

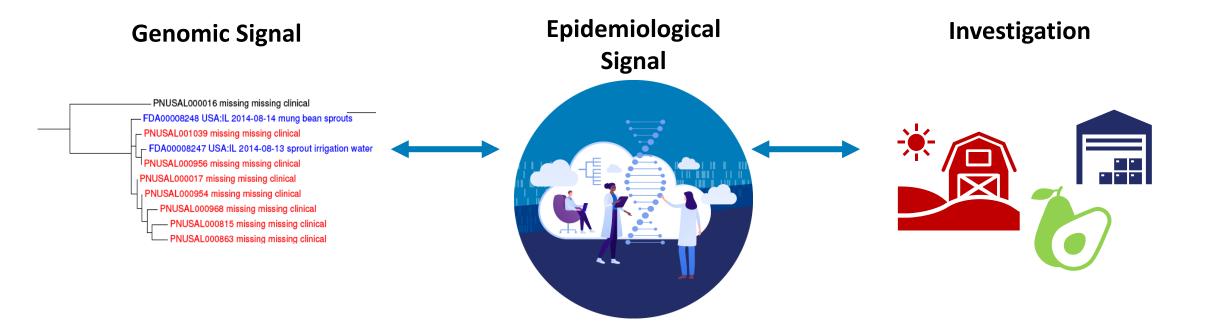
FDA on how it uses WGS Plus, Cost – Benefit - Budget

"Whole Genome Sequencing (WGS) for food safety and its uses in prevention and response of foodborne outbreaks",

The Pathogen Surveillance in Agriculture, Food and the Environment (PATH-SAFE) Programme conference London UK Feb. 28th and 29th, 2024 Marc Allard PhD, Ruth Timme PhD, and Eric Stevens PhD. FDA Center for Food Safety and Applied Nutrition Marc.allard@fda.hhs.gov

New Field: Genomic Epidemiology





Gen Epi regulatory communication

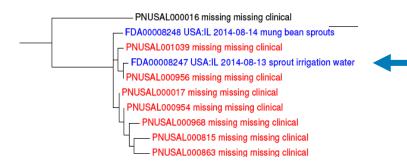


OAO: watches for signal

CORE and OC: to communication to CDC, USDA-FSIS and State DOH + Ag

CORE and OC:

Facility/farm Inspections or product testing by ORA Consumer Safety Officers



ORA / LFFM-funded laboratories

upload genomic data to NCBI

ORA Office of Regulatory Affairs Regulatory arm of FDA Inspections, sequencing



OAO Office of Analytics and Outreach Data interpretation and Risk assessment CORE Outbreaks, OC Compliance

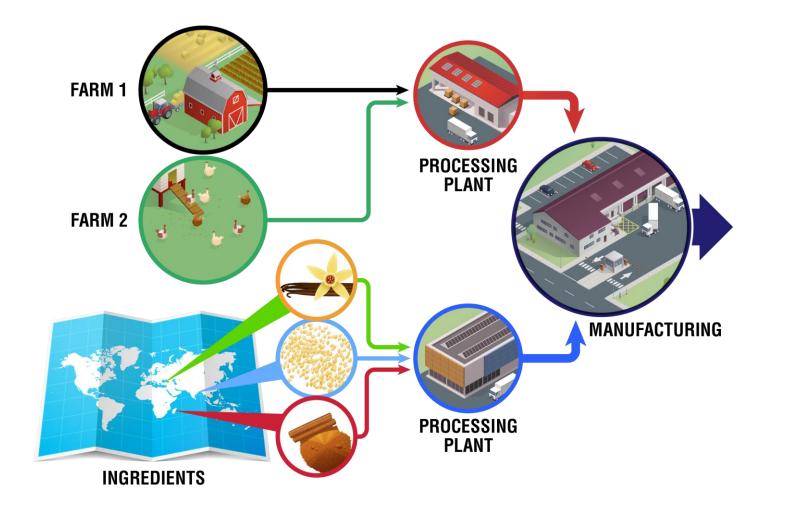


OFS Office of Food Safety, policy OIE Office of International Engagement ORS Office of Regulatory Science, research

Current FDA workflow works for all pathogens and all genomic methods, collected under FDA surveillance and inspection activities (virus, parasite, shellfish, filth, supplement botanicals).

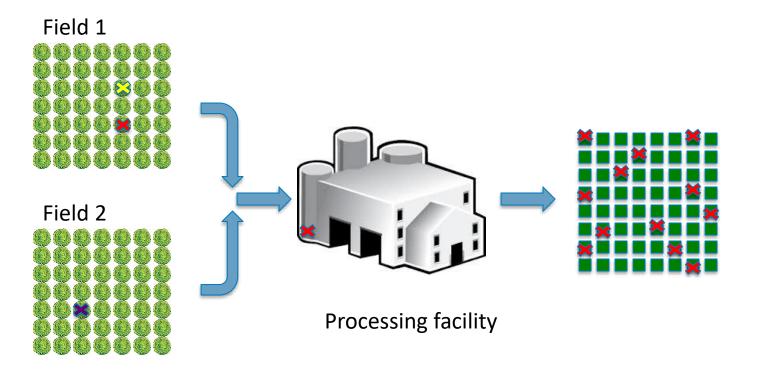
Identifying an Outbreak Vehicle: Trace Forward and Trace Backward



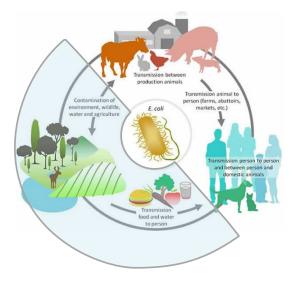




Identifying an Outbreak Vehicle: Determining Resident or Transient pathogen



FDA's GenomeTrakr program



- Sequencing the genomes of foodborne pathogens found in food, food processing facilities, farm environment, water, etc.
- Collaborate with other US agencies and international counterparts to integrate our data with genomic data collected from animals and human clinicals – data made public in real-time.
- Clustering at NCBI Pathogen Detection helps FDA identify causes of foodborne outbreaks and identify other events, like harborage.

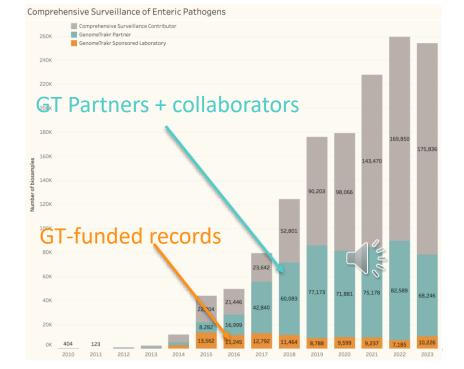
45 funded laboratories:

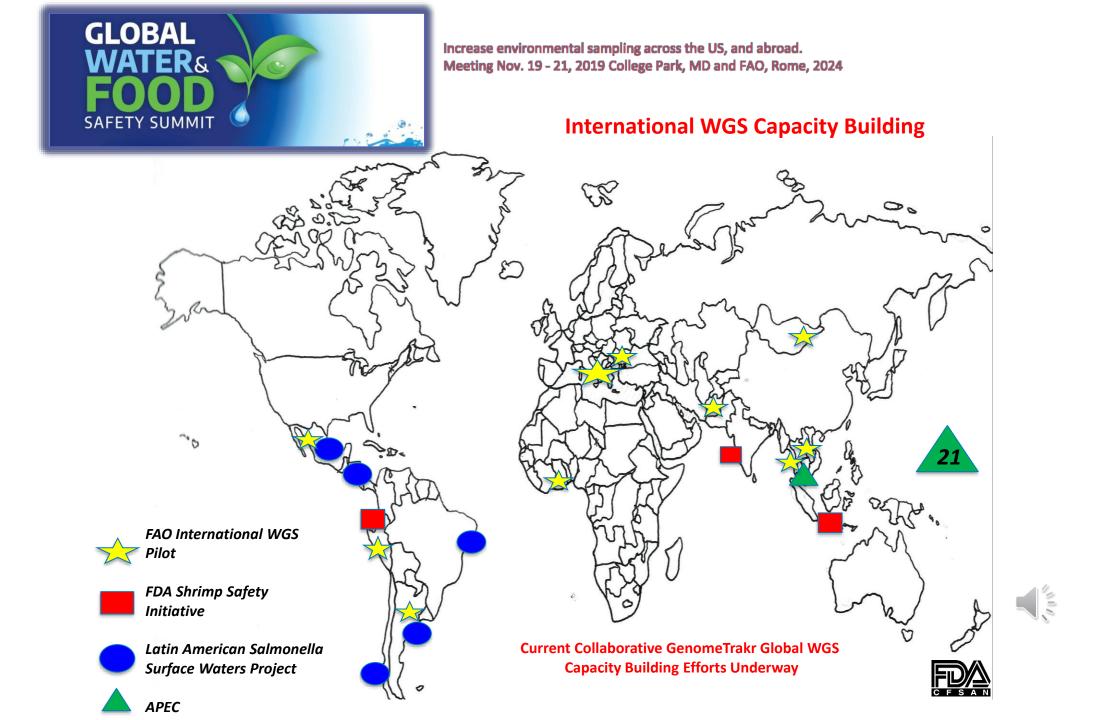


Numerous GT partners and collaborators:

- Public health, Ag, and academic labs
- US agency partners: GenFS and others
- International counterparts







Best Practices from GenomeTrakr



	G F/IR	 Standard metadata required for interoperability FAIR = Findable, accessible, interoperable, reusable
ta ab	protocols.io 🦑	Public, version-controlled protocols ➤ GenomeTrakr workspace:
tal 2011, but asily rk by he ure,	Galaxy	 Open-access analysis platform ➤ "MicroRunQC": QC workflow for microbial pathogens
rrmed by rst, we cover own provide rovide e. These Health	INSUC INSUC	Open data repository for hosting genome + metadata ➤ Enables public/private collaboration

Timme, R.E., Wolfgang, W.J., Balkey, M. *et al. One Health Outlook* **2**, 20 (2020). <u>https://doi.org/10.1186/s42522-020-00026-3</u> https://www.protocols.io/workspaces/genometrakr1/publications

BMC Part of Springer Nature

One Health Outlook

Review Open Access Published: 19 October 2020

Optimizing open data to support one health: best practices to ensure interoperability of genomic dat 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 NCRI and the GenomeTrak represe illustrates how one 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 evert the current community standards for WGS quilty. provide an SOP for sessing 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 etp details. And finally, we provide curration 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 gails of Findable, Accessible, Interoperable and Re-usable (FAIR) data.

One Health Enteric package:

US Interagency Collaboration for Genomics for Food and Feed Safety (Gen-FS)



National Center for Biotechnology Information (NCBI) Centers for Disease Control and Prevention (CDC) Food and Drug Administration (FDA) U.S. Department of Agriculture (USDA)

OHE 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



Food samples

Geographic origin

- Intended consumer
- Collection site description
- Food product type
- Food source
- Food processing types
- Food preservation process
- Food cooking process
- Food additives
- Food contact surface
- Food container wrapping
- Food quality date



Facility type Building setting Food processed Facility location Monitoring zone Surface material Surface material cond.

 Biocide used Animal intrusion

- Indoor sampling surface
- Surface orientation Surface temperature
- Farm equip. used Water samples
 - Extreme weather event

Farm and Environment

Environmental material

Plant growth medium

- Fertilizer administration
- Food cleaning process
- Sanitizer used

- Mechanical damage

Generic template available at NCBI B



Preview BioSample Types and Attributes

* Select the package that best describes your samples.

All packages Packages for MAG submitters Packages for metagenome submitters

(Optional) Filter packages by organism name

Enter the full scientific name of your samples, e.g., Escherichia coli

- 1 To filter for relevant BioSample packages, enter the full scientific name of the organism of your samples
- If your BioSamples are derived from a species not represented in NCBI's Taxonomy database, enter the genus-level name, e.g., Escherichia
- If your BioSamples are derived from more than one organism, enter the common species, genus, or family, e.g., Enterobacteriaceae
- · If your BioSamples are metagenomic/environmental, or metagenome-assembled genomes (MAG), select the appropriate tab above
- For more information about organism names, see Organism information.

NCBI packages More...

GSC MIxS packages for genomes, metagenomes, and marker sequences More...

○ SARS-CoV-2: clinical or host-associated

Use for SARS-CoV-2 samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of SARS-CoV-2 cases.

SARS-CoV-2: wastewater surveillance Use for SARS-CoV-2 wastewater surveillance samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of SARS-CoV-2 cases

O Pathogen

Use for pathogen samples that are relevant to public health Required attributes include those considered useful for the rapid analysis and trace back of pathogens

One Health Enteric

Use for microbial isolates that are collected for genomic surveillance of enteric pathogens.Sample spaces include the following: 1. human/animal hosts; 2. food samples; 3. food facilities; 4. environmental samples (farm, water, and the environment)

US public health agencies have created customized versions of this package that include more specific guidance, controlled vocabulary picklists, and sub-packages for each of the 4 sample types.

GitHub repository Validation for the OHE package

Microbe

Use for bacteria or other unicellular micro when it is not appropriate or advantageous to use MIxS, Pakogen or Virus packages.

MIGS Cultured Bacterial/Archaeal

Use for cultured bacterial or archaeal genomic sequences. Organism must have lineage Bacteria or Archaea.

MIGS Eukaryotic

Use for eukaryotic genomic sequences. Organism must have lineage Eukaryota

MIGS Viral

Use for virus genomic sequences. Organism must have lineage Viruses.

MIMAG Metagenome-assembled Genome

Use for metagenome-assembled genome sequences produced using computational binning tools that group sequences into individual organism genome assemblies starting from metagenomic data sets. Organism cannot contain the term 'metagenome'. Use the MIUVIG package for virus genomes. Before creating BioSamples for prokaryotic and eukaryotic MAGs, please read and follow the MAG submission instructions.

MIMARKS Specimen

Use for any type of marker gene sequences, eg, 16S, 18S, 23S, 28S rRNA or COI obtained from cultured or voucher-identifiable specimens. Organism cannot contain the term 'metagenome'.

MIMARKS Survey related

Use for any type of marker gene sequences, eg. 16, 18S, 23S, 28S rRNA or COI obtained directly from the convironment, without culturing or identification of the organisms. Organismmust be a metagenome, where lineage starts with unclass if the sequences and scientific name ends with 'metagenome'.

MIMS Environmental/Metagenome

Use for environmental and metagenome sequences. Organism



Relative loc of sample

 Host disease Host sex + age Host tissue sampled

Animal housing system

Human/animal host

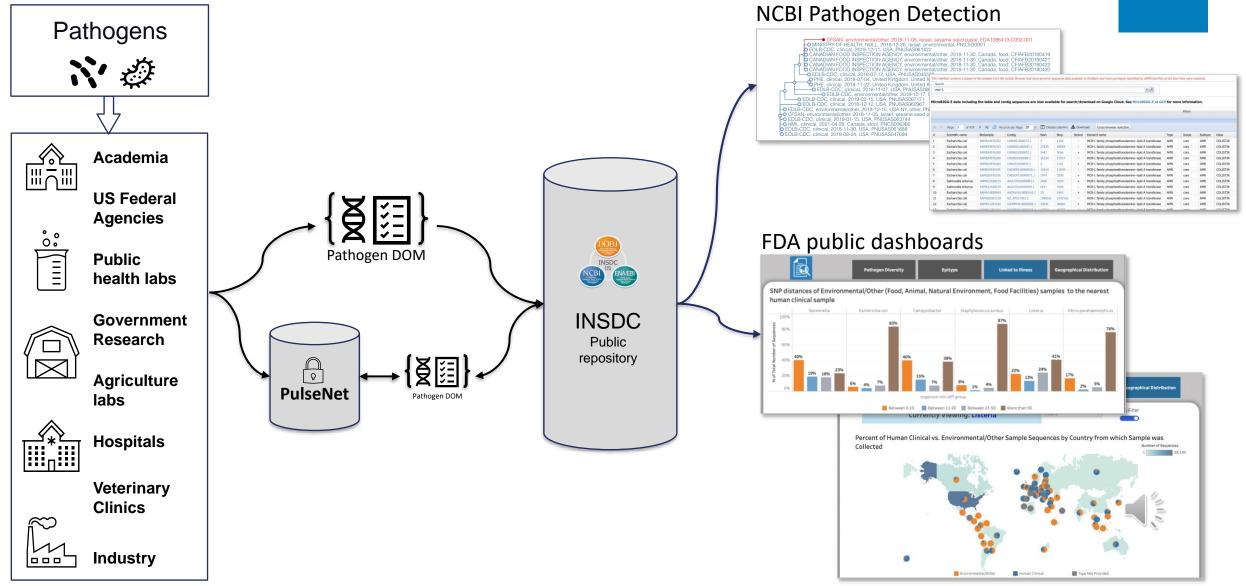
Animal environment

Host

Antimicrobials in food

US enteric pathogen surveillance





Salmonella tahini clusters highlight global contribution

130,935 Clusters currently tracked.

 CFSAN, environmental/other, 2018-11-05, Israel, sesame seed paste, FDA1086413-C002-001 MINISTRY OF HEALTH, NULL, 2018-12-26, Israel, environmental, PNILSO000 EDLB-CDC, clinical, 2018-12-11, USA, PNUSAS061422 CANADIAN FOOD INSPECTION AGENCY, environmental/other, 2018-11-30, Canada, food, CFIAFB20180419 CANADIAN FOOD INSPECTION AGENCY, environmental/other, 2018-11-30, Canada, food, CFIAFB20180421 CANADIAN FOOD INSPECTION AGENCY, environmental/other, 2018-11-30, Canada, food, CFIAFB20180422 CANADIAN FOOD INSPECTION AGENCY, environmental/other, 2018-11-30, Canada, food, CEIAEB20180420 –O EDLB-CDC, clinical, 2018-07-12, USA, PNUSAS045588 PHE, clinical, 2018-07-04, United Kingdom: United Kingdom, human, 553780 PHE, clinical, 2018-11-22, United Kingdom: United Kingdom, human, 550813 EDLB-CDC, clinical, 2018-11-07, USA, PNUSAS060239 EDLB-CDC, environmental/other, 2018-12-17, USA:NY, food, PNUSAS063364 EDLB-CDC, clinical, 2019-02-15, USA, PNUSAS067171 EDLB-CDC, clinical, 2018-12-12, USA, PNUSAS062967 EDLB-CDC, environmental/other, 2018-12-10, USA:NY, other, PNUSAS062706 CFSAN, environmental/other, 2018-11-05, Israel, sesame seed paste, FDA1086413-C001-001 EDLB-CDC, clinical, 2019-01-15, USA, PNUSAS063744 •• NML, clinical, 2021-04-28, Canada, stool, PNCS006366 EDLB-CDC, clinical, 2018-11-30, USA, PNUSAS061688 EDLB-CDC, clinical, 2018-08-24, USA, PNUSAS047694

- United Kingdom (PHE)
- United States (GenomeTrakr, PulseNet)
- Canada (CFIA, NLM)
- Israel

PHE, clinical, 2021-11-30, United Kingdom: United Kingdom, human, 1537643
 PHE, clinical, 2018-09-19, United Kingdom: United Kingdom, human, 572351
 PHE, clinical, 2020-05-22, United Kingdom: United Kingdom, human, 945093
 CFSAN, environmental/other, 2021-01-14, Sudan, sweet tahini, FDA1152646-C001-001

- United Kingdom
- United States (GenomeTrakr)

PHE, clinical, 2018-09-21, United Kingdom: United Kingdom, human, 598916
CFSAN, environmental/other, 2021-02-22, Turkey, dessert, FDA1153560-C002-005
CFSAN, environmental/other, 2015-02-10, Turkey, sesame seed paste, FNE0103
EDLB-CDC, clinical, 2018-08-02, USA, 2013K-0799
CFSAN, environmental/other, 2015-02-11, USA:MN, sesame seed paste, MDH-2013-00199
environmental/other, 2018-11-03, Turkey, ground sesame seed paste tahini, CFSAN076213
CFSAN, environmental/other, 2016-12-06, Turkey, dessert cocoa, FNE0021
CFSAN, environmental/other, 2015-02-10, Turkey, sesame seed paste, FNE0140
CFSAN, environmental/other, 2015-04-12, United Kingdom: London, human, 68664
GBRU, clinical, 2015-04-12, United Kingdom: London, human, 41941
GBRU, clinical, 2015-04-12, United Kingdom: London, human, 45966

- United Kingdom (PHE, GBRU)
- United States (GenomeTrakr, PulseNet)

PULSENET, clinical, 2021-07-22, USA, PNUSAS214535
 PULSENET, clinical, 2021-07-21, USA, PNUSAS214328
 NATIONAL INSTITUTE OF PUBLIC HEALTH - NATIONAL INS, clinical, 2021-05-10, Poland, See PHE, clinical, 2018-09-14, United Kingdom: United Kingdom, human, 579263
 PHE, clinical, 2019-01-30, United Kingdom: United Kingdom, human, 524796
 CFSAN, environmental/other, 2021-07-06, Jordan, sesame seed paste tahini, FDA1162570-C002-005

- United Kingdom (PHE)
- United States (GenomeTrakr, PulseNet)
- Poland

	Avian influenz	a (HPAI)	Influenza A+B	Food-borne pa	athogens ^a			
Case study area								
Institution	APHA (UK)	FLI (DE)	EMC (NL)	IZSLER (IT)	INEI-ANLIS (ARG)	MDH (US)	PHAC (CAN)	PHE (UK)
Outbreak or routine surveillance	Outbreak	Outbreak	Routine surveillance	Routine surveillance	Routine surveillance	Routine surveillance	Routine surveillance	Routine surveillance
	26	30	630	175	320	1,767	8,630	15,791
Number of samples in reference period	in 8 months	3 months	5 months	12 months	12 months	12 months	12 months	12 months
WGS								
Sequencer used	Illumina MiSeq	IonTorrent PGM	Nanopore GridION	Illumina MiSeq	Illumina MiSeq	Illumina MiSeq	Illumina MiSeq	Illumina HiSeq
	1–2	6	30	24	12	24	32	Processing: 40
Batch size for sample processing/sequ encing								Sequencing: 96
Equipment	€ 58.53	€ 210.71	€ 2.50	€ 163.49	€ 43.02	€ 29.53	€ 75.90	€ 35.23
Consumables	€ 830.97	€ 254.88	€ 33.52	€ 165.37	€ 104.62	€ 104.40	€ 69.75	

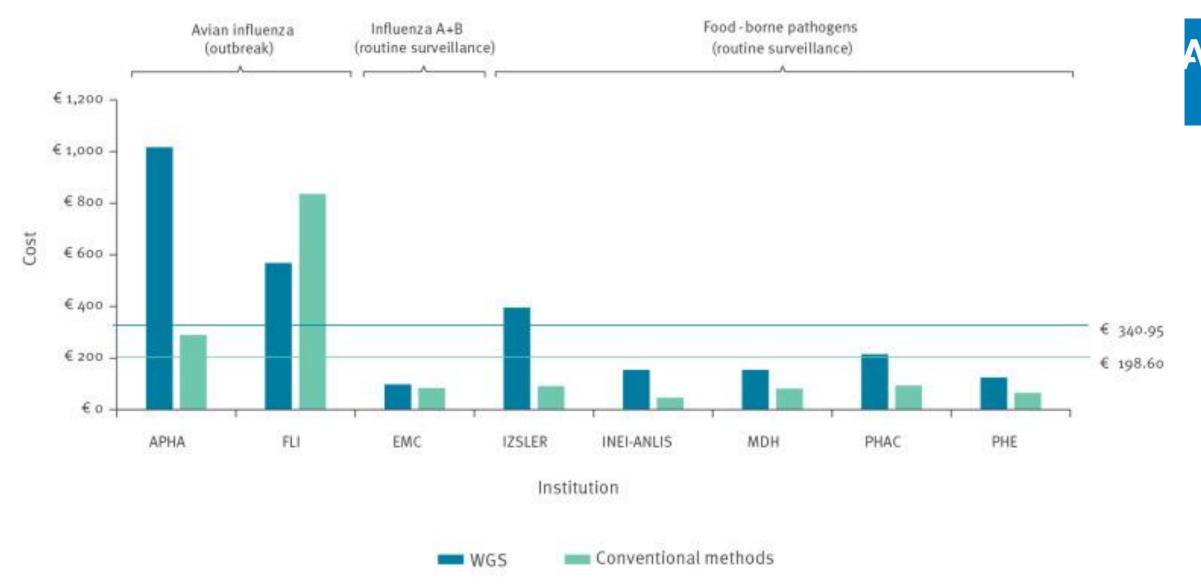
Economic impact studies

Overview of per-sample costs of whole genome sequencing vs conventional methods, by cost type, case studies covering a specified reference period between 2016 and 2019 (n = 8 institutes)

INEI-ANLIS Dr Carlos G Malbrán, Buenos Aires, Argentina

Alleweldt et al. Economic evaluation of whole genome sequencing for pathogen identification and surveillance—results of case studies in Europe and the Americas 2016 to 2019. Euro. Surveill. 2021 Mar 4; 26(9): 1900606 **<u>Celine Nadon</u>,

FDA



Over-all per-sample costs of whole genome sequencing vs conventional methods, case studies covering a specified reference period between 2016 and 2019 (n = 8 institutes).

Case study institution	IZSLER (IT)	INEI-ANLIS (ARG)	MDH (US)	PHAC (CAN)	PHE (UK)	Average
Cost per sample (WGS)	€ 395.14	€ 154.49	€ 154.51	€ 215.36	€ 124.59	€ 208.82
Cost per sample (conventional methods)	€91.87	€ 46.61	€ 81.16	€ 94.29	€ 65.46	€ 75.88
Differential cost of WGS compared with conventional methods	€ 303.27	€ 107.88	€ 73.35	€ 121.07	€ 59.13	€ 132.94
Number of samples per year (<i>Salmonella</i>)	110	128	1,010	8,273	10,147	3,934
Total additional costs per year due to the use of WGS	€ 33,360	€ 13,809	€ 74,084	€ 1,001,623	€ 599,992	€ 344,573
Average cost per reported case of salmonellosis	€ 12,124	€ 11,821	€ 13,225	€ 12,174	€ 12,401	€ 12,349
Number of reported cases of salmonellosis that need to be avoided to break even	2.8	1.2	5.6	82.3	48.3	28.0
Number of cases of salmonellosis reported annually ^a	276 ^b	758	906	7,665	8,770	4,404
Percentage of total number of reported cases of salmonellosis that need to be avoided to	1.0%	0.2%	0.6%	1.1%	0.6%	0.7%

Results of break-even analysis, whole genome sequencing vs conventional methods, case studies covering a specified reference period between 2016 and 2018 (n = 5 institutes).

Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna, Parma, Italy



Results

On a per-sample basis, WGS was between 1.2 and 4.3 times more expensive than routine conventional methods. However, WGS brought major benefits for pathogen identification and surveillance, substantially changing laboratory workflows, analytical processes and outbreaks detection and control. Between 0.2% and 1.1% (on average 0.7%) of reported salmonellosis cases would need to be prevented to break even with respect to the additional costs of WGS.

Conclusions

Even at cost levels documented here, WGS provides a level of additional information that more than balances the additional costs if used effectively. The substantial cost differences for WGS between reference laboratories were due to economies of scale, degree of automation, sequencing technology used and institutional discounts for equipment and consumables, as well as the extent to which sequencers are used at full capacity.

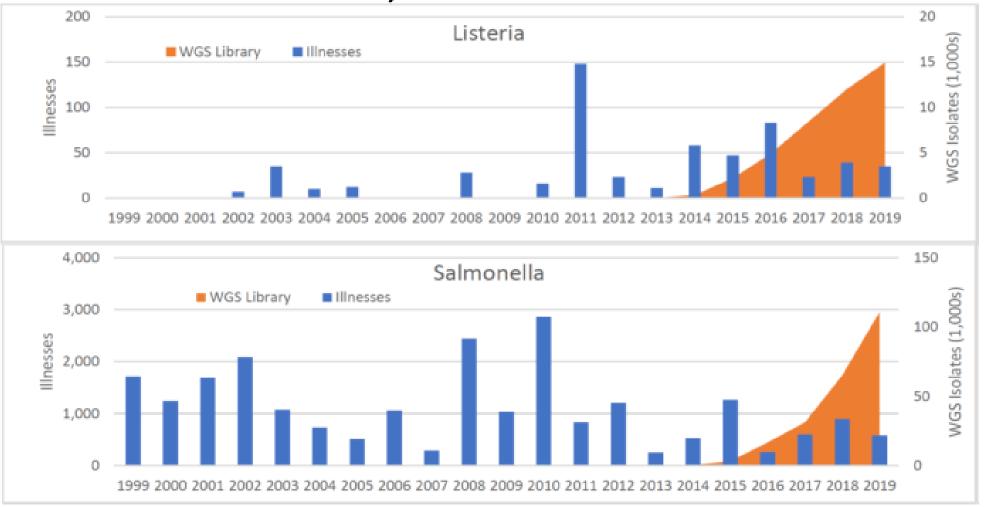
Ford et al. Cost of whole genome sequencing for non-typhoidal Salmonella enterica. PLoS ONE 2021; 16(3):e0248561 For Australia break even is 1.9%



FDA

16

Data and Summary Statistics



Brown et al. (2021) An economic evaluation of the Whole Genome Sequencing source tracking program in the U.S. PLoS ONE 16(10): e0258262.

$$SV = \underbrace{\left[p_x * x - (c_x(x) + c_e(e(WGS))\right]}_{profit \ Function} - \underbrace{\left[C_I * x * \gamma_I(e(WGS)) * n_I(WGS)\right]}_{public \ health \ externatility \ function} - \underbrace{\left[c_{WGS}(WGS)\right]}_{implentation \ cost}$$

$$I_O = \underbrace{x * \gamma_I(e(WGS))}_{probability \ outbreak} * \underbrace{n_I(WGS)}_{number \ of \ illnesses} * \underbrace{\alpha_O(WGS)}_{probability \ illnesses}$$

$$probability \ illnesses}_{are \ observed}$$

$$Y_{p,t} = \beta_0 + \beta_1 WGS_library_{p,t} + \varepsilon_{p,t}$$

$$Y_{p,t} = \beta_0 + \beta_1 WGS_library_{p,t} + \beta_2 X_{p,t} + \varepsilon_{p,t}$$

$$Y_{p,t} = \beta_0 + \beta_1 WGS_{-library_{p,t}} + \beta_2 X_{p,t} + \beta_3 FSMA_t + \varepsilon_{p,t}$$

Benefits = $\hat{\beta}_1 x$ WGS Isolates x Underreporting Multiplier x Monetary Loss

If you want to do these kinds of calculations, please let our PhD economist talk to yours.



Economic Evaluation of WGS Reduces the Burden of Illness

	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)

Total Burden Averted (in millions)

18

Economic Impact

- 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.

PLOS	ONE			
🔓 OPEN ACCESS 💋 PEER-F	REVIEWED			
Source track	king progr rd, Michael C. Bazac	am in the U	.S. Travis Minor	ne Sequencing
Published: October 6, 202	Authors	Metrics	Comments	Media Coverage
Abstract	Abstrac	rt		
Introduction Materials and methods Results Discussion Conclusion Supporting information Acknowledgments References	Sequencinç analysis of related illne tests using 1999 throuy (NCBI) Pati outbreak illi are consisti added to th per WGS p Carlo analy nearly \$500	g (WGS) Network in 2013 Whole Genome source to seese through theoretical data from FDA regulated gh 2019 and examine the hogen detection program nesses for three pilot pade ent with the theoretical m e public NCB database athogen, per year. Empiny risis to estimate benefits a million, compared to an	a, as a tool to improve foo racking implementation or empirical, and cost bene food commodity outbreal effect of the National Ce of source tracking WGS hogens (E. coft, Listeria, a, odel and suggest that eas is associated with a reduc ical results are connected approximately \$22 millior	enomeTrakr Whole Genome d safety. This study presents an opticntial food contamination an fit analyses. We conduct empiric ks gamering FDA response from neter or Biotechnology Informatic isolates collected in the U.S. on and Salmonella, Empirical result. Empirical result. Empirical result on d approximately 6 illnesses tion of approximately 6 illnesses to existing illnerature for a Morte al health benefits are estimated a investment by public health
Reader Comments	agencies. E	Even under conservative	assumptions, the program	n likely broke even in its second



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.

FDA

Price et al. 2023 A systematic review of economic evaluations of whole-genome sequencing for the surveillance of bacterial pathogens. Microb Genom 2023; 9(2). Discussion of 9 different economic impact studies.



There were significant variations in the research questions addressed in the various publications yet, most studies demonstrated cost savings due to WGS that were largely attributed to averted cases of infection.

For this benefit to be realized maximally, WGS needs to be employed early in the analytical pipeline. Conversely, delay in the use of WGS reduces the benefits, as early detection of outbreaks enables timely implementation of interventions to interrupt transmission.

More economic evidence of WGS in public health settings is required to foster wider applications of WGS as a surveillance tool in public health.

We dedicate this work to Robert Stones FERA

