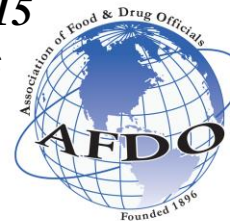*June 20-24, 2015*
*Indianapolis*

*Regulatory Genomics and Beyond:*

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

Eric W. Brown, Ph.D.
Director
Division of Microbiology
Center for Food Safety & Applied Nutrition
U.S. Food & Drug Administration
College Park, Maryland 20740

FDA
Center for Food
Safety and Applied Nutrition

# PFGE v/s WGS

- ## WGS is high resolution
  - ➢ 3-5 million data points are collected for each isolate

- ## WGS analyses are statistically robust
  - ➢ Unlike PFGE patterns, WGS data can be analyzed in its evolutionary context. Accurate and stable genetic changes within pathogen genomes enable us to pin point specific common sources of outbreak strains (farms, processing plants, food types, and geographic regions).

  - ➢ **Source Tracking is Key Application**

# Background: CFSAN SNP Pipeline

Single Nucleotide Polymorphism
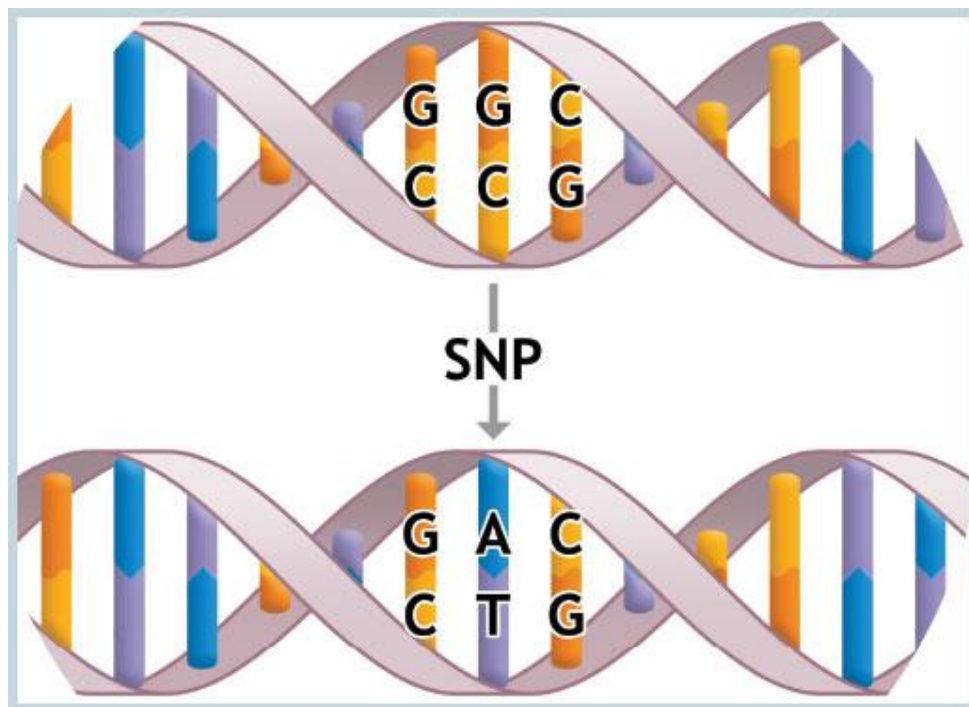


Reference

Food or Environmental Isolate

http://www.ibbl.lu/wp-content/uploads/2012/07/SNPs.jpg

**Circa 2009    IN or OUT?**

**FOODBORNE OUTBREAK INVESTIGATION:**
**WGS analysis of foodborne salmonellae case study**

This investigation focused on *Salmonella* Montevideo samples associated with red and black pepper used in the production of Italian-style spiced meats in a New England processing facility. This manufacturer was implicated in a major salmonellosis outbreak that affected more than 272 people in 44 states and the District of Columbia.

15-20x shot gun sequencing
35 pure culture isolates
from patients, foods and
Environmental samples.

Concatenate 40 variable genes for
Phylogenetic analysis

# FDA WGS Application to Actual Food Contamination Events

Montevideo black and red pepper
Senftenberg black and red pepper
Enteritidis shell/liquid eggs
Heidelberg ground turkey
Heidelberg chicken broilers
Heidelberg chicken livers
Enteritidis custard
Bareilly tuna scrape
Tennessee peanut butter/peanut butter paste
Typhimurium peanut butter
Braenderup peanut butter/nut butter
Tennessee cilantro
Agona dry cereal
Agona papaya
Newport tomatoes
Newport environmental
Kentucky - Cerro dairy/dairy farms
Anatum spices/pepper flakes
Javiana  cantaloupes
Saintpaul hot peppers
4,5,12: i –

Lmono cantaloupes
Lmono queso cheese
Lmono potato salad
Lmono artisanal cheeses
Lmono avocados
Lmono ricotta
Lmono celery/chix salad
Lmono smoked fish
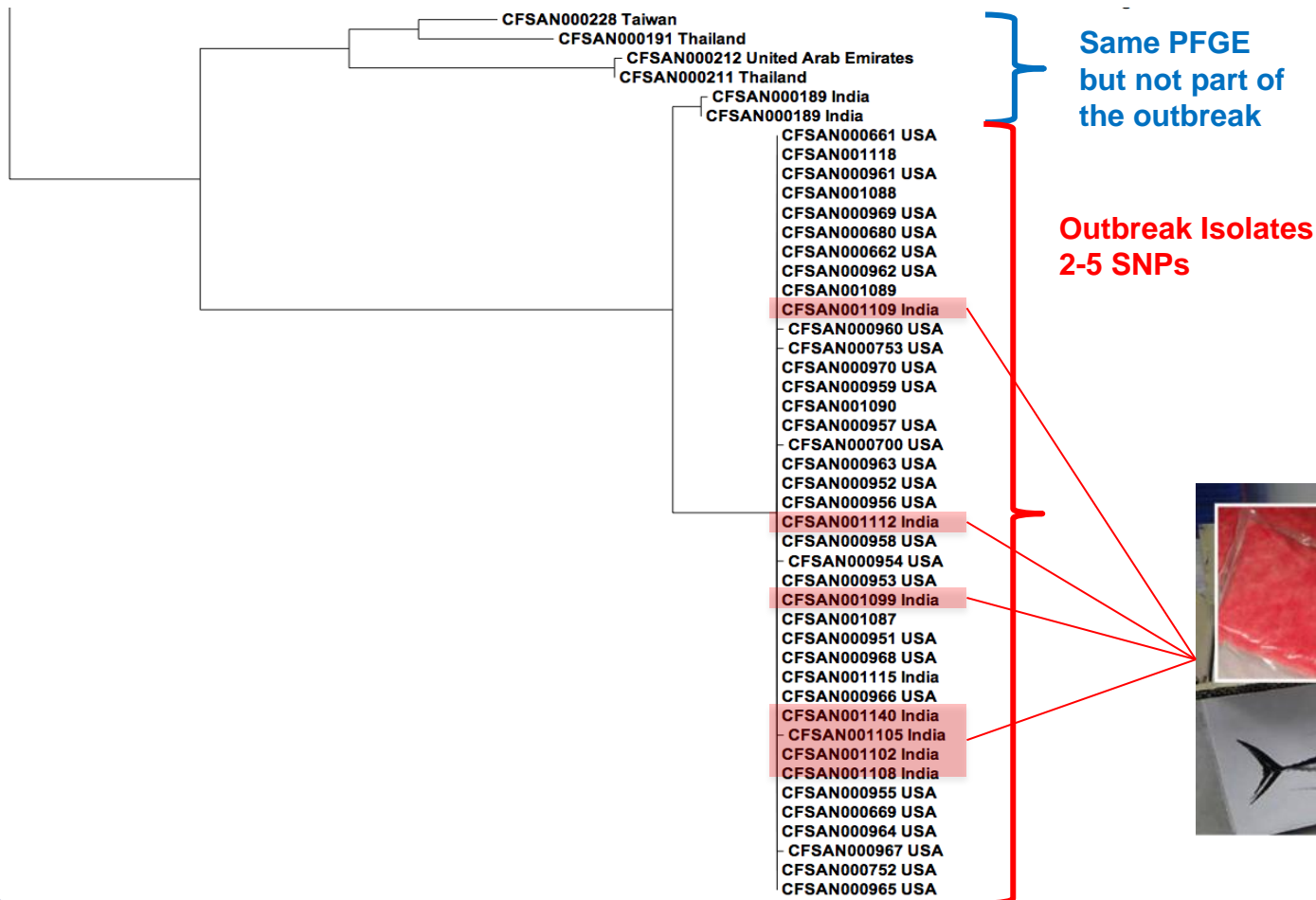Lmono other herbs
Lmono peaches

Cronobacter infant formula
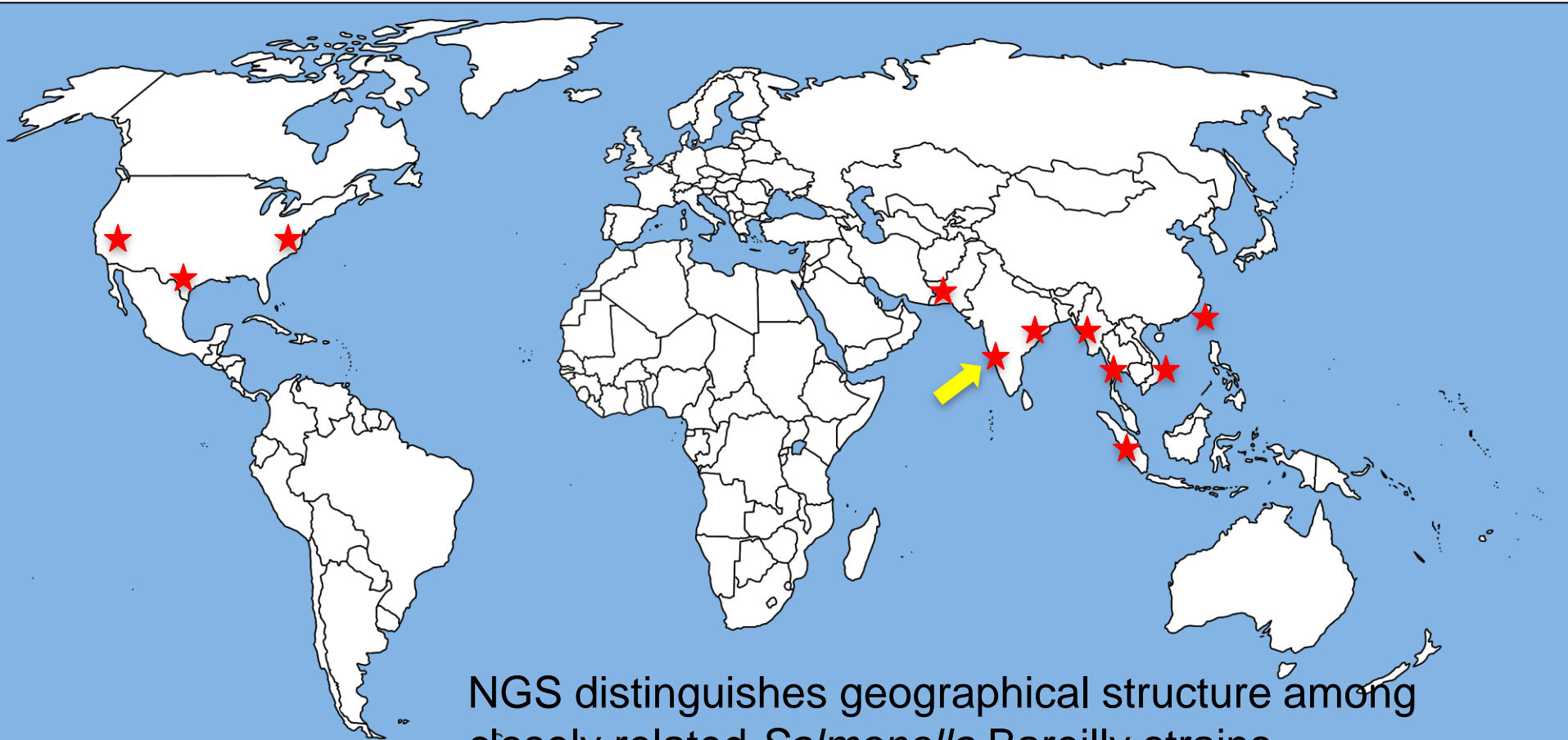
V para oysters

EcO157:H7 lettuce

STEC beef

…Numerous other taxa

# SNP phylogeny for *S.* Bareilly strains



CFSAN000228 Taiwan
CFSAN000191 Thailand
CFSAN000212 United Arab Emirates
CFSAN000211 Thailand
CFSAN000189 India
CFSAN000189 India

**Same PFGE but not part of the outbreak**

CFSAN000661 USA
CFSAN001118
CFSAN000961 USA
CFSAN001088
CFSAN000969 USA
CFSAN000680 USA
CFSAN000662 USA
CFSAN000962 USA
CFSAN001089
CFSAN001109 India
CFSAN000960 USA
CFSAN000753 USA
CFSAN000970 USA
CFSAN000959 USA
CFSAN001090
CFSAN000957 USA
CFSAN000700 USA
CFSAN000963 USA
CFSAN000952 USA
CFSAN000956 USA
CFSAN001112 India
CFSAN000958 USA
CFSAN000954 USA
CFSAN000953 USA
CFSAN001099 India
CFSAN001087
CFSAN000951 USA
CFSAN000968 USA
CFSAN001115 India
CFSAN000966 USA
CFSAN001140 India
CFSAN001105 India
CFSAN001102 India
CFSAN001108 India
CFSAN000955 USA
CFSAN000669 USA
CFSAN000964 USA
CFSAN000967 USA
CFSAN000752 USA
CFSAN000965 USA

**Outbreak Isolates 2-5 SNPs**

0.04

PFGE identical in red



NGS distinguishes geographical structure among closely related *Salmonella* Bareilly strains
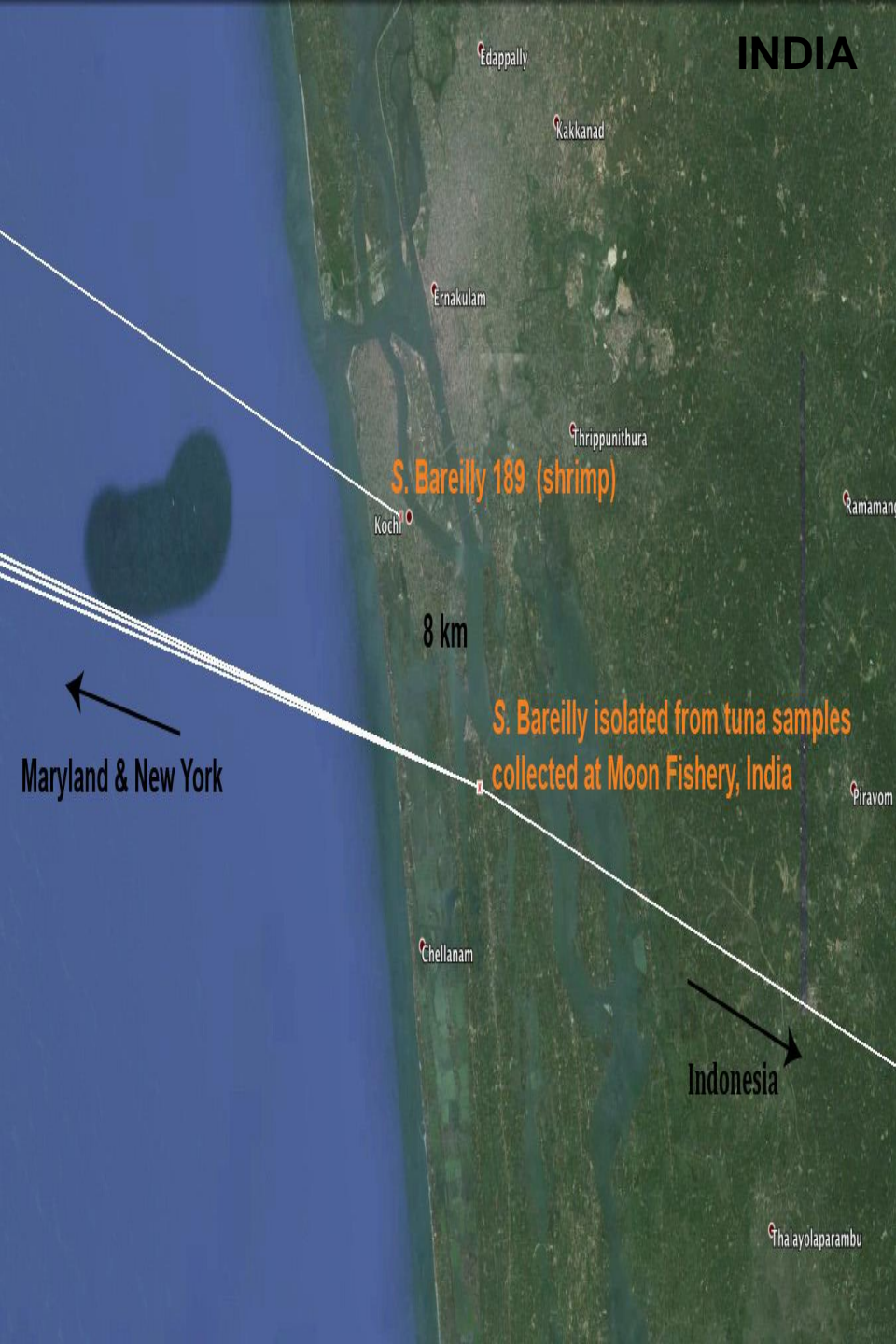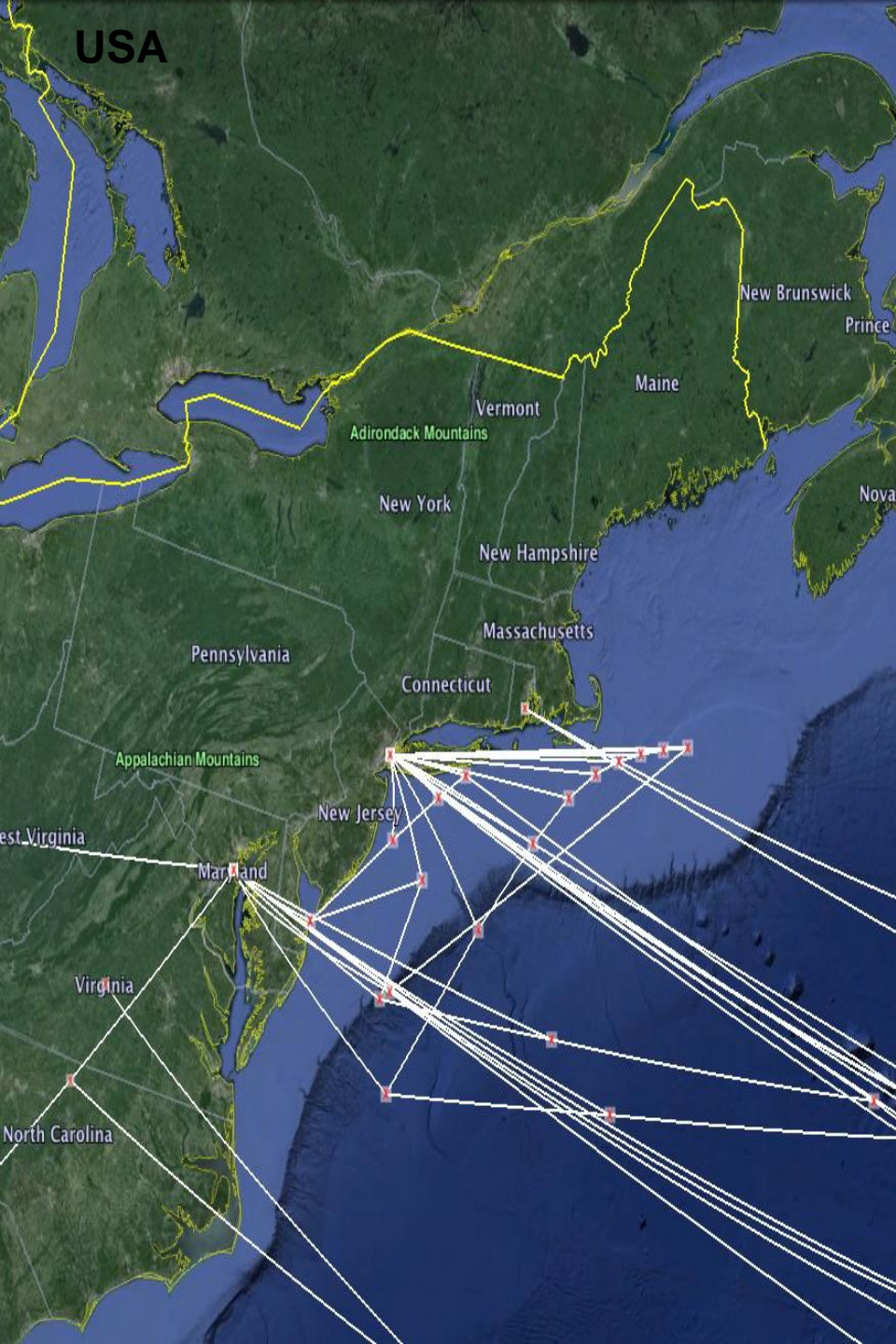
# Is WGS a viable solution?

- Cost
- Increasing ease of operation
- Database longevity
- Comparable times to conventional pipelines
- Sample prep
  - Identical for all pathogens
- Cost savings
  - Resistance, subtyping, virulence factors, more…
- New applications
  - tracking, regulatory/compliance actions, historical trends, more…
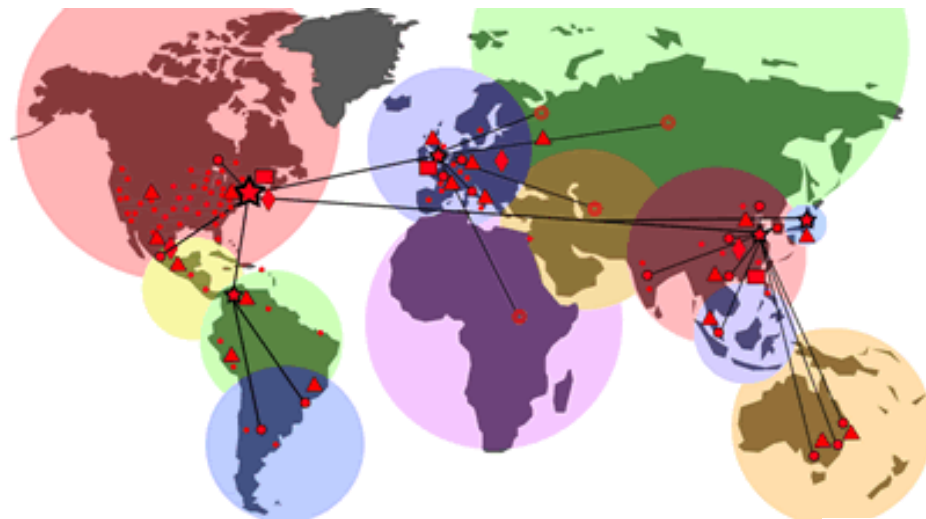
**Cost per bacterial genome**

454

**Illumina Miseq**
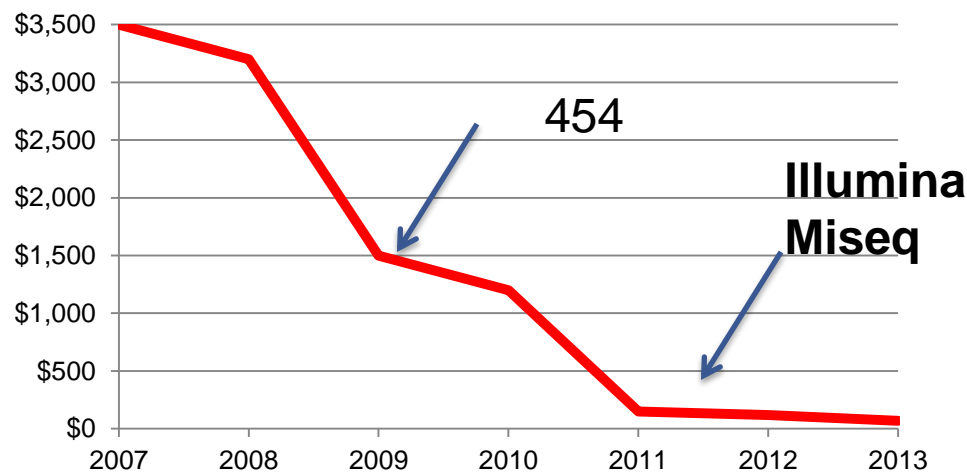
$70/genome in 2014
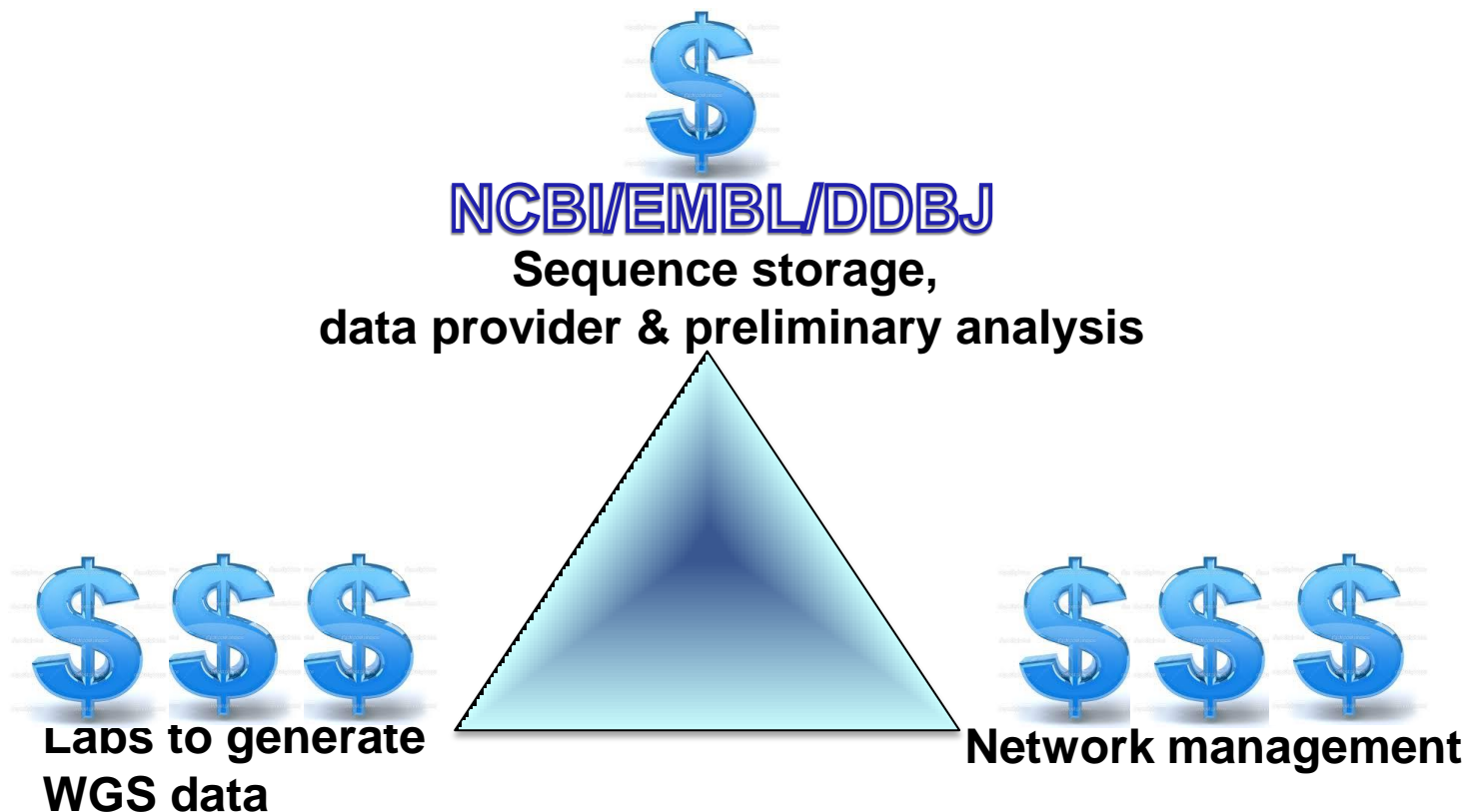
$40/genome in 2015 w/ **Illumina NextSeq** Technology

# Why Develop a WGS based Network?

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

# Key elements of a national/international WGS network

**NCBI/EMBL/DDBJ**

**Sequence storage,
data provider & preliminary analysis**

**Labs to generate WGS data**

**Network management**

# **GenomeTrakr Strategy**

- Develop a <u>distributed</u> sequencing based network, rather than centralized model

- Provide sequence and minimal metadata in a publicly accessible database

  - Partner with NCBI for storage and serving data

  - Cost prohibitive for FDA to establish its own high capacity data site

  - Industry (food, pharma, and methods development), academia, hospitals, clinical public health laboratories, and other government agencies have access to data for individual needs (ie, BioFire Test Film – economic development, job creation)
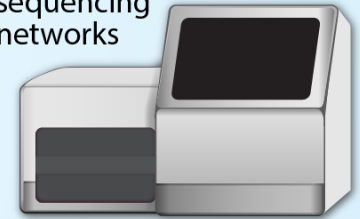
# Basic Data Flow for Global WGS Public Access Databases

**DATA ACQUISITION**

*Sequence and upload genomic and geographic data*
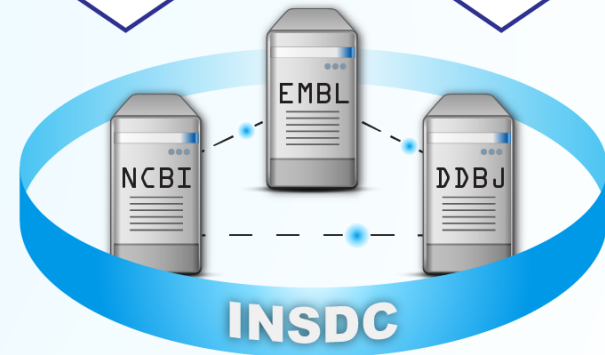
Other distributed sequencing networks

**DATA ASSEMBLY, ANALYSIS, AND STORAGE**

*International Nucleotide Sequence Database Collaboration (INSDC)*
*Shared Public Access Databases*

- *NCBI – National Center for Biotechnology Information*
- *EMBL – European Molecular Biology Laboratory*
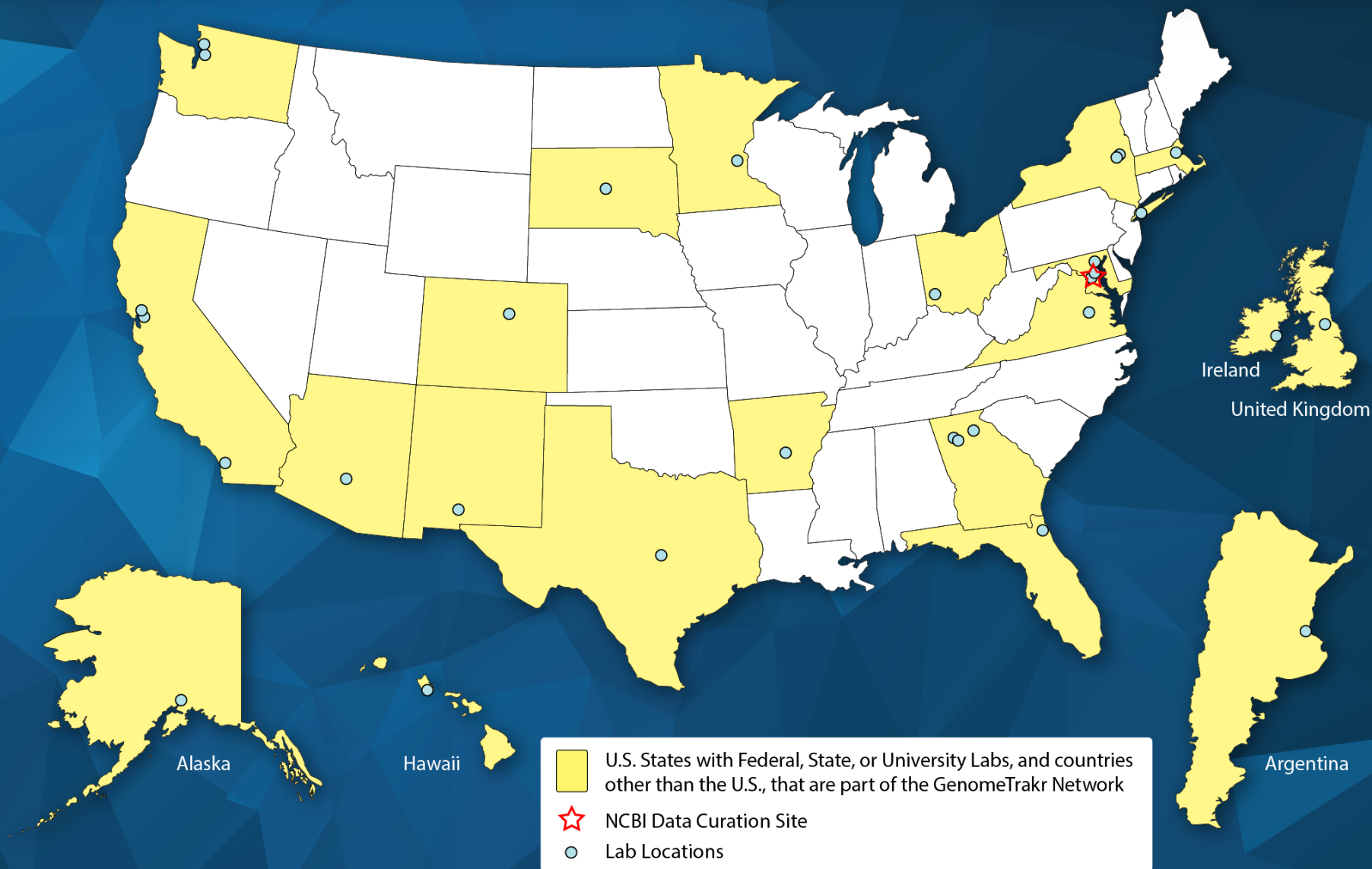- *DDBJ – DNA Databank of Japan*

**INSDC**

**PUBLIC HEALTH APPLICATION AND INTERPRETATION OF DATA**

- *Find clinical links*
- *Identify clusters*
- *Conduct traceback*
- *Develop rapid methods*
- *Develop culture independent tests*
- *Develop new analytical software*

11/2014    **State, Local, Federal, and Foreign Public Health Agencies**    Academia/Industry

# GenomeTrakr Labs

Ireland

United Kingdom

Alaska

Hawaii

Argentina

U.S. States with Federal, State, or University Labs, and countries other than the U.S., that are part of the GenomeTrakr Network

☆ NCBI Data Curation Site

◯ Lab Locations

May 2015 | Map is not to scale

**Cumulative Number of *Salmonella* and *Listeria* Isolates Sequenced Over Time by the GenomeTrakr Network**
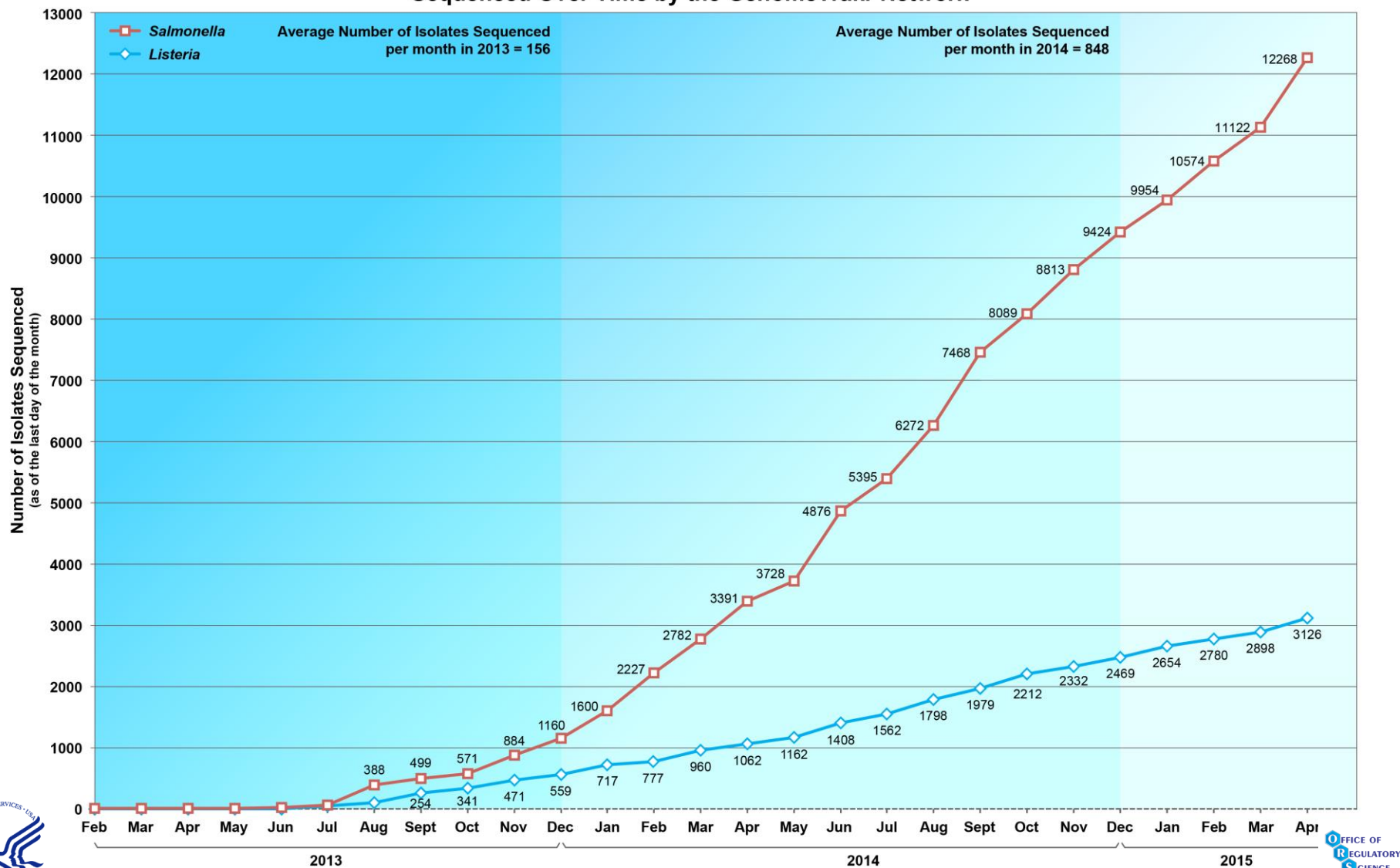
# Real-time Integration of WGS into
# FDA regulatory workflow

**DEPARTMENT OF HEALTH AND HUMAN SERVICES**

Food and Drug Administration
Silver Spring, MD 20993

March 11, 2014

██████████████ t

Roos Foods Inc.
251 Roos Lane
Kenton, DE 19955

ORDER: Suspension of Food Facility Registration
Notice of Opportunity for Hearing

███████████████████

The U.S. Food and Drug Administration (FDA) hereby issues this Order to suspend the registration of your food facility, Roos Foods, Inc. (Roos), located at 251 Roos Lane, Kenton, DE 19955. Your food facility was registered with FDA pursuant to section 415(a) of the Federal Food, Drug, and Cosmetic Act (FD&C Act) (21 U.S.C. 350d(a)) on June 4, 2013. Section 415(b)(1) of the FD&C Act provides, in relevant part, that if FDA determines that a food manufactured, processed, packed, received, or held by a facility registered under section 415 has a reasonable probability of causing serious adverse health consequences or death to humans or animals, FDA may by order suspend the registration of a facility (1) that created, caused, or was otherwise responsible for such reasonable probability; or (2) that knew of, or had reason to know of, such reasonable probability, and packed, received, or held such food.
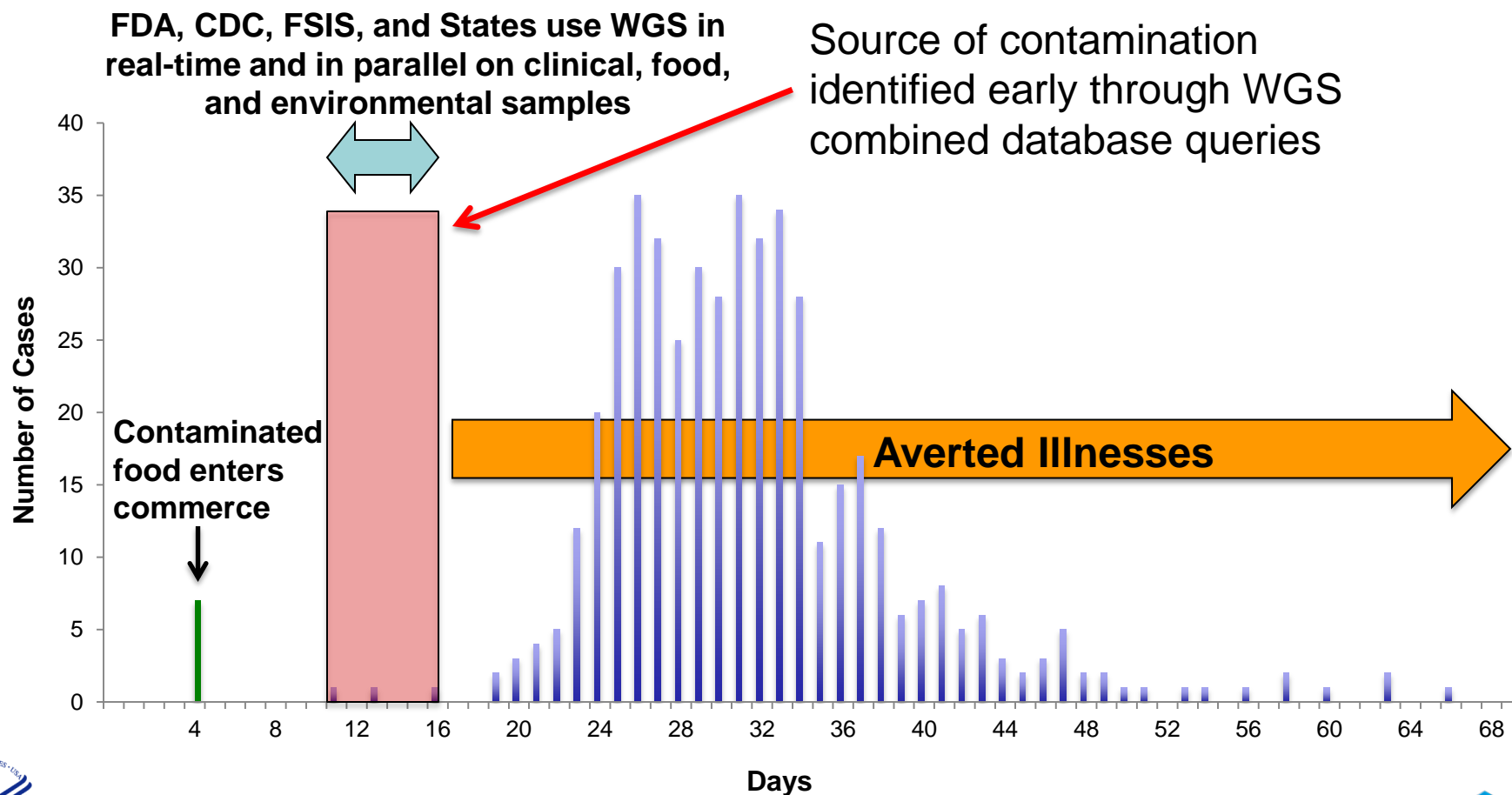
# S. Braenderup
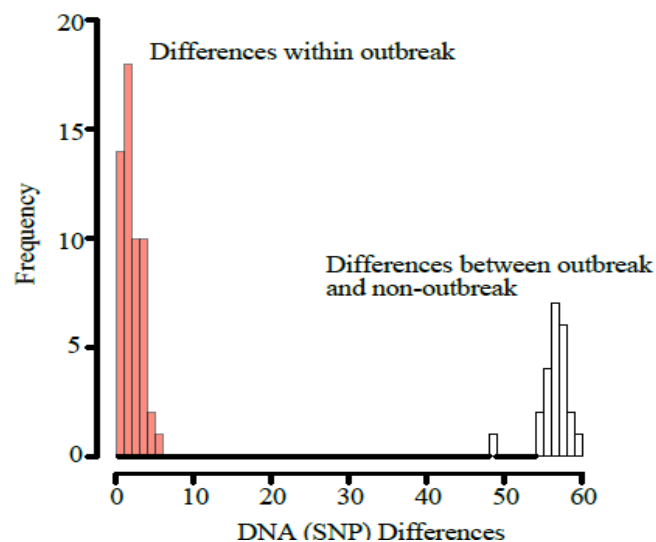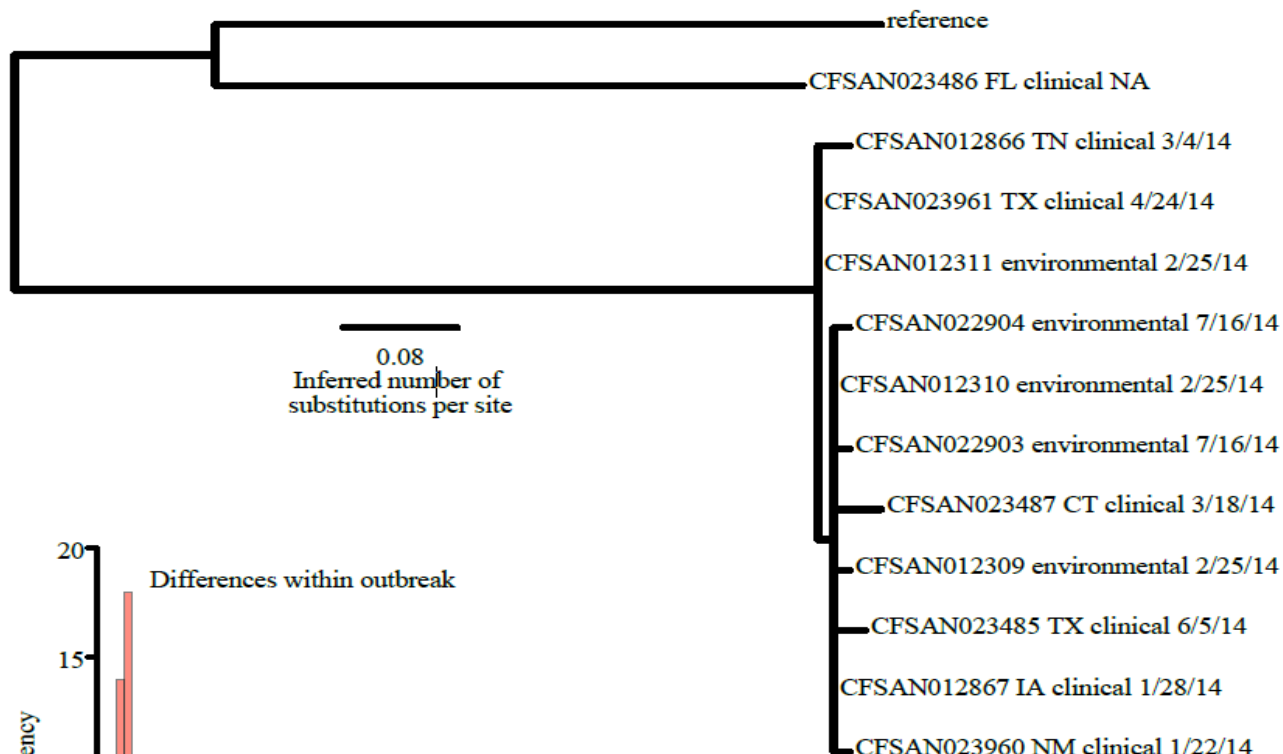
# Health and Economic Impact of Active WGS-based Surveillance

**Comparison of 2 related contamination events.**

❑ **Similar facilities – broad domestic distribution.**

❑ **Nut butter 1:  42 cases and 10 hospitalizations with as many as 1,260 illnesses unreported (Fall 2012)**

❑ **Nut butter 2: – 4 confirmed cases, 1 hospitalization (Summer 2014)**

❑ **WGS informed investigation prevented significant illness and hospitalizations**

  • **lower illness rate and treatment cost ($14,000 per hospitalization and $445 physician visit) + fraction of loss of productivity, long-term and chronic onset complications associated with Salmonella infection**

# **Example: L. mono in sprouts**



1. Food and environmental samples collected as part of a routine inspection at a sprout facility
   - L. mono detected in 3 swabs
2. WGS analysis showed these isolates to be a match to recent clinical isolates sequenced as part of the CDC Real-Time Listeria project
3. Re-inspection of the facility found additional positive environmental L. mono positives
4. WGS analysis showed that the original food/environmental isolates, re-inspection environmental isolates, and clinical isolates from this period to be a match.

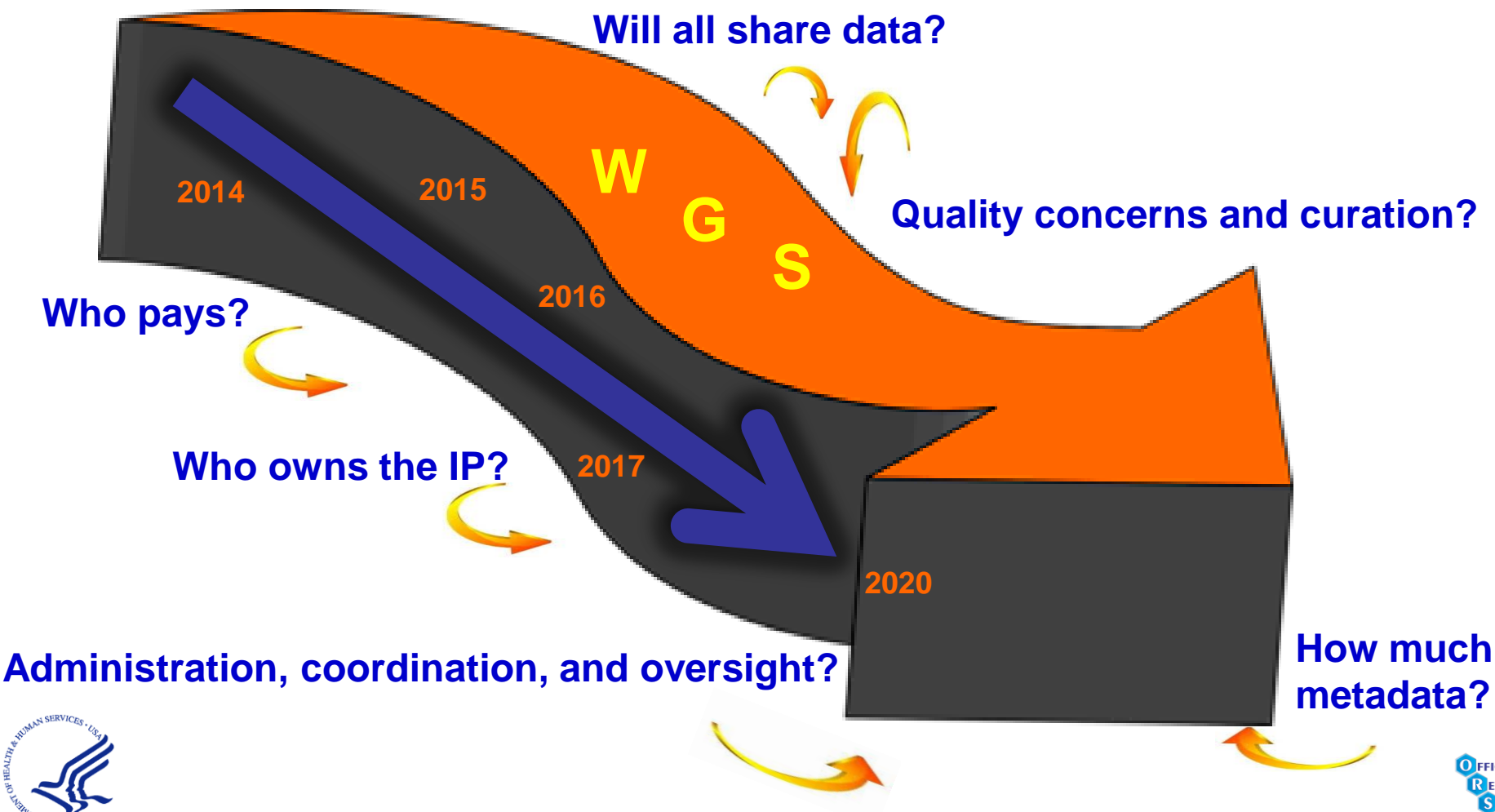Bootstrap >= 95

0-6 SNPs

>99.999% Identity
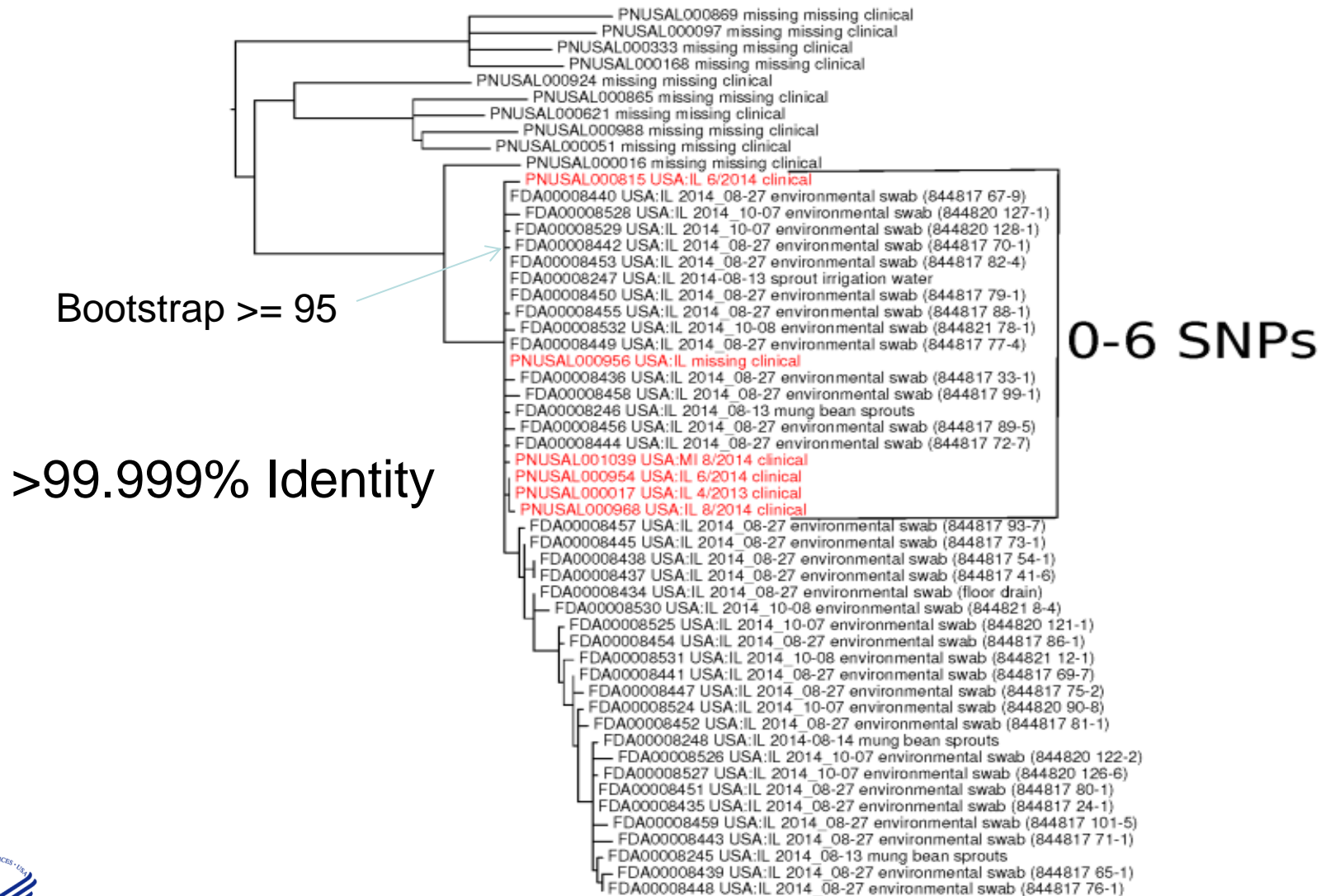
# Current Status

- NGS clearly defines foodborne outbreaks – more than 15 different examples

- NGS network is reliable, efficient and can provide very good location specificity of outbreaks

- GT collaborators have sequenced more than 15,000 *Salmonella*, more than 4,000 *Listeria monocytogenes*, and closed 100 genomes for outbreak mapping.  Our current rate is about 600 draft sequences a month or 1 genome per hour.

- The need for increased number of well characterized environmental (food, water, facility, etc.) sequences may outweigh need for extensive clinical isolates

- Highly successful partnership with CDC on real-time tracking of *Listeria* outbreaks but need to expand into other pathogens like *Salmonella* (#1 most devastating foodborne pathogen)

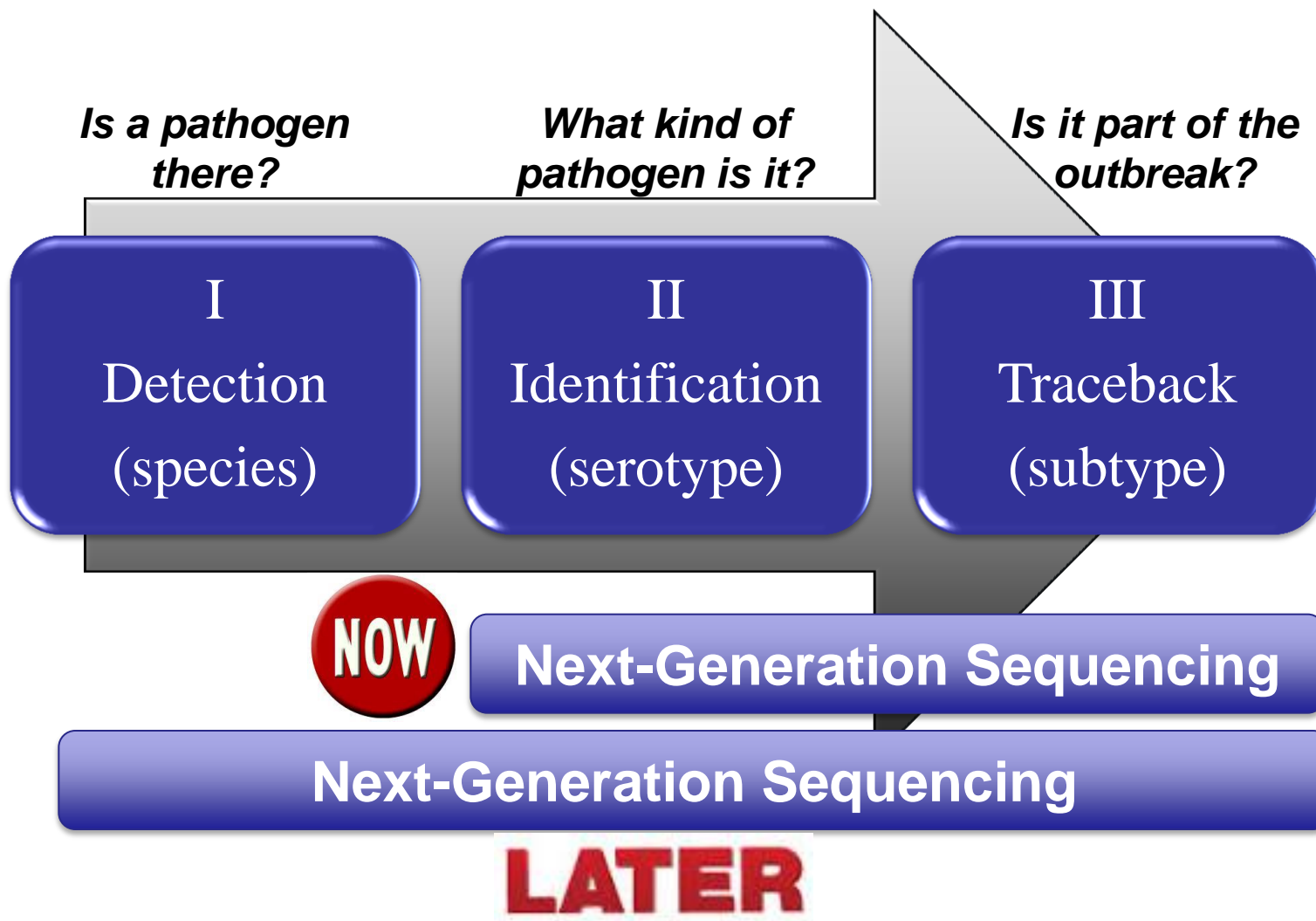## WGS-based monitoring can pinpoint root causes in the Fresh-cut Produce Supply Chain:



Simplified Tomato Distribution Pattern from Farm to Table
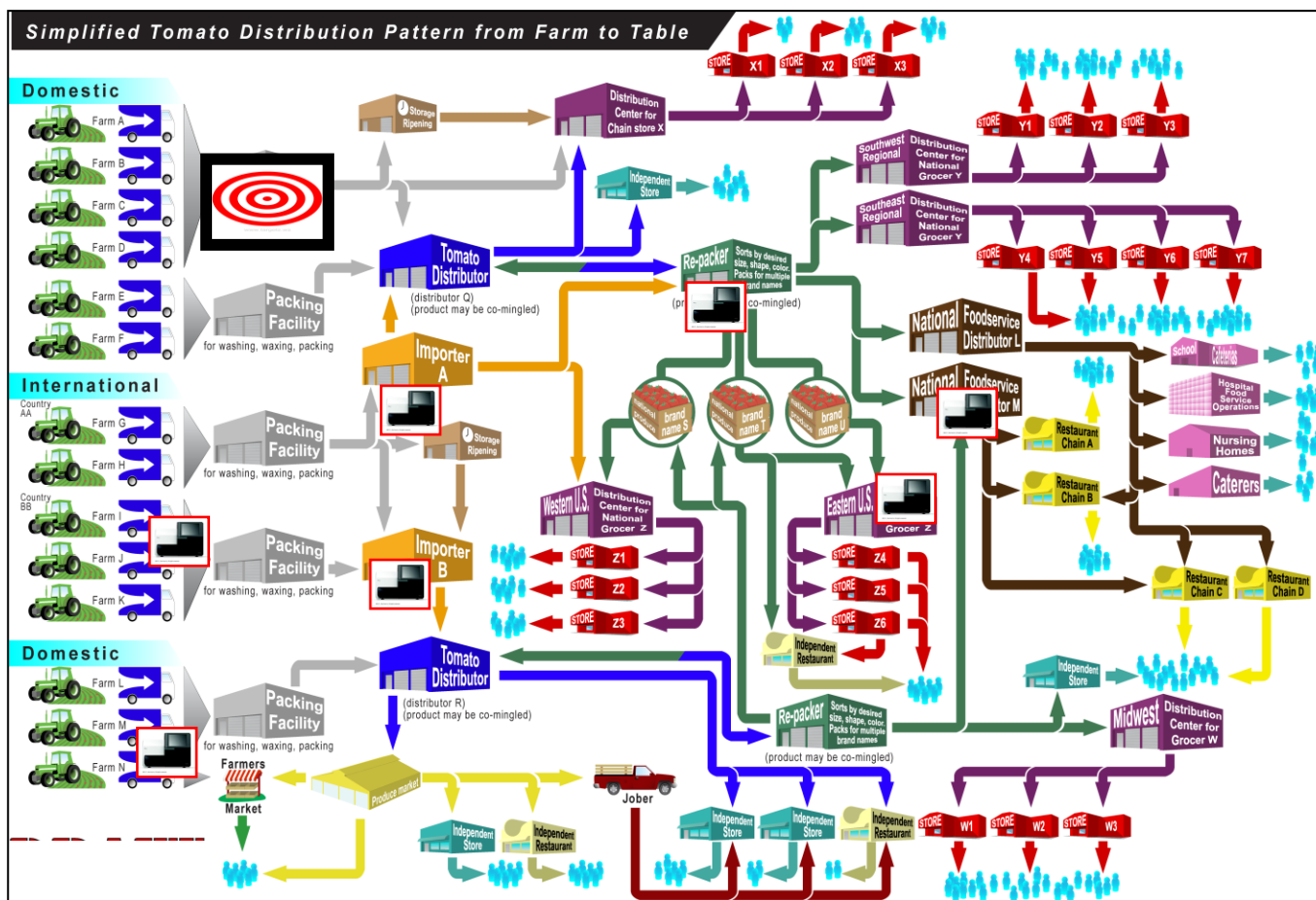
# Next-Generation sequencing faces several large challenges as it deploys to a global public health tool:



Will all share data?

Quality concerns and curation?

Who pays?

Who owns the IP?

W G S

2014

2015

2016

2017

2020

Administration, coordination, and oversight?

How much metadata?

# 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**