

Current Status of Whole Genome Sequencing in the Human Foods Program at FDA

Steven M. Musser, Ph.D.

Deputy Director, Center for Food Safety and Applied Nutrition,
US Food and Drug Administration

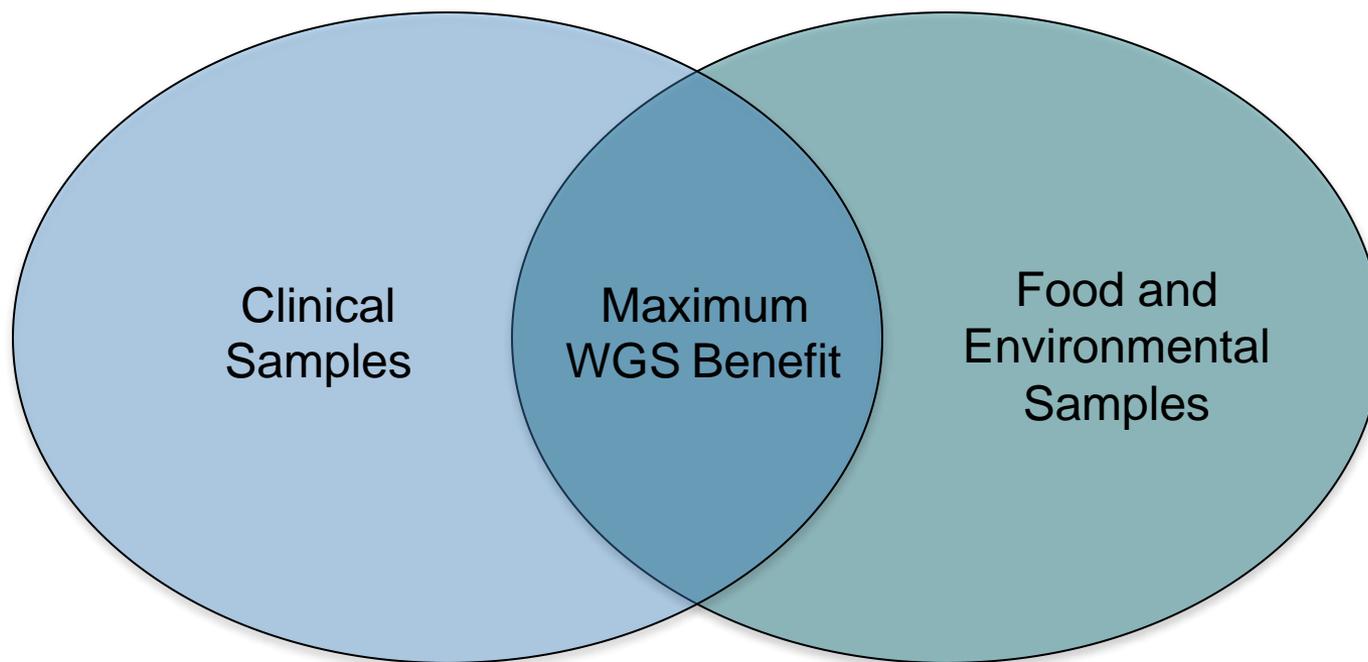
Use of Whole Genome Sequence (WGS) Analysis to Improve Food Safety and Public Health
U.S. Department of Agriculture, Washington, DC October 26-27, 2017

What is GenomeTrakr

- A distributed network of laboratories that utilize whole genome sequencing for pathogen identification. Global database of sequence data from all sources, public health, government, private, academic, etc.
- Network was established by FDA in 2011 and is growing rapidly. The network includes CDC, FSIS and dozens of collaborating institutions around the world.
- Open access data curation, daily data analysis and public data visualization tools are provided by the National Center for Biotechnology Information (NCBI)



Importance of a Balanced Approach



Current Scope of the GenomeTrakr Network

- Network includes labs at FDA, CDC, FSIS, 19 state agriculture, health and university labs, 1 U.S. hospital lab, and 17 labs located outside the U.S.
 - Contributing labs are on 4 continents and in 10 countries
- The network provides high resolution genomic sequences of food pathogens, ex. *Salmonella*, *Listeria*, STEC's, others. Greater than 150,000 sequences in the database
- Partnered with CDC and FSIS in 2012 to study all clinical and environmental isolates of *Listeria monocytogenes*

FDA GenomeTrakr website

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

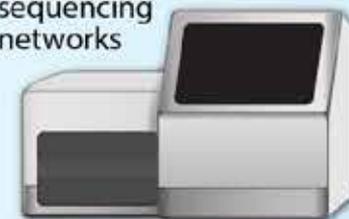
Basic Data Flow for Global WGS Public Access Databases

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



Metadata

Food/environmental Submission

Pathogen: environmental/food/other sample from

Listeria monocytogenes

BioSample: SAMN02709234;

Sample name: FDA00007620

Organism Listeria monocytogenes

cellular organisms; Bacteria; Firmicutes; Bacilli;

Bacillales; Listeriaceae; Listeria

Pathogen: environmental/food/other; v.1.0

strain

FDA00007620

host disease

missing

latitude and longitude

missing

collection date

3/26/2012

isolate

CFSAN003790

geographic location

Italy

isolation source

moliterno al tartufo
cheese

collected by

FDA

PFGE_SecondaryEnzyme_pattern

GX6A12.0280

PFGE_PrimaryEnzyme_pattern

GX6A16.0085

Submission

FDA, Justin Payne; 2014-03-27

ID: 2709234

Clinical Submission

Pathogen: clinical or host-associated sample
from **Salmonella enterica**

BioSample: SAMN02927343;

Sample name:/2013K-0563

Organism Salmonella enterica

cellular organisms; Bacteria; Proteobacteria;

Gammaproteobacteria; Enterobacteriales;

Enterobacteriaceae; Salmonella

Pathogen: clinical or host-associated; version 1.0

strain

2013K-0563

collected by

CDC

collection date

Missing

geographic location

USA

host

Missing

host disease

Missing

isolation source

Missing

latitude and longitude

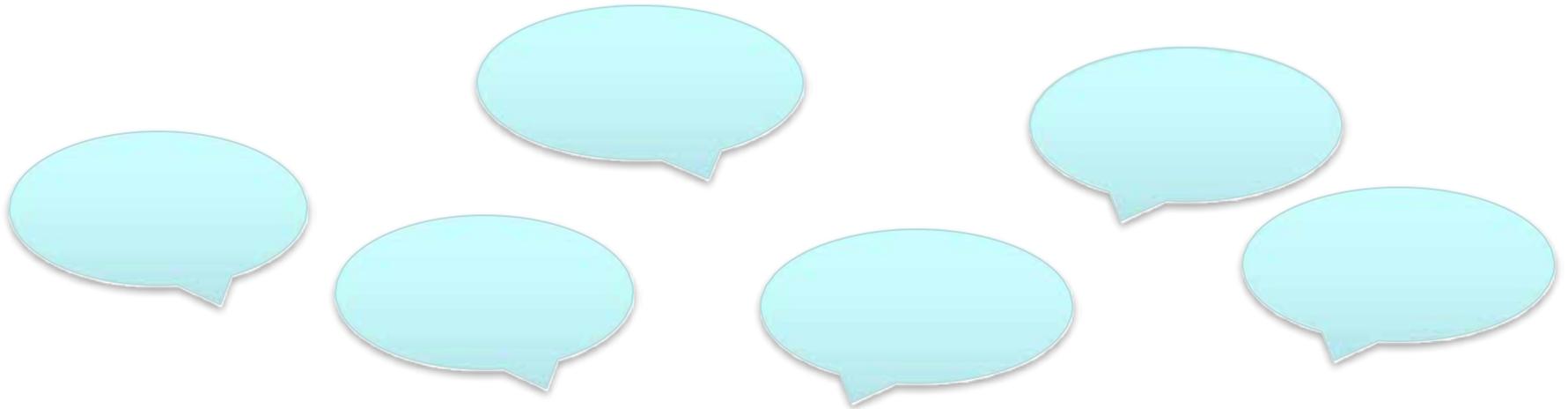
Missing

Submission

Pulsenet, Eija Trees; 2014-07-18

ID: 2927343

One Data Record - Many Possibilities



.....AAGCTTGGAGATCTACGTGTACCTAGTCGAAGCTA....

Unknown

Benefits of a WGS Approach for Food Safety

- More discriminatory and informative than PFGE
- Clues to geographic origin of pathogen
- Still relatively slow, so perhaps not fully ready for certain clinical applications

This means:

- Greater certainty when matching clinical, environmental, and product sample isolates
- Links between illnesses and the potential source of contamination can be made with fewer isolates
- Investigators can be deployed in a more targeted manner, saving resources

End Result:

- Faster identification of the food involved in the outbreak
- Potential to help reduce the number of foodborne illnesses and deaths over the long term both in the U.S. and abroad.

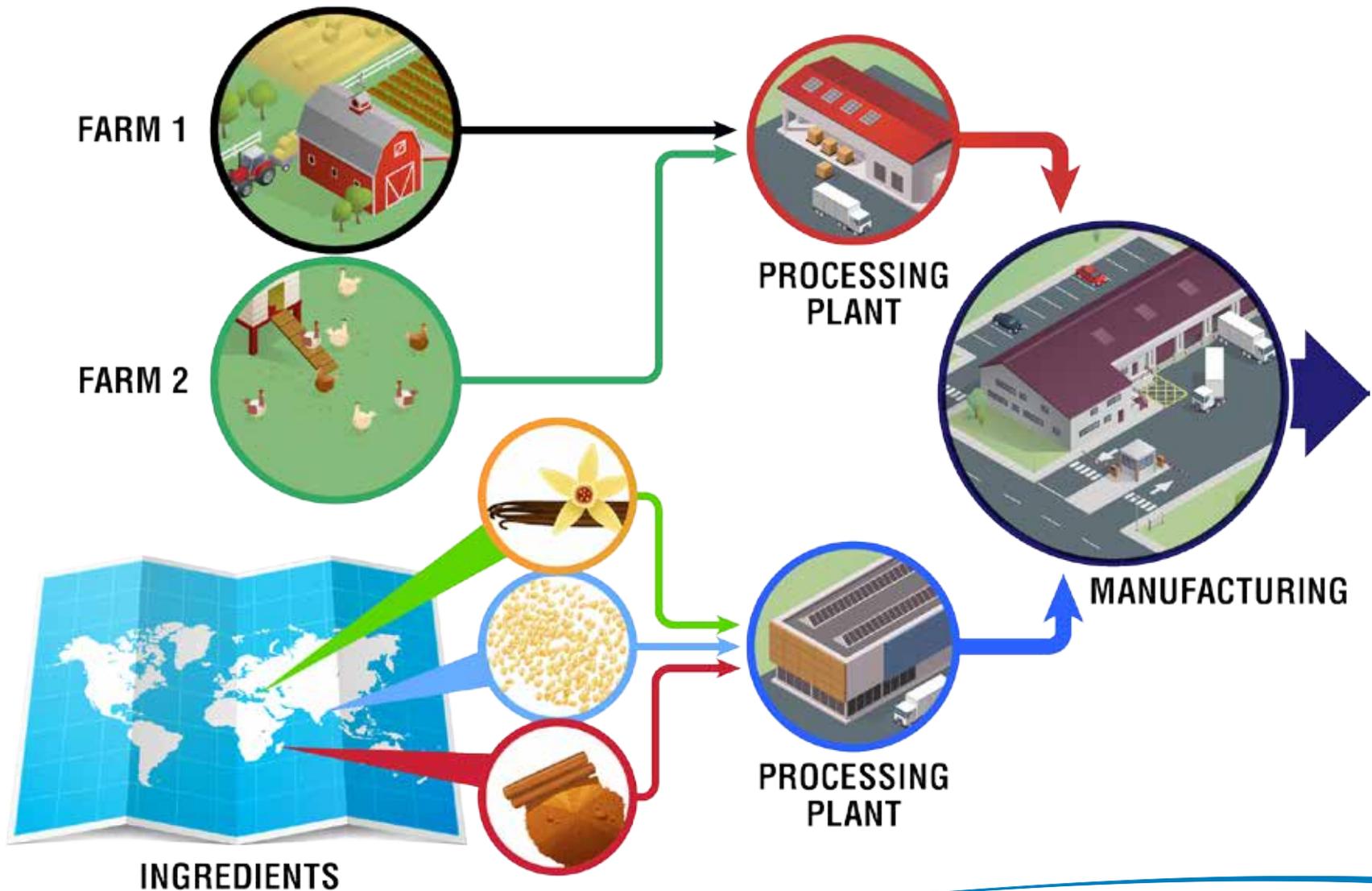
Common questions about data analysis

- What data analysis tool should be used and why (SNP's or wgMLST)?
- What does a match mean?
 - No single threshold for all species/types: rough, conservative guide
 - Inclusion: ≤ 20 SNPs match, virtually identical
- Are the data analysis tools validated?
 - CFSAN's SNP pipeline
 - Documentation: <http://snp-pipeline.rtdf.org>
 - Source Code: <https://github.com/CFSAN-Biostatistics/snp-pipeline>
- What software is available?
 - Commercial – ex. BioNumerics
 - Public Domain – ex. Galaxy
- How do I access the data analysis site at NCBI?
 - <https://www.ncbi.nlm.nih.gov/pathogens/>

What happens with a WGS “match”

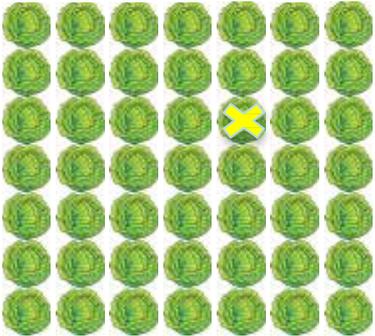
- Reanalysis of the data
- A reinspection and resampling of the facility and/or foods
- Regulatory action depends on results of inspection and follow-up environmental samples
- Supporting epidemiological evidence

Environmental sampling

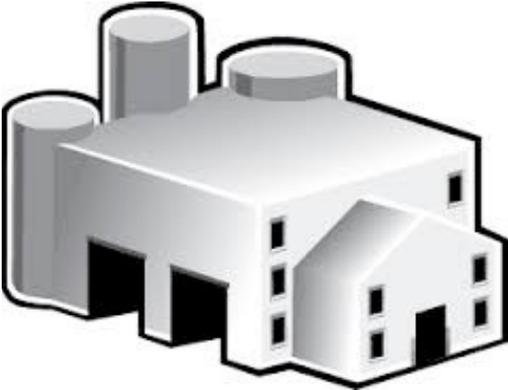
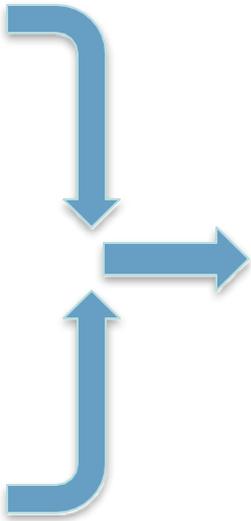
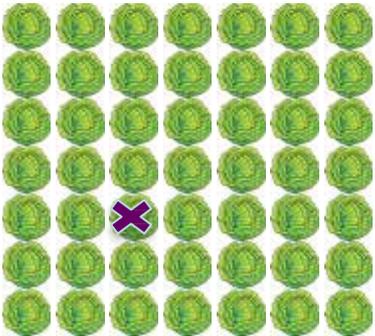


Example

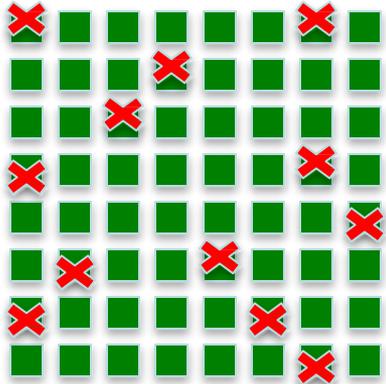
Source 1



Source 2



Processing facility



Enforcement actions

Environmental sampling with WGS

- Total number of facilities inspected
- Total number of WGS bioinformatics analyses requested in the last 12 months: 167
- May 2016- September 2017 > 600 inspections
 - 1 Suspension of FFR 415(b)
 - 18 Regulatory Meetings
 - 6 Voluntary Recalls
 - 1 Mandatory Recall
 - 7 Warning Letters
 - 3 Injunctions

Should we start using WGS?

- Contribute sequences to GenomeTrakr database
- Routine environmental or product sampling or high volume sampling
 - Probably not fast enough or cost effective at this point
- Supply chain management
 - Recommended if classical methods fail to identify the source of the contamination
- Several positive results for the same pathogen or serotype
 - Recommended to identify source or effectiveness of cleaning procedures.
- Following positive results from an inspection
 - Highly recommended

Summary

- WGS is now routine in FDA's outbreak response and compliance/surveillance activities. Internally (across our agency), and in collaboration with FSIS and CDC, WGS has now been deployed and benefitted the traceability of numerous foodborne contamination events.
- WGS can be used to inform traceback investigations and delimit the scope of food contamination events unlike ever before – not just a regulatory tool - numerous offshoot applications exist (i.e., supply chain management, quality assurance, process evaluation, etc.)
- Genome sequences are portable and instantly cross-compatible. One technology approach irrelevant of organism. Slow movement away from PFGE
- Have to balance the need for increased number of well characterized environmental (food, water, facility, etc.) sequences with the need for extensive clinical isolates
- WGS, unlike PFGE, is more than a surveillance tool. It provides information on AMR, virulence, serotype, **resistance to sanitizers**, and other critical factors in one assay, including historical reference to pathogen emergence.

Thank you