Establishing a National Surveillance Network for Foodborne Pathogens Based on Whole Genome Sequencing

Steven Musser, Ph.D.

Deputy Center Director for Scientific Operations
Center for Food Safety and Applied Nutrition, FDA

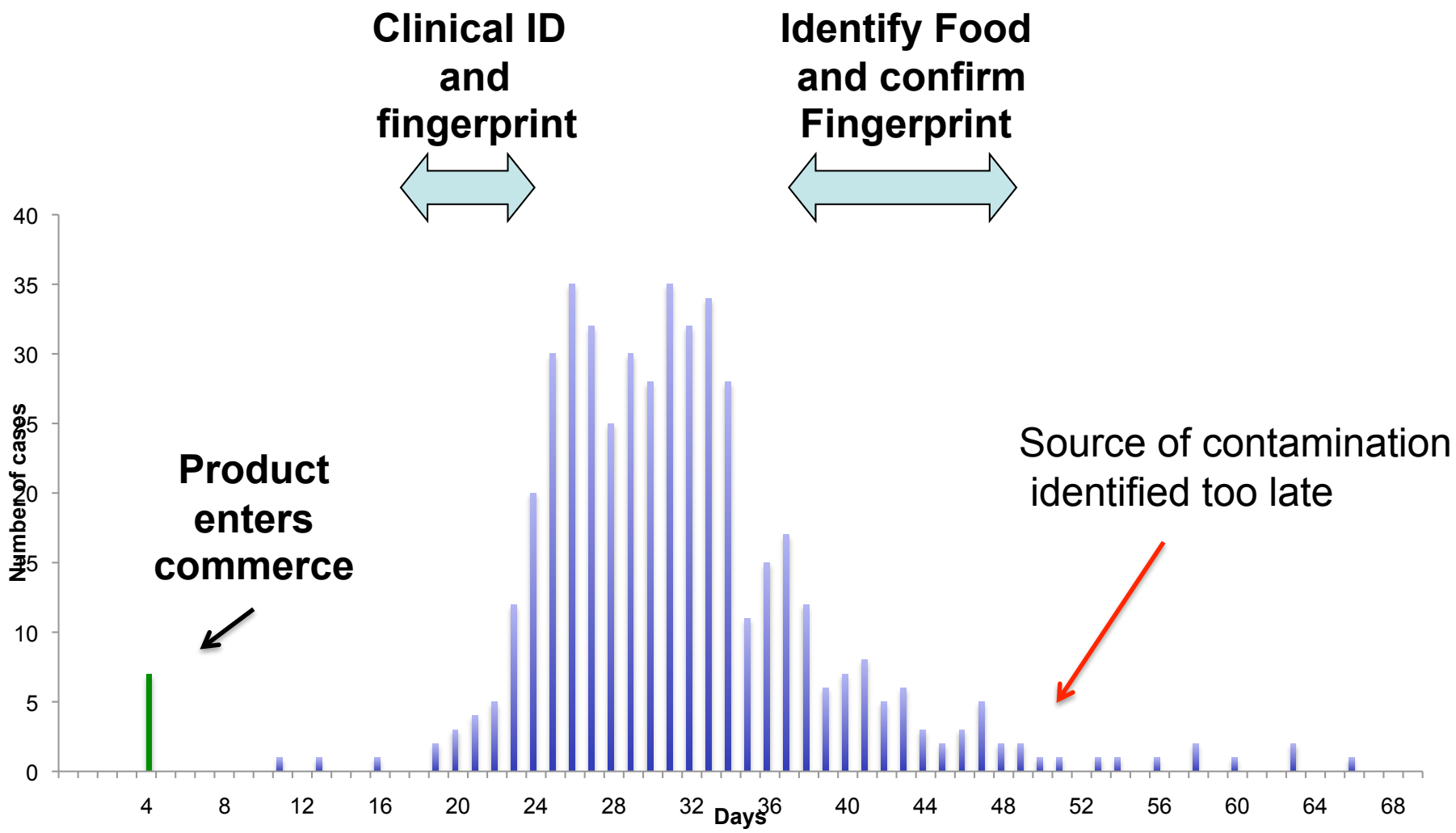
Next-Generation Sequencing for Food Pathogen Traceability
UCD Institute of Food and Health *in conjunction with* UCD Centre for Food Safety and the Food Safety Authority of Ireland - March 24, 2014

Foodborne Illness in the US

- Each year 9.4 million episodes of foodborne illness in the United States
- 55,961 hospitalizations
- 1,351 deaths
- *Salmonella spp. cause* 11% of foodborne illnesses each year

(Scallan et al. 2011 Emerging Infectious Diseases • www.cdc.gov/eid).

The Public Health Need



Some perspective on the food supply

- **Tracking and Tracing of food pathogens**
  - Almost 200,000 registered food facilities (2/14)
    - 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 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.

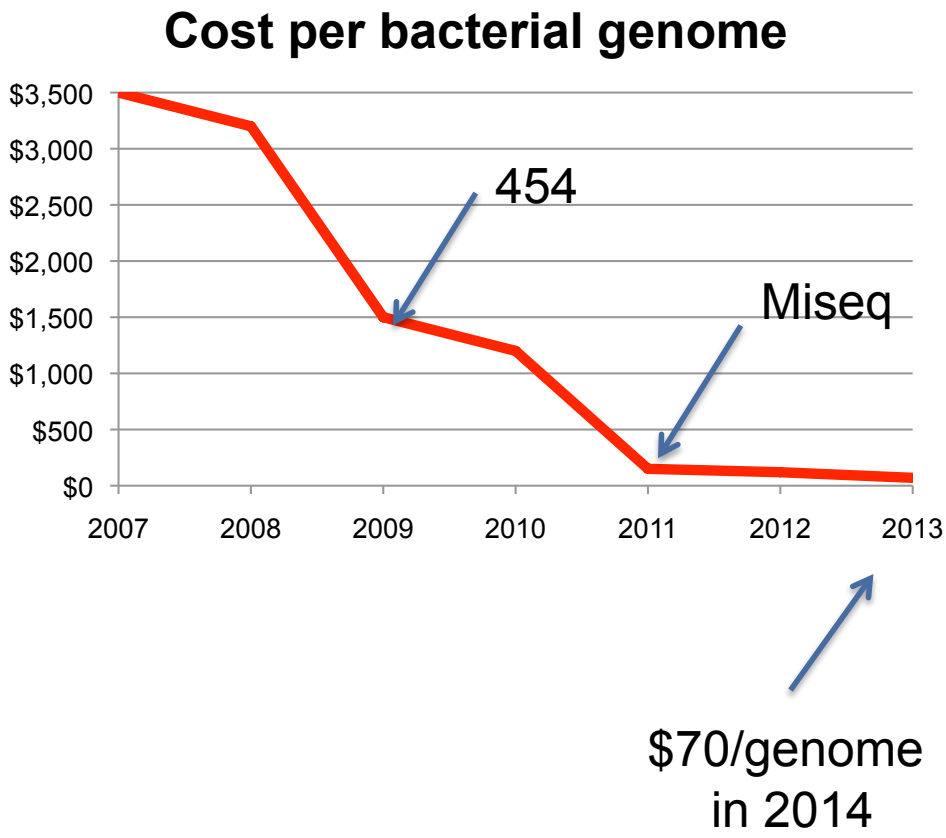
NOTE: Countries are listed in alphabetical order and not by volume of export.

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

www.iom.edu

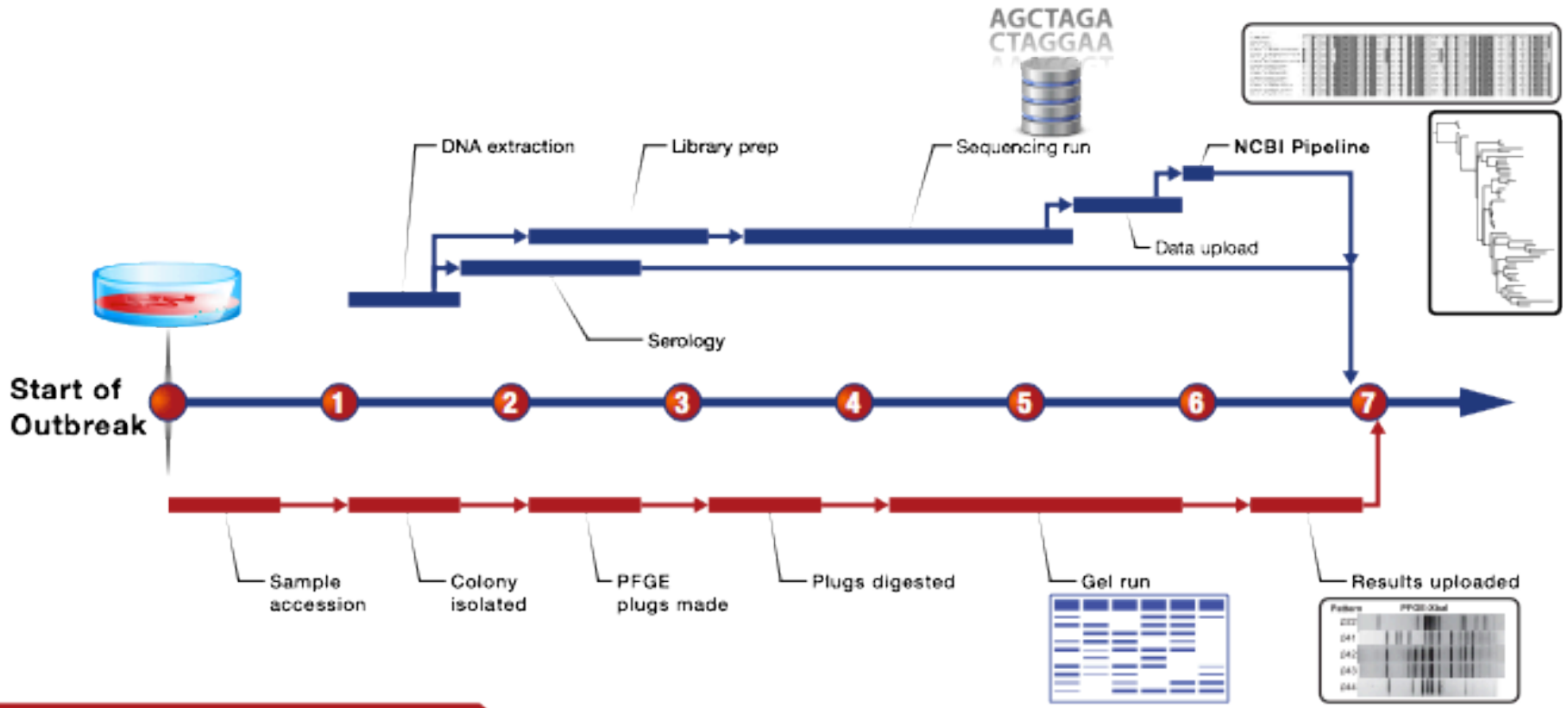
Is WGS a viable solution?

- Cost
- Increasing ease of operation
- Database longevity
- Sample prep
  - Identical for all pathogens
- Cost savings
  - Resistance, subtyping, virulence factors, more…
- New applications
  - tracking, regulatory/compliance actions, historical trends, more…



Next-generation Lab Response vs. Conventional Lab Response

Next-Generation Lab Response



Conventional Lab Response

1 = Day 1

11/27/2013

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"***

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

Can WGS fill a Public Health role?

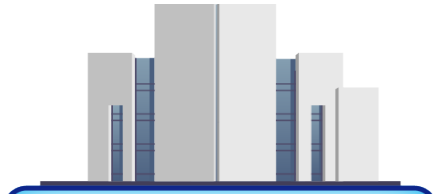
- **If yes, then…**
  - Initiate pilot study
  - Develop collaborations and partnerships
    - NCBI, States, CDC and other Federal partners
  - What infrastructure would be needed?
  - Support multiple sequencing platforms?
    - Multiple data formats
    - How reproducible are the data AND answers?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

Metadata

- **Simple but complete for each Strain**
  - Clinical or environmental (specific source)
    - Environmental swab or type of food
  - Location as accurate as allowable
    - State, Region, Country
  - Submitter name – Usually organization
  - Date of isolation

Network Requirements

- Well characterized strain sets
- A large database of sequences with accurate metadata
- A network of sequencing labs
- Analytical software
- Somewhere to store the data



FDA, USDA, CDC

State, Local, Federal and Foreign Public Health Agencies

Academia

DATA ANALYSIS

NCBI, EMBL DDBJ
(Public Access Database)

GenomeTrakr

DATA ASSEMBLY AND STORAGE

High Capacity Sequencing of Collections

1000 GENOME PROJECT

Network of Sequencers

DATA ACQUISITION

FDA provides

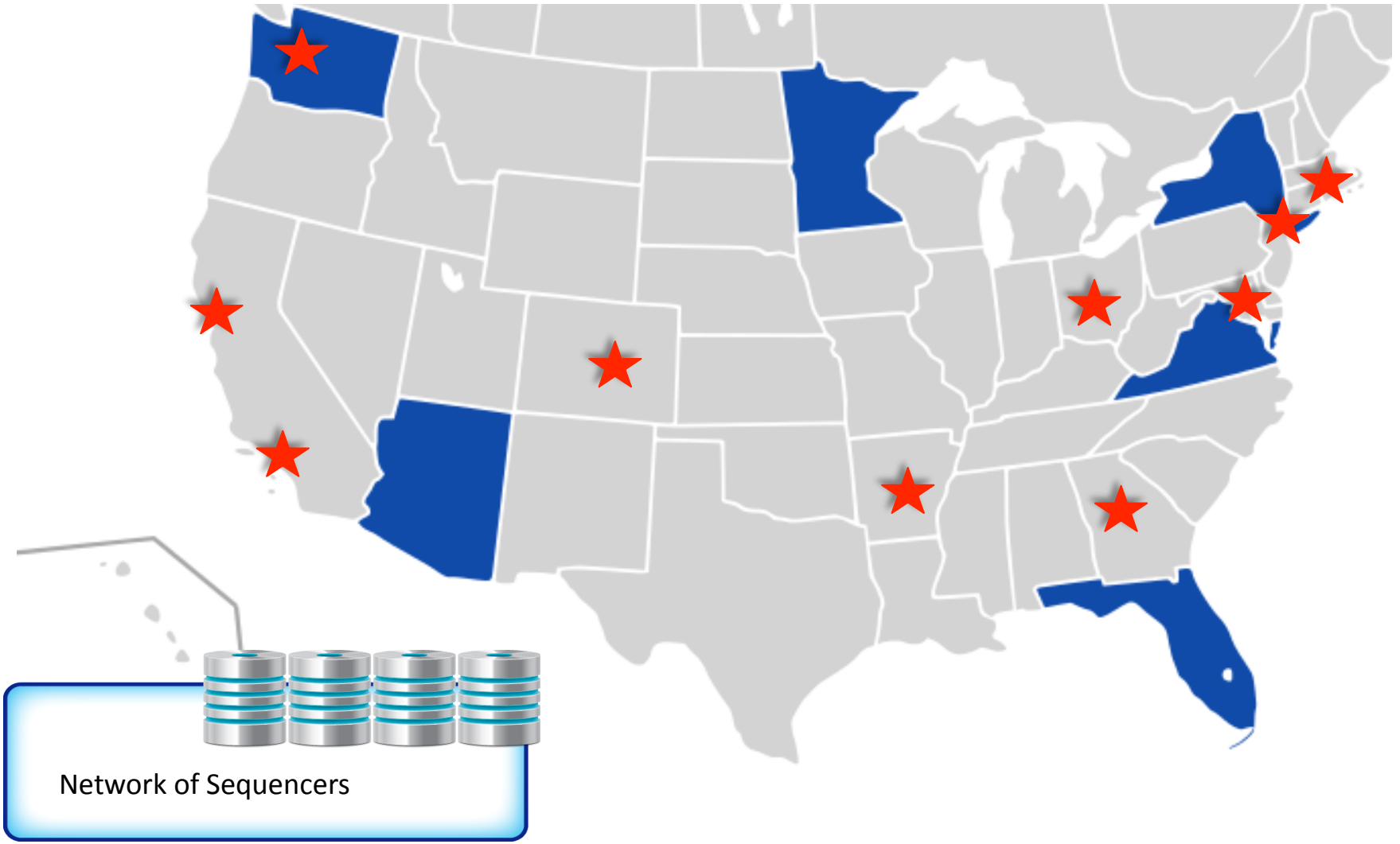
- 1 Miseq system
- Sufficient reagents to sequence > 300 genomes per year
- Dedicated scientific staff (bioinformatics and/or laboratory support) through Oak Ridge Institute for Science and Education (ORISE)
- Bioinformatics and laboratory support, analysis pipeline

Network Lab provides

- Minimum ~300 genomes with metadata uploaded to NCBI per annum, minimum 20X coverage
- food and environmental related bacterial (prefer *Salmonella*) isolates

Cost to FDA ≈ $200k/lab

7 state health depts.
+ 10 FDA-ORA



FDA-State Desktop Pilot called GenomeTrakr

http://www.ncbi.nlm.nih.gov/bioproject/183844

Food and Drug Administration, Center for Food Safety and Applied Nutrition: GenomeTrakr Project

Accession: PRJNA183844 ID: 183844

Currently encompasses whole genome sequencing of cultured pathogens as part of a surveillance project for the rapid detection of outbreaks of foodborne illnesses

Project Type: Umbrella project (***Subtype:*** Disease)

Relevance: Agricultural, Medical, Industrial, Environmental

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
Nucleotide (total)	31172
WGS master	194
Genomic DNA	30978
SRA Experiments	437
Protein Sequences	901056
PUBLICATIONS	
PubMed	12
PMC	9
OTHER DATASETS	
BioSample	1104
GEO DataSets	1

Project Type	Number of Projects
Genome sequencing	
Highest level of assembly :	
SRA or Trace	1
No data links	4
Total	5

BioProject accession	Assembly level	Name	Title
PRJNA183847	-	GenomeTrakr Project: Arizona State Public Health Laboratory	GenomeTrakr Project: Arizona State Public Health Laboratory (Arizona State Public Health...)
PRJNA183848	-	GenomeTrakr Project: Florida Department of Health	GenomeTrakr Project: Florida Department of Health (Florida Department of Health)
PRJNA186035	SRA or Trace	GenomeTrakr Project: Food and Drug Administration, Center for Food Safety and Applied Nutrition	GenomeTrakr Project: Food and Drug Administration, Center for Food Safety and Applied Nutrition (Center for Food Safety and...)
PRJNA183850	-	GenomeTrakr Project: New York State Dept. of Health, Wadsworth Center	GenomeTrakr Project: New York State Dept. of Health, Wadsworth Center (New York State Dept. of Health...)
PRJNA183851	-	GenomeTrakr Project: Washington State Department of Health Public Health Laboratory	GenomeTrakr Project: Washington State Department of Health Public Health Laboratory (Washington State Department...)

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

Expanding the network

Partners with sequencers

United Kingdom - FSA
Canada – CFIA and PHAC
Argentina - WHO
Taiwan

Partners with isolates

Ireland
Mexico
Turkey
Columbia
Chile
Brazil
Thailand
Ethiopia

State Partners

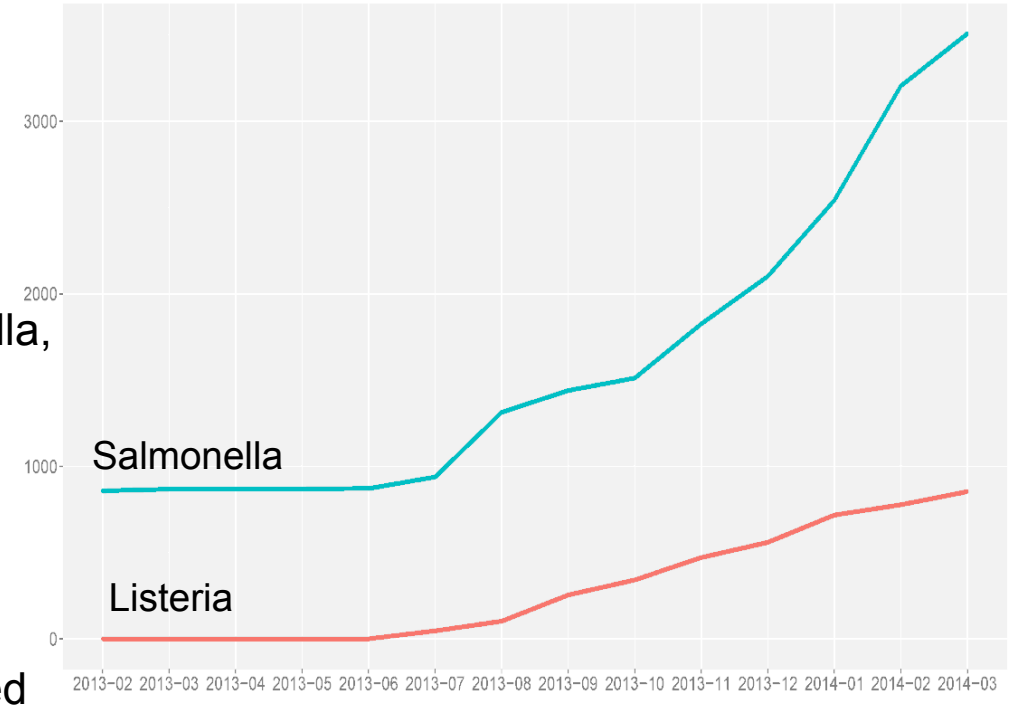
6 States have requested funding

Organizations/Countries joining the network

APHL
WHO
USDA
GMI
Italy
Germany
Denmark
Australia
Spain

Now What?

- NGS clearly defines foodborne outbreaks – more than 15 different examples
- NGS network is reliable, efficient and can provide very good location specificity of outbreaks
- We have sequenced about 2900 Salmonella, more than 900 Listeria, and closed 100 genomes. Our current rate is about 500 sequences a month.
- The need for increased number of well characterized environmental (food, water, facility, etc.) sequences may outweigh need for extensive clinical isolates
- Many requests for information or help from other public health labs



Needs/concerns

- Network security issues
  - Sequencers
  - Software
- Improved informatics and software development
  - Widely available commercial solutions
  - Custom solutions
  - Automated identification of AMR, virulence markers, etc
- Cloud computing and access to HPC
- Data presentation to different groups
  - Physicians
  - Epidemiologists
  - Researchers

FDA -CFSAN
Marc Allard Rebecca Bell
Eric Brown Andrea Ottesen James Pettengill
Ruth Timme Jie Zheng Charlie Wang
Christine Keys *Cong Li*
Errol Strain Yan Luo
Mark Mammel Darcy Hanes
FDA Division of Field Sciences Rebecca Dreisch
NYPH Bill Wolfgang Kimberly Musser and colleagues
MPH Alvina Chu and colleagues
FDH Anita Wright Judy Johnson
ADPH Victor Waddell Dave Engelthaller Paul Keim
WDH Brian Hyatt Chen Li William Glover
CDC John Besser, Eija Trees, Duncan MacCannell and colleagues
National Institutes of Health
David Lipman (NCBI) Martin Shumway (NCBI)
Tatiana Tatusova (NCBI) William Klimke (NCBI)
Illumina
Lisa Alves Susan Knowles Omayma Al-Awar and colleagues
CLC Bio David Michaels Cecilia Boysen and colleagues

Questions