Genomic Epidemiology: Whole-Genome-Sequencing–Powered Surveillance and Outbreak Investigation of Foodborne Bacterial Pathogens

Xiangyu Deng,[1] Henk C. den Bakker,[2] and Rene S. Hendriksen[3]

[1]Center for Food Safety and Department of Food Science and Technology, University of Georgia, Griffin, Georgia 30269; email: xdeng@uga.edu

[2]International Center for Food Industry Excellence, Department of Animal and Food Sciences, Texas Tech University, Lubbock, Texas 79409

[3]National Food Institute, Research Group of Genomic Epidemiology, Technical University of Denmark, Kongens Lyngby, DK-2800 Denmark

Annu. Rev. Food Sci. Technol. 2016. 7:353–74

First published online as a Review in Advance on January 11, 2016

The *Annual Review of Food Science and Technology* is online at food.annualreviews.org

This article's doi:
10.1146/annurev-food-041715-033259

Keywords

food safety, public health, epidemiology, subtyping, bioinformatics, genomics

Abstract

As we are approaching the twentieth anniversary of PulseNet, a network of public health and regulatory laboratories that has changed the landscape of foodborne illness surveillance through molecular subtyping, public health microbiology is undergoing another transformation brought about by so-called next-generation sequencing (NGS) technologies that have made whole-genome sequencing (WGS) of foodborne bacterial pathogens a realistic and superior alternative to traditional subtyping methods. Routine, real-time, and widespread application of WGS in food safety and public health is on the horizon. Technological, operational, and policy challenges are still present and being addressed by an international and multidisciplinary community of researchers, public health practitioners, and other stakeholders.

INTRODUCTION

Despite improvements to ensure microbiological safety of our food supply over the past century, foodborne illness continues to be a major and enduring threat to public health (Doyle et al. 2015). It is estimated that each year roughly 1 in 6 Americans is sickened by foodborne illness, causing 128,000 hospitalizations and 3,000 deaths (Scallan et al. 2011). The majority of the known 31 foodborne pathogens acquired in the United States are bacterial species.

Laboratory identification and differentiation of bacterial infectious agents are critical for timely detection of foodborne outbreaks, accurate pinpointing of transmission vehicles, and fast removal of contaminated food items from circulation. Since the mid-1990s, molecular subtyping has been an instrumental tool for surveillance and outbreak investigation of foodborne illness; its utility is exemplified by the widespread adoption of a bacterial subtyping or fingerprinting technique called pulsed-field gel electrophoresis (PFGE) in public health laboratories across the United States and other countries (Gerner-Smidt et al. 2006). PFGE and other subtyping methods, while having substantially advanced foodborne disease surveillance, were developed well before the emergence of the so-called next-generation sequencing (NGS) technologies for whole-genome sequencing (WGS), which has revolutionized many areas of biomedical research and very recently started to transform public health microbiology.

This review is aimed to provide an introduction and overview of the emerging field herein referred to as genomic epidemiology. It should be noted that public health implementation of WGS is still at an early stage that features a dynamic landscape of technological innovations, novel applications, research, surveillance, and policy initiatives as well as new opportunities and imposing challenges. In recognition of the multidisciplinary nature of the subject, especially its technical roots in genomics and bioinformatics, a brief summary of sequencing platforms, informatics methods, and software tools important for public health applications is also included.

FOODBORNE PATHOGEN SURVEILLANCE AND SUBTYPING

Surveillance of Foodborne Illness

Public health surveillance is defined by the World Health Organization (WHO) as a systematic ongoing collection, collation, analysis, and interpretation of data and a timely dissemination of information (**http://www.who.int/topics/public_health_surveillance/en/**). Surveillance is conducted to facilitate better control of diseases and lead to public health actions such as outbreak detection; to measure the magnitude, burden, and trends of disease; to improve the knowledge of causes, sources, reservoirs, risks, morbidity, and mortality; to guide programs to measure the effectiveness of interventions; and to assist policymakers in setting priorities. To date, four generic types of enteric disease surveillance systems exist: no formal (occasional) surveillance, syndromic surveillance, laboratory-based surveillance, and integrated laboratory-based [One Health (Inst. Med. 2012)] surveillance. The last two types are mostly associated with foodborne pathogens (**Table 1**) and stand out from the first two types of surveillance systems by being based on laboratory results. Integrated foodborne surveillance refers to concomitant testing of isolates from humans, animals, and foods, and may include environmental and animal-feed samples (Jamison 2006).

Molecular Subtyping

In the past decade, our ability to distinguish in a rapid and reliable way between epidemiologically related isolates from the same bacterial species has increased, thereby enhancing our capacity to

Table 1 Major foodborne bacterial pathogens

Pathogen	Primary foodborne species/type	Disease	Incidence	Burden	Source
Salmonella	*Salmonella enterica* ssp. *enterica* serovar Enteritidis and Typhimurium (Hendriksen et al. 2011)	Salmonellosis, mild and self-limiting gastroenteritis, 5% of cases develop bacteremia, typhoid fever (Hohmann 2001)	In the US, 15.2 cases per 100,000 inhabitants in 2013 (Crim et al. 2014)	Annually 1.2 million illnesses and 450 deaths in the US (Scallan et al. 2011)	Food of animal origin, such as beef, poultry, pork, as well as eggs, vegetables, and ready-to-eat food
Campylobacter	*Campylobacter jejuni/ Campylobacter coli*	Campylobacteriosis, mild and self-limiting gastroenteritis, Guillain-Barré syndrome, Miller-Fisher syndrome (Bolton 2015)	In the US, 13.8 cases per 100,000 inhabitants in 2013 (Crim et al. 2014)	An estimated 1.3 million persons are affected each year	Primarily broilers but also in pigs, cattle, wild animals, and pets. Water and ready-to-eat food can also be contaminated (Kaakoush et al. 2015)
Escherichia coli	Shiga toxin–producing *E. coli* (STEC), including serovar O157 as well as a number of non-O157	Mild intestinal discomfort to hemolytic uremic syndrome, renal disease, death (Majowicz et al. 2014)	In the US, 1.15 STEC O157 cases and 1.17 STEC non-O157 cases per 100,000 inhabitants in 2013 (Crim et al. 2014)	An estimated 265,000 illnesses per year (Scallan et al. 2011)	Primarily from ruminant animals
Listeria	*Listeria monocytogenes*	Listeriosis, mild and self-limiting gastroenteritis, sepsis, meningitis, encephalitis, abortion (Farber & Peterkin 1991)	In the US, 0.26 cases per 100,000 inhabitants in 2013 (Crim et al. 2014)	Annually 1,600 illnesses and 260 deaths in the US (Scallan et al. 2011)	Milk, dairy products, meat, eggs, seafood, and ready-to-eat food (Farber & Peterkin 1991)

detect outbreaks, conduct surveillance, and understand or elucidate the epidemiology of certain types or clones. Thus, bacterial typing techniques have been developed to measure genetic relatedness among emerging pathogenic strains, clones, or clusters of bacteria from a single species.

In the 1920s, the beginning of the bacterial typing era, typing systems were based solely on phenotypic methods such as serotyping (Grimont & Weill 2007), phage typing (Smith 1951, Ward et al. 1987), and antibiogram typing. These early subtyping methods are currently still in use in many countries as part of national surveillance programs despite the availability of molecular methods and the rapid advancement of WGS technologies (Didelot et al. 2012). In addition, these early subtyping methods have proven not to have the sufficient discriminatory power in outbreak situations and need to be supplemented with molecular techniques such as PFGE, multiple-loci variable-number tandem repeat (VNTR) analysis (MLVA), or similar (see sidebar, Common Molecular Subtyping Methods in Public Health Surveillance). The number of molecular subtyping methods and their discriminatory power has increased, and several methods have been implemented to meet the public health needs, the latest being the use of WGS (Sabat et al. 2013).

COMMON MOLECULAR SUBTYPING METHODS IN PUBLIC HEALTH SURVEILLANCE

PFGE is an adaptation of conventional agarose gel electrophoresis that allows extremely large DNA fragments to be resolved due to the periodic alternation of the angle of an electric field (van Belkum et al. 2007). PFGE is still the most commonly used molecular subtyping method for surveillance and outbreak detections because of its high discriminating power and reproducibility (Ribot et al. 2006, Tenover et al. 1995, van Belkum et al. 2007).

MLVA is a polymerase chain reaction (PCR)-based method that measures the length of the repetitive DNA due to deletion or insertions (Lindstedt et al. 2003). Amplicons of selected loci are visualized by gel or capillary electrophoresis to estimate the size of the repeats in each locus (Vergnaud & Pourcel 2009). MLVA types are assigned according to designated schemes.

Multilocus sequence typing (MLST) builds on allele variation in the sequences of seven housekeeping genes (Maiden et al. 1998). Pathogen-specific housekeeping genes are sequenced and subsequently submitted to the PubMed-based MLST database (**www.pubmlst.org**), which assigns the identified allele code and sequence type. MLST has a lower discriminatory power than PFGE but is often used to investigate population structures or to identify specific clones or clonal complexes (Ribot et al. 2006).

Before applying a subtyping technique, it is important to understand and evaluate the strengths and weaknesses of the method. Certain criteria are normally used to assess subtyping methods such as discriminatory power, reproducibility, typeability, and repeatability. A good typing method should give the same result each time it is used in the same laboratory (repeatability) and when used in different laboratories (reproducibility). Typically, sequence-based methods are more repeatable and reproducible than DNA gel electrophoresis–based methods. The methods should also be able to assign a specific type to all tested isolates (typeability) and, furthermore, be able to assign a different type to two epidemiologically unrelated isolates sampled randomly from a population of the same bacterial species (discriminatory power) (Foxman et al. 2005, van Belkum et al. 2007). In addition, several other aspects also affect the choice of typing techniques, such as the costs, accessibility, and workload (van Belkum et al. 2007).

Subtyping-Based Foodborne Pathogen Surveillance in the United States and Europe

In the United States, several institutes independently conduct surveillance of foodborne pathogens in humans, food, and animals (Jones et al. 2007). The Centers for Disease Control and Prevention (CDC) collect data on reported laboratory-confirmed cases of human origin to, e.g., the National *Salmonella* Surveillance System (NSSS) through the Public Health Laboratory Information System (PHLIS) (Swaminathan et al. 2006a). The PHLIS database contains data from all state and territorial public health laboratories (Cent. Dis. Control Prev. 2008, Martin & Bean 1995). The US Department of Agriculture Food Safety and Inspection Service (USDA-FSIS), and the US Food and Drug Administration Center for Veterinary Medicine (FDA-CVM) collect foodborne pathogen data from food animals and retail meats, respectively (Gilbert et al. 2007).

PulseNet USA was established in 1996 as the national molecular surveillance network for foodborne infections (Gerner-Smidt et al. 2006). The aim of the network was to rapidly detect outbreaks caused by foodborne pathogens through PFGE and a national database of PFGE patterns. In 2001, full national participation by all 50 states was achieved. By 2005, the network consisted of 65 participating public health laboratories, four countries, three cities, and eight food

safety regulatory laboratories. All PFGE profiles are uploaded to the national database, where local or multistate outbreaks are investigated. In 1999, PulseNet USA harmonized protocols with PulseNet Canada, forming PulseNet International (Swaminathan et al. 2006b). PulseNet USA and PulseNet Canada have investigated numerous multinational outbreaks, showing the advantage of international collaboration in food safety surveillance. In addition to PulseNet USA and PulseNet Canada, PulseNet networks have been established in the Asia Pacific region, Latin America, Europe, and China. In Europe, PulseNet Europe started in 2001 as the Salm-gene network, which was initially meant to build a database for *Salmonella* PFGE profiles among European countries. By the end of 2004, the Salm-gene database contained approximately 20,000 profiles of primarily *Salmonella enterica* serotypes Enteritidis and Typhimurium (Swaminathan et al. 2006b). In 2004, the Salm-gene network merged with Enter-Net, which later became PulseNet Europe in 2004 (Fisher et al. 2005). Surveillance by PulseNet Europe ceased in 2007 because of a lack of funding but was resumed as the European molecular typing program for verotoxin/Shiga toxin–producing *Escherichia coli* (VTEC/STEC), *Salmonella*, and *Listeria*.

In 2003, the European Parliament and Council adopted Directive 2003/99/EC on the monitoring of zoonoses and zoonotic agents (Counc. Eur. Union 2003a, O'Brien et al. 2005), amending Decision 90/424/EEC (Counc. Eur. Union 1990) and repealing Council Directive 92/117/EEC (Counc. Eur. Union 1992). The purpose of the Directive 2003/99/EC was, among other things, to ensure a proper and harmonized surveillance of zoonoses. The directive was later supplemented with the Regulation 2160/2003 on the control of *Salmonella* and other specified foodborne zoonotic agents, which in principle should cover the whole food chain, from farm to table (e.g., laying hens, broilers, turkeys, pigs, veal calves, etc.) (Counc. Eur. Union 2003b, O'Brien et al. 2005). The surveillance should take place on a harmonized basis by all member states according to the detailed rules laid down in the regulation. The data are collected and published annually by the European Food Safety Authority (EFSA) (Eur. Food Saf. Auth. 2010), including human data collected by the European Center for Disease Prevention and Control (ECDC) and through the European Surveillance System. This surveillance has also expanded to include the European molecular typing program for VTEC, *Salmonella*, and *Listeria* through a mandate from the European Commission to provide technical support to the development of a database on molecular typing data, e.g., PFGE on isolates of *Salmonella*, *Listeria monocytogenes*, and Shiga toxin–producing *E. coli* from food, feed, animals, and the related environment (Eur. Food Saf. Auth. 2014). The European molecular typing program is a joint initiative between EFSA and ECDC in which ECDC provides data submitted through, among other sources, The European Surveillance System (TESSy). ECDC collects, analyzes, and disseminates predominantly case-based surveillance data collected at the European level. It retrieves data from all EU countries on a number of communicable diseases, including foodborne infections as described in Decision Number 2119/98/EC (Eur. Cent. Dis. Prev. Control 2013).

GENOMIC EPIDEMIOLOGY: TRANSFORMATION OF PUBLIC HEALTH

Recent drastic advances in sequencing technologies and bioinformatics tools have made WGS a viable and advanced solution for epidemiologic investigation and surveillance of foodborne bacterial pathogens. The term genomic epidemiology has been increasingly used to describe the practice of utilizing WGS to access, index, and analyze DNA sequence features of epidemiologic importance. Genomic elements that vary in rates of mutation in bacterial evolution supply ample targets for epidemiologic investigations at different temporal and geographical scales. Genetic determinants of certain phenotypes, such as serotype and antimicrobial resistance, provide the possibility of in silico prediction of clinically important phenotypic traits. Because whole-genome

sequences contain the entirety of the genetic information of an organism, of which DNA markers of any sort become mere subsets, WGS promises a comprehensive platform for public health microbiology that provides a one-stop shop for various subtyping and characterization targets. For example, WGS can both help solve outbreaks by affording high discriminatory power in differentiating closely related isolates and assist in tracking epidemiologic trends by monitoring antimicrobial resistance. WGS also comes with backward compatibility and future extensibility to any existing and potential typing schemes as the complete set of DNA variations across entire genomes are made available for analysis.

Applications: Case Studies

The first application of WGS in support of a foodborne outbreak investigation was reported in 2010 (Gilmour et al. 2010). The 2008 multiprovince outbreak of listeriosis was the deadliest foodborne outbreak in Canada; its investigation involved the WGS of two outbreak-associated isolates that displayed similar but distinct PFGE patterns. WGS allowed detailed genetic comparison of the isolates, leading to the conclusion that multiple distinct but highly related strains may have been involved in this large outbreak. This study demonstrated that WGS could be applied in high priority public health events and suggested that this technology would inevitably find its use in public health.

Since then, WGS has been frequently implemented in both outbreak investigation and routine surveillance of foodborne pathogens at various geographic scales (e.g., local, national, and international) and in different temporal scenarios (e.g., retrospective and prospective).

The first significant use of WGS that led to the source identification of a foodborne outbreak and regulatory decision making was reported in 2011 (Lienau et al. 2011). A multistate outbreak of *Salmonella enterica* serotype Montevideo infections occurred between July 2009 and May 2010 (**http://www.cdc.gov/salmonella/montevideo/**). Nearly 300 illnesses were reported from 44 states and the District of Columbia. Epidemiologic information, including shopper card records and laboratory testing of products and ingredients, suggested black and red peppers used to prepare salami were the source of the illness. Subsequent investigation into the supply chains of the implicated spices established the link between patients, the food product manufacturer, and the ingredient supplier, which was confirmed by WGS of 35 strains. (Lienau et al. 2011).

A strain of Shiga toxin–producing *E. coli* O104:H4 caused one of the largest foodborne outbreaks of recent history in the summer of 2011 in Germany. More than 3,000 cases of infection were reported with more than 40 deaths. Associated cases, mostly from travelers returning from Germany, were also reported in more than a dozen European and North American countries. The investigation of this outbreak marked a quick and international mobilization of WGS efforts that resulted in real-time and open-source release of sequencing data from multiple sequencing platforms and crowd-sourced analyses by groups in multiple countries (Grad et al. 2012, Klumpp et al. 2012, Mellmann et al. 2011, Rohde et al. 2011). Within a week, WGS revealed that the unusually virulent outbreak strain belonged to a distinct group of enteroaggregative *E. coli* that had acquired the combination of genes for Shiga toxin 2 gene, antimicrobial resistance, and other virulence factors.

S. enterica serotype Enteritidis, especially the PFGE pattern JEGX01.004, is frequently associated with poultry and egg production, but its evolutionary history and population structure have been only recently studied by WGS (Allard et al. 2013, Deng et al. 2014, Zheng et al. 2014). At the state level in the United States, den Bakker et al. (2014) demonstrated the practicality of implementing WGS for outbreak surveillance of this pathogen in a state public health laboratory. By conducting both retrospective and prospective studies based on WGS, the investigators were

able to attribute additional putative cases to confirmed outbreaks, some of which would have been excluded from the outbreak by PFGE because of different PFGE types than the outbreak strain, and detect additional putative outbreak clusters.

Also in the United States, WGS-powered routine and comprehensive surveillance of a foodborne pathogen has been explored in a pilot study through collaboration among the CDC, FDA, USDA, the National Center for Biotechnology Information, and state and local public health departments (**http://www.cdc.gov/listeria/pdf/whole-genome-sequencing-and-listeria-508c.pdf**). Beginning in 2013, the *Listeria* WGS Project sought to sequence every clinical, food, and environmental isolate of *L. monocytogenes* collected in the United States in real-time. By combining WGS and epidemiologic data and building upon existing surveillance infrastructure such as PulseNet, the first year of the project has seen enhanced surveillance of *L. monocytogenes*, as evidenced by an increased number of solved outbreaks, improved detection of listeriosis clusters, and reduced average size of detected clusters.

Joensen et al. (2014) reported a comparative study between the current and WGS-based workflows under the setting of routine and real-time surveillance of VTEC/STEC in Denmark. During a 7-week period in the fall of 2012, all suspected VTEC isolates collected in Denmark (n = 46) were sent to the Statens Serum Institut and subjected, in parallel, both to the current procedure of pathogen subtyping and to identification consisting of a variety of techniques and WGS coupled with analyses by a suite of publicly available web tools (**www.genomicepidemiology.org**), including KmerFinder, multilocus sequence typing (MLST) (Larsen et al. 2012), SeroTypeFinder (Joensen et al. 2015), VirulenceFinder, snpTree (Leekitcharoenphon et al. 2012), and NDtree. The results revealed a clear concordance between the two approaches in terms of strain characterization, but with WGS ultimately providing a higher resolution for subtyping and additional information such as an overall real-time clustering of isolates compared to the conventional approach. Furthermore, NGS proved to be much quicker, reducing the entire workflow to four days and at a lower cost than the conventional approach.

Sequencing Technologies

The past decade has seen the introduction of several high-throughput sequencing technologies, referred to as NGS technologies. Prior to this, automated Sanger technology was dominant (van Dijk et al. 2014). The big advantage of NGS technologies is that they produce massive amounts of data at greatly reduced per base-pair sequencing costs, making it possible to sequence complete microbial genomes at a price comparable to traditional subtyping methods such as PFGE or MLST. Although several technologies have been commercially available (see van Dijk et al. 2014 for a review), three technologies dominate the field of genomic epidemiology currently: (*a*) Illumina sequencing, (*b*) Ion Torrent sequencing, and (*c*) Pacific Biosciences sequencing. Both Illumina and Ion Torrent can be classified as short-read sequence technologies, producing large numbers of short (up to 300 bp for Illumina and 400 bp for Ion Torrent) reads. The Illumina platform relies on a sequencing-by-synthesis–based technology, using reversible (fluorescent) terminators, whereas the Ion Torrent technology relies on so-called semiconductor sequencing. The Ion Torrent is prone to homopolymeric tract length-related errors (Quail et al. 2012), and although reads generated on any of the Illumina sequencers are characterized by a high accuracy, they are not free from errors and may display more subtle sequence motif–related errors (Meacham et al. 2011, Ross et al. 2013). Pacific Biosciences sequencing relies on a single-molecule real-time (SMRT) sequencing technology. This sequencing platform produces long reads with relatively low per-read accuracy; however, consensus accuracy is high (Koren & Phillippy 2015). In addition

to base composition of sequences, base modifications (e.g., methylation) can be inferred (Clark et al. 2012) from data generated with SMRT sequencing.

Bioinformatics

The main objectives of bioinformatics in surveillance and outbreak investigation are to (*a*) infer evolutionary relationships between strains of pathogens (phylogenetics) and to (*b*) infer epidemiologically relevant phenotypic characteristics based on genomic data.

Currently most bioinformatics approaches rely on one or more of the following principles: (*a*) reference-based read mapping, (*b*) de novo assembly, or (*c*) de novo variant callers.

Reference-based read-mapping approaches. In reference sequence–based approaches, reads obtained from NGS are mapped to a reference genome, preferably a high quality, finished genome sequence. The most commonly used read-mapping algorithms for Illumina data are Burrows-Wheeler aligner (BWA) (Li & Durbin 2009) and Bowtie 2 (Langmead & Salzberg 2012). Both algorithms rely on a Burrows-Wheeler transform-based algorithm to index the reference genome, followed by alignment of the sequence reads using a relaxed, quality-aware algorithm (Langmead & Salzberg 2012, Li & Durbin 2009). In principle, mapped reads can be used to infer patterns of gene presence and absence as compared to the reference strain, as regions with no or sparse coverage are likely to be absent from the query strain. Reference-based read-mapping methods are most commonly used as a basis for the discovery of genomic variants [e.g., single nucleotide polymorphisms (SNPs) and small insertions and deletions (indels)], using so-called variant caller algorithms based on frequentist [e.g., Varscan (Koboldt et al. 2012)] and Bayesian [e.g., Freebayes (Garrison & Marth 2012), GATK (McKenna et al. 2010, van der Auwera et al. 2013)] frameworks. Generally, pipelines used in a surveillance and outbreak investigation focus on SNPs (den Bakker et al. 2014, Deng et al. 2014, Stasiewicz et al. 2015), as indels can be less confidently termed (Neuman et al. 2013). The reliance on a reference has disadvantages; only genomic regions that are present in the reference sequence can be queried for putative genomic variants, and highly divergent genomic regions are also missing from the comparison, as sequence divergence causes multiple misalignments, which may exceed the number of mismatches allowed by an algorithm for mapping a read. Major advantages of reference-based methods are that they are computationally relatively inexpensive and therefore can be fast and do not require large investments in powerful computers.

De novo approaches. De novo variant detection methods can be subdivided into two classes of methods: (*a*) those relying on de novo–assembled whole-genome sequences and (*b*) truly de novo–variant discovery methods, which require no or minimal assembly of the original reads. Most de novo short-read assemblers are based on one of two assembly strategies: overlap, layout, consensus (OLC) and de Bruijn graph. OLC-based methods have been traditionally used for Sanger sequencing data sets and are currently applied in assembly software such as Celera (Myers et al. 2000), Edena (Hernandez et al. 2008), and MIRA (Chevreux 2005). Of importance to current surveillance and outbreak investigation efforts are Celera and MIRA, as these assemblers are used in de novo assembly pipelines of PacBio (Koren & Phillippy 2015) and Ion Torrent (Baez-Ortega et al. 2015) data, respectively. Most short read de novo assemblers are based on de Bruijn graph strategies. Central to these approaches is the breakup of all reads into so-called k-mers of a length k, where k is less than the read length (e.g., a 100-bp read can be broken up into 46 overlapping 55-mers), after which a graph is constructed of perfectly overlapping k-mers. In a perfect world without

sequence errors and genomic complexity, this graph would form a perfect circle representing a bacterial chromosome; however, when sequence errors occur or repeat regions are encountered, the graph becomes more complex, displaying, for example, bubbles (caused by sequence errors in the middle of the read), spurs (caused by errors at either end of a read), and cycles (caused by repetitive sequence) in the graph (see Compeau et al. 2011, Miller et al. 2010 for an excellent introduction and explanation of de Bruijn graph–based assembly algorithms). A critical component of using de Bruijn graph–based assembly is the choice of a suitable k-mer for the assembly (see Zerbino 2010 for an explanation of the trade-offs between longer or shorter k-mer lengths), a problem that is circumvented by integrating multiple assemblies using a series of k-mer sizes into a final assembly in more recent de Bruijn graph–based assembly pipelines such as SPAdes (Bankevich et al. 2012). De novo assemblies can be used as reference sequences for reference-based methods, comparative genomics, and variant detection in tools such as Parsnp (Treangen et al. 2014), but they are not as accurate as reference-based variant calling pipelines (Olson et al. 2015).

Reference-free de novo variant callers. Reference-free de novo genomic variant callers are developed to infer genomic variants from raw, unassembled reads, based on k-mers, as in kSNP (Gardner & Hall 2013, Gardner et al. 2015, Timme et al. 2013), or from reads assembled into de Bruijn graphs, as implemented into Cortex_var (Iqbal et al. 2012, 2013). kSNP takes both assembled genomes and raw reads as input and through a k-mer–based process employing the Jellyfish algorithm (Marçais & Kingsford 2011) finds SNPs, creates SNP matrices, and performs additional phylogenetic analyses. When names of finished genomes are provided, kSNP downloads the GenBank annotations of these genomes and attempts to annotate the SNPs found during the reference-free SNP calling procedure (Gardner et al. 2015). Along with SNPs, Cortex_var can call indel and structural variants (e.g., inversions) (Iqbal et al. 2013) and additionally can be used to infer the presence or absence of genes or genomic regions (e.g., prophages, plasmids), as demonstrated in den Bakker et al. (2014) and Stasiewicz et al. (2015).

Whole-Genome-Sequencing–Based Subtyping Tools

Several pipelines or tools have been developed to enable the inference of traditional subtypes (e.g., MLST types, serotypes, etc.), using assembled genomes or unassembled sequence reads as input. The web-based tools of the Center for Genomic Epidemiology (**http://www.genomicepidemiology.org/**) are noteworthy and include, among others, tools to infer MLST types (Larsen et al. 2012) and *E. coli* serotypes (SerotypeFinder; **https://cge.cbs.dtu.dk/services/SerotypeFinder/**). Zhang et al. (2015) developed a tool, SeqSero, which has proven to be highly accurate in predicting traditional serotypes from raw WGS sequence data obtained from pure *Salmonella* cultures and also from murine fecal metagenomic data containing *Salmonella* DNA. The input for this tool can be either raw sequence reads or assembled genomes. SRST2 (Inouye et al. 2014) is a read mapping–based tool for detection of genes, alleles, and multilocus sequence types from WGS data. Although this tool comes with preformatted databases and scripts to conveniently download and format species-specific databases from the MLST database (**http://www.mlst.net/databases/**), custom databases for genes of interest can be easily generated. Recent studies on several species revealed that WGS can accurately predict antimicrobial resistance (Gordon et al. 2014, Tyson et al. 2015).

Phylogenetic Analyses

Although phylogenetic approaches have traditionally been used in macroevolutionary studies, these approaches are increasingly being used to reconstruct the evolutionary history of bacterial populations from their genome sequences (e.g., Zhou et al. 2013). In particular, when the bacterial population of interest for outbreak detection or surveillance is assumed to be largely clonal (i.e., no or very little recombination within the population), the population structure can be expected to follow the structure of a phylogenetic tree (Holmes et al. 1999, Smith et al. 1993). In most studies, the input for the phylogenetic analysis is an SNP matrix, a matrix consisting of only the variable SNPs in the genome sequence, minus regions that occur in a limited number of strains or regions that have been determined to be subject to homologous recombination. Algorithms for tree inference can be subdivided into four different categories: parsimony methods, maximum-likelihood methods, Bayesian methods, and distance methods. Parsimony-based algorithms attempt to find the tree that requires the fewest evolutionary changes (e.g., SNP changes), maximum-likelihood algorithms attempt to find the tree that is most likely to have produced the observed data (given a model of molecular evolution), Bayesian methods find the tree (or parts of a tree) that has the highest posterior probability given the observed data, and distance methods seek the tree that best represents the observed distances (e.g., SNP differences) or the tree whose sum of branch lengths is the minimum. Maximum-likelihood, Bayesian, and parsimony methods are most commonly applied in WGS outbreak investigation studies. Frequently used types of software are the maximum-likelihood methods, including MEGA6 (maximum-likelihood, parsimony, and distance methods) (Tamura et al. 2013), GARLI (Bazinet et al. 2014), RaXML (Stamatakis 2006), FastTree (rapid maximum-likelihood methods) (Price et al. 2010), and BEAST (Drummond et al. 2012) for Bayesian analyses.

Although homologous recombination may, in theory, impair accurate phylogenetic reconstruction (Didelot et al. 2012), Hedge & Wilson (2014) determined that the topology of phylogenetic trees is quite robust in the influence of recombination, whereas branch length is not. Methods, such as ClonalFrameML (Didelot & Wilson 2015), have been developed to simultaneously detect recombination in bacterial genomes and account for this in phylogenetic reconstruction.

Initiatives and Surveillance Networks

In September 2011, the Technical University of Denmark (DTU) and the FDA organized a scientific meeting in Brussels, Belgium, to discuss the possibility of using WGS as the basis for a global microbial surveillance system for conducting national or regional surveillance by public health agencies and practitioners, including front-line diagnostic laboratories and hospitals (Kupferschmidt 2011). It became evident that several preconditions were met for a possible successful initiation, such as WGS being a realistic alternative to current molecular typing techniques, the dramatic decline in sequencing costs, the rapid development of bioinformatics tools, the expansion of fast broadband Internet, and a sincere desire to build a global One Health (Inst. Med. 2012) system to facilitate infectious disease detection and surveillance. This meeting was later referred to as the first international Global Microbial Identifier (GMI) meeting (Glob. Microb. Identif. 2011) and set the vision of GMI: a world in which high quality microbiological genomic information from human, food, animal, and plant domains is shared globally to improve (public) health and healthcare and provide a healthy environment for all. Since this meeting, nine meetings have been held (the latest in May 2015 in Beijing, China) to ensure the momentum of the initiative and share information about its deliverables. Today, GMI has grown into a global network of scientists, experts, and national and regional institutions and organizations, including WHO,

the United Nations Food and Agricultural Organization (FAO), and the World Organisation for Animal Health (OIE); all are committed to improving global infectious disease surveillance using WGS.

The GenomeTrakr network established by the FDA (**http://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/ucm363134.htm**) is the first open-source network of its kind that utilizes WGS for pathogen identification. As of August 2015, GenomeTrakr participants included 14 federal laboratories, 14 state public health and academic laboratories, and 3 laboratories outside the United States. These laboratories collect WGS and metadata of foodborne pathogens, including their geographic origins, and share them via publicly accessible databases housed at the National Center for Biotechnology Information (NCBI) (**http://www.ncbi.nlm.nih.gov/bioproject/?term=genometrakr**). Fast turnaround of isolate sampling, genome sequencing, and data curation and analysis across this nationally distributed network has revealed the possibility of WGS-powered, real-time food safety applications, such as the aforementioned *Listeria* WGS Project in partnership with other federal and state agencies and the real-time surveillance of *S. enterica* serotype Enteritidis in Minnesota, Washington, and New York. As of August 2015, GenomeTrakr has sequenced and made publicly available more than 12,900 *Salmonella* isolates, more than 3,200 *L. monocytogenes* isolates, and hundreds of isolates of other organisms, including *E. coli*, *Shigella*, *Cronobacter*, and *Campylobacter*.

To initiate the utility of WGS in the area of food safety and security, the 100K Foodborne Pathogen Genome Sequencing Project was founded by the University of California-Davis School of Veterinary Medicine, Agilent Technologies, and the FDA in collaboration with partners around the globe to sequence 100,000 genomes of the world's most important foodborne pathogens (**http://100kgenome.vetmed.ucdavis.edu/**). Subsequently, all genome sequences produced will be deposited at NCBI to support the development of new diagnostic tools.

Envisioned by the European Union's Horizon 2020 research and innovation program, the COMPARE [Collaborative Management Platform for Detection and Analyses of (Re-) Emerging and Foodborne Outbreaks in Europe] project aims to establish a global, web-based, open platform for improving rapid identification, containment, and mitigation of emerging infectious diseases and foodborne disease outbreaks across sectors and domains (public health, veterinary health, food safety) (**http://www.compare-europe.eu/People**). As a multidisciplinary research network of 29 European participants from 10 European Union countries and Australia, COMPARE consists of partners from the public health, veterinary health, and food safety sectors, with expertise in WGS, big data analysis and storage, and outbreak detection and response. Examples of core activities within the COMPARE project include development of risk assessment models, risk-based strategies for sample and data collection, harmonized standards for sample processing and sequencing, analytical workflows for generating actionable information, an information sharing platform, and risk communication tools.

In Denmark, The Center for Genomic Epidemiology (CGE) was established in 2010 with the goal of achieving the proof-of-concept for WGS-based infectious disease surveillance and providing web-based solutions toward straightforward characterization of pathogenic and commensal bacterial isolates in real-time by combining bioinformatics and epidemiology (**http://www.genomicepidemiology.org/**). This has enabled novice users and frontline diagnostic laboratories to utilize the benefits of WGS to identify species, antimicrobial resistance, and virulence genes as well as to determine similarity to existing genomes through plain language reports. To date, an abundance of those easy-to-use, free-of-charge, and web-based bioinformatics tools have been developed to replace many conventional public health laboratory workflows for most foodborne pathogens, including *Salmonella*, *E. coli*, and others. The panel of bioinformatics tools developed by CGE includes KmerFinder (species identification), MLST, PlasmidFinder,

VirulenceFinder, ResFinder (detection of antibiotic resistance genes), SeroType Finder, and a variety of phylogenetic tools (**https://cge.cbs.dtu.dk//services/all.php**) (Carattoli et al. 2014; Joensen et al. 2014, 2015; Larsen et al. 2012; Zankari et al. 2012).

OPPORTUNITIES AND CHALLENGES

The transformative potential as well as existing bottlenecks in applying WGS in public health microbiology have been widely recognized and, in multiple cases, well demonstrated (Grad & Lipsitch 2014, Köser et al. 2012, Lipkin 2013, Robinson et al. 2013). Although some common issues apply to infectious disease surveillance in general, characteristics of foodborne illness epidemiology present specific opportunities and challenges for both the integration of WGS into established systems and the exploration of new areas of application. The ubiquitous presence of microbial foodborne pathogens in natural environments and their ability to contaminate the food supply lead to occasional outbreaks and steady numbers of sporadic cases. Effective surveillance therefore requires both robust detection of outbreaks from baseline levels of sporadic cases and effective monitoring of those sporadic cases. Although it is often the surveillance of clinical cases that leads to outbreak detection and identification of food vehicles, the prospect of extending WGS-powered surveillance to food production and distribution may create early warnings and prompt more rapid removal of contaminated products before outbreaks develop.

Transforming Foodborne Disease Surveillance by Routinely Applying Real-Time Whole-Genome Sequencing

WGS in routine and real-time surveillance further and substantially improves the efficacy of outbreak detection and investigation. Transformation of foodborne illness surveillance often comes with the adoption of improved subtyping methods. In the 1960s, routine *Salmonella* serotyping allowed outbreaks of particular serotypes to be identified from the background of all salmonellosis cases. Since 1996, the discriminatory power afforded by PFGE has allowed PulseNet to single out different outbreaks within individual *Salmonella* serotypes. Nevertheless, monomorphic pathogens that are genetically highly homogenous can still defy the resolution of PFGE and other subtyping methods. For example, approximately 45% of *S. enterica* serotype Enteritidis isolates received by PulseNet display the same PFGE pattern using XbaI, rendering PFGE ineffective in some investigations (Deng et al. 2015).

With the ultimate subtyping resolution delivered by interrogating entire genomes, WGS is able to help delineate disease clusters by distinguishing isolates that are descended from common progenitors at multiple evolutionary scales even within the course of an outbreak, which can lead to flexible, sensitive, and specific case definitions. When infections caused by isolates of a common subtype first emerge, they can be lumped together with sporadic cases and remain below an empirical baseline to trigger an investigation of a cluster of cases until more cases are reported in the outbreak. With routine and real-time surveillance powered by WGS, more outbreaks will likely be detected and solved at an earlier stage, when the number of outbreak-associated cases is still small. This will likely increase the demand for more efficient and effective exposure assessments and cluster investigations, as more clusters will be identified. In this sense, traditional shoe-leather epidemiology will benefit from rather than be replaced by WGS, and valuable public health resources can be better allocated to focus on serious and imminent public health threats through more precise and timely outbreak detection powered by WGS.

WGS can also lead to more efficient surveillance of sporadic cases, where no clear link to an outbreak can be established. During the multiagency initiative using WGS for routine and near

real-time surveillance of *L. monocytogenes* in the United States, a single case of listeriosis in Ohio was linked to recalled bagged lettuce (**http://www.cdc.gov/amd/stories/listeria.html**). Even though no definitive evidence that the patient had consumed the recalled product had ever been retrieved, WGS-based analysis found that the isolate from the patient matched closely with that from the recalled lettuce. This provided informative and actionable epidemiologic intelligence that led to the listing of lettuce, a likely new vehicle for *Listeria* infection, on the questionnaires for exposure assessment of *Listeria* infection.

Sporadic cases are less likely to be investigated even though they compose much of the foodborne illness burden and shape the overall foodborne illness landscape. An established surveillance network for sporadic foodborne disease such as FoodNet plays an important role in attributing illness to specific foods and settings, and in monitoring trends of specific illnesses (Allos et al. 2004), both of which can be enhanced by WGS. Prioritization of prevention and control strategies for major foodborne pathogens requires the identification of their primary food sources, which is challenging for sporadic cases. Among common source attribution approaches, subtyping such as serotyping and phage typing (Barco et al. 2013) is often most frequently used (Pires et al. 2014). Subtyping schemes for this purpose should provide a strong association with specific sources or reservoirs as well as appropriate discrimination to differentiate isolates while grouping those attributable to the same source (Hald et al. 2004). WGS enables robust phylogenetic reconstruction and population dissection to reveal lineages associated with particular ecological sources. Access to any sequence marker in the genomes can then allow flexible selection of subtyping targets for developing source attribution models. Although no WGS-based source attribution study has been reported in the literature as of this review, it represents a promising direction for future research, especially as monitoring networks such as the FDA GenomeTrakr are creating large amounts of publicly available WGS data from various food and environmental sources (**http://www.ncbi.nlm.nih.gov/bioproject/183844**).

Systematic WGS of isolates from sporadic cases can complement similar efforts in outbreak investigations. Shifting trends of foodborne disease, such as emerging pathogens and rapidly spreading clones, can be revealed and monitored by WGS using phylogenomic and phylogeographic approaches (Mather et al. 2013, Okoro et al. 2012). The prospect of inferring important phenotypes such as virulence potential (Joensen et al. 2014) and antibiotic resistance (Gordon et al. 2014, Tyson et al. 2015) from WGS will lead to more efficient surveillance of foodborne illness. It can be anticipated that by bridging the data gaps between the epidemiology of outbreaks and sporadic cases, surveillance efforts can be better targeted toward important and emerging pathogens.

Whole-Genome Sequencing in Food Safety Surveillance

Traceback of foodborne pathogens is often challenged by the presence of multiple ingredients in a product and by modern supply chains that can source ingredients from a wide geographic range, including intercontinental trade. The use of WGS in combination with epidemiologic evidence has demonstrated its promising utility in pinpointing the source of foodborne pathogen contamination. As mentioned above, during the investigation of the 2009–2010 multistate outbreak of *S. enterica* serotype Montevideo (Lienau et al. 2011) infections, the resolution of WGS confirmed the epidemiologic evidence and allowed the exclusion of pistachios as a confounding food vehicle of the outbreak strains. During a previous outbreak, strains of the same PFGE pattern had been isolated from pistachios held in the same facility that processed the contaminated spice rub. In another example of a 2012 multistate *Salmonella* outbreak, strains of serotype Bareilly and Nchanga were identified from a raw scraped ground tuna product that caused more than 400 illnesses in 28 states and the District of Columbia (**http://www.cdc.gov/salmonella/bareilly-04-12/**). Although

epidemiologic investigation linked the outbreak to a domestic supplier, WGS provided further evidence to trace the original source of the contaminated ingredient to a precise overseas location (**http://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/ucm422075.htm**) (Hoffmann et al. 2016). Genomes of the outbreak isolates were found to be closely related to the *S. enterica* serotype Bareilly DNA isolated from shrimp in southwest India several years earlier. The plant that processed the contaminated shrimp was five miles away from the plant that processed the scraped ground tuna.

Although efforts have largely focused on the surveillance of clinical cases as the consequences of consuming contaminated foods, WGS also has the potential to monitor the contaminated foods themselves, leading to earlier or even preventive interventions. When gathering baseline data of foodborne pathogen occurrence in nut butter–processing facilities in 2014, the FDA obtained isolates of *S. enterica* serotype Braenderup that displayed the same PFGE pattern as a small number of clinical isolates from the previous few months. Before the epidemiologic link could be established between the few cases and the nut butter product from the facility, investigation by WGS suggested that all the processing facility environmental and clinical isolates were essentially the same strain. Soon after the investigation, related products were recalled for potential *Salmonella* contamination.

It can be predicted that the application of WGS of foodborne pathogens by public health and regulatory agencies will have a major impact on the food industry. Although the prospect of more outbreaks being detected in a timely and precise manner may be viewed as a challenge, the food industry can potentially benefit from the application of WGS in various situations. High-resolution subtyping by WGS can help pinpoint the source of contamination and reduce the possibility of false implication due to indistinguishable subtypes determined by lower resolution methods. Similar principles and methodologies for WGS-powered pathogen traceback in outbreak investigations may be applied in industrial settings, especially when the modes of entry, transmission, and persistence of pathogens in food production and distribution systems remain unclear. For example, in integrated poultry production systems, where the entire chain of operation is coordinated by a single company, a fine-grained understanding of *Salmonella* movement, whether vertical transmission from upstream to downstream facilities or horizontal introduction from a farm's surroundings (Liljebjelke et al. 2005), is still missing. Differentiation between indigenous (e.g., persistent strains) and imported (e.g., ingredients from suppliers) contamination can be achieved by WGS to help design targeted intervention or control strategies. For instance, WGS has provided crucial evidence to identify long-term persistence of individual strains of *L. monocytogenes* (Orsi et al. 2008) and *S. enterica* serotype Agona (**http://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/ucm422075.htm**) at respective food processing plants for more than a decade. In the *Salmonella* case, WGS coupled with the facility's renovation record traced the contamination to the location that likely harbored the strain and revealed a likely and unexpected food safety gap during the facility renovation.

Challenges

Routine and widespread application of WGS in foodborne pathogen surveillance requires standardized methods that can be practiced among laboratories. Rapid evolution of sequencing technologies and their analytical roots in bioinformatics make standardization a particular challenge. Comparative studies have been frequently performed to evaluate major and newly developed sequencers (Loman et al. 2012, Quail et al. 2012). Although Illumina platforms have become the de facto standard for foodborne illness diagnosis and surveillance, new-generation technologies are emerging such as the nanopore-based, thumb drive–sized MinIon device, whose utility in outbreak investigation has been recently demonstrated (Quick et al. 2015). Compared

with the evaluation of sequencing technologies for which explicit quality attributes are available, benchmarking of bioinformatics tools for WGS data analysis can be more challenging. For general analyses such as de novo genome assembly, read mapping, and variant calling, multiple tools exist whereby no single tool outperforms others with all data sets (Earl et al. 2011). Further complexity can be generated when several software tools are combined into a workflow or pipeline—which is commonly practiced in public health laboratories—that turns WGS data into epidemiologically relevant results such as phylogenetic trees. For example, the CDC and FDA both have an individually developed pipeline for SNP-based subtyping and phylogenetic analysis (Katz et al. 2013, Pettengill et al. 2014).

A particular debate regarding the standard bioinformatics approach for WGS-based subtyping of foodborne pathogens has been focused on two methods: whole-genome SNP typing (WGST) and whole-genome MLST (wgMLST). WGST has been the dominant method for phylogenetic analysis of WGS data. As stable phylogenetic markers (Keim et al. 2004), SNPs can be identified from whole genomes to allow robust evolutionary analysis and high-resolution subtyping. However, WGST does not produce easily communicable nomenclatures. The set of polymorphic sites for subtyping, usually SNPs located in conserved parts of the genomes (i.e., core genome) to be analyzed, is subject to change when different groups of strains are investigated, making definitive naming of subtypes difficult. As standard nomenclature is important for coordinated surveillance, wgMLST has been recently proposed as a whole-genome-scale upgrade from the well-established MLST system (Cody et al. 2013). A core genome (i.e., genes conserved among all strains of, typically, a bacterial species) version of wgMLST (core genome MLST or cgMLST) has been developed for *L. monocytogenes* (Ruppitsch et al. 2015). In comparison with WGST, wgMLST allows definitive naming of sequence types by querying against a precompiled database of alleles. However, wgMLST may not be as discriminatory as WGST because intergenic regions with informative mutations are typically excluded from allele databases. Also, wgMLST can lead to simplification of allelic differences as multiple mutations can be collapsed into a single sequence type. Despite the shortcomings and the fact that allele databases for major foodborne pathogens are still being developed, it is anticipated that wgMLST will become the routine and primary method for WGS-powered pathogen surveillance in the United States (P. Gerner-Smidt, personal communication).

In addition to the selection of analytical tools, epidemiologic interpretation of WGS data often relies on empirical evidence as well. The superior discriminatory power of WGS, under which two isolates are rarely identical across entire genomes, often reveals a continuum of genetic variations among isolates instead of classifying them categorically into a few discreet subtypes. Therefore, empirically defined outbreak clusters are often used to ascribe isolates to a point-source contamination, taking into consideration both epidemiologic information (e.g., isolates sampled during the same outbreak) and phylogenetic evidence (e.g., isolates clustered in a well-supported clade) (den Bakker et al. 2011, Deng et al. 2014). Recent studies have begun to develop quantitative and model-based approaches that use SNP distance as an indicator of epidemiologic linkage in both outbreak investigation (Octavia et al. 2015) and food safety surveillance (Stasiewicz et al. 2015). The former study modeled the mutation process during an outbreak based on known rates of *S. enterica* serotype Typhimurium mutation and derived an SNP distance cutoff value as supplementary evidence for case definition.

Advances have been made to address such issues. Principles and guidelines have been proposed regarding the quality assurance of WGS-based testing in clinical laboratories (Gargis et al. 2012). Considered to be a landmark, the FDA authorized the first sequencer for clinical application (Collins & Hamburg 2013) in 2013. Very recently, the College of American Pathologists (CAP) published laboratory standards for NGS clinical tests (Aziz et al. 2015). In moving toward the

interlaboratory standardization and harmonization of using WGS for microbial pathogen surveillance, the GMI initiative has been organizing pilot proficiency tests to review the entire laboratory workflow from wet-lab WGS to bioinformatics analysis and epidemiologic interpretation of WGS data (**http://www.globalmicrobialidentifier.org/Workgroups/About-the-GMI-Proficiency-Test-2015**).

Besides operational challenges, WGS also brings about ethical and legal issues for, for example, outbreak management (Rump et al. 2013). WGS-powered foodborne illness surveillance and outbreak investigation have mainly been public health endeavors driven by government agencies. The current model of collecting isolates, generation of WGS data, and public archiving of the data along with metainformation may not directly apply to the food industry, whose embracement of the technology will likely be beneficial to both itself and public health. Emerging debate on how to use WGS data for pathogen traceback and source attribution may have an impact on industry involvement in public health surveillance networks. For example, the degree to which a genome sequence match between isolates is used as evidence to establish an epidemiologic link, implicate a business, and pursue regulatory action, especially when definitive epidemiologic evidence is missing, will likely affect industry's perception of this technology in the context of liability.

Another imposing challenge for infectious disease surveillance is the growing popularity of culture-independent diagnostics (Cronquist et al. 2012). Surveillance networks are facing an increasing risk of losing the opportunity to obtain cultures as clinical laboratories adopt culture-independent methods. Without isolates to perform subtyping and other assays such as antimicrobial susceptibility testing, the ability to conduct surveillance and outbreak investigation is compromised. Metagenomics is becoming a potential solution to this challenge by characterizing the entirety of genetic materials directly from a specimen with the purpose of extracting as much information about specific pathogens (Loman et al. 2013, Wilson et al. 2014). Although a discussion of technical gaps and ethical issues (e.g., concomitant sequencing and revealing of patient DNA, respectively) of metagenomics diagnostics are beyond the scope of this review, it is worth noting that metagenomics analysis requires comprehensive reference databases of pathogen genomes. From this perspective, ongoing efforts in WGS of major foodborne pathogens are laying the foundation for future application of metagenomics in public health.

SUMMARY POINTS

1. WGS promises the integration of various individual pathogen subtyping and identification workflows commonly practiced in public health laboratories into a comprehensive and cost-effective platform that can enable routine and real-time surveillance of foodborne diseases.
2. WGS tools have matured to the point that they can provide meaningful, low-cost, and timely data for surveillance and outbreak investigation of foodborne pathogens, including genotypic monitoring of antimicrobial resistance.
3. Rapidly evolving technologies and burgeoning initiatives contribute to a fluid and dynamic state of early adoption of WGS in public health laboratories. Increasing reports of successful case studies and pilot projects are accompanied by the lack of methodological standards and imposing and emerging challenges.
4. There is an urgent need for development of bioinformatics tools, extension of the usage of WGS in routine and real-time surveillance, and standardization of all aspects of WGS application in public health laboratories.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors would like to thank Marc Allard and Eric Brown for critical review of the manuscript. This work was supported in part by the USDA National Institute of Food and Agriculture Hatch project 1006141.

LITERATURE CITED

Allard MW, Luo Y, Strain E, Pettengill J, Timme R, et al. 2013. On the evolutionary history, population genetics and diversity among isolates of *Salmonella* Enteritidis PFGE pattern JEGX01.0004. *PLOS ONE* 8(1):e55254

Allos BM, Moore MR, Griffin PM, Tauxe RV. 2004. Surveillance for sporadic foodborne disease in the 21st century: the FoodNet perspective. *Clin. Infect. Dis.* 38(Suppl. 3):S115–20

Aziz N, Zhao Q, Bry L, Driscoll DK, Funke B, et al. 2015. College of American Pathologists' laboratory standards for next-generation sequencing clinical tests. *Arch. Pathol. Lab. Med.* 139(4):481–93

Baez-Ortega A, Lorenzo-Diaz F, Hernandez M, Gonzalez-Vila CI, Roda-Garcia JL, et al. 2015. IonGAP: Integrative bacterial genome analysis for Ion Torrent sequence data. *Bioinformatics* 31:2870–73

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19(5):455–77

Barco L, Barrucci F, Olsen JE, Ricci A. 2013. Salmonella source attribution based on microbial subtyping. *Int. J. Food Microbiol.* 163(2–3):193–203

Bazinet AL, Zwickl DJ, Cummings MP. 2014. A gateway for phylogenetic analysis powered by grid computing featuring GARLI 2.0. *Syst. Biol.* 63(5):812–18

Bolton DJ. 2015. *Campylobacter* virulence and survival factors. *Food Microbiol.* 48:99–108

Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, et al. 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* 58(7):3895–903

Cent. Dis. Control Prev. 2008. Salmonella *Annual Summary 2006*. Atlanta, GA: CDC

Chevreux B. 2005. *MIRA: an automated genome and EST assembler*. PhD Thesis, Ruprecht-Karls-Univ., Heidelberg, Ger.

Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, et al. 2012. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res.* 40(4):e29

Cody AJ, McCarthy ND, Jansen van Rensburg M, Isinkaye T, Bentley SD, et al. 2013. Real-time genomic epidemiological evaluation of human *Campylobacter* isolates by use of whole-genome multilocus sequence typing. *J. Clin. Microbiol.* 51(8):2526–34

Collins FS, Hamburg MA. 2013. First FDA authorization for next-generation sequencer. *N. Engl. J. Med.* 369(25):2369–71

Compeau PEC, Pevzner PA, Tesler G. 2011. How to apply de Bruijn graphs to genome assembly. *Nat. Biotechnol.* 29(11):987–91

Counc. Eur. Union. 1990. *On expenditure in the veterinary field.* Counc. Decis. 90/424/EEC, Counc. Eur. Union., Brussels, Belg.

Counc. Eur. Union. 1992. *Concerning measures for protection against specified zoonoses and specified zoonotic agents in animals and products of animal origin in order to prevent outbreaks of food-borne infections and intoxications.* Counc. Dir. 92/117/EEC, Counc. Eur. Union., Brussels, Belg.

Counc. Eur. Union. 2003a. *On the control of* Salmonella *and other specified food-borne zoonotic agents.* Counc. Dir. 2160/2003, Counc. Eur. Union., Brussels, Belg.

Counc. Eur. Union. 2003b. *On the monitoring of zoonoses and zoonotic agents.* Counc. Dir. 2003/99/EC, Counc. Eur. Union., Brussels, Belg.

Crim SM, Iwamoto M, Huang JY, Griffin PM, Gilliss D, et al. 2014. Incidence and trends of infection with pathogens transmitted commonly through food: Foodborne Diseases Active Surveillance Network, 10 U.S. sites, 2006–2013. *Morb. Mortal. Wkly. Rep.* 63(15):328–32

Cronquist AB, Mody RK, Atkinson R, Besser J, Tobin-D'Angelo M, et al. 2012. Impacts of culture-independent diagnostic practices on public health surveillance for bacterial enteric pathogens. *Clin. Infect. Dis.* 54(Suppl. 5):S432–39

den Bakker HC, Allard MW, Bopp D, Brown EW, Fontana J, et al. 2014. Rapid whole-genome sequencing for surveillance of *Salmonella enterica* serovar Enteritidis. *Emerg. Infect. Dis.* 20(8):1306–14

den Bakker HC, Switt AIM, Cummings CA, Hoelzer K, Degoricija L, et al. 2011. A whole-genome single nucleotide polymorphism–based approach to trace and identify outbreaks linked to a common *Salmonella enterica* subsp. *enterica* serovar Montevideo pulsed-field gel electrophoresis type. *Appl. Environ. Microbiol.* 77(24):8648–55

Deng X, Desai PT, den Bakker HC, Mikoleit M, Tolar B, et al. 2014. Genomic epidemiology of *Salmonella enterica* erotype Enteritidis based on population structure of prevalent lineages. *Emerg. Infect. Dis.* 20(9):1481–89

Deng X, Shariat N, Driebe EM, Roe CC, Tolar B, et al. 2015. Comparative analysis of subtyping methods against a whole-genome-sequencing standard for *Salmonella enterica* serotype Enteritidis. *J. Clin. Microbiol.* 53(1):212–18

Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. 2012. Transforming clinical microbiology with bacterial genome sequencing. *Nat. Rev. Genet.* 13:601–12

Didelot X, Wilson DJ. 2015. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLOS Comput. Biol.* 11(2):e1004041

Doyle MP, Erickson MC, Alali W, Cannon J, Deng X, et al. 2015. The food industry's current and future role in preventing microbial foodborne illness within the United States. *Clin. Infect. Dis.* 61(2):252–59

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29(8):1969–73

Earl DA, Bradnam K, St. John J, Darling A, Lin D, et al. 2011. Assemblathon 1: a competitive assessment of de novo short read assembly methods. *Genome Res.* 21(12):2224–41

Eur. Cent. Dis. Prev. Control. 2013. *Surveillance of communicable diseases in Europe - a concept to integrate molecular typing data into EU-level surveillance.* Stockholm, Sweden: Eur. Cent. Dis. Prev. Control. **http://ecdc.europa.eu/en/publications/Publications/surveillance-concept-molecular%20typing-sept2011.pdf**

Eur. Food Saf. Auth. 2010. Community Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in the European Union in 2008. Parma, Italy: EFSA. **http://www.efsa.europa.eu/en/efsajournal/pub/1496**

Eur. Food Saf. Auth. 2014. Technical specifications for the pilot on the collection of data on molecular testing of food-borne pathogens from food, feed and animal samples. Parma, Italy: EFSA. **http://www.efsa.europa.eu/it/supporting/pub/712e**

Farber JM, Peterkin PI. 1991. *Listeria monocytogenes*, a food-borne pathogen. *Microbiol. Rev.* 55(3):476–511

Fisher IST, Threlfall EJ, Enter-net, Salm-gene. 2005. The Enter-net and Salm-gene databases of foodborne bacterial pathogens that cause human infections in Europe and beyond: An international collaboration in surveillance and the development of intervention strategies. *Epidemiol. Infect.* 133(1):1–7

Foxman B, Zhang L, Koopman JS, Manning SD, Marrs CF. 2005. Choosing an appropriate bacterial typing technique for epidemiologic studies. *Epidemiol. Perspect. Innov.* 2:10

Gardner SN, Hall BG. 2013. When whole-genome alignments just won't work: kSNP v2 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial genomes. *PLOS ONE* 8(12):e81760

Gardner SN, Slezak T, Hall BG. 2015. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* 31:2877–78

Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, et al. 2012. Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat. Biotechnol.* 30(11):1033–36

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907v2

Gerner-Smidt P, Hise K, Kincaid J, Hunter S, Rolando S, et al. 2006. PulseNet USA: a five-year update. *Foodborne Pathog. Dis.* 3(1):9–19

Gilbert JM, White DG, McDermott PF. 2007. The US national antimicrobial resistance monitoring system. *Future Microbiol.* 2(5):493–500

Gilmour MW, Graham M, van Domselaar G, Tyler S, Kent H, et al. 2010. High-throughput genome sequencing of two *Listeria monocytogenes* clinical isolates during a large foodborne outbreak. *BMC Genomics* 11:120

Glob. Microb. Identif. 2011. *Perspectives of a Global, Real-Time Microbiological Genomic Identification System - Implications for National and Global Detection and Control of Infectious Diseases. Consensus Report of an Expert Meeting 1–2 September 2011, Bruxelles, Belgium.* Lyngby, Den.: Natl. Food Inst. Tech. Univ. Den. **http://www.globalmicrobialidentifier.org/-/media/Sites/gmi/News-and-events/2011/1-meeting-2011-report.ashx?la=da**

Gordon NC, Price JR, Cole K, Everitt R, Morgan M, et al. 2014. Prediction of *Staphylococcus aureus* antimicrobial resistance by whole-genome sequencing. *J. Clin. Microbiol.* 52(4):1182–91

Grad YH, Lipsitch M. 2014. Epidemiologic data and pathogen genome sequences: a powerful synergy for public health. *Genome Biol.* 15(11):538

Grad YH, Lipsitch M, Feldgarden M, Arachchi HM, Cerqueira GC, et al. 2012. Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. *PNAS* 109(8):3065–70

Grimont PAD, Weill F. 2007. *Antigenic formulae of the* Salmonella *serovars*. WHO Collab. Cent. Ref. Res. *Salmonella*, Paris, Fr.

Hald T, Vose D, Wegener HC, Koupeev T. 2004. A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. *Risk Anal.* 24(1):255–69

Hedge J, Wilson DJ. 2014. Bacterial phylogenetic reconstruction from whole genomes is robust to recombination but demographic inference is not. *mBio* 5(6):e02158-14

Hendriksen RS, Vieira AR, Karlsmose S, Lo Fo Wong DM, Jensen AB, et al. 2011. Global monitoring of *Salmonella* serovar distribution from the World Health Organization global foodborne infections network country data bank: results of quality assured laboratories from 2001 to 2007. *Foodborne Pathog. Dis.* 8:887–900

Hernandez D, François P, Farinelli L, Osterås M, Schrenzel J. 2008. De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Res.* 18(5):802–9

Hoffmann M, Luo Y, Monday SR, Gonzales-Escalona N, Ottesen AR, Muruvanda T, et al. 2016. Tracing origins of the *Salmonella* Bareilly strain causing a food-borne outbreak in the United States. *J. Infect Dis.* 213(4):502–8

Hohmann EL. 2001. Nontyphoidal salmonellosis. *Clin. Infect. Dis.* 32(2):263–69

Holmes EC, Urwin R, Maiden MC. 1999. The influence of recombination on the population structure and evolution of the human pathogen *Neisseria meningitidis*. *Mol. Biol. Evol.* 16(6):741–49

Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, et al. 2014. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* 6(11):90

Inst. Med. 2012. *Improving Food Safety Through a One Health Approach: Workshop Summary*. Washington, DC: Natl. Acad. Press

Iqbal Z, Caccamo M, Turner I, Flicek P, McVean G. 2012. De novo assembly and genotyping of variants using colored de Bruijn graphs. *Nat. Genet.* 44(2):226–32

Iqbal Z, Turner I, McVean G. 2013. High-throughput microbial population genomics using the Cortex variation assembler. *Bioinformatics* 29(2):275–76

Jamison DT, Bremen JG, Measham AR, Alleyne G, Claeson M, et al., eds. 2006. *Disease Control Priorities in Developing Countries*. Washington, DC: World Bank. 2nd ed.

Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, et al. 2014. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 52(5):1501–10

Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheutz F. 2015. Rapid and easy in silico serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *J. Clin. Microbiol.* 53(8):2410–26

Jones TF, Scallan E, Angulo FJ. 2007. FoodNet: overview of a decade of achievement. *Foodborne Pathog. Dis.* 4(1):60–66

Kaakoush NO, Castaño-Rodríguez N, Mitchell HM, Man SM. 2015. Global epidemiology of *Campylobacter* infection. *Clin. Microbiol. Rev.* 28(3):687–720

Katz LS, Petkau A, Beaulaurier J, Tyler S, Antonova ES, et al. 2013. Evolutionary dynamics of *Vibrio cholerae* O1 following a single-source introduction to Haiti. *mBio* 4(4):e00398-13

Keim P, van Ert MN, Pearson T, Vogler AJ, Huynh LY, Wagner DM. 2004. Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales. *Infect. Genet. Evol.* 4(3):205–13

Klumpp J, Fouts DE, Sozhamannan S. 2012. Next generation sequencing technologies and the changing landscape of phage genomics. *Bacteriophage* 2(3):190–99

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, et al. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22(3):568–76

Koren S, Phillippy A. 2015. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Curr. Opin. Microbiol.* 23:110–20

Köser CU, Ellington MJ, Cartwright EJP, Gillespie SH, Brown NM, et al. 2012. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLOS Pathog.* 8(8):e1002824

Kupferschmidt K. 2011. Epidemiology. Outbreak detectives embrace the genome era. *Science* 30:1818–19

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Meth.* 9(4):357–59

Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, et al. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin. Microbiol.* 50(4):1355–61

Leekitcharoenphon P, Kaas RS, Thomsen MCF, Friis C, Rasmussen S, Aarestrup FM. 2012. snpTree: a web-server to identify and construct SNP trees from whole genome sequence data. *BMC Genomics* 13(Suppl. 7):S6

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–60

Lienau EK, Strain E, Wang C, Zheng J, Ottesen AR, et al. 2011. Identification of a salmonellosis outbreak by means of molecular sequencing. *N. Engl. J. Med.* 364(10):981–82

Liljebjelke KA, Hofacre CL, Liu T, White DG, Ayers S, et al. 2005. Vertical and horizontal transmission of *Salmonella* within integrated broiler production system. *Foodborne Pathog. Dis.* 2(1):90–102

Lindstedt B-A, Heir E, Gjernes E, Kapperud G. 2003. DNA fingerprinting of *Salmonella enterica* subsp. *enterica* serovar Typhimurium with emphasis on phage type DT104 based on variable number of tandem repeat loci. *J. Clin. Microbiol.* 41(4):1469–79

Lipkin WI. 2013. The changing face of pathogen discovery and surveillance. *Nat. Rev. Microbiol.* 11(2):133–41

Loman NJ, Constantinidou C, Christner M, Rohde H, Chan JZ-M, et al. 2013. A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic *Escherichia coli* O104:H4. *JAMA* 309(14):1502–10

Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, et al. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* 30:434–39

Maiden MCJ, Bygraves JA, Feil EJ, Morelli G, Russell JE, et al. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *PNAS* 95(6):3140–45

Majowicz SE, Scallan E, Jones-Bitton A, Sargeant JM, Stapleton J, et al. 2014. Global incidence of human Shiga toxin-producing *Escherichia coli* infections and deaths: a systematic review and knowledge synthesis. *Foodborne Pathog. Dis.* 11(6):447–55

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27(6):764–70

Martin SM, Bean NH. 1995. Data management issues for emerging diseases and new tools for managing surveillance and laboratory data. *Emerg. Infect. Dis.* 1(4):124–28

Mather AE, Reid SWJ, Maskell DJ, Parkhill J, Fookes MC, et al. 2013. Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. *Science* 341(6153):1514–17

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20(9):1297–303

Meacham F, Boffelli D, Dhahbi J, Martin DIK, Singer M, Pachter L. 2011. Identification and correction of systematic error in high-throughput sequence data. *BMC Bioinform.* 12:451

Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, et al. 2011. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLOS ONE* 6(7):e22751

Miller JR, Koren S, Sutton G. 2010. Assembly algorithms for next-generation sequencing data. *Genomics* 95(6):315–27

Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, et al. 2000. A whole-genome assembly of *Drosophila*. *Science* 287(5461):2196–204

Neuman JA, Isakov O, Shomron N. 2013. Analysis of insertion-deletion from deep-sequencing data: software evaluation for optimal detection. *Brief. Bioinform.* 14(1):46–55

O'Brien SJ, Gillespie IA, Adak GK. 2005. Foodborne disease surveillance as a basis for policy-making. In *Food Safety Assurance and Veterinary Public Health: Risk Management Strategies*, *Monitoring and Surveillance*, ed. FJM Smulders, JD Collins, pp. 33–52. Wageningen, Neth.: Wageningen Acad. Publ.

Octavia S, Wang Q, Tanaka MM, Kaur S, Sintchenko V, Lan R. 2015. Delineating community outbreaks of *Salmonella enterica* serovar Typhimurium by use of whole-genome sequencing: insights into genomic variability within an outbreak. *J. Clin. Microbiol.* 53(4):1063–71

Okoro CK, Kingsley RA, Connor TR, Harris SR, Parry CM, et al. 2012. Intracontinental spread of human invasive *Salmonella* Typhimurium pathovariants in sub-Saharan Africa. *Nat. Genet.* 44:1215–21

Olson ND, Lund SP, Colman RE, Foster JT, Sahl JW, et al. 2015. Best practices for evaluating single nucleotide variant calling methods for microbial genomics. *Front. Genet.* 6:235

Orsi RH, Borowsky ML, Lauer P, Young SK, Nusbaum C, et al. 2008. Short-term genome evolution of *Listeria monocytogenes* in a non-controlled environment. *BMC Genomics* 9:539

Pettengill JB, Luo Y, Davis S, Chen Y, González-Escalona N, et al. 2014. An evaluation of alternative methods for constructing phylogenies from whole genome sequence data: a case study with *Salmonella*. *PeerJ* 2:e620

Pires SM, Vieira AR, Hald T, Cole D. 2014. Source attribution of human salmonellosis: an overview of methods and estimates. *Foodborne Pathog. Dis.* 11(9):667–76

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2: approximately maximum-likelihood trees for large alignments. *PLOS ONE* 5(3):e9490

Quail MA, Smith M, Coupland P, Otto TD, Harris SR, et al. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13(1):341

Quick J, Ashton P, Calus S, Chatt C, Gossain S, et al. 2015. Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biol.* 16(1):114

Ribot EM, Fair MA, Gautom R, Cameron DN, Hunter SB, et al. 2006. Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157:H7, *Salmonella*, and *Shigella* for PulseNet. *Foodborne Pathog. Dis.* 3(1):59–67

Robinson ER, Walker TM, Pallen MJ. 2013. Genomics and outbreak investigation: from sequence to consequence. *Genome Med.* 5(4):36

Rohde H, Qin J, Cui Y, Li D, Loman NJ, et al. 2011. Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. *N. Engl. J. Med.* 365(8):718–24

Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, et al. 2013. Characterizing and measuring bias in sequence data. *Genome Biol.* 14(5):R51

Rump B, Cornelis C, Woonink F, Verweij M. 2013. The need for ethical reflection on the use of molecular microbial characterisation in outbreak management. *Eur. Surveill.* 18(4):20384

Ruppitsch W, Pietzka A, Prior K, Bletz S, Fernandez HL, et al. 2015. Defining and evaluating a core genome multilocus sequence typing scheme for whole-genome sequence-based typing of *Listeria monocytogenes*. *J. Clin. Microbiol.* 53(9):2869–76

Sabat AJ, Budimir A, Nashev D, Sa-Leao R, van Dijl JM, et al. 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Eur. Surveill.* 18(4):20380

Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, et al. 2011. Foodborne illness acquired in the United States: major pathogens. *Emerg. Infect. Dis.* 17(1):7–15

Smith HW. 1951. The typing of *Salmonella* Thompson by means of bacteriophage. *J. Gen. Microbiol.* 5(3):472–79

Smith JM, Smith NH, O'Rourke M, Spratt BG. 1993. How clonal are bacteria? *PNAS* 90(10):4384–88

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–90

Stasiewicz MJ, Oliver HF, Wiedmann M, den Bakker HC. 2015. Whole genome sequencing allows for improved identification of persistent *Listeria monocytogenes* in food associated environments. *Appl. Environ. Microbiol.* In press

Swaminathan B, Barrett TJ, Fields P. 2006a. Surveillance for human *Salmonella* infections in the United States. *J. AOAC Int.* 89(2):553–59

Swaminathan B, Gerner-Smidt P, Ng L-K, Lukinmaa S, Kam K-M, et al. 2006b. Building PulseNet International: an interconnected system of laboratory networks to facilitate timely public health recognition and response to foodborne disease outbreaks and emerging foodborne diseases. *Foodborne Pathog. Dis.* 3(1):36–50

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30(12):2725–29

Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, et al. 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J. Clin. Microbiol.* 33(9):2233–39

Timme RE, Pettengill JB, Allard MW, Strain E, Barrangou R, et al. 2013. Phylogenetic diversity of the enteric pathogen *Salmonella enterica* subsp. *enterica* inferred from genome-wide reference-free SNP characters. *Genome. Biol. Evol.* 5(11):2109–23

Treangen TJ, Ondov BD, Koren S, Phillippy A. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* 15(11):524

Tyson GH, McDermott PF, Li C, Chen Y, Tadesse DA, et al. 2015. WGS accurately predicts antimicrobial resistance in *Escherichia coli*. *J. Antimicrob. Chemother.* 70:2763–69

van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, et al. 2007. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clin. Microbiol. Infect.* 13(Suppl. 3):1–46

van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* 11(1110):11.10.1–11.10.33

van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. 2014. Ten years of next-generation sequencing technology. *Trends Genet.* 30(9):418–26

Vergnaud G, Pourcel C. 2009. Multiple locus variable number of tandem repeats analysis. *Methods Mol. Biol.* 551:141–58

Ward LR, de Sa JD, Rowe B. 1987. A phage-typing scheme for *Salmonella enteritidis*. *Epidemiol. Infect.* 99(2):291–94

Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, et al. 2014. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N. Engl. J. Med.* 370(25):2408–17

Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, et al. 2012. Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67(11):2640–44

Zerbino DR. 2010. Using the Velvet de novo assembler for short-read sequencing technologies. *Curr. Protoc. Bioinform.* Chapter 11:Unit 11.5

Zhang S, Yin Y, Jones MB, Zhang Z, Deatherage Kaiser BL, et al. 2015. *Salmonella* serotype determination utilizing high-throughput genome sequencing data. *J. Clin. Microbiol.* 53(5):1685–92

Zheng J, Pettengill J, Strain E, Allard MW, Ahmed R, et al. 2014. Genetic diversity and evolution of *Salmonella enterica* serovar Enteritidis strains with different phage types. *J. Clin. Microbiol.* 52(5):1490–500

Zhou Z, McCann A, Litrup E, Murphy R, Cormican M, et al. 2013. Neutral genomic microevolution of a recently emerged pathogen, *Salmonella enterica* serovar Agona. *PLOS Genet.* 9(4):e1003471

Annual Review
of Food Science
and Technology
Volume 7, 2016

Contents