

The GenomeTrakr Experience



“Science Meets Policy” conference
Parma, Italy
September 5th, 2023



Eric L. Stevens, PhD

International Policy Analyst

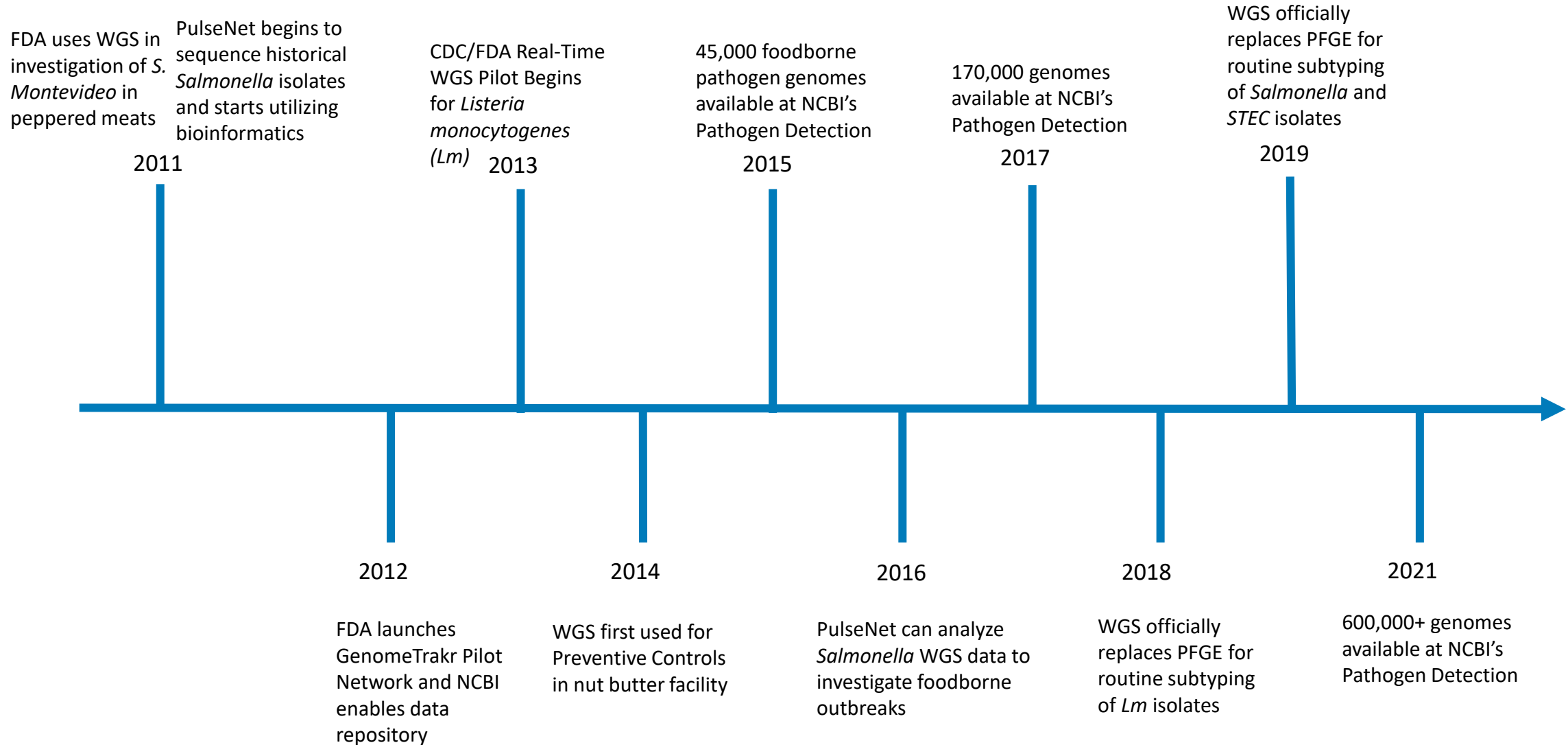
FDA’s Center for Food Safety and Applied Nutrition

Eric.stevens@fda.hhs.gov



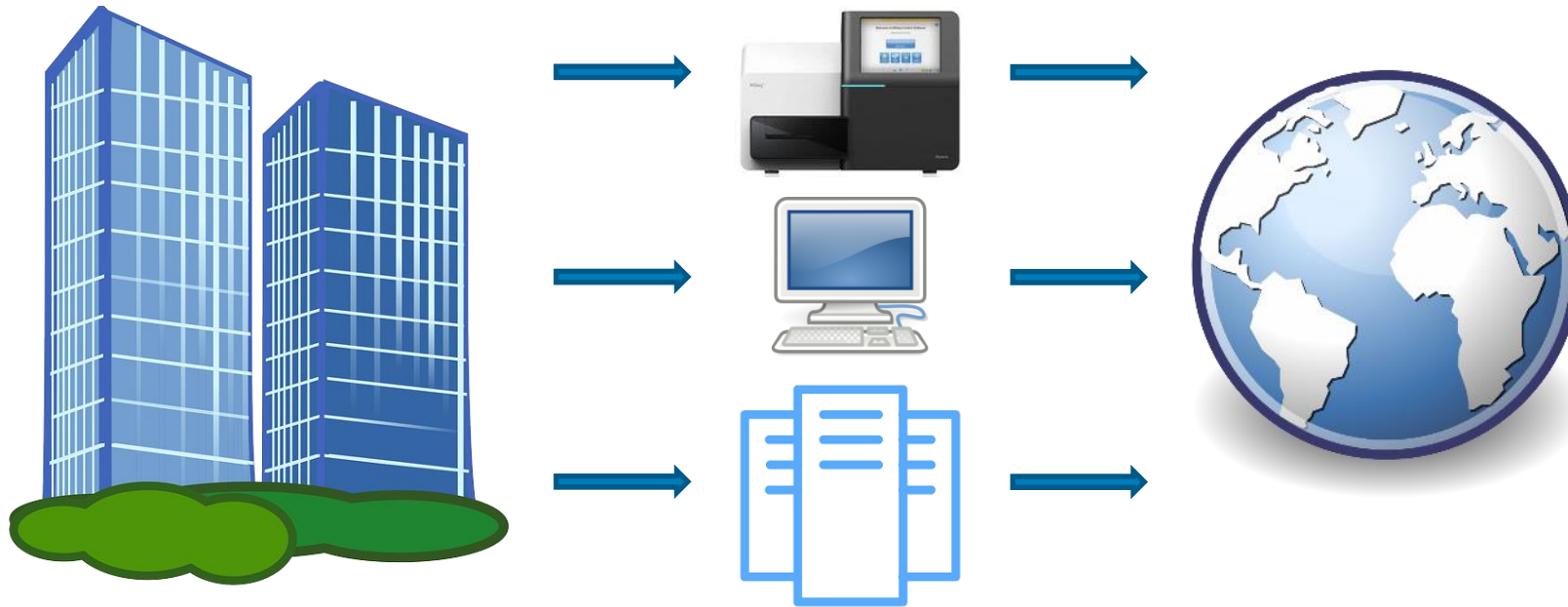
The Past: PFGE surveillance and Private databases

Big Picture: This did not happen overnight



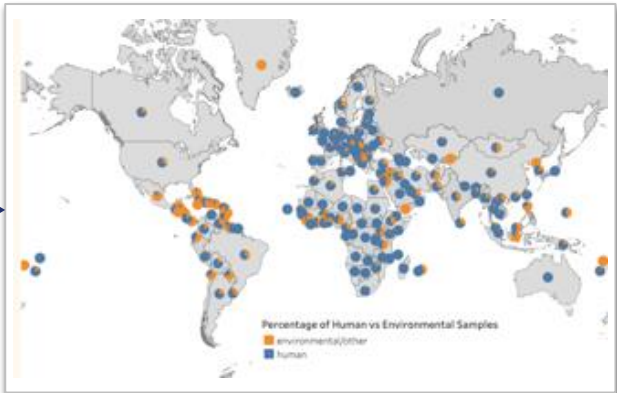
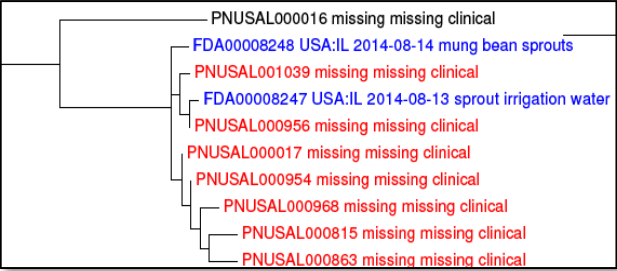
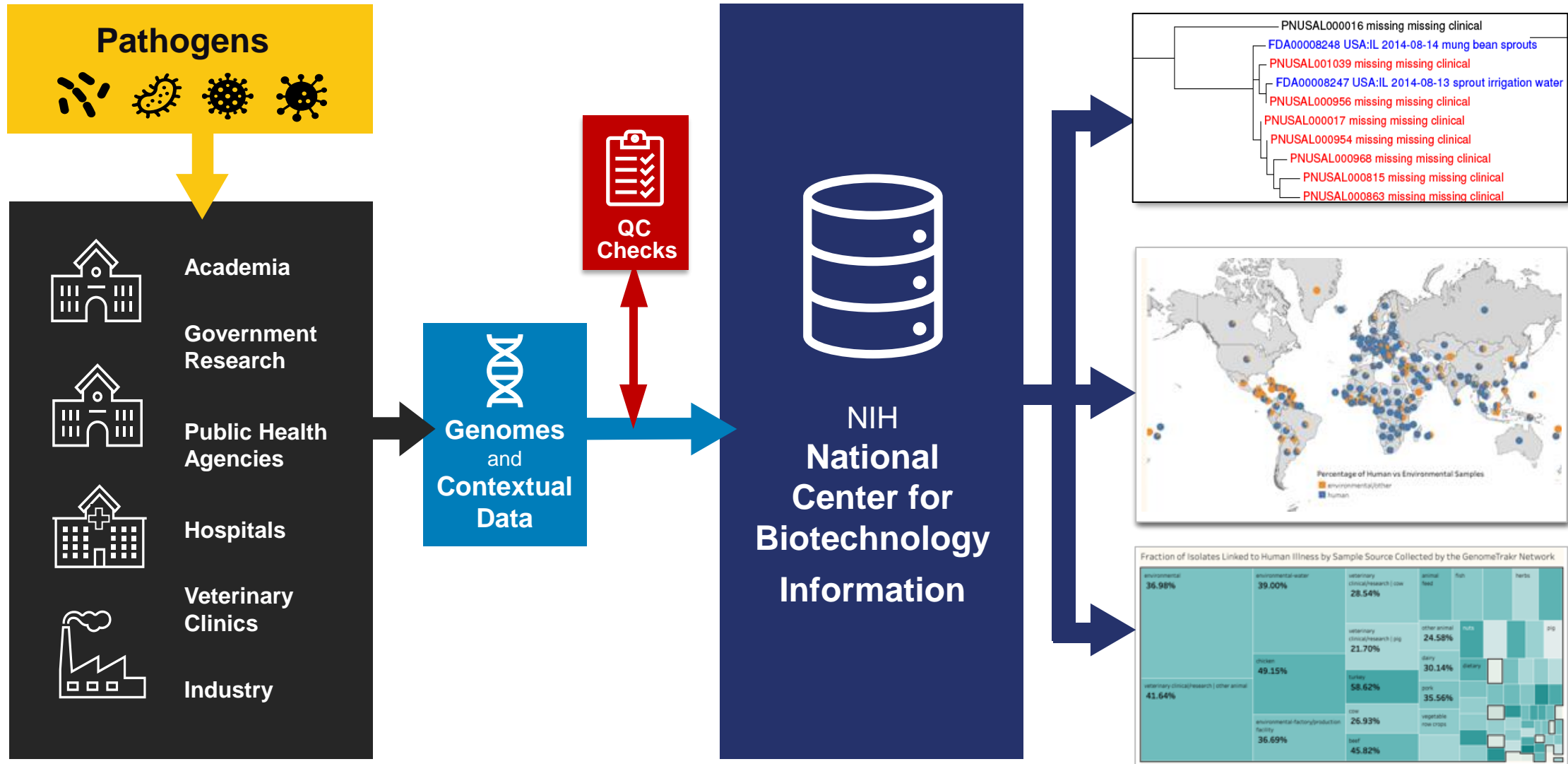
Transition Challenges: Infrastructure

- How is data trafficked in and out of your lab?
- Where is the data stored?
- How is the data analyzed/compared to other data?



The Present:
Open Databases enable real-time linkages

GenomeTrakr: Big Data



Best Practices:



BMC Part of Springer Nature

One Health Outlook

Home About Articles Submission Guidelines

Review | Open Access | Published: 19 October 2020

Optimizing open data to support one health: best practices to ensure interoperability of genomic data from bacterial pathogens

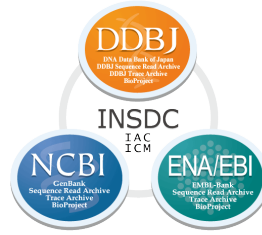
Ruth E. Timme, William J. Wolfgang, Maria Balkey, Sai Laxmi Gubbala Venkata, Robyn Randolph, Marc Allard & Errol Strain

One Health Outlook 2, Article number: 20 (2020) | Cite this article

2353 Accesses | 11 Citations | 35 Altmetric | Metrics

Abstract

The holistic approach of One Health, which sees human, animal, plant, and environmental health as a unit, rather than discrete parts, requires not only interdisciplinary cooperation, but standardized methods for communicating and archiving data, enabling participants to easily share what they have learned and allow others to build upon their findings. Ongoing work by NCBI and the GenomeTrakr project illustrates how open data platforms can help meet the needs of federal and state regulators, public health laboratories, departments of agriculture, and universities. Here we describe how microbial pathogen surveillance can be transformed by having an open access database along with Best Practices for contributors to follow. First, we describe the open pathogen surveillance framework, hosted on the NCBI platform. We cover the current community standards for WGS quality, provide an SOP for assessing your own sequence quality and recommend QC thresholds for all submitters to follow. We then provide an overview of NCBI data submission along with step by step details. And finally, we provide curation guidance and an SOP for keeping your public data current within the database. These Best Practices can be models for other open data projects, thereby advancing the One Health goals of Findable, Accessible, Interoperable and Re-usable (FAIR) data.



Independent, international, open data repository



Public, version-controlled protocols



Set data standards for sequence and contextual data (machine-readable)



Open-access data analysis platform

One Health Enteric package scope



CORE attributes

- Isolate identifiers
- Collected by
- Date of collection
- Geographic location
- Sampling purpose
- Sampling device
- Project name
- IFSAC category
- Source type
- sequenced by



Human/animal host

- Host
- Host disease
- Host sex + age
- Host tissue sampled
- Animal environment
- Antimicrobials in food
- Animal housing system



Food samples

- Geographic origin
- Intended consumer
- Collection site description
- Food product type
- Label claims
- Food source
- Food processing types
- Food preservation process
- Food additives
- Food contact surface
- Food container wrapping
- Food container integrity



Food facility

- Facility type
- Building setting
- Food processed
- Facility location
- Monitoring zone
- Indoor sampling surface
- Surface material
- Surface material cond.
- Surface orientation
- Surface temperature
- Biocide used
- Animal intrusion



Farm and Environment

- ENVO triad
- Farm type
- Plant growth medium
- watering method
- Relative loc of sample
- Fertilizer administration
- Food cleaning process
- Sanitizer used
- Farm equip. used
- Water samples
- Extreme weather event
- Mechanical damage

One Health Enteric



An official website of the United States government [Here's how you know](#) ▾

NIH National Library of Medicine
National Center for Biotechnology Information Log in

BioSample
[Advanced](#)

BioSample Packages

A package represents a type of BioSample and specifies the list of attributes by which it should be described. Several BioSample packages are defined in the database. You can download the package details in [xml format](#).

The following information is stored for BioSample Packages:

Name	Canonical name and version of the package.
DisplayName	Human-readable display name and version of the package.
ShortName	Short name of the core package.
EnvPackage	Ancillary environmental package (used only for MixS packages).
Description	Definition of the package.
Example	Example record.

Unless submitters should use canonical name, for example: MIGS.ba.air.5.0.

Standard Packages

- [SARS-CoV-2: clinical or host-associated; version 1.0](#)
- [SARS-CoV-2: wastewater surveillance; version 1.0](#)
- [One Health Enteric; version 1.0](#) ←
- [Microbe; version 1.0](#)
- [Model organism or animal; version 1.0](#)
- [Metagenome or environmental; version 1.0](#)
- [Invertebrate; version 1.0](#)
- [Human; version 1.0](#)
- [Plant; version 1.0](#)
- [Virus; version 1.0](#)
- [Beta-lactamase; version 1.0](#)

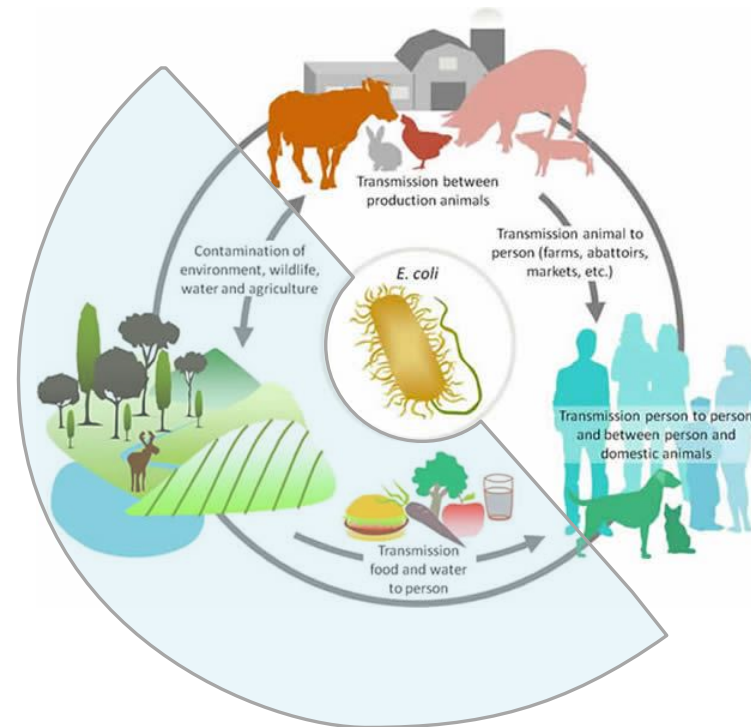
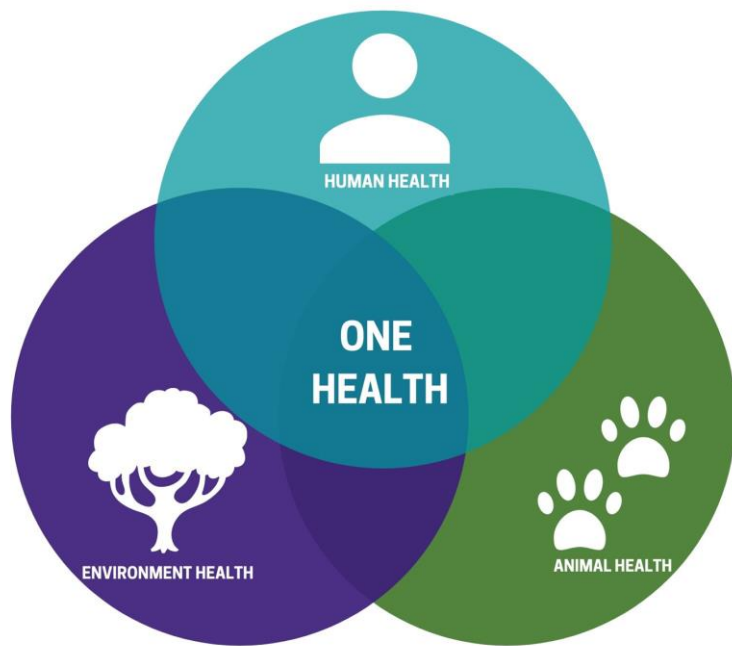
Pathogen

- [Pathogen: clinical or host-associated; version 1.0](#)
- [Pathogen: environmental/food/other; version 1.0](#)

One Health framework: uniquely important for foodborne pathogens

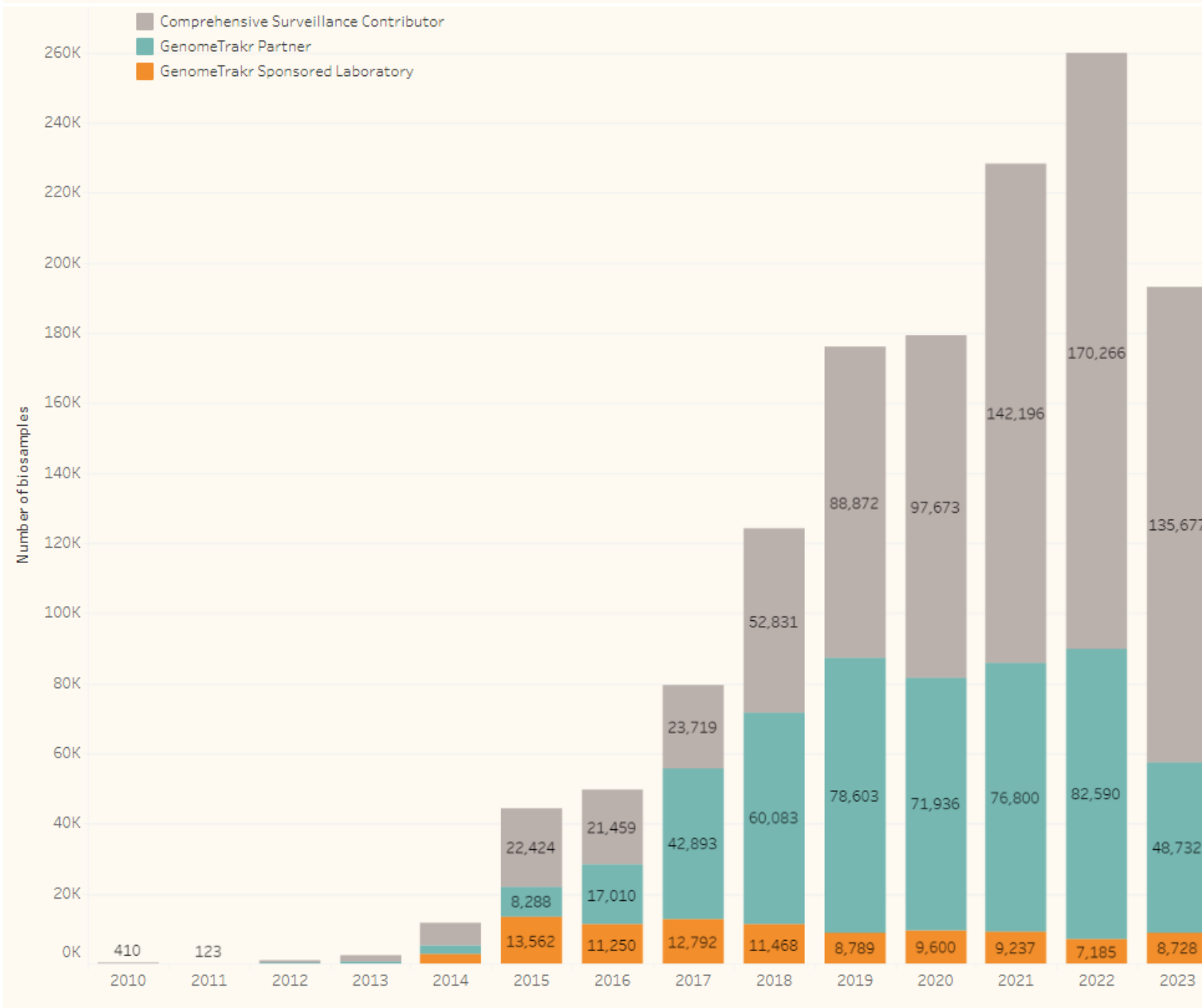


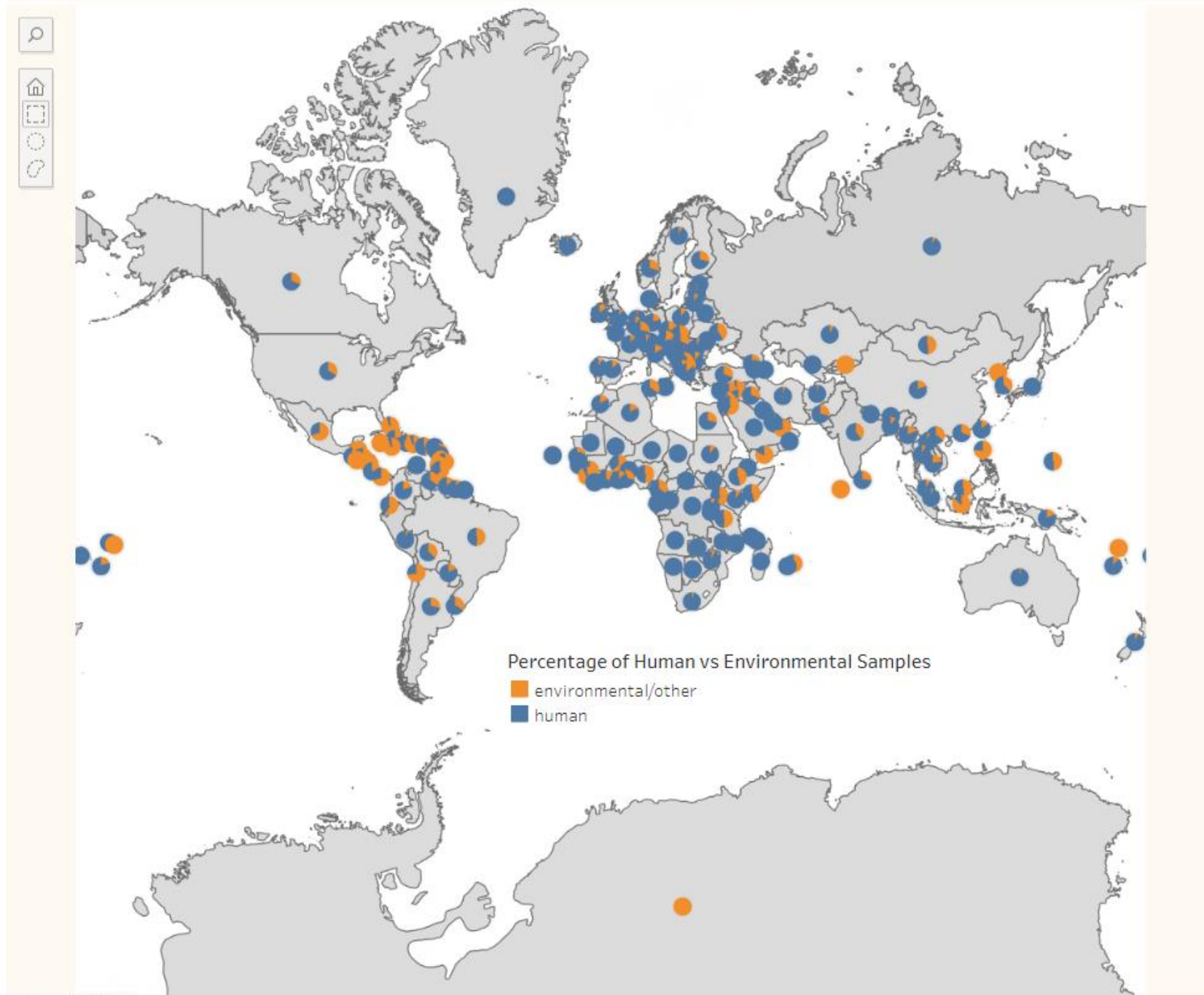
Pathogen surveillance: need to sample the full lifecycle of these pathogens (humans, animals, environment, water, air, food processing plants, food products)



Openly sharing data enables us to traverse the One Health sphere

Comprehensive Surveillance of Enteric Pathogens





It is essential to link clinicals to food/env

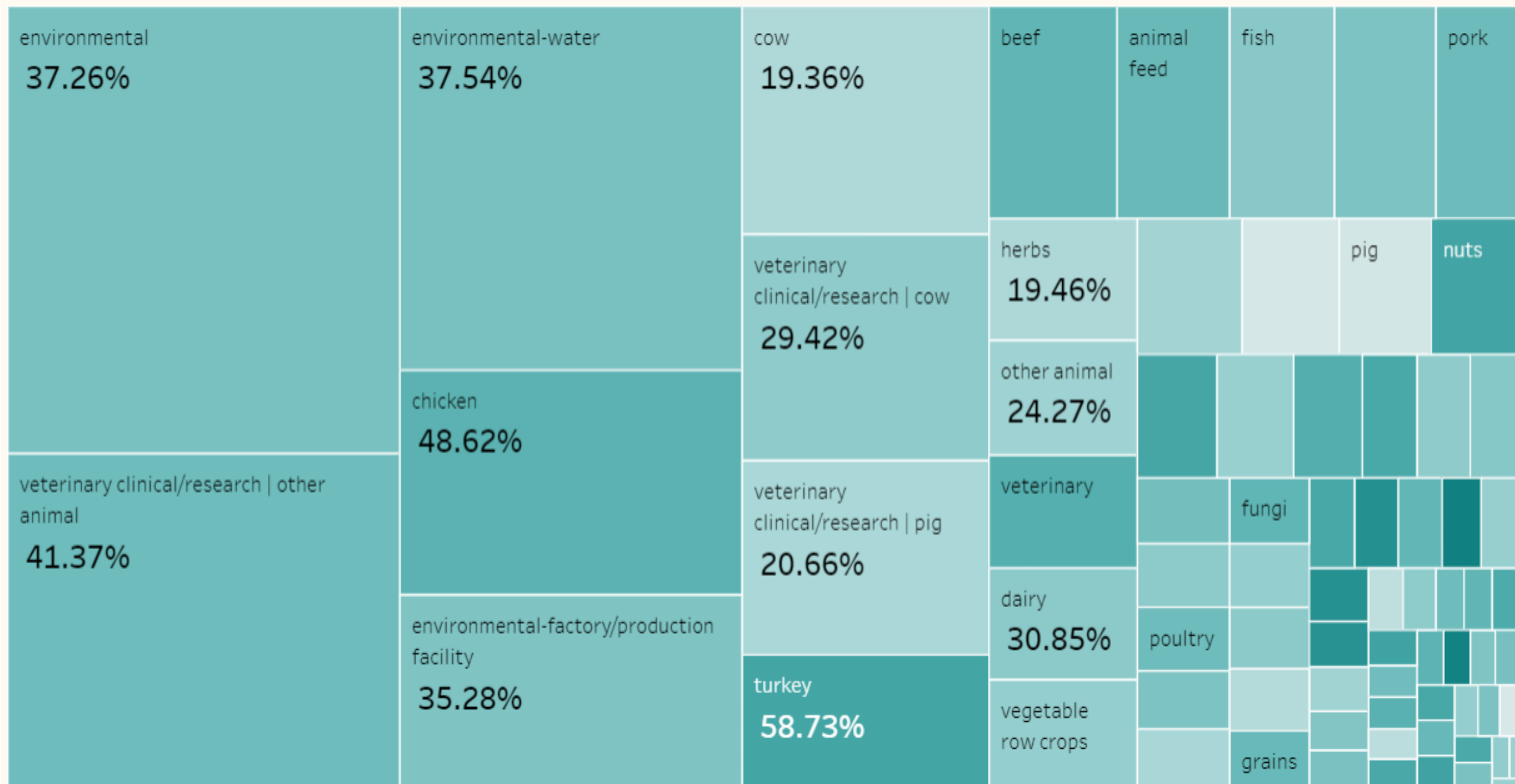


GenomeTrakr Sponsored Surveillance of Enteric Pathogens

Number of sequenced isolates	Number of pathogen species	Broad sample source										
95,755	19	<p>A donut chart illustrating the distribution of 95,755 sequenced isolates across four broad sample sources. The largest portion is Food at 37.29%, followed by Environmental at 30.96%, Animal at 22.01%, and Human at 0.07%.</p> <table border="1"><thead><tr><th>Source</th><th>Percentage</th></tr></thead><tbody><tr><td>Food</td><td>37.29%</td></tr><tr><td>Environmental</td><td>30.96%</td></tr><tr><td>Animal</td><td>22.01%</td></tr><tr><td>Human</td><td>0.07%</td></tr></tbody></table>	Source	Percentage	Food	37.29%	Environmental	30.96%	Animal	22.01%	Human	0.07%
Source	Percentage											
Food	37.29%											
Environmental	30.96%											
Animal	22.01%											
Human	0.07%											

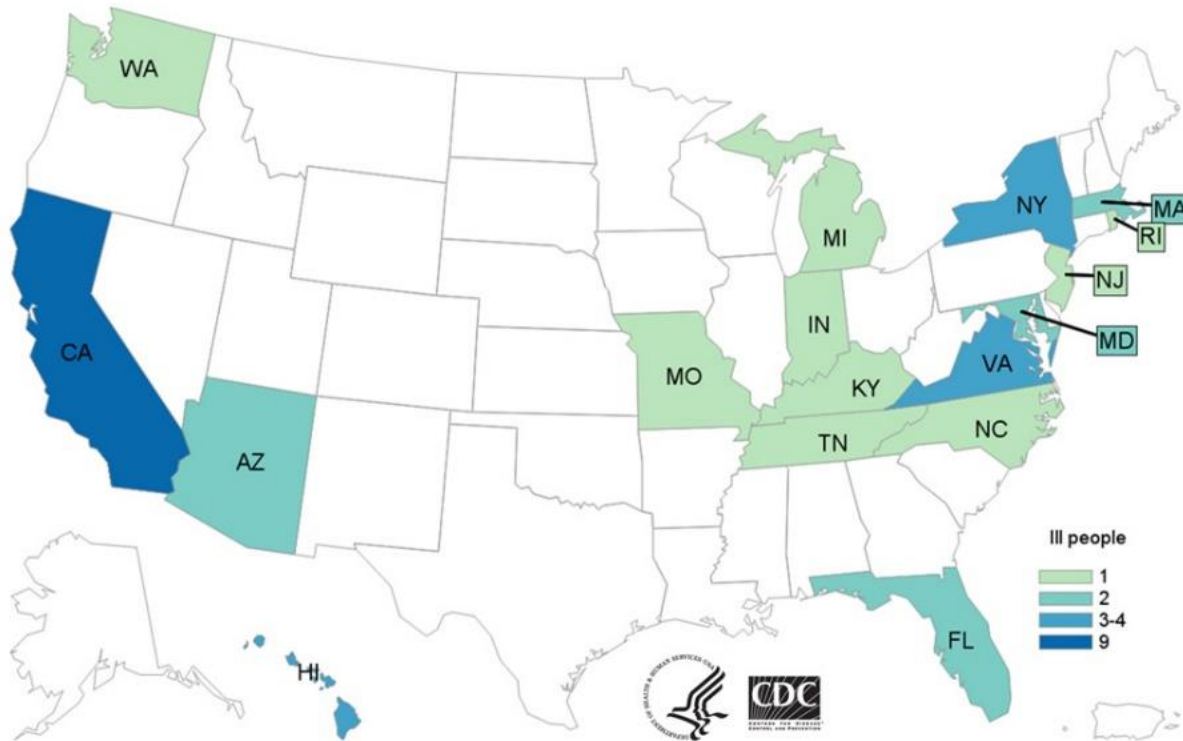
U.S. GenomeTrakr Sponsored Laboratories

Fraction of Isolates Linked to Human Illness by Sample Source Collected by the GenomeTrakr Network



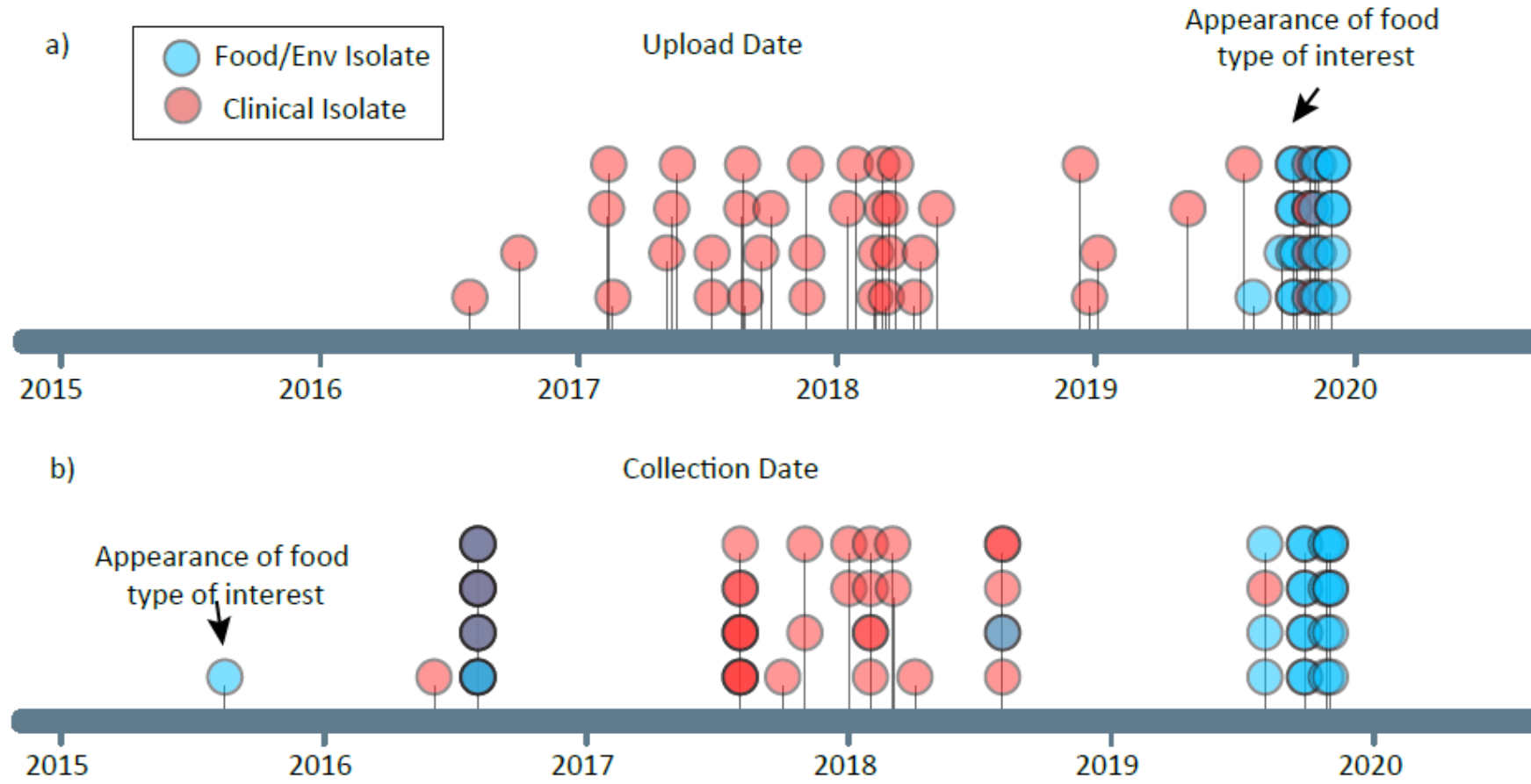
Case Study:
First LM outbreak
(ultimately linked to Enoki Mushrooms)

Early US Investigation



- 36 cases in 17 states
- 6 pregnancy-associated cases; 2 fetal losses
- 4 deaths
- Illness Onset dates: Nov. 23, 2016-Dec.13, 2019
 - 2016: 1 case
 - 2017: 18 cases
 - 2018: 17 cases
 - 2019: 6 cases
- Multiple investigations throughout the years; source was not identified
- Epidemiologic signal for Asian-style foods including fresh produce
- Preliminary traceback record review for a positive US State sample of enoki mushrooms identified a manufacturer of interest in the Republic of Korea

Collection vs. Upload

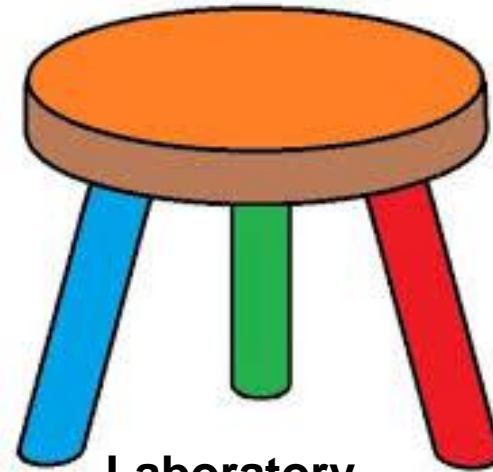


The 3 Legged Stool of Foodborne Outbreak Investigations



Epidemiology

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



Laboratory Support

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

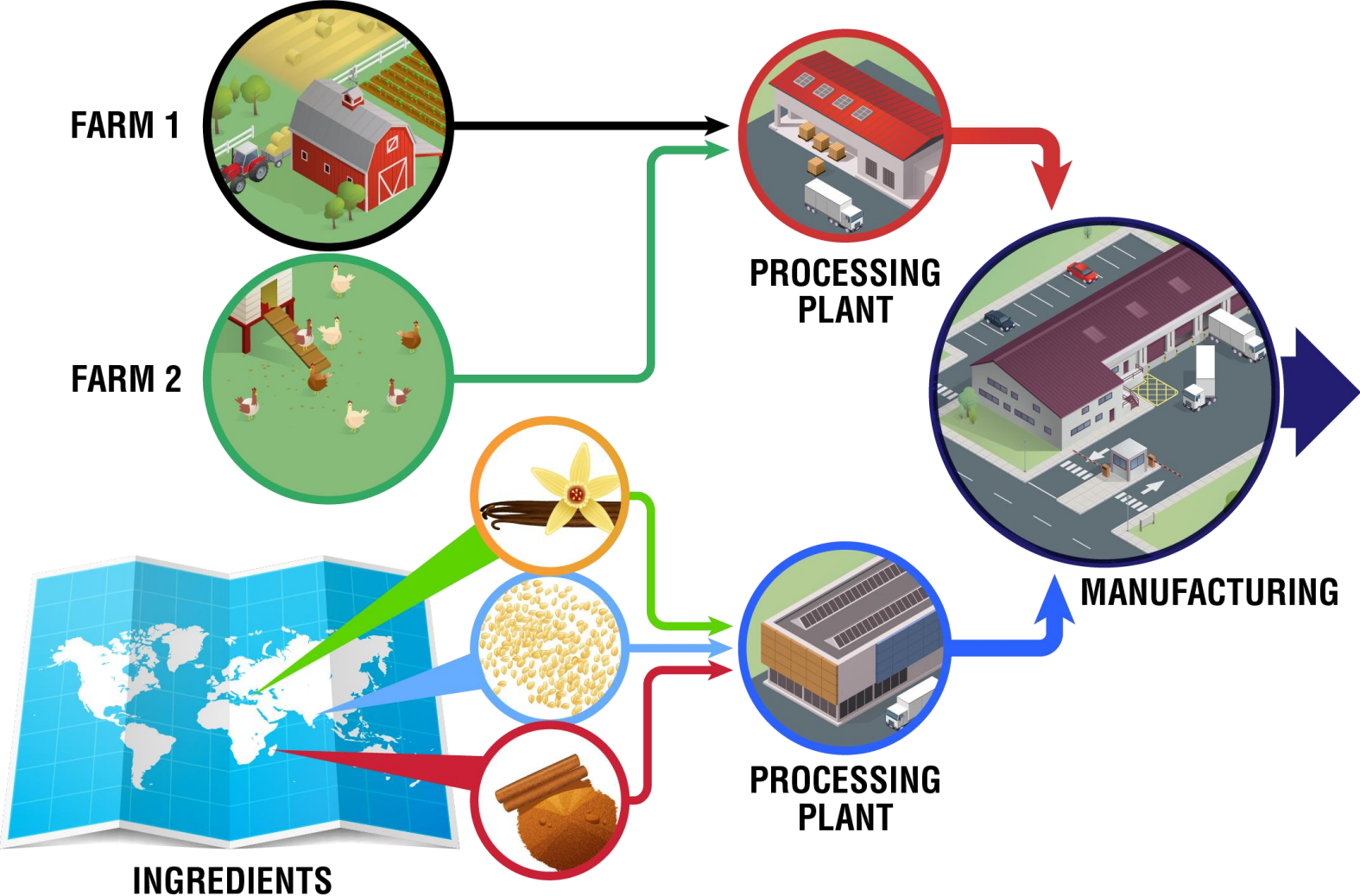
Traceback

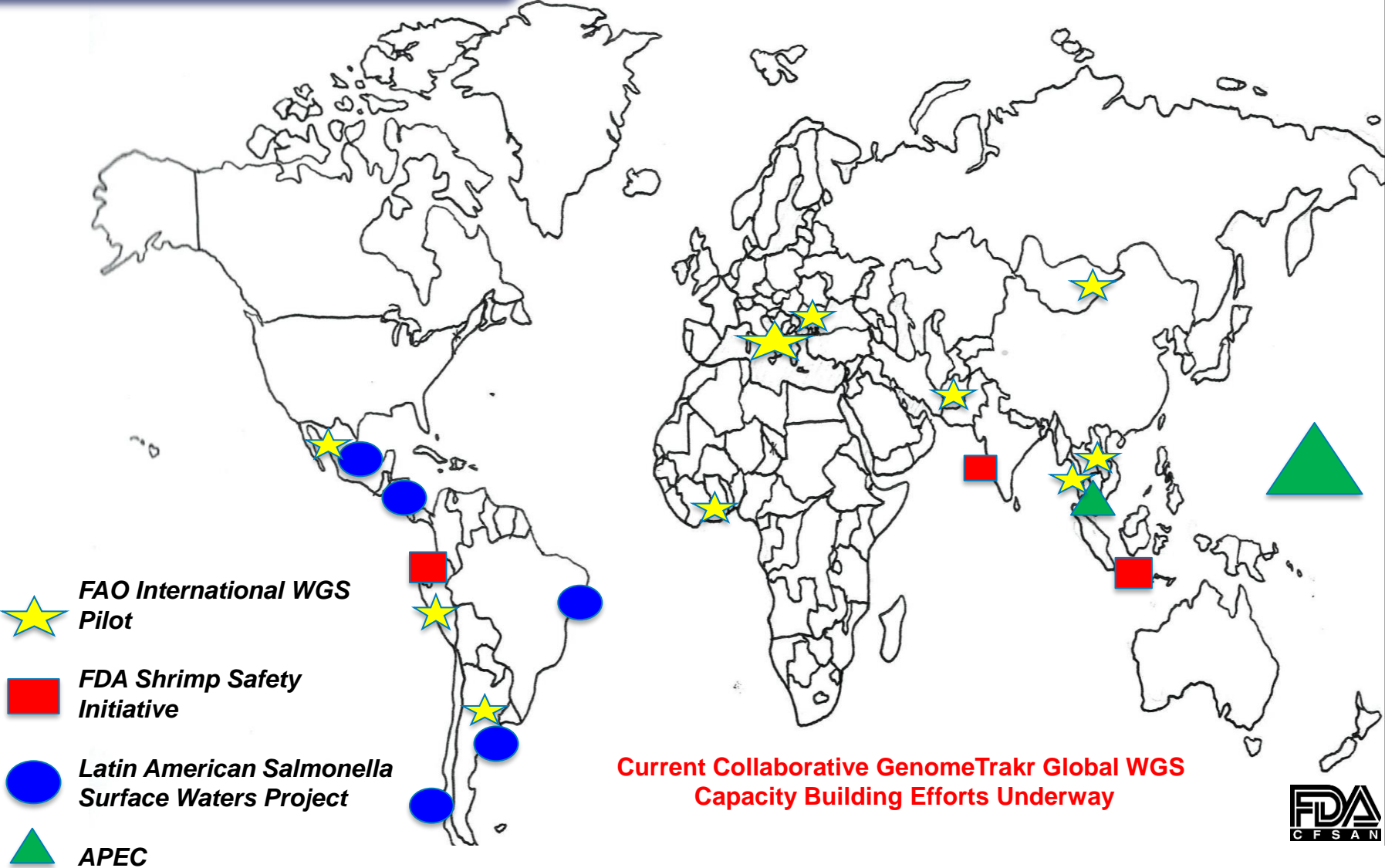
- ~distribution chains/product flow
- ~environmental assessments
- ~potential source reservoirs
- ~GAP/GMP review of HACCPs
- ~Inspectional findings
- ~Risk assessments



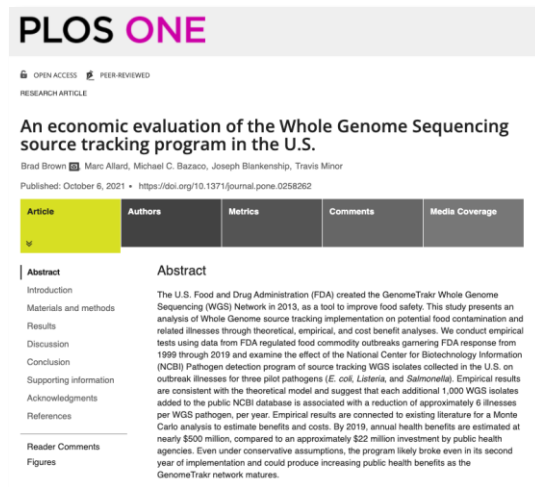
From the Present to the Future:
Where do we go and what can we do?

Identifying Root Cause: Transition to Preventive Controls





- GenomeTrakr program was likely cost effective by its second year of implementation
- \$100 M -> \$450 M in net annual health benefits (est. from 2019). >\$ Billion estimated benefits.



Return on Investment: \$10 dollars in averted human health costs for every \$1 dollar invested. For each additional 1,000 WGS isolates added to the public NCBI database is associated with a reduction of approximately 6 illnesses per WGS pathogen, per year.

Metagenomics

