Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Genomics of foodborne pathogens for microbial food safety

Marc W Allard, Rebecca Bell, Christina M Ferreira, Narjol Gonzalez-Escalona, Maria Hoffmann, Tim Muruvanda, Andrea Ottesen, Padmini Ramachandran, Elizabeth Reed, Shashi Sharma, Eric Stevens, Ruth Timme, Jie Zheng and Eric W Brown

Whole genome sequencing (WGS) has been broadly used to provide detailed characterization of foodborne pathogens. These genomes for diverse species including *Salmonella*, *Escherichia coli*, *Listeria*, *Campylobacter* and *Vibrio* have provided great insight into the genetic make-up of these pathogens. Numerous government agencies, industry and academia have developed new applications in food safety using WGS approaches such as outbreak detection and characterization, source tracking, determining the root cause of a contamination event, profiling of virulence and pathogenicity attributes, antimicrobial resistance monitoring, quality assurance for microbiology testing, as well as many others. The future looks bright for additional applications that come with the new technologies and tools in genomics and metagenomics.

Address
Food and Drug Administration, Office of Regulatory Science, Division of Microbiology, HFS-712, 5001 Campus Drive, College Park, MD 20740, USA

Corresponding author: Allard, Marc W (Marc.Allard@fda.hhs.gov), Allard, Marc W (Marc.Allard@fda.hhs.gov)

Current Opinion in Biotechnology 2018, **49**:224–229

This review comes from a themed issue on **Food biotechnology**

Edited by **Maria Marco** and **Eddy Smid**

For a complete overview see the Issue and the Editorial

Available online 21st November 2017

https://doi.org/10.1016/j.copbio.2017.11.002

0958-1669/Published by Elsevier Ltd.

Introduction

Applications of whole genome sequencing (WGS) for the characterization of foodborne pathogens are relatively recent phenomena. Collecting the entire genome of a foodborne pathogen was rapidly adopted as the ultimate characterization tool for describing what makes up a pathogen. Because genomic data is inherited vertically, from cell to daughter cell, these data can be used to reconstruct the evolutionary history of these pathogens. Phylogenetics is a powerful tool used for many applications in foodborne outbreak detection and source tracking [1••,2••,3••]. The field started with case studies of past outbreaks demonstrating the feasibility of WGS and phylogenetics to provide useful information that could support outbreak investigations [4,5•,6••,7,8•,9–11]. These retrospective studies laid a solid foundation for our current real-time investigations using this technology. The importance of genomic data in understanding foodborne pathogens and the integration of the field cannot be overstated. The common genetic code is shared across all pathogens regardless of the source of the isolates or sequenced technology used. This allows public health laboratories to build local and global integrated networks all utilizing and consuming the same raw data [12••,13,14•,15,16•,17•,18].

The objective nature of the genetic data makes comparisons fully transparent, allowing for broad sharing of data across the affected disciplines. The digitization of this genetic information allows the world to enable Dr. David Lipman's (founder and director of NCBI through 2017) vision of a digital immune system whereby all of the genetic information about pathogens is collected and shared through public access. Where everyone has the same information at the same time to solve common public health problems. The data allow analyses to be directly scalable so that the growth of these databases is exponential and show no signs of decreasing. As new genomic data accumulate and are better characterized, new applications for this data will be discovered and applied to real-world public health problems. Our goals are to build a global one health WGS database where human pathogens are rapidly characterized and linked to closely related isolates. This rapid characterization and linkage of pathogens will speed up investigations and ultimately improve food safety and public health.

GenomeTrakr/PulseNet are WGS networks to combine environmental and clinical genomes for a one health perspective

Pulse Field Gel Electrophoresis (PFGE) technology is rapidly being replaced by WGS methods. The Food and Drug Administration (US FDA) in combination with the

National Center for Biotechnology Information (NCBI) developed the first distributed network of state and federal laboratories using WGS called GenomeTrakr that was then rapidly combined with US Centers for Disease Control and Prevention (US CDC) and US Department of Agriculture (USDA) WGS laboratories [19••,20••]. The genomic data from US surveillance efforts are uploaded immediately after data collection and shared publicly at the NCBI Pathogen Detection site (NCBI URL: https://www.ncbi.nlm.nih.gov/pathogens/) where NCBI provides daily clusters, in the form of phylogenetic trees, that determines the closest matches to newly submitted data. The genetic relatedness revealed by the phylogenies represents potential links between food, environmental and clinical isolates. A genetic 'match' is a hypothesis of shared ancestry but legally a match is not enough for government action. FDA does not move forward with any regulatory actions without separate investigative evidence that products or food facilities are contaminated. Similarly, food exposure discovered through epidemiological investigation is also used to support the genomic links between food, environmental and clinical isolates. To date FDA is sequencing all isolates collected and available in their frozen culture collections spanning over 20 years of curation. Additionally, all positive isolates collected by FDA through surveillance and inspection are sequenced and uploaded to the GenomeTrakr database. The early linkages from WGS data are important physical evidence that helps speed up investigations and targets the limited resources of the federal laboratories. The results of sharing data in real-time are that outbreaks are being recognized earlier, and many more, smaller clusters are being identified [17•,20••].

Power of combining geospatial information on pathogen strains with genomic information for attribution of particular genotypes during outbreaks and other contamination events

With global economies and the daily import and export of foodstuffs, traditional tracebacks may take weeks to accomplish and contaminated foods may remain on the shelves for consumers to purchase, or large recalls of uncontaminated food may be initiated in order to prevent illnesses. None of these scenarios is ideal. Faster source-tracking that hones in on specific food from a specific production facility, packinghouse or agricultural field is desirable. The power of WGS combined with phylogenetics and comprehensive metadata on each isolate makes faster and more accurate source-tracking achievable. In a retrospective study of the large shell-egg outbreak of 2012, WGS and phylogenetics was able to clearly differentiate illnesses and attribute them to several farms [3••]. Further, in Hoffman *et al.* [6••] these analyses were able to pinpoint the origin of the contaminated tuna scrape responsible for over 400 *Salmonella enterica* subsp. *enterica* serovar Bareilly (*S.* Bareilly), illnesses. It has been demonstrated that with a comprehensive database populated with relevant metadata (such as location, collection date, food product etc.) this tool is able to single out exact field or farms where contamination occurred [21–23]. This accuracy and certainty in what is a match has improved source tracking in linking contaminated foods and understanding the root cause of the outbreak or contamination event.

All FDA regulatory actions are subject to a full review to determine the root cause of the outbreak. This allows us to fully understand how the foodborne pathogen managed to evade any preventative controls in place. Trace forward and trace backward investigations should uncover the root cause during the outbreak investigation. Once the root is determined then a forward investigation is conducted to find all of the localities that received contaminated product. Thus investigations often become prolonged depending on the complexity of food distribution chain. While the physical inspection of food facilities is taking place, epidemiological questionnaires are given to anyone who shares a common WGS genotype. If the exposure evidence and field inspection evidence do not coincide then this suggests that additional foods are part of a larger food contamination network by being indirectly connected to the common clonal WGS pathogen. Genomics can quickly show potential linkage and launch investigations based on genetic relatedness, but independent evidence, of exposure and/or documentation of a facility or food contamination, is required by FDA to conduct regulatory actions. Any new discoveries about mechanisms of contamination is further tested and then added to good manufacturing or good farming practice guidelines (GMPs GFPs) so that all producers and manufactures of that food class can learn from the post-investigations.

Toward a single microbiology workflow for pathogen characterization

WGS-based serotyping and automated AMR typing dashboard tools

Traditional microbiological tests like serotyping and phage typing are cumbersome, but genetic and genomic alternatives are being developed [24]. The transformation to a single microbiological workflow based on WGS data requires building smart databases that include both genotype and phenotype of well-characterized pathogens. Antimicrobial resistance (AMR) tests are collected to support the National Antimicrobial Monitoring System [25•]. Several online tools include both antimicrobial resistance reference gene database as well as automatic tools to call the presence of the AMR genotypes. These tools are available at NCBI Pathogen Detection, and DTU CGE [26•,27•].

WGS-based profiling of virulence and pathogenicity attributes

Other genotype to phenotype predictions that should be developed include identifying genes for pathogenicity