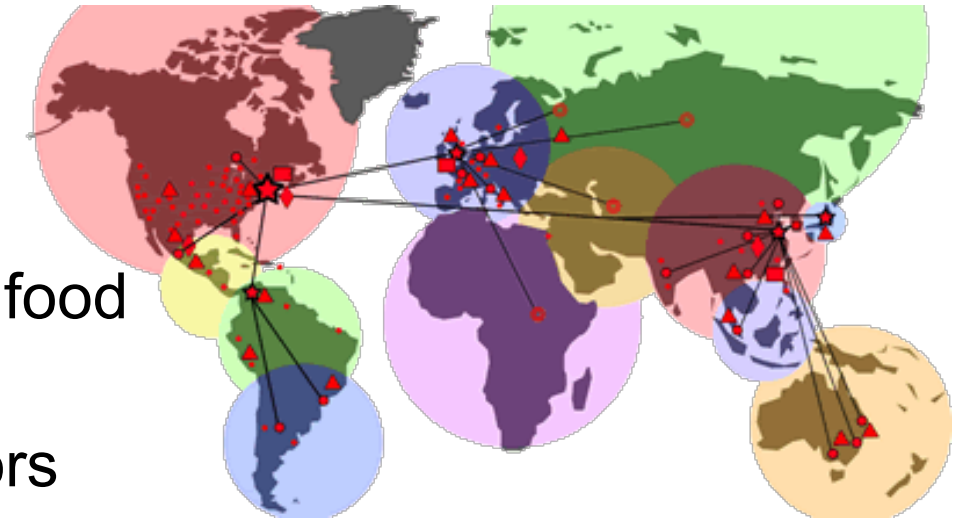


COMPLEX FOOD VEHICLES

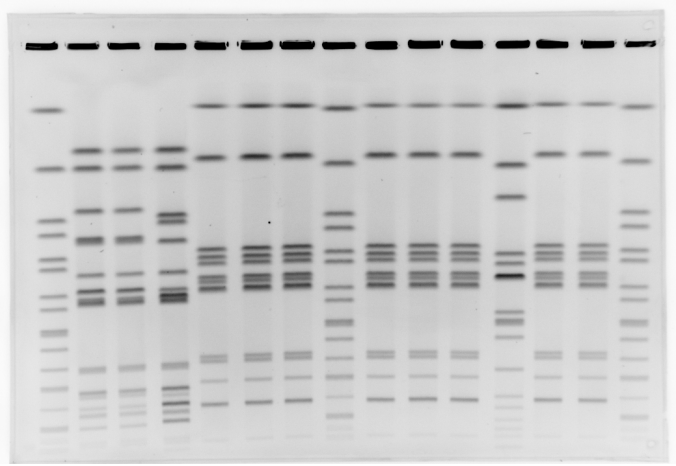
(Slide courtesy of J. Ostell, NCBI)

Why develop a WGS-based Network?

- Tracking and tracing of food pathogens
  - Global & Domestic
  - Faster identification of the food involved in the outbreak
- Limited number of investigators vs. facilities and import lines
- Insufficient resolution of current tools
  - matching clinical to environmental samples



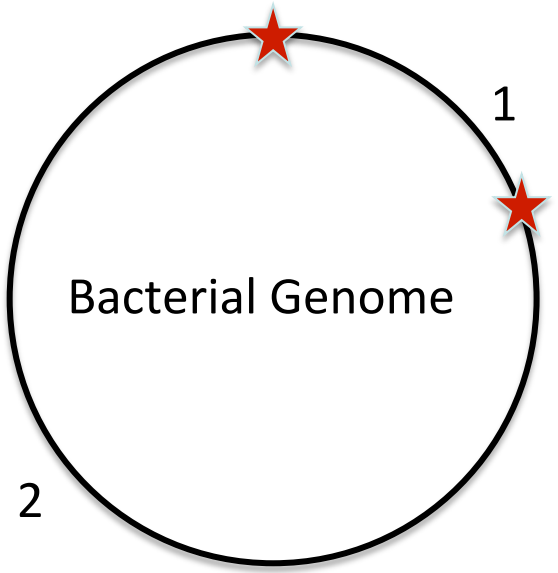
PFGE discrimination level



- Molecular technique that uses restriction enzymes to cut DNA at specific locations
- The resulting pieces are then run on a gel and separate according to size (with the smaller fragments moving further).
- Use the banding pattern of the gel to group bacterial isolates

WGS:
Whole-genome sequencing

- PFGE only gives us information at a restriction site based on the banding pattern

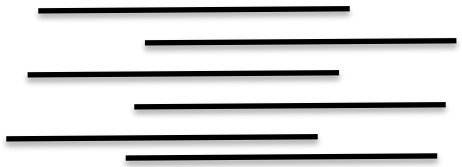


- WGS has the ability to give us information at every position in the bacterial genome

WGS: Very brief overview

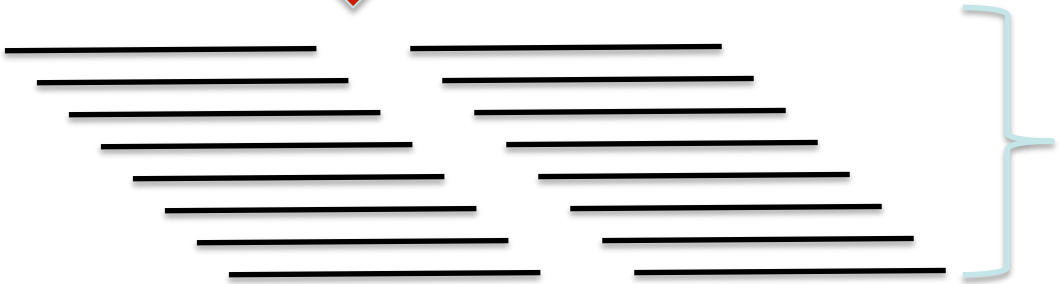
Bacterial Genome

Amplify genomic DNA and prepare for sequencing chemistry



"Reads"
(thousands to millions)
(range from 300 bp to 40 kb)

Assembly Method (many to choose from)



Aligned reads

Derived Genome Sequence

WGS:
Not all of the genomic sequence is needed

In fact, we expect many of the sites to be the same

SNP = Single Nucleotide Polymorphism

TTGTTGGCAG
ATGT**A**GGCAG
ATGTTGGCAG
ATGTTGGCA**C**

4 aligned genome sequences reduced to 3 SNPs

WGS Phylogenetic tree

Clusters similar sequences together under the assumption that it follows evolutionary history

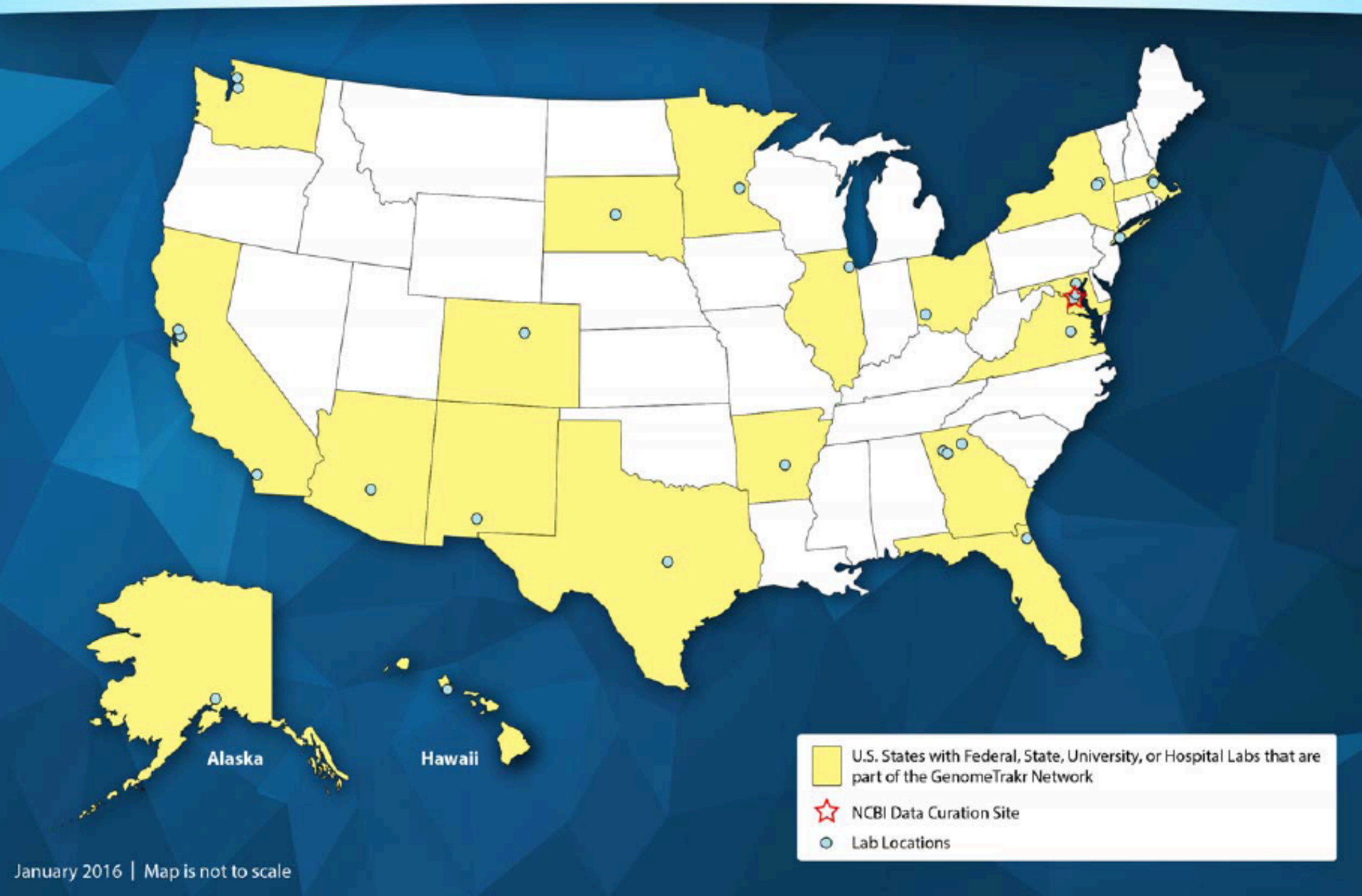
FDA's WGS program

- Most basic application is identifying pathogens from food or environmental samples and then comparing it to clinical isolates
- WGS allows better resolution than PFGE in differentiating related versus unrelated strains
  - Key for epidemiological insights
- However, one of the most promising public health benefits may come from the combination of a pathogen's genomic information with its geographic location

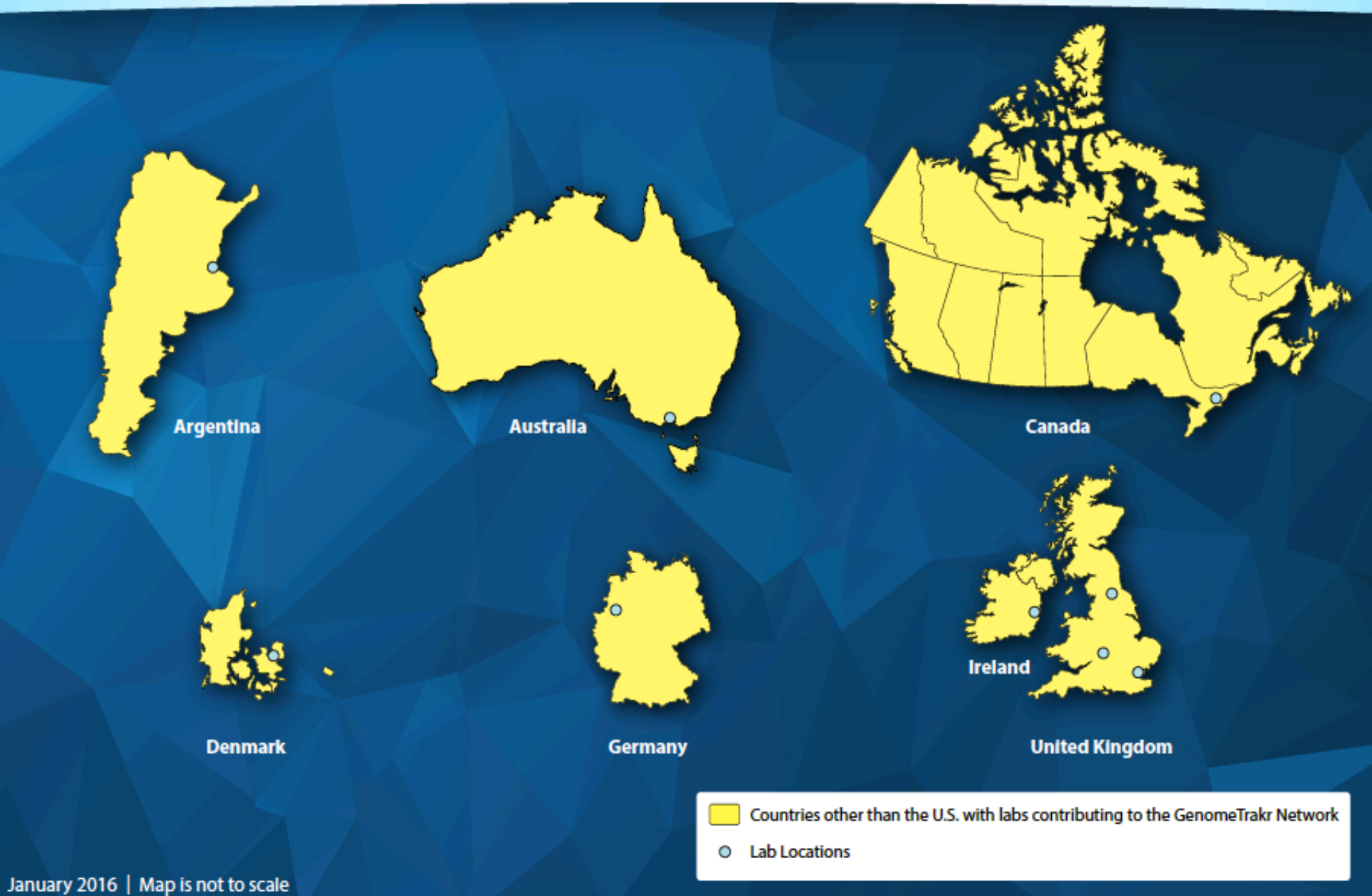
FDA's GenomeTrakr

- First distributed network of labs to use whole genome sequencing
- Consists:
  - 14 federal labs
  - 15 state health and university labs
  - 1 U.S. hospital lab
  - 9 international labs
  - Independent researchers
- Data curation and bioinformatic support/analyses provided by National Center for Biotechnology Information (NCBI)

U.S. GenomeTrakr Labs

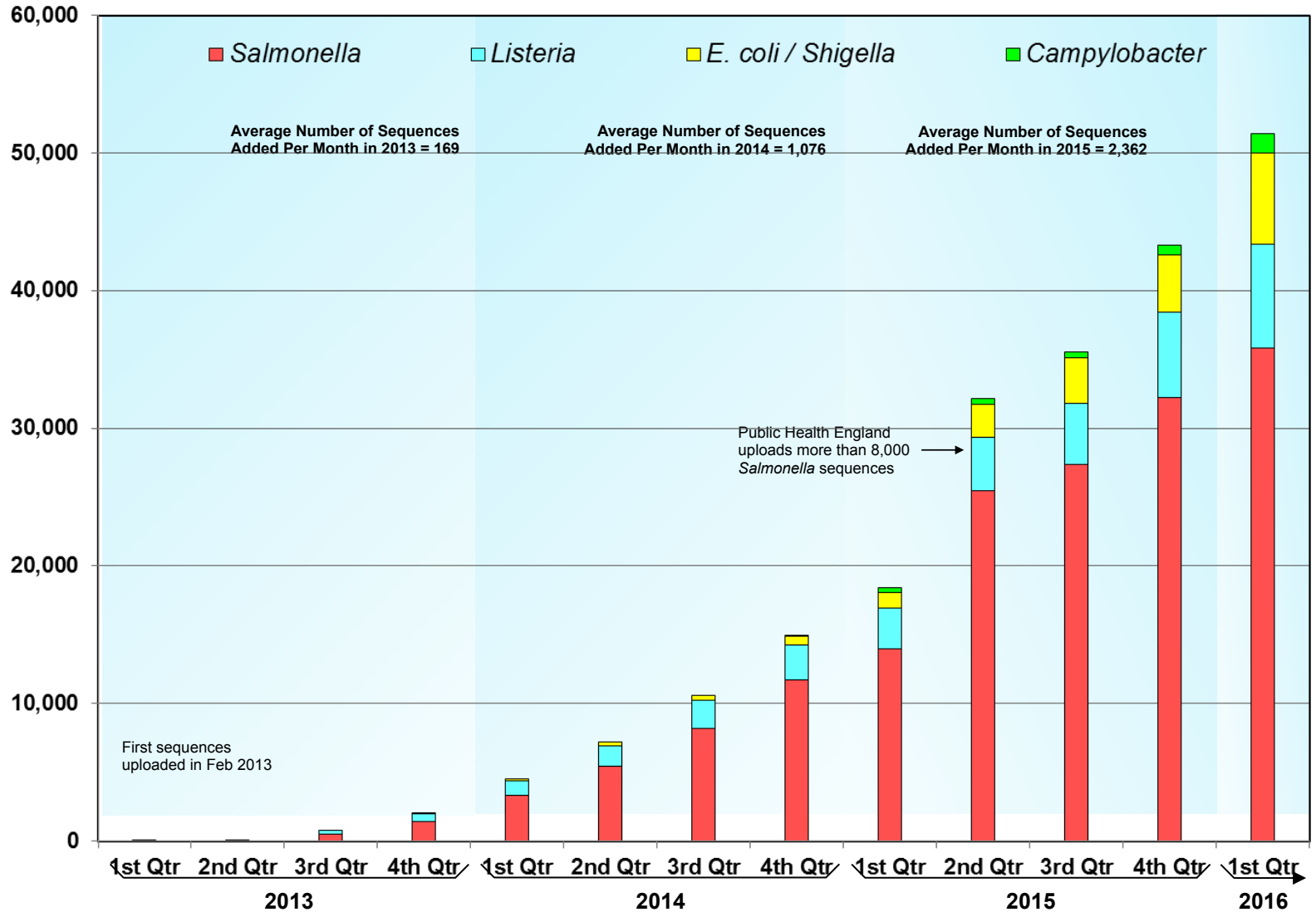


Labs Outside the U.S. Contributing to GenomeTrakr



January 2016 | Map is not to scale

Total Number of Sequences in the GenomeTrakr Database



Minimal pathogen metadata

What

sample_name
organism
strain/isolate

Category (attribute_package)
1a) **Clinical/Host-associated**
1a1) **specific_host**
1a2) **isolation_source**
1a3) **host-disease**
OR
1b) **Environmental/Food/Other**
1b1) **isolation_source**

When

collection_date

Where

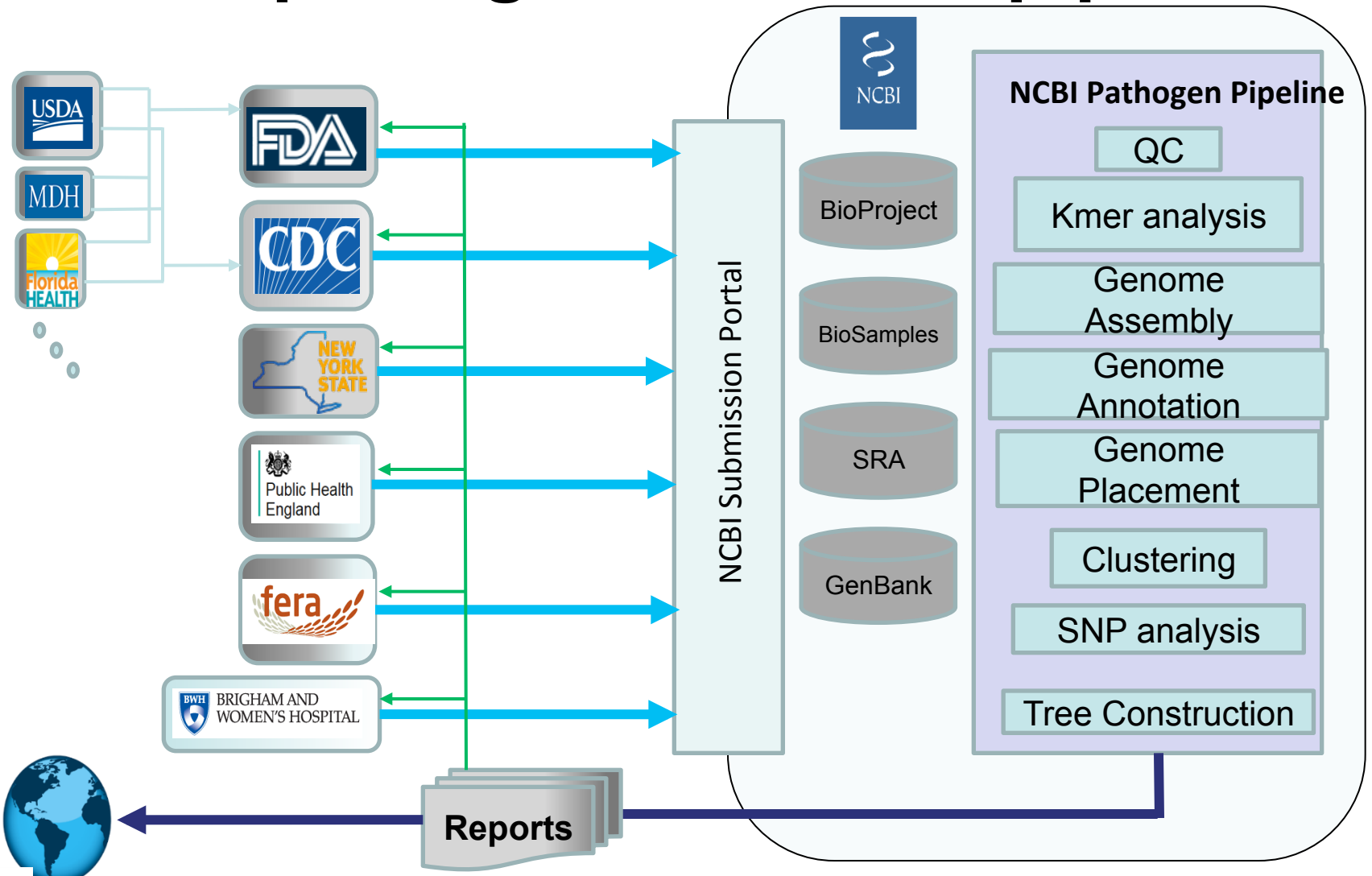
Geographic location

6a) **geo_loc_name**
OR
6b) **lat_lon**

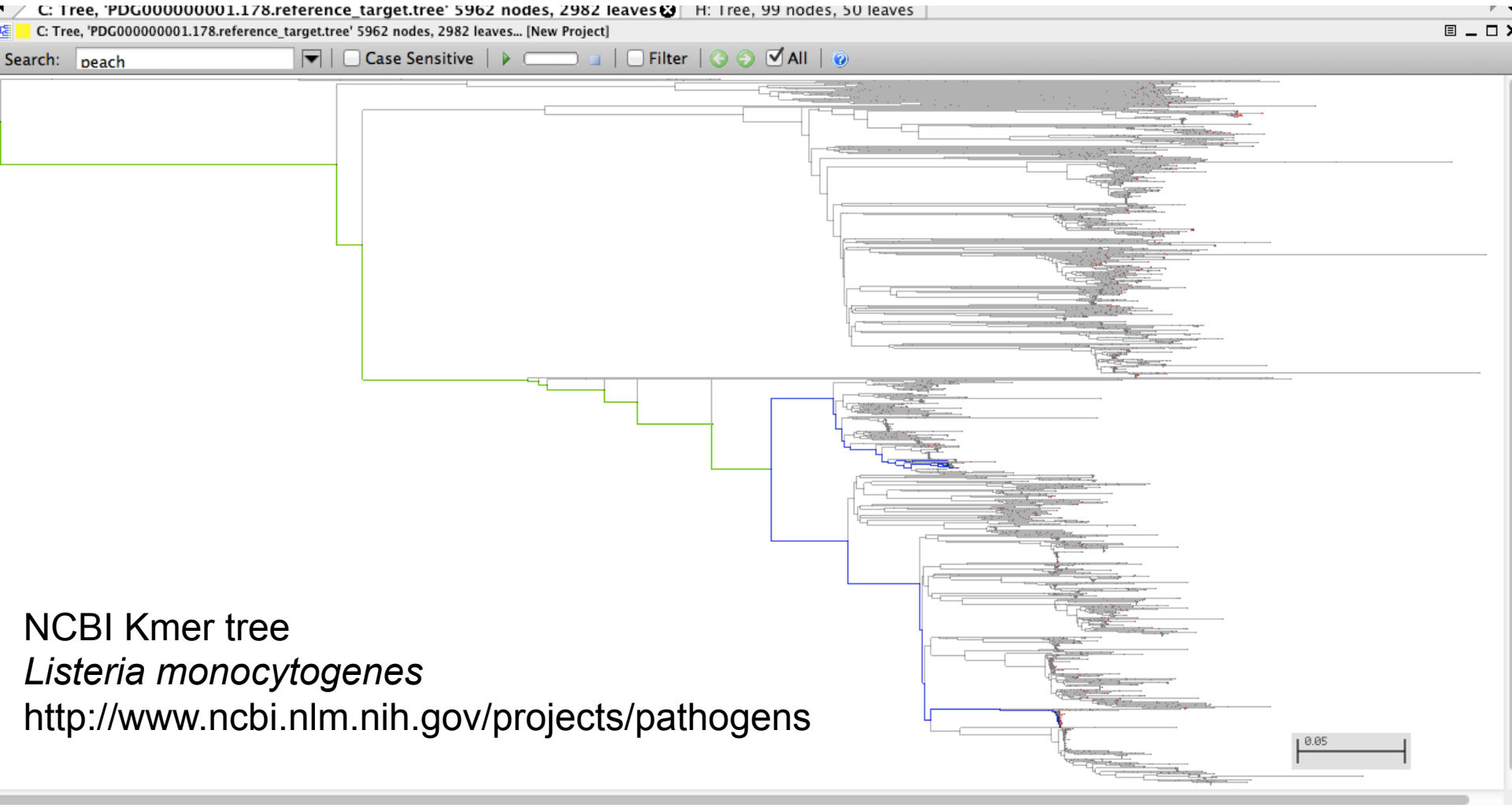
Who

collected by

NCBI pathogen detection pipeline



Daily phylogenetic trees

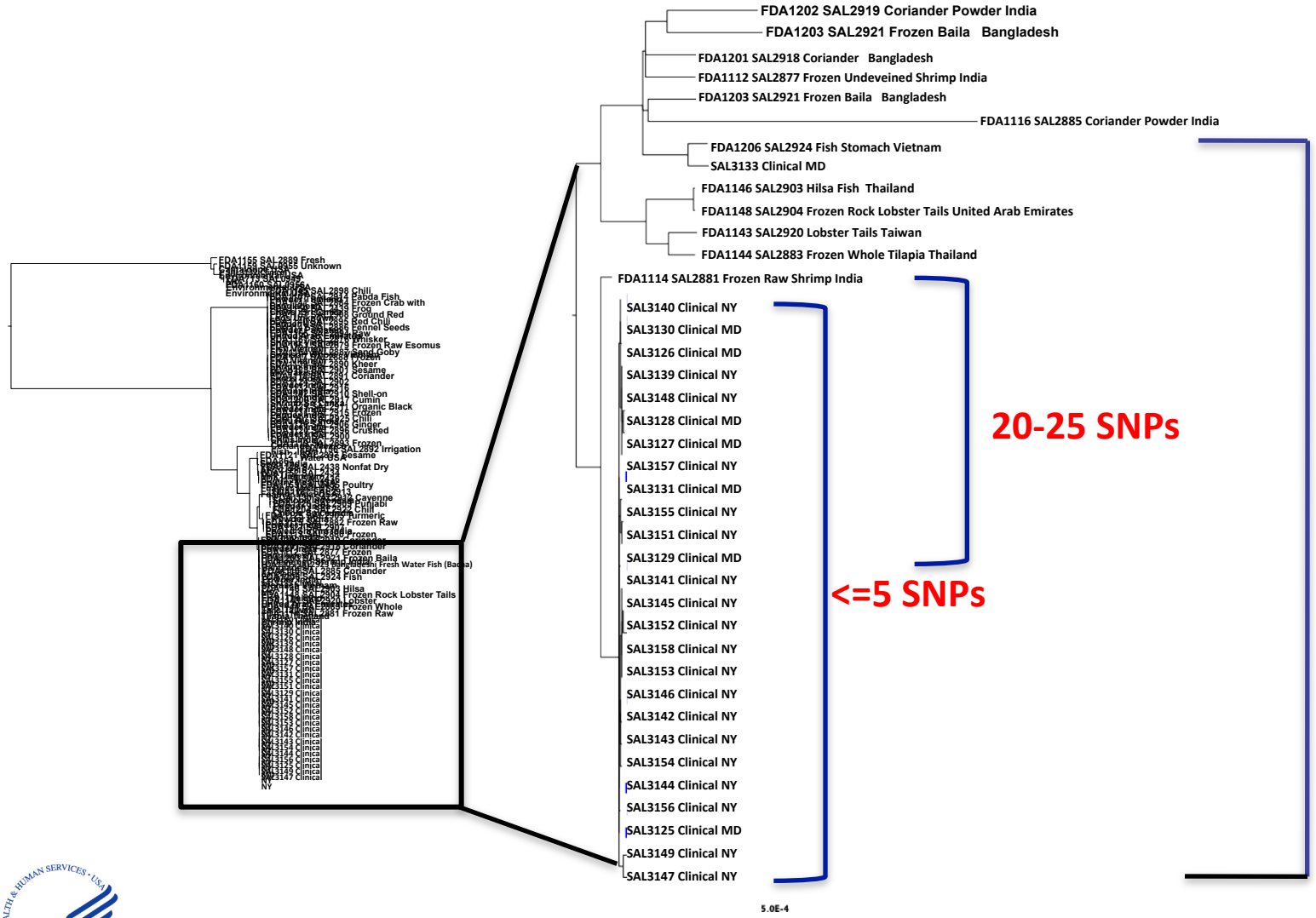


NCBI Kmer tree
Listeria monocytogenes
http://www.ncbi.nlm.nih.gov/projects/pathogens

Example 1: Global source traceback

- In 2012, 410 individuals in the U.S. became sick from ingesting food that contained S. Bareilly
- CDC investigated a multistate (29 states) outbreak
- The illnesses were linked to frozen raw yellowfin tuna (Nakaochi Scrape) which was imported from India
- FDA examined the WGS of these isolates and compared them to previous S. Bareilly events

S. Bareilly Phylogeny



PFGE Match

110-130 SNPs

NGS distinguishes geographical structure among closely related *Salmonella* Bareilly strains

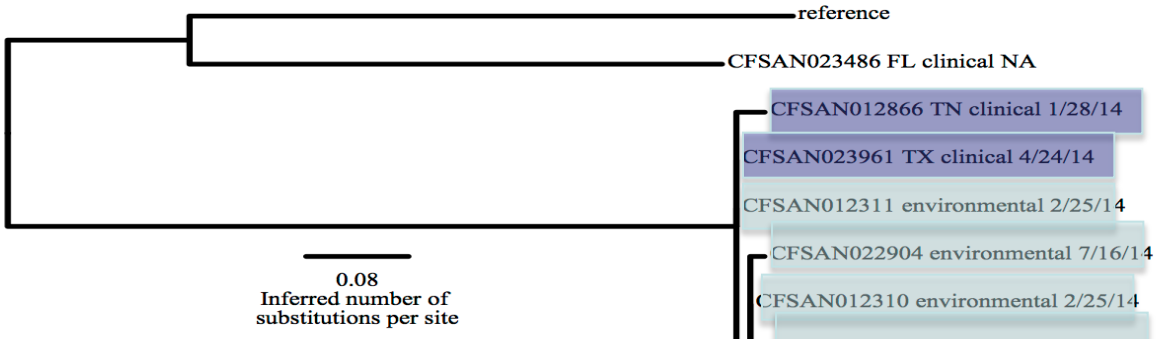
WGS supports preventative controls

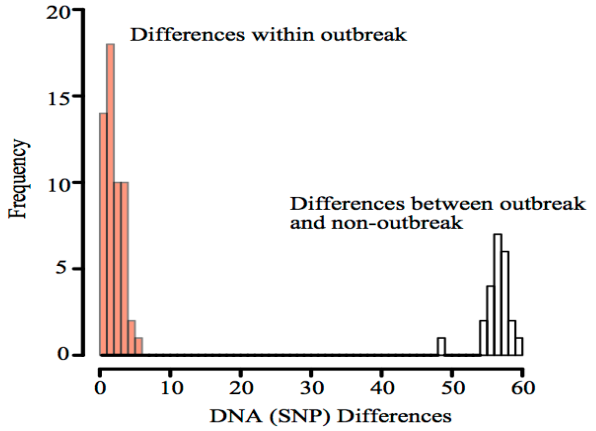
- Permits deep dive to solve persistent/complex problems in a facility or on a farm
- Comparison of internal WGS results to public database of food/environmental isolates
- Public software and analysis tools readily available to industry for viewing of results

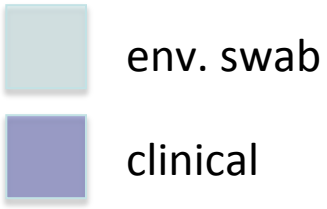
Example 2:
S. Braenderup 2014 pre-outbreak

- In 2014, FDA conducted baseline environmental sampling in nut butter processing facilities
- A few of the samples tested positive for S. Braenderup and a PFGE pattern matched several cases of recent salmonellosis without a common link
- WGS was performed on both environmental and clinical isolates and found to be extremely close (2 SNP differences)

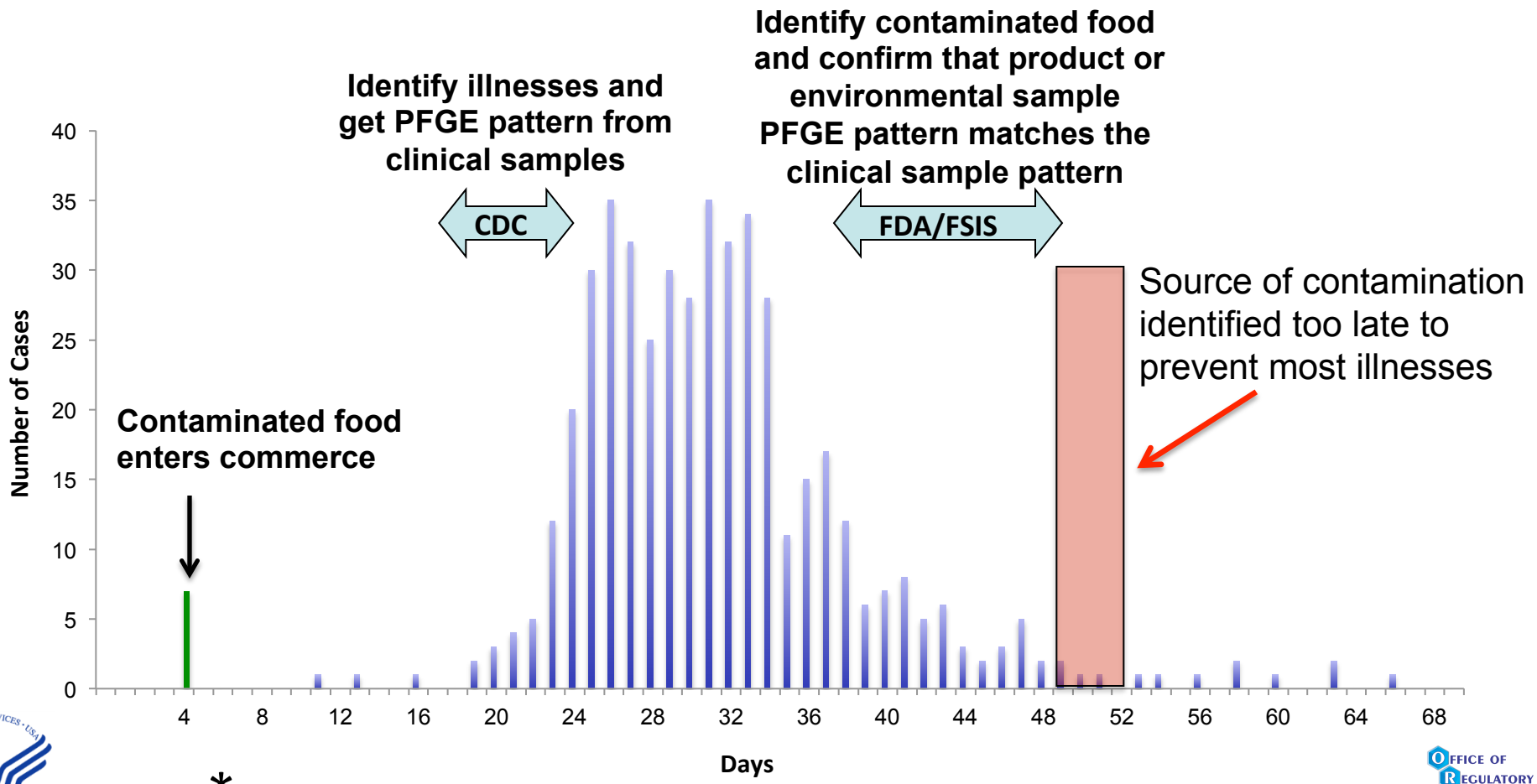
S. Braenderup 2014 pre-outbreak





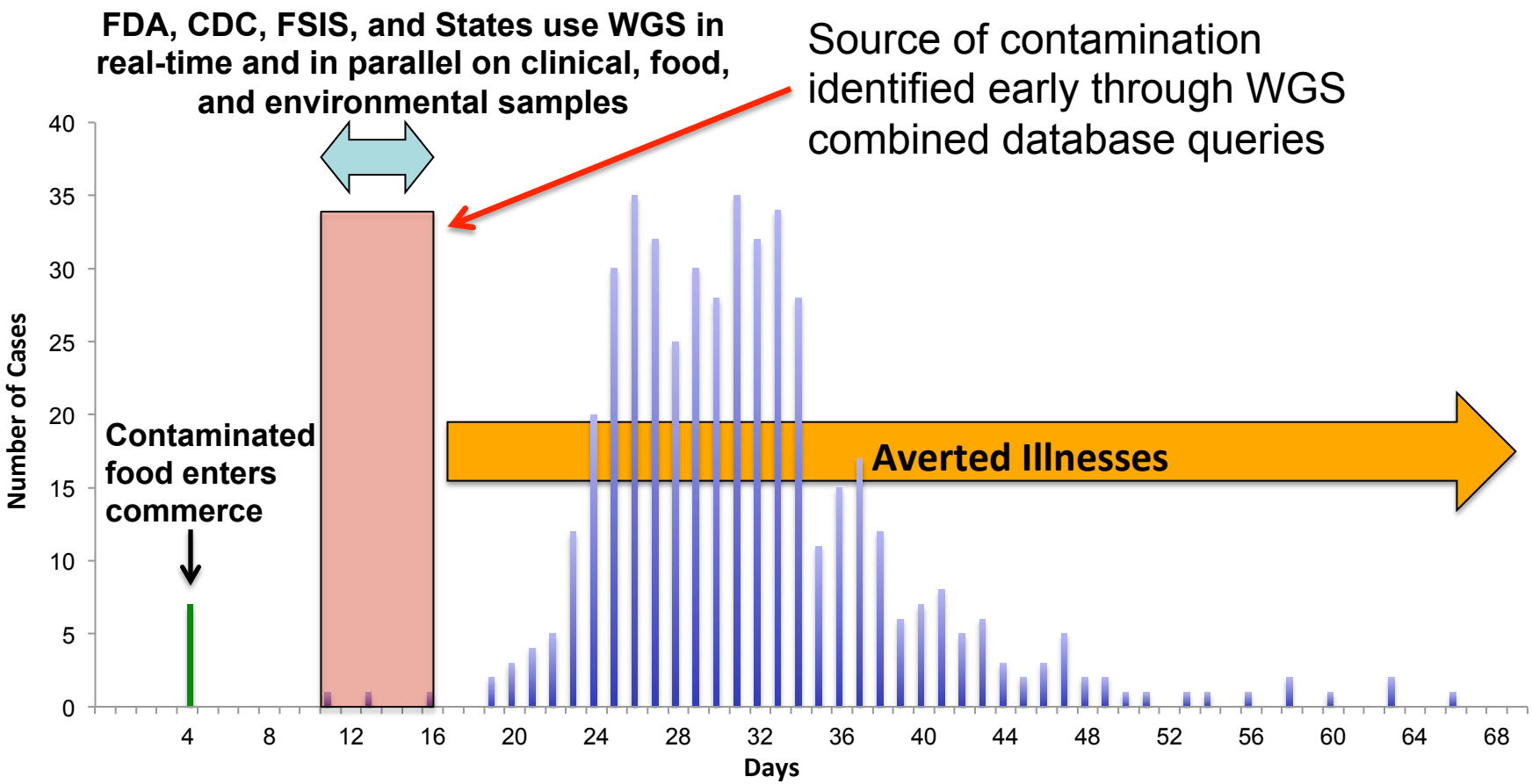


Representative* timeline for traditional approach (PFGE)



*Data is for illustrational purposes and does not represent an actual outbreak

Representative* timeline using WGS



*Data is for illustrational purposes and does not represent an actual outbreak

Immediate impacts of WGS to industry, growers, and distributers, countries, states.

- Earlier intervention means:
  - Reduced amount of recalled product
  - Fewer sick patients which means fewer lawsuits
  - Less impact overall and minimal damage to brand recognition

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**
  - Eastern Laboratory
- **CDC**
  - Enteric Diseases Laboratory
- **INEI-ANLIS "Carolos Malbran Institute," Argentina**
- **Centre for Food Safety, University College Dublin, Ireland**
- **Food Environmental Research Agency, UK**
- **Public Health England, UK**
- **WHO**
- **Illumina**
- **Pac Bio**
- **CLC Bio**
- **Other independent collaborators**