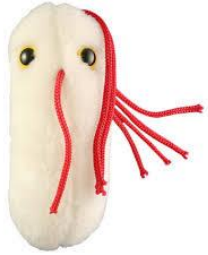
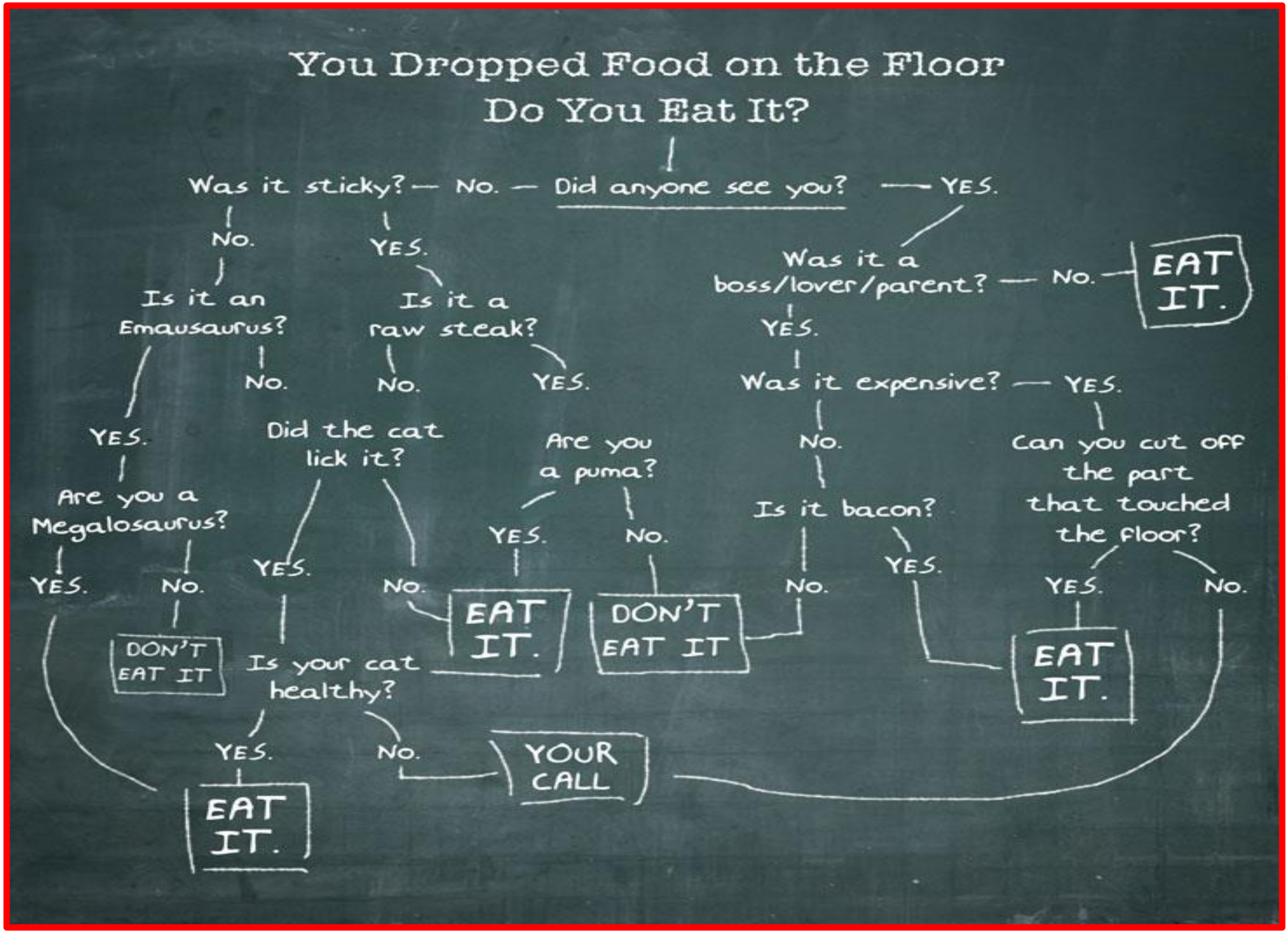


"Integrating Modern Genomic Science into Practical Microbiology: The Case of Food Safety"

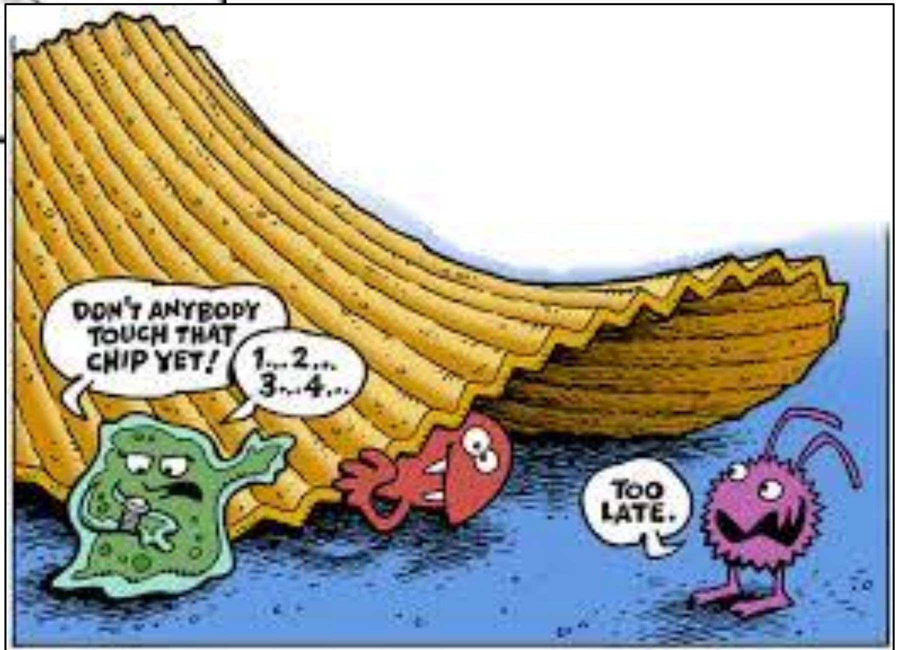
Eric W. Brown, Ph.D., M.Sc., FAAM
Director
Division of Microbiology
Center for Food Safety & Applied Nutrition
U.S. Food & Drug Administration
College Park, Maryland 20740

Recent Advances in Food Safety Science



(Credited to D. Archer, UF)

JAHENRY.COM



"THE 5 SECOND RULE"

What's New in Food Safety?

↓

What's New in Food Microbiology?

↓

What's the New Food Microbiology?

↓

What's the New Microbiology?

Next-Generation Sequencing

A high-throughput sequencing method that parallelizes the sequencing process, producing millions of sequences at once.

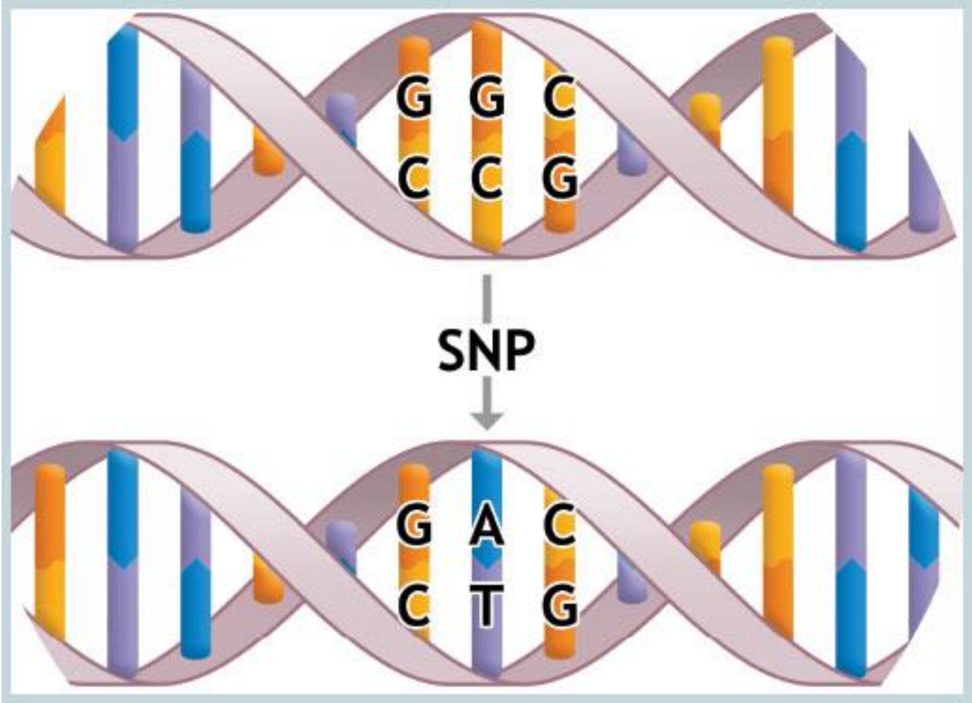
Over the past 7 years, "Next-Generation" sequencing technologies have made accessible data capable of answering questions fundamental to our understanding of life and the factors that govern human health. The combination of the vast increase in data generated, coupled with plummeting costs required to generate these data, has rendered this technology a tractable, general purpose tool for a variety of applications.

"Whole Genome Sequencing Is The Biggest Th
Happen To Food Microbiology Since
Pasteur Showed Us How To Culture Pathogens

Dr. Jorgen Schlundt
Exec Director and Founder
The Global Microbial Identifier

Background: CFSAN SNP Pipeline

Single Nucleotide Polymorphism

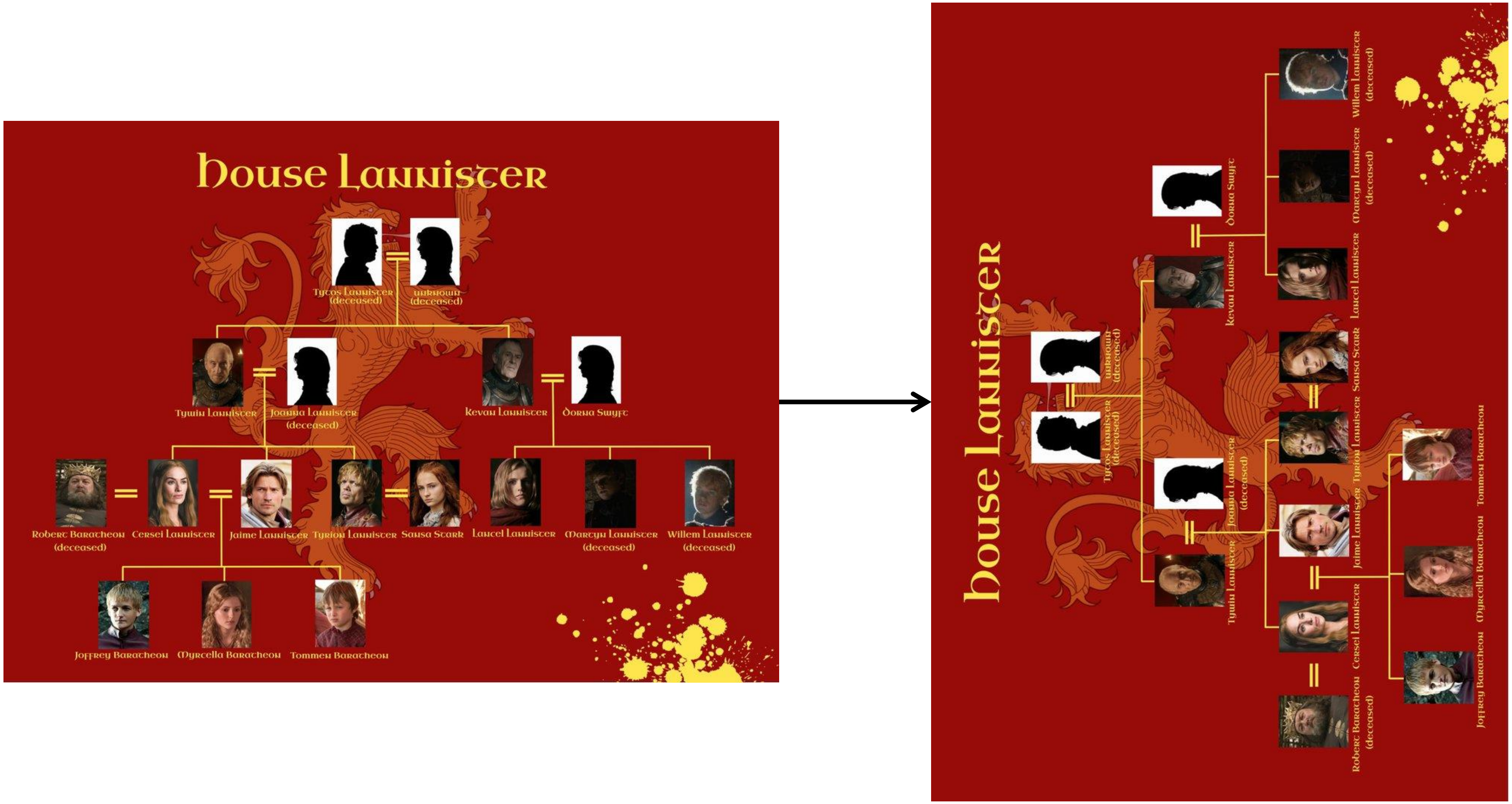


Reference

Food or Environmental Isolate

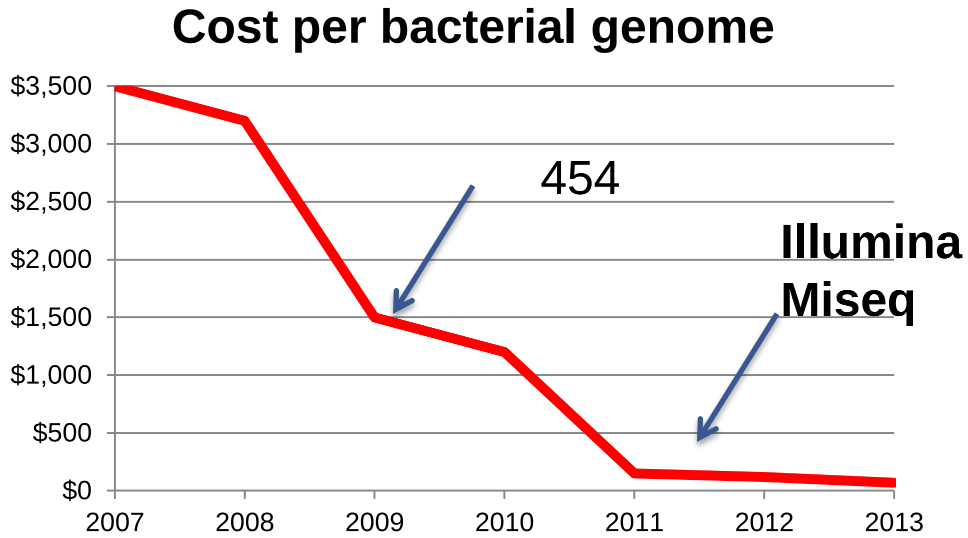
http://www.ibbl.lu/wp-content/uploads/2012/07/SNPs.jpg

Pedigree vs Phylogeny



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…

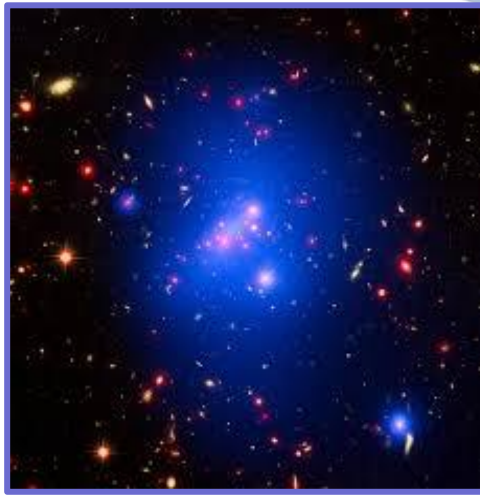


Cost per bacterial genome

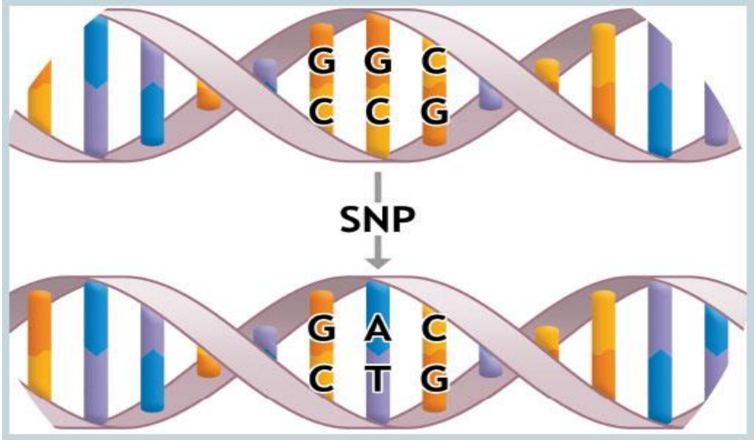
$70/genome
in 2014

$40/genome
in 2015 w/
Illumina NextSeq Technology

PFGE transition to WGS

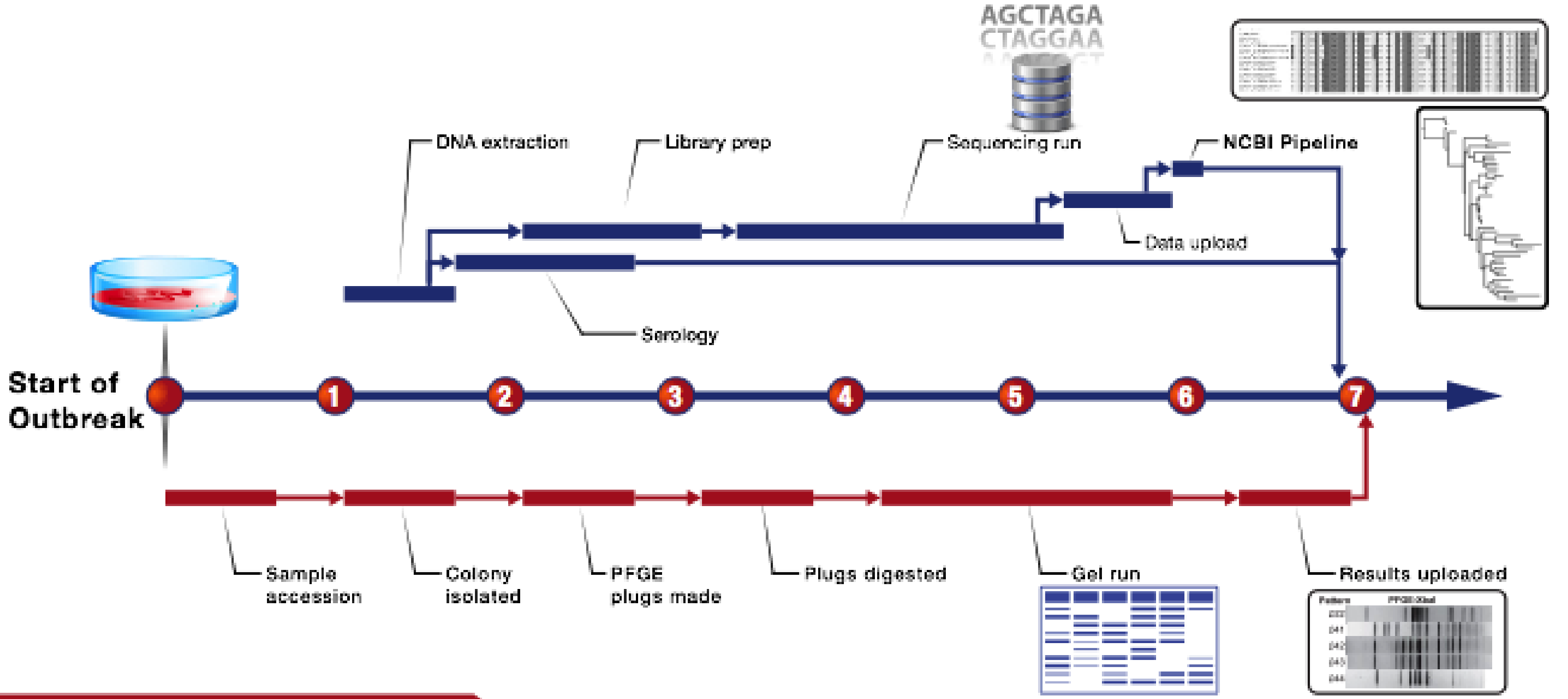


- WGS is high resolution
  - 3-5 million data points are collected for each isolate
- WGS analyses are statistically robust
  - Unlike PFGE patterns, WGS data can be analyzed in its evolutionary context. Accurate and stable genetic changes within pathogen genomes enable us to pin point specific common sources of outbreak strains (farms, processing plants, food types, and geographic regions).
  - ***Source Tracking is Key Application:*** ***moving from PFGE to WGS has been no less impactful than the move from backyard telescopes to the Hubble in terms of resolution in differentiating foodborne outbreaks.***



Next-generation Lab Response vs. Conventional Lab Response

Next-Generation Lab Response



AGCTAGA
CTAGGAA

DNA extraction

Library prep

Sequencing run

NCBI Pipeline

Data upload

Serology

Start of Outbreak

1 2 3 4 5 6 7

Sample accession

Colony isolated

PFGE plugs made

Plugs digested

Gel run

Results uploaded

Conventional Lab Response

1 = Day 1

FDA circa 1906

Topics to be discussed…

- **(1) The rationale for WGS in Food Safety and @ FDA**
- **(2) Augmenting traceback of foodborne outbreaks with WGS**
- **(3) National and International Activities: The GenomeTrakr**
- **(4) Genotypes = Phenotypes, Adaptations and other WGS byproducts**

The Rationale for WGS and other 'Omics in FDA's Food Safety Programs

Safe, Wholesome, Sanitary Foods

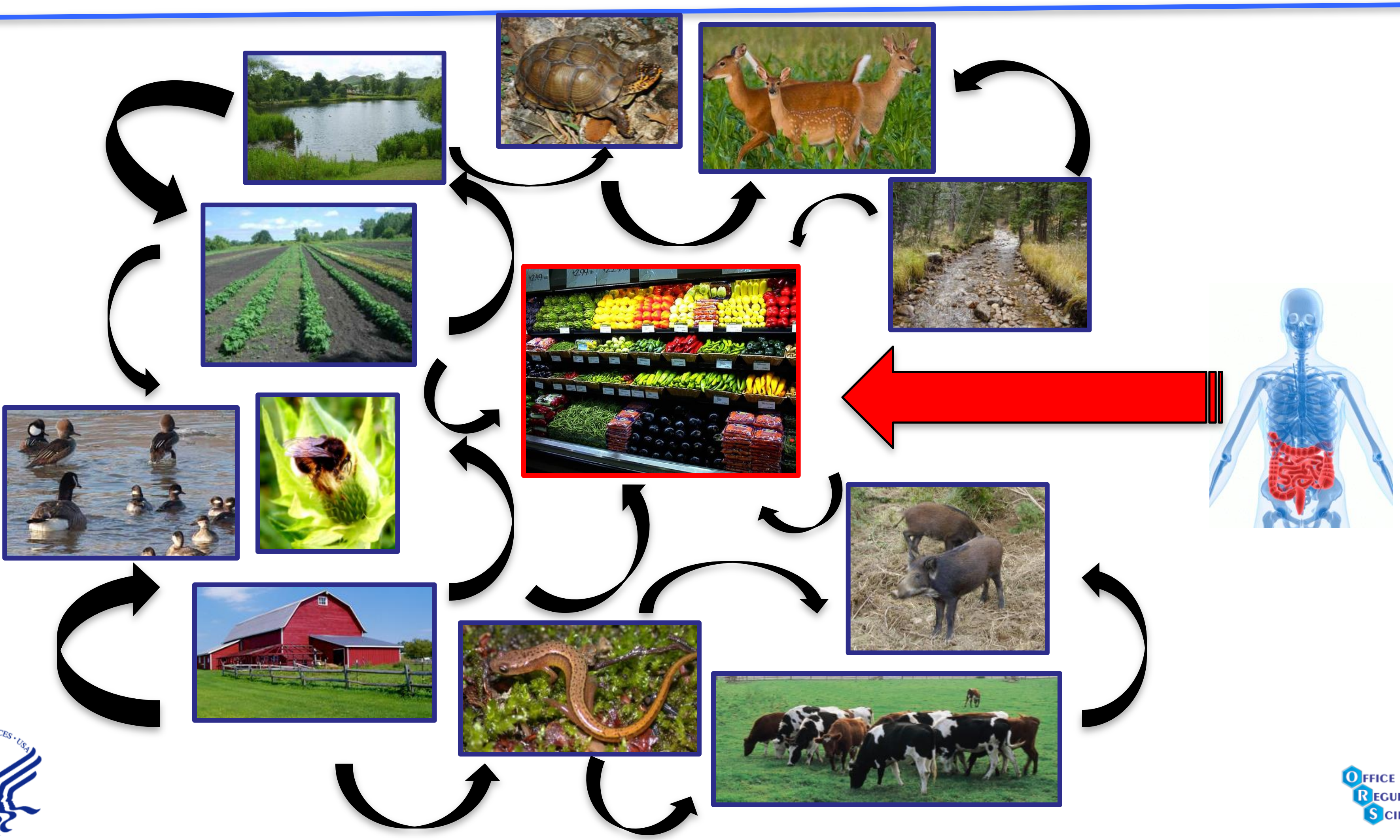
The Salmonella 'radiation'

[redacted]

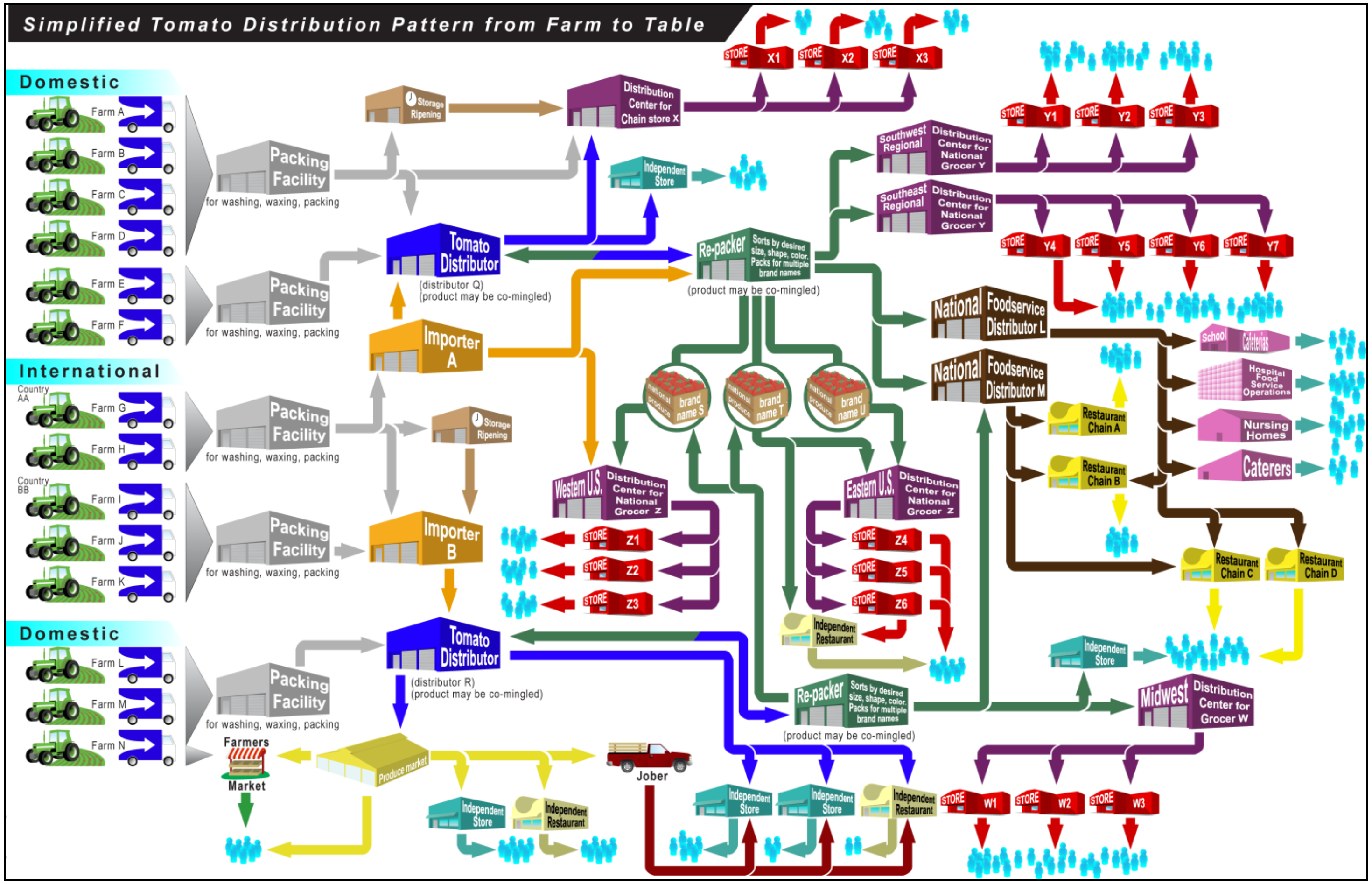
[redacted]

[redacted]

Understanding How *Salmonella* Contamination found its way to the food supply is not an easy task...



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



The Fresh-cut Tomato Supply Chain

Some perspective on the food supply

- **Tracking and Tracing of food pathogens**
  - Almost 200,000 registered food facilities
    - 81,574 Domestic and 115,753 Foreign
  - More than 300 ports of entry
  - More than 130,000 importers and more than 11 million import lines/yr
  - In the US there are more than 2 million farms

"The Smokin' Hot Pepper"

FOOD SAFETY

Salmonella 'smoking gun' located

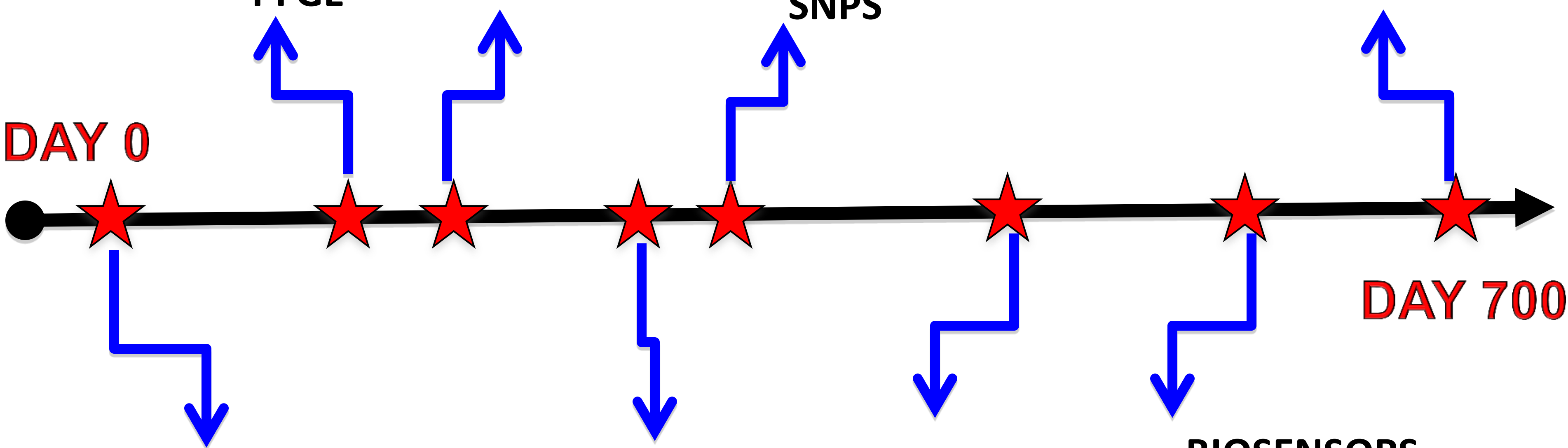
July 31, 2008 | Tiffany Hsu | Times Staff Writer

A strain of the salmonella bacteria that sickened more than 1,300 people has been found in a serrano pepper and a sample of irrigation water at a farm in Mexico, U.S. health officials said Wednesday.

They called the discovery a "breakthrough" but cautioned that tomatoes may still be a culprit in the nearly four-month outbreak that has alarmed consumers and cost the domestic produce industry hundreds of millions of dollars.

(Salmonella Saintpaul outbreak – Summer 2008)

HIGHLIGHTS IN SUBTYPING METHODS DEVELOPMENT IN THE FDA FOODS PROGRAM



DAY 0

CONCATENATED PFGE

MLSA

INTRAGENIC SNPS

WGS

DAY 700

SINGLE MARKER PFGE

HYPERVARIABLE SNPS

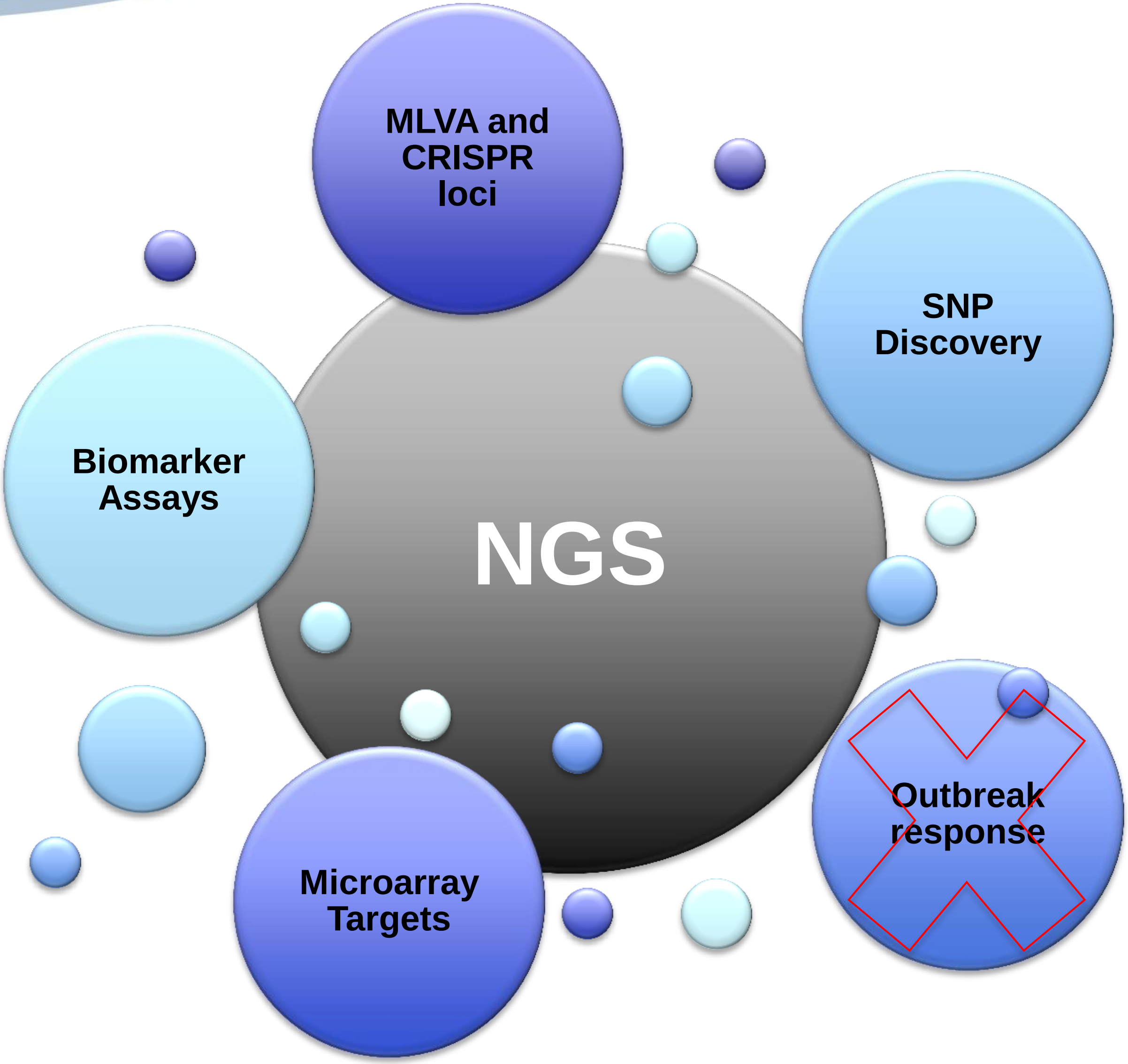
X-MAP

BIOSENSORS

©2011, Illumina Inc. All rights reserved.

Augmenting the Traceability of Foodborne Outbreaks and Other Important Food Safety Applications for WGS

Next-Generation Sequencing (NGS) provides support for other technologies and fosters novel targets and assay design for rapid diagnostics.



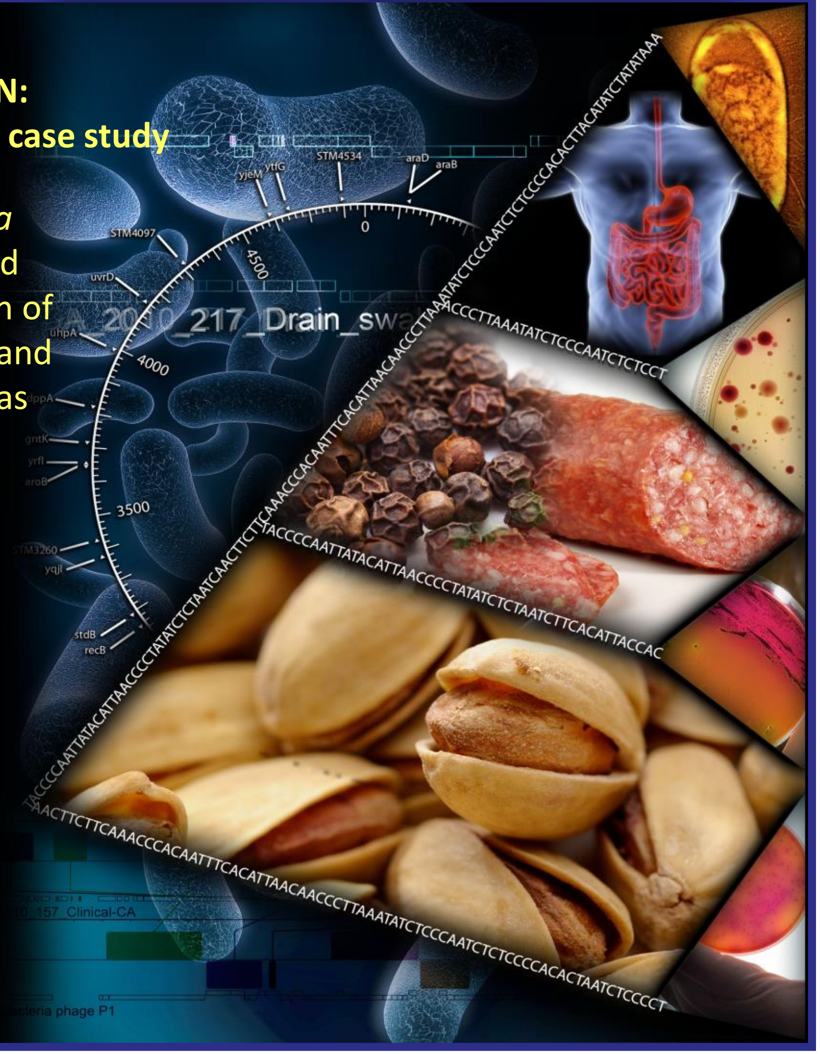
CASE 1: Same PFGE Patterns

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

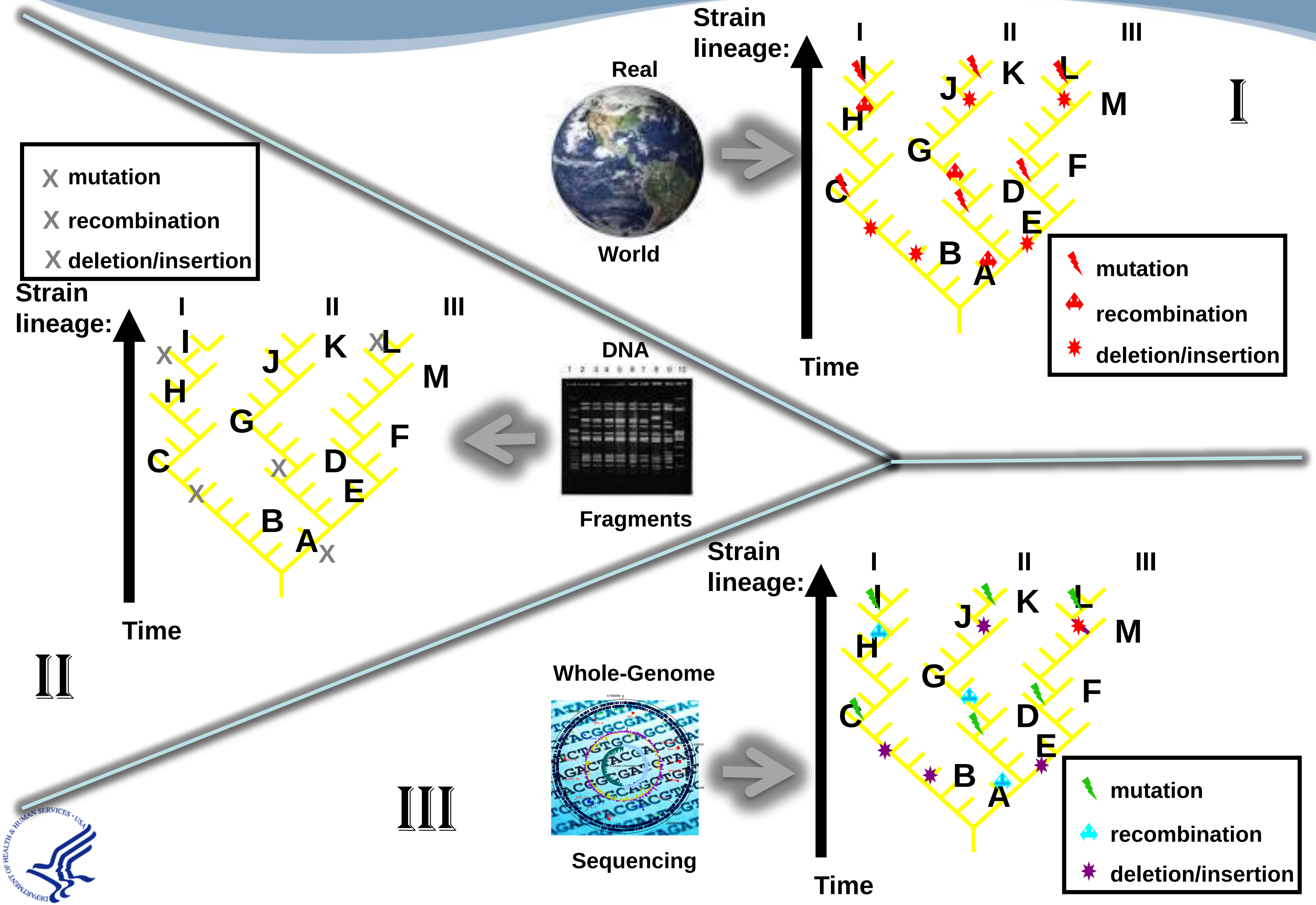


IN or OUT?

This from 1859, Darwin's, On the Origin of Species

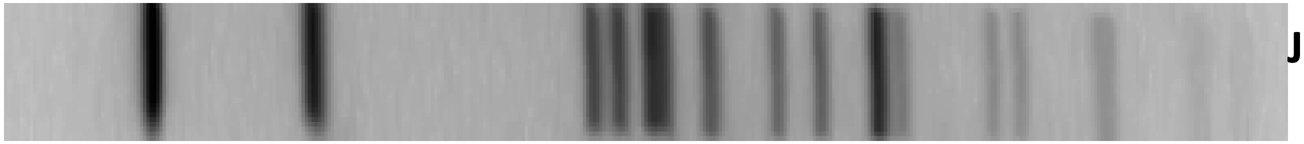
"It is obvious that the Galapagos Islands would be likely to receive colonists, whether by occasional means of transport or by formerly continuous land, from America; and the Cape de Verde Islands from Africa; and that such colonists would be liable to modification;— the principle of inheritance still betraying their original birthplace"

In WGS, we now have the potential to discern those birthplaces…


Real
World
Strain lineage:
I
II
III
I
H
C
B
A
J
G
K
L
M
F
D
E
Time
I
mutation
recombination
deletion/insertion
X mutation
X recombination
X deletion/insertion
DNA
1 2 3 4 5 6 7 8 9 10
Fragments
Strain lineage:
I
II
III
I
H
C
B
A
J
G
K
L
M
F
D
E
X
Time
II
Whole-Genome
Sequencing
III
Strain lineage:
I
II
III
I
H
C
B
A
J
G
K
L
M
F
D
E
Time
mutation
recombination
deletion/insertion

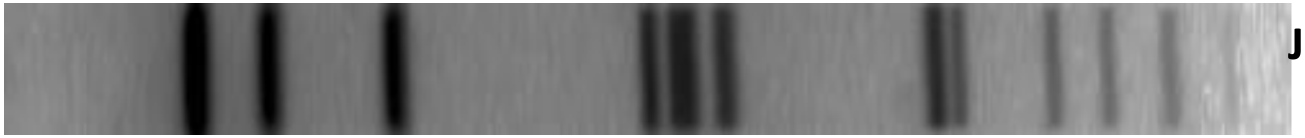
During the S. Montevideo outbreak, all isolates were indistinguishable by 1st, 2nd, and third enzyme PFGE.

PFGE-XbaI



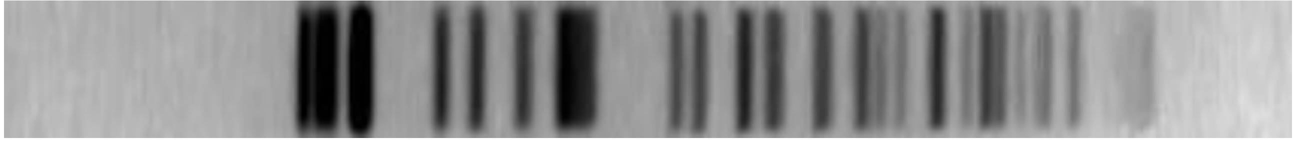
JIXX01.0011

PFGE-BlnI

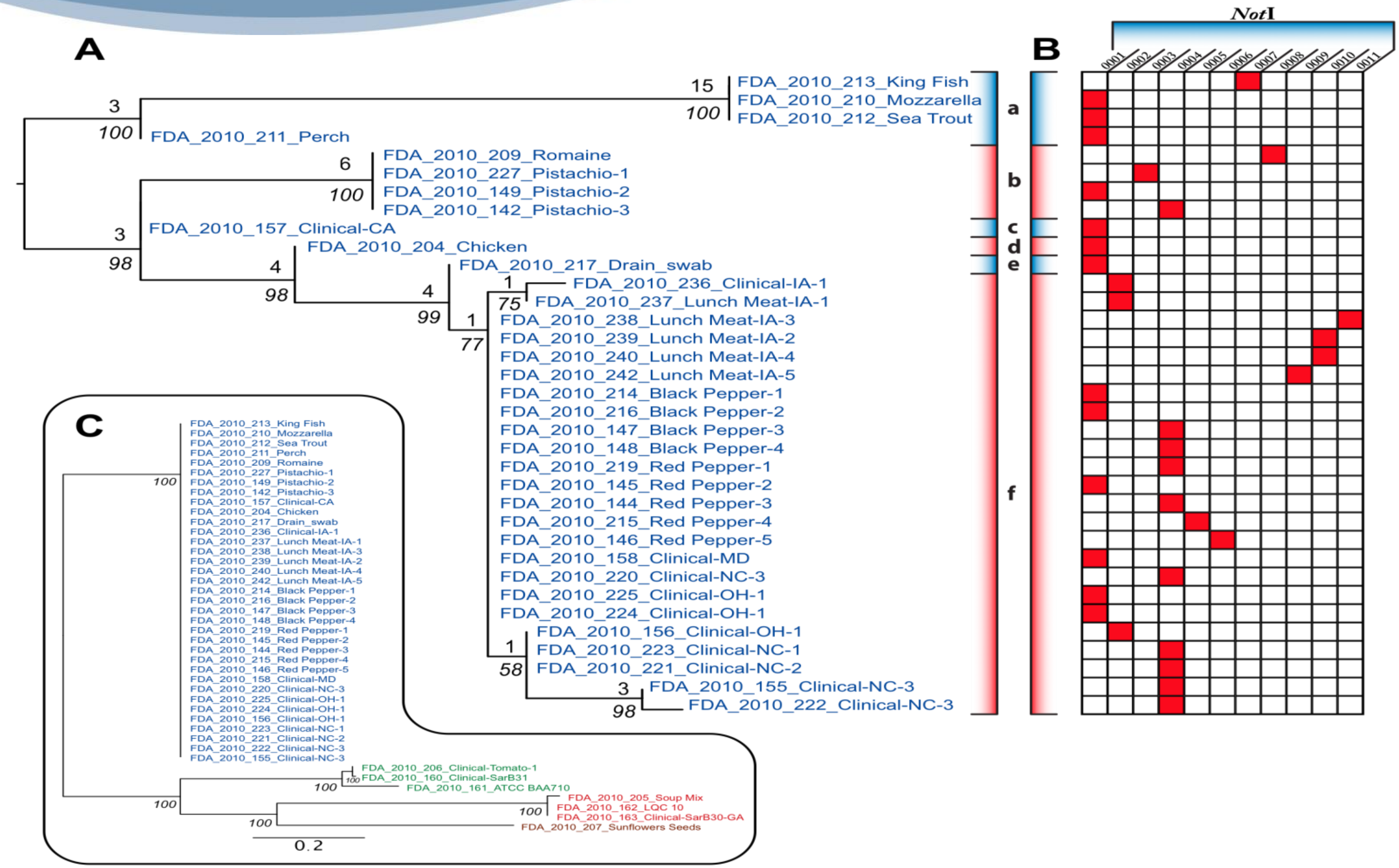


JIXA26.0012

PFGE-SpeI



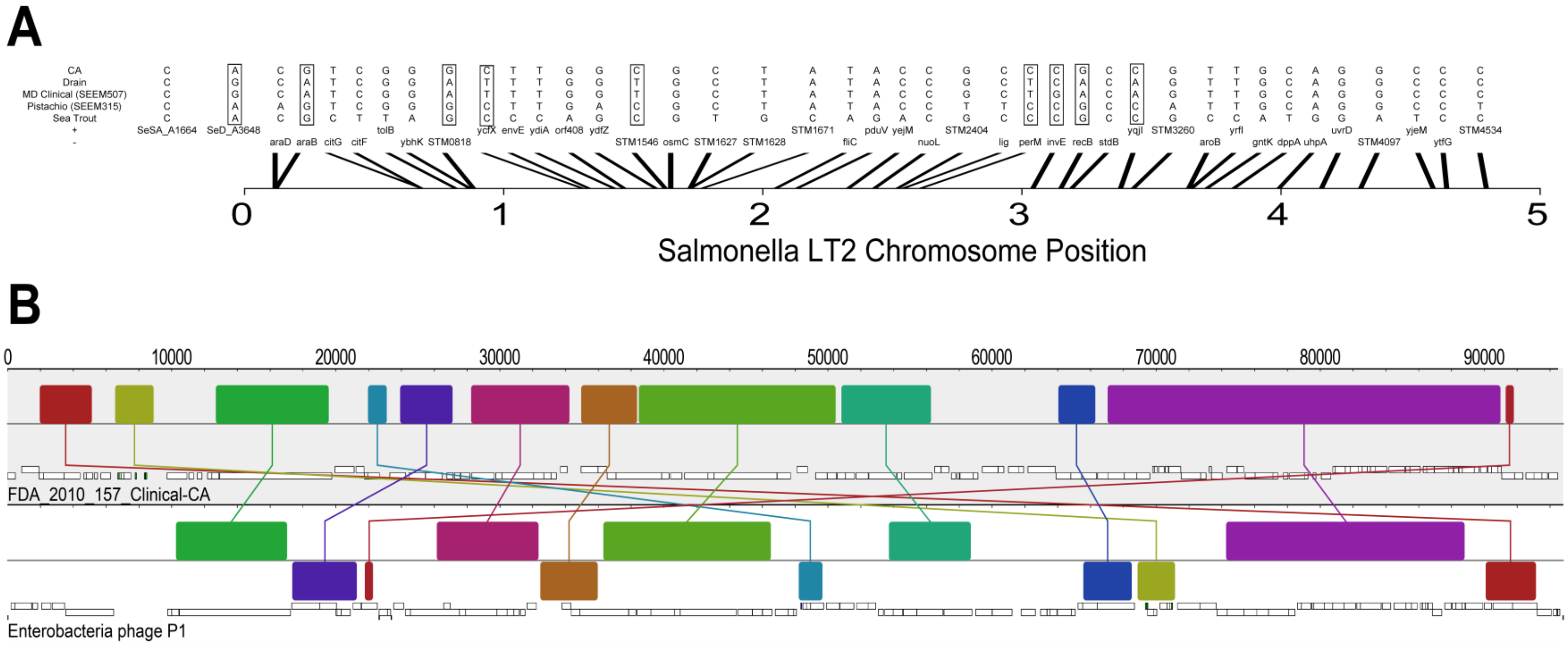
JIXS18.0001



D

	SARB30	SARB31	Sunflower	SeaTrout	Perch	Clinical-CA	Pistachio	Chicken	Drain	Outbreak
SARB30	-	191 (±15)	202 (±14)	248 (±18)	235 (±17)	242 (±17)	248 (±17)	244 (±17)	247 (±18)	249 (±18)
SARB31	24277 (±74)	-	190 (±12)	177 (±16)	164 (±14)	170 (±14)	176 (±15)	172 (±14)	175 (±15)	177 (±15)
Sunflower	22981 (±61)	24200 (±65)	-	255 (±22)	242 (±21)	248 (±20)	254 (±21)	250 (±20)	253 (±21)	255 (±21)
SeaTrout	25946 (±104)	21087 (±81)	25103 (±103)	-	13 (±4)	19 (±4)	25 (±4)	21 (±5)	24 (±4)	26 (±5)
Perch	25948 (±103)	21085 (±79)	25105 (±104)	47 (±6)	-	6 (±2)	12 (±2)	8 (±3)	11 (±3)	13 (±3)
Clinical-CA	25961 (±106)	21101 (±78)	25116 (±104)	61 (±7)	63 (±6)	-	6 (±2)	2 (±2)	5 (±2)	7 (±2)
Pistachio	25942 (±101)	21080 (±80)	25098 (±104)	42 (±5)	43 (±6)	40 (±5)	-	8 (±3)	11 (±3)	13 (±3)
Chicken	25940 (±102)	21079 (±79)	25097 (±106)	39 (±5)	40 (±5)	37 (±7)	18 (±4)	-	3 (±2)	5 (±2)
Drain	25944 (±102)	21083 (±80)	25100 (±105)	43 (±6)	44 (±6)	41 (±6)	22 (±4)	12 (±3)	-	2 (±1)
Outbreak	25960 (±103)	21098 (±79)	25115 (±106)	59 (±6)	60 (±6)	56 (±7)	37 (±5)	28 (±4)	18 (±2)	-

40 genes vary within an *S*. Montevideo outbreak while unique SNPs and a 100kb insertion separates a CA isolate from the outbreak



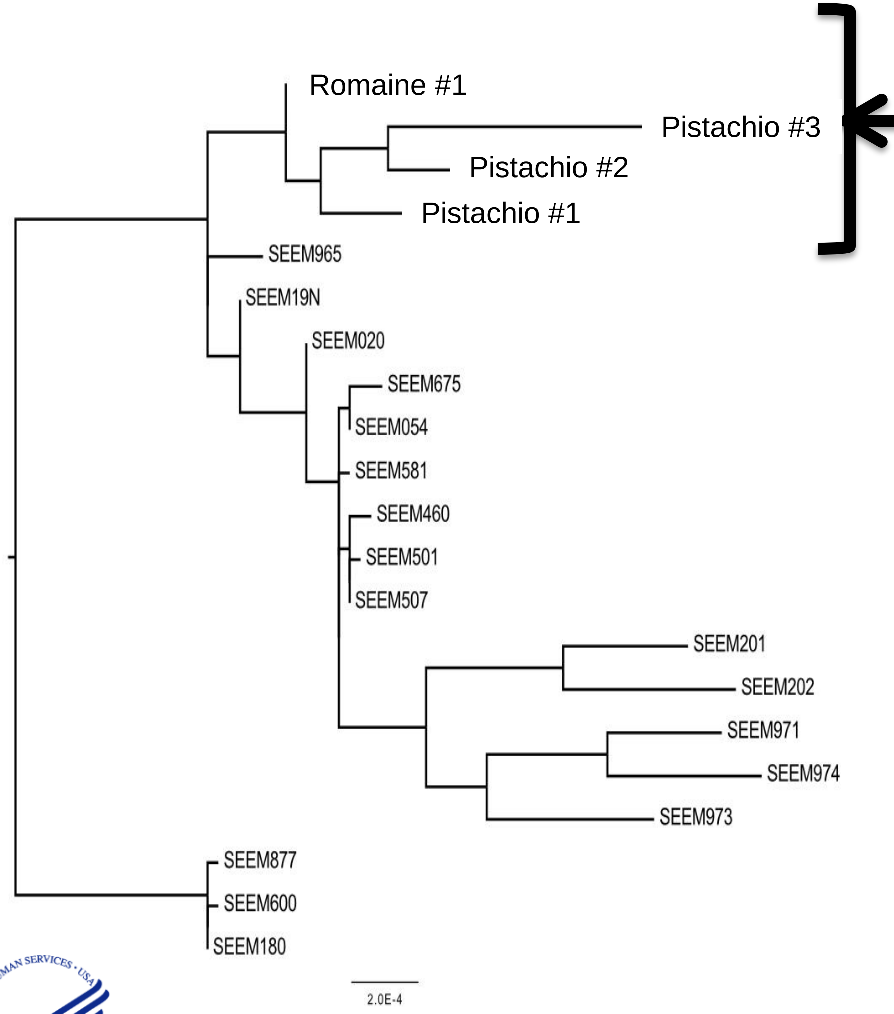
Next-Generation sequencing can be used to address different facets of outbreak response:

- **Have we seen this isolate before?**
  (Compare it to reference isolates)

- **Do these isolates form a cluster? (i.e. is it outbreak or background**
  (Compare them to reference and other outbreak isolates)

- **Is there a similarity between food/environmental and clinical isolates? (Compare them to reference, clinical, and food/environmental isolates)**

Salmonella reveals extensive phylogeographic structure



COMPLEX FOOD VEHICLES

The Well-Traveled Salad.

Do You Know Where Your Food Has Been?

As consumers, many of us fail to recognize that even our domestic and local food supplies are part of a global network. The daily activity of consuming food directly links our health as humans to the health of crops and produce, food animals, and the environments in which they are produced.

LETTUCE
Canada, Chile, Dominican Republic, Mexico, Peru, USA

CUCUMBERS
Canada, Honduras, India, Mexico, Spain, USA

FETA CHEESE
Canada, Denmark, Egypt, Germany, Greece, Israel, Italy, Turkey, UK, USA

VINAIGRETTE
Argentina, Brazil, Canada, Chile, China, France, Germany, Greece, India, Indonesia, Italy, Mexico, Morocco, Peru, Portugal, Spain, Thailand, Tunisia, Turkey, USA, Vietnam

OLIVES
Greece, Israel, Mexico, Spain, USA

SPROUTS
Argentina, Australia, Bangladesh, Canada, China, Egypt, France, India, Morocco, Nepal, Pakistan, South Africa, Spain, Turkey, USA

MANDARIN ORANGES
Israel, Mexico, Morocco, South Africa, Spain

ONIONS
Canada, China, Germany, India, USA

TOMATOES
Canada, Dominican Republic, Holland, Israel, Italy, Mexico, USA

CROUTONS
Argentina, Australia, Brazil, Canada, China, France, India, Mexico, Netherlands, Poland, Russia, Switzerland, Uruguay, USA, Vietnam

A "One Health" approach to food safety—bringing together expertise and resources from the clinical, veterinary, wildlife health, and ecology communities—has the potential to reveal the sources, pathways, and factors driving the outbreaks of foodborne illness and possibly prevent them from occurring in the first place.

***NOTE:** Countries are listed in alphabetical order and not by volume of export.*

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES
www.iom.edu

(Slide courtesy of J. Ostell, NCBI)

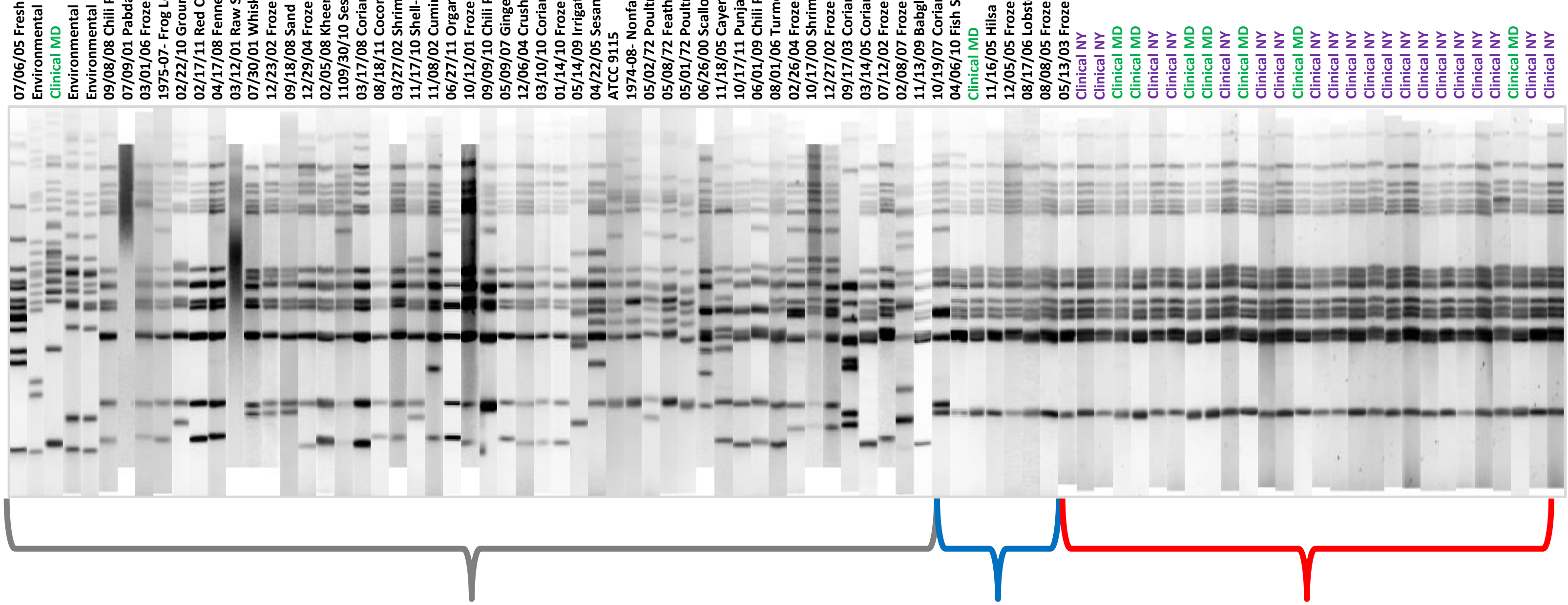
Linking Food Reservoirs Across the Globe?

Salmonella enterica serovar Bareilly

- CDC investigated a multistate (29 states) outbreak
- 410 confirmed cases between January 1st and July 7th, 2012
- Among the 326 case patient, 55 (17%) had been hospitalized
- Tuna was implicated as source of this outbreak
- At this time no reference genome was available at NCBI

S. Bareilly Outbreak April-June 2012

CASE #2: Global Source Tracking



Different PFGE than the outbreak pattern

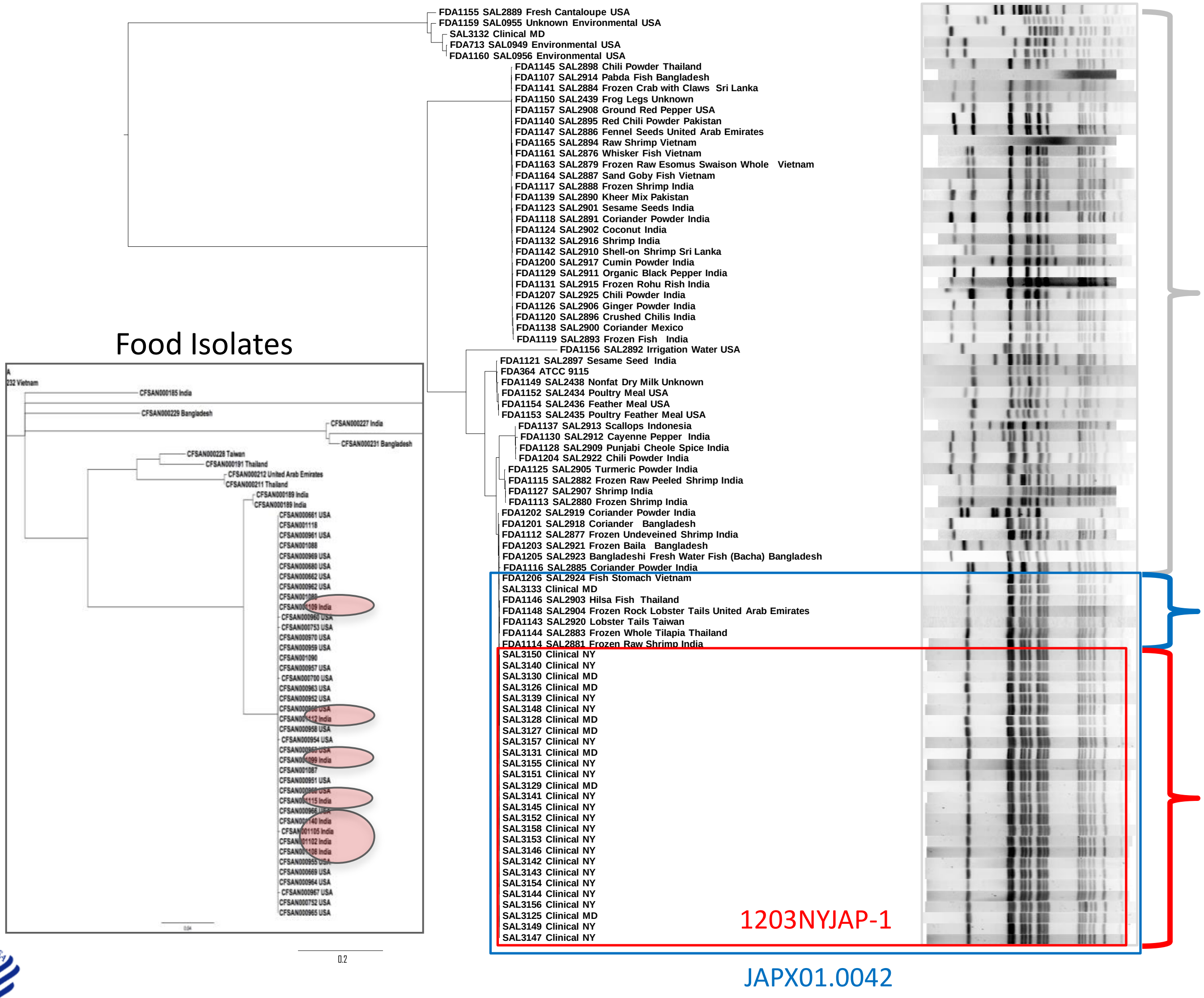
Same PFGE but not part of the outbreak

Outbreak Isolates

MD isolates – in green
NY isolates – in purple

WHERE DO YOU LIVE?

S. Bareilly Phylogeny & PFGE

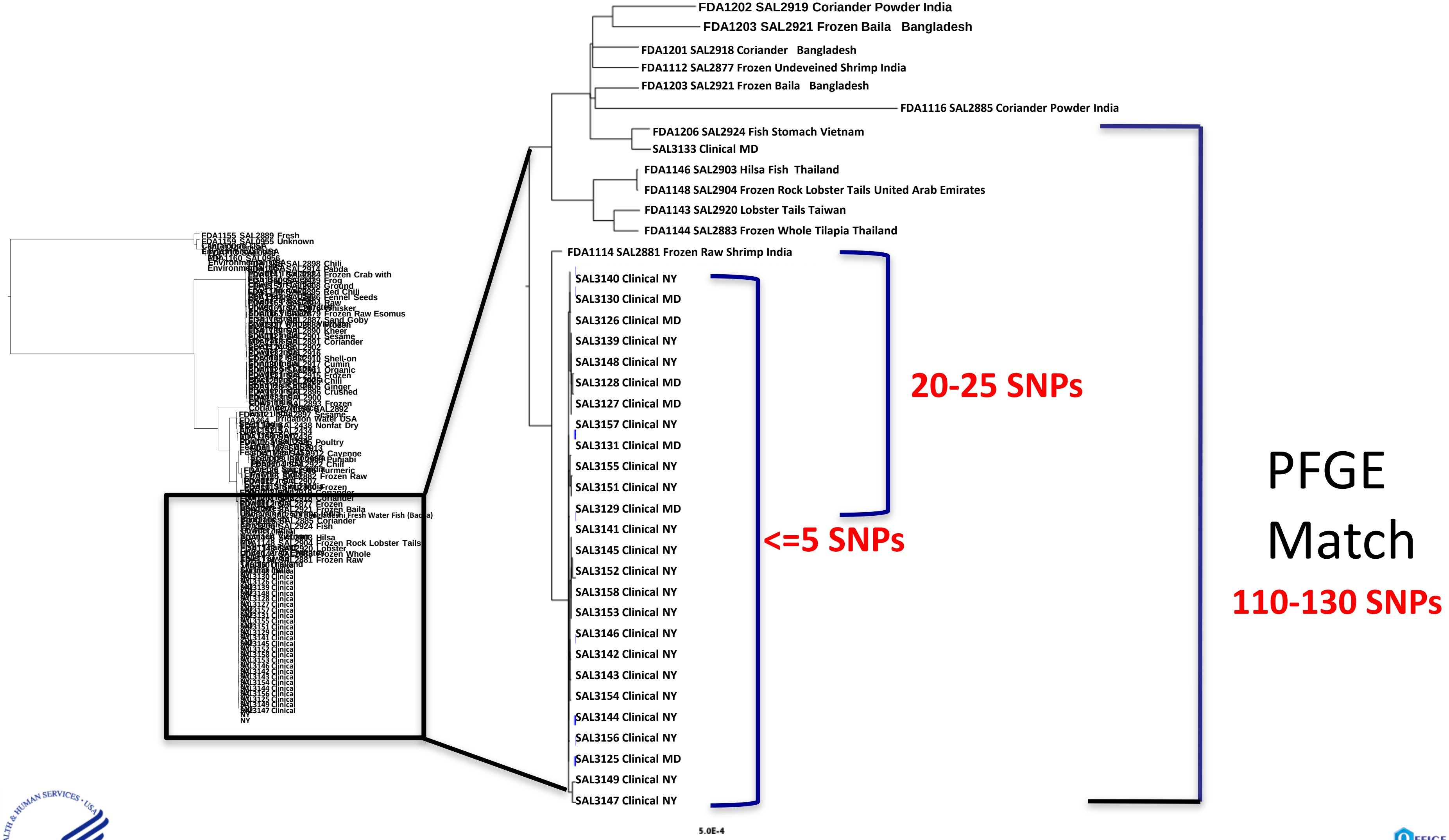


Different PFGE than the outbreak pattern

Same PFGE but not part of the outbreak

Outbreak Isolates

S. Bareilly Phylogeny



20-25 SNPs

<=5 SNPs

PFGE
Match

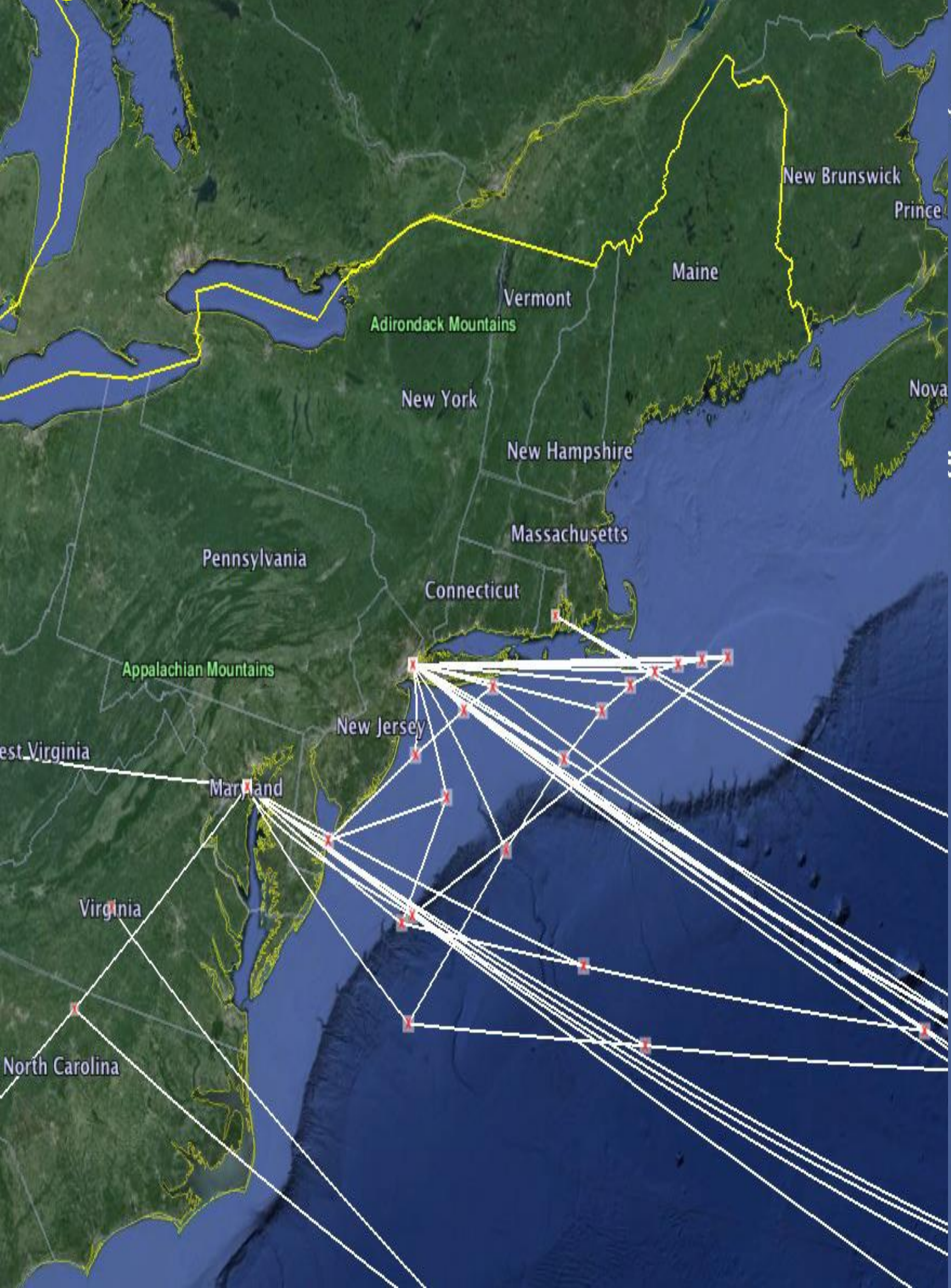
110-130 SNPs

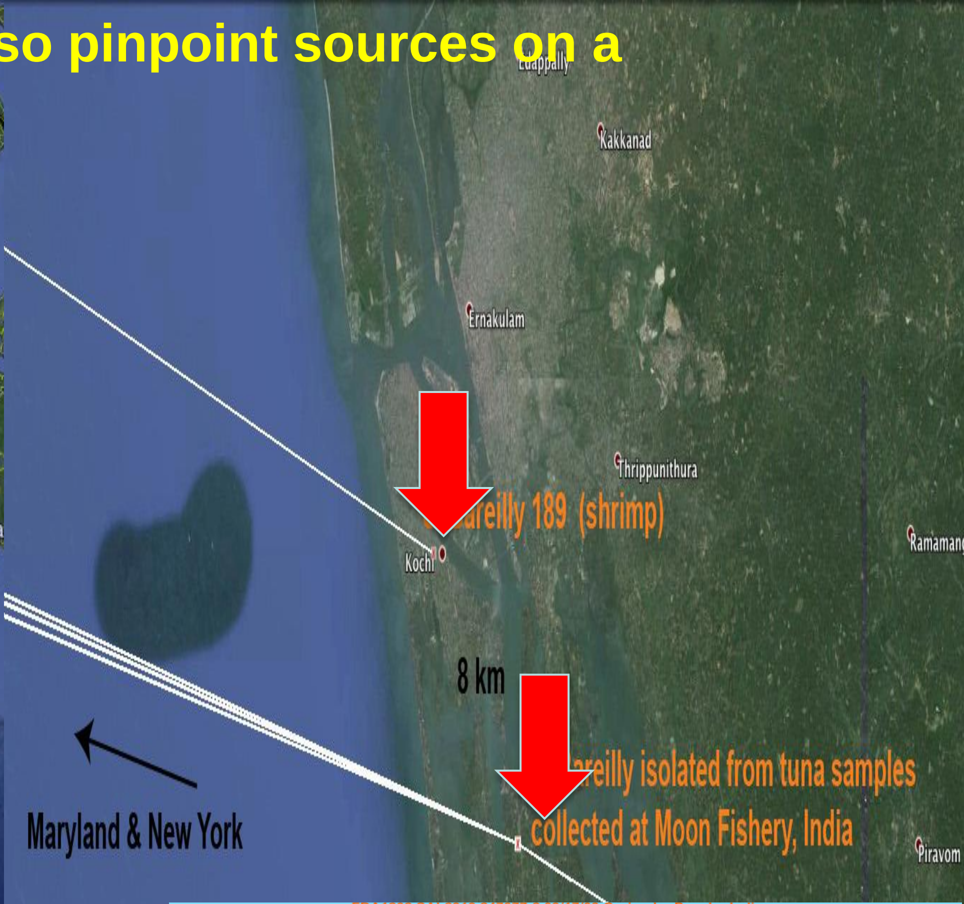
NGS distinguishes geographical structure among closely related *Salmonella* Bareilly strains

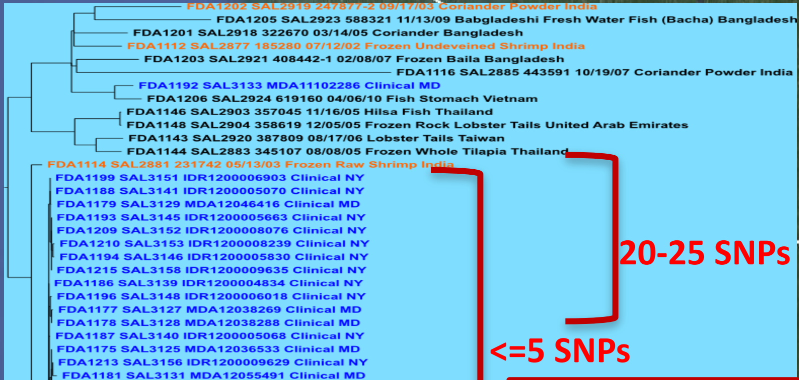
Transmission Networks

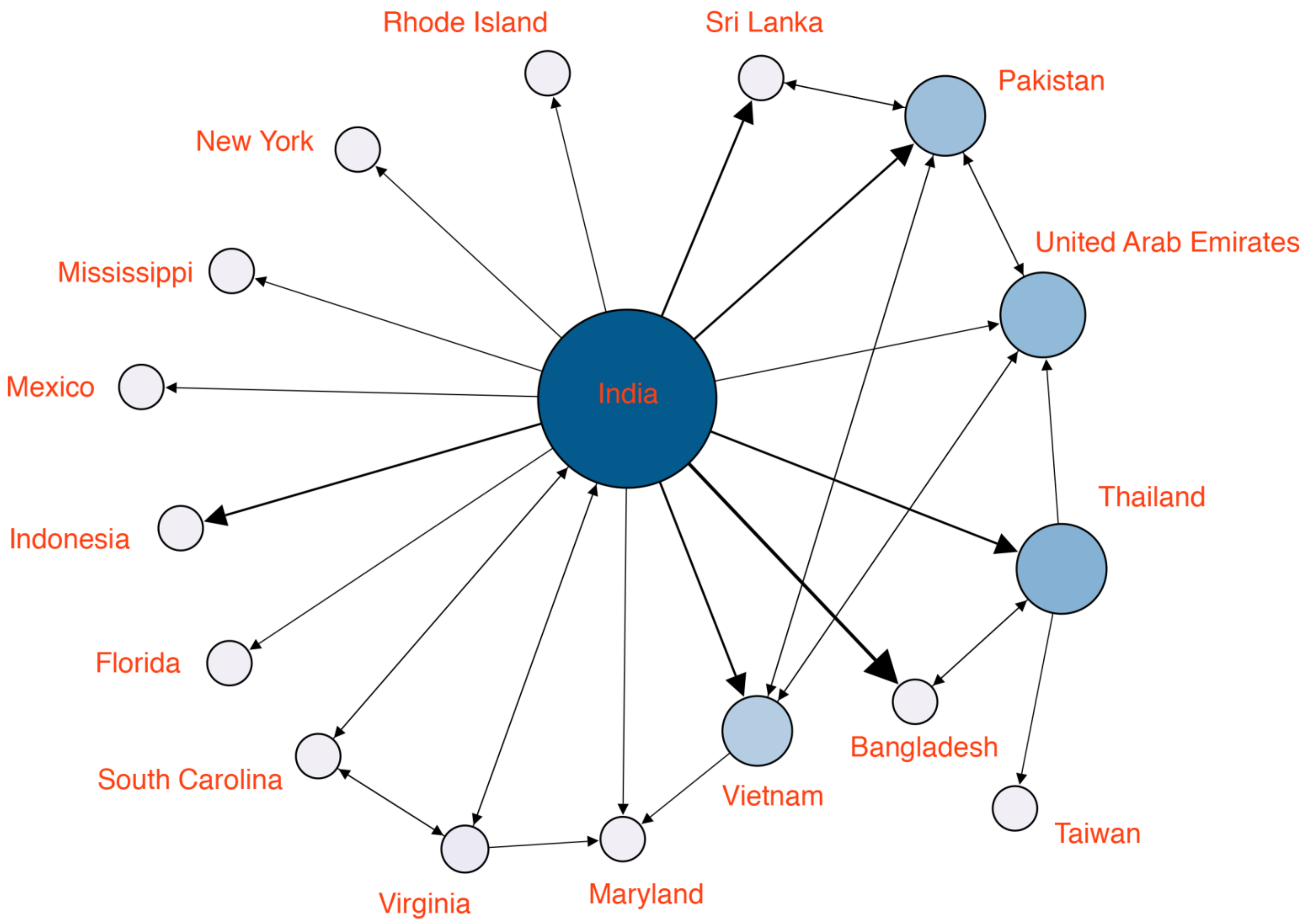
- Useful to identify geographic locations important for historical transmission and spread of *Salmonella*
- Enable food producers and public health officials to target interventions at places where they may have the most impact for improving food safety

WGS Cpmparative genomics can also pinpoint sources on a global scale.







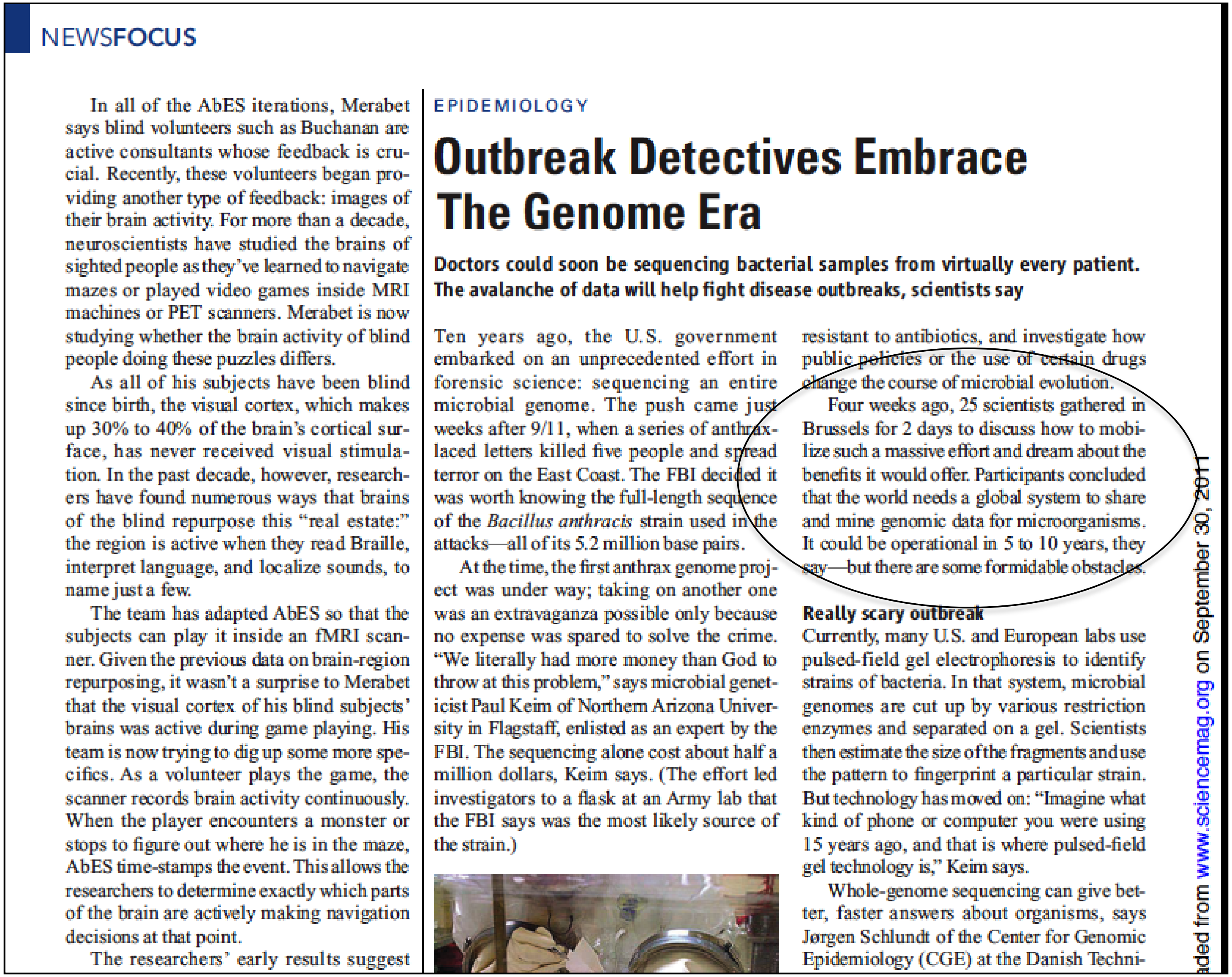

Rhode Island
Sri Lanka
Pakistan
New York
United Arab Emirates
Mississippi
Mexico
India
Thailand
Indonesia
Florida
Bangladesh
South Carolina
Vietnam
Taiwan
Virginia
Maryland

©2011, Illumina Inc. All rights reserved.

National and International WGS Initiatives

September 2011-Birth of "GMI"

The "Brussels Meeting"-Hosted by my friend and colleague, Dr. Jorgen Schlunt



NEWSFOCUS

In all of the AbES iterations, Merabet says blind volunteers such as Buchanan are active consultants whose feedback is crucial. Recently, these volunteers began providing another type of feedback: images of their brain activity. For more than a decade, neuroscientists have studied the brains of sighted people as they've learned to navigate mazes or played video games inside MRI machines or PET scanners. Merabet is now studying whether the brain activity of blind people doing these puzzles differs.

As all of his subjects have been blind since birth, the visual cortex, which makes up 30% to 40% of the brain's cortical surface, has never received visual stimulation. In the past decade, however, researchers have found numerous ways that brains of the blind repurpose this "real estate:" the region is active when they read Braille, interpret language, and localize sounds, to name just a few.

The team has adapted AbES so that the subjects can play it inside an fMRI scanner. Given the previous data on brain-region repurposing, it wasn't a surprise to Merabet that the visual cortex of his blind subjects' brains was active during game playing. His team is now trying to dig up some more specifics. As a volunteer plays the game, the scanner records brain activity continuously. When the player encounters a monster or stops to figure out where he is in the maze, AbES time-stamps the event. This allows the researchers to determine exactly which parts of the brain are actively making navigation decisions at that point.

The researchers' early results suggest

EPIDEMIOLOGY

Outbreak Detectives Embrace The Genome Era

Doctors could soon be sequencing bacterial samples from virtually every patient. The avalanche of data will help fight disease outbreaks, scientists say

Ten years ago, the U.S. government embarked on an unprecedented effort in forensic science: sequencing an entire microbial genome. The push came just weeks after 9/11, when a series of anthrax-laced letters killed five people and spread terror on the East Coast. The FBI decided it was worth knowing the full-length sequence of the *Bacillus anthracis* strain used in the attacks—all of its 5.2 million base pairs.

At the time, the first anthrax genome project was under way; taking on another one was an extravaganza possible only because no expense was spared to solve the crime. "We literally had more money than God to throw at this problem," says microbial geneticist Paul Keim of Northern Arizona University in Flagstaff, enlisted as an expert by the FBI. The sequencing alone cost about half a million dollars, Keim says. (The effort led investigators to a flask at an Army lab that the FBI says was the most likely source of the strain.)

resistant to antibiotics, and investigate how public policies or the use of certain drugs change the course of microbial evolution.

Four weeks ago, 25 scientists gathered in Brussels for 2 days to discuss how to mobilize such a massive effort and dream about the benefits it would offer. Participants concluded that the world needs a global system to share and mine genomic data for microorganisms. It could be operational in 5 to 10 years, they say—but there are some formidable obstacles.

Really scary outbreak

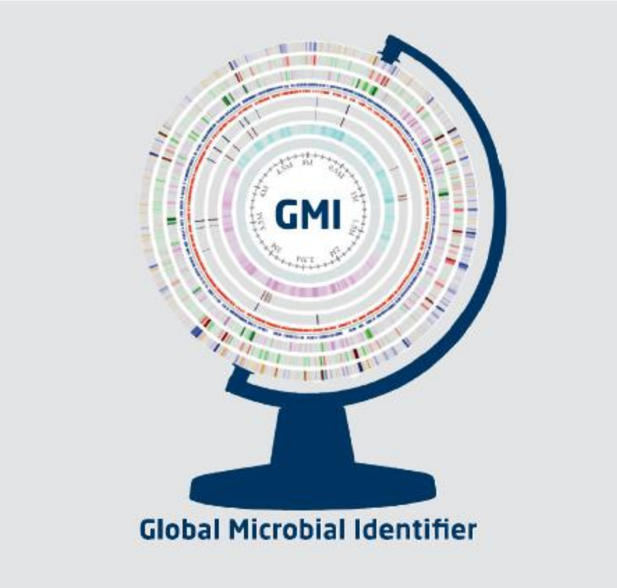
Currently, many U.S. and European labs use pulsed-field gel electrophoresis to identify strains of bacteria. In that system, microbial genomes are cut up by various restriction enzymes and separated on a gel. Scientists then estimate the size of the fragments and use the pattern to fingerprint a particular strain. But technology has moved on: "Imagine what kind of phone or computer you were using 15 years ago, and that is where pulsed-field gel technology is," Keim says.

Whole-genome sequencing can give better, faster answers about organisms, says Jørgen Schlundt of the Center for Genomic Epidemiology (CGE) at the Danish Techni-

Global Microbial Identifier

http://www.g-m-i.org/

- Make novel genomic technologies and informatics tools available for improved global patient diagnostics, surveillance, research and public health response. develop a global system to aggregate, share, mine and use microbiological genomic data to address global public health and clinical challenges, a high impact area in need of focused effort. 500 members in 30 countries

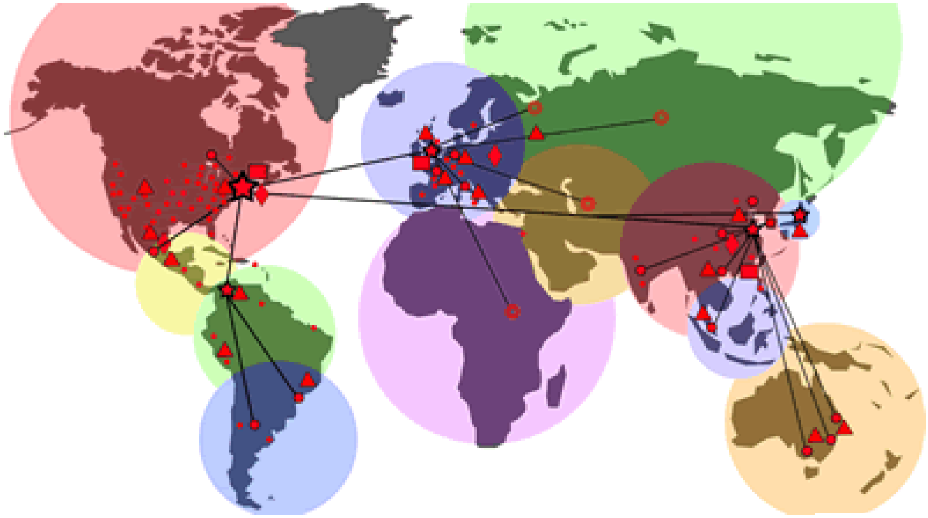


Work groups

1. Political challenges, outreach and building a global network
2. Repository and storage of sequence and meta-data
3. Analytical approaches
4. Ring trials and quality assurance
5. Pilot project

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



Can WGS function on a distributed level (ie, a WGS-based food shield)?

- **If yes, then...**
  - Initiate pilot study - GenomeTrakr
  - Develop collaborations and partnerships
    - NCBI, States, CDC and other Federal partners
  - What infrastructure would be needed?
  - Support multiple sequencing platforms?
    - Multiple data formats
    - How reproducible are the data AND answers?
  - How would data be accessed and stored?
    - Public vs. private **No data hoarding**
  - Metadata

Basic Data Flow for Global WGS Public Access Databases

DATA ACQUISITION

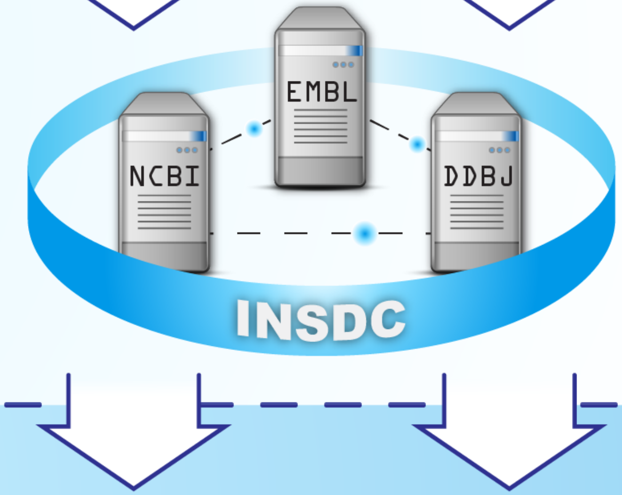
Sequence and upload genomic and geographic data

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*



PUBLIC HEALTH APPLICATION AND INTERPRETATION OF DATA

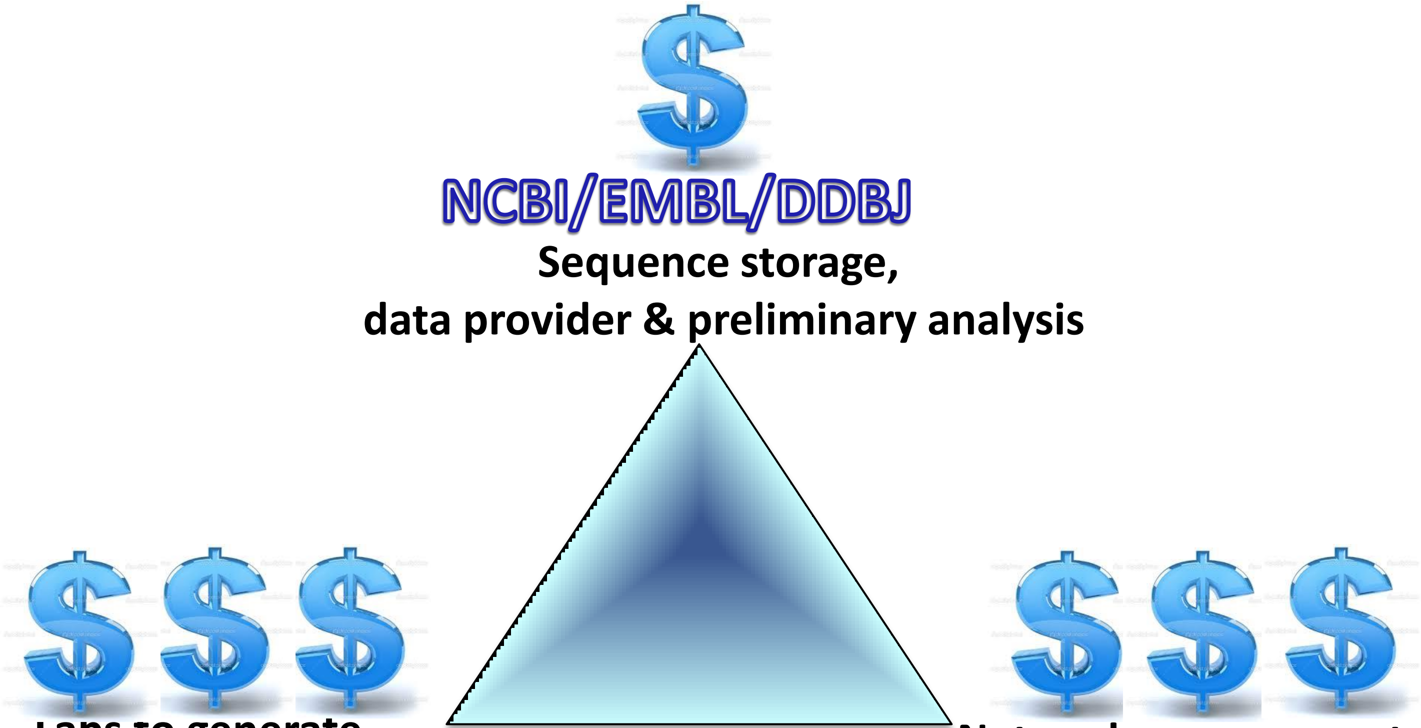
- *Find clinical links*
- *Identify clusters*
- *Conduct traceback*
- *Develop rapid methods*
- *Develop culture independent tests*
- *Develop new analytical software*

State, Local, Federal, and Foreign Public Health Agencies

Academia/Industry

11/2014

Key elements of a national/international WGS network



NCBI/EMBL/DDBJ

Sequence storage,
data provider & preliminary analysis

Labs to generate
WGS data

Network management

U.S. GenomeTrakr Labs

🟨 U.S. States with Federal, State, University, or Hospital Labs that are part of the GenomeTrakr Network

☆ NCBI Data Curation Site

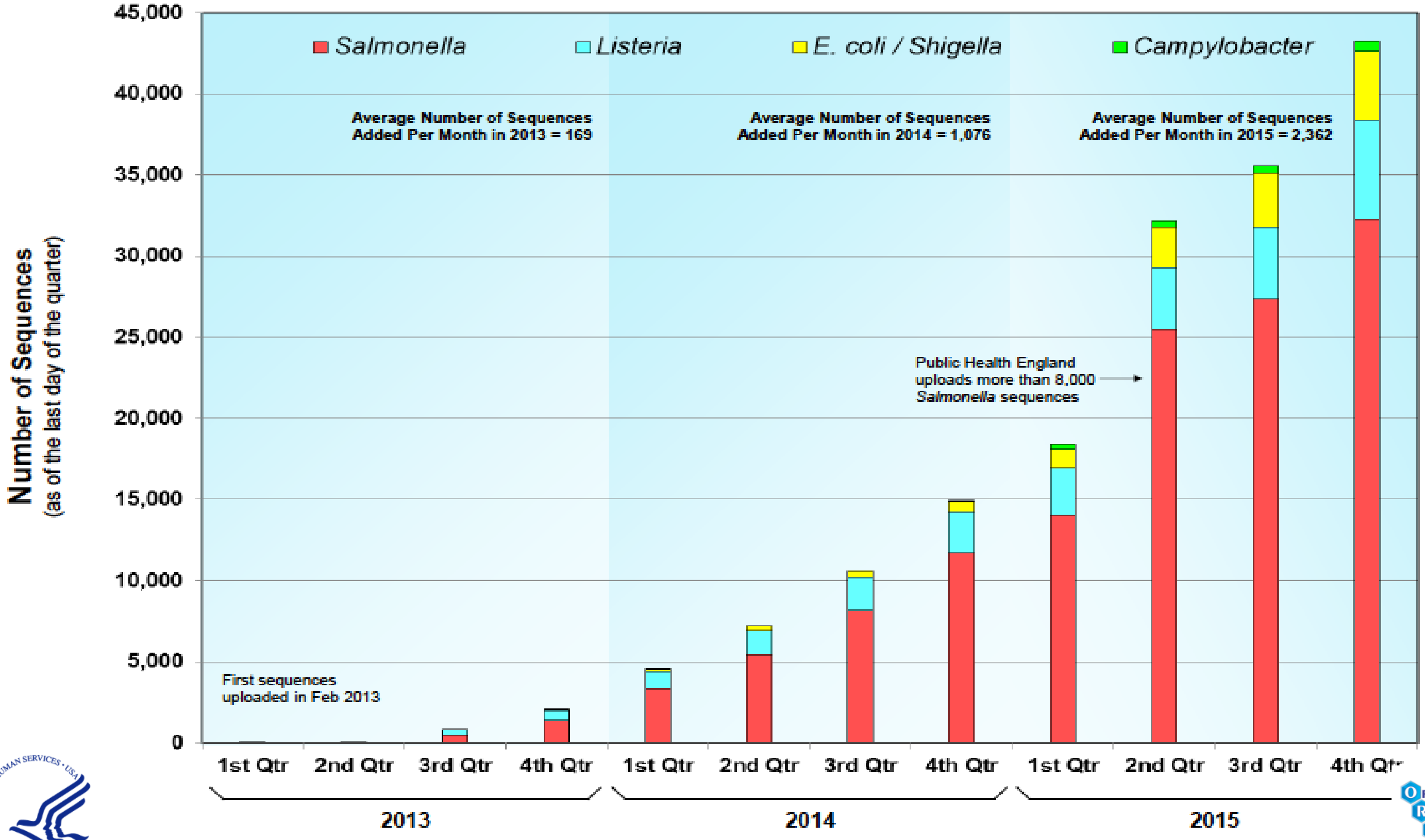
○ Lab Locations

January 2016 | Map is not to scale

Labs Outside the U.S. Contributing to GenomeTrakr

January 2016 | Map is not to scale

Total Number of Sequences in the GenomeTrakr Database



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

Public Health England
uploads more than 8,000
Salmonella sequences

First sequences
uploaded in Feb 2013

Current Status

- **NGS unequivocally supports foodborne contamination investigators – more than 55 different examples at FDA**
- **NGS network is reliable, efficient and can provide very good location/ingredient inferences – expanding to 17 states in 2017.**
- **At the WGS Laboratory in CFSAN, we have sequenced more than 15,000 *Salmonella*, more than 3,000 *Listeria*, and closed 130 genomes. Our current rate is about 600 draft sequences per month.**
- **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 – looking forward to expand into other pathogens like STEC (next joint RT effort) and *Salmonella* (#1 most devastating foodborne pathogen).**

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

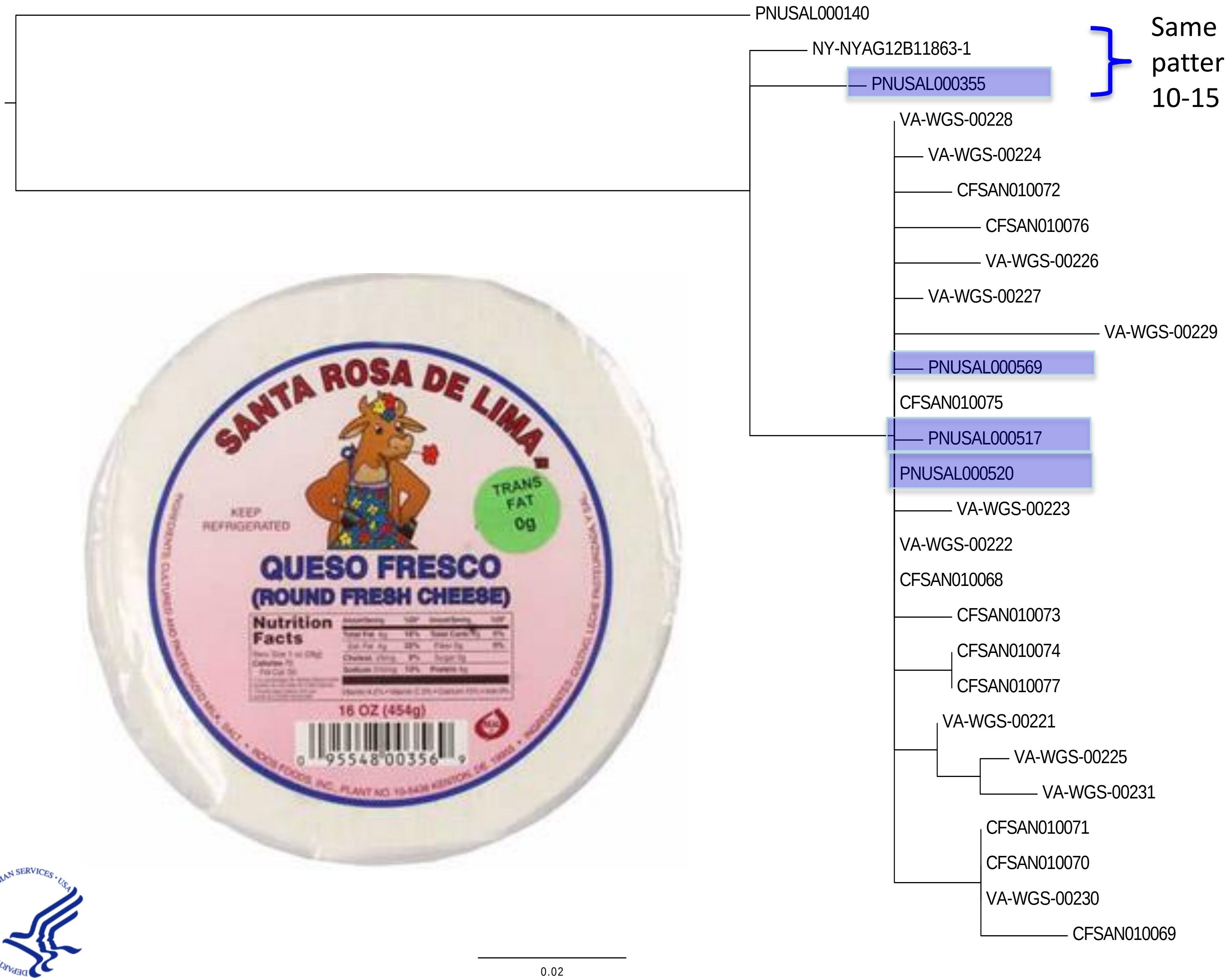
Cronobacter infant formula

V para oysters

EcO157:H7 lettuce

STEC beef

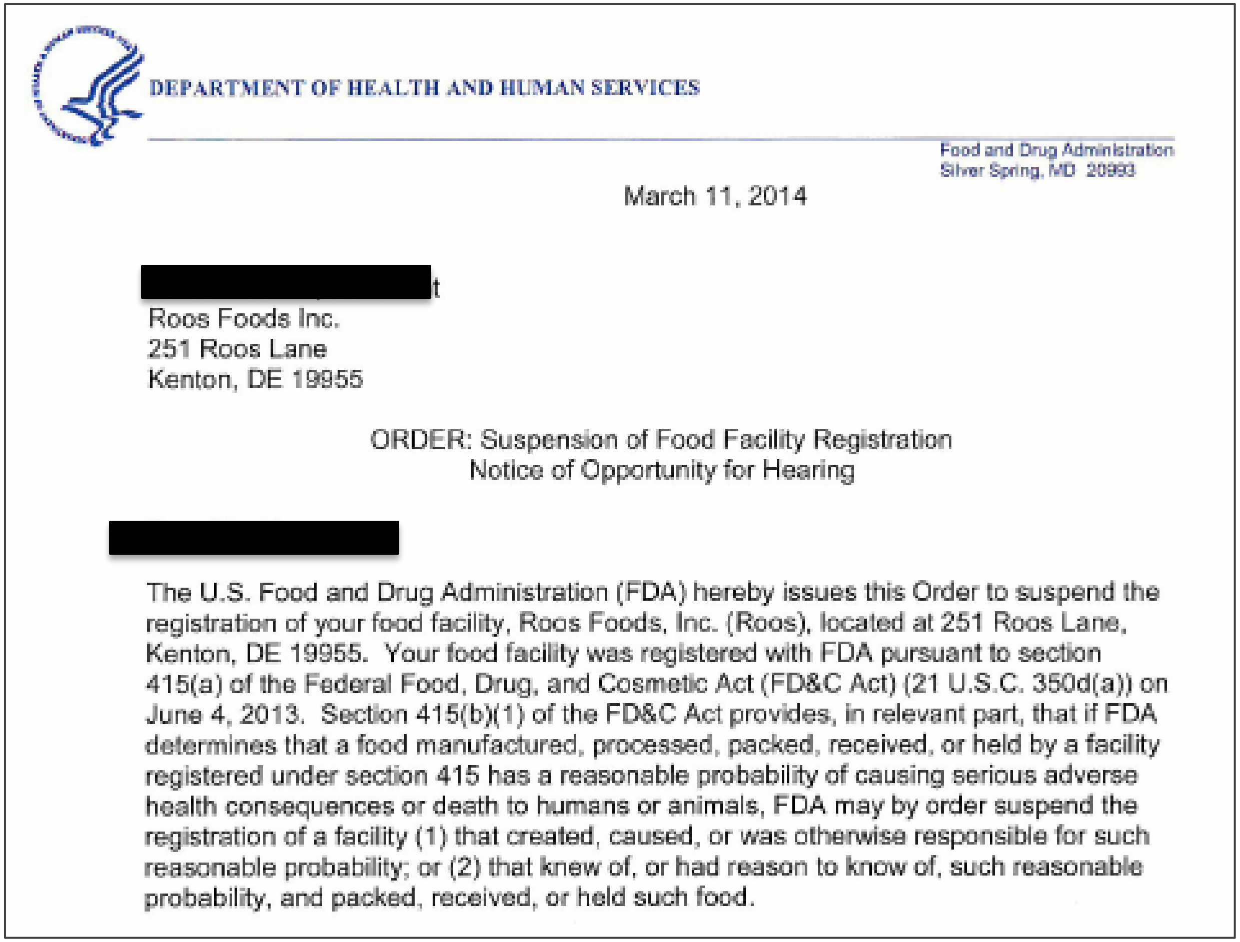
…Numerous other taxa



Same PFGE pattern
10-15 SNPs

2014
outbreak
5 SNPs

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

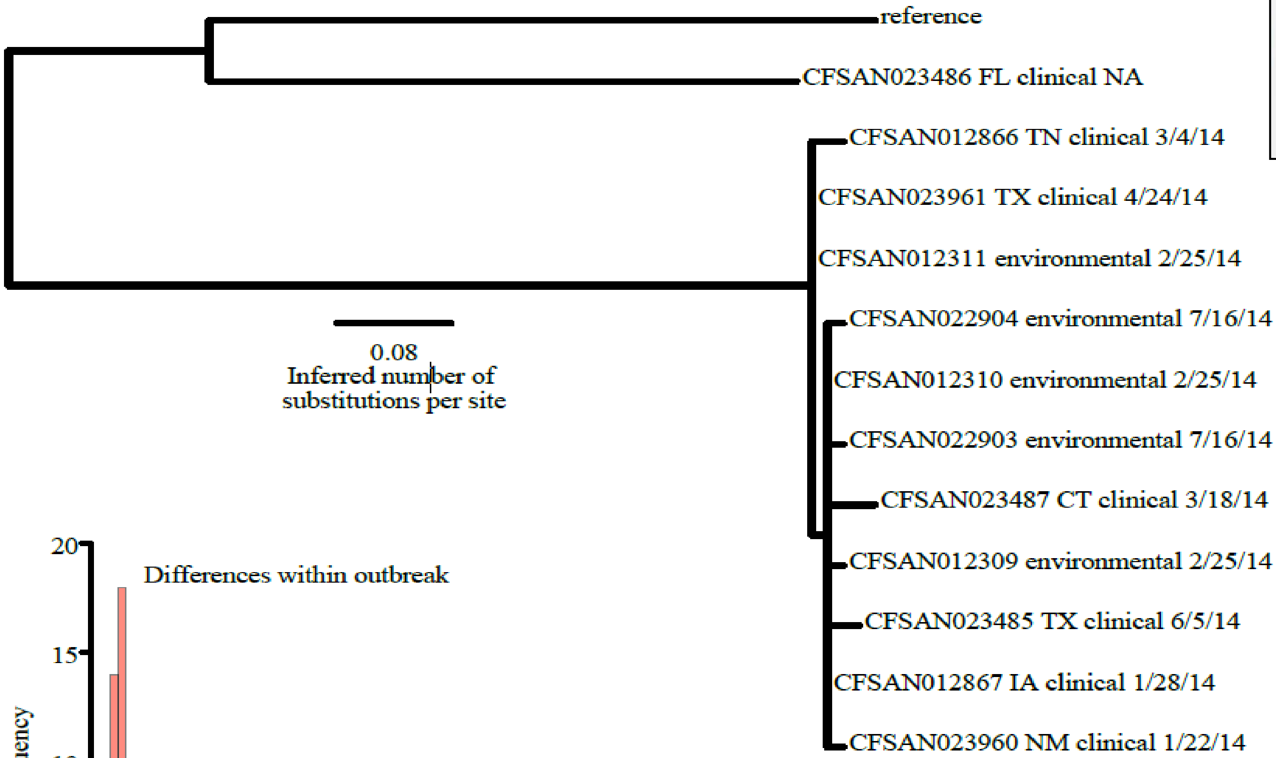
[redacted]
Roos Foods Inc.
251 Roos Lane
Kenton, DE 19955

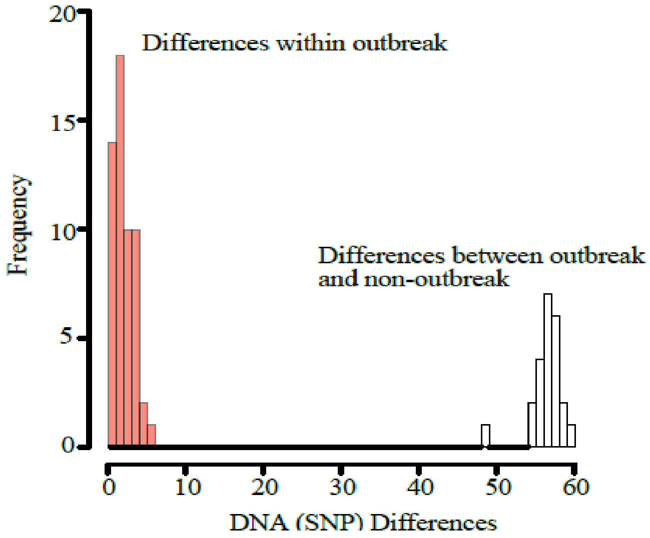
ORDER: Suspension of Food Facility Registration
Notice of Opportunity for Hearing

[redacted]

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






[redacted]

Photo courtesy of the MaraNatha website

Comparison of Nut Butter Outbreaks*

- *Salmonella* Tennessee, Con Agra, Peter Pan Peanut Butter,
  - 2006/2007: **715 cases, 129 hospitalizations, no deaths**
- *Salmonella* Typhimurium, multiple peanut products, Peanut Corporation of America,
  - 2008/2009: **714 cases, 166 hospitalizations, may have contributed to 9 deaths**

Post GenomeTrakr Network –Whole Genome Sequencing

- *Salmonella* Braenderup, nSpired Natural Foods, multiple almond and peanut butters, 2014: **6 cases, 1 hospitalization, no deaths**

* Source: CDC's Foodborne Outbreak Online Database (FOOD Tool)

"A COMPASS THAT POINTS TRUE NORTH and a TELESCOPE FOR THE DEEPEST OF SPACE"

S. Montevideo in salami (2009):

"Sorting through the ingredients"

S. Bareilly in tuna (2012):

"Linking up halfway across the world"

S. Tennessee in peanut butter paste (2007/2009) & S. Agona in dry cereal (1998-2008):

"Probing back in time"

S. Enteritidis in shell eggs (2010):

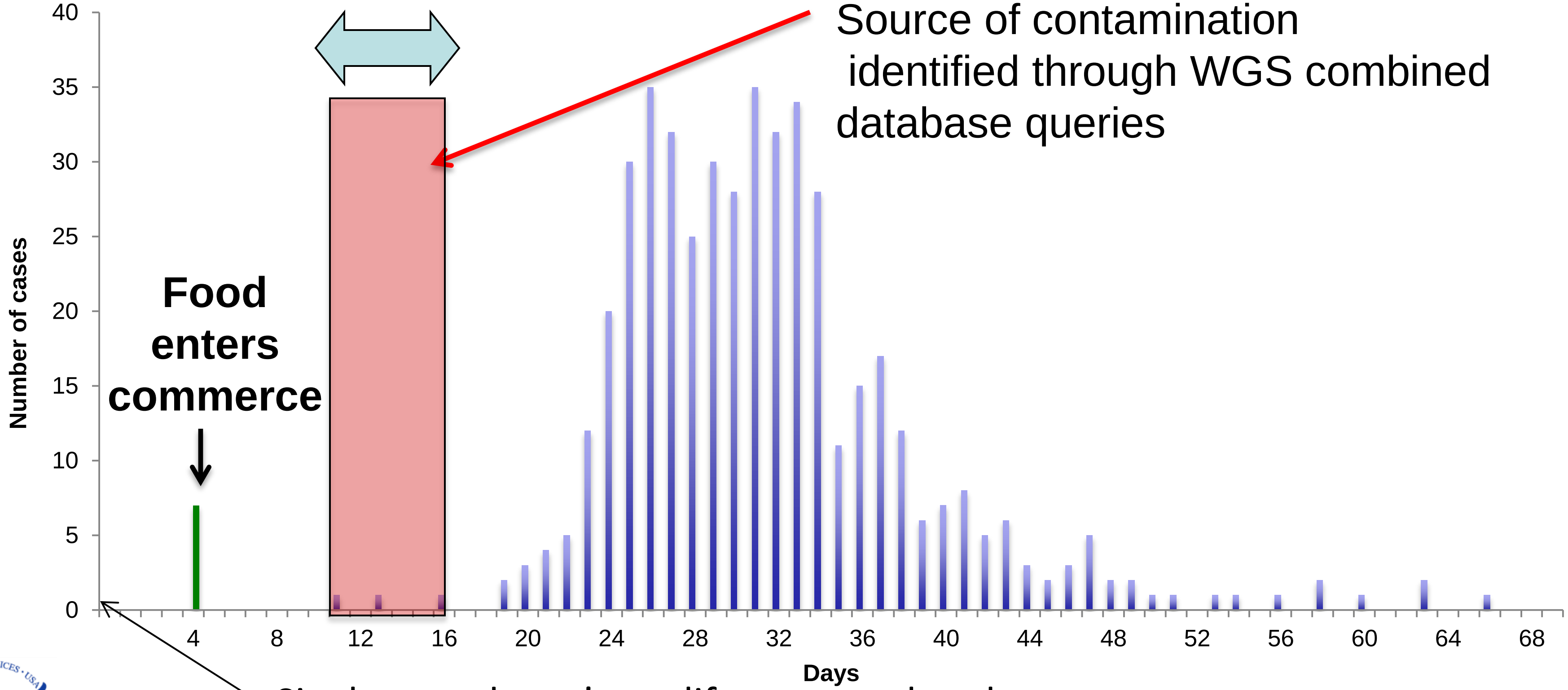
"Sourcing down to the farm"

S. Braenderup in nut butter (2014):

"Bypassing a Food Vehicle Altogether"

Epidemiology w/ WGS

clinical and environmental isolates
enter the database in real-time.
FDA, CDC, USDA, States

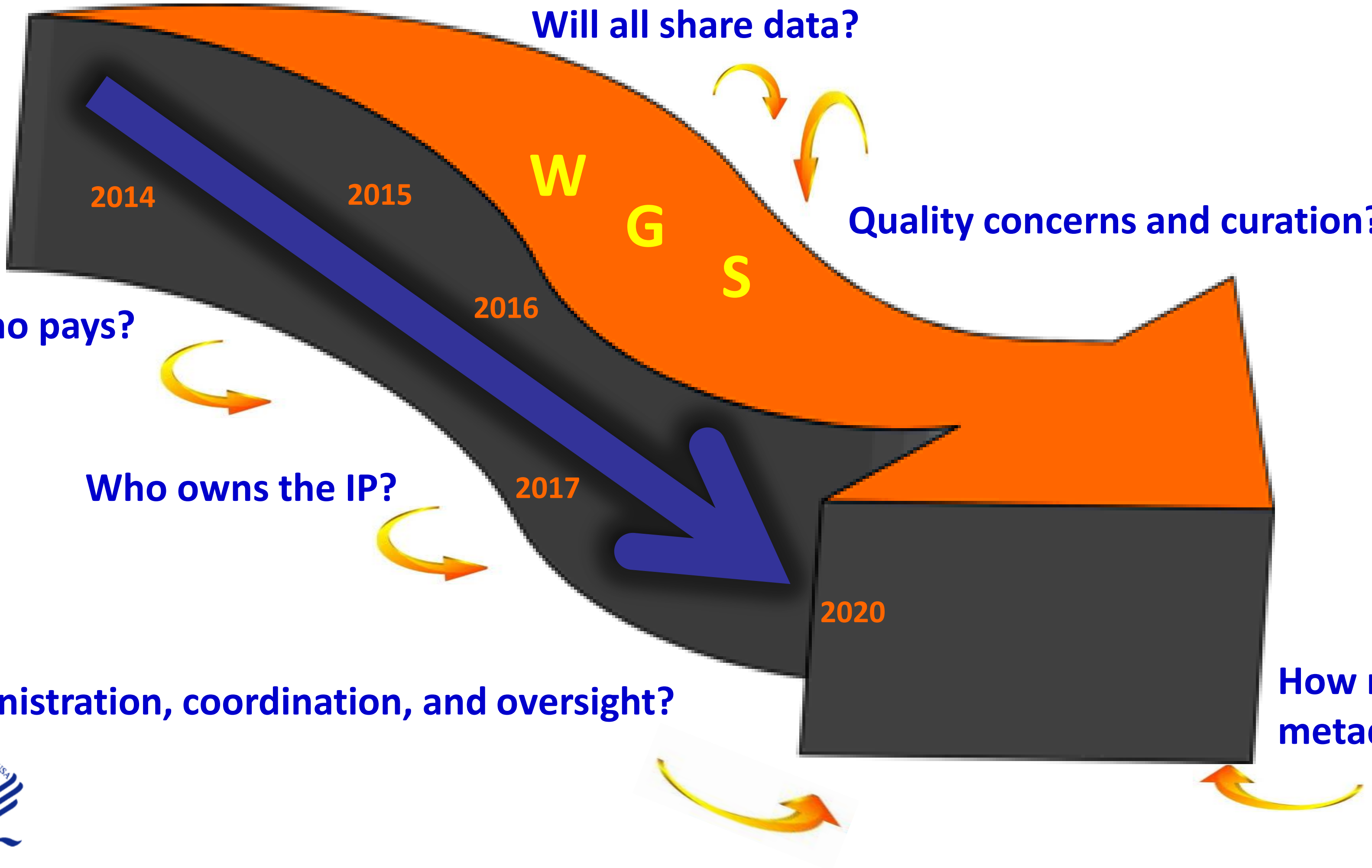


Food enters commerce

Source of contamination identified through WGS combined database queries

Single case doesn't qualify as an outbreak

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?

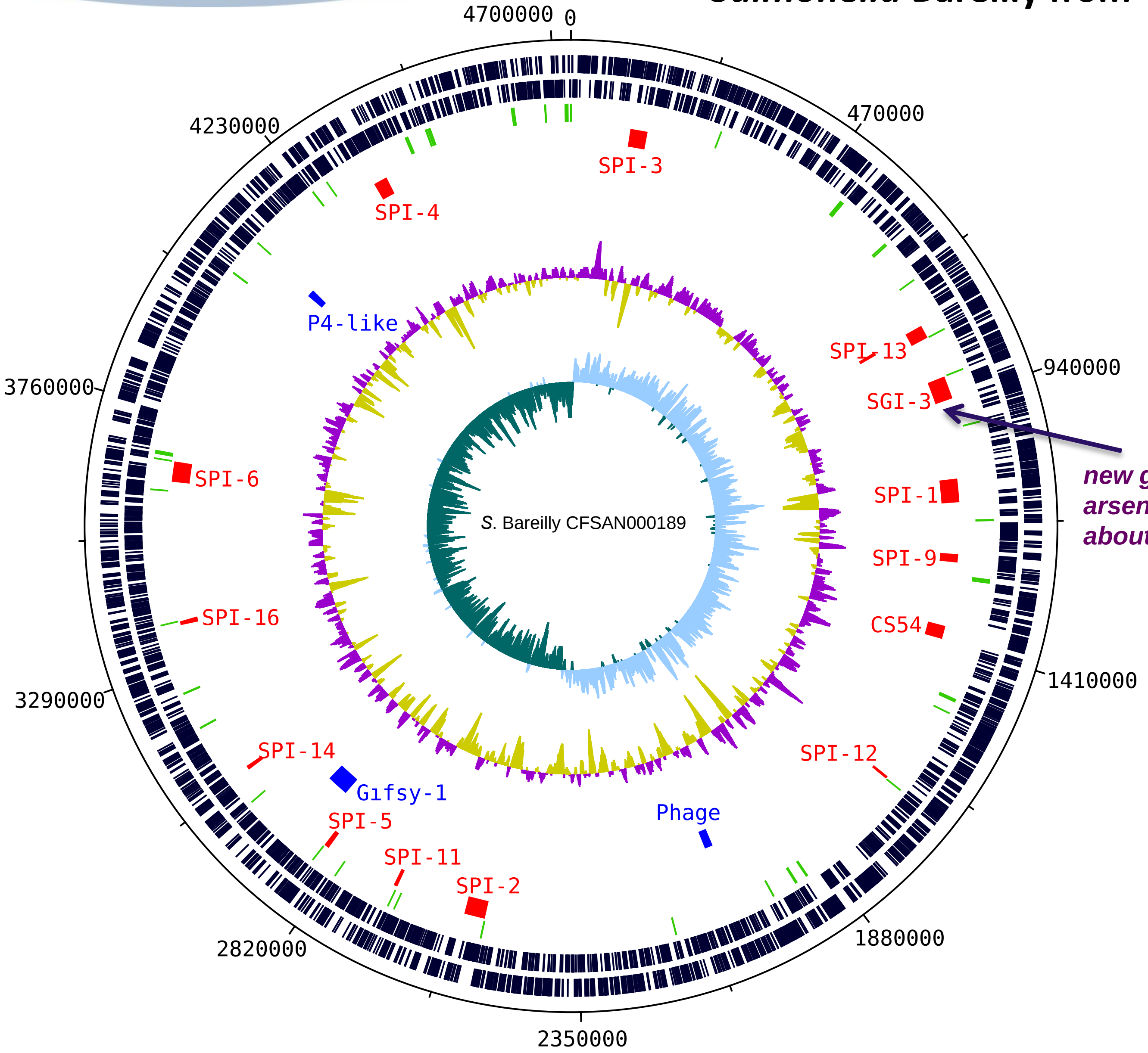
Who owns the IP?

Administration, coordination, and oversight?

How much metadata?

WGS and Beyond:, Epigenetics, Phenotypic Prediction, and Adaptation

Salmonella Bareilly from Tuna



new genomic island
arsenic resistance operon
about 40 kb

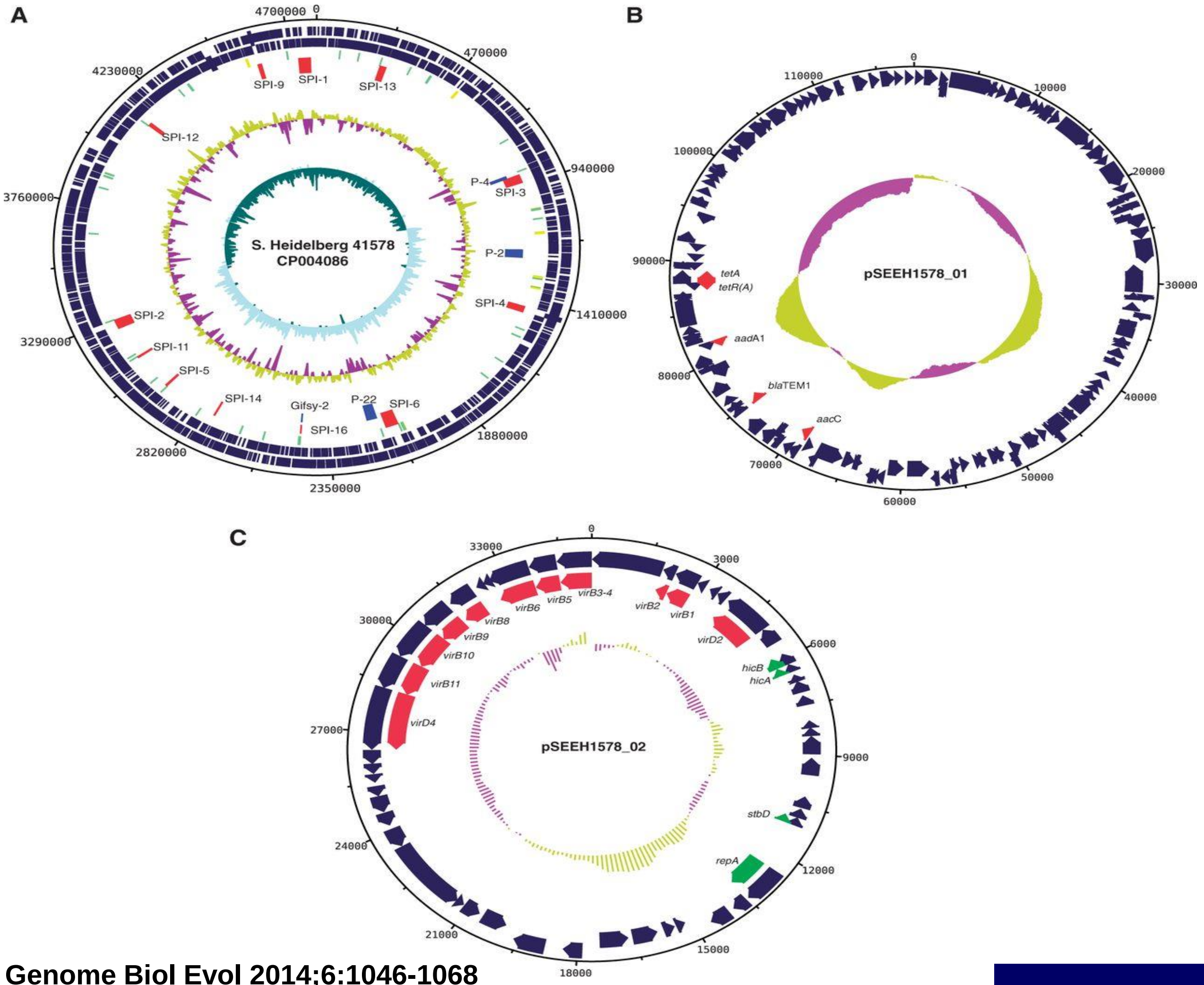
New Genomic Island

Name	Minimum	Maximum	Length	Direction	product
toxin CDS	43,465	43,788	324	forward	toxin
antitoxin CDS	43,124	43,444	321	forward	antitoxin
transcriptional regulator CDS	38,480	39,184	705	forward	transcriptional regulator
fatty acid transporter CDS	36,027	37,214	1,188	forward	fatty acid transporter
mechanosensitive ion channel protein MscS CDS	34,410	35,996	1,587	forward	mechanosensitive ion channel protein MscS
alkyl sulfatase CDS	32,360	34,267	1,908	forward	alkyl sulfatase
histidine phosphatase CDS	30,833	31,447	615	reverse	histidine phosphatase
transcriptional regulator CDS	28,266	28,532	267	reverse	transcriptional regulator
magnesium transporter CDS	24,110	24,766	657	forward	magnesium transporter
fimbrial usher protein CDS	20,181	22,691	2,511	reverse	fimbrial usher protein
fimbrial chaperone protein StdC CDS	19,404	20,129	726	reverse	fimbrial chaperone protein StdC
Positive regulator GrlA CDS	17,713	18,189	477	reverse	Positive regulator GrlA
transcriptional regulator CDS	16,860	17,720	861	reverse	transcriptional regulator
membrane protein CDS	14,535	15,275	741	reverse	membrane protein
ArsR family transcriptional regulator CDS	13,772	14,101	330	reverse	ArsR family transcriptional regulator
NADPH-dependent FMN reductase CDS	13,057	13,770	714	reverse	NADPH-dependent FMN reductase
RNA polymerase sigma 70 CDS	12,503	13,048	546	reverse	RNA polymerase sigma 70
arsenic resistance operon repressor CDS	12,065	12,427	363	reverse	arsenic resistance operon repressor
arsenic transporter ATPase CDS	10,287	12,044	1,758	reverse	arsenic transporter ATPase
ModE family transcriptional regulator CDS	9,845	10,213	369	forward	ModE family transcriptional regulator
arsenate reductase CDS	8,388	8,819	432	forward	arsenate reductase
arylsulfatase CDS	7,086	8,375	1,290	forward	arylsulfatase
arsenic transporter ATPase CDS	5,287	7,038	1,752	forward	arsenic transporter ATPase
arsenic resistance operon repressor CDS	4,907	5,269	363	forward	arsenic resistance operon repressor
arsenic resistance operon repressor CDS	4,506	4,859	354	forward	arsenic resistance operon repressor
nucleotidyltransferase CDS	2,538	4,313	1,776	forward	nucleotidyltransferase
integrase CDS	1	1,194	1,194	forward	integrase

Functional Assays for SNPs

Better understanding of adaptive change in *Salmonella* and *Lm* may provide more accurate risk assessment as well as enhanced preventive control measures on the farm and in the processing plant.

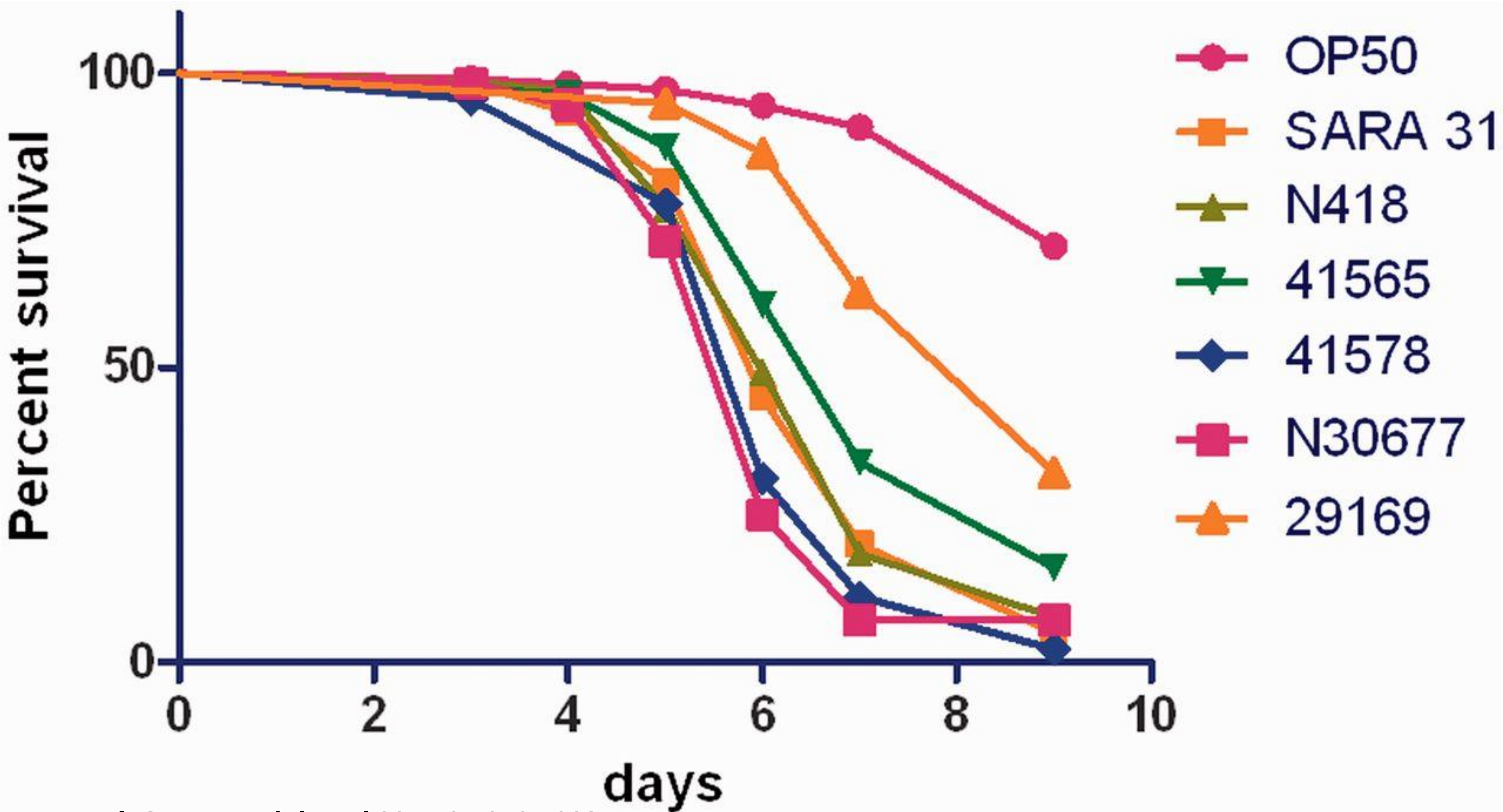
Chromosome and plasmids features of a clinical S. Heidelberg isolate 41578 as characterized by PAC-BIO reads.



Maria Hoffmann et al. Genome Biol Evol 2014;6:1046-1068

Genome Biology and Evolution

Caenorhabditis elegans survival data from six S. Heidelberg isolates shows decreased survival with presence of specific plasmids.



Maria Hoffmann et al. Genome Biol Evol 2014;6:1046-1068

Genome Biology and Evolution

Applications of WGS in the Food Safety Environment

- Delimiting scope and traceback of food contamination events (Track-N-Trace)
- Quality control for FDA testing and surveillance (Confidence in Regulatory Actions)
- Preventive control monitoring for compliance standards (the “repeat offender” project)
- ID, geno/pheno typing schemes (AST,Serotyping, VP) (CVM,CDRH,CFSAN) – risk assessment and adaptive change in *Salmonella* and *Listeria*

SeqSero

***Salmonella* Serotyping by Whole Genome Sequencing**

- **Reads (paired-end & interleaved)**
- **Reads (paired-end)**
- **Reads (single-end)**
- **Genome Assembly**

☐ ***The following formats are supported for raw reads input: .fastq.gz(preferred), .fastq and .sra.**

☐ **Please select your input file:**

☐ ***The following formats are supported for raw reads input: .fastq.gz(preferred), .fastq and .sra.**

☐ **Please select the first reads file:**

☐ **Please select the second reads file:**

☐ ***The following formats are supported for raw reads input: .fastq.gz(preferred), .fastq and .sra.**

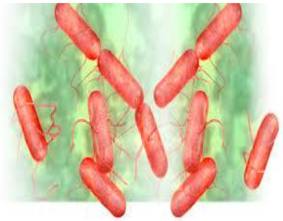
☐ **Please select your input file:**

☐ ***The FASTA format is supported for genome assembly input.**

☐ **Please select your input file:**

http://www.denglab.info/

Table 1. Accuracy of serotype predictions



Result	No. of genomes (% of total)		
	Reads mapping, CDC strains	Reads mapping, GenomeTrakr strains	Assembled genomes
Expected serotype[a]	304 (98.7)	3,061 (92.6)	324 (91.5)
Unexpected serotype	2 (0.65)	205 (6.2)[b]	11 (3.1)[b]
Partial or no serotype[c]	2 (0.65)	40 (1.2)	19 (5.4)
Total tested	308	3306	354

a The identification of the predicted serotype was considered correct when the serotype antigens detected corresponded to the antigens detected by conventional methods. See Text S1 in the supplemental material for a discussion of interpretation of serotype results. For GenomeTrakr and genome assembly datasets, serotype prediction in consensus with annotated serotype was considered correct.

b Numbers represent serotype predictions inconsistent with the annotated serotype; the accuracy of the annotated serotype is unknown.

c Some or all of the expected serotype determinants were not detected.

J Clin Microbiol. 2015 May;53(5):1685-92. doi: 10.1128/JCM.00323-15. Epub 2015 Mar 11.

Zhang S, Yin Y, Jones MB, Zhang Z, Deatherage Kaiser BL, Dinsmore BA, Fitzgerald C, Fields PI, Deng X.

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

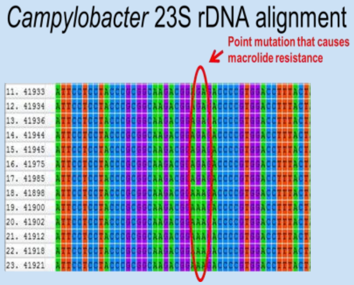
BLAST ®

Local Blast AR
Gene Database

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

Acquired AR genes

Sequences alignment
gyrA gene
23S rRNA gene



Point mutations related
to AR

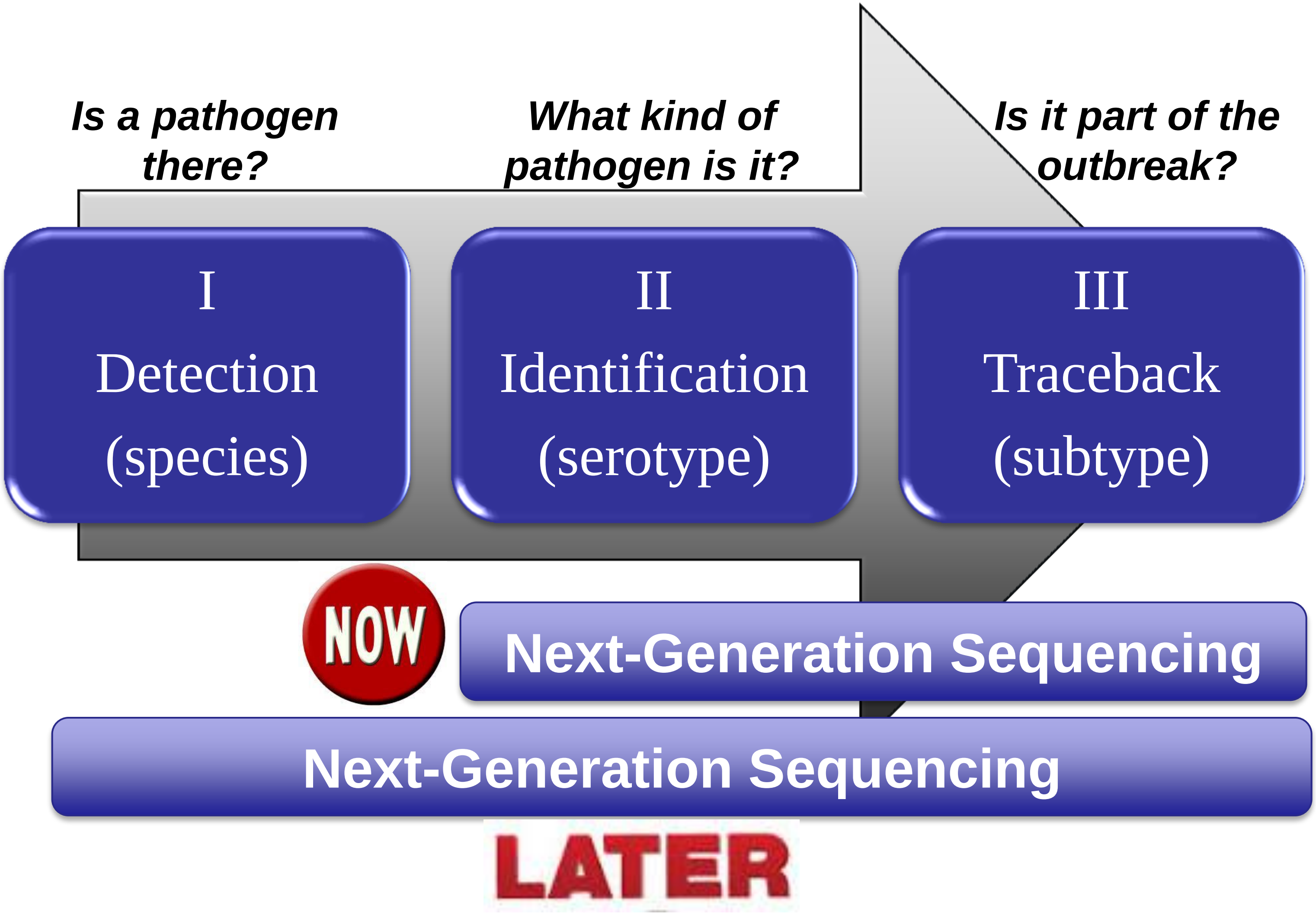
AR Genotype

Phenotype/Genotype Correlation

Antibiotic	Phenotype: resistant		Phenotype: susceptible		Sensitivity (%)	Specificity (%)
	Genotype: resistant	Genotype: susceptible	Genotype: resistant	Genotype: susceptible		
GEN	3	0	0	73	100	100
KAN	10	0	0	66	100	100
STR	52	0	12	12	100	50
AMC	26	0	0	50	100	100
AXO	23	1	0	52	95.8	100
FOX	24	0	2	50	100	96.2
TIO	22	1	1	52	95.7	98.1
COT	13	0	0	63	100	100
FIS	68	0	0	8	100	100
AZI	1	0	0	75	100	100
AMP	55	0	0	21	100	100
CHL	55	0	0	21	100	100
CIP	6	0	0	70	100	100
NAL	21	0	0	55	100	100
TET	67	0	0	9	100	100
Overall	446	2	15	677	99.6	97.8

J Antimicrob Chemother. 2015 Jul 3. pii: dkv186. [Epub ahead of print]
WGS accurately predicts antimicrobial resistance in Escherichia coli.
Tyson GH1, McDermott PF1, Li C1, Chen Y1, Tadesse DA1, Mukherjee S1, Bodeis-Jones S1, Kabera C1, Gaines SA1, Loneragan GH2, Edrington TS3, Torrence M4, Harhay DM5, Zhao S6.

Investigating Food Contamination Events with OMICS Approaches



Is a pathogen there?

What kind of pathogen is it?

Is it part of the outbreak?

I
Detection
(species)

II
Identification
(serotype)

III
Traceback
(subtype)

NOW

Next-Generation Sequencing

Next-Generation Sequencing

LATER

IN THE NOT SO DISTANT FUTURE…..

Species

Resistance

Virulence

Subtype

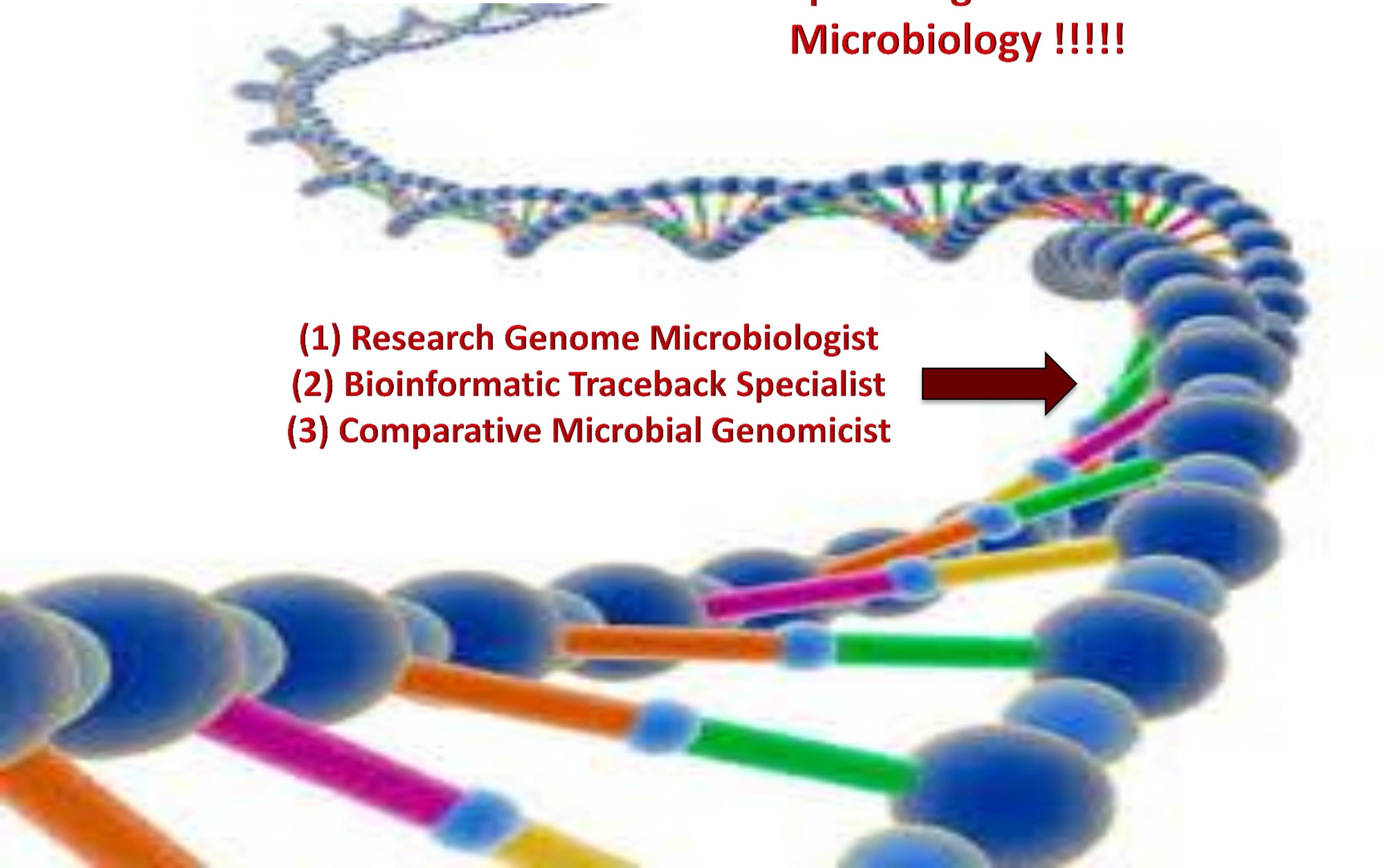
Serotype

Adaptations

ONE MICROBIOLOGICAL WORKFLOW: ONE MICROBIOLOGICAL TOOL BOX

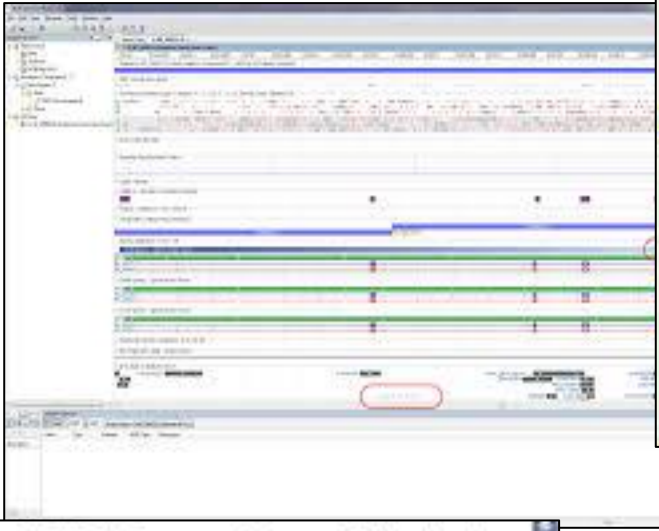
All AT YOUR FINGERTIPS

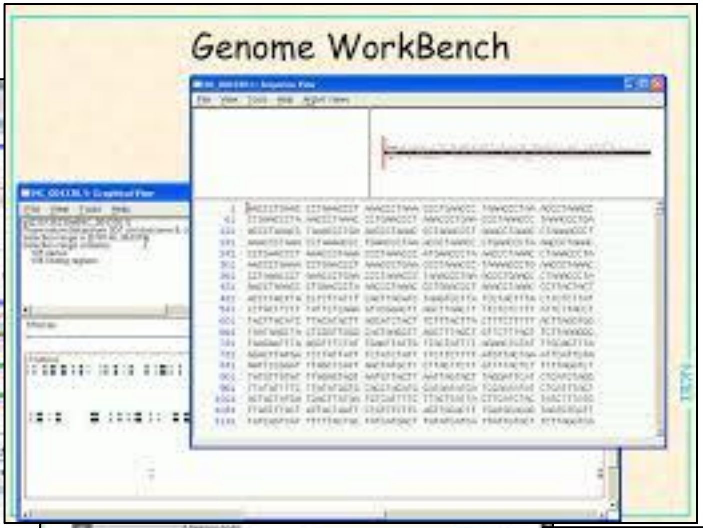
WGS is Spawning New Careers in Microbiology !!!!!

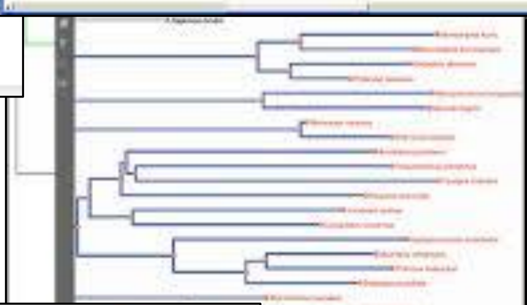


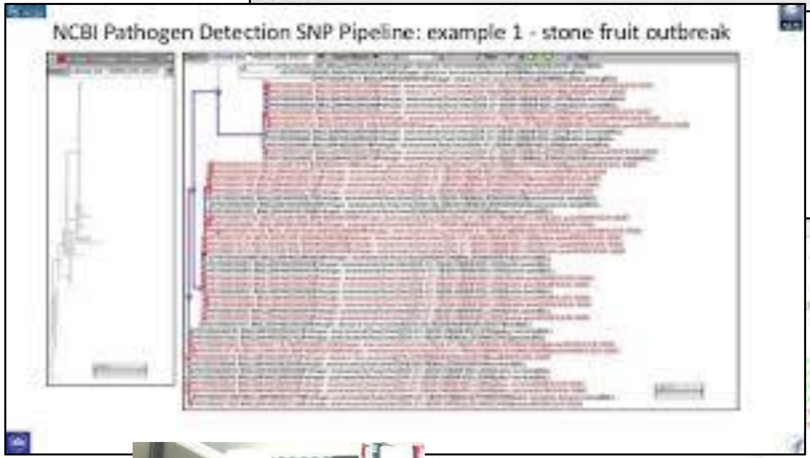
(1) Research Genome Microbiologist
(2) Bioinformatic Traceback Specialist
(3) Comparative Microbial Genomicist

DESKTOP TOOLS FOR ALL LEVELS









Summary

- Conventional approaches to strain subtyping cannot always remain in step with a foodborne investigation. Rapid methods for genotyping outbreak strains in “high resolution” are needed.

- Certain foodborne pathogenic strains are highly clonal in nature, requiring the Application of methods with enhanced sensitivity for differentiating highly homogeneous groups of isolates.

- WGS is the “engine” that will drive gene/SNP discovery and the future of one Microbiology work flow in the diagnostic microbiology laboratory.

- WGS is agnostic, portable, and highly capable of distributed networking

- WGS provides the ultimate in discriminatory power by interrogating millions of nucleotide positions in a matter of just a few hours.

- WGS-based typing approaches are supplanting current Microbiological methods (*i.e.*, phage typing – serotyping, PFGE, etc.) AND WILL YIELD ONE MICROBIOLOGICAL WORKFLOW VERY SOON.

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**