Trends in Analytical Chemistry 96 (2017) 52–61

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

journal homepage: www.elsevier.com/locate/trac

Precision food safety: A systems approach to food safety facilitated by genomics tools

Jasna Kovac [a, b], Henk den Bakker [c], Laura M. Carroll [a], Martin Wiedmann [a, *]

[a] Department of Food Science, Cornell University, Stocking Hall, Ithaca, NY 14853, United States
[b] Department of Food Science, Pennsylvania State University, University Park, State College, PA 16803, United States
[c] Department of Animal and Food Sciences, Texas Tech University, Box 42141, Lubbock, TX 79409, United States

ARTICLE INFO

Article history:
Available online 9 June 2017

Keywords:
Precision food safety
Functional genomics
Metagenomics
Outbreak investigation
Risk assessment

ABSTRACT

Genomics and related tools are beginning to have a tremendous impact on food safety by allowing for a more precise approach to pathogen detection, characterization, and identification. The data produced using these tools are expected to lead to a paradigm shift in modern food safety concepts with impacts similar to those facilitated through development of personalized medicine methods in human medicine. We hence envision a future of food safety that can be best described as "*precision food safety*". While precision food safety will employ a number of new approaches and methods, omics tools will play a particularly important role and will provide for improved and more precise food safety procedures, including increasingly accurate risk assessment that will support evidence-based food safety decision-making. This review will highlight examples where genomics-based precision food safety approaches will have a significant impact.

1. Introduction

As part of the "foodomics" [1], the rapid emergence and adoption of data-intensive tools in food safety is heralding the early stages of a paradigm shift that we propose will usher in an era of implementing new approaches to food safety, which can be best described as "*precision food safety*". While a number of approaches, such as Geographic Information Systems (GIS) technologies [2–6], play essential roles in precision food safety, omics technologies are one of the key drivers of this movement. In particular, whole genome sequencing (WGS) not only allows for highly sensitive "precision" subtyping that facilitates considerably improved detection of foodborne disease outbreaks [7–10], but also allows for comprehensive characterization of foodborne pathogens and identification of strains and clonal groups that differ in virulence and antimicrobial resistance [11–18]. While WGS is becoming increasingly used in routine surveillance of select foodborne pathogens, evaluation of the practicality of metagenomics and metatranscriptomics for detection of pathogens in food and in humans has just begun [19–24]. Despite sensitivity challenges, which are the main weakness of metagenomic methods compared to traditional enrichment-based methods, shifts in microbiome structure may be predictive of known and yet uncharacterized pathogens or food safety hazards. Metagenomic approaches may hence allow for improved detection of food safety issues and other aberrations in foods and ingredients that may be relevant in food production. Genomics and metagenomics approaches can also be complemented with proteomics and metabolomics methods to detect bacterial toxins and mycotoxins in foodstuff [1,25,26].

Considering the rapid establishment of WGS in food safety in the past few years, we anticipate a future where omics-based tools will become an indispensable part of the food safety surveillance and risk assessment toolbox and will allow for improved phylogenetic and genetic classification of pathogens through precise identification of virulence and antimicrobial resistance genes. These omics-derived data will drive subsequent phenotypic characterization and help to better define and identify groups of bacteria that represent foodborne pathogens and public health hazards. For

* Corresponding author. Department of Food Science, Cornell University, 341 Stocking Hall, Ithaca, NY 14853, United States.
E-mail addresses: jk2739@cornell.edu, jzk303@psu.edu (J. Kovac), Henk.C.den-bakker@ttu.edu (H. Bakker), lmc297@cornell.edu (L.M. Carroll), mw16@cornell.edu (M. Wiedmann).

http://dx.doi.org/10.1016/j.trac.2017.06.001

example, isolates in which known or putative virulence and antimicrobial genes are detected based on WGS can be further phenotypically characterized to confirm the role of these genes in phenotypes of interest (e.g., cytotoxic effects, phenotypic antimicrobial resistance). Integrated omics- and phenotype-based data will ultimately allow for quantitative characterization of risks associated with different groups of bacteria, guiding the development of targeted food safety policies and procedures. Hence, food safety measures will be increasingly less reliant on simplifying assumptions, such as fairly homogenous distribution of virulence-related characteristics in a given species, allowing for scientifically rigorous identification and characterization of well-defined subgroups with specific virulence-associated characteristics. Ultimately, precision food safety approaches will allow for development of more targeted policies and procedures focusing on species subgroups (e.g., serotypes) or clonal groups, as detailed below with specific examples. Application of precision tools for monitoring of microbial and pathogen populations in foods, food-associated environments, and in humans will provide rapid feedback supporting identification of new and unrecognized food safety hazards as well as refinement and validation of pathogen definitions and detection and characterization methods. This review will provide an overview of the rapidly improving omics tools and their use in the development of a precision food safety framework (Fig. 1), with examples demonstrating how omics tools are already creating the base knowledge leading food safety into the precision era.

2. Implementation of WGS will considerably improve outbreak and cluster detection

Use of WGS in bacterial population genomics has tremendously increased our understanding of genome evolution and biology of bacterial pathogens [27–29]. While genome sequencing was initially expensive, the introduction of next generation sequencing technologies and less expensive small bench top sequencers [30] has decreased the overall sequencing costs. This brought the per isolate price of microbial WGS to a point where it is comparable or even below the price range of traditional subtyping methods (e.g., Pulsed Field Gel Electrophoresis [PFGE]) and made WGS an indispensible tool in modern outbreak investigations. One of the earliest reports on the use of WGS in support of a foodborne disease outbreak investigation was by Gilmour et al. [10], describing the genome sequences of two distinct *Listeria monocytogenes* strains involved in a multi-province outbreak in Canada in 2008. The first report of the use of WGS to infer the potential source of a foodborne outbreak was by Lienau et al. [9], involving isolates of the multistate outbreak of *Salmonella* Montevideo which occurred between July 2009 and May 2010 (http://www.cdc.gov/salmonella/montevideo/). The feasibility of integrating WGS in the routine outbreak surveillance of *Salmonella* Enteritidis, a pathogen that cannot be discriminated well by PFGE, was first demonstrated by Den Bakker et al. [8]. As of March 2017, WGS is used for routine *L. monocytogenes* surveillance and to support investigations of foodborne disease

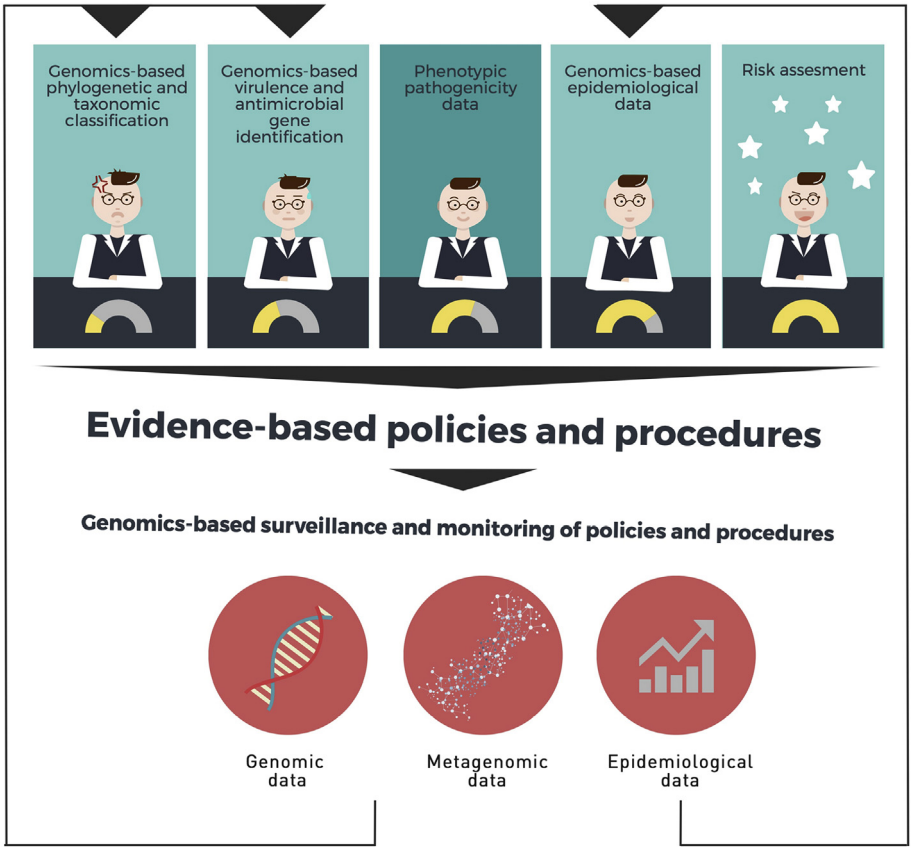

Fig. 1. The principle of genomics-enabled precision food safety. Genomics based phylogenetic and taxonomic classification, as well as identification of virulence and antimicrobial resistance genes provide for precise characterization of pathogens that informs subsequent phenotypic ("functional genomics") characterization for refined definition and identification of pathogens. Integration of these data with functional genomics and epidemiological data and utilization of the combined data sets in risk assessments allows for more precise science based formation of food safety policies (at the governmental level) and food safety procedures (at the firm level). The effectiveness of these policies is monitored with genomics and metagenomics tools that allow for precise identification of known, as well as new and emerging pathogens. This information is feeding back into genomics-based phylogenetic and taxonomic classification, as well as identification of pathogens and pathogen clonal groups that require targeted policies and procedures for their control.

outbreaks caused by a number of other pathogens in the United States [31]. These efforts made by Centers for Disease Control and Prevention (CDC), Food and Drug Administration (FDA) and U.S. Department for Agriculture (USDA) are facilitated by the publicly accessible GenomeTrakr database [7].

WGS, similar to its DNA sequence-based predecessor multilocus sequence typing (MLST), profits in theory from the unambiguous nature and electronic portability of nucleotide sequence data [32]. Additionally, WGS yields an unprecedented resolution because the majority (>90%) of the genome is interrogated for genomic differences, as compared to just a small part (<0.01%) of the genome being used in the traditional 7-gene MLST. In addition to high resolution subtyping, WGS can be used to predict antimicrobial resistance phenotypes [33] and additional phenotypic characteristics [34], supporting precision antibiotic treatment for patients associated.

Current WGS technologies can be subdivided into two categories based on the length of the sequence reads they produce; (i) short read sequencing technologies, producing reads up to 600 base-pairs long (e.g., Illumina, Ion Torrent), and (ii) long read technologies, capable of producing reads longer than 1000 base-pairs and often longer than 70,000 base-pairs (e.g., Pacific Biosciences, Oxford Nanopore). Reads produced by both types of technologies display different degrees of sequencing error [35]. This caveat is overcome by sequencing a genome multiple times per genomic position (coverage), and bioinformatically by inferring the consensus for individual genomic positions. Bioinformatics approaches to infer genomic differences can be subdivided into two categories: (i) *de novo* methods and (ii) reference based methods. Reference based methods rely on alignment of reads to a previously sequenced genome, usually by Burrow-Wheeler transform algorithm to make mapping of large numbers of reads computationally efficient [36,37]. *De novo* methods do not rely on mapping to a reference genome, but instead try to either reconstruct (parts of) the genome or try to find differences in smaller sub-sequences of the genome (i.e., k-mers). Algorithms used for short read data are usually based on a De Bruijn graph-based algorithm [38,39], while data generated with long read technologies generally rely on overlap consensus-based algorithms [40].

Types of WGS analyses are currently used for precision outbreak cluster detection; core genome MLST (cgMLST), whole genome MLST (wgMLST), and reference mapping high quality SNP-based clustering pipelines. For core genome MLST (cgMLST) [41,42] a database is created containing all core genes of a particular taxon (e.g., a species, subspecies or serovar) of interest, and the unique allelic types found for each of the core genes. wgMLST schemes include a large number of loci that are not part of the core genome and hence characterize also the accessory genome. Congruent with MLST, a unique combination of allele sequences is considered a cgMLST type (CT) or a wgMLST type. Because at least a few allele differences are usually observed among typically more than thousand core genes, the criterion of the CT or wgMLST type is relaxed by allowing a certain number of allele (e.g., 7 alleles) differences within a CT or a wgMLST type [43]. The advantage of the cgMLST and wgMLST approaches is that they provide a higher level of discrimination compared to older subtyping methods such as PFGE, and a standardized nomenclature of CTs and wgMLST types, which improves communication between national and international public health centers. A potentially higher discriminatory resolution can be obtained with the reference mapping high quality SNP (hqSNP)-based clustering pipelines [44,45]. While alleles in cgMLST and wgMLST may differ by single nucleotides, insertions and deletions, these pipelines focus on genome wide single nucleotide differences (single nucleotide variants (SNVs) or single nucleotide polymorphisms (SNPs)) inferred based on mapped reads. High quality refers to the fact that only SNPs that comply to specific requirements are used, such as minimum coverage of the polymorphic site, a minimal percentage of reads confirming the alternative allele, etc. Clustering of isolates is then determined by performing a phylogenetic analysis of a matrix based on the SNP sites. Practically, the relative discriminatory power of cg and wgMLST approaches and hqSNP approaches still needs to be evaluated more stringently and for different organisms.

While integration of WGS in surveillance programs of *L. monocytogenes* and *Salmonella enterica* has proven to be successful in resolving outbreaks that previously would not have been detected [31], further standardization of the application of WGS in outbreak detection is still needed. One could think of multiple technical aspects that should be included in such standards, for example the minimal (*de novo*) assembly quality for use in cgMLST, minimal mapping quality for SNPs for reference-based mapping pipelines, and standard practices in sequencing labs. Feasibility of the standardization of cgMLST data using standardized databases and nomenclature for subtypes has been demonstrated for *L. monocytogenes* [43]. Standardization of SNP-based clustering, however, may prove challenging, because a standard cut off for the number of SNP differences for the inclusion or exclusion of an isolate in an outbreak cluster may be difficult to determine. Important factors include the significant differences in mutation rates observed in different species of pathogens and even between subpopulations within species [46] as well as considerable differences in the number of generations per time unit (e.g., year) in different environments, making it impossible to use a generic cut-off for outbreak definitions across all bacterial pathogens, and showing the necessity of including traditional epidemiological methods, and additional metadata in outbreak cluster definitions. Important steps towards standardization of the application of genomic data for food microbiology, pathogen identification and disease outbreak detection have clearly been made on international level by regulatory agencies outlining best practices for data integrity, reproducibility and traceability [47].

3. Metagenomics tools provide a unique opportunity to monitor supply chains and detect supply chain aberrations that may represent food safety hazards

While the implementation of WGS has allowed characterization of foodborne pathogens at unprecedented resolution, WGS still requires the isolation of the pathogen in question via culture-based methods, a process which involves the use of pathogen-specific enrichment media, selective media, and isolation protocols. The idea that one can sequence all microorganisms in a particular food matrix at any point in the food supply chain and detect pathogens and spoilage organisms prior to distribution and consumption without making assumptions about a possible contaminant makes metagenomic sequencing an attractive approach within the context of food safety.

Until recently, high-throughput sequencing of targeted amplicons such as the 16S ribosomal DNA gene (16S rDNA) has been the method of choice for characterizing microbial communities of interest. 16S rDNA sequencing is relatively inexpensive and boasts a well-established repertoire of computational tools, pipelines, and databases with which data can be analyzed [48–50]. In addition, 16S rDNA genes are present in all bacterial species, making them ideal targets for bacterial diversity studies [49], particularly in complex food matrices where an over-abundance of DNA from a eukaryotic host may limit the depth of sequencing [50,51]. This approach has been used to characterize the microbiota of various foods [50,52,53] and has provided new insights into microbial community dynamics during food fermentations [50] and during

enrichment of foods for pathogen detection [21,54]. 16S rDNA amplicon sequencing has also been used to monitor shifts in food processing facility microbiomes, which may pinpoint the sources of pathogen or food spoilage microorganisms that reside in the processing environment [55,56]. While metagenomics approaches have provided novel insights into the microbial community composition of various food matrices and food processing facilities, these approaches have inherent shortcomings that are difficult to overcome when applied for detection of foodborne pathogens along the food supply chain. These include general limitation of amplicon sequencing compared to traditional enrichment methods, such as (i) poor sensitivity of amplicon sequencing for detection of low-level contamination, (ii) amplification bias, and (iii) inability to distinguish between genetic material originating from viable and non-viable cells (e.g. differentiating viable *Salmonella* from inactivated *Salmonella* in poultry products) [57,58]. Further limitations of 16S rDNA amplicon sequencing, compared to other meta-omics methods, include (i) the inability to reliably distinguish pathogenic from non-pathogenic species (e.g. *L. monocytogenes* from *Listeria innocua* [59], human pathogens *Bacillus anthracis* from *Bacillus cereus* and insect pathogen *Bacillus thuringiensis* [12]), and (ii) lack of functional characterization of the microbiome in question [60], such as information about species subtypes (e.g., serotypes or clonal groups), the presence of antimicrobial resistance determinants, and virulence potential.

An alternative to 16S rDNA sequencing comes in the form of metagenomic shotgun sequencing, in which all DNA present in a sample—rather than just a genetic marker of interest—is sequenced. This approach overcomes the amplification bias, taxonomic resolution, and functionality issues of 16S rDNA sequencing, but can be less sensitive for pathogen detection due to untargeted DNA sequencing. Like 16S rDNA sequencing, shotgun metagenomic sequencing also does not allow for differentiation between DNA originating from living and dead organisms. Nevertheless, metagenomic shotgun sequencing has been rapidly gaining popularity, facilitated by decreased sequencing costs and increasingly accessible tools, pipelines, and approaches for analyzing the large quantity of data it produces [60,61]. In its current manifestation, metagenomic shotgun sequencing is performed by extracting all of the genomic DNA from a particular community (with optional steps to deplete background DNA from a eukaryotic host to increase the sensitivity of microbial gene detection), shearing the DNA into smaller fragments, and sequencing these fragments to produce reads [61]. After sequencing, data analysis is carried out according to the goals of the experiment, which may include an assessment of the taxonomic diversity of the community through alignment or assembly [61], gene prediction and functional annotation [61], and/or associating community data with a particular phenotype (a metagenome-wide association study), as commonly done for human disease states [60,62].

Metagenomic shotgun approaches have been recently extended to survey communities in foods [50], with a number of them focusing on characterizing or detecting foodborne pathogens and/or spoilage organisms in a variety of food matrices [21–24]. This approach has also been used to characterize shifts in microbiome composition at different stages of microbiological enrichment of foodborne pathogens from tomatoes and cilantro [21,23], which provided important information regarding competing microbiota that is able to grow or even inhibit foodborne pathogens in selective isolation media. This demonstrates how metagenomics approaches can be used also for development of improved selective media for traditional isolation of foodborne pathogens. To overcome the sensitivity issues, metagenomic sequencing has been combined with short enrichment periods for detection of foodborne pathogens, such as Shiga-toxin producing *Escherichia coli* on fresh-bagged spinach, and showed that 10 CFU/100 g can be detected after 8 h of enrichment [22].

One of the key advantages of shotgun metagenomics is generation of data that allows for functional inference. This potential has been exploited for tracing of foodborne pathogens and antimicrobial resistance genes along the food supply chain, as shown by two studies that specifically monitored beef production chain from cattle entry into the feedlot to market-ready meat production [20,24]. However, current challenges facing such approaches when applied to meat and similar food products is reduced sequencing capacity due to sequencing of host DNA. Over 99% of the reads in these types of samples may map to the host [24], which results in poor sensitivity for detection of genes with low abundance (e.g., antimicrobial resistance genes, virulence genes). To overcome this challenge given current technology, it is beneficial to integrate different methods (e.g., metagenomics, qPCR, phenotypic characterization), each bringing its own strengths to the table. Such a functional omics approach has been successfully used to study causes of pink discoloration defect in cheese [63], and it is easy to imagine how similar strategies could be applied in food safety to characterize antimicrobial and virulence potential of microbial communities residing in foods.

Additionally, some of the shortcomings of metagenomics can be compensated by metatranscriptomic sequencing, which is becoming used in food safety and quality. Unlike metagenomic sequencing, which involves sequencing the DNA present in an entire community, metatranscriptomic sequencing involves sequencing complementary DNA (cDNA) created from RNA extracted from an entire community, typically focusing on mRNA. While metagenomics gives insight into which genes are present in a community, metatranscriptomics allows for specific identification of genes in a population that are being transcribed and possibly translated. While some assume that this approach allows for differentiation between live and dead cells present in the sample, considering that most of the short-lived microbial mRNA is degraded rapidly after microbial cell death, this is not necessarily a given as some RNAs may be highly stable and may remain intact for considerable time after death of an organism. Metatranscriptomics allows for selective sequencing of microbial mRNA after rRNA, tRNA and polyadenylated eukaryotic mRNA depletion [64]. Selective sequencing of cDNA produced from microbial mRNA alone provides for better use of sequencing capacity and, hence, results in a deeper coverage of microbial transcriptome, which increases the sensitivity of this method for detection of foodborne pathogens in food samples rich in eukaryotic host DNA. Nevertheless, metatranscriptomics has limited use in detection of antimicrobial resistance and virulence genes in foodborne pathogens, as these genes are unlikely to be transcribed in the absence of selective pressure or outside of the host. In its current manifestation, metatranscriptomic shotgun sequencing approaches have been applied mostly within the realm of food quality [50,52,65,66], with many early studies being performed to characterize the microbiomes of various cheeses [67–69]. Its utility in detection of foodborne pathogens has been assessed in clinical setting [70] and is currently being investigated for application in food by IBM's Sequencing the Food Supply Chain Consortium (http://www.research.ibm.com/client-programs/foodsafety/), which is employing both metagenomic and metatranscriptomic approaches to characterize baseline microbial communities along the food supply chain and detect safety and quality anomalies [71].

Together, metagenomic and metatranscriptomic sequencing have the potential to address a number of issues facing the global food supply chain at greater precision than their technological predecessors. While more research is needed to optimize these technologies within the realm of food safety, future applications

might include pathogen detection, monitoring of antimicrobial resistance determinants, quality control, and detection of fraudulent products. As sequencing costs continue to decrease and data analysis tools become more accessible, these high-throughput biosurveillance approaches are likely to become more popular for assessing food safety and quality from farm to fork. The greatest challenge with genomics and metagenomics methods is the turnaround time, which at the current state of the development, cannot compete with sensitive and rapid pathogen detection methods that target specific (groups) of microorganisms and are able to differentiate between viable and non-viable microorganisms (e.g., RAPID-B; [72]). Detection methods, such as RAPID-B can therefore be used to complement the (meta)genomics methods to enhance the pathogen detection time and ensure the ability to differentiate between viable and non-viable detected organisms.

4. Functional genomics and analysis of pathogen genomes allow for improved definition of strains, subtypes and public health risks associated with them

Successful implementation of WGS in surveillance of foodborne pathogens through laboratory networks such as GenomeTrakr [7] is generating increasing amounts of genomic data available through open databases. Although the primary rationale for collecting these data is to improve outbreak detection, the practical utility of these data includes many areas of food safety research. Specifically, genomic data is increasingly used for characterization of virulence [11–13,16,18] and antimicrobial resistance [73–76] in foodborne pathogens, and the knowledge generated through these efforts is rapidly changing the way we define microbial food safety hazards. Currently regulatory authorities responsible for food safety still largely define microbial food safety hazards based on taxonomic classification into species, with limited consideration of the disparity in pathogenic potential at a subtype level. Nevertheless, strides towards subtype-specific food safety hazard definition have already been made as demonstrated by the example of certain serotypes of Shiga toxin-producing *E. coli* (STEC), which are considered a higher food safety risky as compared to other *E. coli* pathovars. On the other hand, differentiation among pathogen subtypes is generally not used to assess *Salmonella* as a food safety hazard, despite a growing body of evidence suggesting considerable differences in the ability of different non-typhoidal *Salmonella* serotypes to cause disease in humans and animals [77–79]. Generating data that will allow for reliable differentiation among pathogens on a subtype level is one of the key goals of precision food safety and includes several steps.

Firstly, a stride toward precision food safety requires improved phenotypic and taxonomic definitions of pathogenic microbial isolates, and secondly, it requires an establishment of food safety risks associated with the presence of different subtypes in food (Fig. 2). Functional genomics is already playing an integral role in this process by serving as a powerful tool for elucidating relationships between genotypes and phenotypes (e.g., ability to cause disease or resist antimicrobial treatment), taking individual genetic traits and environmental conditions affecting the expression of genetic traits into consideration [80,81]. Complementary to functional genomics, maximum likelihood and Bayesian phylogenetic inference based on the WGS single nucleotide polymorphisms (SNPs), as well as pairwise genomic similarity analyses (e.g., Average Nucleotide Identity blast [ANIb], *in silico* DNA-DNA hybridization [DDH]) are allowing for more accurate species and subtype definitions when dealing with closely related organisms that cannot be delineated using 16S rDNA sequences [82]. Phylogenomic analyses are also beginning to improve our understanding of the mechanisms driving the evolution and transmission of

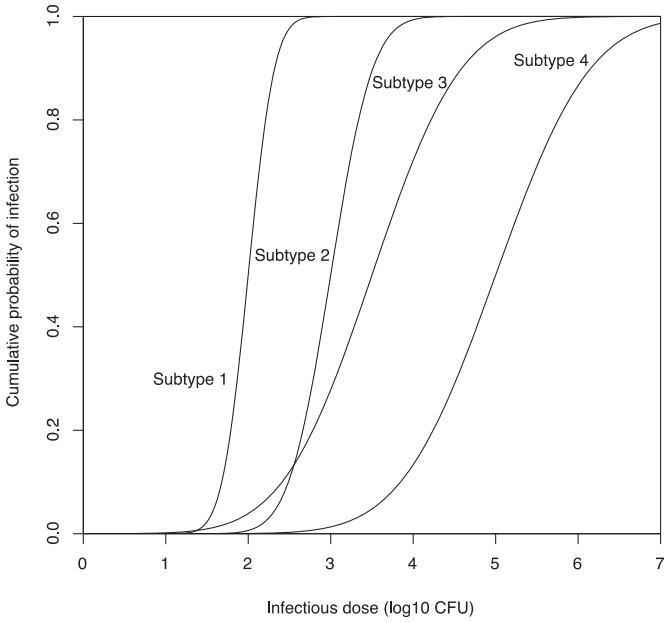

Fig. 2. Dose response curves may differ substantially among specific subtypes of foodborne pathogens within the species, pointing out the need for collecting subtype-specific data, to allow for precise food safety risk assessment. X-axis represents infectious dose in log10 CFU and Y-axis represents the corresponding cumulative probability of infection.

virulence factors [15]. This is particularly informative and important for species in which virulence factors do not correlate well with taxonomic definitions and where virulence potential thus needs to be assessed at the subtype level. In such cases integration of phylogenetic (e.g., inferring taxonomic relationships), functional genomic (e.g., function of genes and regulation of their expression), and phenotypic (e.g., toxicity and invasion assays in tissue culture and animal models) data is essential for development of hazard characterization schemes that can facilitate improved and more precise food safety risk assessments.

Salmonella is one example where omics-enabled precision food safety approaches will allow for improved hazard characterization. Among the >2500 *Salmonella* serotypes, only five have been estimated to cause >60% of human disease cases [77,83]. Although the prevalence of different serotypes in human disease cases may in part be attributed to differences in exposure, a number of studies suggest host adaptation and differences in virulence potential among different serotypes [16,18,77,79]. For example, recent studies suggest that some serotypes possess distinct virulence genes that may lead to specific virulence characteristics. For example serotypes Javiana, Montevideo, and Oranienburg as well as a specific Mississippi clade encode a cytolethal distending toxin [79,84,85] that can cause DNA damage, which may affect disease outcomes and could cause long-term sequelae. Serotypes also have been reported to differ in their invasiveness and antimicrobial resistance patterns, which influence the infection outcomes and the success of clinical treatment [16,77,86]. Some serotypes also appear to be more frequently associated with infection of animals (e.g., Dublin, Cerro). Recent functional genomics studies on *S.* Cerro, which appears to be an emerging serotype in dairy in the United States, led to identification of a point mutation in the virulence gene *sopA*, as well as genomic changes likely responsible for its impaired pathogenic potential in humans [16,18].

Another area where precision food safety approaches will provide important new insights are cases where species definitions for closely related bacterial species do not necessarily correlate well

with virulence and where clonal groups or strains that have been linked to foodborne illness cases can be found across multiple species [12,15,87]. The potential for horizontal transmission of virulence genes among closely related species is further obstructing assessment of pathogenic potential and associated food safety risk in such cases. One example of this scenario is represented by *B. cereus* and closely related species, a group of Gram-positive foodborne pathogens for which a number of recent studies suggest that precision food safety approaches are of particular value [12,15,80,87–89]. *B. cereus* and eight closely related species (*Bacillus pseudomycoides*, *B. anthracis*, *B. thuringiensis*, *Bacillus wiedmannii*, *Bacillus toyonensis*, *Bacillus weihenstephanensis*, *Bacillus mycoides*, and *Bacillus cytotoxicus*), which form the *B. cereus* group species complex are challenging in terms of food safety risk assessment due to their diverse pathogenic potential. Traditionally, only the species *B. cereus* (i.e., *B. cereus* sensu stricto [s.s.]) was considered as a foodborne pathogen, but several other species in the group were shown in the last decade to have pathogenic potential [87,90]. However, species classification among isolates representing the *B. cereus* group is extremely challenging as some species have been defined predominantly based on unique phenotypic characteristics; hence phylogenetic clustering is not consistent with phenotypic species definitions for a number of members of this group [12]. Accessibility of WGS helped in overcoming species definition issues in *B. cereus* group by enabling robust reconstruction of phylogenetic relationships that supported establishment of consistent phylogenetic clades [12,15] using genomic methods, such as ANIb, *in silico* DDH and/or core genome phylogeny [82]. These data also further supported that isolates classified into different *B. cereus* group species are genetically closely related and fall into the same clades (Fig. 3). For example, functional genomics approaches provided clear evidence that isolates classified as *B. thuringiensis* (an insecticidal biocontrol agent) and isolates classified as the foodborne pathogen *B. cereus* are genetically the same species [12,15,89]. Furthermore, isolates phenotypically classified into either of these species were found to carry genes encoding toxins linked to the ability to cause foodborne disease, suggesting that isolates representing both of these species may be able to cause foodborne disease in humans. These findings demonstrate the public health relevance of functional genomics studies in the area of food safety and suggest the need for improved assessment of *B. cereus* use as a bioinsecticide in food crops. These findings are consistent with previous studies that also reported human outbreaks associated with *B. thuringiensis* [91]. The recent studies on *B. cereus* discussed above demonstrate how genomics tools can be used to improve taxonomic and virulence classification, and guide precise experimental design to characterize virulence phenotypes of isolates representing bacterial groups that are potential food safety hazard. In cases of toxin-producing foodborne pathogens, such as *B. cereus*, which can cause foodborne intoxication, food safety can also greatly benefit from using metabolomics methods for detection of toxins [25].

5. Outline of a framework for a systems approach to omics-enabled precision food safety

Building on the recent advances in omics approaches to food safety, we propose a framework for a systems approach to omics-enabled food safety that will integrate genomic, metagenomic, phenotypic and epidemiological data, as well as risk assessments informed by these data to facilitate improved evidence-based food safety policies and procedures implemented and used by both

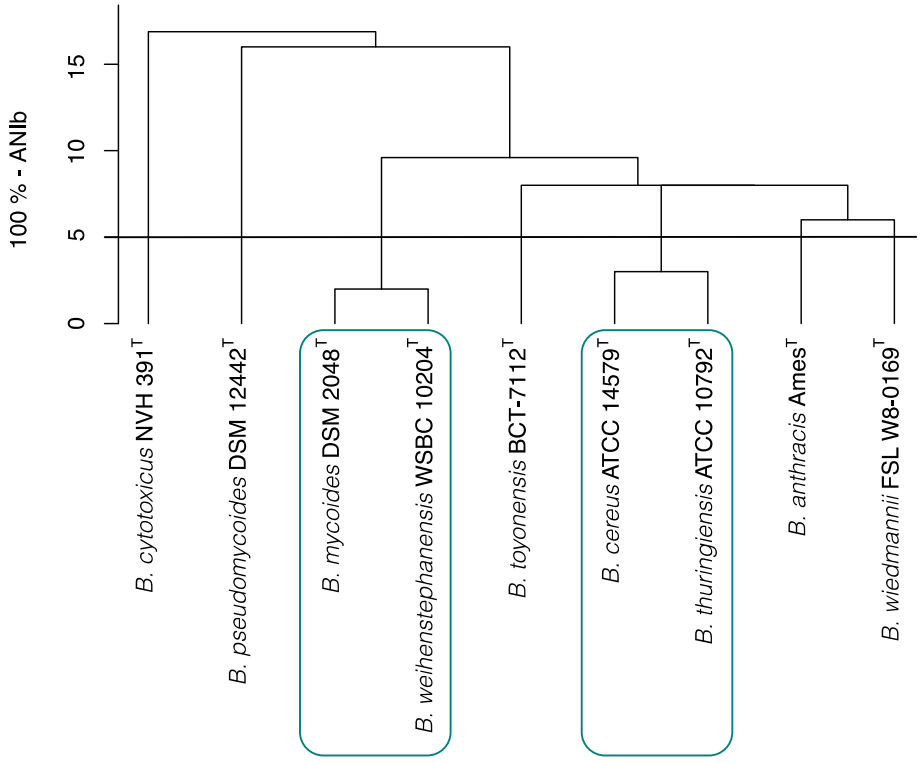

Fig. 3. Pairwise Average Nucleotide Identity BLAST (ANIb) values demonstrating genome-wide similarity among *B. cereus* group species type strains. ANIb value of 95% has been established as a genomic species cut-off that is equivalent to traditional 70% DNA-DNA hybridization cut-off [97]. As demonstrated in the UPGMA dendrogram constructed based on pairwise ANIb values, the types strains representing the species *B. cereus* and *B. thuringiensis*, as well as the types strains representing the species *B. weihenstephanensis* and *B. mycoides* fail to meet this criterion due to high genome-wide similarity. These findings show that at least some isolates currently classified as *B. cereus* and *B. thuringiensis* represent a single species. Similarly, isolates representing *B. weihenstephanensis* and *B. mycoides* likely also represent a single species.

industry and government agencies. This framework includes (i) genomics-based phylogenetic and taxonomic classification, (ii) genomics-based virulence and antimicrobial gene identification, (iii) functional genomics, including design of phenotypic pathogenicity experiments guided by genomics data, (iv) genomics-based epidemiological data, and (iv) integration of genomic, phenotypic and epidemiological data to support food safety risk assessment (Fig. 1). These inputs will allow for evidence-based formation and/or refinement of policies and procedures with genomics-based epidemiological monitoring of their effectiveness, as well as monitoring for new and emerging food safety hazards. New data obtained through this approach will provide a feedback loop that allows for continued food safety improvements through evidence-based decision-making. This includes improved risk assessment, policies and procedures resulting in increased food safety and security, and decreased morbidity and mortality attributed to foodborne diseases. The paragraphs below briefly outline the key components of this framework.

(i) *Genomics-based phylogenetic and taxonomic classification*. A first step in precision food safety efforts will often involve aligning phenotype-based taxonomy with phylogeny, which will facilitate accurate identification of species and subtypes and will also facilitate addressing situations where phenotypic species or subtype definitions do not match with phylogeny (e.g., *B. cereus* and *B. thuringiensis* species, polyphyletic *Salmonella* serotypes). The use of WGS is proposed for this purpose, as it provides superior phylogenetic resolution and allows for differentiation among closely related species that cannot be distinguished from each other based on 16S rDNA sequences. We propose that validation of currently established taxonomic species and subtypes includes identification of core genome SNPs that are not subjected to horizontal gene transfer and construction of core genome maximum likelihood phylogeny to identify phylogenetic groups. Furthermore, computing average nucleotide identity BLAST (ANIb $\leq$ 95% used as a species cut-off [92]) and *in silico* DNA-DNA hybridization (DDH $\leq$ 70% used as a species cut-off [93]) values will facilitate classification in cases where incongruences between taxonomy and phylogeny are identified.

(ii) *Genomics-based virulence and antimicrobial gene identification*. Following phylogeny-based and taxonomic classification, we propose to further characterize isolates by examining their genomic sequences for presence of virulence genes, genes conferring antimicrobial resistance, and other genes or gene functional categories that may be associated with specific phenotypes of interest. For phylogenetic groups (e.g., species or subtypes) where associations between relevant genes and phenotypes have been established, presence of these genes can be used as a factor to help predict food quality, food safety and/or public health risk.

(iii) *Functional genomics, including design of phenotypic pathogenicity experiments guided by genomics data*. In cases where associations between geno- and phenotypes have not yet been established, the information obtained from analyses of genomic data outlined in (i) and (ii) will guide hypothesis-driven design of laboratory experiments to identify and characterize virulence genes and other genes affecting infection and disease outcomes (e.g., antimicrobial resistance genes) at the transcriptional, proteomic and phenotypic level, including assessment of cytotoxicity and dose-response in tissue culture and animal models. This will allow for identification of species or subtypes of pathogens with higher virulence potential and allow for development of targeted rapid assays for detection of specific subtypes, enhancing pathogen detection and improving the clinical treatment decision-making.

(iv) *Genomics-based epidemiological data*. Surveillance data that link genomic data, including possible metagenomics data created on patient specimens and WGS data of pathogen isolates, with clinical data (e.g., disease symptoms, severity, clinical outcome) will represent a key data set that will facilitate precision food safety approaches. These types of data will not only allow for identification of previously unrecognized pathogens, but will also identify pathogen subtypes or genotypes associated with specific (e.g., more severe) disease symptoms or outcomes.

(v) *Integration of genomic, phenotypic and epidemiological data to perform risk assessment*. Consolidating data obtained through functional genomics approaches with epidemiological data obtained through surveillance, such as frequency of specific subtypes in clinical cases, patient conditions, comorbidities and food exposures will allow for food safety risk assessment on a subtype or genotype level as appropriate (Fig. 2).

6. Discussions

Genomics-based approaches to food safety are rapidly advancing our ability to detect food safety issues, particularly through WGS-based surveillance strategies. For example, routine implementation of WGS for characterization of *L. monocytogenes* from human clinical cases, food, and food associated sources, initiated in the U.S. in 2013, has considerably improved and enhanced detection of foodborne disease outbreaks, to a point where it is not uncommon to see outbreaks with two associated human cases identified and traced back to a specific food source [94]. Some even propose that these tools will facilitate trace-back of sporadic cases to specific food sources. However, use of WGS for detection of foodborne disease outbreaks and identification of outbreak sources requires integration of subtyping tools with epidemiological data in order to achieve the high level of precision needed for risk assessment on a pathogen subtype level. Importantly, WGS is rapidly being implemented as a routine tool to characterize human, food, and environmental isolates of all key foodborne pathogens, bringing a high level of precision to foodborne disease surveillance with anticipated broad implications. While metagenomics and metatranscriptomics approaches would typically not detect low-level contamination in food matrices, as detection of 1 CFU per 25 g of food via metagenomics tools is essentially unachievable, these tools are complementary to WGS and other rapid detection methods (e.g., immunoassay, PCR, flow cytometry, mass spectrometry/metabolomics) and will make increasingly important contributions to precision food safety. For example, metagenomics tools will facilitate detection of unculturable and previously unrecognized pathogens, particularly from human specimens where pathogen DNA would typically be present at high levels. Similarly, when combined with enrichment procedures, metagenomics tools will allow for improved detection of foodborne pathogens. Importantly, metagenomics and metatranscriptomics approaches allow for improved characterization of overall microbiota of a food and their application in this context will give a more precise picture of the microbial hygienic status of foods and processing plant environments. Shifts in microbiota composition detected using metagenomic and metatranscriptomic tools may be indicative of potential food safety issues, which can then be addressed before a given food product is released on the market. Furthermore, proteomics and metabolomics methods can be applied to detect microbial metabolites, including bacterial

toxins and mycotoxins, as well as other potential food safety hazards, such as antibiotic or pesticide residuals [25]. Overall, these applications may replace current hygienic monitoring methods, such as coliforms or *Enterobacteriaceae* testing, and provide for more precise detection of hygiene issues allowing for timely and more targeted corrective actions.

Application of genomics tools, including comprehensive data analyses, will provide important new insights into the biology and virulence of foodborne pathogens allowing for precise definitions of foodborne pathogens on a subtype level. While definition of what constitutes a foodborne pathogen, and hence a food safety hazard, may be increasingly less reliant on taxonomic units, genomics-based data provide unique opportunities for improved species definitions, which may be useful for some pathogens where virulence characteristics are clearly linked to a specific species. However, in many cases, genomics-based approaches may help define pathogens with criteria that do not rely solely on species definitions. For example, in the case of *B. cereus* and related organisms, a pathogen may be defined as a member of any of the species that represent *B. cereus* group and that carry a certain set of virulence genes. Additionally, members of this group that also carry genetic markers predictive of their ability to grow at low temperatures may be classified into a separate group considered a more severe hazard when present in refrigerated ready-to-eat foods that support growth of these organisms. In other cases, the definition of pathogen may be based solely on the presence of a certain set of genes in a given organism. While we appreciate that food safety policies that utilize these approaches comprehensively may take decades to materialize, these ideas will likely be implemented and utilized in different aspects of food safety in the near future, as supported by the fact that some firms already use molecular screens for a specific target gene or genes, such as *stx1* and *stx2*, for rational decision-making on utilization of different supply streams. Further enhancement of the use of omics tools and bioinformatics in food safety will require global scientific and regulatory collaborative efforts, such as those put forth by the Global Coalition for Regulatory Science Research [95,96].

7. Conclusion

We anticipate that the concept of precision food safety will rapidly gain momentum and support, particularly since some of the tools and approaches that fall under this concept are already being used in public health and industry, including routine use of WGS for foodborne disease surveillance in a number of countries. While many or some of the ideas detailed in this review may seem futuristic or unrealistic to some, continued rapid technological advances, including, but not limited to the development of rapid and more affordable sequencing technologies and sample preparation protocols will continue to move this field ahead. In addition, use of other precision food safety tools, such as GIS technologies, in conjunction with omics-based technologies is expected to have considerable impact. Importantly, precision food safety will become critical as we will need to address trade-offs between food safety and other human health impacts, such as negative environmental and food security impacts associated with recalls and destruction of food. In some of these cases the potential or measurable consequences of corrective actions may actually represent a larger human health risk that could outweigh extremely small public health risks that are being averted by the corrective actions.

Acknowledgments

We acknowledge the long-term support of dairy food safety research at Cornell from the New York State Dairy Promotion Board, representing New York farmers and their unwavering dedication and commitment to the quality and safety of milk and dairy products. Some of the work reviewed here was based upon work supported by the USDA National Institute of Food and Agriculture Hatch project NYC-143436. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the National Institute of Food and Agriculture (NIFA) or the United States Department of Agriculture (USDA).

References

[1] D.J.J. Giacometti, Foodomics in microbial safety, Trends Anal. Chem. 52 (2013) 16–22.
[2] D. Weller, S. Shiwakoti, P. Bergholz, Y. Grohn, M. Wiedmann, L.K. Strawn, Validation of a previously developed geospatial model that predicts the prevalence of *Listeria monocytogenes* in New York state produce fields, Appl. Environ. Microbiol. 82 (2016) 797–807.
[3] S.Y. Wang, D. Weller, J. Falardeau, L.K. Strawn, F.O. Mardones, A.D. Adell, A.I.M. Switt, Food safety trends: from globalization of whole genome sequencing to application of new tools to prevent foodborne diseases, Trends Food Sci. Tech. 57 (2016) 188–198.
[4] L.K. Strawn, E.D. Fortes, E.A. Bihn, K.K. Nightingale, Y.T. Grohn, R.W. Worobo, M. Wiedmann, P.W. Bergholz, Landscape and meteorological factors affecting prevalence of three food-borne pathogens in fruit and vegetable farms, Appl. Environ. Microbiol. 79 (2013) 588–600.
[5] L. Hashemi Beni, S. Villeneuve, D.I. LeBlanc, K. Cote, A. Fazil, A. Otten, R. McKellar, P. Delaquis, Spatio-temporal assessment of food safety risks in Canadian food distribution systems using GIS, Spat. Spatiotemporal Epidemiol. 3 (2012) 215–223.
[6] L.H. Beni, S. Villeneuve, D.I. LeBlanc, P. Delaquis, A GIS-based approach in support of an assessment of food safety risks, Trans. Gis 15 (2011) 95–108.
[7] M.W. Allard, E. Strain, D. Melka, K. Bunning, S.M. Musser, E.W. Brown, R. Timme, Practical value of food pathogen traceability through building a whole-genome sequencing network and database, J. Clin. Microbiol. 54 (2016) 1975–1983.
[8] H.C. den Bakker, M.W. Allard, D. Bopp, E.W. Brown, J. Fontana, Z. Iqbal, A. Kinney, R. Limberger, K.A. Musser, M. Shudt, E. Strain, M. Wiedmann, W.J. Wolfgang, Rapid whole-genome sequencing for surveillance of *Salmonella enterica* serovar Enteritidis, Emerg. Infect. Dis. 20 (2014) 1306–1314.
[9] E.K. Lienau, E. Strain, C. Wang, J. Zheng, A.R. Ottesen, C.E. Keys, T.S. Hammack, S.M. Musser, E.W. Brown, M.W. Allard, G. Cao, J. Meng, R. Stones, Identification of a salmonellosis outbreak by means of molecular sequencing, N. Engl. J. Med. 364 (2011) 981–982.
[10] M.W. Gilmour, M. Graham, G. Van Domselaar, S. Tyler, H. Kent, K.M. Trout-Yakel, O. Larios, V. Allen, B. Lee, C. Nadon, High-throughput genome sequencing of two *Listeria monocytogenes* clinical isolates during a large foodborne outbreak, BMC Genomics 11 (2010) 120.
[11] L. Grande, V. Michelacci, R. Bondi, F. Gigliucci, E. Franz, M.A. Badouei, S. Schlager, F. Minelli, R. Tozzoli, A. Caprioli, S. Morabito, Whole-genome characterization and strain comparison of VT2f-producing *Escherichia coli* causing hemolytic uremic syndrome, Emerg. Infect. Dis. 22 (2016) 2078–2086.
[12] J. Kovac, R.A. Miller, L.M. Carroll, D.J. Kent, J. Jian, S.M. Beno, M. Wiedmann, Production of hemolysin BL by *Bacillus cereus* group isolates of dairy origin is associated with whole-genome phylogenetic clade, BMC Genomics 17 (2016) 581.
[13] R.L. Lindsey, H. Pouseele, J.C. Chen, N.A. Strockbine, H.A. Carleton, Implementation of Whole Genome Sequencing (WGS) for identification and characterization of Shiga toxin-producing *Escherichia coli* (STEC) in the United States, Front. Microbiol. 7 (2016) 766.
[14] M.J. Jansen van Rensburg, C. Swift, A.J. Cody, C. Jenkins, M.C. Maiden, Exploiting bacterial whole-genome sequencing data for evaluation of diagnostic assays: *Campylobacter* species identification as a case study, J. Clin. Microbiol. 54 (2016) 2882–2890.
[15] M.E. Bohm, C. Huptas, V.M. Krey, S. Scherer, Massive horizontal gene transfer, strictly vertical inheritance and ancient duplications differentially shape the evolution of *Bacillus cereus* enterotoxin operons *hbl, cytK* and *nhe*, BMC Evol. Biol. 15 (2015) 246.
[16] L.D. Rodriguez-Rivera, A.I. Moreno Switt, L. Degoricija, R. Fang, C.A. Cummings, M.R. Furtado, M. Wiedmann, H.C. den Bakker, Genomic characterization of *Salmonella* Cerro ST367, an emerging *Salmonella* subtype in cattle in the United States, BMC Genomics 15 (2014) 427.
[17] O. Lukjancenko, T.M. Wassenaar, D.W. Ussery, Comparison of 61 sequenced *Escherichia coli* genomes, Microb. Ecol. 60 (2010) 708–720.
[18] J. Kovac, Kevin J. Cummings, Lorraine D. Rodriguez-Rivera, Laura M. Carroll, Anil Thachil, Martin Wiedmann, Temporal genomic phylogeny reconstruction indicates a geospatial transmission path of *Salmonella* Cerro in the United States and a clade-specific loss of hydrogen sulfide production, Front. Microbiol. 8 (2017) 737.
[19] D.F. Nieuwenhuijse, M.P. Koopmans, Metagenomic sequencing for surveillance of food- and waterborne viral diseases, Front. Microbiol. 8 (2017) 230.

[20] X. Yang, N.R. Noyes, E. Doster, J.N. Martin, L.M. Linke, R.J. Magnuson, H. Yang, I. Geornaras, D.R. Woerner, K.L. Jones, J. Ruiz, C. Boucher, P.S. Morley, K.E. Belk, Use of metagenomic shotgun sequencing technology to detect foodborne pathogens within the microbiome of the beef production chain, Appl. Environ. Microbiol. 82 (2016) 2433–2443.
[21] K.G. Jarvis, J.R. White, C.J. Grim, L. Ewing, A.R. Ottesen, J.J. Beaubrun, J.B. Pettengill, E. Brown, D.E. Hanes, Cilantro microbiome before and after nonselective pre-enrichment for *Salmonella* using 16S rRNA and metagenomic sequencing, BMC Microbiol. 15 (2015) 160.
[22] S.R. Leonard, M.K. Mammel, D.W. Lacher, C.A. Elkins, Application of metagenomic sequencing to food safety: detection of Shiga toxin-producing *Escherichia coli* on fresh bagged spinach, Appl. Environ. Microbiol. 81 (2015) 8183–8191.
[23] A.R. Ottesen, A. Gonzalez, R. Bell, C. Arce, S. Rideout, M. Allard, P. Evans, E. Strain, S. Musser, R. Knight, E. Brown, J.B. Pettengill, Co-enriching microflora associated with culture based methods to detect *Salmonella* from tomato phyllosphere, PLoS One 8 (2013) e73079.
[24] N.R. Noyes, X. Yang, L.M. Linke, R.J. Magnuson, A. Dettenwanger, S. Cook, I. Geornaras, D.E. Woerner, S.P. Gow, T.A. McAllister, H. Yang, J. Ruiz, K.L. Jones, C.A. Boucher, P.S. Morley, K.E. Belk, Resistome diversity in cattle and the environment decreases during beef production, Elife 5 (2016) e13195.
[25] Yong-Jiang Xu, Chengshu Wang, Wanxing Eugene Ho, Choon Nam Ong, Recent developments and applications of metabolomics in microbiological investigations, Trends Anal. Chem. 56 (2014) 37–48.
[26] W.R. Russell, S.H. Duncan, Advanced analytical methodologies to study the microbial metabolome of the human gut, Trends Anal. Chem. 52 (2013) 54–60.
[27] L.S. Katz, A. Petkau, J. Beaulaurier, S. Tyler, E.S. Antonova, M.A. Turnsek, Y. Guo, S. Wang, E.E. Paxinos, F. Orata, L.M. Gladney, S. Stroika, J.P. Folster, L. Rowe, M.M. Freeman, N. Knox, M. Frace, J. Boncy, M. Graham, B.K. Hammer, Y. Boucher, A. Bashir, W.P. Hanage, G. Van Domselaar, C.L. Tarr, Evolutionary dynamics of *Vibrio cholerae* O1 following a single-source introduction to Haiti, MBio 4 (2013).
[28] N.J. Croucher, S.R. Harris, C. Fraser, M.A. Quail, J. Burton, M. van der Linden, L. McGee, A. von Gottberg, J.H. Song, K.S. Ko, B. Pichon, S. Baker, C.M. Parry, L.M. Lambertsen, D. Shahinas, D.R. Pillai, T.J. Mitchell, G. Dougan, A. Tomasz, K.P. Klugman, J. Parkhill, W.P. Hanage, S.D. Bentley, Rapid pneumococcal evolution in response to clinical interventions, Science 331 (2011) 430–434.
[29] G. Morelli, Y. Song, C.J. Mazzoni, M. Eppinger, P. Roumagnac, D.M. Wagner, M. Feldkamp, B. Kusecek, A.J. Vogler, Y. Li, Y. Cui, N.R. Thomson, T. Jombart, R. Leblois, P. Lichtner, L. Rahalison, J.M. Petersen, F. Balloux, P. Keim, T. Wirth, J. Ravel, R. Yang, E. Carniel, M. Achtman, *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity, Nat. Genet. 42 (2010) 1140–1143.
[30] N.J. Loman, R.V. Misra, T.J. Dallman, C. Constantinidou, S.E. Gharbia, J. Wain, M.J. Pallen, Performance comparison of benchtop high-throughput sequencing platforms, Nat. Biotechnol. 30 (2012) 434–439.
[31] B.R. Jackson, C. Tarr, E. Strain, K.A. Jackson, A. Conrad, H. Carleton, L.S. Katz, S. Stroika, L.H. Gould, R.K. Mody, B.J. Silk, J. Beal, Y. Chen, R. Timme, M. Doyle, A. Fields, M. Wise, G. Tillman, S. Defibaugh-Chavez, Z. Kucerova, A. Sabol, K. Roache, E. Trees, M. Simmons, J. Wasilenko, K. Kubota, H. Pouseele, W. Klimke, J. Besser, E. Brown, M. Allard, P. Gerner-Smidt, Implementation of nationwide real-time whole-genome sequencing to enhance Listeriosis outbreak detection and investigation, Clin. Infect. Dis. 63 (2016) 380–386.
[32] M.C. Maiden, J.A. Bygraves, E. Feil, G. Morelli, J.E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D.A. Caugant, I.M. Feavers, M. Achtman, B.G. Spratt, Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 3140–3145.
[33] M. Inouye, H. Dashnow, L.A. Raven, M.B. Schultz, B.J. Pope, T. Tomita, J. Zobel, K.E. Holt, SRST2: rapid genomic surveillance for public health and hospital microbiology labs, Genome Med. 6 (2014) 90.
[34] A. Weimann, K. Mooren, J. Frank, P.B. Pope, A. Bremges, A.C. McHardy, From genomes to phenotypes: traitar, the microbial trait analyzer, mSystems 1 (2016).
[35] E.L. van Dijk, H. Auger, Y. Jaszczyszyn, C. Thermes, Ten years of next-generation sequencing technology, Trends Genet. 30 (2014) 418–426.
[36] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (2012) 357–359.
[37] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics 25 (2009) 1754–1760.
[38] D.R. Zerbino, E. Birney, Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs, Genome Res. 18 (2008) 821–829.
[39] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (2012) 455–477.
[40] C.S. Chin, D.H. Alexander, P. Marks, A.A. Klammer, J. Drake, C. Heiner, A. Clum, A. Copeland, J. Huddleston, E.E. Eichler, S.W. Turner, J. Korlach, Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data, Nat. Methods 10 (2013) 563–569.
[41] M. de Been, M. Pinholt, J. Top, S. Bletz, A. Mellmann, W. van Schaik, E. Brouwer, M. Rogers, Y. Kraat, M. Bonten, J. Corander, H. Westh, D. Harmsen, R.J. Willems, Core genome multilocus sequence typing scheme for high-resolution typing of *Enterococcus faecium*, J. Clin. Microbiol. 53 (2015) 3788–3797.
[42] W. Ruppitsch, A. Pietzka, K. Prior, S. Bletz, H.L. Fernandez, F. Allerberger, D. Harmsen, A. Mellmann, Defining and evaluating a core genome multilocus sequence typing scheme for whole-genome sequence-based typing of *Listeria monocytogenes*, J. Clin. Microbiol. 53 (2015) 2869–2876.
[43] A. Moura, A. Criscuolo, H. Pouseele, M.M. Maury, A. Leclercq, C. Tarr, J.T. Bjorkman, T. Dallman, A. Reimer, V. Enouf, E. Larsonneur, H. Carleton, H. Bracq-Dieye, L.S. Katz, L. Jones, M. Touchon, M. Tourdjman, M. Walker, S. Stroika, T. Cantinelli, V. Chenal-Francisque, Z. Kucerova, E.P. Rocha, C. Nadon, K. Grant, E.M. Nielsen, B. Pot, P. Gerner-Smidt, M. Lecuit, S. Brisse, Whole genome-based population biology and epidemiological surveillance of *Listeria monocytogenes*, Nat. Microbiol. 2 (2016) 16185.
[44] S. Davis, James B. Pettengill, Yan Luo, Justin Payne, Al Shpuntoff, Hugh Rand, Errol Strain, CFSAN SNP pipeline: an automated method for constructing SNP matrices from next-generation sequence data, PeerJ Comput. Sci. 1 (2015) e20.
[45] L.S. Katz, T. Griswold, A.J. Williams-Newkirk, D. Wagner, A. Petkau, C. Sieffert, G. Van Domselaar, X. Deng, H.A. Carleton, A comparative analysis of the Lyve-Set phylogenomics pipeline for genomic epidemiology of foodborne pathogens, Front. Microbiol. 8 (2017).
[46] S. Duchêne, K.E. Holt, F.-X. Weill, S. Le Hello, J. Hawkey, D.J. Edwards, M. Fourment, E.C. Holmes, Genome-scale rates of evolutionary change in bacteria, Microb. Genomics 2 (2016).
[47] D. Lambert, A. Pightling, E. Griffiths, G. Van Domselaar, P. Evans, S. Berthelet, D. Craig, P.S. Chandry, R. Stones, F. Brinkman, A. Angers-Loustau, J. Kreysa, W. Tong, B. Blais, Baseline practices for the application of genomic data supporting regulatory food safety, J. AOAC Int. 100 (2017) 721–731.
[48] R. Ranjan, A. Rani, A. Metwally, H.S. McGee, D.L. Perkins, Analysis of the microbiome: advantages of whole genome shotgun versus 16S amplicon sequencing, Biochem. Biophys. Res. Commun. 469 (2016) 967–977.
[49] J. Kuczynski, C.L. Lauber, W.A. Walters, L.W. Parfrey, J.C. Clemente, D. Gevers, R. Knight, Experimental and analytical tools for studying the human microbiome, Nat. Rev. Genet. 13 (2011) 47–58.
[50] F. De Filippis, E. Parente, D. Ercolini, Metagenomics insights into food fermentations, Microb. Biotechnol. 10 (2017) 91–102.
[51] G.R. Feehery, E. Yigit, S.O. Oyola, B.W. Langhorst, V.T. Schmidt, F.J. Stewart, E.T. Dimalanta, L.A. Amaral-Zettler, T. Davis, M.A. Quail, S. Pradhan, A method for selectively enriching microbial DNA from contaminating vertebrate host DNA, PLoS One 8 (2013) e76096.
[52] G. Kergourlay, B. Taminiau, G. Daube, M.C. Champomier Verges, Metagenomic insights into the dynamics of microbial communities in food, Int. J. Food Microbiol. 213 (2015) 31–39.
[53] D. Ercolini, High-throughput sequencing and metagenomics: moving forward in the culture-independent analysis of food microbial ecology, Appl. Environ. Microbiol. 79 (2013) 3148–3155.
[54] T.S. Lusk, A.R. Ottesen, J.R. White, M.W. Allard, E.W. Brown, J.A. Kase, Characterization of microflora in Latin-style cheeses by next-generation sequencing technology, BMC Microbiol. 12 (2012) 254.
[55] G. Stellato, A. La Storia, F. De Filippis, G. Borriello, F. Villani, D. Ercolini, Overlap of spoilage-associated microbiota between meat and the meat processing environment in small-scale and large-scale retail distributions, Appl. Environ. Microbiol. 82 (2016) 4045–4054.
[56] J. Hultman, R. Rahkila, J. Ali, J. Rousu, K.J. Bjorkroth, Meat processing plant microbiome and contamination patterns of cold-tolerant bacteria causing food safety and spoilage risks in the manufacture of vacuum-packaged cooked sausages, Appl. Environ. Microbiol. 81 (2015) 7088–7097.
[57] P. Wolffs, B. Norling, P. Radstrom, Risk assessment of false-positive quantitative real-time PCR results in food, due to detection of DNA originating from dead cells, J. Microbiol. Methods 60 (2005) 315–323.
[58] R.L. Bell, K.G. Jarvis, A.R. Ottesen, M.A. McFarland, E.W. Brown, Recent and emerging innovations in *Salmonella* detection: a food and environmental perspective, Microb. Biotechnol. 9 (2016) 279–292.
[59] J. Czajka, N. Bsat, M. Piani, W. Russ, K. Sultana, M. Wiedmann, R. Whitaker, C.A. Batt, Differentiation of *Listeria monocytogenes* and *Listeria innocua* by 16S rRNA genes and intraspecies discrimination of *Listeria monocytogenes* strains by random amplified polymorphic DNA polymorphisms, Appl. Environ. Microbiol. 59 (1993) 304–308.
[60] J. Wang, H. Jia, Metagenome-wide association studies: fine-mining the microbiome, Nat. Rev. Microbiol. 14 (2016) 508–522.
[61] T.J. Sharpton, An introduction to the analysis of shotgun metagenomic data, Front. Plant Sci. 5 (2014) 209.
[62] S.V. Lynch, O. Pedersen, The human intestinal microbiome in health and disease, N. Engl. J. Med. 375 (2016) 2369–2379.
[63] L. Quigley, D.J. O'Sullivan, D. Daly, O. O'Sullivan, Z. Burdikova, R. Vana, T.P. Beresford, R.P. Ross, G.F. Fitzgerald, P.L. McSweeney, L. Giblin, J.J. Sheehan, P.D. Cotter, *Thermus* and the pink discoloration defect in cheese, mSystems 1 (2016).
[64] S. Bashiardes, G. Zilberman-Schapira, E. Elinav, Use of metatranscriptomics in microbiome research, Bioinform. Biol. Insig. 10 (2016) 19–25.
[65] W. Alkema, J. Boekhorst, M. Wels, S.A. van Hijum, Microbial bioinformatics for food safety and production, Brief. Bioinform. 17 (2016) 283–292.
[66] A. Valdés, C. Ibáñez, C. Simó, V. García-Cañas, Recent transcriptomics advances and emerging applications in food science, Trends Anal. Chem. 52 (2013) 142–154.

[67] M.H. Lessard, C. Viel, B. Boyle, D. St-Gelais, S. Labrie, Metatranscriptome analysis of fungal strains *Penicillium camemberti* and *Geotrichum candidum* reveal cheese matrix breakdown and potential development of sensory properties of ripened Camembert-type cheese, BMC Genomics 15 (2014) 235.
[68] F. De Filippis, A. Genovese, P. Ferranti, J.A. Gilbert, D. Ercolini, Metatranscriptomics reveals temperature-driven functional changes in microbiome impacting cheese maturation rate, Sci. Rep. 6 (2016) 21871.
[69] C. Monnet, E. Dugat-Bony, D. Swennen, J.M. Beckerich, F. Irlinger, S. Fraud, P. Bonnarme, Investigation of the activity of the microorganisms in a reblochon-style cheese by metatranscriptomic analysis, Front. Microbiol. 7 (2016) 536.
[70] J. Xu, S. Thakkar, B. Gong, W. Tong, The FDA's experience with emerging genomics technologies-past, present, and future, AAPS J. 18 (2016) 814–818.
[71] S.B. Edlund, K.L. Beck, N. Haiminen, L.P. Parida, D.B. Storey, B.C. Weimer, J.H. Kaufman, D.D. Chambliss, Design of the MCAW compute service for food safety bioinformatics, Ibm J. Res. Dev. 60 (2016).
[72] D.A. Buzatu, T.J. Moskal, A.J. Williams, W.M. Cooper, W.B. Mattes, J.G. Wilkes, An integrated flow cytometry-based system for real-time, high sensitivity bacterial detection and identification, PLoS One 9 (2014) e94254.
[73] G.H. Tyson, S. Zhao, C. Li, S. Ayers, J.L. Sabo, C. Lam, R.A. Miller, P.F. McDermott, Establishing genotypic cutoff values to measure antimicrobial resistance in *Salmonella*, Antimicrob. Agents Chemother. 61 (2017).
[74] P.F. McDermott, G.H. Tyson, C. Kabera, Y. Chen, C. Li, J.P. Folster, S.L. Ayers, C. Lam, H.P. Tate, S. Zhao, Whole-genome sequencing for detecting antimicrobial resistance in nontyphoidal *Salmonella*, Antimicrob. Agents Chemother. 60 (2016) 5515–5520.
[75] G.H. Tyson, P.F. McDermott, C. Li, Y. Chen, D.A. Tadesse, S. Mukherjee, S. Bodeis-Jones, C. Kabera, S.A. Gaines, G.H. Loneragan, T.S. Edrington, M. Torrence, D.M. Harhay, S. Zhao, WGS accurately predicts antimicrobial resistance in *Escherichia coli*, J. Antimicrob. Chemother. 70 (2015) 2763–2769.
[76] S. Zhao, G.H. Tyson, Y. Chen, C. Li, S. Mukherjee, S. Young, C. Lam, J.P. Folster, J.M. Whichard, P.F. McDermott, Whole-genome sequencing analysis accurately predicts antimicrobial resistance phenotypes in *Campylobacter* spp, Appl. Environ. Microbiol. 82 (2015) 459–466.
[77] T.F. Jones, L.A. Ingram, P.R. Cieslak, D.J. Vugia, M. Tobin-D'Angelo, S. Hurd, C. Medus, A. Cronquist, F.J. Angulo, Salmonellosis outcomes differ substantially by serotype, J. Infect. Dis. 198 (2008) 109–114.
[78] R.A. Miller, M. Wiedmann, Dynamic duo-the *Salmonella* cytolethal distending toxin combines ADP-Ribosyl transferase and nuclease activities in a novel form of the cytolethal distending toxin, Toxins (Basel) 8 (2016).
[79] R.A. Miller, M. Wiedmann, The cytolethal distending toxin produced by nontyphoidal *Salmonella* Serotypes Javiana, Montevideo, Oranienburg, and Mississippi induces DNA damage in a manner similar to that of serotype Typhi, MBio 7 (2016).
[80] N. Jessberger, V.M. Krey, C. Rademacher, M.E. Bohm, A.K. Mohr, M. Ehling-Schulz, S. Scherer, E. Martlbauer, From genome to toxicity: a combinatory approach highlights the complexity of enterotoxin production in *Bacillus cereus*, Front. Microbiol. 6 (2015) 560.
[81] K. Goswami, C. Chen, L. Xiaoli, K.A. Eaton, E.G. Dudley, Coculture of *Escherichia coli* O157:H7 with a nonpathogenic *E. coli* strain increases toxin production and virulence in a germfree mouse model, Infect. Immun. 83 (2015) 4185–4193.
[82] R.A. Miller, S.M. Beno, D.J. Kent, L.M. Carroll, N.H. Martin, K.J. Boor, J. Kovac, *Bacillus wiedmannii* sp. nov., a psychrotolerant and cytotoxic *Bacillus cereus* group species isolated from dairy foods and dairy environments, Int. J. Syst. Evol. Microbiol. 66 (2016) 4744–4753.
[83] S. Issenhuth-Jeanjean, P. Roggentin, M. Mikoleit, M. Guibourdenche, E. de Pinna, S. Nair, P.I. Fields, F.X. Weill, Supplement 2008-2010 (no. 48) to the White-Kauffmann-Le Minor scheme, Res. Microbiol. 165 (2014) 526–530.
[84] L.D. Rodriguez-Rivera, B.M. Bowen, H.C. den Bakker, G.E. Duhamel, M. Wiedmann, Characterization of the cytolethal distending toxin (typhoid toxin) in non-typhoidal *Salmonella* serovars, Gut Pathog. 7 (2015) 19.
[85] H.C. den Bakker, A.I. Moreno Switt, G. Govoni, C.A. Cummings, M.L. Ranieri, L. Degoricija, K. Hoelzer, L.D. Rodriguez-Rivera, S. Brown, E. Bolchacova, M.R. Furtado, M. Wiedmann, Genome sequencing reveals diversification of virulence factor content and possible host adaptation in distinct subpopulations of *Salmonella enterica*, BMC Genomics 12 (2011) 425.
[86] L.M. Carroll, Martin Wiedmann, Henk den Bakker, Julie Siler, Steven Warchocki, David Kent, Svetlana Lyalina, Margaret Davis, William Sischo, Thomas Besser, Lorin Warnick, Richard Pereira, Whole-genome sequencing of drug-resistant *Salmonella enterica* isolated from dairy cattle and humans in New York and Washington states reveals source and geographic associations, Appl. Environ. Microbiol. (June 6, 2017) [in press].
[87] R.A. Miller, Jiahui Jian, Sarah M. Beno, Martin Wiedmann, Jasna Kovac, Molecular, phenotypic, and cytotoxic profiles indicate that multiple *Bacillus cereus* group species may cause foodborne illness, Appl. Environ. Microbiol. (2017) [in press].
[88] M.E. Bohm, V.M. Krey, N. Jessberger, E. Frenzel, S. Scherer, Comparative bioinformatics and experimental analysis of the intergenic regulatory regions of *Bacillus cereus hbl* and *nhe* enterotoxin operons and the impact of cody on virulence heterogeneity, Front. Microbiol. 7 (2016) 768.
[89] K. Zhu, C.S. Holzel, Y. Cui, R. Mayer, Y. Wang, R. Dietrich, A. Didier, R. Bassitta, E. Martlbauer, S. Ding, Probiotic *Bacillus* cereus strains, a potential risk for public health in China, Front. Microbiol. 7 (2016) 718.
[90] M.H. Guinebretiere, P. Velge, O. Couvert, F. Carlin, M.L. Debuyser, C. Nguyen-The, Ability of *Bacillus cereus* group strains to cause food poisoning varies according to phylogenetic affiliation (groups I to VII) rather than species affiliation, J. Clin. Microbiol. 48 (2010) 3388–3391.
[91] L. McIntyre, K. Bernard, D. Beniac, J.L. Isaac-Renton, D.C. Naseby, Identification of *Bacillus cereus* group species associated with food poisoning outbreaks in British Columbia, Canada, Appl. Environ. Microbiol. 74 (2008) 7451–7453.
[92] M. Richter, R. Rossello-Mora, Shifting the genomic gold standard for the prokaryotic species definition, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 19126–19131.
[93] A.F. Auch, M. von Jan, H.P. Klenk, M. Goker, Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison, Stand Genomic Sci. 2 (2010) 117–134.
[94] CDC, The *Listeria* Whole Genome Sequencing Project, 2016.
[95] M.J. Healy, W. Tong, S. Ostroff, H.G. Eichler, A. Patak, M. Neuspiel, H. Deluyker, W. Slikker Jr., Regulatory bioinformatics for food and drug safety, Regul. Toxicol. Pharmacol. 80 (2016) 342–347.
[96] W. Tong, S. Ostroff, B. Blais, P. Silva, M. Dubuc, M. Healy, W. Slikker, Genomics in the land of regulatory science, Regul. Toxicol. Pharmacol. 72 (2015) 102–106.
[97] J.Z. Chan, M.R. Halachev, N.J. Loman, C. Constantinidou, M.J. Pallen, Defining bacterial species in the genomic era: insights from the genus *Acinetobacter*, BMC Microbiol. 12 (2012) 302.