



## Review

# Whole genome sequencing uses for foodborne contamination and compliance: Discovery of an emerging contamination event in an ice cream facility using whole genome sequencing



Marc W. Allard\*, Errol Strain, Hugh Rand, David Melka, William A. Correll, Leslie Hintz, Eric Stevens, Ruth Timme, Sara Lomonaco, Yi Chen, Steven M. Musser, Eric W. Brown

US Food and Drug Administration, 5001 Campus Drive, College Park, MD, USA

## ARTICLE INFO

## Keywords:

Compliance  
Infection control  
Whole genome sequencing  
GenomeTrakr  
Foodborne pathogen  
NCBI pathogen detection

## ABSTRACT

We review how FDA surveillance identifies several ways that whole genome sequencing (WGS) improves actionable outcomes for public health and compliance in a case involving *Listeria monocytogenes* contamination in an ice cream facility. In late August 2017 FDA conducted environmental sampling inside an ice cream facility. These isolates were sequenced and deposited into the GenomeTrakr databases. In September 2018 the Centers for Disease Control and Prevention contacted the Florida Department of Health after finding that the pathogen analyses of three clinical cases of listeriosis (two in 2013, one in 2018) were highly related to the aforementioned *L. monocytogenes* isolates collected from the ice cream facility. In 2017, FDA returned to the ice cream facility in late September 2018 and conducted further environmental sampling and again recovered *L. monocytogenes* from environmental subsamples that were genetically related to the clinical cases. A voluntary recall was issued to include all ice cream manufactured from August 2017 to October 2018. Subsequently, FDA suspended this food facility's registration. WGS results for *L. monocytogenes* found in the facility and from clinical samples clustered together by 0–31 single nucleotide polymorphisms (SNPs). The FDA worked together with the Centers for Disease Control and Prevention, as well as the Florida Department of Health, and the Florida Department of Agriculture and Consumer Services to recall all ice cream products produced by this facility. Our data suggests that when available isolates from food facility inspections are subject to whole genome sequencing and the subsequent sequence data point to linkages between these strains and recent clinical isolates (i.e., < 20 nucleotide differences), compliance officials should take regulatory actions early to prevent further potential illness. The utility of WGS for applications related to enforcement of FDA compliance programs in the context of foodborne pathogens is reviewed.

## 1. Introduction

Whole genome sequencing (WGS) has been used to provide detailed characterization of foodborne pathogens. Genomes comprising diverse groups of pathogenic genus and species including *Listeria monocytogenes*, *E. coli*, *Salmonella*, *Campylobacter* and *Vibrio* have provided insight into the genetic make-up and relatedness of these pathogens. Numerous government agencies, industry, and academia have developed applications in food safety using WGS approaches such as outbreak detection and characterization, source tracking, and determining the root cause of a contamination event. In this particular case study, the FDA GenomeTrakr database, PulseNet and The NCBI Pathogen Detection Portal were used to cluster and characterize low levels of

sporadic illnesses caused by *L. monocytogenes* that had been documented in the United States. By fusing WGS with the GenomeTrakr database, a relationship emerged early between sequences from environmental swabs and those derived from clinical *L. monocytogenes* samples. This observation supported a potential relationship between illnesses and non-food contact environmental samples obtained as part of FDA inspection activities at this food manufacturing facility. This was an important lead as traditional cluster detection methodologies based on PFGE did not identify these cases as an outbreak cluster so they had not previously been investigated for a common vehicle, prior to the linkage referred to here with the FDA samples. As WGS testing and additional follow-up questioning ensued, public health officials concluded that the illnesses were, in fact, related exposure to the suspect

\* Corresponding author.

E-mail address: [Marc.allard@fda.hhs.gov](mailto:Marc.allard@fda.hhs.gov) (M.W. Allard).

<https://doi.org/10.1016/j.meegid.2019.04.026>

Received 5 March 2019; Received in revised form 25 April 2019; Accepted 26 April 2019

Available online 27 April 2019

1567-1348/ Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

**Table 1**Metadata of environmental and clinical isolates of *Listeria* associated with ice cream. Sequences are deposited at NCBI.<sup>a</sup>

Strain	Isolate	Create date	Location	Isolation type	BioSample	Figure
CFSAN004306	PDT000102694.2	1/21/2016	USA:OK	bbq beans w/beef	SAMN03354331	1
FDA00011473	PDT000200081.2	4/5/2017	USA:VA	Environmental swab	SAMN06649805	1
FDA00012171	PDT000244002.2	9/26/2017	USA:FL	Environmental swab	SAMN07702403	1
FDA00012170	PDT000244003.2	9/26/2017	USA:FL	Environmental swab	SAMN07702404	1
FDA00012169	PDT000244004.2	9/26/2017	USA:FL	Environmental swab	SAMN07702405	1
FDA00012166	PDT000244007.2	9/26/2017	USA:FL	Environmental swab	SAMN07702408	1
FDA00012164	PDT000244009.2	9/26/2017	USA:FL	Environmental swab	SAMN07702410	1
FDA00013579	PDT000400899.3	11/8/2018	USA:FL	Environmental swab	SAMN10368217	1
FDA00013578	PDT000400902.3	11/8/2018	USA:FL	Environmental swab	SAMN10368218	1
PNUSAL000220	PDT000000969.3	11/19/2013	USA	Clinical	SAMN02361174	2
PNUSAL000373	PDT000001329.3	11/19/2013	USA	Clinical	SAMN02400182	2
FDA00012168	PDT000244005.2	9/26/2017	USA:FL	Environmental/other	SAMN07702406	2
FDA00012167	PDT000244006.2	9/26/2017	USA:FL	Environmental/other	SAMN07702407	2
FDA00012165	PDT000244008.2	9/26/2017	USA:FL	Environmental/other	SAMN07702409	2
PNUSAL004217	PDT000362180.1	8/13/2018	USA	Clinical	SAMN09813030	2
FDA00013582	PDT000393756.1	10/21/2018	USA:KS	Environmental/other	SAMN10251102	2
FDA00013583	PDT000393775.1	10/21/2018	USA:KS	Environmental/other	SAMN10251093	2
FDA00013584	PDT000393777.1	10/21/2018	USA:KS	Environmental/other	SAMN10251092	2
FDA00013577	PDT000400898.3	11/8/2018	USA:FL	Environmental/other	SAMN10368216	2

<sup>a</sup> Genome Accessions are provided for FDA isolates from the inspections involving ice cream that were assembled and submitted to GenBank; SRA accessions are provided for all other isolates. Metadata and WGS associated with clinical isolates were provided by CDC PulseNet.

ice cream eaten by ill consumers. The authorities reported this information to the food company, and the firm issued a voluntary recall notice for all of their ice cream products. This was followed by suspension of the facility's registration. In this case, strong genomic evidence mobilized the Office of Compliance early, documenting an example where application of WGS and the resultant sequence evidence derived from it provided clear and actionable data to support further investigation.

WGS has been applied to the traceback of foodborne pathogens where insufficient genetic resolution bogged down existing subtyping tools (Lienau et al., 2011; Allard et al., 2013; Hoffmann et al., 2014). The primary application of using WGS for pathogen surveillance was to look for close matches between clinical and environmental isolates, although any WGS linkage can support an investigation and direct additional inquiry (Mellmann et al., 2014; Ashton et al., 2015; Dallman et al., 2016a). In addition to finding similar "matches" in the database, another advantage of WGS data as the premier subtyping tool for foodborne pathogen traceability was the ability to reconstruct the evolutionary history within a cluster of clonal isolates, enabling the identification of a recent common ancestor or source of a contamination event (Chen et al., 2017a; Chen et al., 2017b; Hoffmann et al., 2016). These phylogenetic cluster analyses are highly accurate and can be made readily available, often making the genomic linkage the first lead in an outbreak or compliance investigation, whether this is linking clinical, environmental or both kinds of isolates. In an ideal pathogen surveillance network, these genetic linkages need to be analyzed, investigated and acted on rapidly. Additionally, WGS data diversity in the database should represent the real-world, global, microbial diversity and should include sequences from all countries currently gathering pathogen genomes. Such a database should be manifold in structure and comprise genome sequences from food, environmental and clinical sources (Allard et al., 2016; Gardy and Loman, 2018). A contamination event discovered in the United States may involve foods that were traded on the other side of the world as documented by a recent case study involving trade with Australia (Kwong et al., 2016a). A dearth of real-time WGS data sharing remains in some parts of the world, and despite several conspicuous reasons for this lack of open data availability (Aarestrup and Koopmans, 2016; Nadon et al., 2017), the concern remains that some contamination events likely will not be fully elucidated thus allowing globally traded contaminated product to linger in the food supply chain for extended periods. Restated, the sharing of all WGS data in real-time globally enables timely access for

analysis of the data to all relevant stakeholders engaged in food safety surveillance and mitigation (Allard et al., 2016). Moreover, the continued pooling of data to an open and common curated source(s) allows for refinement of geographical mapping of foodborne pathogen species on a global scale. Unfortunately, many countries still do not actively share genomic data or metadata, an act that clearly stifles growth of global open-source WGS databases, and potentially limits insight into the sources of foodborne contamination and outbreaks on an international scale (Van Walle et al., 2018; Yozwiak et al., 2015). Both the Food and Agricultural Organization (FAO, <http://www.fao.org/food/food-safety-quality/a-z-index/wgs/en/>) and World Health Organization (WHO, [http://www.who.int/foodsafety/publications/foodborne\\_disease/wgs\\_landscape/en/](http://www.who.int/foodsafety/publications/foodborne_disease/wgs_landscape/en/)) are developing guidance documents on the value and barriers of WGS technologies for many countries, and the intrinsic value of sharing these data. As an example, the GenomeTrakr database is made publicly available in real-time in support of these lofty goals (Allard et al., 2016). The database is housed within the National Center for Biotechnology Information (NCBI) and their Pathogen Detection web tools (<https://www.ncbi.nlm.nih.gov/pathogens/>) which provide WGS genomic linkages and phylogenetic trees daily, funded by NCBI. Additionally, the Centers for Disease Control and Prevention (CDC) led PulseNet Network also uploads all SRA genomic data to NCBI as does the United States Department of Agriculture's Food Safety and Inspection Services (USDA-FSIS). Together these data comprise a one-health oriented database where clinical, veterinary, food and environmental WGS samples are combined to discover novel linkages among foodborne pathogens from various niches. Each of these governmental agencies releases minimal metadata for the isolates including who, what, when and where. For example, NCBI biosample SAMN07702403 for strain FDA00012171 lists this isolate as an environmental swab from FL, with a collection date of 2017-08-30 and sequenced from FDA lab 0973150-044-001. The federal laboratories also keep confidential data that is not made publicly available such as firm names, specific locations of food manufacturers, and commercial details of ingredient sources for each specific food product. Confidential data often is shared among professional public health agencies in-order to solve outbreaks. The *Listeria monocytogenes* ice cream case study presented here documents the value of intervention, even when only a few isolates match genomically and or are observed over a long period of time (Table 1). It also highlights the critical need to conduct WGS testing on all clinical and environmentally derived isolates of *Listeria* infected and ill patients in real-time as these data have the potential to provide critical evidence

in support of compliance investigation and regulatory response.

## 2. Materials and methods

Minimal methods are provided for this case study to document how FDA surveillance identifies several ways that WGS improves actionable outcomes for public health and compliance in cases involving *Listeria monocytogenes* contamination. More detailed methods have been published for several *Listeria* cases (Chen et al., 2016, Chen et al., 2017b, Li et al., 2017, Chen et al., 2017c; Burall et al., 2016). In late August 2017 FDA conducted environmental sampling inside an ice cream facility as part of an FDA sampling assignment designed to gather baseline environmental surveillance data and other inspection information on ice cream facilities. When investigators collect an environmental sample this generally consists of one hundred subsamples across zones 1 (e.g. a food contact surfaces) and 2 (e.g. likely to contaminate a food contact surface). At each facility roughly 200 environmental swabs are collected and assessed for the presence of foodborne pathogens. If positive isolates are observed, then often a second inspection is conducted where roughly 300 to 400 additional environmental swabs are collected and assessed for the presence of foodborne pathogens. The timing and extent of additional inspections depends on the circumstances observed. Libraries of genomic DNA were prepared with the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA) and subsequently sequenced on a MiSeq (Illumina, San Diego, CA) according to the manufacturer's instruction. Quality control and assessment of the genomic data followed FDA validated and published methods (Timme et al., 2019a, 2019b). Paired-end reads were assembled using SPAdes software v 3.X with default settings and invoking the `-careful` and `-cov-cutoff auto` options (Bankevich et al., 2012). Annotations of assemblies were processed using the NCBI's Prokaryotic Genome Annotation Pipeline (PGAP, Tatusova et al., 2016). Fastq files obtained from MiSeq run were used as input of CFSAN SNP pipeline v 0.8 (Davis et al., 2015). The CFSAN SNP pipeline was used to generate a SNP matrix, with high-density SNP regions filtered (i.e., those with > 3 SNPs within 1Kb). The GARLI (Genetic Algorithm for Rapid Likelihood Inference) v 2.01 was used to reconstruct the maximum likelihood (ML) phylogenetic tree. We searched the best tree with 10 ML replicates and conducted bootstrap analysis for 1000 replicates (ratematrix = 6rate; ratehetmodel = gamma). Python program SumTrees was used to generate the ML phylogenetic tree with bootstrap values (<https://pythonhosted.org/DendroPy/programs/sumtrees.html>). The phylogenetic relationships among the strains in comparison with others isolated worldwide and SNP Cluster nomenclature were checked at the NCBI Pathogen Detection url (<https://www.ncbi.nlm.nih.gov/pathogens>). Identification of acquired antimicrobial resistance genes (Clausen et al., 2016; Knowles et al., 2016; McDermott et al., 2016) and plasmids within the outbreak isolates were performed using ResFinder v3.0 (Zankari et al., 2012; Zankari et al., 2017) and PlasmidFinder v1.3 (Carattoli et al., 2014), respectively. As an aside, we believe that general food microbiologists will greatly benefit from a thorough understanding of precisely how to these powerful new genomic tools are being used. Part of learning a new field is also learning the language. Detailed glossaries are readily available online and provide definitions for molecular biology, genomics and phylogenetics terminology for example see <https://www.ncbi.nlm.nih.gov/projects/genome/glossary.shtml> and <https://ucmp.berkeley.edu/glossary/gloss1phylo.html> as examples."

## 3. Results

*Listeria monocytogenes* derived from five environmental subsamples from FDA's 2017 sample 0973150 (subs FDA00011473, FDA00012171, FDA00012170, FDA00012169, FDA00012166, FDA00012164), and 2018 samples FDA00013579 and FDA00013578 are genetically related based on SNP distance ( $\leq 20$  SNPs) and a monophyletic relationship (NCBI cluster PDS000004821.11, Fig. 1, Table 1) in the resultant

phylogenetic tree. Three subsamples from FDA's 2017 sample 973,150 (subsamples FDA00012168, FDA00012167, and FDA00012165) and 2018 sample FDA00013577 formed a second distinct cluster (NCBI cluster PDS000025188.7, Fig. 2) and were determined to be the same strain as three CDC-sequenced clinical isolates (two clinical isolates from 2013 (PNUSAL000220, PNUSAL000373) and one clinical isolate from 2018 (PNUSAL004217)) based on SNP distance ( $\leq 20$  SNPs) and a monophyletic relationship with perfect bootstrap support (100%). The NCBI pathogen detection data analysis pipeline clusters isolates that are between 0 and 50 SNP differences apart. These strains are distinct and only represented by these 7 isolates out of  $\sim 21,000$  *L. monocytogenes* isolates in the NCBI Pathogen database at the time of analysis. These strains had not been isolated previously from food or environmental samples from any other firm for which WGS data is available. Two strains of *L. monocytogenes* isolated from environmental samples collected by FDA investigators during an ice cream facility inspection in 2018 (FDA00013579 and FDA00013578) were determined to be the genetically related to strains isolated from the 2017 inspection of the same facility, indicating the presence of multiple resident *L. monocytogenes* strains (Table 1, Fig. 1). Other strains were determined to be genetically related to clinical isolates from 2013 and 2018 (Table 1, Fig. 2) indicating that this strain of *L. monocytogenes* can cause illness and has potentially been resident since at least 2013. The WGS findings strongly predicted that an epidemiological link would be found, and this was confirmed after exposure findings were completed that pointed to these contaminated ice cream products as the causative agent for the three clinical illnesses reported in 2013 and 2018. FDA, CDC and the Florida Department of Health discussed these findings with the firm and the firm issued a voluntary recall notice for all ice cream products that was then followed by FDA suspending the firm. For a detailed description of FDA safety alerts and advisories see (<https://www.fda.gov/Food/RecallsOutbreaksEmergencies/SafetyAlertsAdvisories/ucm624256.htm>). The phylogenetic results also are available at The NCBI Pathogen Detection web site by searching for the NCBI clusters PDS000004821.11 and PDS000025188.7 ([https://www.ncbi.nlm.nih.gov/pathogens/isolates/#/search/taxgroup\\_name:%22Listeria%20monocytogenes%22%20AND%20PDS000025188.7](https://www.ncbi.nlm.nih.gov/pathogens/isolates/#/search/taxgroup_name:%22Listeria%20monocytogenes%22%20AND%20PDS000025188.7)). This case study shows that low levels of sporadic *L. monocytogenes* contamination induced illnesses that were linked by WGS to predict a common source and root cause. With application of WGS, a genomic linkage to a food manufacturing location was quickly and convincingly established allowing for mobilization of FDA's compliance officials to act rapidly. This is not the first instance of the use of WGS to identify contamination events caused by very low levels of sporadic contamination persisting in a food facility, nor is it the first reporting of *L. monocytogenes* in an ice cream product (Pouillot et al., 2016, Chen et al., 2016, Chen et al., 2017b, Li et al., 2017, Chen et al., 2017c; Burall et al., 2016). PulseNet has recently replaced PFGE with WGS as the primary molecular surveillance tool for *Listeria* (Jackson et al., 2016), and other foodborne pathogens such as enterohemorrhagic *E. coli* and *Salmonella* are soon to follow. It is notable that WGS evidence has improved most contamination investigations as some PFGE matches may not trigger cluster investigation since related cases spread out over several months or years and often do not rise above the background illness. With the adoption of WGS analysis for all clinical, food and environmental isolates, the potential for more refined and accurate cluster detection now exists. This approach, which integrates WGS into investigatory workflow will only further strengthen compliance efforts as more data are shared in real-time and globally. Several bottlenecks to the investigatory process persist as part of current compliance and outbreak investigational workflow. Such checkpoints in the process include facility inspections, epidemiological food exposure determinations with statistical confidence, and record review and auditing all of which are slower and more interactive processes requiring investigators on the ground, patient questionnaires, and/or follow up additional inspections and laboratory analyses. WGS-based phylogenetic clustering provides a



**Fig. 1.** Phylogenetic results of *Listeria* clustering environmental isolates, NCBI Pathogen detection cluster PDS000004821 (as of date of April 23, 2019). The number line corresponds to SNPs, where the branch length is proportional to the number of SNPs present. Detailed SNP differences also are available at the NCBI Pathogen Detection site by searching for either strain IDs or SNP cluster IDs. \* indicates contamination cluster.



**Fig. 2.** Phylogenetic results of *Listeria* clustering environmental swabs and clinical isolates, NCBI Pathogen detection cluster PDS0000025188.7 (as of date of April 23, 2019). The number line corresponds to SNPs, where the branch length is proportional to the number of SNPs present. Detailed SNP differences also are available at the NCBI Pathogen Detection site by searching for either strain IDs or SNP cluster IDs. \* indicates contamination cluster.

new and valuable step in this workflow because of the rapidity and accuracy of the resultant genetic linkages. Additional methods have been proposed to speed up investigations such as predicting zoonotic source attribution using machine learning from WGS data (Zhang et al., 2019).

## 4. Discussion

### 4.1. What is unique about this case?

WGS provided detailed information and an increased degree of certainty in identifying the source of this sporadic contamination cluster. WGS sequencing mitigated the associated compliance investigation substantially. Indeed, its specific role augmented FDA's decision-making processes in a series of observations including supporting the thesis that these illnesses were a cluster and linked to specific *L. monocytogenes* isolates found in an ice cream facility – likely source of this contamination event. It is noteworthy that traditional epidemiology, after exposure data was collected and analyzed, further supported the relationship between the illness isolates and isolates from the processing facility. Arguably, this particular event is a poignant example of where WGS has aided in solving a food safety concern that may not have been solved previously. The continued adoption of WGS applications makes it far more likely that contamination sources associated with sporadic foodborne illnesses will now be able to be identified (Van Walle et al., 2018; Franz et al., 2016; Ronholm et al., 2016; Dailman et al., 2016b). FDA and their regulatory offices now include genomic evidence when disclosing compliance investigation details to firms. With this additional information, companies can determine whether a voluntary recall notice is appropriate and can also act to improve preventative controls that may have broken down in production processes. Voluntary recall is an important method to speed up the removal of contaminated product and quickly reduce the public health burden during a contamination event.

The GenomeTrakr and PulseNet databases further leverage WGS technology in an open-access platform which enables more rapid identification of the root sources of foodborne illness and, in turn, allows for faster public health response and reduced numbers of illnesses. WGS has been fully deployed across FDA field laboratories, and this detailed genomic approach is now being applied to all FDA-derived

isolates of *Salmonella*, *E. coli*, *Campylobacter* and *Listeria* in real-time as they are collected from food and environmental sources. To this end, FDA WGS analysts welcome legitimate scientific criticism to improve its analytical methods, but to be convincing, such criticism should include transparent side-by-side analyses and results that compare methods (i.e., between K-mer, SNP and MLST approaches) clustering the same genomic dataset(s). As an example, Nadon et al. (2017) recently reported several weaknesses of a SNP-based analytic approach, yet scientific evidence supporting these claims remains scant. On the contrary, FDA's GenomeTrakr workflow ascribes several beneficial attributes to its own SNP-based validated pipeline including the ability to provide for: 1) stable strain nomenclature; 2) international standardization; 3) curation; and 4) scalability (Davis et al., 2015; Allard et al., 2016; Timme et al., 2019b).

Public health professionals understand that the speed at which one can identify the cause of a contamination event has health consequences. This work identifies ways that WGS improves actionable outcomes for regulatory and public health entities. One way is to sequence more foodborne pathogens from food, environmental and clinical sources and to share them in real-time in publicly available databases. Global comparisons ensure that contamination due to international travel and trade is identified rapidly. Another way is when isolates are available from food facility inspections and the WGS data links these isolates to clinical isolates by a few nucleotide differences, FDA's Compliance officials may be activated early to launch regulatory actions. The ice cream *Listeria* case reported here, catalyzed by WGS, was an example of an early-warning match observed by FDA compliance investigators despite the sporadic nature of the illness and the length of time spanned. The continued success and support of WGS technology in food safety has made it indispensable for recognizing other contamination events. Indeed, it is now common practice to surveil for early matches between foods, environmental samples and clinical cases (Gwinn et al., 2017; Jackson et al., 2016; Kwong et al., 2016b; Lambert et al., 2017; Moura et al., 2017; Waldram et al., 2018).

In conjunction with this practice, FDA currently inspects roughly 800 facilities per year with 200 environmental swabs collected at each inspection for a total of 160,000 samples. From this effort, roughly one in four firms produce positive foodborne pathogen isolates all of which are sequenced and uploaded (2000 to 4000) into the GenomeTrakr database and NCBI Pathogen Detection where they are shared publicly,

including with PulseNet and USDA FSIS, as well as global public health laboratories. Depending on the phylogenetic clusters observed, FDA may act with further inspections, communications and advisories to coordinate public health response. The cumulative number of compliance actions/cases supported by WGS since 2014, in conjunction with the Office of Analytics and Outreach, is 370 cases, with 29 new cases in FY18 Q4, ending in October.

Over the past four years (2014–2017) the FDA office of compliance and office of regulatory affairs has collected and sequenced 11,672 isolates gathered from contaminated facilities during routine surveillance, as well as outbreak and post-outbreak response. As all of these isolates have known geographic locations and provenance, our biostatistics have been able to calculate several probabilities including:  $P(F|D \leq d)$ : What is the probability that 2 isolates are collected from the same food facility if their genetic distance is no more than  $d$  SNPs?;  $P(D \leq d|F)$ : What is the probability that the genetic distance of 2 isolates is no more than  $d$  SNPs if both of them are collected from the same food facility?; and  $P(D \leq d|F)$ : What is the probability that the genetic distance of 2 isolates is no more than  $d$  SNPs if they are collected from different food facilities? The details of this analytical work are published (Wang et al., 2018) and includes two data sets, one for *Salmonella* ( $n = 6351$ ) and one for *Listeria* ( $n = 5321$ ), including isolates sequenced from our freezer collections as well as recent samplings, 04/04/1999 to 07/24/2017. In the Wang et al. (2018) manuscript, the authors clearly define the parameters in the probability equations, and clearly state how their hypothesis testing is structured. For results after bias adjustment,  $P = .91$  for *Salmonella* and  $P = 100.00$  for *Listeria* when SNP distance is 0. The facility match probability reduces with increase SNP distance on a continuum, with the various probability distributions varying between the two species. The take home message is that when SNP distances are low there is a high probability that the isolates came from the same facility. These foodborne pathogens are evolving uniquely in the facilities that they reside, allowing FDA to leverage this bacterial characteristic to use in source tracking and root cause analysis. Closely related isolates ( $< 10$ – $20$  SNPs) generally originate from the same facility (Pightling et al., 2018). Facilities with multiple positive findings often had multiple strains.

This ice cream case study documents that, as a tool for food safety and compliance, WGS case reports are rare. Food facilities are faced with the similar infection control problems that hospitals and hospital networks are also dealing with, and that WGS could provide solutions for termed “precision epidemiology” (Ladner et al., 2019). The detailed metadata associated with isolates and the clusters into which they group into also supports efforts to improve risk assessment and risk management (Sanaa et al., 2018). This analysis provides further evidence that the high resolution and greater certainty of linkage from WGS, by better informing source attribution and root cause analyses, greatly enables government regulators to aid in decisions regarding removing contaminants from the food supply through source attribution and root cause analysis. The added value of a distributed network of desktop sequencers uploading WGS data from all contaminants available from the inspections of food facilities was documented as was the need for the regular inspection of high-risk commodities and the associated facilities that produce them. This study documented the need for a more rapid process that later included more automation and integration that is now part of the current methods supported by the GenomeTrakr and the NCBI Pathogen Detection tools. As speed and action are needed to reduce the impact and public health consequences for any contamination event, there was value of network support for both data collection, as well as bioinformatic support was obvious. The NCBI Pathogen detection web site currently provides automated phylogenetic analysis from validated data-analysis pipelines to regularly cluster all incoming data. This is further confirmed by FDA's validated CFSAN- SNP pipeline. The NCBI systems are designed to identify all clusters of 50 SNPs or less. Careful screening of all new genomes for phylogenetic clusters of 0–20 SNPs away is also regularly conducted, as

is the ability to look for any cluster that include at least one food and environmental isolate with at least one clinical isolate. Because of the compliance work documented in this study here, FDA bioinformaticists now watch for any new isolates clinical or veterinary strains that match FDA isolate genomes to determine whether additional investigation is warranted.

Another strength of making the genomic data public available is that it allows food production and testing companies the ability to download all GenomeTrakr data and validated software locally so that they can identify potential issues and minimize the impact to their brand without a need to upload their data to the public database. Therefore, companies can take advantage directly of the data for the good of public health and their customers. Several companies and third-party providers have already recreated the GenomeTrakr and NCBI Pathogen Detection data and tools thus permitting them to apply these important data directly toward augmenting their own preventive controls programs. FDA also provides the data publicly to further foster innovation and encourage more proactive stances among food industry partners. Several diagnostic companies have used these genomic data already to design better tests for enteric foodborne pathogens.

The economic savings gained from the adoption of WGS methods has been published (Jain et al., 2019) and suggests that Canada, as one example, could be saving from \$5,000,000 to \$90,000,000 USD annually in controlling and responding to *Salmonella*-derived food safety concerns alone. If one extrapolates the Canadian study to the United States costs for the burden of foodborne illnesses (Minor et al., 2015) then the savings annually would be estimated at billions of US dollars across the food safety sector. A more detailed economic estimate must be assessed to determine the exact figure, but the idea is clear that adoption of WGS both nationally and internationally could provide significant savings to public health, the burden of foodborne illness, and the costs associated therewith.

## 5. Conclusions

In this case study and review, we have shown how regulatory authorities and public health officials in government can leverage the power of WGS to identify, characterize and remove contaminants that emerge in the farm to fork continuum. These Data sharing and compliance work flows involving WGS are mechanisms which are highly coordinated both within the agency as well as across the sectors of food, environmental and clinical public health including the CDC, FDA, NCBI, USDA-FSIS and state departments of health and agriculture laboratories. The FDA GenomeTrakr database and network are specific for food safety, while the NCBI Pathogen Detection site also is available for other human pathogen groups including nosocomial and veterinary pathogens, so these mechanisms can be utilized both by other countries and for additional pathogenic bacterial species. The potency, degeneracy, and agnostic nature of WGS data allows for academicians, industry scientists, and or government officials to document and leverage common and specific applications for this powerful tool to improve food safety. This then can be adopted across the globe and allow for rapid identification, linkage, and prevention for both compliance actions and outbreak detection of foodborne contamination events as they emerge globally in the food and feed supply. It is noteworthy that the crosscutting application of WGS data inherently provides a mechanism to expand global one health objectives for full molecular epidemiological integration and linkage across the different disparate public health sectors including those comprising clinical, foods, and environmental pathogens. When WGS is tied to global real-time data-sharing timely discovery of foodborne pathogens can be tracked at a global level. These growing efforts suggest that it is time for global comprehensive international discussions, perhaps through the World Health Organization and or the Food and Agriculture Organization for coordinated efforts among governments to wholly exploit the full potential of this powerful new tool for improved food safety and public

health.

## Declaration of interests

No conflicts of interest, financial or otherwise, are declared by the authors. Trade names mentioned in the manuscript do not constitute an endorsement.

## Funding

Work was funded by internal research funding from the US Food and Drug Administration (FDA). This project was also supported in part by an appointment to the Research Participation Program at the CFSAN, FDA, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and FDA.

## Acknowledgements

We wish to thank our partners at United States Food and Drug Administration (FDA). We would like to thank the many collaborators in the GenomeTrakr network, CDC PulseNet, USDA-FSIS, Florida Department of Health, and the Florida Department of Agriculture and Consumer Services who provide their data for broad use and applications, and the data support that we all receive at the FDA and the NCBI. We thank Sebastian Cianci for editorial support and clear language recommendations.

## References

- Aarestrup, F.M., Koopmans, M.G., 2016. Sharing data for global infectious disease surveillance and outbreak detection. *Trends Micro.* 24 (4), 241–245.
- Allard, M.W., Luo, Y., Strain, E., Pettengill, J., Timme, R., Wang, C., Li, C., Keys, C.E., Zheng, J., Stones, R., Wilson, M.R., Musser, S.M., Brown, E.W., 2013. On the evolutionary history, population genetics and diversity among isolates of *Salmonella* Enteritidis PFGE pattern JEGX01.0004. *PLoS One* 8 (1), e55254.
- Allard, M.W., Strain, E., Melka, D., Bunning, V., Musser, S.M., Brown, E.W., Timme, R., 2016. Practical value of food pathogen traceability through building a whole-genome sequencing network and database. *J. Clin. Microbiol.* 54, 197.
- Ashton, P.M., Peters, T., Ameh, L., McAleer, R., Petrie, S., Nair, S., Muscat, I., de Pinna, E., Dallman, T., 2015. Whole genome sequencing for the retrospective investigation of an outbreak of *Salmonella* typhimurium DT 8. *PLOS Curr. Outbreaks* 10, 7.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotnik, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19 (5), 455–477.
- Burall, L.S., Grim, C.J., Mammel, M.K., Datta, A.R., 2016. Whole genome sequence analysis using JSpecies tool establishes clonal relationships between *Listeria* monocytogenes strains from epidemiologically unrelated listeriosis outbreaks. *PLoS One* 11 (3).
- Carattoli, A., Zankari, E., García-Fernández, A., Larsen, M.V., Lund, O., Villa, L., Aarestrup, F.M., Hasman, H., 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* 58, 3895–3903.
- Chen, Y., Gonzalez-Escalona, N., Hammack, T.S., Allard, M.W., Strain, E.A., Brown, E.W., 2016. Core genome multilocus sequence typing for the identification of globally distributed clonal groups and differentiation of outbreak strains of *Listeria* monocytogenes. *Appl. Environ. Microbiol.* 82 (20), 6258–6272.
- Chen, Y., Luo, Y., Carleton, H., Timme, R., Melka, D., Muruvanda, T., Wang, C., Kastanis, G., Katz, L.S., Turner, L., Fritzing, A., Moore, T., Stones, R., Blankenship, J., Salter, M., Parish, M., Hammack, T.S., Evans, P.S., Tarr, C.L., Allard, M.W., Strain, E.A., Brown, E.W., 2017a. Whole genome and core genome multilocus sequence typing and single nucleotide polymorphism analyses of *Listeria* monocytogenes associated with an outbreak linked to cheese, United States, 2013. *Appl. Environ. Microbiol.* 83 e00633–17.
- Chen, Y., Luo, Y., Curry, P., Timme, R., Melka, D., Doyle, M., Parish, M., Hammack, T., Allard, M.W., Brown, E.W., Strain, E., 2017b. Assessing the genome level diversity of *Listeria* monocytogenes from contaminated ice cream and environmental samples linked to a listeriosis outbreak in the United States. *PLoS One* 12 (2), e0171389.
- Chen, Y., Luo, Y., Pettengill, J., Timme, R., Melka, D., Doyle, M., Jackson, A., Parish, M., Hammack, T.S., Allard, M.W., Brown, E.W., Strain, E.A., 2017c. Singleton sequence type 382, an emerging clonal group of *Listeria* monocytogenes associated with three multistate outbreaks linked to contaminated stone fruit, caramel apples, and leafy green salad. *J. Clin. Microbiol.* 55, 931–941.
- Clausen, P.T., Zankari, E., Aarestrup, F.M., Lund, O., 2016. Benchmarking of methods for identification of antimicrobial resistance genes in bacterial whole genome data. *J. Antimicrob. Chemother.* 71, 2484–2488.
- Dallman, T.J., Chattaway, M.A., Mook, P., Godbole, G., Crook, P.D., Jenkins, C., 2016a. Use of whole-genome sequencing for the public health surveillance of *Shigella sonnei* in England and Wales. *J. Med. Microbiol.* 65, 882–884.
- Dallman, T., Innis, T., Jombart, T., 2016b. Phylogenetic structure of European *Salmonella* Enteritidis outbreak correlates with national and international egg distribution network. *Microb. Genom.* 2, 8.
- Davis, S., Pettengill, J.B., Luo, Y., Payne, J., Shpuntoff, A., Rand, H., Strain, E., 2015. CFSAN SNP pipeline: an automated method for constructing SNP matrices from next-generation sequence data. *Peer J. Comput. Sci.* 1 e20.
- Franz, E., Mughini Gras, L., Dallman, T., 2016. Significance of whole genome sequencing for surveillance, source attribution and microbial risk assessment of foodborne pathogens. *Curr. Opin. Food Sci.* 8, 74–79.
- Gardy, J., Loman, N., 2018. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat. Rev. Genet.* 19, 9–20.
- Gwinn, M., MacCannell, D.R., Khabbaz, R.F., 2017. Integrating advanced molecular technologies into public health. *J. Clin. Micro.* 55, 703–714.
- Hoffmann, M., Zhao, S., Pettengill, J., Luo, Y., Monday, S.R., Abbott, J., Ayers, S.L., Cinar, H.N., Muruvanda, T., Li, C., Allard, M.W., Whichard, J., Meng, J., Brown, E.W., McDermott, P.F., 2014. Comparative genomic analysis and virulence differences in closely related *Salmonella enterica* serotype Heidelberg isolates from humans, retail meats, and animals. *Genome Biol. Evol.* 6 (5), 1046–1068.
- Hoffmann, M., Luo, Y., Monday, S.R., Gonzalez-Escalona, N., Ottesen, A.R., Muruvanda, T., Wang, C., Kastanis, G., Keys, C., Janies, D., Senturk, I., Catalyurek, U.V., Wang, H., Hammack, T.S., Wolfgang, W.J., Schoonmaker-Bopp, D., Chu, A., Myers, R., Haendiges, J., Evans, P., Meng, J., Strain, E., Allard, M.W., Brown, E.W., 2016. Tracing origins of the *Salmonella* bareilly strain causing a food-borne outbreak in the United States. *J. Infect. Dis.* 213 (4), 502–508.
- Jackson, B., Tarr, C., Strain, E., Jackson, K.A., Conrad, A., Carleton, H., Katz, L.S., Stroika, S., Gould, L.H., Mody, R.K., Silk, B.J., Beal, J., Chen, Y., Timme, R., Doyle, M., Fields, A., Wise, M., Tillman, G., Defibaugh-Chavez, S., Kucerova, Z., Sabol, A., Roache, K., Trees, E., Simmons, M., Wasilenko, J., Kubota, K., Pousee, H., Klimke, W., Besser, J., Brown, E., Allard, M., Gerner-Smidt, P., 2016. Implementation of nationwide real-time whole-genome sequencing to enhance listeriosis outbreak detection and investigation. *Clin. Infect. Dis.* 63, 380–386.
- Jain, S., Mukhopadhyay, K., Thomassin, P.J., 2019. An economic analysis of salmonella detection in fresh produce, poultry, and eggs using whole genome sequencing technology in Canada. *Food Res. Int.* 116, 802–809.
- Knowles, M., Stinson, S., Lambert, D., Carrillo, C., Kazil, A., Gauthier, M., Blais, B., 2016. Genomic tools for customized recovery and detection of foodborne Shiga toxinigenic *Escherichia coli*. *J. Food Protect.* 79, 2066–2077.
- Kwong, J.C., Mercouliu, K., Tomita, T., Easton, M., Li, H.Y., Bulach, D.M., Stinear, T.P., Seemann, T., Howden, B.P., 2016a. Prospective whole-genome sequencing enhances national surveillance of *Listeria monocytogenes*. *J. Clin. Microbiol.* 54, 333–342.
- Kwong, J.C., Stafford, R., Strain, E., Stinear, T.P., Seemann, T., Howden, B.P., 2016b. Sharing is caring: international sharing of data enhances genomic surveillance of *Listeria monocytogenes*. *Clin. Inf. Dis.* 63 (6), 846–848.
- Ladner, J.T., Grubaugh, N.D., Pybus, O.G., Andersen, K.G., 2019. Precision epidemiology for infectious disease control. *Nat. Med.* 25, 206–211.
- Lambert, D., Pightling, A., Griffiths, E., Van Domselaar, G., Evans, P., Berthelet, S., Craig, D., Chandry, P.S., Stones, R., Brinkman, F., Angers-Loustau, A., Kreysa, J., Tong, W., Blais, B., 2017. Baseline practices for the application of genomic data supporting regulatory food safety. *J. AOAC Int.* 100, 721–731.
- Li, Z., Pérez-Osorio, A., Wang, Y., Eckmann, K., Glover, W.A., Allard, M.W., Brown, E.W., Chen, Y., 2017. Whole genome sequencing analyses of *Listeria* monocytogenes that persisted in a milkshake machine for a year and caused illnesses in Washington state. *BMC Microbiol.* 17, 134.
- Lienau, E.K., Strain, E., Wang, C., Zheng, J., Ottesen, A., Keys, C.E., Stones, R., Hammack, T., Musser, S., Brown, E.W., Allard, M.W., 2011. Identification of a salmonellosis outbreak by means of molecular sequencing. *N. Engl. J. Med.* 364 (10), 981–982.
- McDermott, P.F., Tyson, G.H., Kabera, C., Chen, Y., Li, C., Folster, J.P., Ayers, S.L., Lam, C., Tate, H.P., Zhao, S., 2016. Whole-genome sequencing for detecting antimicrobial resistance in nontyphoidal *Salmonella*. *Antimicrob. Agents Chemother.* 60, 5515–5520.
- Mellmann, A., Harmsen, D., Cummings, C.A., Zentz, E.B., Leopold, S.R., Rico, A., Prior, K., Szczepanowski, R., Ji, Y., Zhang, W., McLaughlin, S.F., Henkhaus, J.K., Leopold, B., Bielaszewska, M., Prager, R., Brzoska, P.M., Moore, R.L., Guenther, S., Rothberg, J.M., Karch, H., 2014. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS One* 6 (7), e22751.
- Minor, T., Lasher, A., Klontz, K., Brown, B., Nardinelli, C., Zorn, D., 2015. The per case and total annual costs of foodborne illness in the United States. *Risk Anal.* 35 (6), 1125–1139.
- Moura, A., Tourdjman, M., Leclercq, A., Hamelin, E., Laurent, E., Fredriksen, N., Van Cauteren, D., Bracq-Dieye, H., Thouvenot, P., Vales, G., Tessaud-Rita, N., Maury, M.M., Alexandru, A., Criscuolo, A., Quevillon, E., Donguy, M.P., Enouf, V., de Valk, H., Brisse, S., Lecuit, M., 2017. Real-time whole-genome sequencing for surveillance of *Listeria* monocytogenes, France. *Emerg. Infect. Dis.* 23 (9), 1462–1470.
- Nadon, C., Van Walle, I., Gerner-Smidt, P., Campos, J., Chinen, I., Concepcion-Acevedo, J., Gilpin, B., Smith, A.M., Kam, K.M., Perez, E., Trees, E., Kubota, K., Takkinen, J., Møller Nielsen, E., Carleton, H., 2017. PulseNet international: vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. *Euro. Surveill.* 22 (23), 30544.
- Pightling, A.W., Pettengill, J.B., Luo, Y., Baugher, J.D., Rand, H., Strain, E., 2018. Interpreting whole-genome sequence analyses of foodborne bacteria for regulatory

- applications and outbreak investigations. *Front. Microbiol.* 9, 1482.
- Pouillot, R., Klontz, K.C., Chen, Y., Burall, L.S., Macarasin, D., Doyle, M., Bally, K.M., Strain, E., Datta, A.R., Hammack, T.S., Van Doren, J.M., 2016. Infectious dose of *Listeria monocytogenes* in outbreak linked to ice cream, United States, 2015. *Emerg. Infect. Dis.* 22 (12), 2113–2119.
- Ronholm, J., Nasheri, N., Petronella, N., Pagotto, F., 2016. Navigating microbiological food safety in the era of whole-genome sequencing. *Clin. Microbiol. Rev.* 29, 837–857.
- Sanaa, M., Pouillot, R., Garces-Vega, F.J., Strain, E., Van Doren, J.M., 2018. GenomeGraphR: a user-friendly open-source web application for foodborne pathogen whole genome sequencing data integration, analysis, and visualization. *BioRxiv*. <https://doi.org/10.1101/495309>.
- Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., Lomsadze, A., Pruitt, K.D., Borodovsky, M., Ostell, J., 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44 (14), 6614–6624.
- Timme, R.E., Sanchez Leon, M., Allard, M.W., 2019a. Utilizing the public GenomeTrakr database for foodborne pathogen Traceback. In: Bridier, A. (Ed.), *Foodborne Bacterial Pathogens. Methods in Molecular Biology*. vol. 1918 Humana Press, New York, NY.
- Timme, R.E., Strain, E., Baugher, J., Davis, S., Gonzalez-Escalona, N., Sanchez-Leon, M., Allard, M.W., Brown, E.W., Tallent, S., Rand, H., 2019b. Phylogenomic pipeline validation for foodborne pathogen disease surveillance. *J. Clin. Microbiol.* <https://doi.org/10.1128/JCM.01816-18>.
- Van Walle, I., Torgny Björkman, J., Cormican, M., Dallman, T., Mossong, J., Moura, A., Pietzka, A., Ruppitsch, W., Takkinen, J., 2018. Retrospective validation of whole genome sequencing-enhanced surveillance of listeriosis in Europe, 2010 to 2015. *Euro. Surveill.* 23 (33), 1700798.
- Waldram, A., Dolan, G., Ashton, P.M., Jenkins, C., Dallman, T.J., 2018. Epidemiological analysis of *Listeria* clusters identified by whole genome sequencing, England and Wales 2014. *Food Microbiol.* 71, 39–45.
- Wang, Y., Pettengill, J.B., Pightling, A., Timme, R., Allard, M.W., Strain, E., Rand, H., 2018. Genetic diversity of *Salmonella* and *Listeria* isolates from food facilities. *J. Food Protect.* 81 (12), 2082–2089.
- Yozwiak, N.L., Schaffner, S.F., Sabeti, P.C., 2015. Data sharing: make outbreak research open access. *Nature* 518, 477–479.
- Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., Larsen, M.V., 2012. Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644.
- Zankari, E., Allesoe, R., Joensen, K.G., Cavaco, L.M., Lund, O., Aarestrup, F.M., 2017. PointFinder: a novel web tool for WGSbased detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *J. Antimicrob. Chemother.* 72, 2764–2768.
- Zhang, S., Li, S., Gu, W., den Bakker, H., Boxrud, D., Taylor, A., Roe, C., Driebe, E., Engelthaler, D.M., Allard, M.W., Brown, E.W., McDermott, P., Zhao, S., Bruce, B.B., Trees, E., Fields, P.I., Deng, X., 2019. Zoonotic source attribution of *Salmonella enterica* serotype typhimurium using genomic surveillance data, United States. *Emerg. Infect. Dis.* 25 (1), 82–91.