Food Factory Genomics: Where Big Data Drives Quality and Food Safety

Sapna Chitlapilly Dass and Angela Anandappa*

Alliance for Advanced Sanitation, Dept. of Food Science and Technology, University of Nebraska-Lincoln, Food Innovation Center, Lincoln, NE 68588, USA

SUMMARY

Recent advances in tools for use in molecular genetics have opened new doors for developing safer foods. More information is available not only about the essence of the composition of our foods but also about the genetics of the unseen microbial world that influences the safety and quality of our food. Microbial analysis is an integral part of maintaining food quality and safety from farm to table. Isolation, enumeration and cultural methods have long been the pillars of the methodology that food microbiologists use to measure quality and safety parameters. Although they are powerful tools, they are time consuming and cumbersome, making it difficult to keep pace with the rapid advances in molecular genetics. Moreover, the need for rapid turnaround in pathogen detection, especially with regard to potential production failures, can be expensive when business processes are slowed by the observation time for a microbial colony's growth. Genomic sequencing technologies have revolutionized the availability of information, giving depth to our understanding of the microbiological world, and food factories are ripe for using this information to use their testing budgets more efficiently. We outline some of the principles of genomic tools that might benefit the intelligent factory of the future.

OVERVIEW

Food and food processing environments harbor complex microbial communities composed of bacteria and fungi. Some of these microorganisms can have undesirable effects on food quality (spoilage microorganisms) and safety (pathogenic microorganisms). The traditional way of deciphering the microbiota associated with food is by utilizing culture-based techniques, which are based on growing organisms on general or selective media, visualizing growth and enumerating the viable microbial colonies. Microorganisms that are slow growers, uncultivable or present in low numbers can be outcompeted by numerically more abundant species, impeding their detection in culture *(2, 20)*. Issues associated with culture dependency can be overcome by genomic sequencing, which allows identification based on the genomic content of the diversity of complex microbial communities. These are rapid and culture-independent methods.

Genomic sequencing of microbial communities is a high-throughput technology which has seen the emergence of two substantially different sequencing technologies: gene specific sequencing (targeted/amplicon metagenomics) or total microbial genomic sequencing, also called metagenomics sequencing. Both of these techniques have been used as investigative tools for pathogen detection, source tracking, microbial profiling, determining the cause or source of spoilage organisms and pathogens in food processing *(6, 17–19, 21–25)*.

Genomics-based approaches to study the microorganisms associated with food and the food processing environment are listed in *Fig. 1*.

GENOMICS IMPACT ON FOOD SAFETY AND QUALITY

The food industry is concerned about how spoilage and pathogenic microorganisms enter into the final packed product. Current practice includes adopting a variety of testing protocols and programs to monitor and verify ingredients, finished goods, and environmental samples, as well as sanitation program verification testing and various types of ad hoc testing when an issue arises. Generally, best practices include minimizing the testing of finished product and taking a program-centric approach that holistically prevents contamination through optimally managing hurdles that are introduced to mitigate contamination. Verifying the proper functionality of these hurdles is the favored approach so far.

Genomic sequencing can aid in tracing spoilage and pathogenic microorganisms by shining a spotlight on the whole bacterial community associated with the food and the food processing environment. This connection is a powerful tool that has many benefits with current technology, but even greater potential as this rapidly developing set of tools becomes ever more advanced, giving microbiologists highly complex data that can give new

Author for correspondence:
Phone: +1 847.778.0567
E-mail: angela.anandappa@unl.edu

meaning to data-based decision making and the context in which risks are assessed.

The food processing environment as a whole consists of the processing facilities, surrounding environment, and the ingredients and packaging materials that go into making the finished product. The complex matrix of a food has microorganisms that are specific to that ecology, in that environment and in that specific geographic location. The processing facility is a man-made environment and carries microorganism communities reflecting the processing activities in the facility. Each of these ecosystems has intersecting parts where they engage with each other, sharing fragments of DNA, whole living organisms or particles that can influence the function of the adjacent ecosystem.

The collection of microorganisms in a community within an ecosystem is referred to as the microbiota of that specific ecosystem. The genome marker present within each microorganism in the microbiota is used as an identification tool. Collective genomic markers in a community are referred to as their microbiome.

These genomic markers can yield various information about the microbiome. *Figure 2* presents the questions that can be posed to the microbiome and the inferences that can be obtained by use of different genomic markers.

Tools to study microbiota by the use of genomics are powerful. They can identify all microorganisms in a complex matrix without isolation or any unusual growth supplements or temperatures. There are two specific methods of interest: (a) the targeted metagenomics approach and (b) metagenomics employing whole genomic DNA sequencing.

The targeted approach is used for very specific identification needs and can be employed to specifically target bacteria (16S rRNA gene sequencing) or fungi (18S rRNA gene sequencing) or can be used to identify virulent factors or spoilage specific genes. The results are obtained in the form of relative abundance, which represents the abundance of specific organisms in the sample relative to other organisms. This provides a reference point as to how common or how rare the genus is relative to other species *(20)*. The metagenomics approach employing whole genomic DNA sequencing examines all the DNA in the sample, which includes DNA of bacteria, fungi, plants, animals and viruses. This tool provides information on the plant and animal species used as ingredients.

The use of the term rRNA with respect to 16s rRNA gene sequencing is often confused with 16S rDNA and a cause for misperception. The method for 16S rRNA gene sequencing widely uses DNA as a template for amplification and sequencing. The confusion arises because the method

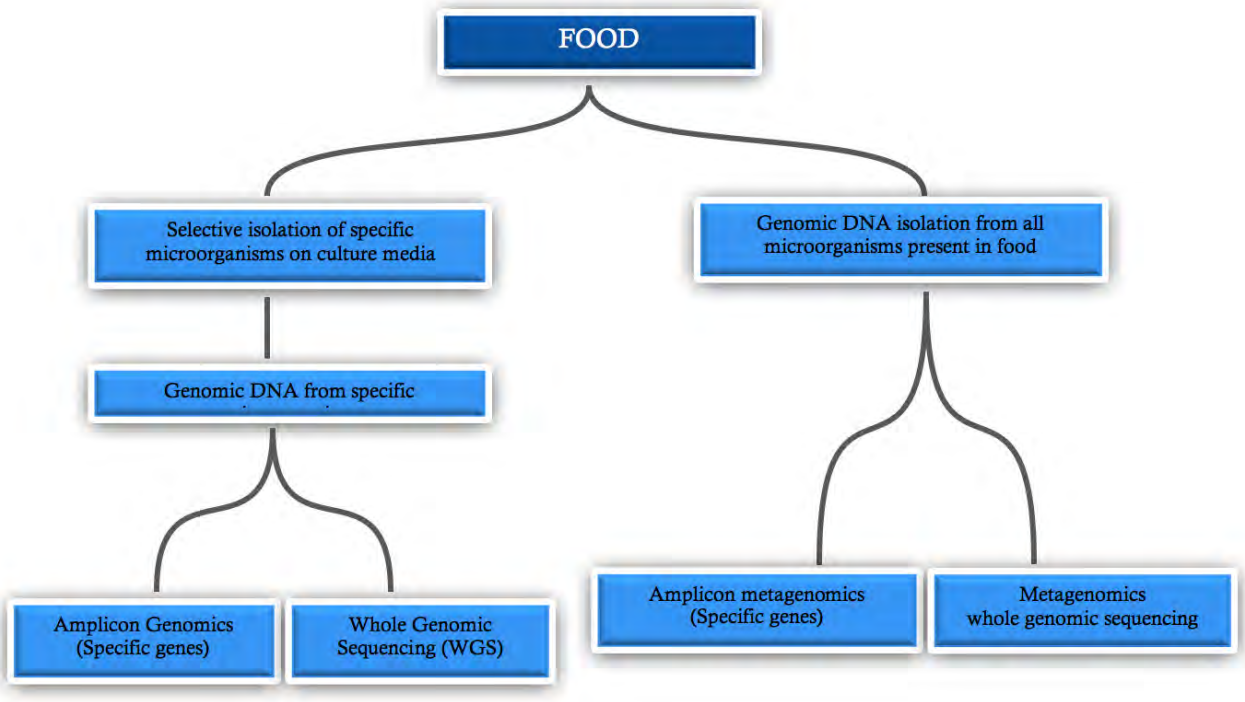

Figure 1. Approach to study microorganisms associated with food or food environment

for sequencing is often referred to as rRNA, on the basis of the fact that the gene portion on the DNA which encodes for 16S rRNA is being sequenced and used as a reference in conducting the analysis. This sequencing method is named after the gene product, which is 16S rRNA, while 16S rDNA is the transcribed DNA, giving a rRNA. Thus, it is not unusual to hear of 16S rRNA and 16S rDNA, and they both refer to analysis of the same gene product being performed on a sample.

Depending on the need, the sequencing method can be chosen to obtain the required level of detail. For example, if a food production facility is interested in understanding the genus of each bacterium and fungus present in the food production facility to validate their sanitation procedures, they can use a targeted/amplicon metagenomics approach to look at genes that are specific to bacteria (16S rRNA) or fungi (18S rRNA). This would yield information on the relative abundance of genera present in the representative samples, for example, *Listeria, Lactobacillus, Eurotium, Aspergillus* or *Salmonella* that are present after sanitation, thus providing vital information on the efficacy of the sanitation procedures. Using a metagenomics approach allows the user to gather the entire ecosystem's signature, and when analyzed over time this information can prove highly valuable in explaining the changing microbiota of the facility. *Figure 3* describes the workflow of the two approaches to study the microbiomes; the targeted metagenomics approach looks at the bacterial population and the metagenomics approach looks all microorganisms present in the flour.

APPLICATION OF GENOMIC APPROACHES IN FOOD SAFETY AND QUALITY

The current use of genomics approaches in food safety have primarily taken the form of quality issue trouble-shooting, food safety monitoring, verification testing in select groups of products and, in a handful of cases, source identification, especially in the context of recall-related troubleshooting. However, most microbiological testing that occurs in the context of food safety and quality testing takes the forms of culture-dependent targeted verification by rapid and plate-based methods that rarely extend beyond the usual list of microorganisms (Aerobic Plate Counts, Yeasts and Molds, *Listeria, E. coli, Salmonella, Enterobacter* and *Lactobacilli*).

Amplicon metagenomics or metagenomics can be employed in food processing facilities as follows:

- In periodic surveillance of the production facility and products
- To check for the efficacy of the processing and sanitation program

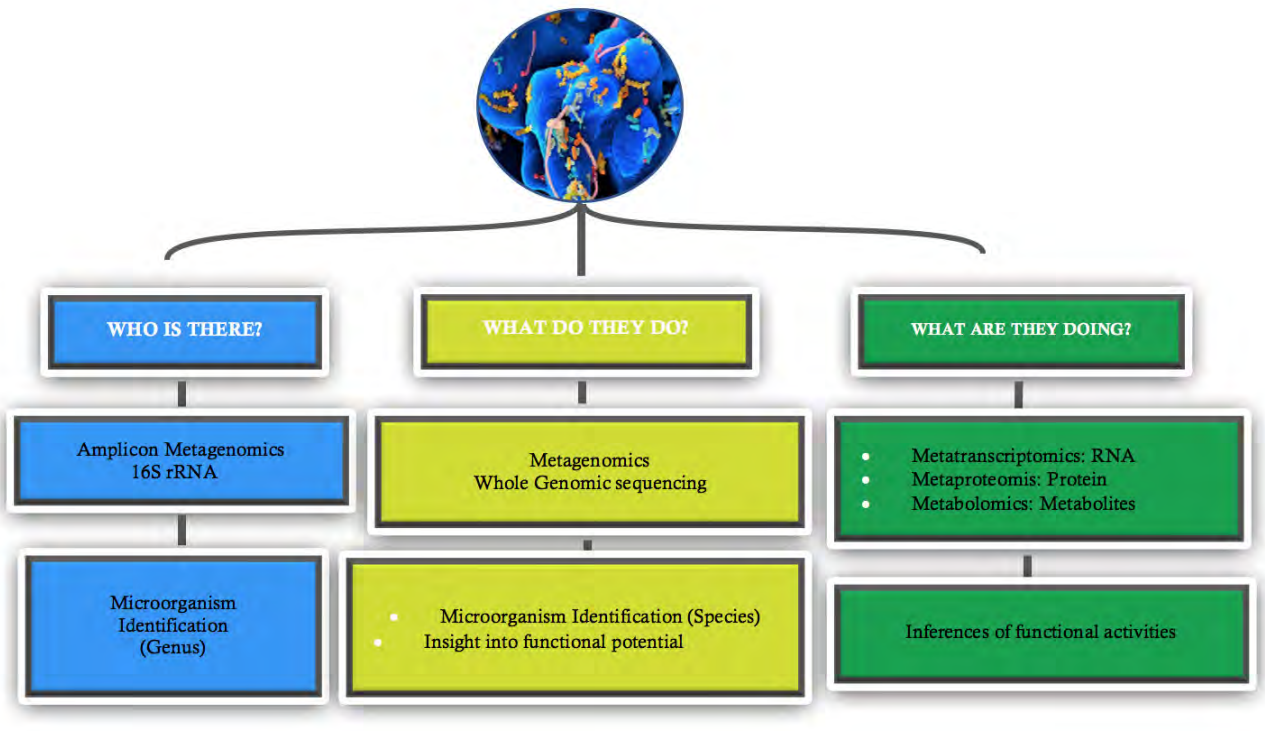

Figure 2. Approaches to study microbiota by use of different types of genomic markers

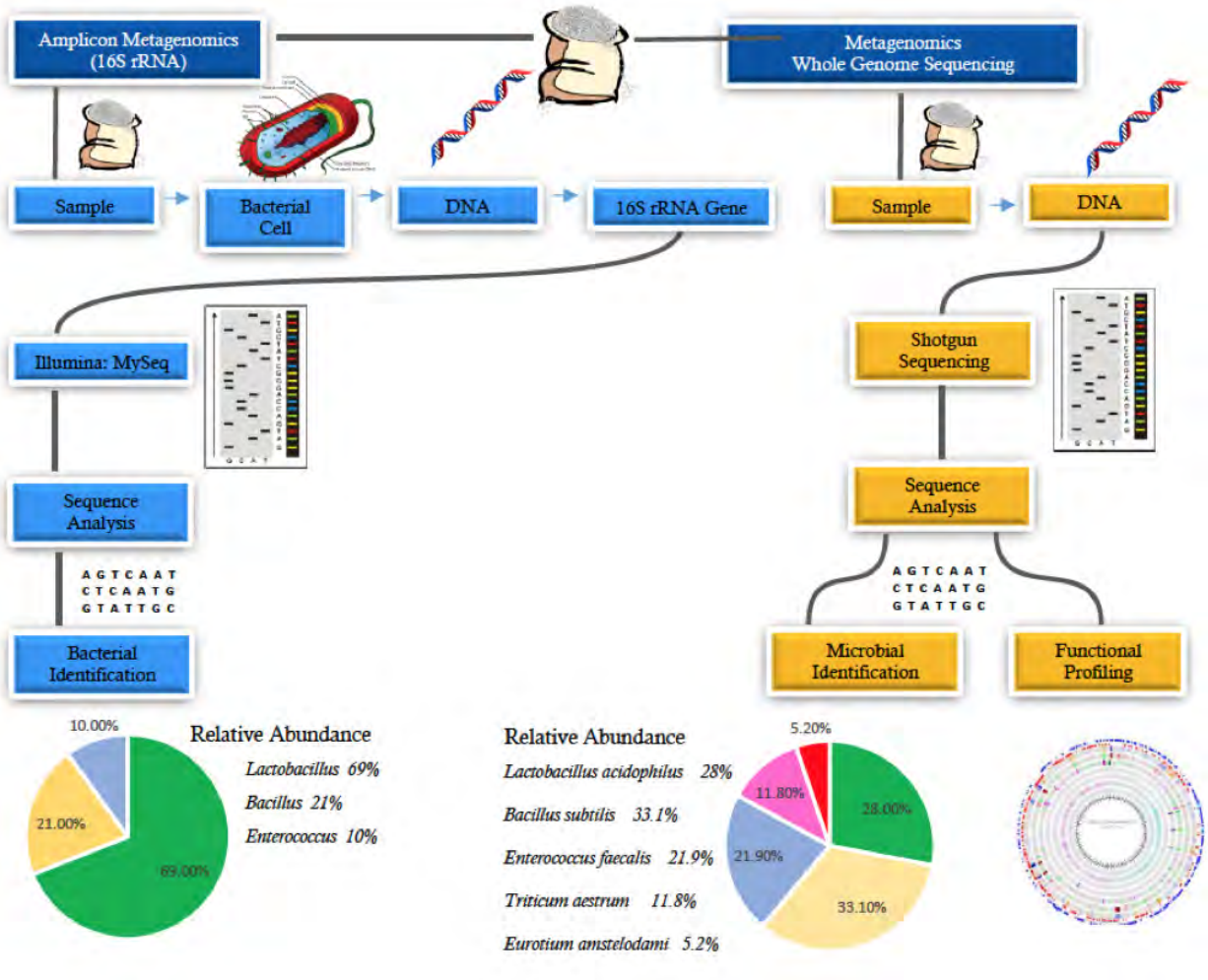

Figure 3. Workflow of amplicon metagenomics and metagenomics approach

- As a verification tool for raw materials obtained from suppliers
- For trouble shooting any persisting spoilage or pathogenic bacteria entering the final packed product
- To understanding changes in microbial populations and recovery of populations of concern based on the rotations in the chemical sanitizer program
- In designing cleaning or sanitizer applications to eradicate and monitor microbial harborage sites
- To differentiate precisely between persistent microorganisms and transients
- In surveillance of new food-production facilities and novel processing operations

Genome tool used by regulatory agencies in foodborne pathogen and outbreak surveillance.

Foodborne disease outbreaks remain a significant global challenge to public health and pose a huge economic burden *(1)*. To tackle these growing challenges, the U.S. Food and Drug Administration (FDA) has created an open-source whole-genomic sequencing network Genome Trakr and an investigative sampling plan, often referred to as a "Swab-a-thon." Whole genome sequencing, (WGS), which is considered a rapid, accurate, cutting-edge technology for investigating food pathogens, is the basis of the FDA's 5-year program of developing a microbiological profile of every processing facility and their products. Food pathogens isolated as part of this sampling will be compared to Genome Trackr run by national laboratories and matched against patient and food sequences from outbreaks. These test results will serve as warning signs for early foodborne outbreak indicators and as well as a tool for the food processing facility to use in taking rigorous preventive actions, which often means stepping up their sanitation program. *Figure 4* describes the workflow for WGS sequence pathways toward incorporating the data into Genome Trackr.

The nature of FDA inspections of food production facilities has changed drastically. In the past, results of visual observations and review of food safety record keeping were the main criteria for FDA inspection activities. If there were any discrepancies, form 483s were issued and the manufacturers would have to employ corrective action and send a written letter to FDA, including descriptions of actions that had been undertaken. In the past couple of years, with the new Food Safety Modernization act (FSMA), along with visual observations and review of food safety records, FDA would look for invisible evidence by employing a Swab-a-thon, or heavy swabbing approach. These Swab-a-thons include obtaining a wide range of samples from all parts of the facility (hygienic zones 1, 2, 3, 4) and can collect up to 200 swabs from various zones, depending on the size of the processing facility. Depending on the number and location of positive samples obtained for pathogenic bacteria of concern, this can result in a recall or even a temporary or long-term shutdown of the facility. The positive swabs are subjected to WGS and compared to Genome Trakr, and if there is a match with the clinical WGS reported, FDA may take legal action leading to criminal prosecution.

CASES WHERE WGS WAS USED AS A TOOL IN IDENTIFICATION OF PATHOGENS

The FDA has used WGS to match pathogens found in food to pathogens collected from people who became ill from consuming the food or pathogens found as a part of FDA's routine surveillance. Following are cases where FDA and CDC employed WGS analysis as part of their outbreak or routine investigation:

1. Fresh fruits and vegetables: There have been a number of outbreaks/ recalls related to fresh produce *(1)*. During their inspection of a fresh produce facility, FDA isolated 19 strains of *Listeria monocytogenes* and performed WGS on the isolates. Seven strains were from food contact surfaces and the rest were either in close proximity to the food contact surface or from a non-food contact surface. The WGS analysis of the 19 strains fell into two distinct strain types, one strain type dominating, with seventeen identical WGS, and the other strain type with two identical WGS. When all the WGS sequences were compared with the WGS sequences on Genome Trakr, the strain type with two identical WGS were found to be associated with eight cases of human illness. Six of the 8 ill individuals were hospitalized for *L. monocytogenes* related illness *(15)*. The company received a form 483 from the FDA requiring a response for corrective action.

   FDA and Center for Disease Control (CDC) performed WGS of isolates obtained from the clinical outbreaks and matched it with a specific brand of salad *(12)*. FDA urged for a recall of the salad, curbing the outbreak.

   Wholesome soy was under scrutiny when FDA and CDC matched WGS of *L. monocytogenes* isolated

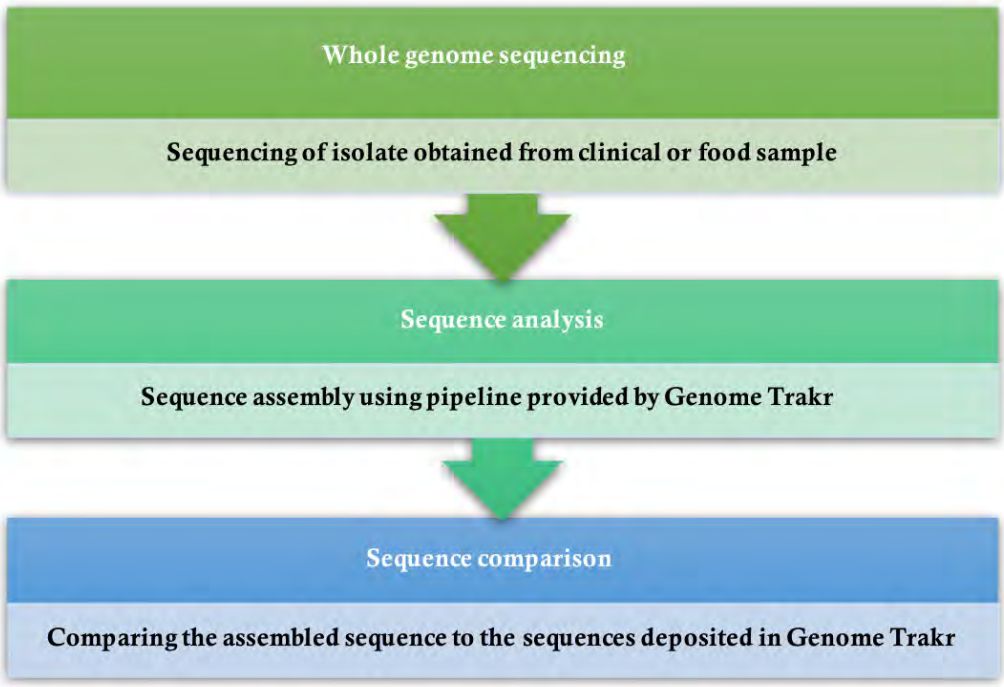

Figure 4. Workflow for WGS sequence comparison to Genome Trakr

from the firm's production facility to WGS of *L. monocytogenes* from five cases who were ill. Overall there were five confirmed cases of illness, and two deaths were reported *(8)*.

2. Flour: FDA/CDC traced back a multistate outbreak of Shiga toxin-producing *E. coli* infections linked to a specific brand of flour. FDA and CDC matched WGS of the pathogen isolated from the patient to the WGS isolated from an open flour pack. Following this revelation, a recall for the product was issued *(11)*.
3. Dairy products: FDA's inspection of an ice cream manufacturing facility isolated 15 isolates of *L. monocytogenes* and found them be comprised of one strain group. On comparison of the strains isolated from the facility to the Genome Trakr database, it was seen that they matched three other isolates: two isolates from finished products tested by a commercial laboratory, and one isolate from an ingredient. FDA urged for a recall and issued the 483 form for corrective actions *(9)*.

   Inspection by FDA in an ice cream plant found *L. monocytogenes* isolated twice over a two-year period. WGS showed that the same strain was isolated both times. FDA suggested that repeated isolation was from an in-house strain that had grown resistant to cleaning and sanitization in place. The company halted its production and incurred a $2.5 million loss from recall, destroying 265 tons of ice cream *(14)*.
4. FDA urged a voluntary recall of milk after finding *Salmonella* from a routine inspection. WGS of strains isolated from this visit was compared to that of strains collected over time from the facility. The strains were identical, pointing to the presence of a persistent strain of *Salmonella* contaminating the facility for nearly 7 years *(10)*.
5. Poultry: CDC reported that a multistate *Salmonella*-related outbreak that sickened 300 people was associated with handling of live poultry *(4)*. WGS, along with other detection methods, was used in the detection of the *Salmonella* outbreak.
6. Restaurants: An outbreak of Shiga-toxin producing *E. coli* linked to a Mexican grill caused illness in 55 people in multiple states, as per CDC reports *(5)*. WGS performed on isolates from ill people showed them to be highly related. Interviews of ill people affected by this multistate event revealed that they had consumed food from this particular Mexican grill chain *(13)*. The chain had to shut its outlets in many states until the root cause of the outbreak could be found.
7. Catering: FDA inspected an airline catering facility and found *L. monocytogenes* isolates from food environmental swabs. WGS analysis confirmed that the isolates collected were identical to the *L. monocytogenes* isolated previously by FDA from the same facility. FDA suggested that the pathogen has become established and has persisted over time and declared that the cleaning procedures in the facility were inadequate to remove the pathogen *(7)*.
8. Seafood: FDA suspended the seafood facility registration because of repeated finding of *L. monocytogenes* in their facility. WGS matched strains isolated from the seafood to the strains isolated from ill people who had consumed the seafood produced by the facility *(16)*.
9. Bridging past and new cases: CDC and FDA matched WGS of *L. monocytogenes* isolates obtained from FDA inspection in 2015 of a soft cheese distribution facility to WGS of a multistate *L. monocytogenes* outbreak caused by contaminated soft cheese. When ill people were interviewed, the only link between all the cases was that they had consumed a specific brand of cheese distributed by the same cheese distributer *(3)*.

WORKING TOWARDS AN FDA VISIT

FDA has stepped up its inspection and foodborne-pathogen surveillance by employing new cutting-edge molecular technologies and the Swab-a-thon sampling program. It is up to the food processing facilities to improve their sanitation and routine surveillance of the processing facilities, incoming ingredients and finished products. Food manufacturers can be prepared for the FDA visit by employing an internal Swab-a-Thon approach to finding potential issues and taking action accordingly. The advantage of conducting periodic swabbing with a view to gaining a thorough understanding of the facility can be best described as surveillance pathways to potential quality concerns. While it is likely that spoilage organisms and pathogens of concern will be found, this data is far more useful in identifying the modes by which these may be transported in the facility, as well as identifying the characteristics of the niches in which problems may persist. It is vital to know where ingredients, packaging and supplies come from and ensure the integrity of the supplier's food safety compliance to protect the business. From the many cases reported by FDA, there is a pattern indicating that many food production facilities could have harborage points and niches that protect pathogens from being removed through standard sanitation protocols. They may persist in the environment and evolve to acquire a unique genetic signature pertinent to that environment. To circumvent this issue of harborage, amplicon sequencing can be used as an investigative tool to study the efficacy of sanitation procedures so as to have a more targeted approach towards the battle against harborage. Employing these genomics tools in a periodic surveillance testing program can avoid any surprises during the FDA visit, and provide the facility with valuable data for internal improvement, serve as a source of historic data and demonstrate one measure of a preventive culture.

CONCLUSIONS

The cost of genome sequencing has been greatly scaled down, making genome-based technologies more affordable for broad adoption for strategic purposes as well as for routine monitoring. Food production facilities can utilize this option for 'smart sampling' as an investigative tool for pathogen detection, source tracking, microbial profiling, determining the fate of spoilage and pathogen profiling during food processing, pre-op, post sanitization, and general facility surveillance activities. Adopting these methods will allow the food processing facility to employ corrective actions to control/eliminate the pathogen or spoilage organisms by giving greater visibility to its pathway into the facility and its prevalence in the sample(s) or in niches. Implementing 'smart sampling' could potentially avoid a product recall or prevent a regulatory finding and allow manufacturers to avoid incurring heavy product losses and their impact on business and brands.

REFERENCES

1. Allard, M. W., E. Strain, D. Melka, K. Bunning, S. M. Musser, E. W. Brown, and R. Timme. 2016. Practical value of food pathogen traceability through building a whole-genome sequencing network and database. *J. Clin. Microbiol.* 54:1975–1983.
2. Amann, R. I., W. Ludwig, and K. H. Schleifer. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Reviews* 59:143–169.
3. CDC. 2015. Multistate outbreak of listeriosis linked to soft cheeses distributed by Karoun Dairies, Inc. https://www.cdc.gov/listeria/outbreaks/soft-cheeses-09-15/index.html:page Accessed 02/23/17.
4. CDC. 2016. Eight multistate outbreaks of human *Salmonella* infections linked to live poultry in backyard flocks. https://www.cdc.gov/salmonella/live-poultry-05-16/index.html. Accessed 02/16/17.
5. CDC. 2016. People infected with the outbreak strain of *E. coli* O26, by state of residence, as of January 27, 2016. https://www.cdc.gov/ecoli/2015/o26-11-15/map.html:Accessed 02/22/17.
6. Dzieciol, M., E. Schornsteiner, M. Muhterem-Uyar, B. Stessl, M. Wagner, and S. Schmitz-Esser. 2016. Bacterial diversity of floor drain biofilms and drain waters in a *Listeria monocytogenes* contaminated food processing environment. *Int. J. Food. Microbiol.* 223:33–40.
7. FDA. 2015. Flying food group, LLC. 12/22/15. https://www.fda.gov/iceci/enforcementactions/ warningletters/2015/ucm483252.htm. Accessed 02/16/17.
8. FDA. 2015. Illinois food company agrees to stop production of contaminated sprouts. https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm444178.htm. Accessed 02/15/17.
9. FDA. 2015. FDA investigates *Listeria monocytogenes* in ice cream products from Blue Bell creameries.https://www.fda.gov/food/recallsoutbreaksemergencies/outbreaks/ucm438104.htm. Accessed 02/15/17.
10. FDA. 2016. Food regulators seize adulterated milk products for food safety violations. https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm531188.htm Accessed 02/16/17.
11. FDA. 2016. FDA investigated multistate outbreak of Shiga toxin-producing *E. coli* infections linked to flour. https://www.fda.gov/food/recallsoutbreaksemergencies/outbreaks/ ucm504192.htm - recalled. Accessed 02/15/17.
12. FDA. 2016. FDA investigated multistate outbreak of the Dole facility in Springfield, Ohio. https://www.fda.gov/food/recallsoutbreaksemergencies/outbreaks/ucm482807.htm. Accessed 02/16/2017.
13. FDA. 2016. FDA investigates multistate outbreak of *E. coli* O26 infections linked to chipotle Mexican grill restaurants. https://www.fda.gov/food/recallsoutbreaksemergencies/outbreaks/ucm470410.htm. Accessed 02/22/17.
14. FDA. 2016. Jeni's splendid ice creams, llc 8/9/16. https://www.fda.gov/iceci/enforcementactions/ warningletters/2016/ucm516395.htm. Accessed 02/15/17.
15. FDA. 2016. Oregon potato company 7/15/16. https://www.fda.gov/iceci/enforcementactions/ warningletters/2016/ucm512776.htm. Accessed 02/16/2017.
16. FDA. 2016. FDA suspends food facility registration of SM Fish Corp. https://www.fda.gov/food/ recallsoutbreaksemergencies/safetyalertsadvisories/ucm520985.html. Accessed 02/15/17.
17. Hultman, J., R. Rahkila, J. Ali, J. Rousu, and K. J. Bjorkroth. 2015. 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.* 81:7088–7097.
18. Kergourlay, G., B. Taminiau, G. Daube, and M. C. Champomier Verges. 2015. Metagenomic insights into the dynamics of microbial communities in food. *Int. J. Food Microbiol.* 213:31–39.
19. Lucking, G., M. Stoeckel, Z. Atamer, J. Hinrichs, and M. Ehling-Schulz. 2013. Characterization of aerobic spore-forming bacteria associated with industrial dairy processing environments and product spoilage. *Int. J. Food Microbiol.* 166:270–279.
20. Mayo, B., C. T. C. C. Rachid, Á. Alegría, A. M. O. Leite, R. S. Peixoto, and S. Delgado. 2014. Impact of next generation sequencing techniques in food microbiology. *Current Genomics.* 15:293–309.
21. Pothakos, V., G. Stellato, D. Ercolini, and F. Devlieghere. 2015. Processing environment and ingredients are both sources of *Leuconostoc gelidum*, which emerges as a major spoiler in ready-to-eat meals. *Appl. Environ. Microbiol*. 81:3529–354.
22. Santo Domingo, J. W., D. G. Bambic, T. A. Edge, and S. Wuertz. 2007. Quo vadis source tracking? Towards a strategic framework for environmental monitoring of fecal pollution. *Water Res.* 41:3539–3552.
23. Stellato, G., F. De Filippis, A. La Storia, and D. Ercolini. 2015. Coexistence of lactic acid bacteria and potential spoilage microbiota in a dairy processing environment. *Appl. Environ. Microbiol.* 81:7893–7904.
24. Stellato, G., A. La Storia, F. De Filippis, G. Borriello, F. Villani, and D. Ercolini. 2016. Overlap of spoilage-associated microbiota between meat and the meat processing environment in small-scale and large-scale retail distributions. *Appl. Environ Microbiol.* 82:4045–4054.
25. Weber, M., J. Geissert, M. Kruse, and A. Lipski. 2014. Comparative analysis of bacterial community composition in bulk tank raw milk by culture-dependent and culture-independent methods using the viability dye propidium monoazide. *J. Dairy Sci.* 97:6761–6776.