Journal of Food Protection
International Association for Food Protection

GENERAL INTEREST | MARCH 08 2022

Use of Whole Genome Sequencing by the Federal Interagency Collaboration for Genomics for Food and Feed Safety in the United States

ERIC L. STEVENS ; HEATHER A. CARLETON; JENNIFER BEAL; GLENN E. TILLMAN; REBECCA L. LINDSEY; A. C. LAUER; ARTHUR PIGHTLING; KAREN G. JARVIS; ANDREA OTTESEN; PADMINI RAMACHANDRAN; LESLIE HINTZ; LEE S. KATZ; JASON P. FOLSTER; JEAN M. WHICHARD; EIJA TREES; RUTH E. TIMME; PATRICK McDERMOTT; BEVERLY WOLPERT; MICHAEL BAZACO; SHAOHUA ZHAO; SABINA LINDLEY; BEAU B. BRUCE; PATRICIA M. GRIFFIN; ERIC BROWN; MARC ALLARD; SANDRA TALLENT; KARI IRVIN; MARIA HOFFMANN; MATT WISE; ROBERT TAUXE; ... Show more

J Food Prot (2022) 85 (5): 755–772.

https://doi.org/10.4315/JFP-21-437 Article history

ABSTRACT

This multiagency report developed by the Interagency Collaboration for Genomics for Food and Feed Safety provides an overview of the use of and transition to whole genome sequencing (WGS) technology for detection and characterization of pathogens transmitted commonly by food and for identification of their sources. We describe foodborne pathogen analysis, investigation, and harmonization efforts among the following federal agencies: National Institutes of Health; Department of Health and Human Services, Centers for Disease Control and Prevention (CDC) and U.S. Food and Drug Administration (FDA); and U.S. Department of Agriculture, Food Safety and Inspection Service, Agricultural Research Service, and Animal and Plant Health Inspection Service. We describe single nucleotide polymorphism, core-genome, and whole genome multilocus sequence typing data analysis methods as used in the PulseNet (CDC) and GenomeTrakr (FDA) networks, underscoring the complementary nature of the results for linking genetically related foodborne pathogens during outbreak investigations while allowing flexibility to meet the specific needs of Interagency Collaboration partners. We highlight how we apply WGS to pathogen characterization (virulence and antimicrobial resistance profiles) and source attribution efforts and increase transparency by making the sequences and other data publicly available through the National Center for Biotechnology Information. We also highlight the impact of current trends in the use of culture-independent diagnostic tests for human diagnostic testing on analytical approaches related to food safety and what is next for the use of WGS in the area of food safety.

- We summarize the transition to WGS for monitoring food safety within the United States.
- We provide a critical review of data analysis methods and applications to outbreaks.
- We outline possible future applications of WGS in the area of food safety.

Keywords: Foodborne outbreak, Food safety, Molecular subtyping, U.S. public health agencies, Whole genome sequencing

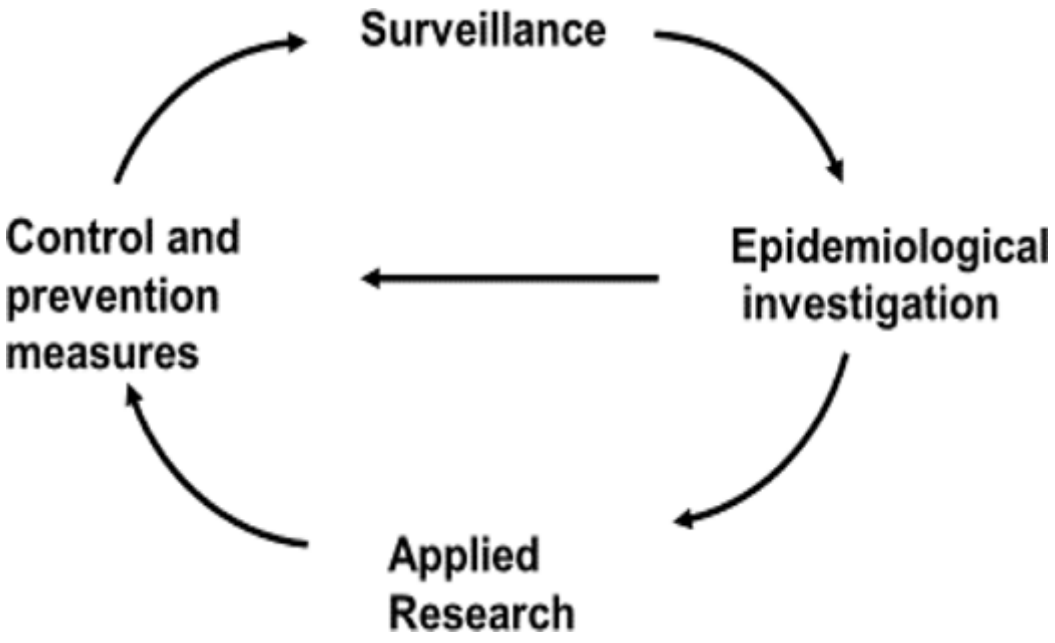
INTRODUCTION TO FOOD SAFETY

Foodborne infections are a common and substantial burden on public health. Each year in the United States, an estimated 48 million persons become ill, 128,000 are hospitalized, and 3,000 die from consuming foods contaminated with microbial pathogens *(68, 69)*. The annual economic burden (e.g., cost of medical care and lost wages) caused by 15 principal foodborne pathogens was estimated as $15.5 billion in 2015 *(35)*. Many thousands of foodborne illnesses are diagnosed each year in clinical laboratories in which *Salmonella,* Shiga toxin–producing *Escherichia coli* (STEC), *Listeria monocytogenes,* and other pathogens are isolated from clinical specimens. Clinical illnesses are then reported to public health departments for epidemiologic follow-up, and the isolated strain is sent to a public health laboratory for further characterization and subtyping. Although most infections that occur each year in the United States are sporadic events (i.e., without a known connection with other infections or outbreaks), public health officials investigate and report an average of ca. 900 foodborne outbreaks (defined as two or more people with the same illness linked to consumption of the same contaminated food or drink) *(9)*. A focus in public health investigations on those illnesses that are most likely to be part of a foodborne outbreak prov investigators more food exposure and laboratory information and enables those investigators to be identify, control, and prevent food safety problems.

Food can become contaminated with various pathogens, and the contamination may occur at any point in the distribution chain from the farm or fishery where it was grown and harvested to the final preparation by the retail establishment or consumer. When a contaminated food is consumed by susceptible individuals, the size and extent of the resulting outbreak is dependent on where in the food production chain contamination occurred, how frequently the initial contamination event occurred, the length of time the contamination goes unresolved, and the subsequent distribution and shelf life of the contaminated food.

which food can become contaminated. Public health activities associated with foodborne outbreak investigations form a cycle that drives continuous improvements in food safety (Fig. 1). When public health surveillance reveals an increase in a specific type of infection, an epidemiologic investigation ensues. A typical investigation may include interviewing ill (and sometimes well) people about exposures to define the association between exposure and illness, determining venues of exposure and tracing the food production chain, collecting food and environmental samples for testing, and inspecting processing facilities and farms. Combining epidemiologic, microbiologic, traceback, and environmental assessment data allows for the identification of the food that was the source of the outbreak. Control measures can then be implemented to stop the outbreak and prevent additional illnesses, and ongoing surveillance can verify that the outbreak has stopped when infections become less frequent. Outbreak investigations can also drive advances in food safety by revealing previously unknown food safety hazards and identifying areas of research that are important for developing effective long-term prevention and inspection strategies.

FIGURE 1



View large Download slide

The cycle of public health control and prevention measures.

Multiple challenges are associated with outbreak detection and the subsequent investigation and response. The primary challenge is that the food safety landscape is constantly changing as food

Another challenge is detection of the signal of related cases amidst the background noise of sporadic cases. Food vehicles and sources of food contamination responsible for sporadic cases, which account for the clear majority of estimated foodborne illnesses, are rarely identified. Estimates of the number or proportion of foodborne illnesses attributed to specific food commodities are often based on outbreak-related cases for which the food vehicle was identified, but how well these outbreak-related cases represent the illness from sporadic cases is unknown *(22)*. The growing use of culture-independent diagnostic tests in clinical laboratories means that additional culture of positive specimens is now needed to isolate the pathogen for molecular surveillance *(71),* which places an additional burden on the clinical or public health laboratory to culture the isolate, and in some cases the organism is not recovered. Some infections are not easily or routinely diagnosed, and the etiologies of some illnesses remain undiscovered.

To overcome these food safety challenges, the public health community strives for continual advances in methods and technologies to characterize etiologic agents and exposures to them. This information must then be integrated into surveillance and investigation systems. Current methods depend on first using traditional microbiological methods to isolate a bacterial pathogen then using molecular subtyping to characterize the pathogen and allow for grouping of illnesses likely caused by a common source. Whole genome sequencing (WGS) offers higher resolution for characterizing outbreak-related foodborne pathogens than do previous subtyping methods, replacing pulsed-field gel electrophoresis (PFGE), which has been the “gold standard” for the past 20 years *(8)*. WGS and its practical and potential applications are discussed in greater detail in the following sections.

INTRODUCTION TO AND OVERVIEW OF WGS

WGS explained

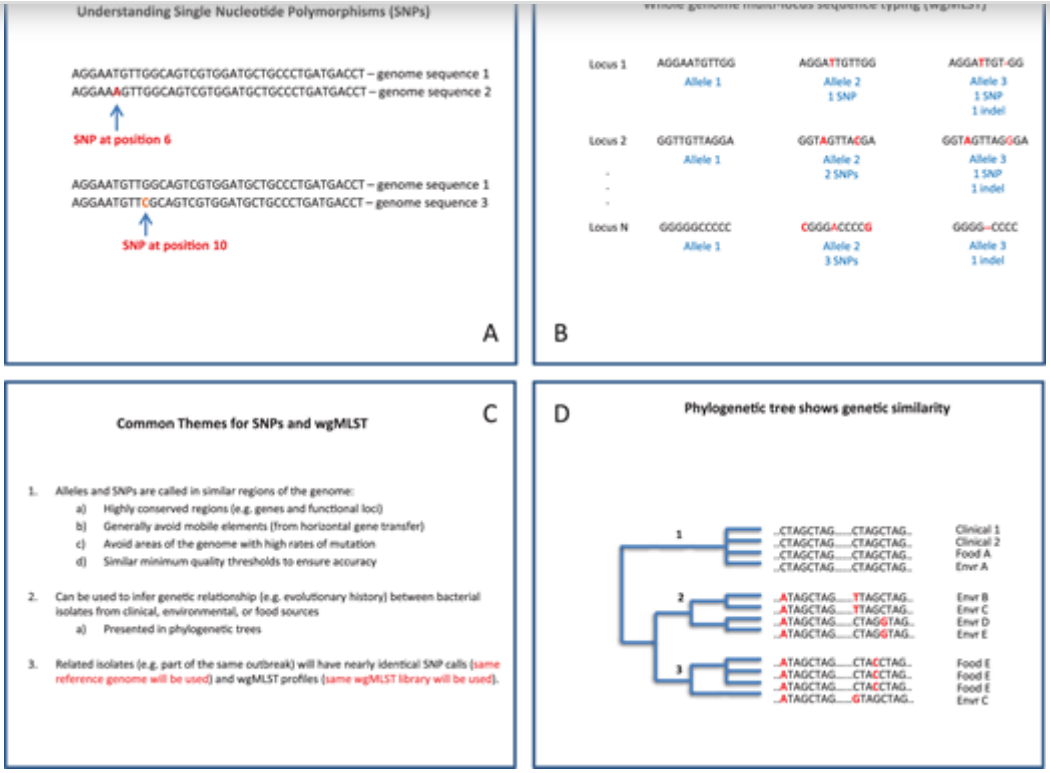
DNA sequencing methods determine the number and order of the four nucleotides (A, G, C, and T) that make up a given strand of DNA. The purpose and function of WGS is to reveal the complete genomic sequence of an organism. WGS is a specific application of next-generation sequencing technologies, which have rapidly decreased in cost and time, thereby increasing their practicality for use in public health settings. For example, WGS of foodborne pathogens results in the identification of ca. 3 to 5 million nucleotides (or base pairs) that make up the genome of a bacterial isolate. Current WGS technologies cannot sequence the complete genome of a bacterial isolate in a single read (i.e., a single unbroken string in which the uninterrupted order of the 3 to 5 million sequenced nucleotides is known). Therefore, to generate a genomic sequence of a bacterial isolate that is close to complete, thousands to millions of

identical stretches within the DNA sequence, and computer algorithms are used to reconstitute the complete genomic sequence of a single bacterial genome *(67),* leaving as few gaps in the sequence as possible. Once the genomic sequence has been reassembled, additional bioinformatics pipelines (e.g., a workflow protocol in which the sequence data are analyzed by the sequential use of validated computer programs) can be used to compare the number and order of nucleotides between genomes of foodborne pathogens of the same species (e.g., isolates derived from clinical, food, and environmental samples). Both single nucleotide polymorphism (SNP) and multilocus sequence type (MLST) subtyping methods are used to find those genomes that are genetically related and may have a common source within the context of a foodborne outbreak.

WGS is rapidly changing the practice of microbiology and public health surveillance and investigation of foodborne illnesses. The use of these techniques to sequence and analyze the whole genome of foodborne pathogens provides a greater level of detail about the bacteria that make people sick than have previous characterization methods. WGS also may provide a one-step characterization of bacteria by identifying the species, serotype, genotype, and resistance and virulence genes all within a single laboratory workflow. Given the higher resolution of WGS and its ability to replace more time-consuming and costly traditional microbiology methods, health and regulatory agencies in the United States have heavily invested in transitioning over the last several years from PFGE to WGS for routine characterization of foodborne pathogens. The result of these efforts has been a substantial growth in the capacity of federal, state, and local health and regulatory agencies to use WGS to identify and address food safety problems.

SNP approach

One method used for molecular subtyping and genomic characterization is a SNP-based approach. A SNP is a polymorphism relative to a reference genome and results when a mutation at one position changes a single nucleotide (e.g., at genomic position 6, bacterial genome 1 has nucleotide T and bacterial genome 2 has nucleotide A). Figure 2A illustrates this concept and further demonstrates how SNPs can be called in different positions depending on the genomes that are being compared (e.g., at position 10, G in bacterial genome 1 and C in bacterial genome 3). In general, the fewer SNPs that two bacterial isolates have between them, the more genetically related they are because their evolutionary relationship is closer. SNP-based approaches are often used when the highest level of specificity in linking two or more isolates is needed. The U.S. Food and Drug Administration (FDA), because of its regulatory authority, uses a SNP-based approach within its GenomeTrakr network for connecting clinical isolates to a food or environmental isolate based on experience and a robust body of research and validation *(2, 55, 78)*.



View large Download slide

Overview of SNP and wgMLST analysis. (A) Examples of SNPs at positions 6 and 10 for three genome sequences. (B) Examples of three alleles with different DNA sequences across multiple loci. (C) Common themes for SNP and wgMLST analysis. (D) Typical phylogenetic tree constructed using SNP and/or wgMLST approaches. Cluster 1 isolates (clinical 1, clinical 2, food A, and envr A) are grouped together because they have identical nucleotide sequences, whereas isolates in clusters 2 and 3 have nucleotide differences (identified in red) that make them more distantly related to the isolates in cluster 1.

Gene-by-gene typing approaches: cgMLST and wgMLST

MLST analytic approaches compare gene variants between isolates to measure relatedness. These genes are compared against a reference scheme of known gene variants. Schemes can be built on the core genome (cg), which contain those genes that are found in 95% of genomes investigated, or on the whole genome (wg), which includes the core genome as well as the pan genome (i.e., genes that may be present in some strains but not in others). The creation of an MLST scheme begins by generating a genus- or species-specific library of known genes (loci) using well-annotated genomes then identifying variants for those loci (categorized as alleles) based on similarity and other thresholds. After a scheme has been constructed, the genomes of new isolates can be quickly characterized based on their alleles compared with those of the scheme. The cgMLST approach is well suited for quickly clustering clinical illnesses that

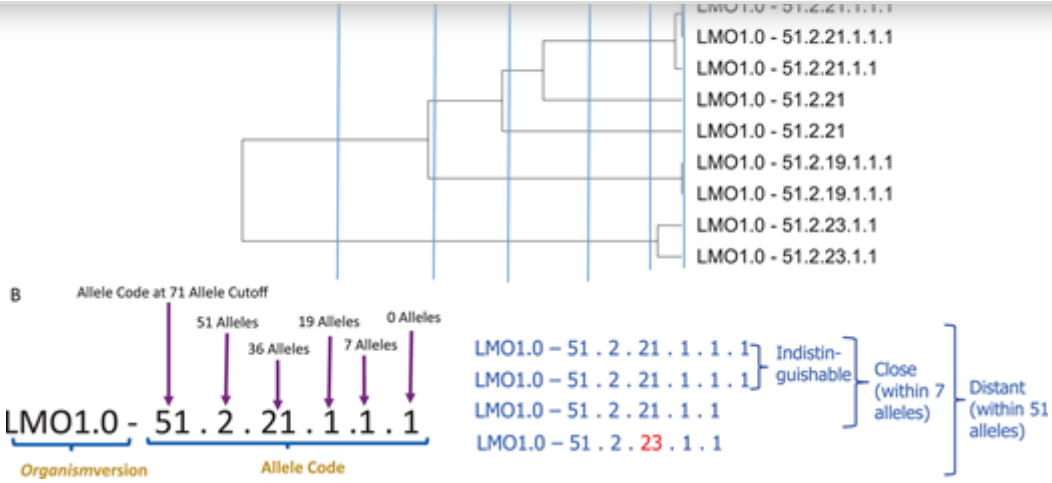
separate genomic loci. Allele 1 for locus 1, locus 2, and locus N could be viewed as the reference allele (i.e., the prototype allele for that locus). Allele 2 for locus 1 has a SNP, and allele 3 for locus 1 has both a SNP and a deletion of a single nucleotide compared with allele 1. Similar representations of alleles are shown for locus 2 and locus N.

For example, a single *L. monocytogenes* genome contains ca. 3,000 unique genes. When >200 complete genome sequences of *L. monocytogenes* were used to construct a reference allele library, ca. 5,000 unique genes were analyzed and catalogued to define the wgMLST scheme, and 1,748 genes were identified for the cgMLST scheme *(40, 56)*. The variants for all the informative loci were then assigned an allele number. The string of allele numbers was then compared between strains, and the differences in allele numbers at the same locus between strains was used to measure similarity (relatedness) between isolates.

Communicating genome relatedness based on MLST with allele codes

PulseNet uses allele codes *(80),* a strain nomenclature built from single linkage trees based on cgMLSTs, to communicate about genome relatedness and detect outbreak clusters. Allele codes (Fig. 3A) are assigned based on allele difference thresholds that overlay a single linkage tree and are hierarchical, so relatedness of isolates can be inferred from shared allele code numbers. The *Listeria* allele code (LMO1.0) contains six numbers, with the first representing ca. 71 allele differences and each remaining number representing a lower threshold difference until the sixth number, which represents 0 allele differences (Fig. 3B). Isolates that do not have a match at any of the lower allele difference thresholds can receive a partial name (isolate 5; Fig. 3A). Isolates with ≤71 allele differences between genomes based on cgMLST (isolates 1 through 8; Fig. 3A) share the same first number. Isolates with the same six-number allele code are considered indistinguishable (isolates 1 through 3 and isolates 7 and 8; Fig. 3A). Isolates 7 and 8 (Fig. 3A) share first two numbers with isolates 1, 2, and 3, indicating that they are distantly related, differing by approximately 51 alleles. Allele codes allow communication about isolates that may be part of a potential outbreak cluster on a tree, similar to how zip codes are used to group street addresses in the same geographic region. All the genomes that share the same allele code to the fourth, fifth, or sixth number are considered closely related and may be investigated epidemiologically to determine whether relatedness between cases, i.e., exposure to the same suspected outbreak source. Allele codes also allow communication of phylogenetic relatedness in text form and can be easily shared between public health scientists.

FIGURE 3



View large Download slide

Overview of PulseNet allele codes. (A) Single linkage tree is used to form an allele code; blue lines indicate approximate cutoffs for different numbers in the allele code at 71, 51, 36, 19, 7, and 0 allele differences, as indicated by top scale bar. Singletons on a tree receive partial allele codes. (B) PulseNet allele code interpretation at different cutoffs. For example, 51 is the allele code at the 71-allele cutoff; it does not mean 51 of 71 allele codes.

Common themes for SNP and cg-wgMLST approaches

Both cg-wgMLST and SNP approaches analyze similar regions of the genome, including genes and functional loci that are highly conserved. Additionally, cg-wgMLST and SNP analyses can be designed to avoid genome regions that may confound phylogenetic analyses, such as mobile elements or areas with high rates of mutation. Both methods maintain identical minimum quality thresholds of the sequence data to ensure accuracy. A general overview of the similarities of SNP and cg-wgMLST approaches is given in Figure 2C. Both approaches have the goal of gathering variant sites, which results in the constructio nearly identical phylogenetic trees *(33)* that are visual representations of the inferred evolutionary relationships among the individual clinical, food, or environmental isolates (Fig. 2D). In other words, both cg-wgMLST and SNP approaches can be used to extract data for successful reconstruction of the evolutionary history of a set of bacterial isolates *(26)* and can allow U.S. public health agencies to better detect and respond to the causes of foodborne illnesses.

Complementary analysis using SNPs and wgMLSTs

U.S. public health agencies are frequently asked about the comparability of the results of SNP and cg-wgMLST approaches. These approaches are complementary but have two major differences: (i) the SNP approach assess variations both within (coding) and between genes (noncoding), whereas MLST

MLST approach all of the genes contained within the reference scheme will be used to call alleles within the test genomes. For that reason, SNP and MLST analyses will not always produce identical results. When two identical strains are compared, both SNP and wgMLST approaches will have zero SNP and wgMLST allele differences, respectively. However, for strains that are closely related genetically and thus suspected to be part of the same outbreak event, the exact number of SNP or wgMLST allele differences may vary, but the total number of differences will be quite similar (e.g., three SNP and two wgMLST allele differences). In either case, the interpretation of the analysis will agree that the isolates are closely related in terms of separation by time or number of bacterial cell divisions and thus are likely to be part of the same outbreak event *(65)*. Determination of a fixed value that defines the number of allele or SNP differences that define an isolate as being related to an outbreak is not possible; instead, a generalized rule of thumb of ≤20 SNP or allele differences paired with exposure information and WGS data from a food source suspected to be linked to the illnesses can be used as an initial screen to aid in determining which isolates are associated with an outbreak. This rule of thumb can vary by organism and serotype and by the source and length of the outbreak. When the strains being compared are not closely related (and likely not part of the outbreak), the total number of SNP and wgMLST allele differences might not be close, but interpreation will agree that the strains are unrelated. In summary, the two approaches are compatible when used to consider whether isolates are part of the same foodborne outbreak event and thus very closely genetically related *(63)*.

TRANSITION OF LABORATORY-BASED FOODBORNE DISEASE SURVEILLANCE TO WGS IN THE UNITED STATES

WGS data collection and analysis—validation and harmonization across U.S. public health agencies

To help improve the transition from previous subtyping methodologies (e.g., PFGE) to WGS, U.S. pu health agencies have formed various collaborations, such as the Interagency Collaboration for Genomics for Food and Feed Safety (Gen-FS) *(83)*. Gen-FS is composed of representatives from the Centers for Disease Control and Prevention (CDC); FDA; U.S. Department of Agriculture (USDA), Food Safety and Inspection Service (FSIS), Agricultural Research Service, and Animal and Plant Health Inspection Service; and the National Institutes of Health (NIH). Their working groups have focused in part on harmonizing sequencing protocols across PulseNet and GenomeTrakr, including quality assurance measures and accompanying quality control checks, to ensure all WGS data generated by Gen-FS members in the Pathogen Detection database at the National Center for Biotechnology Information (NCBI) meet the minimum quality standards set by Gen-FS. All downstream analyses, including cluster analysis presented

events from GenomeTrakr, are harmonized such that the results are accurate and comparable across the analytical pipelines. Benchmark data sets derived from empirical data *(78)* and simulated data *(55)* are used in this harmonization effort. Gen-FS also runs an annual multilaboratory proficiency test across the PulseNet and GenomeTrakr laboratory networks *(76, 77)*. This test measures proficiency for each lab and serves as a multilaboratory validation exercise by assessing the accuracy and reproducibility of WGS data collection across the entire network. All WGS data are then deposited and made available at the NCBI Pathogen Detection Web site.

Transition to WGS in PulseNet, the national molecular subtyping network for foodborne disease surveillance

The formation of PulseNet, a national network of local, state, and federal public health laboratories, and implementation of real-time PFGE in 1996 revolutionized foodborne disease surveillance in the United States, leading to a dramatic increase in the number of multistate outbreaks being recognized, investigated, and resolved *(75)*. Over the last two decades, PFGE has been the gold standard subtyping method used by U.S. public health and food safety agencies to characterize bacteria most often associated with outbreaks, particularly *Salmonella,* STEC O157, and *L. monocytogenes*. PulseNet used internationally standardized PFGE protocols, a certification and proficiency testing program, and a standardized data analysis platform based on commercial off-the-shelf software. PulseNet member state and local public health and agriculture laboratories and federal laboratories analyzed their pulsed-field gels and identified isolates from cases that were part of a local outbreak. PulseNet laboratories also uploaded gel images and metadata about isolate sources to a national database where multistate outbreaks could be identified. The PulseNet database submission is limited to federal, state, and local public health laboratories that pass a certification process, participate in proficiency testing, and sign a memorandum of understanding. PulseNet created a communication workflow between epidemiologists and laboratory scientists at local, state, and federal levels to alert each other about outbreaks, investigations, and clusters of foodborne isolates with indistinguishable patterns. This initial communication approach has evolved into new and sophisticated outbreak communication workflows managed by multiple public health partners.

PFGE-based subtyping allowed investigators to better detect and define outbreaks than was possible with previous methods (e.g., serotyping and plasmid profiling) because PFGE is a more sensitive and specific method for grouping clinical isolates that may be related and part of an outbreak. As a result, PFGE allowed connections to be made between illnesses that were separated in both time of onset (e.g., days, weeks, or months) and geographic location (e.g., counties, states, or countries). FDA and FSIS laboratories

public health agencies to identify and resolve many more widely dispersed foodborne outbreaks and implement preventive control measures based on the investigation findings. The more rapid control of foodborne outbreaks has the direct impact of preventing additional illnesses caused by foodborne pathogens. The use of PFGE in the national PulseNet network was estimated to prevent annually at least 270,000 illnesses caused by *Salmonella,* STEC O157, and *L. monocytogenes* and saved more than $507 million in medical costs and lost productivity *(70)*.

WGS provides even greater discriminatory power and precision than PFGE, comparing isolates based on millions of base pairs rather than 15 to 30 gel bands, to help determine which ill people to include in an outbreak investigation and to assess the similarity of bacteria isolated from ill people, foods, food production environments, animals, and farms. In a 3-year pilot study in which U.S. public health agencies used real-time WGS for subtyping all *L. monocytogenes* isolates, the number of outbreaks resolved each year increased from two in the PFGE era to six, and the median size of clusters investigated decreased from six cases to four cases *(40)*. Thus, use of WGS has made it possible to find, resolve, and stop more outbreaks while they were smaller than was possible with PFGE data alone.

In 2019, PulseNet was already using WGS for national surveillance and cluster identification of *L. monocytogenes* and transitioned to WGS for national real-time surveillance of *Salmonella, Escherichia* (STEC and *Shigella*), and *Campylobacter*. Before this transition, in 2013 PulseNet began using WGS for outbreak-associated *Salmonella* isolates, representative outbreak and sporadic STEC and *Shigella* isolates, and outbreak and sporadic *Campylobacter* isolates. PulseNet uses cgMLSTs and allele codes as its primary method for cluster detection and has adopted internationally standardized cgMLST schemes for foodborne surveillance and outbreak detection *(56, 64)*. PulseNet through its work with PulseNet International network of >80 countries, has also worked to standardize analysis approaches internationally for foodborne bacteria sequence analysis *(57)*.

To be certified and participate in the PulseNet WGS network, each PulseNet member laboratory first must demonstrate that it can produce sequence data that meet quality thresholds for submission to the national database. Each member also participates in annual proficiency testing. PulseNet laboratories currently submit raw sequence reads and limited metadata to NCBI databases in near real time and upload extended isolate metadata and analyzed cg-wgMLST results to the PulseNet national database. Only limited metadata (organism and country of origin) are uploaded initially to NCBI to protect patient confidentiality. Originally, some PulseNet members released additional limited metadata after 6 months, including clinical source (e.g., blood, stool, or urine), serotype, isolation date, patient age category, and geographic region.

the FDA and FSIS in combination with metadata submitted to the PulseNet national database to identify clusters that may represent outbreaks. These clusters may be further analyzed using the CDC SNP pipeline Lyve-SET *(43),* the FDA Center for Food Safety and Applied Nutrition (CFSAN) SNP pipeline *(20),* or the NCBI Pathogen Detection pipeline when the cgMLST results need to be confirmed by another method or to provide additional information. These SNP and cg-wgMLST analyses are shared with PulseNet members and public health agencies and discussed weekly with CDC epidemiologists and the FDA and FSIS in addition to other pertinent information (e.g., traceback data and food samples). In 2018, PulseNet implemented cgMLST allele code nomenclature for naming isolates for *L. monocytogenes*. Similar nomenclature schemes were implemented for *E. coli* and *Salmonella* in 2021. PulseNet also routinely extracts information about microbial identification, serotype, virulence genes, antimicrobial resistance (AMR) determinants, and plasmid profiles from the sequence data of all isolates subtyped by network participants. This information is used to prioritize and investigate clusters that may represent outbreaks but also has wider food safety applications beyond outbreaks.

FDA GenomeTrakr inception and the increasing importance of food and environmental isolates

With the decreased cost of WGS technology, the FDA began a pilot project called GenomeTrakr in 2012 to build a public genomic reference database of historical food and environmental *Salmonella* isolates. The goal of this project was to improve the accuracy and response time for identifying the contamination source in foodborne outbreaks, identifying harborage sites in facilities, and establishing preventive controls *(3)*. During the pilot project, WGS data were collected by a distributed set of public health laboratories, transferred to the FDA for quality screening, and then uploaded to a dedicated BioProject at the NCBI Sequence Read Archive (SRA) database *(47)*. Over 1,000 *Salmonella* genomes were collect first year, ca. 10,000 the second year, and currently, along with many new contributors (including P other U.S. agencies, and international partners), the maturing public *Salmonella* genome database is >385,000 genomes. After the initial success of building a public WGS database for *Salmonella,* similar efforts were launched for *L. monocytogenes (40)* in 2013 and soon thereafter for pathogenic *E. coli* and *Shigella* species, *Campylobacter jejuni, Vibrio parahaemolyticus,* and *Cronobacter* species. The Pathogen Detection portal at NCBI is now the central public repository for foodborne pathogen genomes used for real-time surveillance in the United States.

Some foodborne pathogen isolates, such as *Salmonella,* collected by the FDA and FSIS field laboratories were the primary source of the food and environmental isolates in the PulseNet PFGE database. These

The number of food and environmental isolate sequences contributed by state public laboratories varied widely depending on sampling effort, level of collaboration with respective state agriculture laboratories, and funding.

The shift to storing WGS data in an open, public database and the easily portable nature of sequence reads versus a PFGE image created an opportunity to significantly expand the number and diversity of isolates by recruiting collaborators beyond the PulseNet community. FDA scientists recognized this advantage and worked to leverage the GenomeTrakr network to include state agriculture laboratories, academic laboratories, private laboratories, and international collaborators with the overall goal of more thoroughly capturing the global population diversity of food and environmental isolates across the One Health arena for key foodborne pathogen species. The probability of a food-environment "match" for any new clinical isolate added to the database increases as more food and environmental isolates are added to the database, supporting the mission of pinpointing the likely causes of foodborne outbreaks, identifying harborage sites in facilities, and using WGS data to establish preventive controls.

These developments have increased the utility of food and environmental isolates in outbreak investigations. For the past 20 years, outbreak investigations typically have begun by identifying groups of individuals (identified as a cluster through PFGE of isolates) who may have been exposed to a common source (e.g., contaminated food), and then epidemiologic, laboratory, and traceback investigations were used to identify and confirm a vehicle as described previously. Evaluation of food and environmental isolates matching the PFGE pattern in question sometimes provided clues about which vehicles to include in the investigation; but frequently the relevance of matching food and environmental isolates was unclear, especially for isolates with common PFGE patterns. This investigation model also required data fro enough individuals to provide a statistically relevant signal, i.e., a given PFGE pattern "flagging" ab expected baseline for that pattern. In contrast, with a technology such as WGS that has higher specificity and sensitivity, suspected outbreaks can be detected based on clusters of fewer cases with greater confidence that they are related, which means that investigations and subsequent corrective actions can be initiated sooner. Identification of food and environmental isolates that are highly related to clinical isolates yields more information about a possible food vehicle than does a PFGE comparison. This potentially actionable information can provide clues as to what type of questions investigators should ask ill persons with regard to their exposure history or can provide additional evidence when epidemiology reveals a general signal but but this signal does not coalesce around a specific brand or type of food (e.g., when patients simply report eating leafy greens and a highly related isolate is recovered from bagged salad

Currently, raw sequence data plus a standard suite of metadata are collected in laboratories across the GenomeTrakr network, assessed for data quality *(29),* and submitted directly to public repositories at NCBI in real time *(79)*. The data are then processed through the NCBI Pathogen Detection analysis pipeline, and within 1 to 2 days the sequences appear in the NCBI Pathogen Detection browser where results of cluster analyses are available for searching and browsing. Detailed reproducible protocols for data collection through submission are publicly available at the GenomeTrakr workspace (https://www.protocols.io/workspaces/genometrakr1/publications). On average, the FDA GenomeTrakr network submits >1,000 isolates per month to the NCBI Pathogen Detection database and pipeline.

The FDA evaluates the Pathogen Detection site daily for mission-relevant clustering results, such as a close match between a food isolate and an isolate collected from a patient or a match between an environmental swab isolate and an isolate collected from the same location in a previous year. When such results are found, scientists download the sequence data associated with a particular cluster of interest then rerun the SNP analysis with the CFSAN open source SNP pipeline *(20)*. Depending on the nature of the cluster, appropriate stakeholders are contacted for follow-up. For example, a cluster of clonal isolates collected from the same facility over multiple years might be sent to the CFSAN Office of Compliance, where the data will become part of routine compliance evaluations. New food and clinical matches might be forwarded to the CFSAN Office of Coordinated Outbreak and Response for evaluation of a possible outbreak in collaboration with the CDC. A particular state laboratory may also be contacted when the cluster appears to be contained within the state's boundaries. In either routine compliance activities or an outbreak investigation, a regulatory response by the FDA will include additional types of evidence, such as epidemiologic, traceback, and inspection data, and supporting data from WGS cluster analysis *(1, 6*

Transition to WGS by the FSIS

The FSIS has been a part of the CDC PulseNet surveillance program since inception of the program in the 1990s. Throughout the partnership, the FSIS has utilized PFGE as the primary subtyping tool in outbreak investigations and for further characterization of pathogens. As part of their 2011 to 2016 strategic plan *(82),* the FSIS began using WGS in 2014 to characterize *L. monocytogenes* isolates from routine sampling programs and to characterize other pathogen isolates related to potential outbreak investigations *(18, 40)*. In 2015, the FSIS began sequencing bacterial isolates collected by the National Antimicrobial Resistance Monitoring System (NARMS) *(53)* and STEC isolates from routine sampling programs. By 2017, the FSIS had

The FSIS routinely uploads to the NCBI Pathogen Detection database the raw sequence files and minimal metadata, which include collection year, serovar for *Salmonella* isolates, serogroup for STEC isolates, species for *Campylobacter* isolates, isolation source (e.g., environment or food), and geographic location. The FSIS actively shares NCBI accession numbers and additional metadata with public health partners during outbreak investigations for inclusion in WGS analyses and routinely performs and/or shares SNP analyses obtained from the NCBI Pathogen Detection browser to aid in internal investigations for outbreak analyses or environmental assessment at federal establishments. The FSIS utilizes pairwise character differences (wgMLST alleles and/or SNPs) to communicate potential relatedness and allele code nomenclature currently available for *L. monocytogenes* through PulseNet. The FSIS process is similar to that of the FDA for identifying potential links between food and/or environmental isolate sequences within an establishment to understand conditions of harborage or cross-contamination and food, environmental, and clinical isolate sequences for potential links to illnesses. For *L. monocytogenes,* FSIS communicates with the FDA and other partners about any relationships between isolates recovered from distinct sampling events, and character differences and the CDC PulseNet nomenclature scheme are used to assess harborage or cross-contamination events within an establishment.

Like the regulatory approach used by the FDA, the FSIS approch includes all the evidence gathered, such as epidemiologic and traceback data, for WGS cluster analysis and response. For example, when epidemiology is not able to identify a specific brand or type of product, the FSIS has been successful in using WGS data from nonclinical isolates that are highly related to clinical isolates (e.g., pork product isolates that are highly related to environmental isolates from a production facility) to guide outbreak investigations. The actual process and the use of subtyping information to inform regulatory action change with the transition from PFGE to WGS. Directive 8080.3 *(84)* describes the FSIS approach to foodborne illness investigations in collaboration with the CDC and state health departments.

Coordination between PulseNet and GenomeTrakr networks

Both GenomeTrakr and PulseNet laboratories can detect foodborne outbreaks with more precision using WGS and have been harmonizing approaches through a Gen-FS working group. PulseNet and GenomeTrakr laboratories in various states often overlap, although PulseNet primarily focuses on laboratories that receive clinical specimens and GenomeTrakr focuses on laboratories that receive food and environmental isolates. Both networks submit sequence read sets and limited metadata to NCBI, and their isolates are included on the NCBI Pathogen Detection page. PulseNet has a national database to

laboratories that are PulseNet certified have access to the PulseNet database and can use the NCBI Pathogen Detection pipeline to identify clusters of illnesses that contain food or environmental isolates and further characterize these clusters using the well-validated CFSAN SNP pipeline, which produces results comparable to those of the CDC Lyve-SET SNP pipeline and the NCBI SNP pipeline. These laboratory networks are separate and complementary, and their collaboration has allowed U.S. investigators to detect foodborne outbreaks and identify sources with more precision.

NCBI Pathogen Detection

Among its many tasks, NCBI (at the National Library of Medicine, NIH) creates and maintains sequence databases, including GenBank and the SRA *(6, 45)*. The sequences from the GenomeTrakr and PulseNet surveillance networks are predominantly submitted to the SRA, and a small number of assembled genomes are submitted to GenBank. Sample metadata describing the pathogens are submitted to BioSample using the pathogen packages (metadata templates) for clinical or environmental samples and include such data as the type of organism, geographical location, and time of sampling *(58)*. Sequence data and sample metadata are organized under BioProjects (Table 1), which are then flagged internally to monitor new data uploads for processing in the Pathogen Detection pipeline. Thus, NCBI serves as the public repository of the sequence and metadata for these surveillance networks and is part of the International Nucleotide Sequence Database Collaboration. Sequences that are submitted to NCBI and partner databases in Europe (European Nucleotide Archive, European Bioinformatics Institute, Hinxton, UK) and Japan (DNA Databank of Japan, National Institute of Genetics, Shizuoka, Japan) are exchanged on a daily basis *(15)*. As of 7 December 2021, the total number of genomes available from isolates of bacteria associated with foodborne, zoonotic, and other sources available was 741,710: 409,749 *Salmonella enterica,* 204,517 *E. coli* and *Shigella,* 70,604 *Campylobacter jejuni,* 48,475 *L. monocytogenes,* 4,922 *parahaemolyticus,* 807 *Cronobacter,* 804 *Vibrio vulnificus,* 1,106 *Clostridium botulinum,* and 726 *Clostridium perfringens*.

TABLE 1

BioProject numbers for PulseNet and GenomeTrakr sequences and metadata at NCBI

PRJNA753476	Umbrella
PRJNA230403	*Salmonella* spp.
PRJNA212117	*Listeria monocytogenes*
PRJNA266293	*Vibrio* spp.
PRJNA218110	*E. coli* and *Shigella* spp.
PRJNA239251	*Campylobacter* spp.

View large

Discussions pertaining to the transition from PFGE to WGS with FDA, CDC, and FSIS scientists coincided with efforts at NCBI to provide enhanced analysis of archival data. Based on these efforts, NCBI decided to develop the Pathogen Detection page based on feedback from public health partners and other NCBI contributors. The Pathogen Detection project provides two enhancements, both of which are integrated into an easy-to-use Web interface: (i) rapid analysis of the genetic relationships among pathogen isolates and (ii) analysis of AMR genetic determinants. The rapid clustering system solves several problems due to real-time surveillance by sequencing. For any new isolate, the nearest sequence neighbors need to be identified in an ever-growing database, and efficient analysis is needed to support real-time surveillance. An analysis that takes days to complete is not useful and limits the effectiveness of the system. The transition to WGS for both surveillance networks was been completed in 2019, and the number of submissions is expected to reach 90,000 per year in the United States.

To meet the requirements for foodborne pathogen surveillance for *Campylobacter* spp., *E. coli* and spp., *L. monocytogenes,* and *S. enterica* using short-read sequencing, NCBI developed a de novo assembler (SKESA) *(74)* and a set of wgMLST schemes. SKESA produces high-confidence calls (or certainty that a nucleotide at a certain position in a sequenced genome is correct) in the assemblies, which is critically important for wgMLST allele calling and SNP calling of clonal isolates for outbreak and traceback purposes. Rapid reports are produced for SKESA assemblies once the wgMLST allele calls are made, typically within 1 h, for *E. coli,Shigella* spp., *Listeria,* and *Salmonella* isolates sequenced in the surveillance network. Full clustering using SNPs, which occurs at least once per day when new data are submitted and sometimes more frequently, is based on all the SKESA and GenBank assemblies, including those submitted as part of research programs (i.e., not collected as part of government surveillance systems). For both processes, assemblies are compared with wgMLST schemes developed at NCBI to determine the number

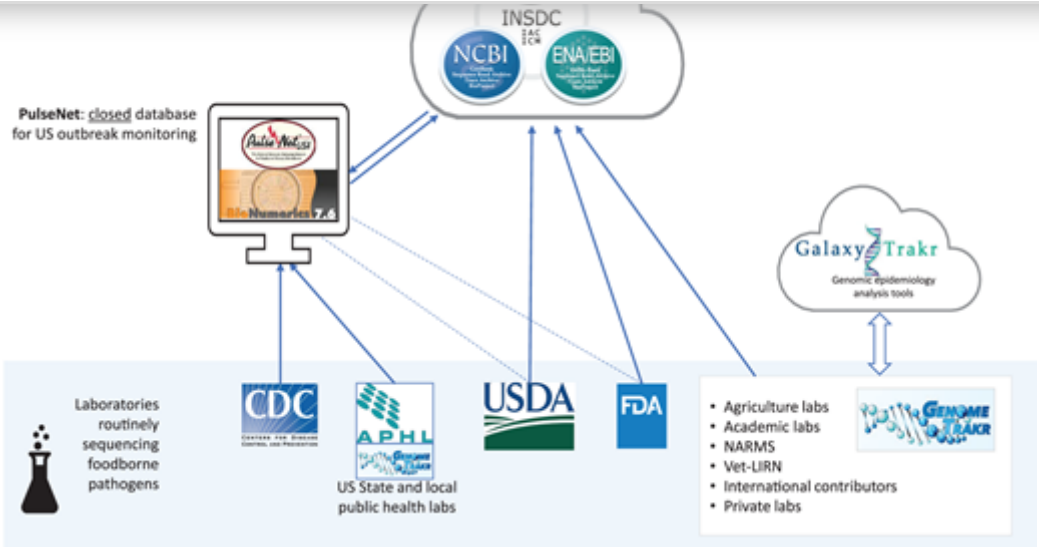
The result is that for each organism, hundreds to thousands of clusters are generated from this process depending on the number of genomes available, from two to several thousand isolates per cluster and some singleton genomes not closely related to anything else in the database. Within each cluster, the most contiguous complete genome is chosen as a reference, and all other genomes in the cluster are compared with this master reference to generate a set of SNPs. Regions of dense SNPs, likely due to recombination, are filtered out. Phylogenetic trees of each SNP matrix are produced using the maximum compatibility algorithm *(14)*. The full list of isolates and the SNP trees are available on the NCBI Pathogen Detection Web site each day that new data are submitted.

The Pathogen Detection Web site consists of an isolates browser (https://www.ncbi.nlm.nih.gov/pathogens/isolates/) and, for those isolates that end up in a cluster, a SNP Tree Viewer. For each assembled pathogen, genome metadata are provided based on information submitted into the archival databases, including the type of organism, geographical location, isolation source, and analysis results. This information also includes links to the SNP Tree Viewer when the isolate is a member of a cluster and two SNP distances when appropriate: the minimum SNP distance to an isolate of the opposite type (clinical versus food or environmental) and of the same type (clinical versus clinical, and food or environmental versus food or environmental). The AMR genetic determinants also are listed for any isolate that has them. The SNP Tree Viewer provides the phylogenetic tree of each cluster, the SNP distances of any isolates in that SNP cluster, and the metadata. These two interfaces provide a streamlined system for browsing and searching the SNP clusters, which allows collaborators at state and federal public health laboratories to quickly determine isolate relatedness for outbreak investigations and other activities.

Figure 4 provides a generalized model for how FDA, CDC, USDA, and NCBI sharing of WGS data fits t within the U.S. food safety system. Figure 5 provides a timeline of important developments in the transition to WGS for foodborne disease surveillance and response.

FIGURE 4

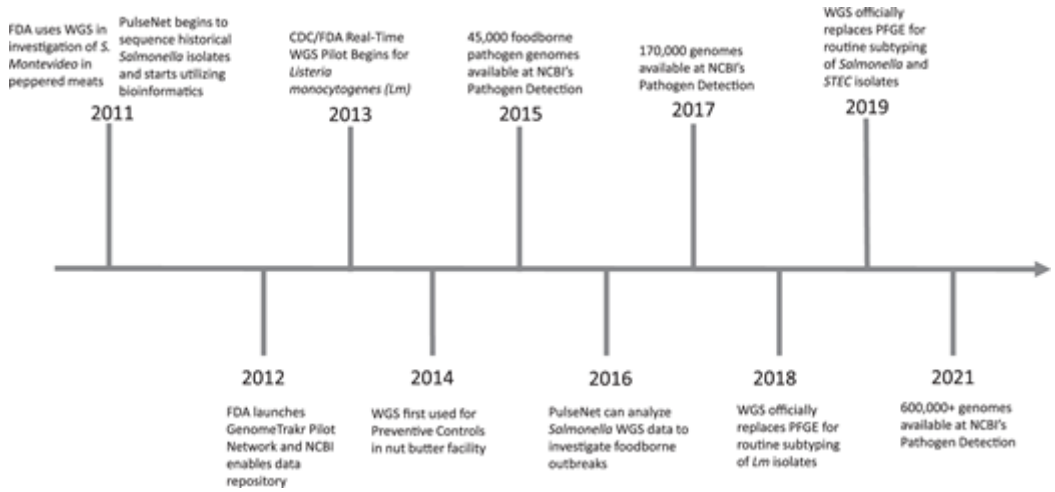
S



View large Download slide

Overview of PulseNet and GenomeTrakr networks.

FIGURE 5



View large Download slide

Key dates during the transition to whole genome sequencing of isolates associated with foodborne outbreaks in the United States.

PRACTICAL APPLICATIONS OF WGS IN FOOD SAFETY

evidence; when the strain causing illness is also identified in a food or processing environment, the greater sensitivity and especially specificity of WGS means a higher likelihood that this particular food, firm, or establishment could be related to the illnesses. Use of WGS indirectly strengthens the epidemiologic and traceback lines of evidence by reducing the likelihood of incorporating unrelated cases in epidemiologic or traceback investigations. However, WGS analyses alone cannot exclusively indicate causality. For example, when WGS analysis reveals a cluster of clinical isolates in a foodborne outbreak closely related to isolates from foods or firms, the WGS data cannot be used alone to conclusively determine the source of the illnesses. To confirm the cause of the outbreak, other lines of evidence from epidemiologic and/or traceback investigations are needed to corroborate the WGS results and patient exposure information and conclusively link the illnesses to a specific food or firm. Additional evidence, such as regulatory inspection findings or environmental assessments, may also be used to support WGS results in both outbreak and nonoutbreak situations. Interpretation of WGS results in the context of these other types of investigational data is critical to ensuring that public health action and regulatory decision making are as accurate and evidence based as possible.

The higher resolution of the WGS approach compared with the previously used PFGE approach is advantageous in several important ways in outbreak investigations for (i) excluding cases that are unrelated to the outbreak but have isolates with indistinguishable PFGE patterns; (ii) combining cases that have isolates with different PFGE patterns; (iii) identifying cases retrospectively; (iv) identifying outbreak investigations that would not have been triggered with traditional PFGE cluster detection methods; and (v) providing other important genetic information beyond relatedness.

Outbreak investigations illustrating the exclusion of cases with isolates that have th same PFGE patterns

WGS data allow for more specific case definitions that can be used to determine which ill people should be considered part of an outbreak. By using WGS and phylogenetic analysis to characterize the outbreak strain (i.e., by considering only closely related isolates or those within a defined clade to be part of the investigation), other isolates can be excluded that would otherwise be indistinguishable and thus included based on the PFGE subtyping method. Because clusters identified by WGS are more likely to include only those cases that share a common exposure (i.e., only those with closely related strains), this approach reduces misclassification of unrelated cases, which can dilute an epidemiologic signal. One example of this exclusionary power occurred during a 2015 outbreak of *Salmonella* serotype I 4,[5],12:i:− infections linked to pork *(44)*. The outbreak strain had a PFGE pattern that was very common nationally but relatively

allowed for exclusion of many illnesses that occurred in other states.

WGS cluster detection based on common exposures or closely related strains can help reduce noise and improve the epidemiologic signals from an outbreak, thus helping to focus an investigation. The agencies involved must carefully evaluate such information in a broader context of inclusivity and exclusivity to ensure that signals or sources are not inadvertently missed.

Outbreak investigations illustrating the inclusion of cases with isolates that have different PFGE patterns

A second way the higher resolution of WGS data impacts outbreak investigations is by grouping genetically related isolates together that appear different based on PFGE data. It may seem counterintuitive that WGS data may indicate that isolates are actually closely related when they appear unrelated based on the PFGE pattern, but this determination is related to the improved evolutionary models used for WGS analysis. Mobile genetic elements such as plasmids and phages, which may code for important phenotypic traits such as virulence or antibiotic resistance, may provide little information about the vertical evolution of a bacterial strain. However, these elements can affect the PFGE banding patterns. By masking mobile genetic elements in the WGS analysis pipeline, a more accurate picture of evolutionarily relevant differences between isolates can emerge. During an outbreak of *L. monocytogenes* infections linked to Middle Eastern–style cheeses *(12), L. monocytogenes* isolates with multiple PFGE pattern combinations were recovered from samples collected from the cheese production environment in both 2010 and 2015. Clinical isolates were identified throughout that time period with the same and other PFGE pattern combinations. WGS data revealed that the isolates with these different PFGE pattern combinations were highly genetically related. Combined with epidemiologic and traceback data, these WGS data helped prove that the fa had a resident strain of *Listeria* that persisted in the production environment and had been the sou human illnesses for at least 5 years. WGS verified that the chromosomal *Listeria* genome remained the same despite changing plasmids or phage elements that yielded different PFGE patterns.

Outbreak investigations illustrating retrospective identification of cases

A third means by which the higher resolution of WGS has affected investigations is for identification of outbreaks or other groups of illnesses tied to a common source, which would not have been identified with other subtyping methods. One example is the identification of an isolate with a PFGE pattern has a sufficiently high expected baseline frequency that only very large illness events would signal that an outbreak is occurring. Another example is instances in which illnesses linked to a common source occur

clinical isolates can initiate an investigation. In a 2015 outbreak of *L. monocytogenes* infections linked to ice cream *(11),* a multistate investigation began after multiple *L. monocytogenes* isolates were identified in ice cream collected for routine testing by the South Carolina Department of Public Health and Environment. This finding led investigators to look for potentially related isolates in the PulseNet database. Sequencing of these isolates and the subsequent epidemiologic, traceback, and environmental investigations ultimately linked illnesses occurring over a 5-year period to ice cream from two production facilities in Texas and Oklahoma that were owned by a single company.

Going beyond traditional cluster detection—use of WGS to link single illnesses to possible outbreak sources

The higher resolution of WGS has allowed much smaller groups of illnesses and even a single illness to be linked to a food or other product. When evaluating a link between illnesses and a food item, investigators consider the collective weight of three lines of evidence: epidemiologic, traceback, and microbiologic. When WGS-based microbiologic evidence supports a possible link between ill people and a food, less specific epidemiologic or traceback evidence is required to confirm the link than is needed for PFGE-based microbiologic evidence because of the stronger WGS evidence of genetic relatedness. For example, during a 2016 outbreak of STEC O121 infections linked to flour, STEC O26 was also isolated from the implicated flour *(17)*. This isolate was closely related, as deteremined by WGS, to an isolate from a single person who was ill with an STEC O26 infection and who reported exposure to the implicated flour in the week before becoming ill. Based on both the epidemiologic and WGS evidence, this ill person was considered part of the outbreak.

WGS provides other genetic information important in outbreak investigations

The final way WGS impacts outbreak investigations is by detection of genes known to be related to antimicrobial resistance, serotype, and virulence. By identifying all of these genes in a single workflow for all sequenced isolates, investigators obtain more timely data to use in an outbreak investigation. Traditional phenotypic resistance testing is generally limited to a smaller subset of outbreak isolates, but with WGS, predicted resistance data are available for all sequenced isolates. Genotypic detection of AMR genes played a prominent role in a 2017 outbreak of multidrug-resistant *C. jejuni* infections linked to contact with puppies purchased at pet stores *(35)*. In this investigation, WGS was fast and provided complete information about the resistance pattern of the isolates, including the fact that they were resistant to commonly recommended first-line antibiotics, which in turn informed treatment recommendations made to health care providers managing patients with suspected or confirmed

patient treatment recommendations.

OTHER USES OF WGS BEYOND OUTBREAK DETECTION

Foodborne illness source attribution

Foodborne illness source attribution analyses are used to identify which foods are the most important sources of foodborne illnesses in a population. Source attribution models have been created to estimate the percentage of illnesses caused by each pathogen transmitted by various food and nonfood categories *(74)*. Federal agencies and food safety experts rely on attribution analyses to inform strategic planning and risk-based decision making, estimate potential benefits of interventions, and evaluate the impact of current interventions, such as new or revised regulations, policies, and performance standards *(13)*.

Attribution analyses for the United States have included models that rely primarily on outbreak data. Future models should incorporate WGS, epidemiologic, and other data linked to isolates from sporadic and outbreak-associated illnesses and from food, water, animals, and environments. The advantage of WGS is illustrated by the challenge of attributing the sources of sporadic illnesses caused by the highly clonal *Salmonella* Enteritidis. Whereas PFGE and phage typing have been ineffective, WGS has higher discriminatory power, allowing specific lineages to be characterized *(21)*. Phylogenetic studies have indicated that lineages can be characterized by SNPs associated with specific host species. These lineages appear to exhibit genetic adaptation through the acquisition of virulence genes or pseudogenes *(31, 46)*.

Supervised machine learning (e.g., random forest and support vector machines) may be useful for predicting probable sources of sporadic infections caused by various pathogens *(51, 52, 88)*. Expanding the pool of sequenced isolates with relevant metadata (especially source and exposure information) could help refine attribution models and improve our understanding of the many transmission modes for human illness. To make attribution estimates more useful for informing regulatory policy, detailed information about food isolates with appropriate regulatory food category assignments *(66)* and adequate numbers of isolates for training predictive models are needed. These advances in data integration and modeling techniques will facilitate development of attribution models based on sporadic illnesses for which transmission routes may differ in important ways from those of outbreak-associated illnesses *(22)*.

Substantial overlap exists in the methodologies for using WGS and other data in source attribution models and those that could be used to generate hypotheses about the possible source of an outbreak. Improvements in data collection methods and analysis will likely improve the usefulness for both

other data could be added to determine whether other factors improve source prediction. Although WGS analysis of bacterial pathogens is transforming our ability to link illnesses to sources, interpretation requires flexibility to take into account all relevant information and implications of findings while protecting the food supply *(65)*.

Use in preventive control: an FDA and FSIS perspective

Some genetic adaptations in *Salmonella* and other foodborne pathogens associated with food production and processing environments may be responsible for foodborne illnesses and have become an additional public health concern *(5, 28, 46, 89)*. The emergence of outbreaks attributed to some strains may indicate changes in the adaptive characteristics of these organisms in a particular food commodity or environment. For example, a recent outbreak strain of *Salmonella* Bareilly recovered from tuna carried a novel genetic island containing a unique arsenic resistance operon *(34)*. Arsenic is a toxic metalloid present in the natural environment and is often detected at high concentrations in fish and seafood, which absorb the arsenic from the water in which they live. The acquisition of the resistance operon highlights the evolutionary capability of *Salmonella*; the operon confers a selective advantage for survival, persistence, and growth in some food matrices and environments, increasing the possibility for host morbidity and mortality *(87)*.

In the last few years, foodborne pathogens have adapted to various preventive controls originally used by food manufacturers to significantly minimize or prevent hazards. For example, some pathogens have become more tolerant and/or resistant to quaternary ammonium chloride compounds, biocides, chlorine, heavy metals, antibiotics, heat, salts, and acids *(16, 27, 42, 50, 60, 73, 81, 85)*. By integrating WGS data, especially long-read sequence data, with transcriptomic data, the genes, genomic cassettes, and other elements capable of conferring adaptive advantages to foodborne and environmental pathogens c identified. These new technologies will continue to improve the ability of food safety scientists to p and ultimately exploit key phenotypic and genotypic characteristics of isolates recovered from food processing environments and farms and allow development of newer and more effective preventive controls. An understanding of how some genes in pathogenic bacteria, including *Salmonella,* help them survive, adapt to, and evade preventive controls is important for preventing, predicting, monitoring, and containing these hazards with precision.

WGS for AMR surveillance

DNA sequencing technologies greatly enhance our ability to predict functional traits of bacteria traditionally captured by other methods such as antimicrobial susceptibility testing. In many foodborne

using WGS data alone (e.g., the FDA Resistome Tracker). WGS data can also be screened for genes and mutations that confer resistance to drugs that are not included in traditional phenotypic susceptibility tests. Traditional antimicrobial susceptibility testing will be important for discovering new AMR determinants that can be added to genetic screening tools and provides MIC confirmation of genetic predictions. However, WGS will complement and perhaps someday largely replace traditional phenotypic AMR screening.

The Center for Genomic Epidemiology (CGE) has developed a ResFinder database (https://cge.cbs.dtu.dk/services/ResFinder/) that can be used for detecting the presence and absence of AMR genes and chromosomal mutations conferring resistance. WGS data also reveal the genomic context of the resistance determinants, which often are located on plasmids and other mobile elements. This information is critical to our understanding of possible sources of resistance determinants, the capacity for horizontal spread, and coselection pressures driving AMR, including temporal and geographic distribution of specific genes. In some cases, multiple genes can code for resistance to an antimicrobial agent. The CGE Web site offers a PlasmidFinder tool that identifies plasmids in *Enterobacteriaceae* and gram-positive bacteria (https://cge.cbs.dtu.dk/services/PlasmidFinder/). This genetic information about AMR genes and mobile elements is also extracted as part of the routine PulseNet workflow and included in the NARMS Now Web page (https://wwwn.cdc.gov/narmsnow/) with both predicted and phenotypic resistance data.

NCBI has built the AMRFinderPlus tool to identify and assemble AMR genes *(24, 25)*. The results of the AMRFinder tool are added to the Isolates browser as gene calls for any isolate found to contain AMR genes. Specific isolates with specific AMR genes can be searched for in the interface.

Given the widespread application of WGS to identify resistance determinants and the high degree c concordance of WGS results with results of traditional antimicrobial susceptibility testing, WGS-bas prediction of AMR will undoubtedly be useful to regulatory agencies such as the FDA and the FSIS and will contribute to development and promotion of acceptance of new WGS methods and interpretations and applications of the data. By providing tools for monitoring the distribution of AMR determinants in pathogens across various agricultural commodities, geographic regions, and time, WGS methods and data will be useful for informing proper regulatory standards for agricultural practices both domestically and internationally. The WGS approach is a powerful tool for advancing the risk assessment processes associated with approval of new animal antimicrobial drugs and use of antimicrobials in consumer products.

Metagenomic sequencing is the collective identification of an entire microbial community within a sample.

The term “metagenome” was first published by Handelsman et al. *(32)* in a 1998 study in which bacterial cloning methods were used to identify the uncultured bacteria in soil, the soil metagenome. Advances in sequencing technology and lower costs have resulted in rapid expansion of metagenomic sequencing to include samples such as food and environmental and clinical specimens, and this technology can be used to inform public health agencies about pathogen presence and persistence. Samples can be processed without the need for cultures (culture independent) to release the genomic material from their respective microbial communities, which can then be sequenced and analyzed with bioinformatics to identify genetic markers specific to viruses, bacteria, and fungi, including virulence factors and AMR markers. The Human Microbiome and American Gut projects are example of the utilization of metagenomic sequencing to investigate human health and disease and have provided groundbreaking tools to support the use of this technology to investigate the prevalence and persistence of pathogens in the food supply *(38, 39, 54)*.

Limited shotgun metagenomic and targeted 16S rRNA amplicon sequencing methods are routinely used for food safety and clinical applications to test the specificity and sensitivity of pathogen detection methods in specific environments *(9, 23, 37, 48)*. However, current foodborne pathogen detection workflows rely on culture-based methods that can take weeks to complete. A typical workflow for *Salmonella, Listeria, Campylobacter,* or *E. coli* detection may include a series of nonselective and selective culture steps followed by plating on selective media to obtain a bacterial isolate, from which a genome can be determined and uploaded to the GenomeTrakr or PulseNet databases. The less time-intensive process application of metagenomics has been useful for food safety applications by improving pathogen detection. Some examples of applications focused on food safety, improving pathogen detection, and traceback analyses include the characterization of ice cream contaminated with *Listeria,* the detection of *Salmonella* in tomato and cilantro enrichment cultures, and the detection of *E. coli* in bagged fresh [illegible] *(19, 41, 48, 49, 61, 62)*. Metagenomics has also been successful in linking clinical *Salmonella* Heidelberg isolates from stool samples and food isolates in two 2017 foodborne outbreaks *(36)*. A significant finding in that study was that the two *Salmonella* Heidelberg isolates were indistinguishable by PFGE, but WGS revealed that the isolates were not closely related, thereby providing evidence for two separate outbreaks *(36)*.

Additional food safety applications of these limited metagenomic and targeted amplicon sequencing methods are identification of spoilage bacteria and risk points in food production, characterization or monitoring of fermentation processes, and identification of microbiome variations from the farm-to-consumer continuum *(7)*. For example, an assessment of a beef processing facility revealed significant

samples and significant reductions in *Campylobacter* levels during the journey from farm to consumer *(59)*. Microbiome profiling studies are also an efficient way to observe community profile shifts induced by temperature abuse and the impact of these shifts on pathogen persistence and to study the succession of bacteria throughout commercial processing of packaged leafy greens *(30, 72)*.

Complete metagenomic sequencing could transform future foodborne pathogen source tracking efforts because it eliminates the need for a pure culture, a requirement that currently is essential for GenomeTrakr and PulseNet WGS analyses and investigations. This approach makes possible a significant reduction in the time needed to initiate a traceback analysis and a significant increase in the number of bacterial and viral genomes identified in a sample. To keep pace with the rapid developments in the field of metagenomics and to systematically collect and store data related to the food microbiome, the FDA has launched a new MetagenomeTrakr pilot project (https://www.ncbi.nlm.nih.gov/bioproject/530970).

For PulseNet, an immediate need is the transition to culture-independent subtyping approaches as clinical laboratories adopt culture-independent diagnostic tests *(10)*. These molecular tests can be used to detect a pathogen directly from a sample without isolating a particular strain. Without an isolate, public health laboratories are often left with the burden of trying to isolate the bacterial strain from a patient sample, often days to weeks after it was collected. This delay in obtaining an isolate can delay the ability of public health agencies to identify potential outbreak clusters and sources. PulseNet is exploring highly multiplexed amplicon sequencing approaches that target hundreds to thousands of pathogen-specific gene targets for subtyping. These targets include AMR, virulence, and serotyping genes and genes used for cluster detection. PulseNet also is exploring targeted and shotgun metagenomics approaches for subtyping. The CDC, FDA, and FSIS laboratories all communicate and collaborate as they develop th… generation of subtyping approaches.

WHAT IS NEXT FOR WGS IN FOOD SAFETY

The continued application of advanced sequencing technologies and subsequent data analysis pipelines will contribute to the realization of greater food safety and enhanced public health by making it possible to identify, characterize, and subtype multiple pathogens directly from patient or environmental samples using metagenomic and other culture-independent methods. From a better understanding of the mechanisms allowing the introduction of pathogens into foods to the various manifestations of clinical disease, sequencing technologies will help scientists develop ways to prevent and reduce foodborne illness. In some ways, we are already using WGS to better understand how a pathogen's genome relates to

pathogenicity, survival, AMR profile, biocide resistance, etc. will lead to improvements in foodborne illness source attribution estimates and risk assessments and inform clinical treatments and preventive control options.

A global commitment to sharing data and developing better WGS tools will hasten decreases in foodborne contamination and improve public health protection. For example, PulseNet USA is working with PulseNet International, a network of >80 international public health laboratories that is also undergoing a transition to WGS *(9)*. This group is collaborating to validate the common cgMLST schemes for foodborne bacterial pathogens used by all members of the network and to create standardized analysis approaches for rapid comparison of isolates by sharing cgMLST allele calls and allele codes across borders. The improved ability to identify growing clusters of clinical illnesses with PulseNet USA and PulseNet International combined with the increased specificity in matching clinical isolates to food or environmental isolates (as with the FDA GenomeTrakr) makes U.S. public health agencies better positioned to respond to food safety problems. The use of WGS technologies by all stakeholders from the global food production chain can help prevent food from becoming contaminated and thus reduce the number of foodborne illnesses and improve global health. Improvements in WGS technology also will augment illness prevention efforts by reducing the time needed to get more accurate and user-friendly results.

The power of WGS and its significance in the fields of food safety and public health are already being realized, and U.S. public health agencies are committed to utilizing the full potential of WGS and recognize its potential impact on food safety and public health, both within the United States and around the world.

ACKNOWLEDGMENTS

This work was supported in part by the intramural research program of the National Library of Med NIH. The authors thank April Pradier for her coordinated management of the manuscript and all participants and stakeholders of PulseNet, GenomeTrakr, and NCBI. The findings and conclusions in this article are those of the authors and do not necessarily represent the official positions of the CDC, FDA, NIH, or participating USDA agencies.

REFERENCES

1. Allard, M. W., Bell R., Ferreira C. M., Gonzalez-Escalona N., Hoffmann M., Muruvanda T., Ottesen A., Ramachandran P., Reed E., Sharma S., Stevens E., Timme R., Zheng J., and Brown E. W.. 2018. Genomics of foodborne pathogens for microbial food safety. *Curr. Opin. Biotechnol.* 49: 224– 229.

High resolution clustering of *Salmonella enterica* serovar Montevideo strains using a next-generation sequencing approach. *BMC Genom.* 13: 32.
Google Scholar

3. Allard, M. W., Strain E., Melka D., Bunning K., Musser S. M., Brown E. W., and Timme R.. 2016. Practical value of food pathogen traceability through building a whole-genome sequencing network and database. *J. Clin. Microbiol.* 54: 1975– 1983.
Google Scholar

4. Asakura, H., Takahashi N., Yamamoto S., and Maruyama H.. 2017. Draft genome sequence of *Campylobacter jejuni* CAM970 and *C. coli* CAM962, associated with a large outbreak of foodborne illness in Fukuoka, Japan, in 2016. *Genome Announc.* 5: e00508– 17.
Google Scholar

5. Baumler, A. J., Tsolis R. M., Ficht T. A., and Adams L. G.. 1998. Evolution of host adaptation in *Salmonella enterica*. *Infect. Immun.* 66: 4579– 4587.
Google Scholar

6. Benson, D. A., Cavanaugh M., Clark K., Karsch-Mizrachi I., Ostell J., Pruitt K. D., and Sayers E. W.. 2018. GenBank. *Nucleic Acids Res.* 46: D41– D47.
Google Scholar

7. Bokulich, N. A., Lewis Z. T., Boundy-Mills K., and Mills D. A.. 2016. A new perspective on microbial landscapes within food production. *Curr. Opin. Biotechnol.* 37: 182– 189.
Google Scholar

8. Brown, E., Dessai U., McGarry S., and Gerner-Smidt P.. 2019. Use of whole-genome sequencing for food safety and public health in the United States. *Foodborne Pathog. Dis.* 16: 441– 450.
Google Scholar

9. Buytaers, F. E., Saltykova A., Mattheus W., Verhaegen B., Roosens N. H. C., Vanneste K., Laisnez V., Hammami N., Pochet B., Cantaert V., Marchal K., Denayer S., and De Keersmaecker S. C. J.. 2021. Application of a strain-level shotgun metagenomics approach on food samples: resolution of the source of a *Salmonella* food-borne outbreak. *Microb. Genom.* 7 (4): 000547.
Google Scholar

10. Carleton, H. A., Besser J., Williams-Newkirk A. J., Huang A., Trees E., and Gerner-Smidt P.. 2019. Metagenomic approaches for public health surveillance of foodborne infections: opportunities and challenges. *Foodborne Pathog. Dis.* 16: 474– 479.
Google Scholar

m-03-15/index.html. Accessed 29 March 2022.

12. Centers for Disease Control and Prevention. 2015. Multistate outbreak of listeriosis linked to soft cheeses distributed by Karoun Dairies, Inc. (final update). Available at: https://www.cdc.gov/listeria/outbreaks/soft-cheeses-09-15/index.html. Accessed 29 March 2022.

13. Centers for Disease Control and Prevention. 2017. Interagency Food Safety Analytics Collaboration strategic plan 2017–2021. Available at: https://www.cdc.gov/foodsafety/pdfs/IFSAC-Strategic-Plan-2017-2021.pdf. Accessed 29 March 2022.

14. Cherry, J. L. 2017. A practical exact maximum compatibility algorithm for reconstruction of recent evolutionary history. *BMC Bioinformatics* 18: 127.
Google Scholar

15. Cochrane, G., Karsch-Mizrachi I., Takagi T., and International Nucleotide Sequence Database Collaboration. 2016. The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Res.* 44: D48– D50.
Google Scholar

16. Conficoni, D., Losasso C., Cortini E., Di Cesare A., Cibin V., Giaccone V., Corno G., and Ricci A.. 2016. Resistance to biocides in *Listeria monocytogenes* collected in meat-processing environments. *Front. Microbiol.* 7: 1627.
Google Scholar

17. Crowe, S. J., Bottichio L., Shade L. N., Whitney B. M., Corral N., Melius B., Arends K. D., Donovan D., Stone J., Allen K., Rosner J., Beal J., Whitlock L., Blackstock A., Wetherington J., Newberry L. A., Schroeder M. N., Wagner D., Trees E., Viazis S., Wise M. E., and Neil K. P.. 2017. Shiga toxin–producing *E. coli* infections associated with flour. *N. Engl. J. Med.* 377: 2036– 2043.
Google Scholar

18. Crowe, S. J., Green A., Hernandez K., Peralta V., Bottichio L., Defibaugh-Chavez S., Douris A., Gieraltowski L., Hise K., La-Pham K., Neil K. P., Simmons M., Tillman G., Tolar B., Wagner D., Wasilenko J., Holt K., Trees E., and Wise M. E.. 2017. Utility of combining whole genome sequencing with traditional investigational methods to solve foodborne outbreaks of *Salmonella* infections associated with chicken: a new tool for tackling this challenging food vehicle. *J. Food Prot.* 80: 654– 660.
Google Scholar

19. Daquigan, N., Grim C. J., White J. R., Hanes D. E., and Jarvis K. G.. 2016. Early recovery of *Salmonella* from food using a 6-hour non-selective pre-enrichment and reformulation of tetrathionate broth. *Front. Microbiol.* 7: 2103.
Google Scholar

PeerJ Comput. Sci. 1: e20.
Google Scholar

21. Deng, X., den Bakker H. C., and Hendriksen R. S.. 2016. Genomic epidemiology: whole-genome-sequencing–powered surveillance and outbreak investigation of foodborne bacterial pathogens. *Annu. Rev. Food Sci. Technol*. 7: 353– 374.
Google Scholar

22. Ebel, E. D., Williams M. S., Cole D., Travis C. C., Klontz K. C., Golden N. J., and Hoekstra R. M.. 2016. Comparing characteristics of sporadic and outbreak-associated foodborne illnesses, United States, 2004–2011. *Emerg. Infect. Dis*. 22: 1193– 1200.
Google Scholar

23. Fay, M., Salazar J. K., Ramachandran P., and Stewart D.. 2021. Microbiomes of commercially-available pine nuts and sesame seeds. *PLoS One* 16: e0252605.
Google Scholar

24. Feldgarden, M., Brover V., Gonzalez-Escalona N., Frye J. G., Haendiges J., Haft D. H., Hoffmann M., Pettengill J. B., Prasad A. B., Tillman G. E., Tyson G. H., and Klimke W.. 2021. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci. Rep*. 11: 12728.
Google Scholar

25. Feldgarden, M., Brover V., Haft D. H., Prasad A. B., Slotta D. J., Tolstoy I., Tyson G. H., Zhao S., Hsu C. H., McDermott P. F., Tadesse D. A., Morales C., Simmons M., Tillman G., Wasilenko J., Folster J. P., and Klimke W.. 2019. Validating the AMRFinder Tool and Resistance Gene Database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob. Agents Chemother*. 63 (11): e00483– 19.
Google Scholar

26. Felsenstein, J. 2004. *Inferring phylogenies*. Sinauer, Sunderland, MA.

27. Foster, J. W. 1995. Low pH adaptation and the acid tolerance response of *Salmonella* Typhimurium. *Crit. Rev. Microbiol*. 21: 215– 237.
Google Scholar

28. Franklin, K., Lingohr E. J., Yoshida C., Anjum M., Bodrossy L., Clark C. G., Kropinski A. M., and Karmali M. A.. 2011. Rapid genoserotyping tool for classification of *Salmonella* serovars. *J. Clin. Microbiol*. 49: 2954– 2965.
Google Scholar

29. Gangiredla, J., Rand H., Benisatto D., Payne J., Strittmatter C., Sanders J., Wolfgang W. J., Libuit K., Herrick J. B., Prarat M., Toro M., Farrell T., and Strain E.. 2021. GalaxyTrakr: a distributed analysis tool for

30. Gu, G., Ottesen A., Bolten S., Ramachandran P., Reed E., Rideout S., Luo Y., Patel J., Brown E., and Nou X.. 2018. Shifts in spinach microbial communities after chlorine washing and storage at compliant and abusive temperatures. *Food Microbiol*. 73: 73– 84.
Google Scholar

31. Guard, J., Morales C. A., Fedorka-Cray P., and Gast R. K.. 2011. Single nucleotide polymorphisms that differentiate two subpopulations of *Salmonella* Enteritidis within phage type. *BMC Res. Notes* 4: 369.
Google Scholar

32. Handelsman, J., Rondon M. R., Brady S. F., Clardy J., and Goodman R. M.. 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol*. 5: R245– R249.
Google Scholar

33. Henri, C., Leekitcharoenphon P., Carleton H. A., Radomski N., Kaas R. S., Mariet J. F., Felten A., Aarestrup F. M., Gerner Smidt P. Roussel S., Guillier L., Mistou M. Y., and Hendriksen R. S.. 2017. An assessment of different genomic approaches for inferring phylogeny of *Listeria monocytogenes*. *Front. Microbiol*. 8: 2351.
Google Scholar

34. 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. F., Catalyurek U. V., Wang H., Hammack T. S., Wolfgang W. J., Schoonmaker-Bopp D., Chu A., Myers R., Haendiges J., Evans P. S., Meng J., Strain E. A., Allard M. W., and Brown E. W.. 2016. Tracing origins of the *Salmonella* Bareilly strain causing a food-borne outbreak in the United States. *J. Infect. Dis*. 213: 502– 508.
Google Scholar

35. Hoffmann, S., Maculloch B., and Batz M.. 2015. *Economic burden of major foodborne illnesses acquired in the United States. EIB-140*. U.S. Department of Agriculture, Economic Research Service, Washington, DC.

36. Huang, A. D., Luo C., Pena-Gonzalez A., Weigand M. R., Tarr C. L., and Konstantinidis K. T.. 2017. Metagenomics of two severe foodborne outbreaks provides diagnostic signatures and signs of coinfection not attainable by traditional methods. *Appl. Environ. Microbiol*. 83 (3): e02577– 16.
Google Scholar

37. Huang, L., Hong B., Yang W., Wang L., and Yu R.. 2021. *Snipe: highly sensitive pathogen detection from metagenomic sequencing data. Brief Bioinformatics 22:bbab064.*

38. Human Microbiome Project Consortium. 2012. A framework for human microbiome research. *Nature* 486: 215– 221.

40. Jackson, B. R., 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., Pouseele H., Klimke W., Besser J., Brown E., Allard M., and 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.
Google Scholar

41. Jarvis, K. G., White J. R., Grim C. J., Ewing L., Ottesen A. R., Beaubrun J. J., Pettengill J. B., Brown E., and Hanes D. E.. 2015. Cilantro microbiome before and after nonselective pre-enrichment for *Salmonella* using 16S rRNA and metagenomic sequencing. *BMC Microbiol*. 15: 160.
Google Scholar

42. Kalburge, S. S., Whitaker W. B., and Boyd E. F.. 2014. High-salt preadaptation of *Vibrio parahaemolyticus* enhances survival in response to lethal environmental stresses. *J. Food Prot*. 77: 246– 253.
Google Scholar

43. Katz, L. S., Griswold T., Williams-Newkirk A. J., Wagner D., Petkau A., Sieffert C., Van Domselaar G., Deng X., and Carleton H. A.. 2017. A comparative analysis of the Lyve-SET phylogenomics pipeline for genomic epidemiology of foodborne pathogens. *Front. Microbiol*. 8: 375.
Google Scholar

44. Kawakami, V. M., Bottichio L., Angelo K., Linton N., Kissler B., Basler C., Lloyd J., Inouye W., Gonzales E., Rietberg K., Melius B., Oltean H., Wise M., Sinatra J., Marsland P., Li Z., Meek R., Kay M., Duchin J., and Lindquist S.. 2016. Notes from the field: outbreak of multidrug-resistant *Salmonella* infections linked to pork—Washington, 2015. *Morb. Mortal. Wkly. Rep*. 65: 379– 381.
Google Scholar

45. Kodama, Y., Shumway M., Leinonen R., and International Nucleotide Sequence Database Collaboration. 2012. The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res*. 40: D54– D56.
Google Scholar

46. Langridge, G. C., Fookes M., Connor T. R., Feltwell T., Feasey N., Parsons B. N., Seth-Smith H. M., Barquist L., Stedman A., Humphrey T., Wigley P., Peters S. E., Maskell D. J., Corander J., Chabalgoity J. A., Barrow P., Parkhill J., Dougan G., and Thomson N. R.. 2015. Patterns of genome evolution that have accompanied host adaptation in *Salmonella*. *Proc. Natl. Acad. Sci. USA* 112: 863– 868.
Google Scholar

Google Scholar

48. Leonard, S. R., Mammel M. K., Lacher D. W., and Elkins C. A.. 2015. Application of metagenomic sequencing to food safety: detection of Shiga toxin–producing *Escherichia coli* on fresh bagged spinach. *Appl. Environ. Microbiol*. 81: 8183– 8191.
Google Scholar

49. Leonard, S. R., Mammel M. K., Lacher D. W., and Elkins C. A.. 2016. Strain-level discrimination of Shiga toxin–producing *Escherichia coli* in spinach using metagenomic sequencing. *PLoS One* 11: e0167870.
Google Scholar

50. Li, H., and Ganzle M.. 2016. Some like it hot: heat resistance of *Escherichia coli* in food. *Front. Microbiol*. 7: 1763.
Google Scholar

51. Lupolova, N., Dallman T. J., Holden N. J., and Gally D. L.. 2017. Patchy promiscuity: machine learning applied to predict the host specificity of *Salmonella enterica* and *Escherichia coli*. *Microb. Genom*. 3 (10): e000135.
Google Scholar

52. Lupolova, N., Dallman T. J., Matthews L., Bono J. L., and Gally D. L.. 2016. Support vector machine applied to predict the zoonotic potential of *E. coli* O157 cattle isolates. *Proc. Natl. Acad. Sci. USA* 113: 11312– 11317.
Google Scholar

53. Marano, N. N., Rossiter S., Stamey K., Joyce K., Barrett T. J., Tollefson L. K., and Angulo F. J.. 2000. The National Antimicrobial Resistance Monitoring System (NARMS) for enteric bacteria, 1996–1999: surveillance for action. *J. Am. Vet. Med. Assoc*. 217: 1829– 1830.
Google Scholar

54. McDonald, D., Hyde E., Debelius J. W., Morton J. T., Gonzalez A., Ackermann G., Aksenov A. A., Behsaz B., Brennan C., Chen Y., DeRight Goldasich L., Dorrestein P. C., Dunn R. R., Fahimipour A. K., Gaffney J., Gilbert J. A., Gogul G., Green J. L., Hugenholtz P., Humphrey G., Huttenhower C., Jackson M. A., Janssen S., Jeste D. V., Jiang L., Kelley S. T., Knights D., Kosciolek T., Ladau J., Leach J., Marotz C., Meleshko D., Melnik A. V., Metcalf J. L., Mohimani H., Montassier E., Navas-Molina J., Nguyen T. T., Peddada S., Pevzner P., Pollard K. S., Rahnavard G., Robbins-Pianka A., Sangwan N., Shorenstein J., Smarr L., Song S. J., Spector T., Swafford A. D., Thackray V. G., Thompson L. R., Tripathi A., Vazquez-Baeza Y., Vrbanac A., Wischmeyer P., Wolfe E., Zhu Q., American Gut Consortium, and Knight R.. 2018. American Gut: an open platform for citizen science microbiome research. *mSystems* 3 (3): e00031– 18.
Google Scholar

Google Scholar

56. 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., and Lecuit M.. 2017. Real-time whole-genome sequencing for surveillance of *Listeria monocytogenes,* France. *Emerg. Infect. Dis*. 23: 1462– 1470.
Google Scholar

57. Nadon, C., Van Walle I., Gerner-Smidt P., Campos J., Chinen I., Concepcion-Acevedo J., Gilpin B., Smith A. M., Man Kam K., Perez E., Trees E., Kubota K., Takkinen J., Nielsen E. M., Carleton H., and FWD-NEXT Expert Panel. 2017. PulseNet International: vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. *Euro Surveill*. 22 (23): 30544.
Google Scholar

58. NCBI Resource Coordinators. 2017. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 45: D12– D17.

59. Oakley, B. B., Morales C. A., Line J., Berrang M. E., Meinersmann R. J., Tillman G. E., Wise M. G., Siragusa G. R., Hiett K. L., and Seal B. S.. 2013. The poultry-associated microbiome: network analysis and farm-to-fork characterizations. *PLoS One* 8: e57190.
Google Scholar

60. Ortiz, S., Lopez-Alonso V., Rodriguez P., and Martinez-Suarez J. V.. 2015. The connection between persistent, disinfectant-resistant *Listeria monocytogenes* strains from two geographically separate Iberian pork processing plants: evidence from comparative genome analysis. *Appl. Environ. Microbiol*. 82: 308– 317.
Google Scholar

61. Ottesen, A., Ramachandran P., Reed E., White J. R., Hasan N., Subramanian P., Ryan G., Jarvis K
Grim C., Daquiqan N., Hanes D., Allard M., Colwell R., Brown E., and Chen Y.. 2016. Enrichment dyna
of *Listeria monocytogenes* and the associated microbiome from naturally contaminated ice cream
linked to a listeriosis outbreak. *BMC Microbiol*. 16: 275.
Google Scholar

62. Ottesen, A. R., Gonzalez A., Bell R., Arce C., Rideout S., Allard M., Evans P., Strain E., Musser S., Knight R., Brown E., and Pettengill J. B.. 2013. Co-enriching microflora associated with culture based methods to detect *Salmonella* from tomato phyllosphere. *PLoS One* 8: e73079.
Google Scholar

63. Page, A. J., Alikhan N. F., Carleton H. A., Seemann T., Keane J. A., and Katz L. S.. 2017. Comparison of classical multi-locus sequence typing software for next-generation sequencing data. *Microb. Genom*. 3: e000124.

analysis of core genome MLST and SNP typing within a European *Salmonella* serovar Enteritidis outbreak. *Int. J. Food Microbiol*. 274: 1– 11.
Google Scholar

65. Pightling, A. W., Pettengill J. B., Luo Y., Baugher J. D., Rand H., and Strain E.. 2018. Interpreting whole-genome sequence analyses of foodborne bacteria for regulatory applications and outbreak investigations. *Front. Microbiol*. 9: 1482.
Google Scholar

66. Richardson, L. C., Bazaco M. C., Parker C. C., Dewey-Mattia D., Golden N., Jones K., Klontz K., Travis C., Kufel J. Z., and Cole D.. 2017. An updated scheme for categorizing foods implicated in foodborne disease outbreaks: a tri-agency collaboration. *Foodborne Pathog. Dis*. 14: 701– 710.
Google Scholar

67. Ronholm, J., Nasheri N., Petronella N., and Pagotto F.. 2016. Navigating microbiological food safety in the era of whole-genome sequencing. *Clin. Microbiol. Rev*. 29: 837– 857.
Google Scholar

68. Scallan, E., Griffin P. M., Angulo F. J., Tauxe R. V., and Hoekstra R. M.. 2011. Foodborne illness acquired in the United States—unspecified agents. *Emerg. Infect. Dis*. 17: 16– 22.
Google Scholar

69. Scallan, E., Hoekstra R. M., Angulo F. J., Tauxe R. V., Widdowson M. A., Roy S. L., Jones J. L., and Griffin P. M.. 2011. Foodborne illness acquired in the United States—major pathogens. *Emerg. Infect. Dis*. 17: 7– 15.
Google Scholar

70. Scharff, R. L., Besser J., Sharp D. J., Jones T. F., Peter G. S., and Hedberg C. W.. 2016. An economic evaluation of PulseNet: a network for foodborne disease surveillance. *Am. J. Prev. Med*. 50: S66– S7
Google Scholar

71. Shane, A. L., Mody R. K., Crump J. A., Tarr P. I., Steiner T. S., Kotloff K., Langley J. M., Wanke C., Warren C. A., Cheng A. C., Cantey J., and Pickering L. K.. 2017. 2017 Infectious Diseases Society of America clinical practice guidelines for the diagnosis and management of infectious diarrhea. *Clin. Infect. Dis*. 65: 1963– 1973.
Google Scholar

72. Soderqvist, K., Osman O. Ahmed, Wolff C., Bertilsson S., Vagsholm I., and Boqvist S.. 2017. Emerging microbiota during cold storage and temperature abuse of ready-to-eat salad. *Infect. Ecol. Epidemiol*. 7: 1328963.
Google Scholar

bacteria following exposure to quaternary ammonium compounds. *J. Appl. Microbiol.* 121: 1275– 1281.
Google Scholar

74. Souvorov, A., Agarwala R., and Lipman D. J.. 2018. SKESA: strategic k-mer extension for scrupulous assemblies. *Genome Biol.* 19: 153.
Google Scholar

75. Swaminathan, B., Barrett T. J., Hunter S. B., Tauxe R. V., and CDC PulseNet Task Force. 2001. PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. *Emerg. Infect. Dis.* 7: 382– 389.
Google Scholar

76. Timme, R. E., Lafon P. C., Balkey M., Adams J. K., Wagner D., Carleton H., Strain E., Hoffmann M., Sabol A., Rand H., Lindsey R., Sheehan D., Baugher J. D., and Trees E.. 2020. Gen-FS coordinated proficiency test data for genomic foodborne pathogen surveillance, 2017 and 2018 exercises. *Sci. Data* 7: 402.
Google Scholar

77. Timme, R. E., Rand H., Sanchez Leon M., Hoffmann M., Strain E., Allard M., Roberson D., and Baugher J. D.. 2018. GenomeTrakr proficiency testing for foodborne pathogen surveillance: an exercise from 2015. *Microb. Genom.* 4 (7): e000185.
Google Scholar

78. Timme, R. E., Rand H., Shumway M., Trees E. K., Simmons M., Agarwala R., Davis S., Tillman G. E., Defibaugh-Chavez S., Carleton H. A., Klimke W. A., and Katz L. S.. 2017. Benchmark datasets for phylogenomic pipeline validation, applications for foodborne pathogen surveillance. *PeerJ* 5: e3893.
Google Scholar

79. Timme, R. E., Sanchez Leon M., and Allard M. W.. 2019. Correction to: Utilizing the public GenomeTrakr database for foodborne pathogen traceback. *Methods Mol. Biol.* 1918: C1.
Google Scholar

80. Tolar, B., Joseph L. A., Schroeder M. N., Stroika S., Ribot E. M., Hise K. B., and Gerner-Smidt P.. 2019. An overview of PulseNet USA databases. *Foodborne Pathog. Dis.* 16: 457– 462.
Google Scholar

81. Usaga, J., Worobo R. W., and Padilla-Zakour O. I.. 2014. Effect of acid adaptation and acid shock on thermal tolerance and survival of *Escherichia coli* O157:H7 and O111 in apple juice. *J. Food Prot.* 77: 1656– 1663.
Google Scholar

82. U.S. Department of Agriculture, Food Safety and Inspection Service. 2016. FSIS strategic framework for FY 2011–2016. Available at: https://www.fsis.usda.gov/shared/PDF/Strategic_Plan_2011

Inspection Service (FSIS) 2017–2021 strategic plan. Available at: https://www.fsis.usda.gov/sites/default/files/media_file/2020-07/Strategic-Plan-2017-2021.pdf. Accessed 29 March 2022.

84. U.S. Department of Agriculture, Food Safety and Inspection Service. 2017. FSIS directive 8080.3, rev. 2. Foodborne illness investigations. Available at: https://www.fsis.usda.gov/wps/wcm/connect/1bffb125-cd80-4b3a-ab45-1f4c0ad863d9/8080.3.pdf?MOD=AJPERES. Accessed 29 March 2022.

85. Wiktorczyk-Kapischke, N., Skowron K., Grudlewska-Buda K., Walecka-Zacharska E., Korkus J., and Gospodarek-Komkowska E.. 2021. Adaptive response of *Listeria monocytogenes* to the stress factors in the food processing environment. *Front. Microbiol*. 12: 710085.
Google Scholar

86. Yang, X., Noyes N. R., Doster E., Martin J. N., Linke L. M., Magnuson R. J., Yang H., Geornaras I., Woerner D. R., Jones K. L., Ruiz J., Boucher C., Morley P. S., and Belk K. E.. 2016. Use of metagenomic shotgun sequencing technology to detect foodborne pathogens within the microbiome of the beef production chain. *Appl. Environ. Microbiol*. 82: 2433– 2443.
Google Scholar

87. Yuk, H. G., and Schneider K. R.. 2006. Adaptation of *Salmonella* spp. in juice stored under refrigerated and room temperature enhances acid resistance to simulated gastric fluid. *Food Microbiol*. 23: 694– 700.
Google Scholar

88. Zhang, S., Li S., Gu W., den Bakker H., Boxrud D., Taylor A., Roe C., Driebe E., Engelthaler D. M., Allard M., Brown E., McDermott P., Zhao S., Bruce B. B., Trees E., Fields P. I., and Deng X.. 2019. Zoonotic source attribution of *Salmonella enterica* serotype Typhimurium using genomic surveillance data, United States. *Emerg. Infect. Dis*. 25: 82– 91.
Google Scholar

89. Zheng, J., Allard S., Reynolds S., Millner P., Arce G., Blodgett R. J., and Brown E. W.. 2013. Colonization and internalization of *Salmonella enterica* in tomato plants. *Appl. Environ. Microbiol*. 79: 2494– 2502.
Google Scholar

Author notes

† These authors contributed equally.

for you:

Retail Deli Slicer Inspection Practices: An EHS-Net Study
Editors' Choice.
LAUREN E. LIPCSEI et al., Journal of Food Protection, 2018

Initial Reports of Foodborne Illness Drive More Public Attention Than Do Food Recall Announcements
JINHO JUNG et al., Journal of Food Protection, 2021

Characterization of Listeria monocytogenes Recovered from Imported Cheese Contributed to the National PulseNet Database by the U.S. Food and Drug Administration from 2001 to 2008
BABGALEH B. TIMBO et al., Journal of Food Protection, 2010

Potential Use of DNA Barcodes in Regulatory Science: Identification of the U.S. Food and Drug Administration's “Dirty 22,” Contributors to the Spread of Foodborne Pathogens
YOLANDA L. JONES et al.,

Listeriosis outbreak associated with packaged leafy greens
Vanessa Raabe et al., AAP News, 2016

GalaxyTrakr: a distributed analysis tool for public health whole genome sequence data accessible to non-bioinformaticians
Jayanthi Gangiredla et al., BMC Genomics, 2021

How Can New and Emerging Therapies Improve Outcomes for Patients with Cervical Cancer?

ReachMD

68 deaths associated with foodborne diseases in 2012
Larry K. Pickering, AAP News, 2013

FDA Researchers Demonstrate Utility of Nanopore Sequencing for Foodborne Pathogen Detection

Huanjia Zhang, GenomeWeb, 2022

Skip to Main Content

Utility of Combining Whole Genome Sequencing with Traditional Investigational Methods To Solve Foodborne Outbreaks of Salmonella Infections Associated with Chicken: A New Tool for Tackling This Challenging Food Vehicle
Samuel J. Crowe et al., Journal of Food Protection, 2017

Skip to Main Content

New Content Alert

Journal of Food Protection®

International Association for Food Protection®

eISSN: 1944-9097 **ISSN:** 0362-028X