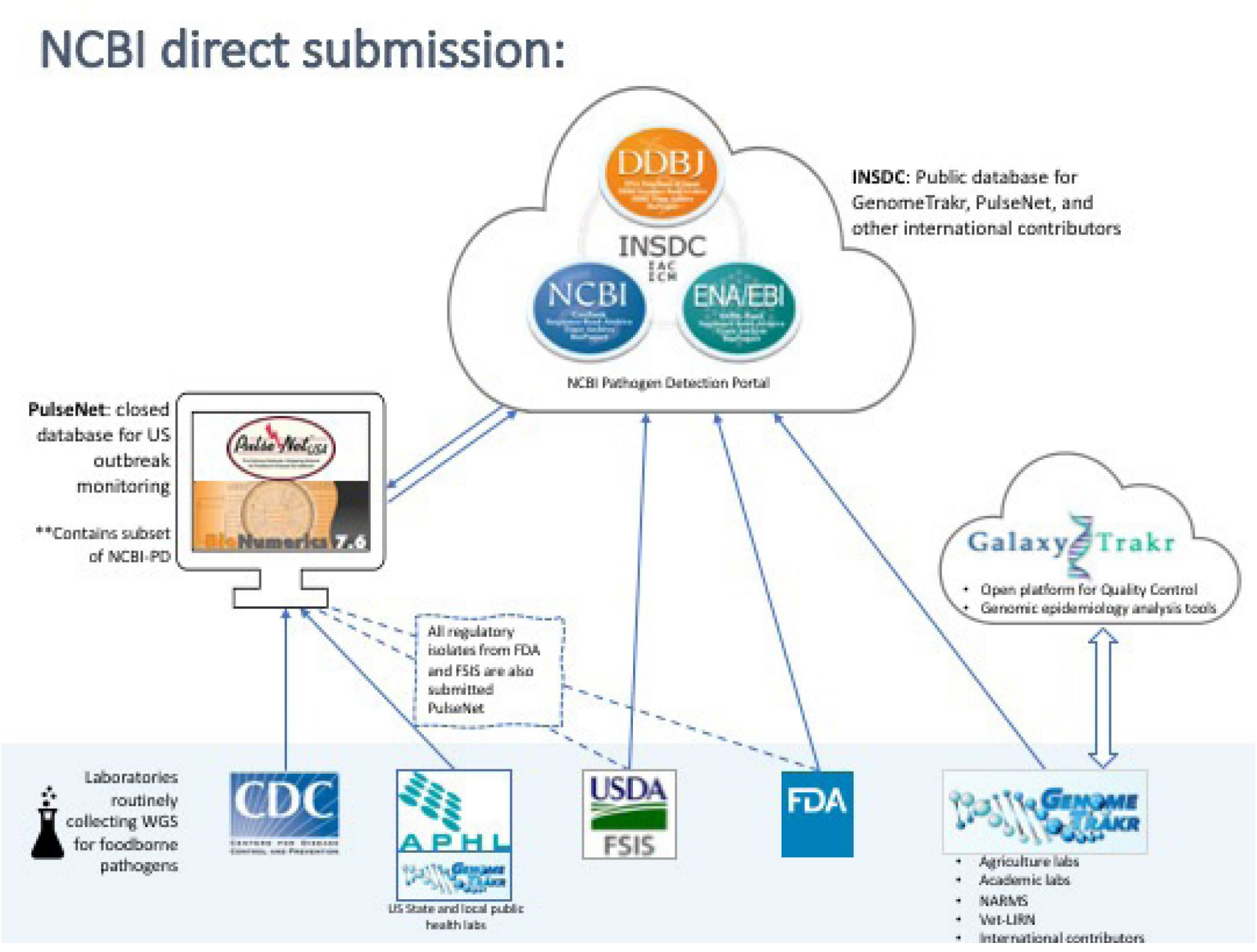
GenomeTrakr database and network updates: Lessons on what can be done with a half million foodborne pathogen genomes

M. Allard, R. Timme, M. Balkey, E. Stevens, M. Hoffmann, G. Kastanis, Tim Muruvanda, J. Payne, A. Pightling, H. Rand, J. Pettengill, Y. Luo, N. Gonzalez-Escalona, M. MaguireThon, D. Melka, P. Curry, Y. Chen, S. Tallent and E. Brown
Food and Drug Administration, College Park MD USA
Marc.allard@fda.hhs.gov

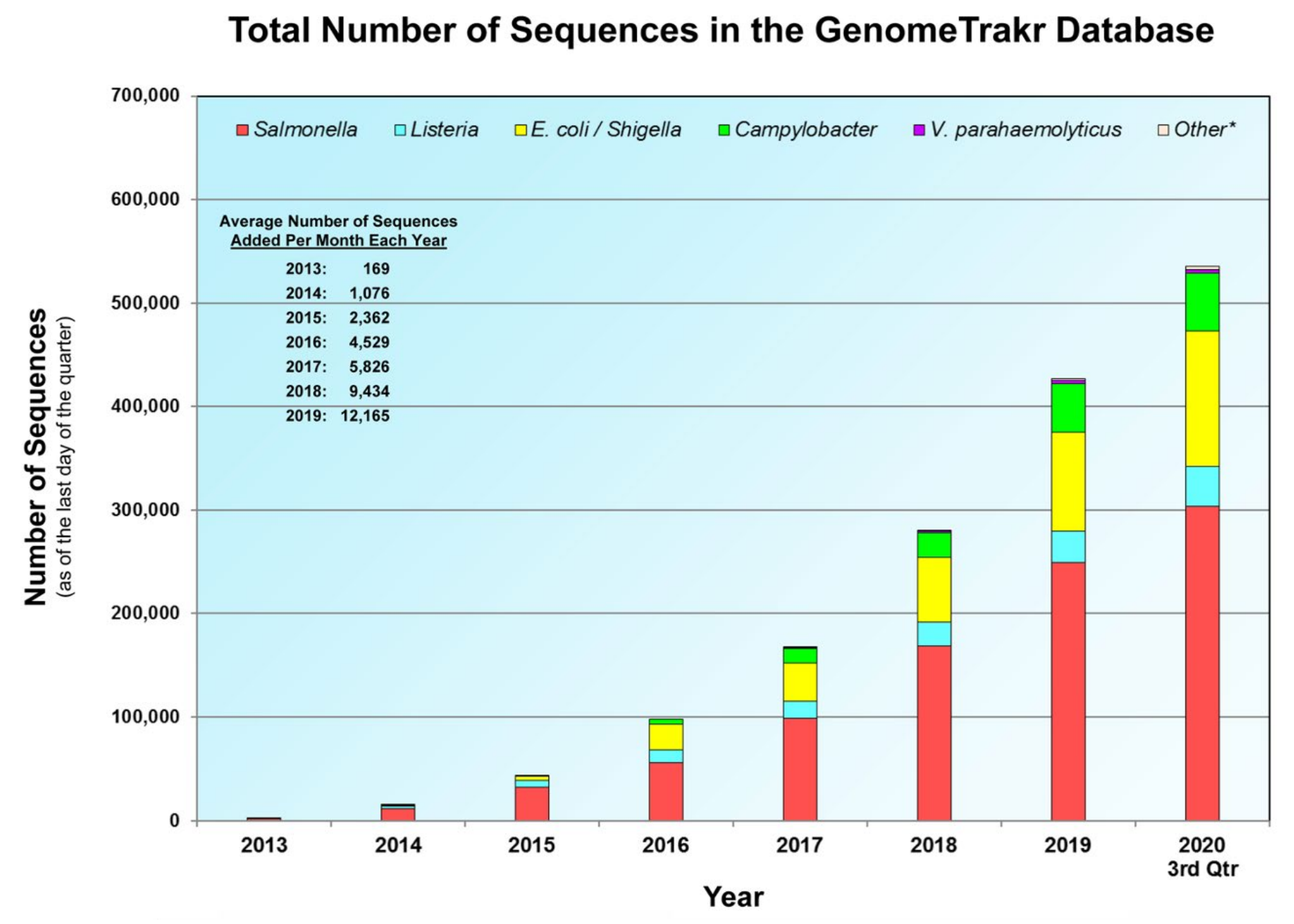
Methods

An international database of federal, state, academic and international laboratories has been using WGS data to rapidly characterize pathogens since 2012. This GenomeTrakr network is part of NCBI Pathogen Detection web site. Public health agencies (FDA, CDC and USDA-FSIS) collect and share data in real time. This high-resolution, rapidly growing database is actively being used in outbreak investigations at state, national, and international levels. GenomeTrakr database has demonstrated how a distributed network of desktop WGS sequencers can be used in concert with traditional epidemiology and investigation for source tracking of foodborne pathogens. This "open data" model allows greater transparency between federal/state agencies, industry partners, academia, and international collaborators. This database has continued to grow and diversify with the foodborne pathogen database increasing in the last year to ~500,000 draft genomes. This report documents the value that is gained from analyses of a half million foodborne pathogen genomes. NCBI, currently is producing daily phylogenetic results for 33 bacterial pathogens including: *Salmonella*, *Listeria*, *E. coli* and *Campylobacter* for >47,000 clusters of < 50 SNPs, plus pairwise SNP count differences for cluster members. A second NCBI product called AMRFinderPlus provides genotype calls of presence/absence for known antimicrobial resistance (AMR) genes as well as genotypes related to virulence and stress. Beyond outbreak and cluster detection, this high-resolution data is harvested by numerous independent public health entities that add value to what is publicly available. We discuss Resistome Tracker software which provides visually informative displays of antibiotic resistance genes in bacteria supported by the Center for Veterinary Medicine; and GenomeGraphR which presents foodborne pathogen WGS data and associated curated metadata in a user-friendly interface that allows users to query a variety of research questions such as, transmission sources and dynamics, global reach, and persistence of genotypes associated with contamination in the food supply and foodborne illness across time or space. The new laboratory flexible funding model has distributed funds to double the existing network of domestic partners. Results demonstrate global benefits of having an open data model. Understanding root causes of foodborne contamination assists our academic, public health and industry partners to develop preventative controls to make food safer globally.

Dataflow and Contents



GenomeTrakr submissions over time



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

Supporting Citations:

Sanaa M, Pouillot R, Garces-Vega FJ, Strain E, Van Doren JM. GenomeGraphR: a user-friendly open-source web application for foodborne pathogen whole genome sequencing data integration, analysis, and visualization. BioRxiv. 2018 https://doi.org/10.1101/495309.

Allard, Stevens and Brown. All for one and one for all: the true potential of whole-genome sequencing. The Lancet Infectious Diseases 2019b;19:683-684

Nguyen et al. 2019. Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal Salmonella. J Clin Microbiol 57:e01260-18. https://doi.org/10.1128/JCM.01260-18.

Zhang et al. 2019. Zoonotic Source Attribution of Salmonella enterica Serotype Typhimurium Using Genomic Surveillance Data, United State. Emerg Infect Dis. 25(1): 82–91. doi: 10.3201/eid2501.180835

Supporting Webliography

GenomeTrakr Videos:
https://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/ucm422080.htm

Galaxy GenomeTrakr:
https://www.galaxytrakr.org/root/login?redirect=%2F

Center for Veterinary Medicine, Resistome Tracker:
https://www.fda.gov/AnimalVeterinary/SafetyHealth/AntimicrobialResistance/NationalAntimicrobialResistanceMonitoringSystem/ucm570694.htm

FDA Whole Genome Sequencing Program:
http://www.fda.gov/Food/FoodScienceResearch

GenomeGraphR: WGS data integration, analysis, and visualization for risk assessment and management:
https://fda-riskmodels.foodrisk.org/genomegraphr/

NCBI Pathogen Detection website and tree visualization:
http://www.ncbi.nlm.nih.gov/projects/pathogens/

CFSAN SNP pipeline:
https://github.com/CFSAN-Biostatistics/snp-pipeline

NCBI Pathogen Detection gene:
A partial gene list is now incorporated into AMRFinderPlus v3.2 (https://github.com/ncbi/amr/wiki) and the Reference Gene Catalog as of 2019-10-31 (https://www.ncbi.nlm.nih.gov/pathogens/isolates#/refgene/)

Food Ontology:
https://foodon.org/

Genomic Epidemiology Ontology:
https://genepio.org/

Collaborations with Independent Academic Researchers:

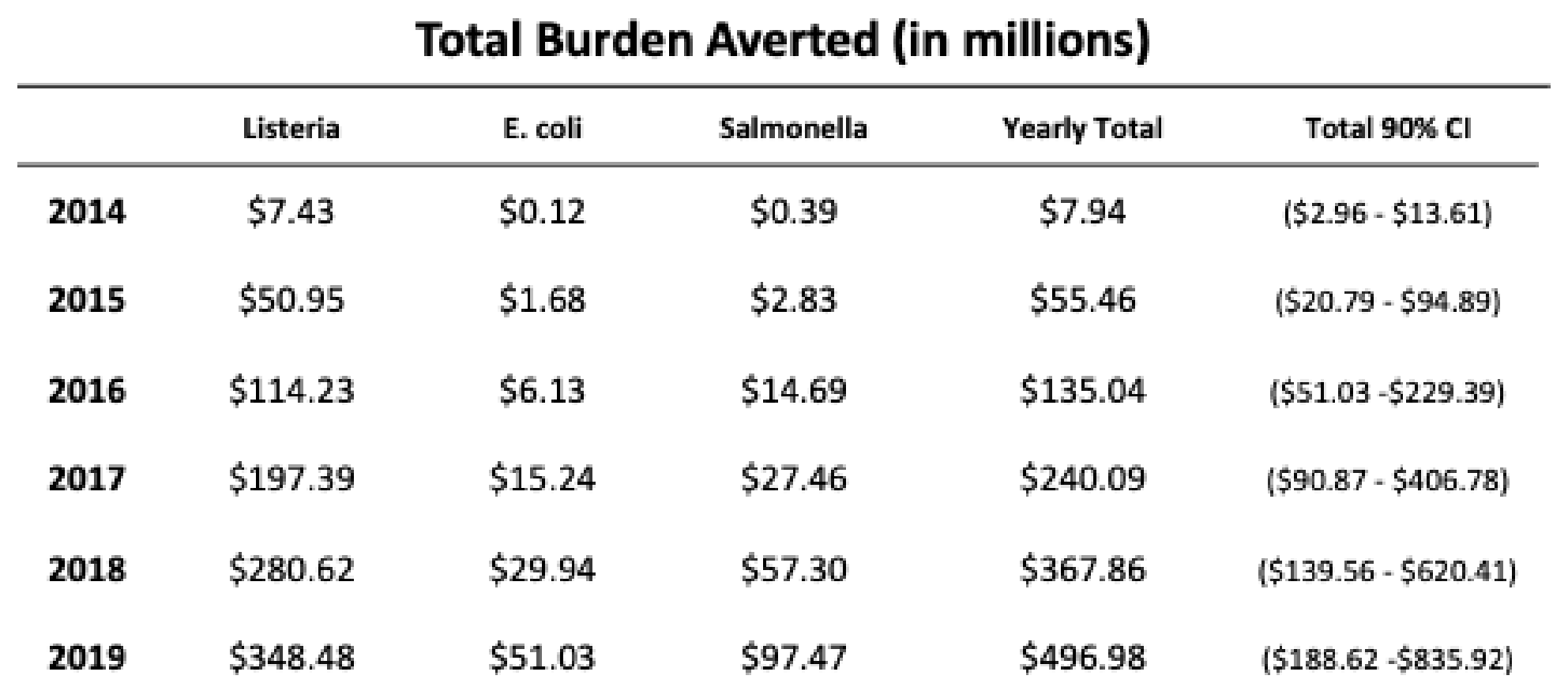
In addition to state labs, university labs, federal labs, and labs located outside of the U.S., the GenomeTrakr network has collaborative relationships with independent researchers, including graduate students, who supply foodborne pathogen isolates and/or perform whole genome sequencing on the isolates.

Data Curation:

Data curation and bioinformatic analyses and support are provided by the National Center for Biotechnology Information (NCBI) in Bethesda, MD.

Metadata improvement include two new fields on every GenomeTrakr BioSample. This will enable computers to more easily interpret source information for the 90K+ existing GenomeTrakr records. FoodOn Ontology term and IFSAC Category term.

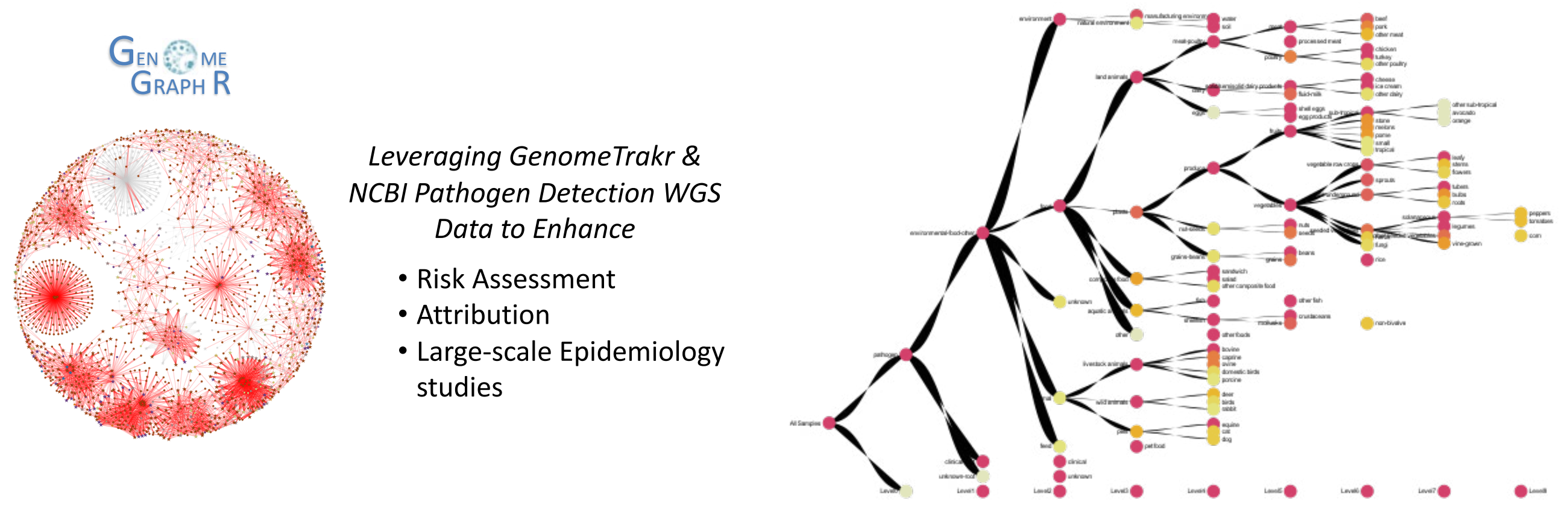
Economic Evaluation of WGS Reduces the Burden of Illness

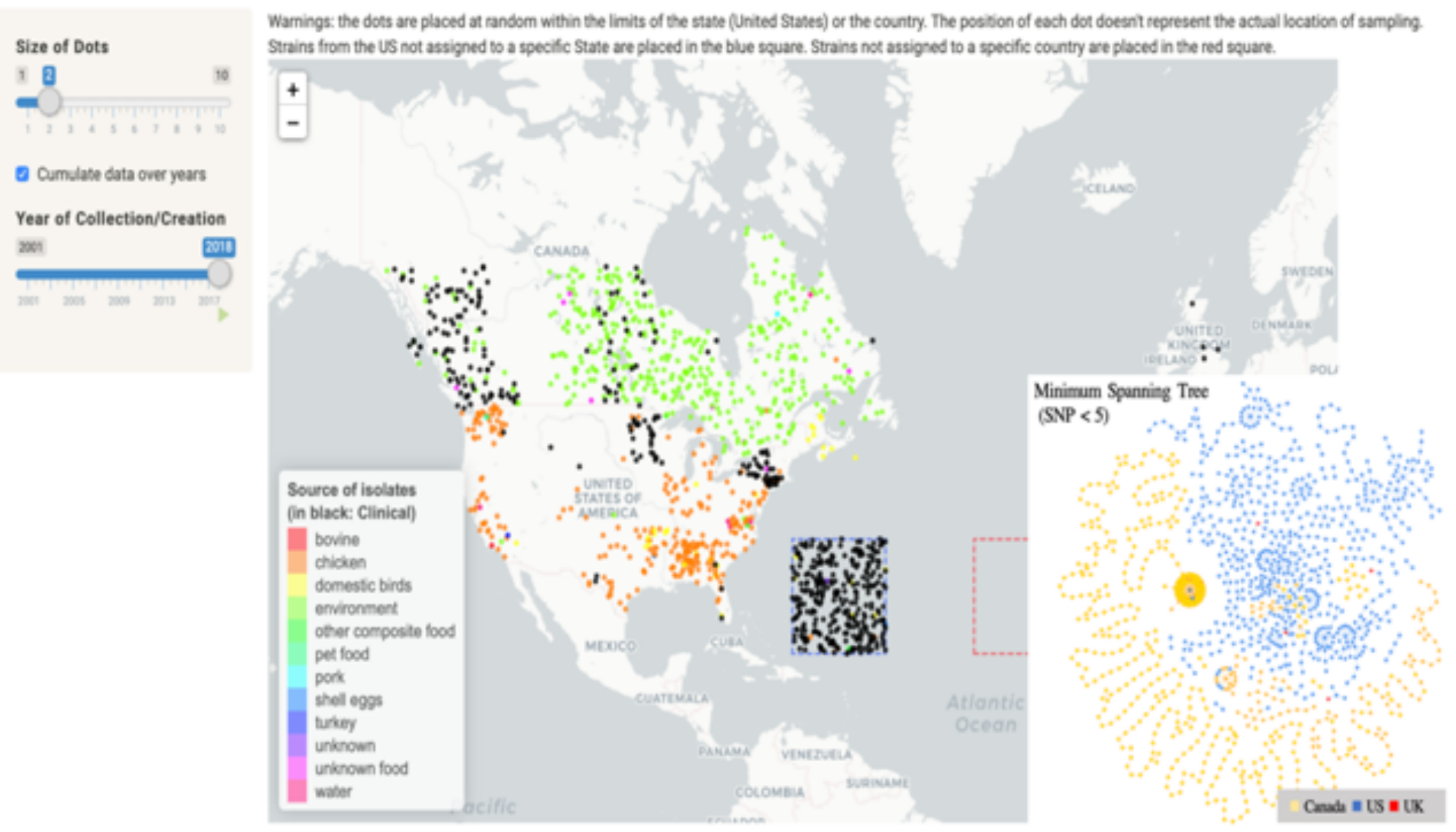


Total Burden Averted (in millions)

	Listeria	E. coli	Salmonella	Yearly Total	Total 90% CI
2014	$7.43	$0.12	$0.39	$7.94	($2.96 - $13.61)
2015	$50.95	$1.68	$2.83	$55.46	($20.79 - $94.89)
2016	$114.23	$6.13	$14.69	$135.04	($51.03 -$229.39)
2017	$197.39	$15.24	$27.46	$240.09	($90.87 - $406.78)
2018	$280.62	$29.94	$57.30	$367.86	($139.56 - $620.41)
2019	$348.48	$51.03	$97.47	$496.98	($188.62 -$835.92)

GenomeTrakr contributors





Warnings: the dots are placed at random within the limits of the state (United States) or the country. The position of each dot doesn't represent the actual location of sampling. Strains from the US not assigned to a specific State are placed in the blue square. Strains not assigned to a specific country are placed in the red square.

Joining and Using the GenomeTrakr Network

For information about joining the GenomeTrakr network as a sequencing lab, providing isolates to a current member lab for sequencing, or using the GenomeTrakr database as a research tool, please utilize the following resources:

- Access the GenomeTrakr data via NCBI's Pathogen Detection portal: https://www.ncbi.nlm.nih.gov/pathogens/
- NCBI's instructions for submitting data: https://www.ncbi.nlm.nih.gov/pathogens/submit-data
- Establish your own WGS submissions to GenomeTrakr: Timme RE, Sanchez Leon M, Allard MW 2019. Utilizing the Public GenomeTrakr Database for Foodborne Pathogen Traceback. Methods in molecular biology (Clifton, N.J.) 1918:201–212. DOI: 10.1007/978-1-4939-9000-9_17.
- Allard MW, Zheng J, Cao G, Timme R, Stevens E, Brown EW. Food Safety Genomics and Connections to One Health and the Clinical Microbiology Laboratory. 2020, Clinics in Laboratory Medicine, 40(4): 553-563.
- Pettengill JB, Markell A, Conrad A, Carleton HA, Beal J, Rand H, Musser S, Brown EW, Allard MW, Huffman J, Harris S, Wise M, Locas A. A multinational listeriosis outbreak and the importance of sharing genomic data. 2020 The Lancet Microbe 1(6), e233-e234.
- Timme RE, Wolfgang WJ, Balkey M, Gubbala Venkata SL, Randolph R, Allard MW, Strain E. Optimizing Open Data to Support One Health: Best Practices to Ensure Interoperability of Genomic Data from Microbial Pathogens. *Preprints* **2020**, 2020040253 (doi: 10.20944/preprints202004.0253.v1). published in One Health Outlook 2(1):1-11.
- Brown B, Minor T, Allard MW, Bazaco MC, Blankenship J. An Economic Evaluation of the Whole Genome Sequencing Program in the U.S. 2021. In review.
- contact GenomeTrakr team at the FDA: GenomeTrakr@fda.hhs.gov