



# THE PROMISE OF MICROBIAL GENOMICS: HOW MICROBIOLOGY IS STANDING UP TO THE MANY CHALLENGES OF A 21<sup>ST</sup> CENTURY FOOD SUPPLY

**Eric W. Brown, Ph.D., M.Sc., FAAM**

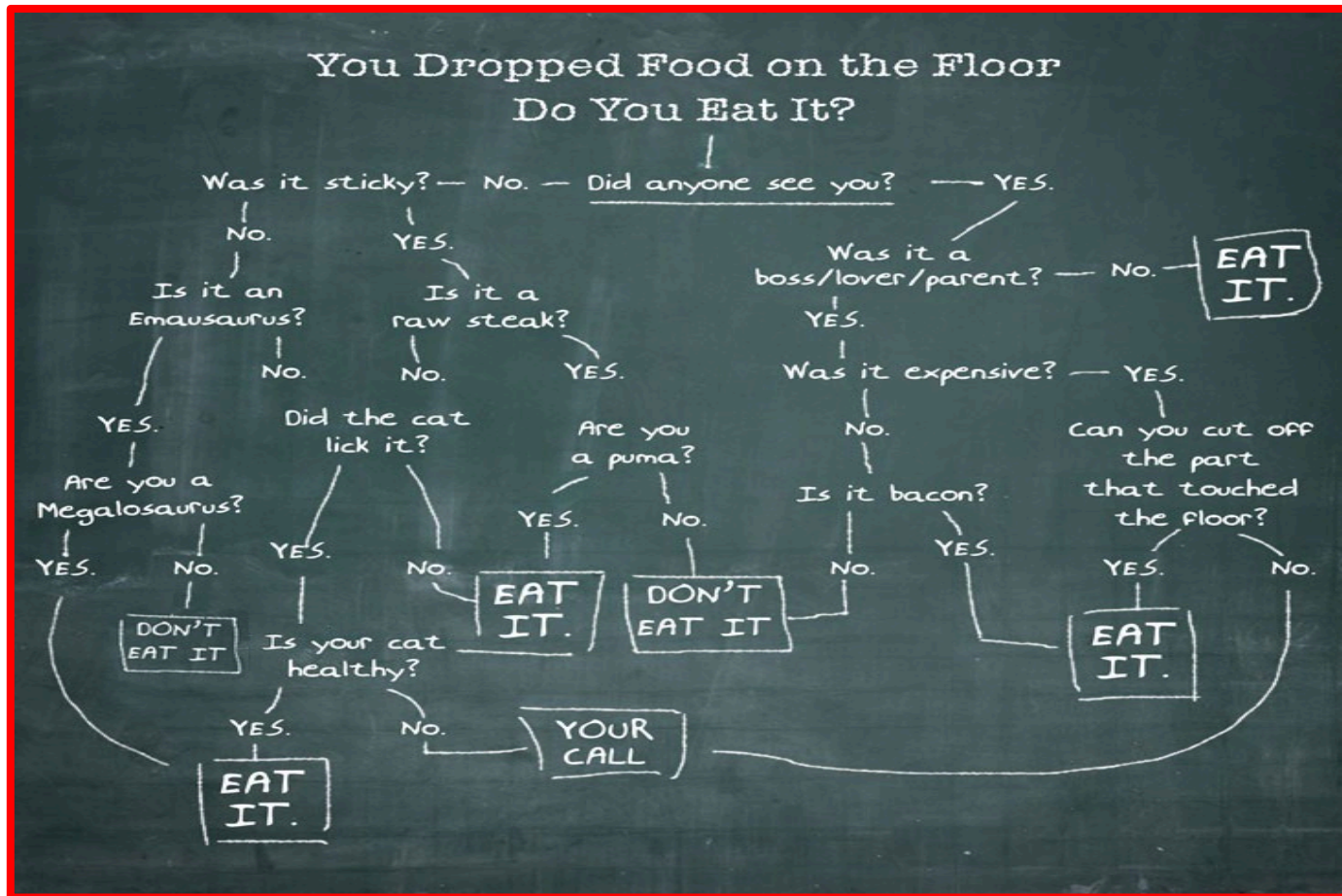
*Director, Division of Microbiology*

*Office of Regulatory Science, Center for Food Safety and Applied Nutrition*

*US Food and Drug Administration, Washington, DC*



## Recent Advances in Food Safety Science



(Credited to D. Archer, UF)

# THE ORIGINS OF WGS

*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.*

AKA

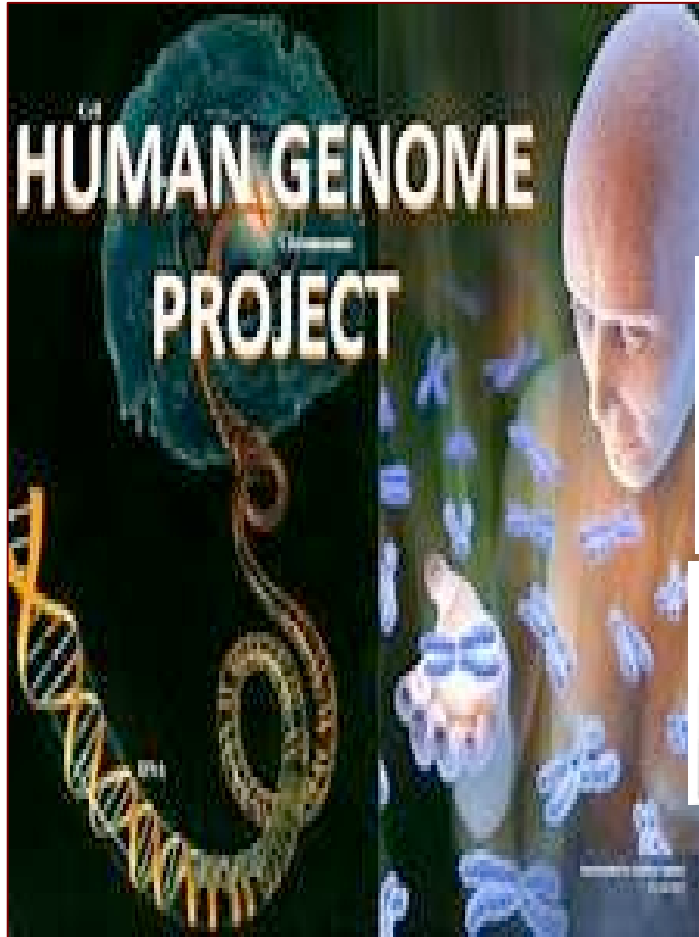
Next-Generation Sequencing (NGS)

High Throughput Sequencing (HTS)

Massive Parallel Sequencing (MPS)

Whole Genome Sequencing (WGS)

AND... “Shotgun Sequencing”





**“We Have Met the Enemy and He is Us...”**  
**~Walt Kelly (Pogo)**

**(1) Adding a new word when the old word would do just fine.**

**subtyping, genotyping, genomo-typing, seque-typing, molecular typing, DNA typing, **NANOtyping from...****

**High resolution clustering of *Salmonella enterica* serovar Montevideo strains using a next-generation sequencing approach.**

**Allard MW, Luo Y, Strain E, Li C, Keys CE, Son I, Stones R, Musser SM, Brown EW.  
BMC Genomics. 2012 Jan 19;13:32. doi: 10.1186/1471-2164-13-32.**

**(2) Not one more Acronym, please!**

**(SNP, hqSNP, SNV, mutation, polymorphism, substitution, synapomorphy, signature synapomorphy,**

**(3) How many more \_\_\_\_\_ (Fill-in-theblank) – OMICS partitions are genuinely essential to the effort?**

**(contaminOMICS, SalmonellOMICS, vegetablOMICS + those already established trialomics, connectomics, cytomics, Editomics, embryomics, enviromics, epigenomics, exomics, foodomics, glycomics, mechanomics, metabolomics, Microbiomics, ORFeomics, phenomics, physiomics, proteomics, regulomics, transcriptomics, toponomics**



# ***“Whole Genome Sequencing Is The Biggest Thing Happen To Food Microbiology Since Pasteur Showed Us How To Culture Pathogens***



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



©2011, Illumina Inc. All rights reserved.



## WHY DOES WGS MATTER?

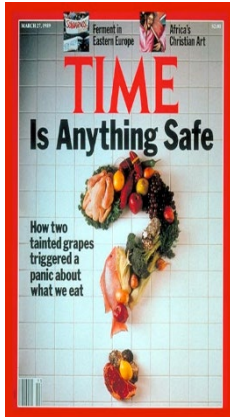




# *Safe, Wholesome, Sanitary Foods*



# Battling Foodborne Illness

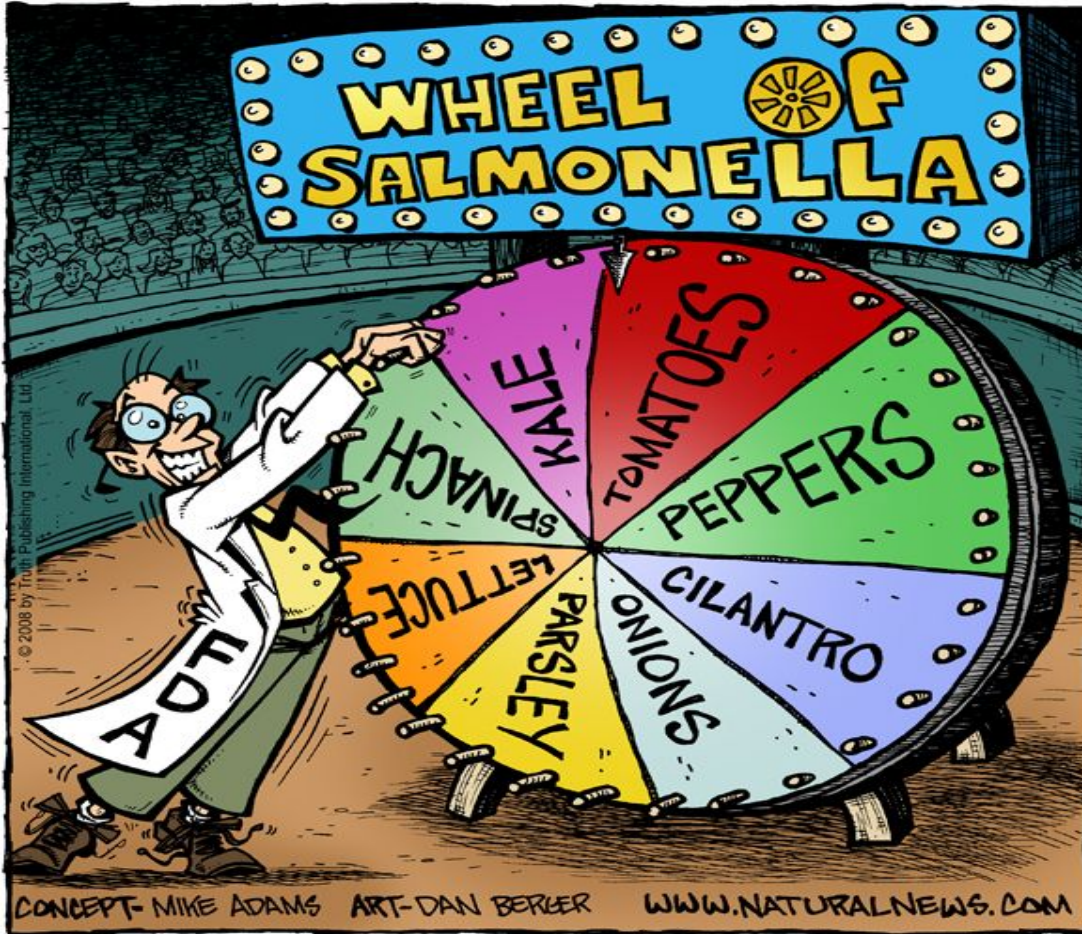


- Each year nearly 10 million episodes of foodborne illness in the United States
- 60,000 hospitalizations
- More than 1300 deaths
- *Salmonella* cause 11% of foodborne illnesses each year





## COUNTERTHINK - "WHEEL OF SALMONELLA"

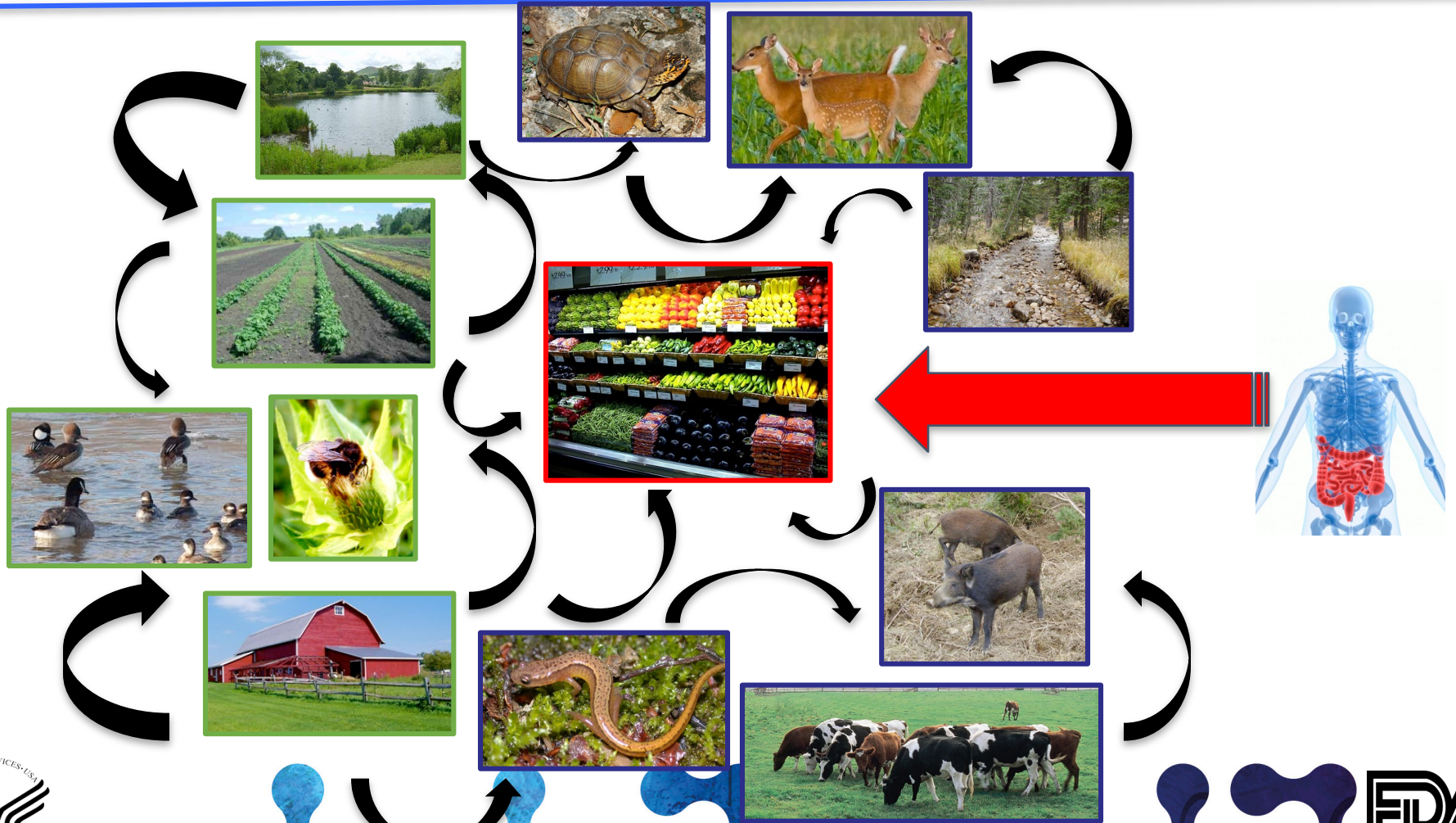


In recent times, *Salmonella* has emerged to contaminate a wide variety of produce and other RCR commodities, revealing its extraordinary adaptability for persistence in numerous niches that comprise the farm-to-fork continuum.

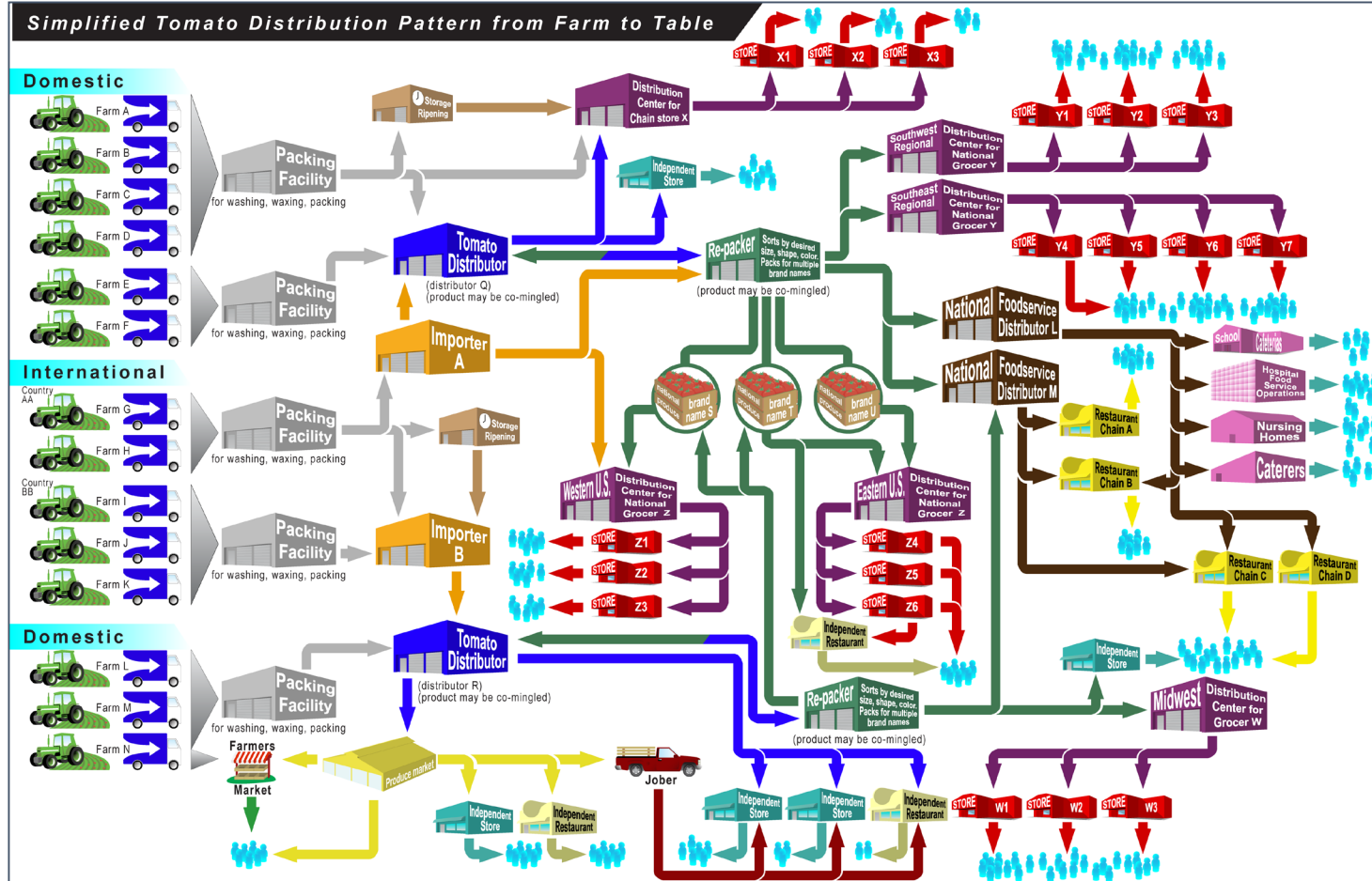
## ***The Salmonella 'radiation'***



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”.



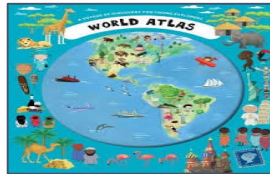
The Fresh-cut Tomato Supply Chain

# The Complex and Global Etiology of Foods

## Salad



Shrimp – India  
Cilantro – Mexico  
Romaine – Salinas, CA  
Cheddar – Wisconsin  
Carrots – Idaho  
Gruyere – Switzerland  
Pecans – Georgia  
Sprouts – Chicago  
Red Cabbage - NY



## Sushi



Shrimp – Indonesia  
Imitation Crab – Alaska  
Tuna Scrape – India  
Fish Roe – Seychelles  
Salmon – Puget Sound  
Soy Sauce – China  
Rice – Thailand  
Seaweed Wrap – CA  
Avocado – Mexico  
Cucumber – Maryland  
Wasabi – Japan  
Pepper – Vietnam

## Fruit platter

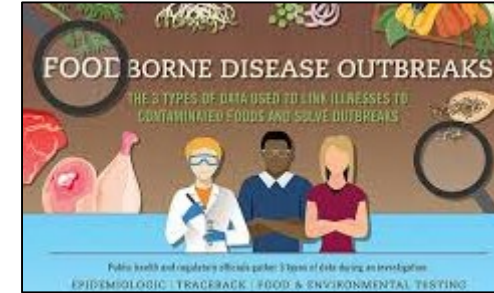
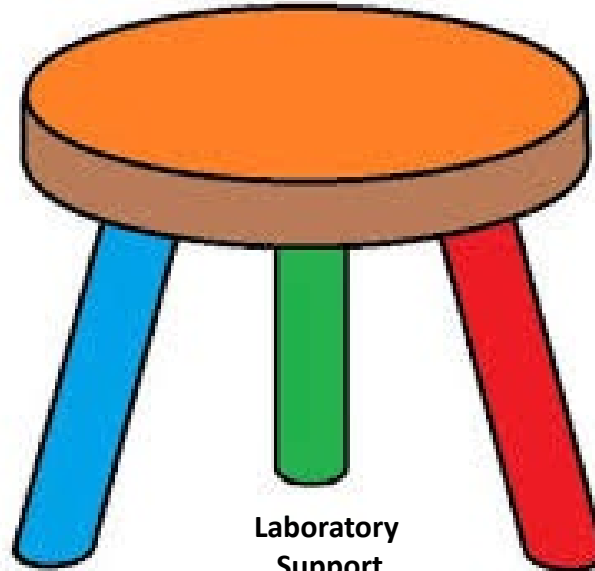
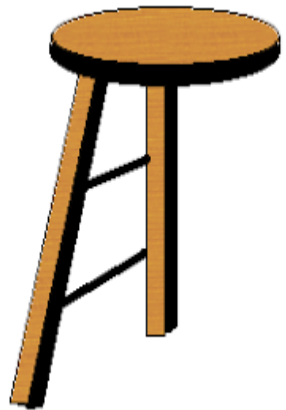


Watermelon – Delaware  
Blackberries – Guatemala  
Blueberries – New Jersey  
Pineapple – Guam  
Grapes – California  
Kiwi – New Zealand  
Apples – New York  
Pears – Oregon  
Cantaloupe – Costa Rica  
Honeydew – Arizona  
Papaya – Mexico  
Banana – Costa Rica

## Some perspective on the US 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 3 Legged Stool of Foodborne Outbreak Investigations



### Epidemiology

- ~questionnaires
- ~interviews
- ~case review
- ~surveillance and reporting
- ~illness clusters
- ~food associations
- ~geographic patterns

### Laboratory Support

- ~pathogen isolation
- ~identification
- ~serology
- ~subtyping
  - PFGE
  - WGS

### Traceback

- ~distribution chains/product flow
- ~environmental assessments
- ~potential source reservoirs
- ~GAP/GMP review of HACCPs
- ~Inspectional findings
- ~Risk assessments

### Global Point Source



### Import Lines



### Ecologic Reservoirs



Tracking contamination down and FAST!



Finished Product



Processing Facility



Farm

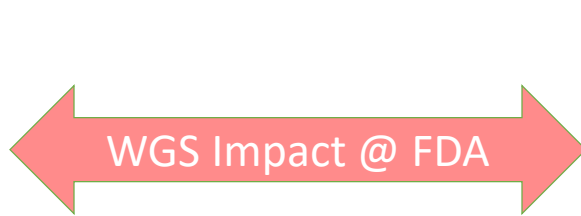
**SAVES LIVES**



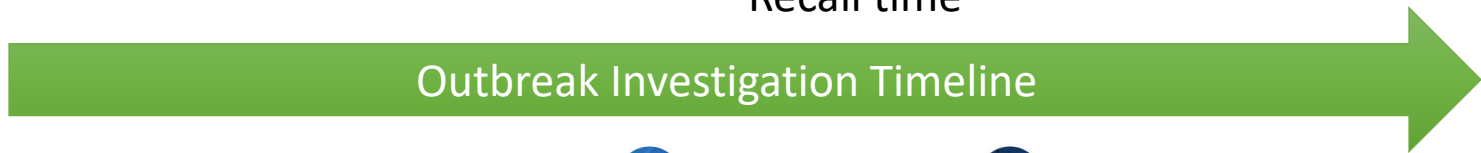
# Outbreak Investigation Response Timelines



Questionnaires  
Laboratory results  
Patient interviews



Inspection  
Sample collection  
Laboratory results  
Recall time





©2011, Illumina Inc. All rights reserved.



## THE BEGINNING



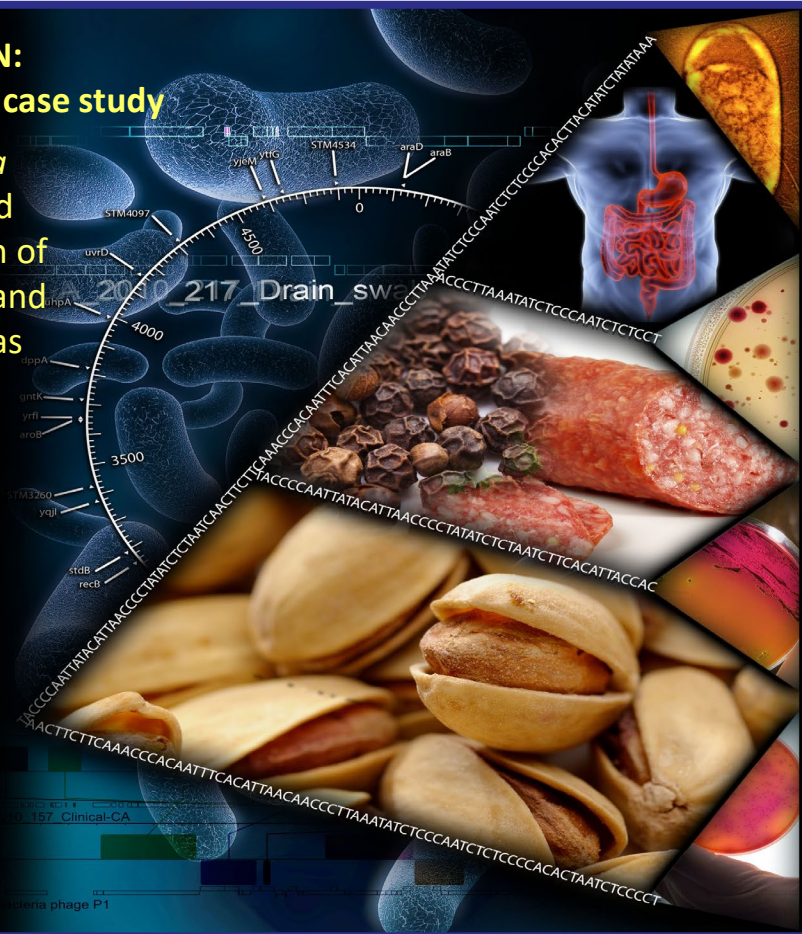
CASE 1: Same PFGE Patterns

**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.

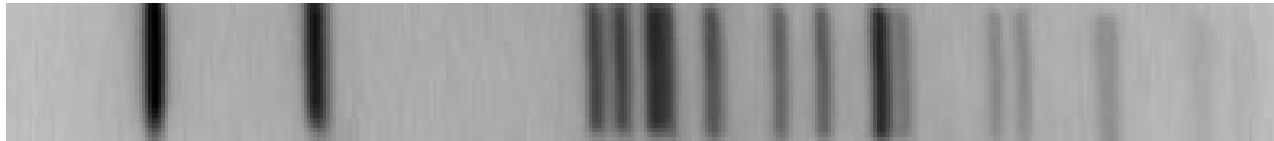
15-20x shot gun sequencing  
35 pure culture isolates  
from patients, foods and  
Environmental samples.

Concatenate 40 variable genes for  
Phylogenetic analysis



**During the S. Montevideo outbreak, all isolates were indistinguishable by 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> enzyme PFGE.**

**PFGE-XbaI**



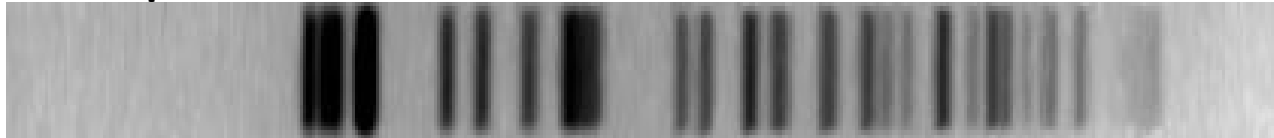
**JIXX01.0011**

**PFGE-BlnI**

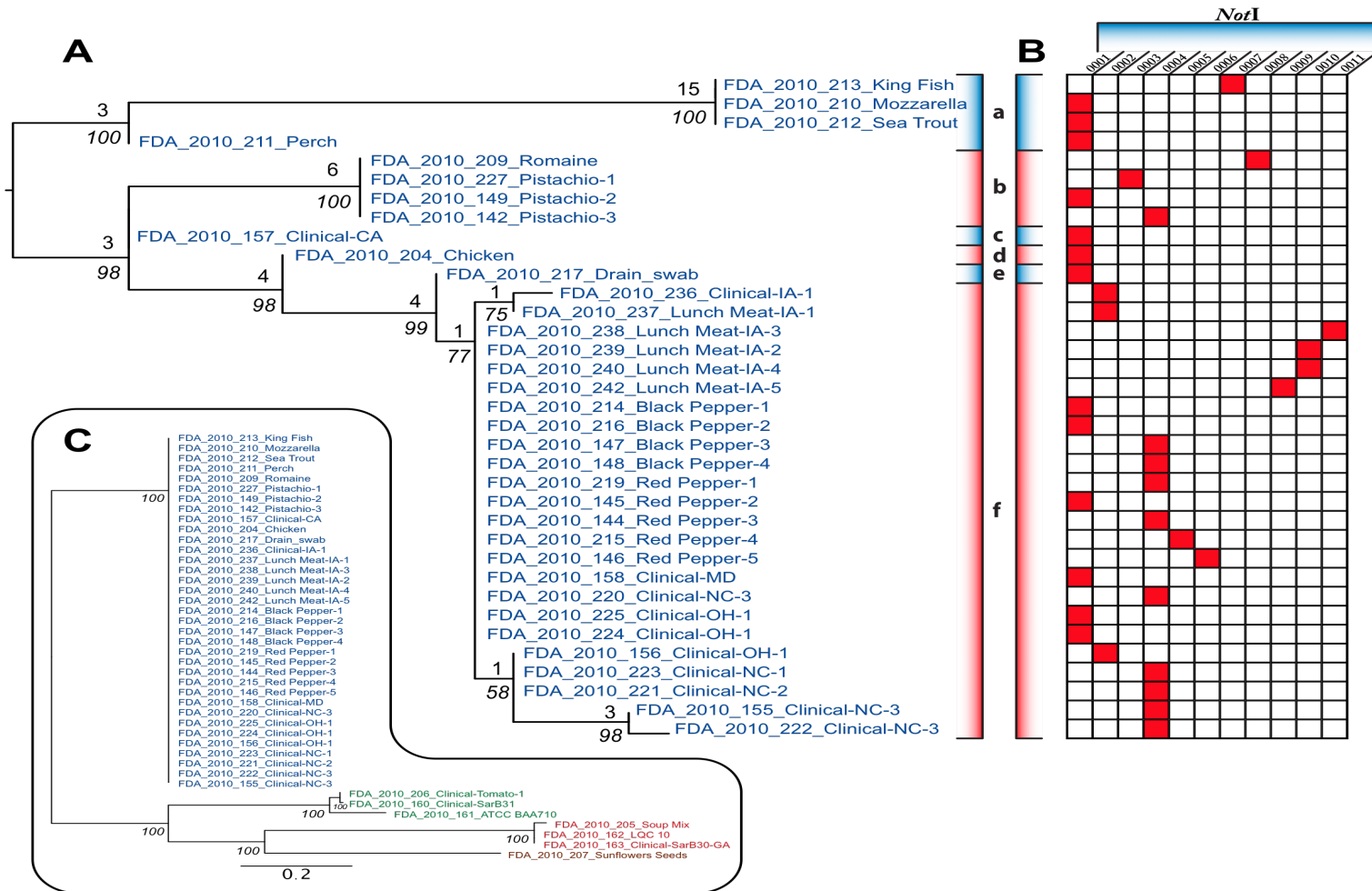


**JIXA26.0012**

**PFGE-SpeI**



**JIXS18.0001**



*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...**

**STAR-GAZING**



**LIGHT-TELESCOPE**

**MODERN REFRACTION TELESCOPE**



**RADIO TELESCOPE**



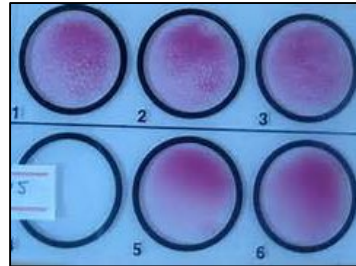
**HUBBLE**

**PATHOGEN PLATING**

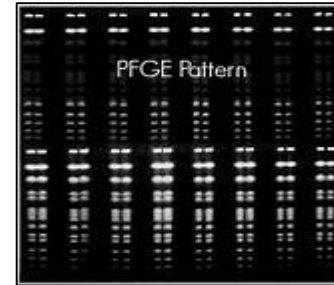


**BIOTYPING SPECIATION**

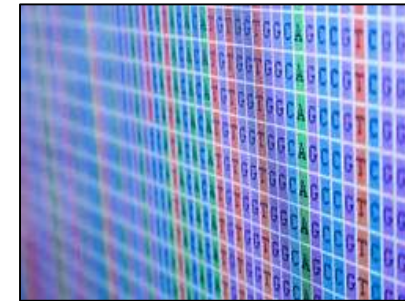
**SEROTYPING**



**PFGE**

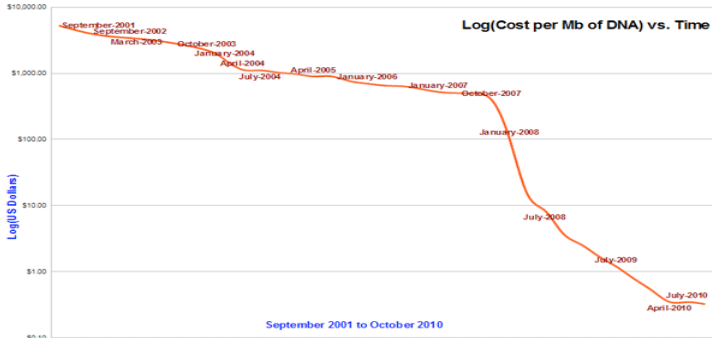


**WGS**



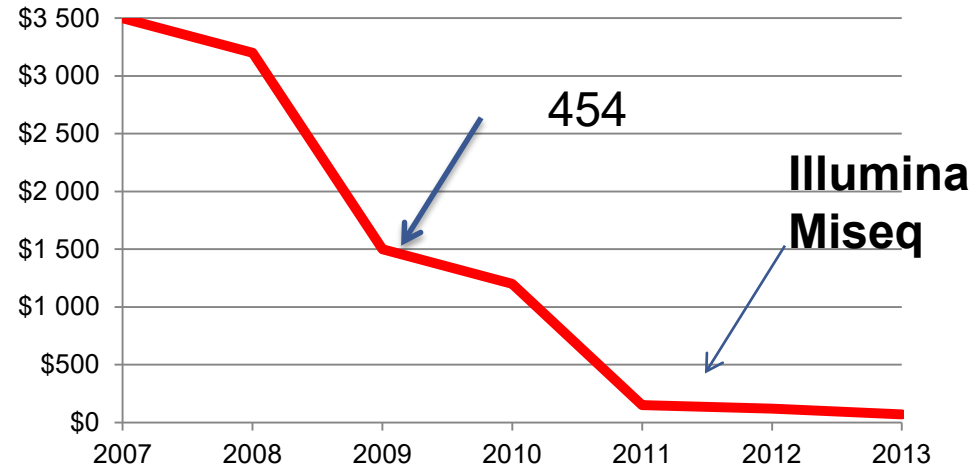
time →

**THE EVOLUTION OF SUBTYPING TOOLS FOR BACTERIAL PATHOGENS**



- 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



\$150/genome  
in 2014

\$75/genome  
in 2019 w/  
Illumina NextSeq Technology

**Illumina NextSeq Technology**





***Third- and fourth- generation sequencing (also known as long-read sequencing) is a class of DNA sequencing methods currently under active development. ...this sequencing works by reading the nucleotide sequences at the single molecule level, in contrast to existing methods that require breaking long strands of DNA into small segments then inferring nucleotide sequences by amplification and synthesis.***



## Brussels-September 2011-Birth of “GMI”

### 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.

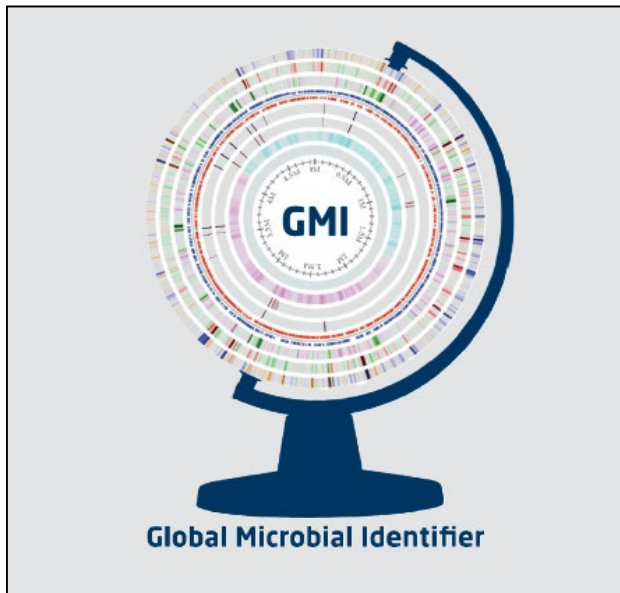
Whole-genome sequencing can give better, faster answers about organisms, says Jørgen Schlundt of the Center for Genomic Epidemiology (CGE) at the Danish Techni-

ated from [www.sciencemag.org](http://www.sciencemag.org) on September 30, 2011

# 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



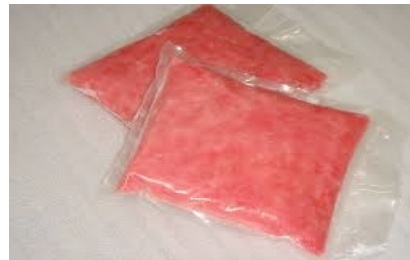
## Work groups

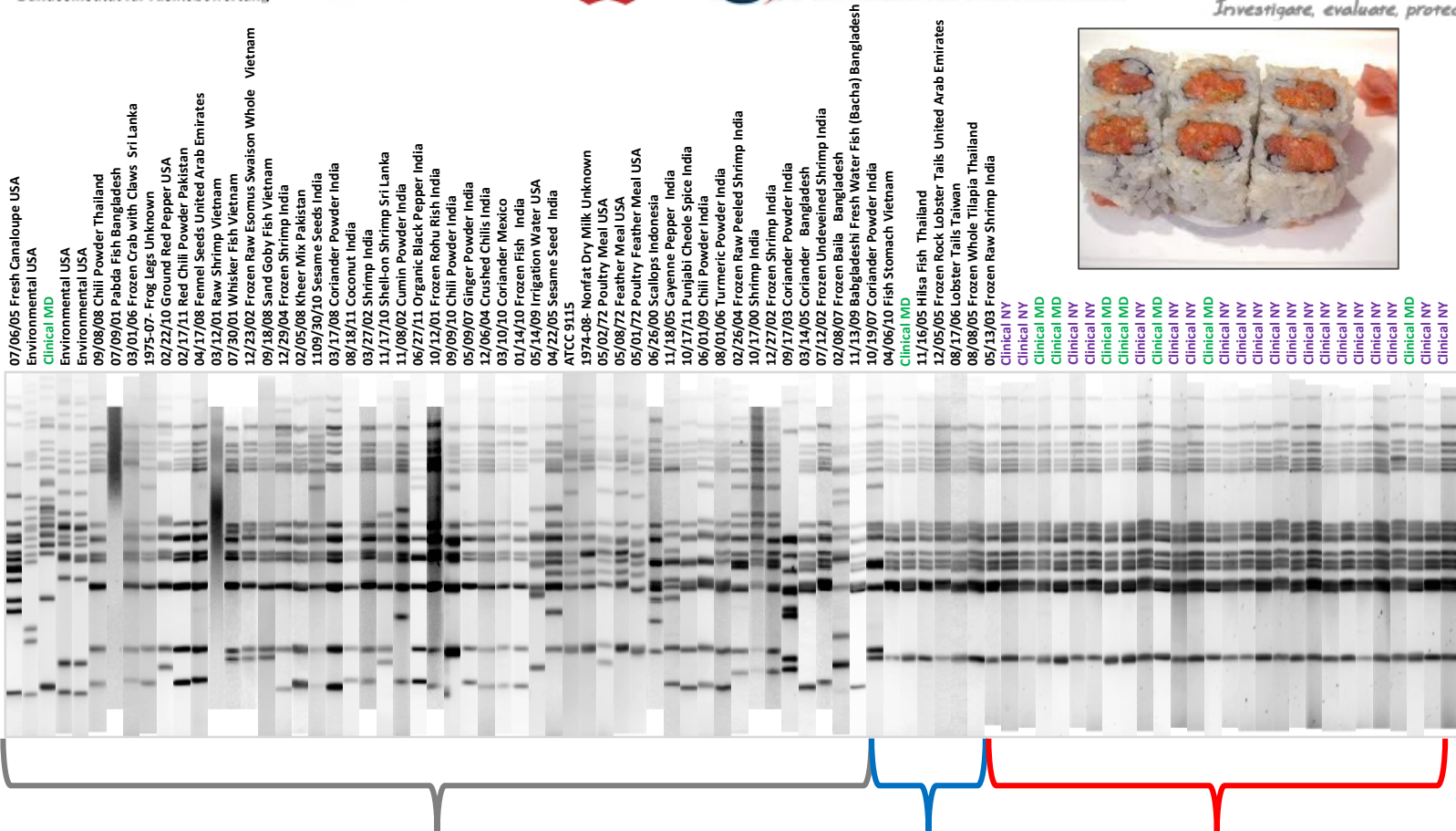
1. Political challenges, outreach and building a global network
2. Repository and storage of sequence and meta-data
3. Analytical approaches
4. Ring trials and quality assurance
5. Pilot project

## 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





- 07/06/05 Fresh Canaloupe USA  
Environmental USA  
Clinical MD
- Environmental USA
- Environmental USA
- Environmental USA
- 07/09/01 Pabda Fish Bangladesh
- 09/08/08 Chili Powder Thailand
- 03/01/06 Frozen Crab with Claws Sri Lanka
- 1975-07 - Frog Legs Unknown
- 02/22/10 Ground Red Pepper USA
- 02/17/11 Red Chili Powder Pakistan
- 04/17/08 Fenmel Seeds United Arab Emirates
- 03/12/01 Raw Shrimp Vietnam
- 07/30/01 Whisker Fish Vietnam
- 12/23/02 Frozen Raw Esomus Swaison Whole Vietnam
- 09/18/08 Sand Goby Fish Vietnam
- 12/29/04 Frozen Shrimp India
- 02/05/08 Kheer Mix Pakistan
- 11/09/30/10 Sesame Seeds India
- 03/17/08 Coriander Powder India
- 08/18/11 Coconut India
- 03/27/02 Shrimp India
- 11/17/10 Shell-on Shrimp Sri Lanka
- 11/08/02 Cumin Powder India
- 06/27/11 Organic Black Pepper India
- 10/12/01 Frozen Rohu Fish India
- 09/09/10 Chili Powder India
- 05/09/07 Ginger Powder India
- 12/06/04 Crushed Chilis India
- 03/10/10 Coriander Mexico
- 01/14/10 Frozen Fish India
- 05/14/09 Irrigation Water USA
- 04/22/05 Sesame Seed India
- ATCC 9115
- 1974-08 - Nonfat Dry Milk Unknown
- 05/02/72 Poultry Meal USA
- 05/08/72 Feather Meal USA
- 05/01/72 Poultry Feather Meal USA
- 06/26/00 Scallops Indonesia
- 11/18/05 Cayenne Pepper India
- 10/17/11 Punjabi Cheole Spice India
- 06/01/09 Chili Powder India
- 02/26/04 Frozen Raw Peeled Shrimp India
- 10/17/00 Shrimp India
- 12/27/02 Frozen Shrimp India
- 09/17/03 Coriander Powder India
- 07/14/05 Coriander Bangladesh
- 07/12/02 Frozen Undeveloped Shrimp India
- 02/08/07 Frozen Bailla Bangladesh
- 11/13/09 Bangladesh Fresh Water Fish (Bacha) Bangladesh
- 10/19/07 Coriander Powder India
- 04/06/10 Fish Stomach Vietnam  
Clinical MD
- 11/16/05 Hilsa Fish Thailand
- 12/05/05 Frozen Rock Lobster Tails United Arab Emirates
- 08/17/06 Lobster Tails Taiwan
- 08/08/05 Frozen Whole Tilapia Thailand
- 05/13/03 Frozen Raw Shrimp India  
Clinical NY
- Clinical NY
- Clinical NY
- Clinical MD
- Clinical MD
- Clinical NY
- Clinical NY
- Clinical MD
- Clinical MD
- Clinical NY
- Clinical MD
- Clinical NY
- Clinical MD
- Clinical NY
- Clinical NY
- Clinical NY
- Clinical NY
- Clinical NY
- Clinical NY
- Clinical NY
- Clinical NY
- Clinical NY
- Clinical NY
- Clinical NY
- Clinical NY
- Clinical NY
- Clinical MD
- Clinical NY
- Clinical NY



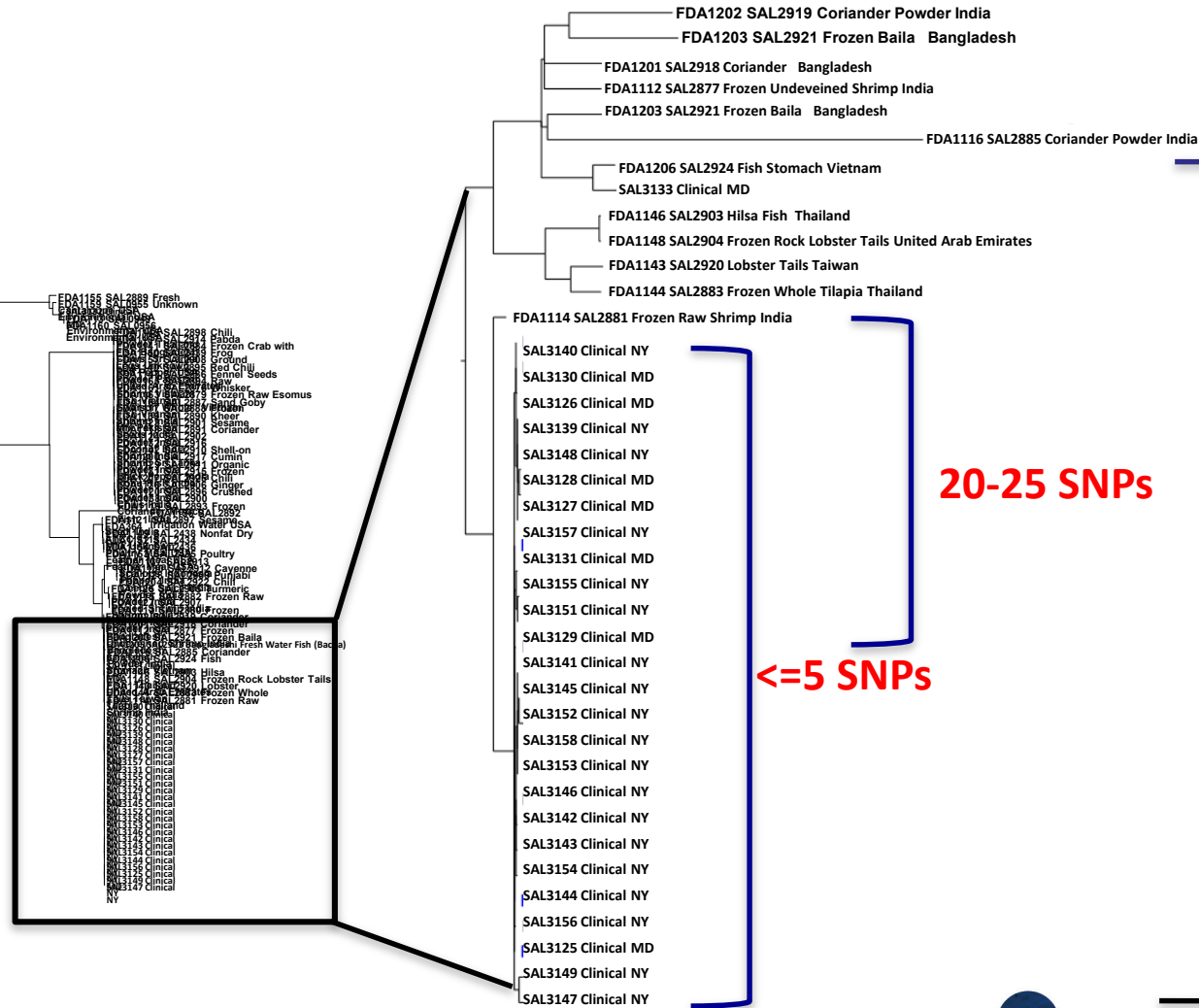
**S. Bareilly Outbreak  
April-June 2012**

Different PFGE than the outbreak pattern

Same PFGE but not part of the outbreak

Outbreak Isolates

MD isolates – in green  
NY isolates – in purple



**20-25 SNPs**

**<=5 SNPs**

**PFGE  
Match  
110-130 SNPs**

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





©2011, Illumina Inc. All rights reserved.



## THE NETWORK





# 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



# Basic Data Flow for Global WGS Public Access Databases

## DATA ACQUISITION

Sequence and upload genomic and geographic data



Other distributed sequencing networks

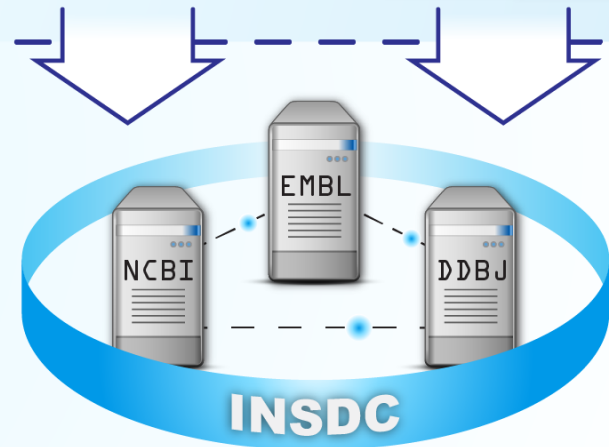


## DATA ASSEMBLY, ANALYSIS, AND STORAGE

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

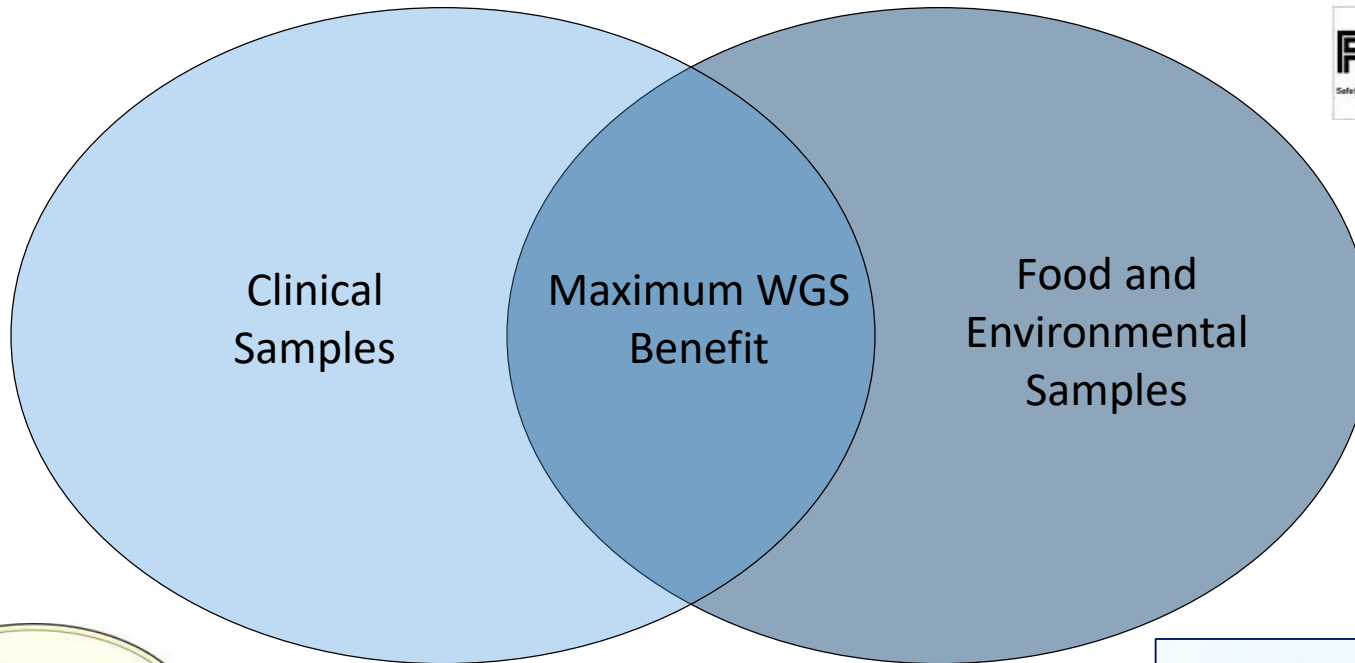
- Find clinical links
- Identify clusters
- Conduct traceback
- Develop rapid methods
- Develop culture independent tests
- Develop new analytical software



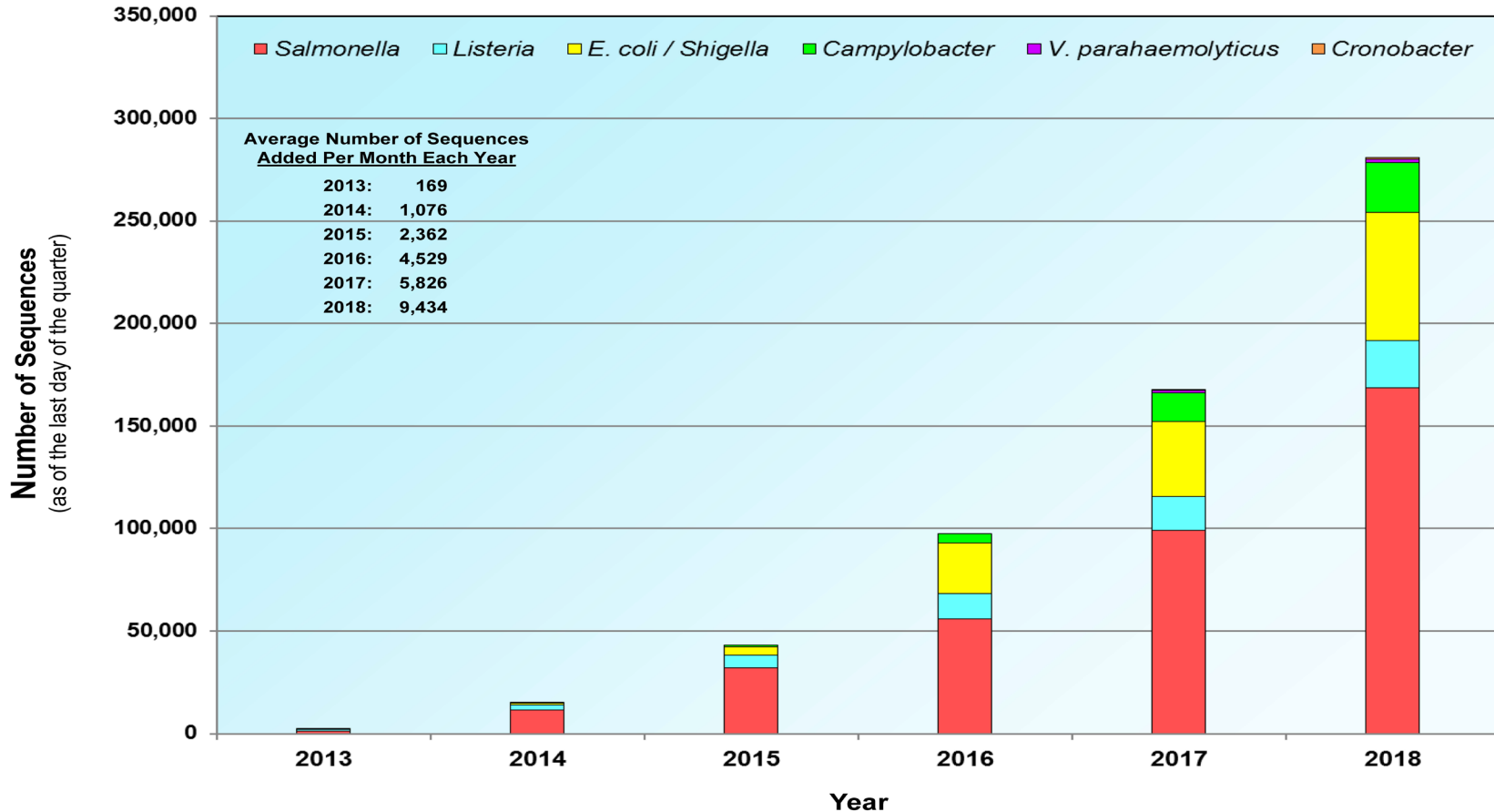
# GenomeTrakr

- First distributed sequencing based network
- State and Federal laboratory network collecting and sharing genomic data from foodborne pathogens
- Partner with NIH/NCBI for storage and serving data
- Partner with CDC for human real-time surveillance
- Partner with FSIS/USDA to better cover the food supply
- Partner with the food industry to expand use to industry
- Partner with international organizations to expand use worldwide
- Open-access genomic reference database

# Importance of a Balanced Approach



## Total Number of Sequences in the GenomeTrakr Database



First sequences uploaded in February 2013

# December 31, 2018 GenomeTrakr Numbers

Species	Total Isolates
<i>Salmonella enterica</i>	168,465
<i>E.coli and Shigella</i>	62,469
<i>Campylobacter jejuni</i>	24,188
<i>Listeria monocytogenes</i>	23,217
<i>Vibrio parahaemolyticus</i>	1,806
<i>Cronobacter</i>	607
<b>Total</b>	<b>280,752</b>

**All Organism Group Sequences**  
**Curated and Analyzed by NCBI – December 31, 2018**

Species	Version	Publication Date*	Latest Isolate Creation Date*	New Isolates	New Clinical Isolates	New Environmental Isolates	Total Isolates
<i>Salmonella enterica</i>	PDG000000002.1274	12/30/2018 7:59	12/29/2018 0:55	348	68	280	168,465
<i>E.coli and Shigella</i>	PDG000000004.1106	12/31/2018 10:12	12/31/2018 1:05	11	11	0	62,469
<i>Campylobacter jejuni</i>	PDG000000003.710	12/28/2018 9:36	12/27/2018 18:52	43	1	42	24,188
<i>Listeria monocytogenes</i>	PDG000000001.1087	12/29/2018 5:10	12/28/2018 18:50	11	11	0	23,217
<i>Klebsiella pneumoniae</i>	PDG000000012.370	12/30/2018 14:59	12/29/2018 22:54	302	1	301	9,246
<i>Mycobacterium tuberculosis</i>	PDG000000034.85	11/5/2018 8:47	10/31/2018 8:50	2	2	0	7,378
<i>Acinetobacter baumannii</i>	PDG000000010.223	12/28/2018 9:25	12/27/2018 16:56	118	3	115	4,784
<i>Neisseria</i>	PDG000000032.88	12/28/2018 9:26	12/27/2018 16:57	38	5	33	4,656
<i>Pseudomonas aeruginosa</i>	PDG000000036.159	12/28/2018 9:26	12/27/2018 16:57	276	94	182	4,131
<i>Vibrio parahaemolyticus</i>	PDG000000023.183	12/13/2018 5:51	12/12/2018 12:46	2	2	0	1,806
<i>Clostridioides difficile</i>	PDG000000045.54	12/11/2018 11:53	12/10/2018 10:51	5	5	0	1,700
<i>Enterobacter</i>	PDG000000028.158	12/29/2018 7:15	12/28/2018 11:59	1	1	0	1,699
<i>Legionella pneumophila</i>	PDG000000026.48	12/28/2018 9:13	12/24/2018 11:24	5	0	5	758
<i>Cronobacter</i>	PDG000000043.29	12/23/2018 8:57	12/19/2018 8:51	1	1	0	607
<i>Serratia marcescens</i>	PDG000000016.115	12/28/2018 9:04	12/24/2018 12:19	7	1	6	544
<i>Staphylococcus pseudintermedius</i>	PDG000000042.64	12/21/2018 11:58	12/20/2018 13:48	0	0	0	358
<i>Klebsiella oxytoca</i>	PDG000000030.92	12/29/2018 5:03	12/28/2018 12:00	18	18	0	318
<i>Citrobacter freundii</i>	PDG000000039.86	12/29/2018 7:16	12/28/2018 12:00	16	16	0	274
<i>Providencia alcalifaciens</i>	PDG000000021.81	12/28/2018 9:11	12/24/2018 10:09	1	0	1	160
<i>Elizabethkingia anophelis</i>	PDG000000014.49	9/26/2018 7:28	8/2/2018 18:43	1	1	0	151
<i>Morganella morganii</i>	PDG000000020.62	12/28/2018 9:07	12/25/2018 0:02	3	2	1	70
<i>Kluyvera intermedia</i>	PDG000000040.19	12/28/2018 9:17	12/24/2018 9:45	1	0	1	15
<b>Total</b>				<b>1,210</b>	<b>243</b>	<b>967</b>	<b>316,994</b>

## Foodborne illness acquired in US – Major pathogens (CDC, 2011)

Pathogens	Foodborne illnesses/year	Note
Bacterial	3.6 million (39%)	
Parasitic & others	0.2 million (2%)	
Viral	5.5 million (59%)	<b>Norovirus: 58%</b>
Total (31 major pathogens*)	9.4 million	

*Slide courtesy of Dr. Zhihui Yang, OARSA/CFSAN/FDA*



# FDA NCBI BioProject for Foodborne Viruses

NCBI Resources How To Sign in to NCBI

BioProject

[Create alert](#) [Advanced](#) [Browse by Project attributes](#) [Help](#)

Display Settings: ▾

Send to: ▾

## foodborne viruses

Accession: PRJNA396739 ID: 396739

### Genome sequencing of foodborne viruses by FDA

Foodborne viruses are the leading cause of foodborne illnesses outpacing all other pathogens combined including bacteria. This project, initiated by the Molecular Virology Team in the Division of Molecular Biology/ Office of Applied Research and Safety Assessment, aims to (1) sequence a wide range of foodborne viruses (both commonly documented and less commonly documented) from clinical specimens, food and water; and (2) thus provide additional reference sequences for phylogenetic analysis and epidemiologic studies linked to foodborne illnesses.

Accession	PRJNA396739
Type	Umbrella project
Submission	Registration date: 1-Aug-2017 <b>OARSA/CFSAN/FDA</b>
Relevance	Foodborne viruses

### foodborne viruses encompasses the following 3 sub-projects:

Project Type		Number of Projects
<b>Umbrella project</b>		3
BioProject accession	Name	Title
PRJNA433976	Metagenomic sequences from food	Metagenomic sequences from food (OARSA/CFSAN/FDA)
PRJNA433975	Metagenomic sequences from human clinical samples	Metagenomic sequences from human clinical samples (OARSA/CFSAN/FDA)
PRJNA433977	Metagenomic sequences from water	Metagenomic sequences from water (OARSA/CFSAN/FDA)

### Related information

BioProject

Data projects

### Recent activity

[Turn Off](#) [Clear](#)

- foodborne viruses BioProject
- PRJNA396739 (1) BioProject
- Gyrase A Mutations in Campylobacter Associated with Decreased Susceptibi PubMed
- Use of whole-genome sequencing for Campylobacter surveillance from NAR PubMed
- Whitehouse CA (96) PubMed

### IMPACT:

- **Sequence database of a wide range of foodborne viruses**
- **Publicly accessible**
- **Phylogenetic analysis**
- **Epidemiologic studies**
- **Monitor potential outbreaks**

## 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

PNUSAL000140

NY-NYAG12B11863-1

PNUSAL000355

VA-WGS-00228

VA-WGS-00224

CFSAN010072

CFSAN010076

VA-WGS-00226

VA-WGS-00227

VA-WGS-00229

PNUSAL000569

CFSAN010075

PNUSAL000517

PNUSAL000520

VA-WGS-00223

VA-WGS-00222

CFSAN010068

CFSAN010073

CFSAN010074

CFSAN010077

VA-WGS-00221

VA-WGS-00225

VA-WGS-00231

CFSAN010071

CFSAN010070

VA-WGS-00230


CFSAN010069

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

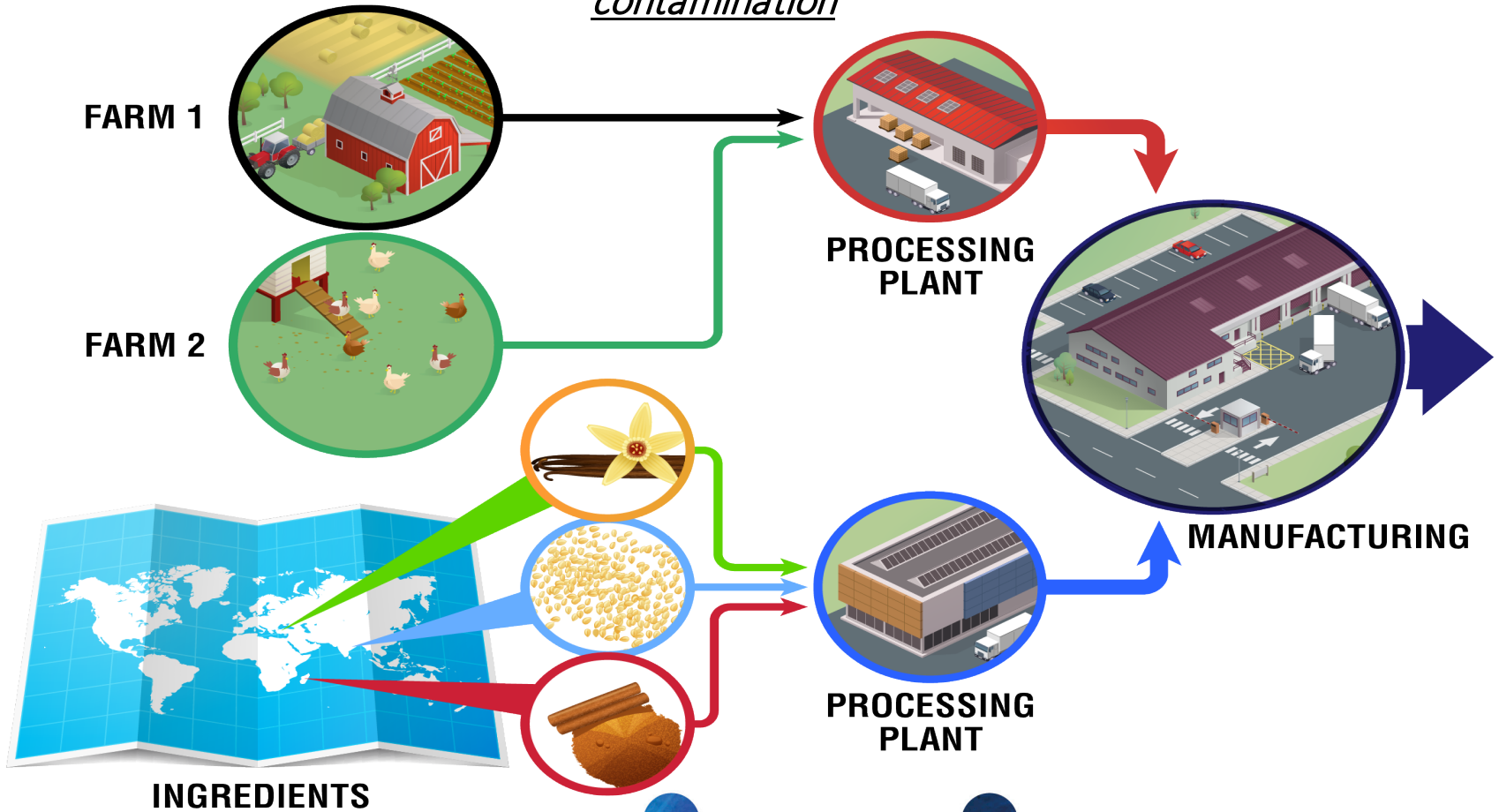
████████████████████  
████████████████████  
████████████████████  
251 Roos Lane  
Kenton, DE 19955

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

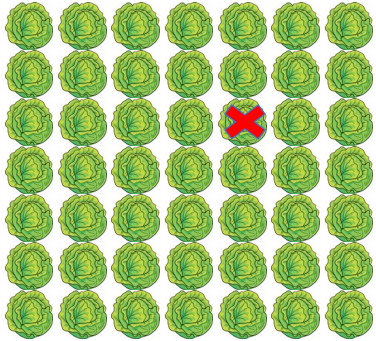
████████████████████

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.

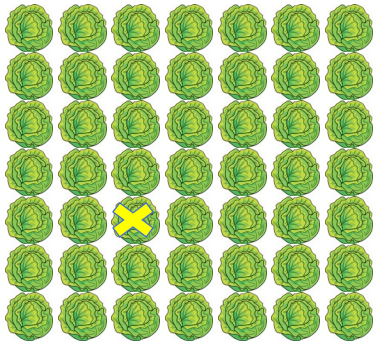
Environmental sampling combined with WGS can help point to root cause of the contamination



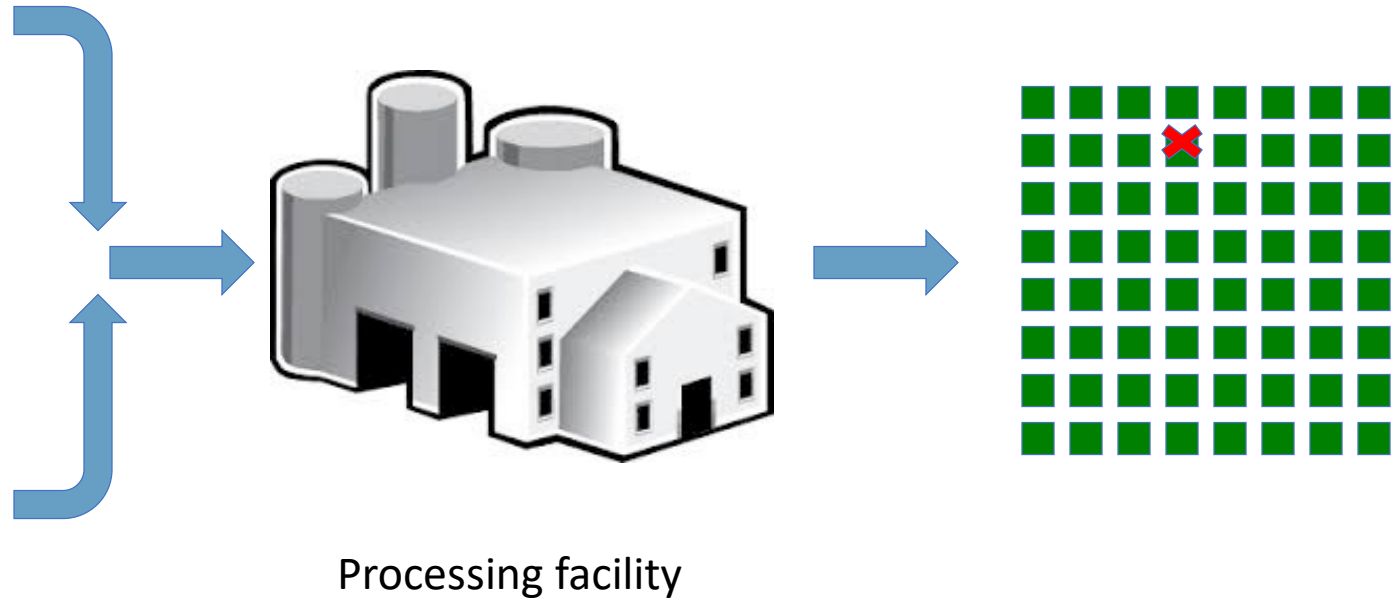
Field 1



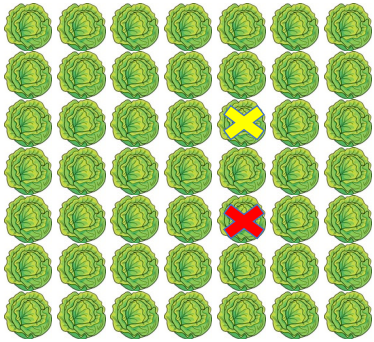
Field 2



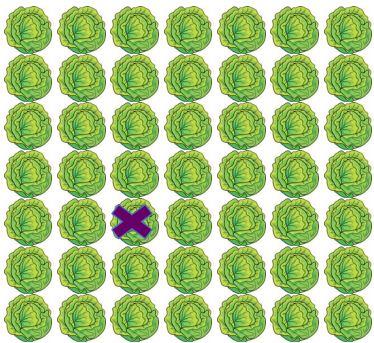
## Scenario 1 (pass through)



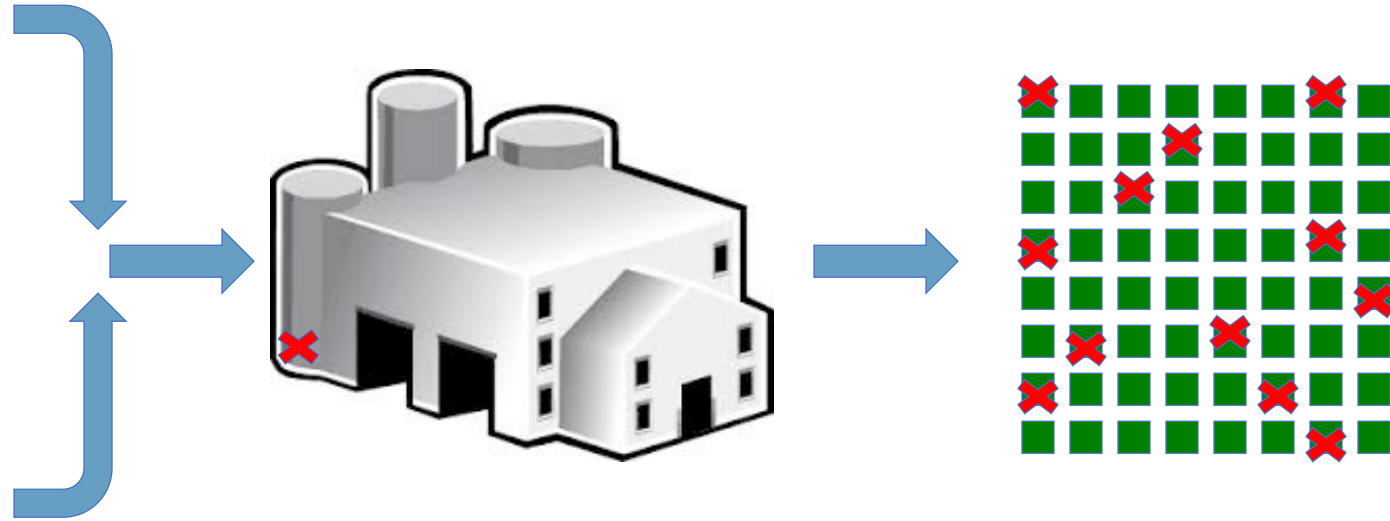
### Field 1



### Field 2



## Scenario 2: Resident Contamination



Processing facility



## ***L. monocytogenes* found on finished product samples in 2014 and through environmental sampling in 2015:**

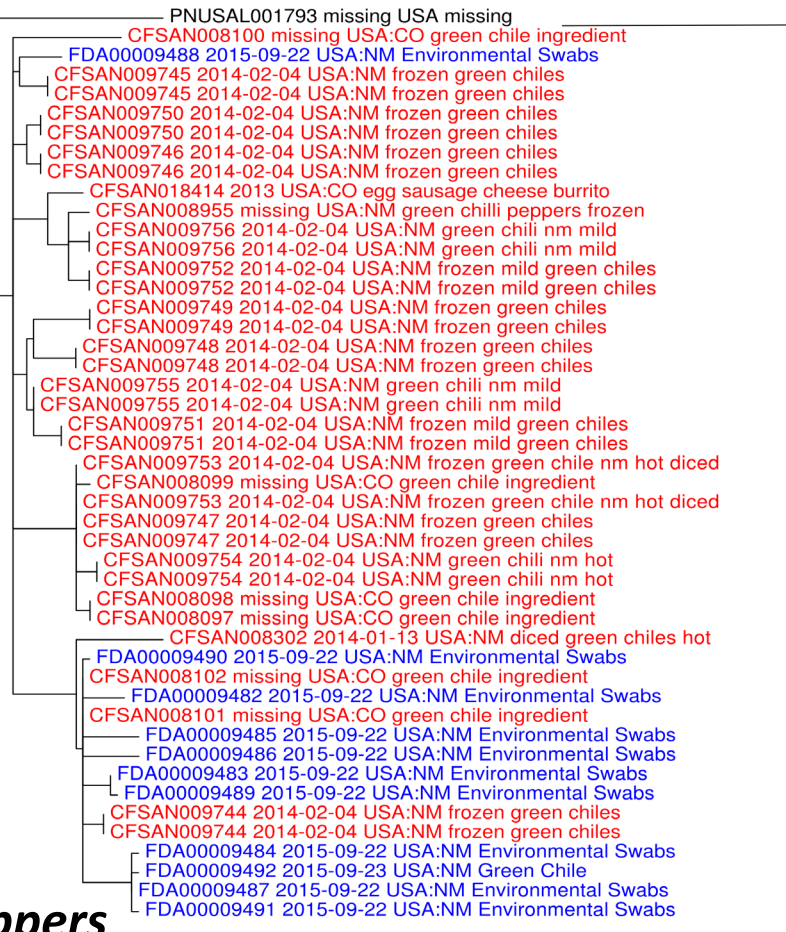
- **Did contamination originate on peppers, or during manufacturing process? Was roasting/steaming controlling for *Listeria*?**
- **Contamination originating from agricultural/farm environments is generally diverse, we would expect to see multiple clones.**
- **Extremely low level of genetic diversity observed using WGS suggests contamination coming from facility.**





Clone A

*L. monocytogenes*  
from 2014 and 2015 are  
virtually identical by WGS,  
contamination originated  
from same source



Red = 2014

Blue = 2015

**Example: Hot Peppers**

# What Can WGS Tell Us?

1. Are bacteria found in food/environmental samples a “match” to clinical isolates?
  - Match may not imply causation, contaminated ingredient
2. Endemic contamination in a facility
  - Bacteria isolated are identical over time, before and after cleanup efforts
3. Source of contamination
  - Database contains isolates from different countries, regions, states
4. Anti-Microbial Resistance, Virulence, Pathogenicity



*“A COMPASS THAT POINTS TRUE NORTH and a TELESCOPE  
FOR THE DEEPEST OF SPACE”*



S. Montevideo in salami (2009):

***“Sorting through the ingredients”***

***“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”***

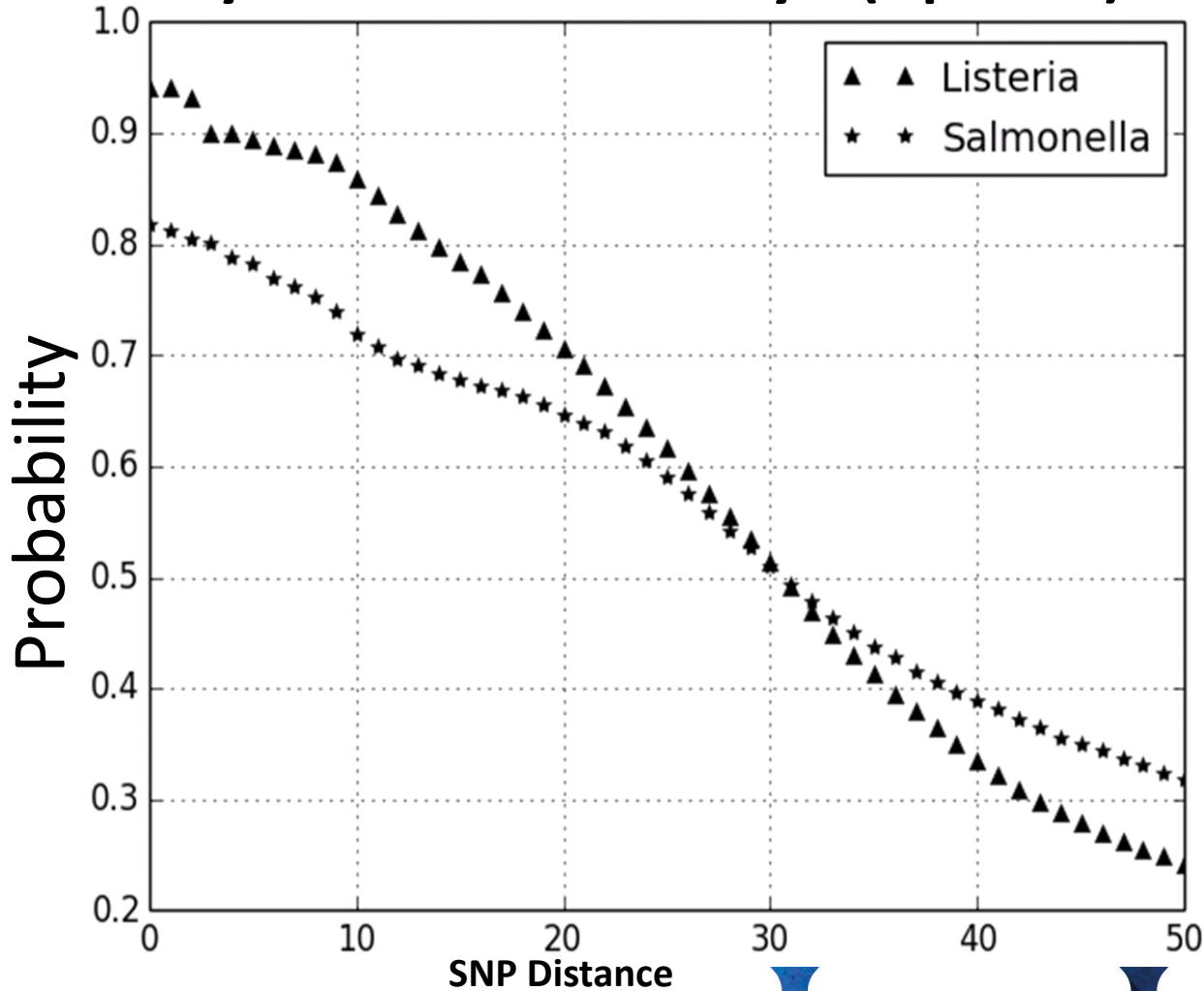
S. Enteritidis in shell eggs (2010):

***“Sourcing down to the farm”***

S. Braenderup in nut butter  
(2014):

***“Bypassing a Food Vehicle  
Altogether”***

# Facility Match Probability $P(F | D \leq d)$



SNP cutoff	<i>Sal.</i>	<i>Lis.</i>
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.



©2011, Illumina Inc. All rights reserved.



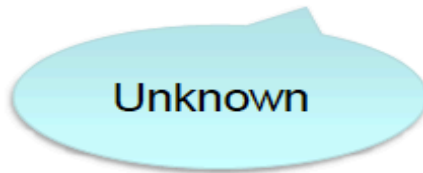
## THE MANY APPLICATIONS



# One Data Record - Many Possibilities



.....AAGCTTGGAGATCTACGTGTACCTAGTCGAAGCTA....



**THUS, Very important to ensure complete and open access to the WGS data**

# Applications of WGS in the Food Safety Environment



*“The right key to open the lock”*

- ❖ **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***



*“The right tool for the right job”*

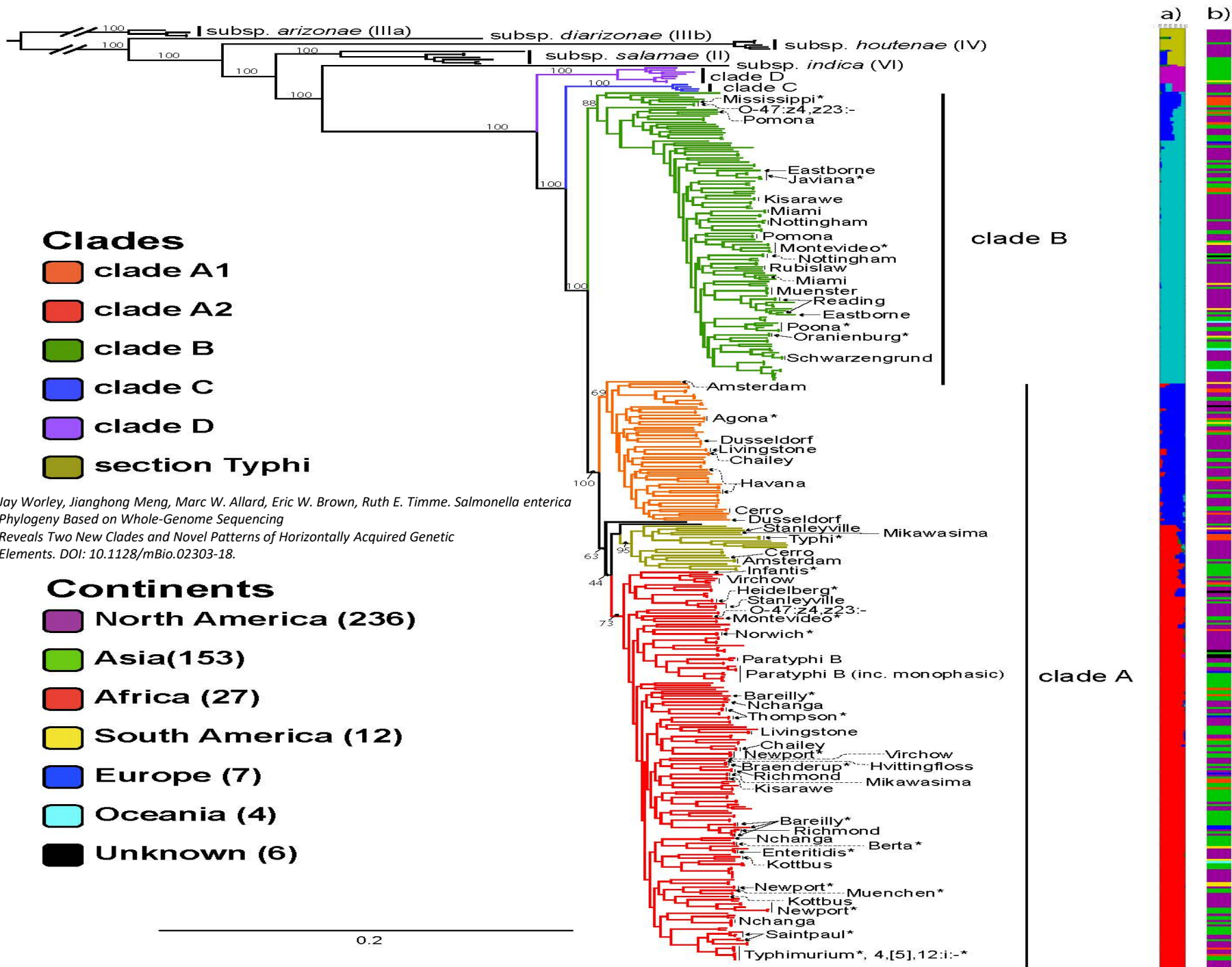


## SeqSero Salmonella Serotyping by Whole Genome Sequencing

- [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:





Jay Worley, Jianghong Meng, Marc W. Allard, Eric W. Brown, Ruth E. Timme. *Salmonella enterica* Phylogeny Based on Whole-Genome Sequencing Reveals Two New Clades and Novel Patterns of Horizontally Acquired Genetic Elements. DOI: 10.1128/mBio.02303-18.

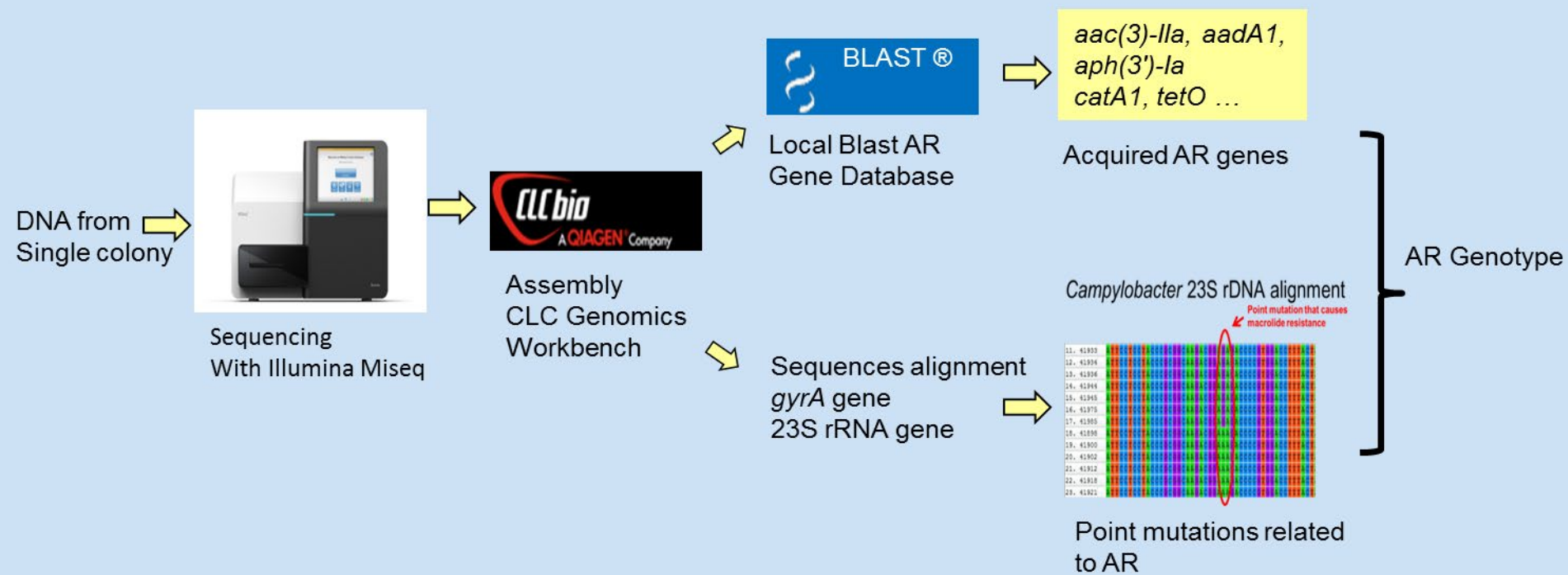


# NARMS

National Antimicrobial Resistance Monitoring System  
 Food and Drug Administration • Centers for Disease Control and Prevention • United States Department of Agriculture



## From WGS to Antibiotic Resistance Genotype



# RESISTOME TRACKER

## Salmonella

Select an icon or alert below to get started.



CUSTOMIZE



COMPARE



DISCOVER

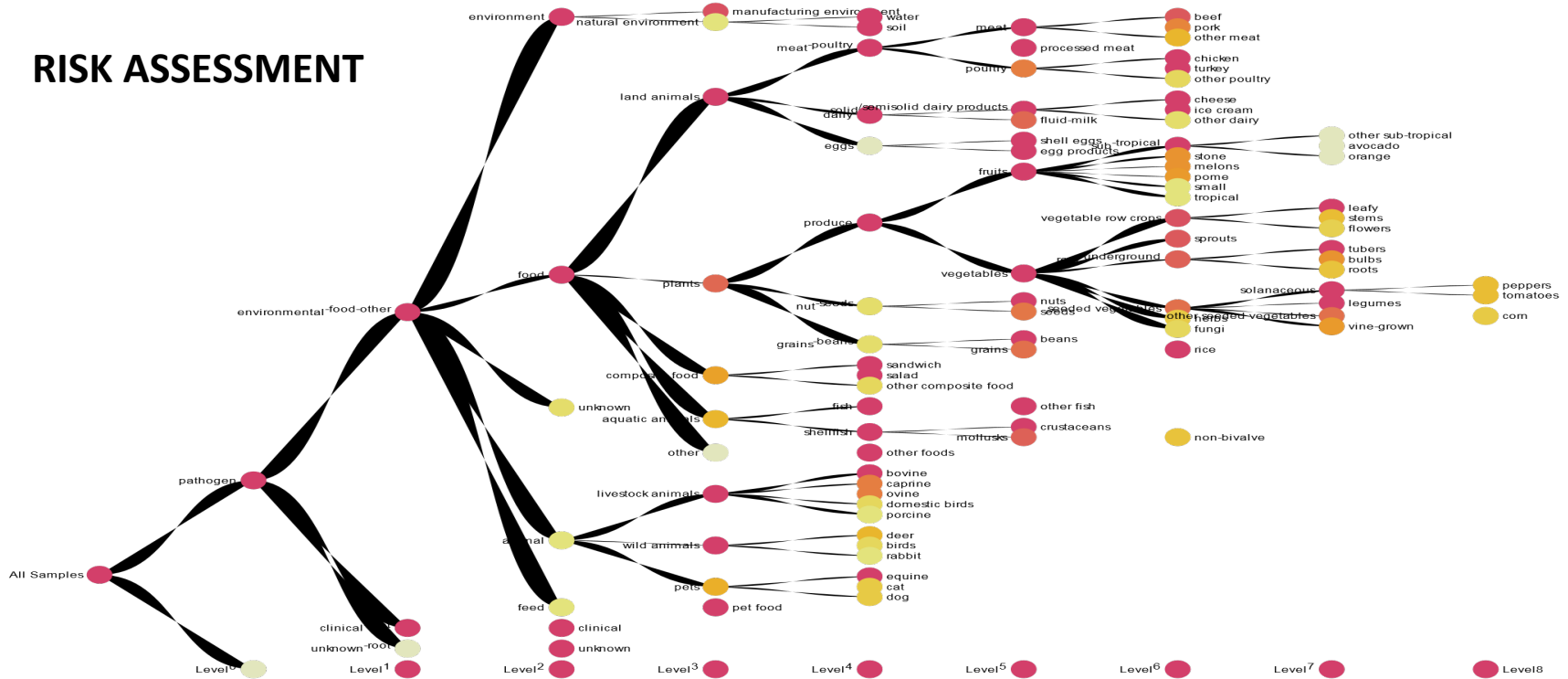


EXPLORE

### ALERTS

There are **11** records with **2** flagged genes uploaded in the last 30 days

# RISK ASSESSMENT



**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.

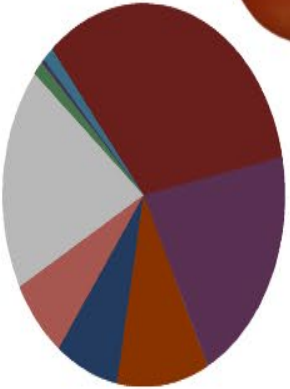
**Dominant Bacterial Families**  
**CA tomato carposphere**  
 16S rRNA gene V1-V2

*Baseline survey of the anatomical microbial ecology of an important food plant: Solanum lycopersicum (tomato). Ottesen AR, González Peña A, White JR, Pettengill JB, Li C, Allard S, Rideout S, Allard M, Hill T, Evans P, Strain E, Musser S, Knight R, Brown E.; BMC Microbiol. 2013 May 24;13:114. doi: 10.1186/1471-2180-13-114.*

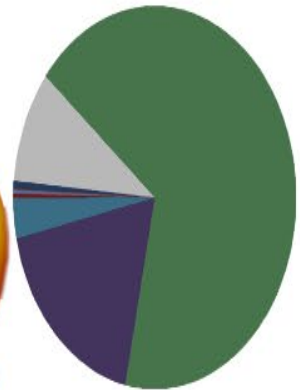
**Dominant Bacterial Families**  
**MD, VA and NC tomato carposphere**  
 16S rRNA gene V1V2



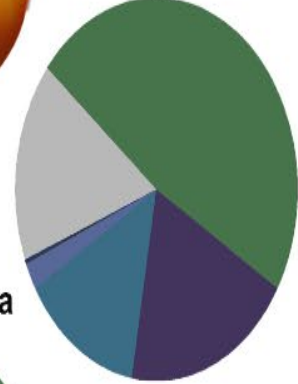
**California**



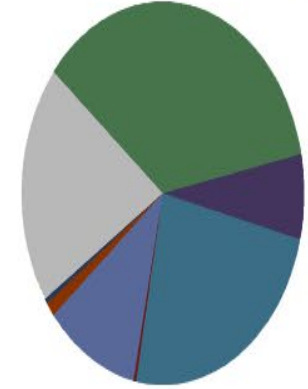
**Maryland**



**Virginia**



**North Carolina**



- Sphingomonas*
- Rhizobium*
- Pseudomonas*
- Erwinia*
- Tatumella*
- Xanthomonas*
- Enterobacter*
- Bacteria, unassigned*
- Lactococcus*
- Other Genera < 5%*

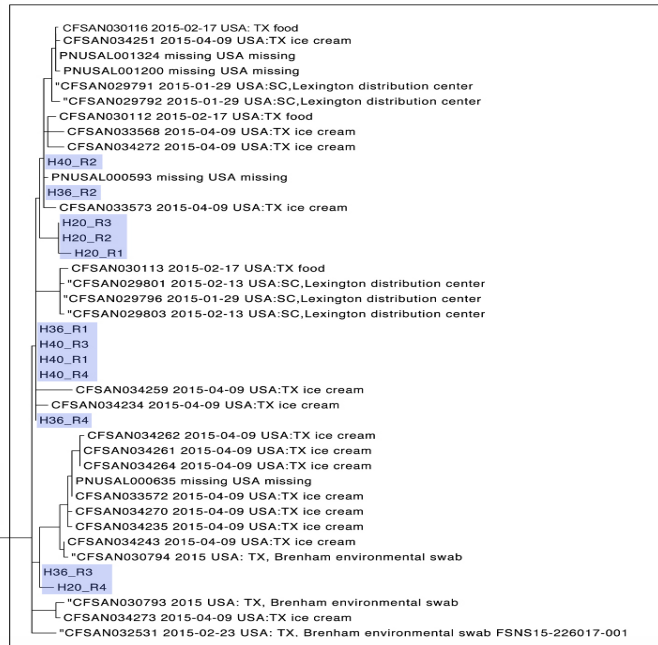
# Genomes from different hours of enrichment (H20, H36, & H40) cluster with WGS of pure isolates from ice cream listeriosis outbreak



*Lm* strains from h 20, 36 and 40 of enrichments

**H 20** —

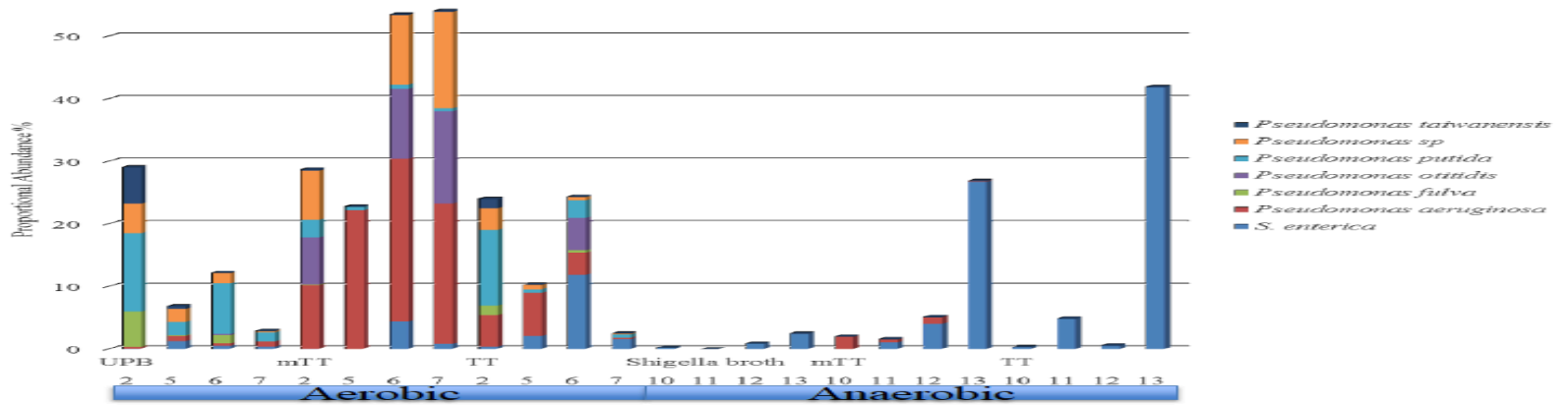
**H 36, 40** —



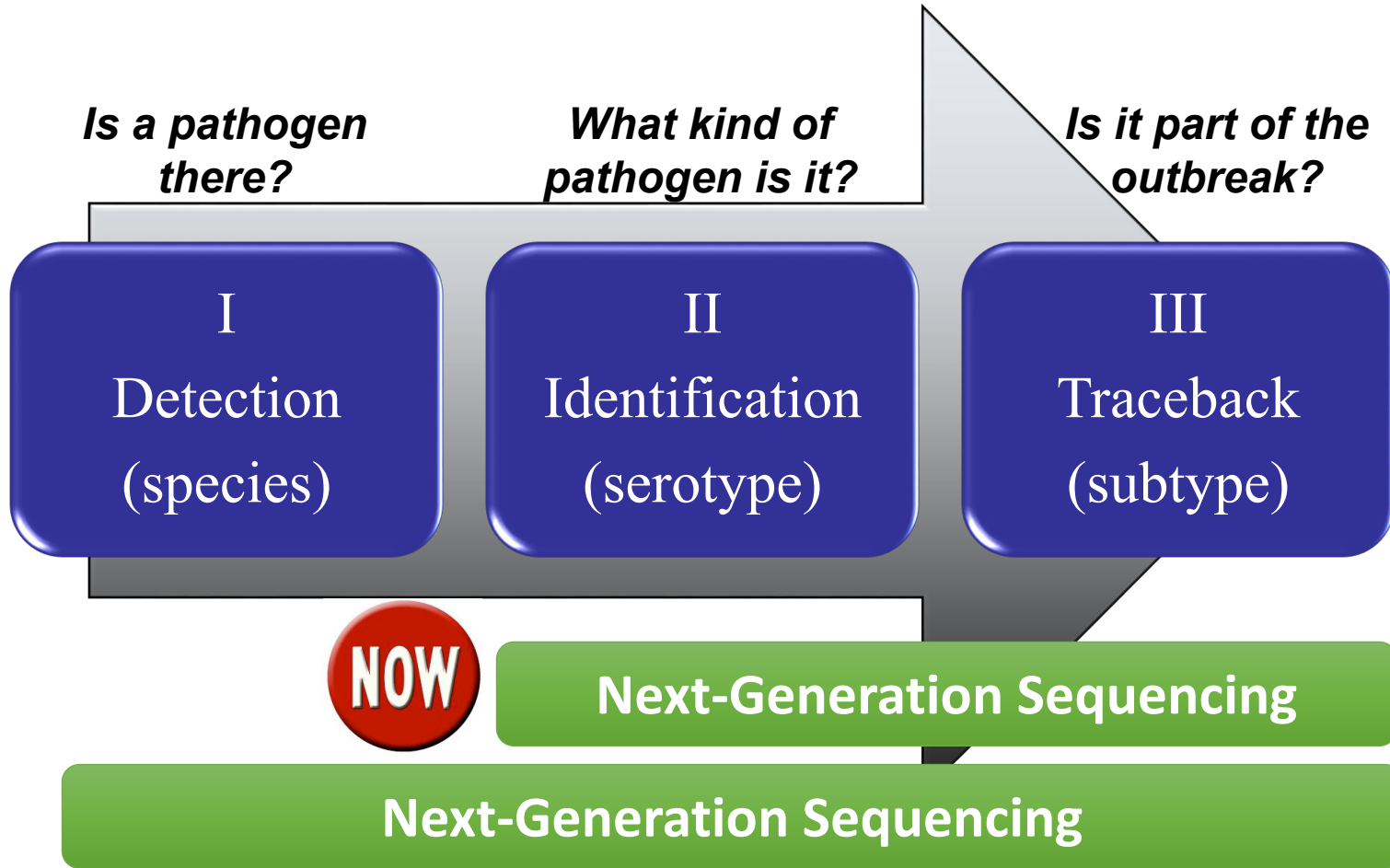
*Lm* strains not associated with the outbreak

**quasiMetaGenomic Sequencing...**

## QUASI-METAGENOMIC ANALYSIS FOR SALMONELLA



## Investigating Food Contamination Events with OMICS Approaches





## 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.**

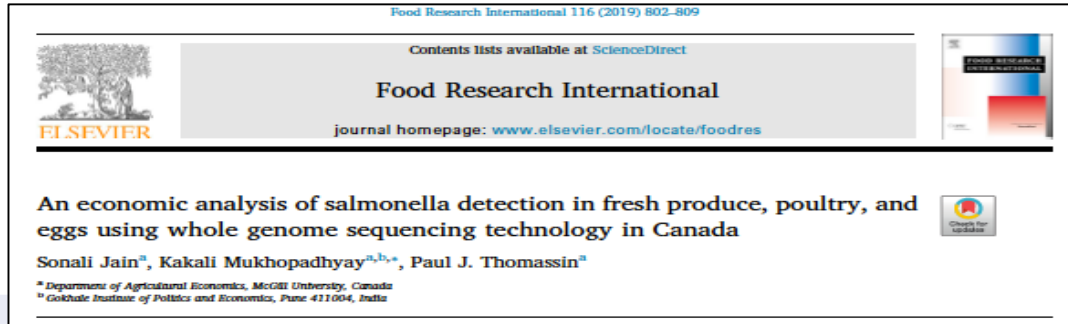


## ***Salmonella* Adaptations of particular interest to food safety specialists:**

- (1) Thermal tolerance
- (2) Dessication resistance
- (3) Osmotic/Ionic tolerance
- (4) Quat resistance
- (5) Chlorine resistance
- (6) Biofilm persistence
- (7) Surface adherence
- (8) Antibiotic resistance
- (9) Antimicrobial resistance
- (10) Ecological fitness
- (11) Heavy metal resistance
- (12) Metabolic persistence
- (13) Enhanced hydrophobic fitness
- (14) Produce invasiveness
- (15) Flower invasiveness
- (16) Root system invasiveness
- (17) Acid resistance
- (18) Surface water fitness
- (19) In vivo plant migratory fitness
- (20) Soil fitness
- (21) Capsaicin resistance
- (22) Swarming
- (23) Trans-ovarian poultry colonization
- (24) Fecal persistence (poultry)
- (25) Yolk content invasion
- (26) Multidrug resistance
- (27) External amoeba harborage
- (28) Internal amoeba harborage
- (29) Acyl-homoserine lactone (AHL)
- (30) KatE stationary-phase catalase
- (31) In vivo migratory fitness
- (32) RDAR phenotype
- (33) The 'Weltevreden' type
- (34) Persistence within the tomato\*\*

**GOAL = <5 years have first 25 mapped**

# Economic IMPACT on Foodborne *Salmonella* When Using WGS



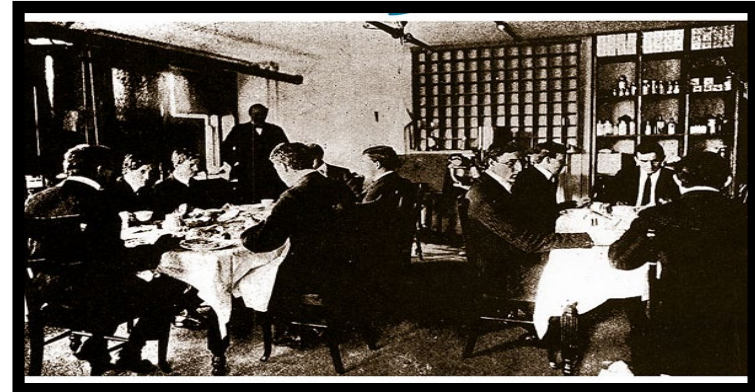
	Canada	United States
Incidence of illness	47,028	1,200,000
Costs to adopt WGS	\$158,340,000	\$100,000,000
QUALY lost	469.75	16,782
Total illness costs	\$287,770,000	\$3,300,000,000
<b>Total net benefit of adopting WGS</b>	<b>\$90,250,000</b>	<b>\$1,000,000,000</b>

\*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. Based on Jain et al., 2019 An economic analysis of salmonella detection in fresh produce, poultry, and eggs using whole genome sequencing technology in Canada. Food Res. Int. 116: 802-809.

## *FDA circa 1906 –FOOD SAFETY*

### PRESENTATION POINTS

Various desktop NGS platforms now exist and are dropping precipitously in price – and per reaction cost making the technology largely accessible for public health applications.



WGS is fast becoming an integral part of the science of food safety, both for more comprehensive characterization and testing of foods and for providing insight about the scope and sources of outbreaks and other food contamination events. Development of international open source databases will empower WGS for sentinel surveillance work on a global scale.

WGS, as part of a laboratory next-generation analysis pipeline, can augment food safety investigations, particularly in cases where strain homogeneity is a problem, by (i) delimiting the scope of a contamination event; (ii) affirming a cluster from common background genotypes; and (iii) source-tracking by comparative genomics of food and environmental isolates of *Salmonella*. Recall, sequences are agnostic.

## 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 “Carlos 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**