Food SafetyTech

Food Genomics
November 2016

Introducing a New Column: Food Genomics

By Gregory Siragusa, Douglas Marshall

Genomics has several powerful applications that food safety professionals can leverage.

DNA sequencing can be used to determine the names, types, and proportions of microorganisms, the component species in a food sample, and track foodborne disease agents. Here we introduce a column exploring aspects and applications of these new techniques, known collectively as food genomics. Each month we will provide take-home knowledge in which every food safety scientist should be familiar.

We live in an exciting time of great change in all of biological and food sciences. In fact, it is not an overstatement to claim that a large portion of the fields of food science, biology, agriculture and medicine will be reformed in what has been called the post-genomics era or simply the genomics era. Food science and food microbiology are major players in this pack and moving in the fast track of these changes. This game-changing technology is fueled by the convergence of two rapidly evolving fields: DNA sequencing and the analysis of that sequencing data (i.e., bioinformatics).

The common jargon uses the acronym NGS for Next Generation Sequencing. NGS refers to the most updated automated DNA sequencing technology available. In several ways, sequencing can be considered a commodity service; hence its price has dropped and its availability is now widespread. What does this mean? A useful analogy is the following: Think of trying to publish a book you wrote. Would you go out, buy a printing press, paper, ink, binding machinery, and produce thousands of copies of your book, or, would you go to a professional printer and get them to print and manufacture copies? For most, the simplicity and experience of the professional print master trumps the do-it-yourself route. Once sequence data is obtained, what is next in the process of using that data? Analysis of sequence data is a specialized field called bioinformatics and has its own expert practitioners. It is a field of study that is a hybrid combination of mathematics, statistics, computer science, and biology. Bioinformatics analyzes the very large datasets produced by NGS and will be increasingly dependent on the internet cloud for its utility to be fully realized.

How will food genomics impact food safety and quality? How will it help in identifying the sources of outbreaks in a fraction of the time it once took? What will this mean for zero-tolerance, for pathogen control, and for responsibilities and liabilities of food producers and processors? There is a growing body of examples and literature that begins to apply genomics and microbiomics to the quality of food and sources of its microbial populations.[5-7]

Over the course of this column, we will be exploring several examples to alert the reader to the myriad of uses of genomics for solving food production issues.

Genomics (NGS and Bioinformatics) are the basis of the US-FDA GenomeTrakr program.[1] Genomics offers an alternative means to serotype *Salmonella* isolates using DNA sequencing.[2] There are several examples of using sequencing of solving the epidemiological source of foodborne microbial outbreaks by comparing the entire bacterial genomes of clinical and food isolates.[3,4]

One powerful application of genomics is to conduct the census of microbial communities to identify the microbial members and their relative proportions, an outcome called a microbiome, all from a single tube! The technique itself is termed *microbiomics*. Just think, we can now identify all bacteria in a complex mixture without isolating what will grow, as well as the many microorganisms we have not yet learned to culture or require unusual temperatures, nutrients, and atmospheres! Can you feel the excitement? Hopefully with knowledge of the power of food genomics you will begin to see the true utility of this technology and begin to appreciate its awesome power. Most importantly, you will begin to see how food genomics is a useful tool for the food science professional.

The microbiome field is changing as of this writing and moving toward using a technique known as whole shotgun metagenome (WSM) analysis in which all of the DNA in a sample is sequenced and not just bacterial, fungal, or specific genes; i.e., a metagenome approach vs. a targeted approach to determining the microbiome of a sample.[8,9] The whole genome shotgun approach is also a powerful tool not only for creating food microbiomes, but can help in the identification of the plant and animal species used as ingredients in foods. WSM requires relatively advanced and sophisticated bioinformatics tools and at the same time sequencing chemistry is advancing, so is bioinformatics. For example, there is an online tool suite known as NEPHELE, which offers publically available online programs, software, and data handling capacity for sequence analysis.[10,11]

So here we are with some brand new shiny tools in the kit. Now the question is, how can the food safety professional begin to use these tools? More to the point is to understand when food genomic data is called for. The first step is to grasp some of the terminology and basic processes. Table 1 lists a few starter terms to become familiar with as well as some web resources that might be helpful to you in understanding these immensely powerful tools.[12,13]

Table 1. Starter Terms in Food Genomics

Annotated Whole Bacterial Genome	High-quality, low-error, gap-free DNA sequence of an entire genome of an organism, in this case, an isolated bacterium, indicating genes and their locations. This can be considered a complete road map of an organism's genetic makeup as expressed in the nucleotides Adenine, Thymine, Cytosine, and Guanine (ATCG's). Can be referred to as WGS or Whole Genome Sequencing.
Bioinformatics	The science of managing and analyzing biological data using advanced computing techniques. Especially important in analyzing genomic research data.
Metagenomes or Whole Shotgun Sequencing	Sequences of Genetic material recovered directly from food, animal, plant, or environmental samples with no foreknowledge of the source of living materials therein. For instance, the metagenome of a yogurt sample will harbor DNA sequences characteristic of starter culture bacteria and bovine DNA (assuming it is bovine milk yogurt). This is another approach to obtaining a microbiome.[6]
Microbiome	A community of microorganisms that inhabit a particular environment or sample. For example a plant microbiome includes all the microorganisms that colonize a plant's surfaces and int passages. This can be a Targeted (Amplicon Sequencing Based) or a Metagenome (Whole Shotgun Metagenome based) microbiome.[6]
Microbiomics	The process of determining a microbiome.
Microbiota	The ecological community of commensal, symbiotic, and pathogenic microorganisms that literally share a space or are within a sample. Formerly the term 'microflora' was used, term is waning in usage.[14]
NGS (Next Generation Sequencing)	High throughput automated sequencing of nucleic acids DNA or RNA.

Finally, in the reference section we have tried to provide you with some useful online reference sources. The U.S. Department of Energy has perhaps the most intuitive, user-friendly and informative sites we have encountered as of late ("Genome Glossary," 2016). The same source also published a talking glossary ("Talking Glossary of Genetic Terms," 2016). The reader should be advised that genomic terminology and nomenclature is still not fully mature. In fact, the number of vague meanings, cross references, and acronyms can sometimes be frustrating; but fear not, as one reads and discusses the terms, they will become clearer. As a start we recommend downloading a helpful reference that follows.[15] There are many other sites you will locate by performing a single web-search. If you would like to share your favorite genomics sites, please drop a line to either author and we will try to compile them into a single electronic document.

We hope this first column will find you coming back for more as we explore this burgeoning field and learn how it is being linked to food safety. Look for future articles on specific food applications, methods, and hot topics in food genomics. Goodbye for now.

References

1. Allard, M. W., Strain, E., Melka, D., Bunning, K., Musser, S. M., Brown, E. W., & Timme, R. (2016). Practical Value of Food Pathogen Traceability through Building a Whole-Genome Sequencing Network and Database. *Journal of Clinical Microbiology*, 54(8), 1975–1983. https://doi.org/10.1128/JCM.00081-16
2. Zhang, S., Yin, Y., Jones, M. B., Zhang, Z., Deatherage Kaiser, B. L., Dinsmore, B. A., … Deng, X. (2015). Salmonella serotype determination utilizing high-throughput genome sequencing data. *Journal of Clinical Microbiology*, 53(5), 1685–1692. https://doi.org/10.1128/JCM.00323-15
3. Burall, L. S., Grim, C. J., Mammel, M. K., & Datta, A. R. (2016). Whole Genome Sequence Analysis Using JSpecies Tool Establishes Clonal Relationships between Listeria monocytogenes Strains from Epidemiologically Unrelated Listeriosis Outbreaks. *PloS One*, 11(3), e0150797. https://doi.org/10.1371/journal.pone.0150797
4. Chen, Y., Burall, L. S., Luo, Y., Timme, R., Melka, D., Muruvanda, T., Brown, E. W. (2016). Isolation, enumeration and whole genome sequencing of Listeria monocytogenes in stone fruits linked to a multistate outbreak. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.01486-16
5. Bokulich, N. A., Lewis, Z. T., Boundy-Mills, K., & Mills, D. A. (2016). A new perspective on microbial landscapes within food production. *Current Opinion in Biotechnology*, 37, 182–189. https://doi.org/10.1016/j.copbio.2015.12.008
6. Bokulich, N. A., & Mills, D. A. (2012). Next-generation approaches to the microbial ecology of food fermentations. BMB Reports, 45(7), 377–389.
7. Zarraonaindia, I., Owens, S. M., Weisenhorn, P., West, K., Hampton-Marcell, J., Lax, S., … Gilbert, J. A. (2015). The soil microbiome influences grapevine-associated microbiota. *mBio*, 6(2). https://doi.org/10.1128/mBio.02527-14
8. Microbial Foods – The Science Of Fermented Foods. (n.d.). Retrieved November 21, 2016, from http://microbialfoods.org/
9. Ranjan, R., Rani, A., Metwally, A., McGee, H. S., & Perkins, D. L. (2016). Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochemical and Biophysical Research Communications*, 469(4), 967–977. https://doi.org/10.1016/j.bbrc.2015.12.083
10. Colosimo, M. E., Peterson, M. W., Mardis, S., & Hirschman, L. (2011). Nephele: genotyping via complete composition vectors and MapReduce. Source Code for Biology and Medicine, 6, 13. https://doi.org/10.1186/1751-0473-6-13
11. Weber, N. (n.d.). Cloud Computing for Scientific Research The NIH Nephele Project for Microbiome Analysis. Accessed November 21, 2016. Retrieved from https://www.google.com/url?q=http://casc.org/meetings/14sep/CASC-NIH-Microbiome-Cloud-Project-20140917.pdf&sa=U&ved=0ahUKEwj2mvTt1LrQAhXMC8AKHUiCBLsQFggHMAE&client=internal-uds-cse&usg=AFQjCNGt_lx2zw4qLNcHuYwZvSg10ivp5Aabout:blank

12. Genome Glossary. (n.d.). Accessed November 21, 2016. Retrieved from http://doegenomestolife.org/glossary/index.shtml
13. Talking Glossary of Genetic Terms. (n.d.). Accessed November 21, 2016. Retrieved from https://www.genome.gov/glossary/
14. Microbiota. (2016, November 14). In Wikipedia. Retrieved from https://en.wikipedia.org/w/index.php?title=Microbiota&oldid=749454552
15. Marchesi, J. R., & Ravel, J. (2015). The vocabulary of microbiome research: a proposal. *Microbiome*, 3, 31. https://doi.org/10.1186/s40168-015-0094-5

Resource

Nutrition, C. for F. S. and A. (n.d.). Whole Genome Sequencing (WGS) Program – GenomeTrakr Network [WebContent]. Accessed November 21, 2016. Retrieved from http://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/ucm363134.htm

Food SafetyTech

Food Genomics
January 3, 2017

Microbiomes Move Standard Plate Count One Step Forward

Reprinted from: *Food Safety Tech*

By Gregory Siragusa, Douglas Marshall*

Last month we introduced several **food genomics terms** including the microbiome. Recall that a microbiome is the community or population of microorganisms that inhabit a particular environment or sample. Recall that there are two broad types of microbiomes, a targeted (e.g., bacteria or fungi) or a metagenome (in which all DNA in a sample is sequenced, not just specific targets like bacteria or fungi). This month we would like to introduce the reader to uses of microbiomes and how they augment standard plate counts and move us into a new era in food microbiology. Before providing examples, it might be useful to review a diagram explaining the general flow of the process of determining a microbiome (See Figure 1).

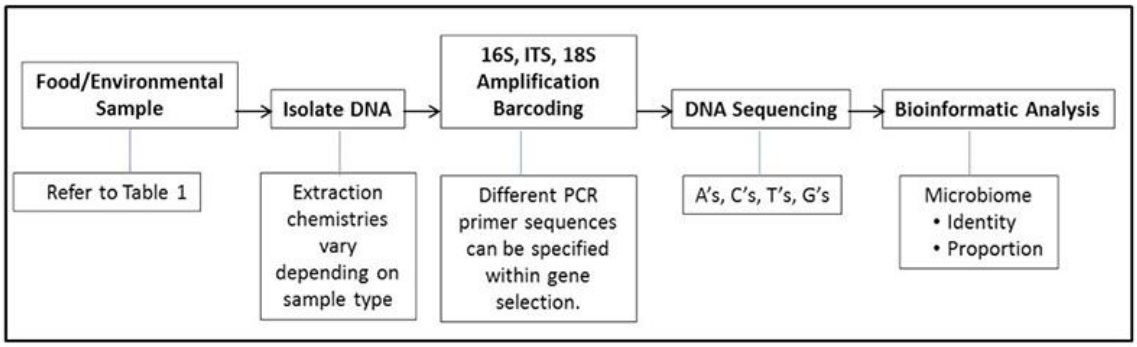


Figure 1. General process for performing a targeted microbiome (bacterial or fungal)

By analogy, if one thinks of cultural microbiology and plate counts as a process of counting colonies of microbes that come from a food or environmental sample, microbiome analysis can be thought of as identifying and counting signature genes, such as the bacterial specific 16S gene, from the microbes in a food or environmental sample. Plate counts have been and remain a food microbiologist most powerful indicator tool in the tool kit; however, we know there are some limitations in their use. One limitation is that not all bacterial or fungal cells are capable of outgrowth and colony formation on specific media under a set of incubation conditions (temperature, time, media pH, storage atmosphere, etc.). Individual plate count methods cannot cover the nearly infinite number of variations of growth atmospheres and nutrients. Because of these limitations microbiologists understand that we have not cultured but many different types of bacteria on the planet (this led to the term "The Great Plate Count Anomaly" (Staley & Konopka, 1985). Think of a holiday party where guests were handed nametags on which was printed: "Hello, I grow on Standard Methods Agar" or "Hello, I grow at 15°C", etc. We can group the partygoers by ability to grow on certain media; we can also count partygoers, but they still do not have names. As effective as our selective and differential media have become, bacterial colonies still do not come with their own "Hello, My Name Is XYZ" name tags. Therefore, in the lab, once a plate is counted it is generally tossed into the autoclave bag, along with unnamed colonies and all they represent. Microbiomes can provide a nametag of sorts as well as what proportion of people at that party share a certain name. For instance: "Hello, My Name is Pseudomonas" or "Hello, My Name is Lactobacillus", etc. The host can then say "Now we are going to count you; would all Pseudomonas pleased gather in this corner?" or "All Lactobacillus please meet at the punch bowl".

A somewhat overly simplified analogy, but it makes the point that microbiome technology gives names and proportions. Microbiomes too have limitations. First, with current technologies microbiomes need a relatively large threshold of organisms of a specific group to appear in the microbiome pie chart— approximately 103. In theory, a colony on a plate of agar medium can be derived from a single cell or colony-forming unit (CFU). Not all amplified genes in a microbiome are necessarily from viable cells (A topic that will be covered later in this series of articles). Forming a colony on an agar surface on the other hand requires cell viability. Finally, the specificity of a microorganism name assigned to a group in a microbiome depends on the size of the sequenced amplicon (an amplicon is a segment of DNA, in this case the 16S gene DNA, resulting from amplification by PCR before sequencing) and how well our microbial databases cover different subtypes in a species. Targeted microbiomes can reliably name the genus of an organism, however resolution to the species and subspecies is not guaranteed. (Later in this series we will discuss metagenomes and how they have the potential to identify to a species or even subspecies level). Readers can find very informative reviews on microbiome specificity in the following cited references: Bokulich, Lewis, Boundy-Mills, & Mills, 2016; de Boer et al., 2015; Ercolini, 2013; Kergourlay, Taminiau, Daube, & Champomier Vergès, 2015.

When we consider the power of using cultural microbiology for quantitative functional indicators of microbial quality together with microbiomic analysis, with limitations and all for both, microbiomes have opened a door to the vast and varied biosphere of our food's microbiology to a depth never before observed. This all sounds great, but how will we benefit and use this information? We have constructed Table 1 with examples and links of microbiome applications to problems that would have required years to study by cultural microbiology techniques alone.

Please note this is by no means an exhaustive list, but it serves to illustrate the very broad and deep potential of microbiomics to food microbiology. We encourage the reader to email the editors or authors with questions regarding any reference. Using PubMed and the search terms "Food AND microbiome" will provide abstracts and a large variety of applications of this technology. These microorganisms open the door to the vast and varied biosphere of the microbiology of food.

Foodstuff	Reference
Ale	(Bokulich, Bamforth, & Mills, 2012)
Beef Burgers	(Ferrocino et al., 2015)
Beefsteak	(De Filippis, La Storia, Villani, & Ercolini, 2013)
Brewhouse and Ingredients	(Bokulich et al., 2012)
Cheese	(Wolfe, Button, Santarelli, & Dutton, 2014)
Cheese and *Listeria* growth	(Callon, Retureau, Didienne, & Montel, 2014)
Cherries, Hydrostatic Pressure	(del Árbol et al., n.d.)
Cocoa	(Illeghems, De Vuyst, Papalexandratou, & Weckx, 2012)
Dairy Starters and Spoilage Bacteria	(Stellato, De Filippis, La Storia, & Ercolini, 2015)
Drinking Water Biofilms	(Chao, Mao, Wang, & Zhang, 2015)
Fermented Foods	(Tamang, Watanabe, & Holzapfel, 2016)
Foodservice Surfaces	(Stellato, La Storia, Cirillo, & Ercolini, 2015)
Fruit and Vegetables	(Leff & Fierer, 2013)
Insect Protein	(Garofalo et al., 2017)
Kitchen surfaces	(Flores et al., 2013)
Lamb	(Wang et al., 2016)
Lobster	(Tirloni, Stella, Gennari, Colombo, & Bernardi, 2016)
Meat and storage atmosphere	(Säde, Penttinen, Björkroth, & Hultman, 2017)
Meat spoilage and processing plant	(Pothakos, Stellato, Ercolini, & Devlieghere, 2015)
Meat Spoilage Volatiles	(Casaburi, Piombino, Nychas, Villani, & Ercolini, 2015)
Meat Stored in Different Atmospheres	(Ercolini et al., 2011)
Milk	(Quigley et al., 2011)
Milk and Cow Diet	(Giello et al., n.d.)
Milk and Mastitis	(Bhatt et al., 2012)
Milk and Teat Preparation	(Doyle, Gleeson, O'Toole, & Cotter, 2016)
Natural starter cultures	(Parente et al., 2016)
Olives	(Abriouel, Benomar, Lucas, & Gálvez, 2011)
Pork Sausage	(Benson et al., 2014)
Spores in complex foods	(de Boer et al., 2015)
Tomato Plants	(Ottesen et al., 2013)

Winemaking (Marzano et al., 2016)

Table 1. Examples of microbiome analysis of different foods and surfaces.

References

- Abriouel, H., Benomar, N., Lucas, R., & Gálvez, A. (2011). Culture-independent study of the diversity of microbial populations in brines during fermentation of naturally-fermented Aloreña green table olives. International Journal of Food Microbiology, 144(3), 487–496. https://doi.org/10.1016/j.ijfoodmicro.2010.11.006
- Benson, A. K., David, J. R. D., Gilbreth, S. E., Smith, G., Nietfeldt, J., Legge, R., … Singh, I. (2014). Microbial successions are associated with changes in chemical profiles of a model refrigerated fresh pork sausage during an 80-day shelf life study. Applied and Environmental Microbiology, 80(17), 5178–5194. https://doi.org/10.1128/AEM.00774-14
- Bhatt, V. D., Ahir, V. B., Koringa, P. G., Jakhesara, S. J., Rank, D. N., Nauriyal, D. S., … Joshi, C. G. (2012). Milk microbiome signatures of subclinical mastitis-affected cattle analyzed by shotgun sequencing. Journal of Applied Microbiology, 112(4), 639–650. https://doi.org/10.1111/j.1365-2672.2012.05244.x
- Bokulich, N. A., Bamforth, C. W., & Mills, D. A. (2012). Brewhouse-resident microbiota are responsible for multi-stage fermentation of American coolship ale. PloS One, 7(4), e35507. https://doi.org/10.1371/journal.pone.0035507
- Bokulich, N. A., Lewis, Z. T., Boundy-Mills, K., & Mills, D. A. (2016). A new perspective on microbial landscapes within food production. Current Opinion in Biotechnology, 37, 182–189. https://doi.org/10.1016/j.copbio.2015.12.008
- Callon, C., Retureau, E., Didienne, R., & Montel, M.-C. (2014). Microbial biodiversity in cheese consortia and comparative Listeria growth on surfaces of uncooked pressed cheeses. International Journal of Food Microbiology, 174, 98–109. https://doi.org/10.1016/j.ijfoodmicro.2014.01.003
- Casaburi, A., Piombino, P., Nychas, G.-J., Villani, F., & Ercolini, D. (2015). Bacterial populations and the volatilome associated to meat spoilage. Food Microbiology, 45(Pt A), 83–102. https://doi.org/10.1016/j.fm.2014.02.002
- Chao, Y., Mao, Y., Wang, Z., & Zhang, T. (2015). Diversity and functions of bacterial community in drinking water biofilms revealed by high-throughput sequencing. Scientific Reports, 5. https://doi.org/10.1038/srep10044
- de Boer, P., Caspers, M., Sanders, J.-W., Kemperman, R., Wijman, J., Lommerse, G., … Kort, R. (2015). Amplicon sequencing for the quantification of spoilage microbiota in complex foods including bacterial spores. Microbiome, 3(1). https://doi.org/10.1186/s40168-015-0096-3

- De Filippis, F., La Storia, A., Villani, F., & Ercolini, D. (2013). Exploring the sources of bacterial spoilers in beefsteaks by culture-independent high-throughput sequencing. PloS One, 8(7), e70222. https://doi.org/10.1371/journal.pone.0070222
- del Árbol, J. T., Pulido, R. P., La Storia, A., Burgos, M. J. G., Lucas, R., Ercolini, D., & Gálvez, A. (n.d.). Microbial diversity in pitted sweet cherries (Prunus avium L.) as affected by high-hydrostatic pressure treatment. Food Research International. https://doi.org/10.1016/j.foodres.2016.10.014
- Doyle, C. J., Gleeson, D., O'Toole, P. W., & Cotter, P. D. (2016). High-throughput sequencing highlights the significant influence of seasonal housing and teat preparation on the raw milk microbiota. Applied and Environmental Microbiology, AEM.02694-16. https://doi.org/10.1128/AEM.02694-16
- Ercolini, D. (2013). High-throughput sequencing and metagenomics: moving forward in the culture-independent analysis of food microbial ecology. Applied and Environmental Microbiology, 79(10), 3148–3155. https://doi.org/10.1128/AEM.00256-13
- Ercolini, D., Ferrocino, I., Nasi, A., Ndagijimana, M., Vernocchi, P., La Storia, A., … Villani, F. (2011). Monitoring of microbial metabolites and bacterial diversity in beef stored under different packaging conditions. Applied and Environmental Microbiology, 77(20), 7372–7381. https://doi.org/10.1128/AEM.05521-11
- Ferrocino, I., Greppi, A., La Storia, A., Rantsiou, K., Ercolini, D., & Cocolin, L. (2015). Impact of Nisin-Activated Packaging on Microbiota of Beef Burgers during Storage. Applied and Environmental Microbiology, 82(2), 549–559. https://doi.org/10.1128/AEM.03093-15
- Flores, G. E., Bates, S. T., Caporaso, J. G., Lauber, C. L., Leff, J. W., Knight, R., & Fierer, N. (2013). Diversity, distribution and sources of bacteria in residential kitchens. Environmental Microbiology, 15(2), 588–596. https://doi.org/10.1111/1462-2920.12036
- Garofalo, C., Osimani, A., Milanović, V., Taccari, M., Cardinali, F., Aquilanti, L., … Clementi, F. (2017). The microbiota of marketed processed edible insects as revealed by high-throughput sequencing. Food Microbiology, 62, 15–22. https://doi.org/10.1016/j.fm.2016.09.012
- Giello, M., La Storia, A., Masucci, F., Di Francia, A., Ercolini, D., & Villani, F. (n.d.). Dynamics of bacterial communities during manufacture and ripening of traditional Caciocavallo of Castelfranco cheese in relation to cows' feeding. Food Microbiology. https://doi.org/10.1016/j.fm.2016.11.016
- Illeghems, K., De Vuyst, L., Papalexandratou, Z., & Weckx, S. (2012). Phylogenetic analysis of a spontaneous cocoa bean fermentation metagenome reveals new insights into its bacterial and fungal community diversity. PloS One, 7(5), e38040. https://doi.org/10.1371/journal.pone.0038040
- Kergourlay, G., Taminiau, B., Daube, G., & Champomier Vergès, M.-C. (2015). Metagenomic insights into the dynamics of microbial communities in food. International Journal of Food Microbiology, 213, 31–39. https://doi.org/10.1016/j.ijfoodmicro.2015.09.010
- Leff, J. W., & Fierer, N. (2013). Bacterial communities associated with the surfaces of fresh fruits and vegetables. PloS One, 8(3), e59310. https://doi.org/10.1371/journal.pone.0059310
- Marzano, M., Fosso, B., Manzari, C., Grieco, F., Intranuovo, M., Cozzi, G., … Santamaria, M. (2016). Complexity and Dynamics of the Winemaking Bacterial

Communities in Berries, Musts, and Wines from Apulian Grape Cultivars through Time and Space. PloS One, 11(6), e0157383. https://doi.org/10.1371/journal.pone.0157383

- Ottesen, A. R., González Peña, A., White, J. R., Pettengill, J. B., Li, C., Allard, S., … Brown, E. (2013). Baseline survey of the anatomical microbial ecology of an important food plant: Solanum lycopersicum (tomato). BMC Microbiology, 13, 114. https://doi.org/10.1186/1471-2180-13-114
- Parente, E., Guidone, A., Matera, A., De Filippis, F., Mauriello, G., & Ricciardi, A. (2016). Microbial community dynamics in thermophilic undefined milk starter cultures. International Journal of Food Microbiology, 217, 59–67. https://doi.org/10.1016/j.ijfoodmicro.2015.10.014
- Pothakos, V., Stellato, G., Ercolini, D., & Devlieghere, F. (2015). Processing Environment and Ingredients Are Both Sources of Leuconostoc gelidum, Which Emerges as a Major Spoiler in Ready-To-Eat Meals. Applied and Environmental Microbiology, 81(10), 3529–3541. https://doi.org/10.1128/AEM.03941-14
- Quigley, L., O'Sullivan, O., Beresford, T. P., Ross, R. P., Fitzgerald, G. F., & Cotter, P. D. (2011). Molecular approaches to analysing the microbial composition of raw milk and raw milk cheese. International Journal of Food Microbiology, 150(2–3), 81–94. https://doi.org/10.1016/j.ijfoodmicro.2011.08.001
- Säde, E., Penttinen, K., Björkroth, J., & Hultman, J. (2017). Exploring lot-to-lot variation in spoilage bacterial communities on commercial modified atmosphere packaged beef. Food Microbiology, 62, 147–152. https://doi.org/10.1016/j.fm.2016.10.004
- Staley, J. T., & Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annual Review of Microbiology, 39, 321–346. https://doi.org/10.1146/annurev.mi.39.100185.001541
- Stellato, G., De Filippis, F., La Storia, A., & Ercolini, D. (2015). Coexistence of Lactic Acid Bacteria and Potential Spoilage Microbiota in a Dairy Processing Environment. Applied and Environmental Microbiology, 81(22), 7893–7904. https://doi.org/10.1128/AEM.02294-15
- Stellato, G., La Storia, A., Cirillo, T., & Ercolini, D. (2015). Bacterial biogeographical patterns in a cooking center for hospital foodservice. International Journal of Food Microbiology, 193, 99–108. https://doi.org/10.1016/j.ijfoodmicro.2014.10.018
- Tamang, J. P., Watanabe, K., & Holzapfel, W. H. (2016). Review: Diversity of Microorganisms in Global Fermented Foods and Beverages. Frontiers in Microbiology, 7. https://doi.org/10.3389/fmicb.2016.00377
- Tirloni, E., Stella, S., Gennari, M., Colombo, F., & Bernardi, C. (2016). American Lobsters (Homarus Americanus) not Surviving During Air Transport: Evaluation of Microbial Spoilage. Italian Journal of Food Safety, 5(2), 5620. https://doi.org/10.4081/ijfs.2016.5620
- Wang, T., Zhao, L., Sun, Y., Ren, F., Chen, S., Zhang, H., & Guo, H. (2016). Changes in the microbiota of lamb packaged in a vacuum and in modified atmospheres during chilled storage analysed by high-throughput sequencing. Meat Science, 121, 253–260. https://doi.org/10.1016/j.meatsci.2016.06.021
- Wolfe, B. E., Button, J. E., Santarelli, M., & Dutton, R. J. (2014). Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. Cell, 158(2), 422–433. https://doi.org/10.1016/j.cell.2014.05.041

Food SafetyTech

Food Genomics
February 17, 2017

Microbiomes a Versatile Tool for FSMA Validation and Verification

By Douglas Marshall, Ph.D., Gregory Siragusa

The use of genomics tools are valuable additions to companies seeking to meet and exceed validation and verification requirements for FSMA compliance (21 CFR 117.3). In this installment of Food Genomics, we present reasons why microbiome analyses are powerful tools for FSMA requirements currently and certainly in the future.

Recall in the first installment of Food Genomics we defined a microbiome as the community of microorganisms that inhabit a particular environment or sample. For example, a food plant's microbiome includes all the microorganisms that colonize a plant's surfaces and internal passages. This can be a targeted (amplicon sequencing-based) or a metagenome (whole shotgun metagenome-based) microbiome. Microbiome analysis can be carried out on processing plant environmental samples, raw ingredients, during shelf life or challenge studies, and in cases of overt spoilage.

As a refresher of FSMA requirements, here is a brief overview. Validation activities include obtaining and evaluating scientific and technical evidence that a control measure, combination of control measures, or the food safety plan as a whole, when properly implemented, is capable of effectively controlling the identified microbial hazards. In other words, can the food safety plan, when implemented, actually control the identified hazards? Verification activities include the application of methods, procedures, tests and other evaluations, in addition to monitoring, to determine whether a control measure or combination of control measures is or has been operating as intended, and to establish the validity of the food safety plan. Verification ensures that the controls in the food safety plan are actually being properly implemented in a way to control the hazards.

Validation establishes the scientific basis for food safety plan process preventive controls. Some examples include using scientific principles and data such as routine indicator microbiology, using expert opinions, conducting in-plant observations or tests, and challenging the process at the limits of its operating controls by conducting challenge studies. FSMA-required validation frequency first includes before the food safety plan is implemented (ideally), within the first 90 calendar days of production, or within a reasonable timeframe with written justification by the preventive controls qualified individual. Additional validation efforts must occur when a change in control measure(s) could impact efficacy or when reanalysis indicates the need.

FSMA requirements stipulate that validation is not required for food allergen preventive controls, sanitation preventive controls, supply-chain program, or recall plan effectiveness. Other preventive controls also may not require validation with written justification. Despite the lack of regulatory expectation, prudent processors may wish to validate these controls in the course of developing their food safety plan. For example, validating sanitation-related controls for pathogen and allergen controls of complex equipment and for how long a processing line can run between cleaning are obvious needs.

There are many routine verification activities expected of FSMA-compliant companies. For process verification, validation of effectiveness, checking equipment calibration, records review, and targeted sampling and testing are examples. Food allergen control verification includes label review and visual inspection of equipment; however, prudent manufacturers using equipment for both allergen-containing and allergen-free foods should consider targeted sampling and testing for allergens. Sanitation verification includes visual inspection of equipment, with environmental monitoring as needed for RTE foods exposed to the environment after processing and before packaging. Supply-chain verification should include second- and third-party audits and targeted sampling and testing. Additional verification activities include system verification, food safety plan reanalysis, third-party audits and internal audits.

Verification procedures should be designed to demonstrate that the food safety plan is consistently being implemented as written. Such procedures are required as appropriate to the food, facility and nature of the preventive control, and can include calibration of process monitoring and verification instruments, and targeted product and environmental monitoring testing.

Food SafetyTech

Food Genomics
March 14, 2017

How is DNA Sequenced?

By Sanjay K. Singh, Douglas Marshall, Gregory Siragusa

Here is a prediction. Within the next year or years, at some time in your daily work life as a food safety professional you will be called upon to either use genomic tools or to understand and relay information based on genomic tools for making important decisions about food safety and quality. Molecular biologists love to use what often seems like a foreign or secret language. Rest assured dear reader, these are mostly just vernacular and are easily understood once you get comfortable with a bit of the vocabulary. In this the fourth installment of our **column** we progress to give you another tool for your food genomics tool kit. We have called upon a colleague and sequencing expert, Dr. Sanjay Singh, to be a guest co-author for this topic on sequencing and guide us through the genomics language barrier.

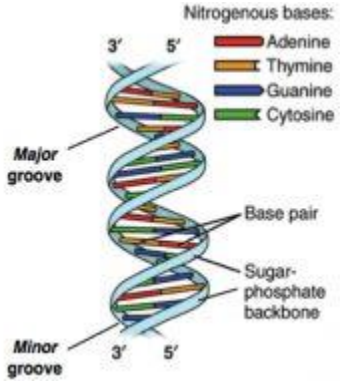
The first report of the annotated (labeled) sequence of the human genome occurred in 2003, 50 years after the discovery of the structure of DNA. In this genome document all the genetic information required to create and sustain a human being was provided. The discovery of the structure of DNA has provided a foundation for a deeper understanding of all life forms, with DNA as a core molecule of genetic information. Of course that includes our food and our tiny friends of the microbial world. Further molecular technological advances in the fields of agriculture, food science, forensics, epidemiology, comparative genomics, medicine, diagnostics and therapeutics are providing stunning examples of the power of genomics in our daily lives. We are only now beginning to harvest the fruits of sequencing and using that knowledge routinely in our respective professions.

In our first column we wrote, "**DNA sequencing** can be used to determine the names, types, and proportions of microorganisms, the component species in a food sample, and track foodborne diseases agents." In this month's column, we present a basic guide to how DNA sequencing chemistry works.

Image courtesy of US Human Genome Project Knowledge base

DNA sequencing is the process of determining the precise order of four nucleotide bases, adenine or A, cytosine or C, guanine or G, and thymine or T in a DNA molecule. By knowing the linear sequence of A, C, G, and T in a DNA molecule, the genetic information carried in that particular DNA molecule can be determined.

DNA sequencing happened from the intersections of different fields including biology, chemistry, mathematics, and physics.[1,2] The critical breakthrough was provided in 1953 by James Watson, Francis Crick, Maurice Wlkins and Rosalind Franklin when they resolved the now familiar double helix structure of DNA.[3] Each helical strand was a polynucleotide, which consists of repeating monomeric units called nucleotides. A nucleotide consists of a sugar (deoxyribose), a phosphate moiety, and one of the four nitrogenous bases—the aforementioned A, C, G, and T. In the double helix, the strands run opposite to each other, commonly referred as anti-parallel. Repeating units of base-pairs (bp), where A always pairs with T and C always pairs with G, are arranged within the double helix so that they are slightly offset from each other like steps in a winding staircase. On a piece of paper, the double helix is often represented by scientists as a flat ladder-like structure, where the base pairs (bp) form the rungs of the ladder while the sugar-phosphate backbone form the antiparallel rails (see Figure 1).



Artistic representation of DNA Double Helix. Source: Eurofins

The two ends of each polynucleotide strand are called 5′ or 3′-end, a nomenclature that represents the chemical structure of the deoxyribose sugar at that terminus. The lengths of a single- or double-stranded DNA are often measured in bases (b) or bases pairs (bp), respectively. The two polynucleotide strands can be readily unzipped by heating, and on cooling, the initial double-helix structure is re-formed or re-annealed. The ability to rezip the initial ladder-like structure can be attributed to the phenomenon of base pairing, which merits repetition—the base A always pairs with T and the base G always with C. This rather innocuous phenomenon of base pairing is the basis for the mechanism by which DNA is copied when cells divide and is also the

theoretical basis on which most traditional and modern DNA sequencing methodologies have been developed.

Other biological advancements also paved the way towards the development of sequencing technologies. Prominent amongst these were the discovery of enzymes that allowed a scientist to manipulate the DNA. For example, restriction enzymes that recognize and cleave DNA at specific short nucleotide sequences can be used to fragment a long duplex strand of DNA.[4] The DNA polymerase enzyme, in the presence of the deoxyribose nucleotide triphosphates (dNTPs: Chemically reactive forms of the nucleotide monomers), can use a single DNA strand to fill in the complementary bases and extend a shorter rail strand (primer extension) of a partial DNA ladder.[5] A critical part of the primer extension is the 'primer', which are short single-stranded DNA pieces (15 to 30 bases long) that are complementary to a segment of the target DNA. These primers are made using automated high-throughput synthesizer machines. Today, such primers can be rapidly manufactured and delivered on the following day. When the primer and the target DNA are combined through a process called annealing (heat and then cool), they form a structure that shows a ladder-like head and a long single-stranded tail. In 1983, Kary Mullis developed an enzyme-based process called Polymerase Chain Reaction (PCR). Using this protocol, one can pick a single copy of DNA and amplify the same sequence an enormous number of times. One can think of PCR as molecular photocopier in which a single piece of DNA is amplified up to approximately 30 billion copies!

The other critical event that changed the course of DNA sequencing efforts was the publication of the 'dideoxy chain termination' method by Dr. Frederick Sanger in December 1977.[6] This marked the beginning of the first generation of DNA sequencing techniques. Most next-generation sequencing methods are refinements of the chain termination, or "Sanger method" of sequencing.

Frederick Sanger chemically modified each base so that when it was incorporated into a growing DNA chain, the chain was forcibly terminated. By setting up a primer extension reaction where in one of the chemically modified 'inactive' base in smaller quantity is mixed with four active bases, Sanger obtained a series of DNA strands, which when separated based on their size indicated the positions of that particular base in the DNA sequence. By analyzing the results from four such reactions run in parallel, each containing a different 'inactive' base, Sanger could piece together the complete sequence of the DNA. Subsequent modifications to the method allowed for the determination of the sequence using dye-labeled termination bases in a single reaction. Since, a sequence of less than <1000 bases can be determined from a single such reaction, the sequence of longer DNA molecules have to be pieced together from many such reads.

Using technologies available in the mid-1990's, as many as 1 million bases of sequence could be determined per day. However, at this rate, determining the sequence of the 3 billion bp human genome required years of sequencing work. By analogy, this is equivalent to reading the Sunday issue of *The New York Times*, about 300,000 words, at a pace of 100 words per day. The cost of sequencing the human genome was a whopping $70 million. The human genome project clearly brought forth a need for technologies that could deliver fast, inexpensive and accurate genome sequences. In response, the field initially exploded with modifications to the Sanger method.

The impetus for these modifications was provided by advances in enzymology, fluorescent detection dyes and capillary-array electrophoresis. Using the Sanger method of sequencing, one can read up to ~1,000 bp in a single reaction, and either 96 or 384 such reactions (in a 96 or 384 well plate) can be performed in parallel using DNA sequencers. More recently a new wave of technological sequencing advances, termed NGS or next-generation sequencing, have been commercialized. NGS is fast, automated, massively parallel and highly reproducible. NGS platforms can read more than 4 billion DNA strands and generate about a terabyte of sequence data in about six days! The whole 3 billion base pairs of the human genome can be sequenced and annotated in a mere month or less.

Our objective here is to provide a brief introduction to aspects of the technologies that are used for NGS. Execution of a sequencing project using any of the NGS technologies involves three steps:

1. **Library preparation:** Generating small pieces of DNA so that they can be read in parallel
2. **Sequencing and imaging:** Determining the sequence of the bases in immobilized DNA molecules in a massively parallel manner
3. **Data analysis a.k.a. bioinformatics:** Piecing together the bits and pieces of the sequence collected in the second step into one logical, massive and contiguous sequence.

Before going much further, we have constructed a table of some important terms for your reference (see Table 1).

Term	Brief Definition/Translation
Read Depth (or Sequencing Depth)	Number of times a sequence is determined for a single sample. A single read can have errors so multiple reads are desired for data quality.
Read Length	Length (bp) of an individual read
Coverage	A measure to determine the fraction of the total genome represented in the sequence data with a particular level of accuracy.
Library Preparation	The first step in the NGS workflow, which involves fragmenting the target DNA to a size compatible with the NGS platform and prepping the same for sequencing, i.e., by attaching adaptors.
Bp, Kb, Mb	A measure of read size or genome size: Base Pair, Kilobases (1,000 bp), Megabases (1 million bp).
Read Quality	Number of bp read errors in a sequence
FASTA and FASTQ files	Computer files containing the sequence
DNA Extraction	Wet chemistry protocol to remove high-quality DNA from a specimen
"Quality of DNA"	Indicators of quantity (ng/ml or ng's) , purity and molecular weight of DNA extracted from a sample
"Just send me your DNA's"	Refers to mailing or bringing DNA extracted from a sample to the sequencing lab

Table I.

During library preparation, genomic DNA is randomly broken into pieces typically <1,000 bp long, followed by ligation of adaptors (synthetic double stranded (ds) DNA fragments of known sequence) to the ends of the sheared DNA. A common theme across the NGS technologies is that millions of these adaptor-flanked DNA templates are attached to solid supports using different methods. This spatial distribution of immobilized templates allows for millions to billions of sequencing reactions to be run simultaneously. For example, in the first next-gen sequencers launched by the company 454 Life Sciences, tiny beads are used that contain several DNA strands complementary to a segment of the added-on adaptor, where the attachment of one template (piece of DNA to be sequenced) to one bead is achieved. Using PCR, multiple copies (millions) of each fragment of DNA tied to a bead are then generated on the surface of each bead.

While different NGS technologies use different sequencing chemistries to determine the sequence, all NGS protocols use smaller quantities of reagent per sequencing reaction than Sanger techniques and allow for multiple orders of increase in the amount of sequence data collected. Each of these advancements helps lower the cost of sequencing. Since sequencing reactions are performed using immobilized DNA fragments, the features of the recorded signal (typically fluorescence or light emitted during the extension of the primer) are on the scale of microns (i.e., smaller than the thickness of a human hair). Therefore, an image of reasonable surface area can provide information on millions of sequencing reactions being run in parallel. Picture a screen with many different colored dots appearing/disappearing in all parts of the screen, each representing a nucleotide base being detected and recorded into a sequence.

In case of the 454 Life Science sequencers, sequencing is conducted by a process called pyrosequencing, where a clever use of the luciferase enzyme makes every base incorporated give off a burst of light. In a single run, the 454 instrument can obtain around 400,000 reads at lengths of 200 to 400 bp. Several NGS platforms have emerged and have further reduced the cost of sequencing a genome (see Table 2).

Platform	Instruments	Read Lengths (bp)
Illumina	MiniSeq, MiSeq, NextSeq, HiSeq, HiSeqX	125–600
Ion Torrent	Proton, PGM	200–400
Pacific Biosciences	PacBio RS, PacBio RS II	4,600–14,000
Roche 454	GS FLX, GS FLX+	400–700
SOLiD	5500, 5500xI, 5500 W	100

Table 2. NGS Sequencing Platforms

In the end, all of these instruments spit out a result that is generally in the form of a file type known as a FASTA or FASTQ (refer to Table 1). These files contain the sequence of ATCG's in

a sample and are the start of the bioinformatics process to be covered in a forthcoming addition to this column.

For the food safety professional, genomics investigations require accurate sequence information for reliable interpretation. Professionals are urged to consider certified sequencing providers that offer strong customer orientation, impeccable quality, fast service and high reliability. Poor quality sequence information can lead to poor quality species assignments in public databases. Faulty assignments lead to wrong bioinformatics interpretations. Recent highly sensationalized food genomics press releases showing the presence of difficult-to-believe contaminants, such as human or rat DNA in highly processed foods, may be due to analysis of poor quality sequence information. It is also recommended that professionals consult with organizations that know something about food science and technology to make sure sequence-based conclusions are based on a foundation of real and sound data.

References

1. Hutchinson, C. A. III. (2007) DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res*. 35, 6227–6237.
2. Lee, T. F. (1991). The Human genome project; Cracking the genetic code of life.
3. Watson, J. D. and Crick, F.H. (1953). A structure for deoxyribose nucleic acid. *Nature*. 171 (4356): 737–738.
4. Smith, H. O. and Wilcox, K. W.(1970) A restriction enzyme from Hemophilus influenza I. Purification and general properties. *J. Mol. Biol*. 51, 379–391.
5. Kaiser A D, Wu R (1968) Structure and function of DNA cohesive ends. Cold Spring Harb. Symp. Quant. Biol 1968;33:729-734.
6. Sanger, F. et al. (1977) Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 265, 687–696.

Additional Resources

1. Smith LM, Sanders JZ, Kaiser RJ, et al. (1986). "Fluorescence detection in automated DNA sequence analysis". Nature. 321 (6071): 674–9.
2. Ewing, B.; Hillier, L.; Wendl, M. C.; and Green, P. (1998). Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res. 8, 175–185.

Food SafetyTech

Food Genomics
April 24, 2017

Metagenomes and Their Utility

By Gregory Siragusa and Douglas Marshall

Guest Co-Author Nur Hasan, Ph.D., VP CosmosID, Inc.

Editor Maria Fontanazza

Recall that in article two of this series we wrote: "…. there are two broad techniques to obtaina microbiome, a targeted (e.g., bacteria or fungi) or a metagenome (in which all DNA in a sample is sequenced, not just specific targets like bacteria or fungi)". In this column we will now explore metagenomes and some applications to food safety and quality. We have invited Dr. Nur Hasan of CosmosID, Inc., an expert in the field of microbial metagenomics, to share his deep knowledge of metagenomics. Our format will be an interview style.

Safe food production and preservation is a balancing act between food enzymes and microbes. We will start with some general questions about the microbial world, and then proceed deeper into why and how tools such as metagenomics are advancing our ability to explore this universe. Finally, we will ask Dr. Hasan how he sees all of this applying to food microbiology and safe food production.

GS/DM: Thank you for joining us. Dr. Hasan, please give us a brief statement of your background and current position.

NH: Thanks for having me. I am a molecular biologist by training. I did my Bachelor and Masters in Microbiology, M.B.A in Marketing, and Ph.D. in Molecular Biology. My current position is Vice President and Head of Research and Development at CosmosID, Inc., where I am leading the effort on developing the world's largest curated genome databases and ultra rapid bioinformatics tools to build the most comprehensive, actionable and user-friendly metagenomic analysis platform for both pathogen detection and microbiome characterization.

GS/DM: The slogan for CosmosID is *"Exploring the Universe of Microbes"*. What is your estimate of the numbers of bacterial genera and species that have not yet been cultured in the lab?

NH: Estimating the number of uncultured bacteria on earth is an ongoing challenge in biology. The widely accepted notion is more than 99% of bacteria from environmental samples remain 'unculturable' in the laboratory; however, with improvements in media design, adjustment of nutrient compositions and optimization of growth conditions based on the ecosystem these bacteria are naturally inhabiting, scientists are now able to grow more bacteria in the lab than we anticipated. Yet, our understanding is very scant on culturable species diversity across diverse ecosystems on earth. With more investigators using metagenomics tools, many ecosystems are being repeatedly sampled, with ever more microbial diversity revealed. Other ecosystems remain ignored, so we only have a skewed understanding of species diversity and what portion of such diversity is actually culturable. A report from Schloss & Handelsman, 2004 highlighted the limitations of sampling, and the fact that it is not possible to estimate the total number of bacterial species on Earth. Despite the limitation, they took a stab at the question and predicted minimum bacterial species richness to be 35,498. A more recent report by Hugenholtz et al., 2009, estimated that there are currently 61 distinct bacterial phyla, of which 31 have no cultivable representatives. Currently NCBI has about 16,757 bacterial species listed, which represent less than 50% of minimum species richness as predicted by Schloss & Handelsman, 2004, and only a fraction of all global species richness of about 107 to 109 estimated by Curtis et al., 2002 & Dykhuizen, 1998.

GS/DM: In generic terms will you tell us what exactly a metagenome is? Also, please explain the meaning of the terms "shotgun sequencing", "shotgun metagenomes", and "metagenomes". How are they equivalent, similar or different?
NH: *Metagenome* is actually an umbrella term. It refers to the collection of genetic content of all organisms present in a given sample. It is studied by a method called metagenomics that involves direct sequencing of a heterogeneous population of DNA molecules from a biological sample all at once. Although in most applications, metagenome is often used to refer to microbial metagenome (the genes and genomes of microbial communities of given sample), in a broader sense, it actually represents total genetic makeup of a sample including genomes and gene sequences of other materials in the sample, such as nucleic acids contributed by other food ingredients of plant and animal origin. The *metagenome* provides an in-depth understanding of the composition, structure, functional and metabolic activities of food, agricultural, and human communities.

Shotgun sequencing is a method where long strands of DNA (such as an entire genome of a bacterium) are randomly shredded ("shotgunning") into smaller DNA fragments, so that they can be sequenced individually. Once sequenced, these small fragments are then assembled together into contigs by computer programs that find overlaps in the genetic code, and the complete sequence of the bacterial genome is generated. Now, instead of one genome, if you directly sequence entire assemblage of genomes from a metagenome using such shotgun approach, it's called *shotgun metagenomics* and resulting output is termed a *shotgun metagenome.* By this method, you are literally sequencing thousands of genomes simultaneously from a given metagenome in one assay and get the opportunity to reconstruct individual genomes or genome fragments for investigation and comparison of the genetic consortia and taxonomic composition of complete communities and their predicted functions.

Whereas *targeted 16S rRNA* or *targeted 16S amplicon sequencing* relies on amplification and sequencing of one target region, the 16S gene region, shotgun metagenomics is actually target free, it is aimed at sequencing entire genomes of every organism present in a sample and gives a more accurate, and unbiased biological representation of a sample. As an analogy of *shotgun metagenomics,* lets think about your library where you may have multiple books (like as different organisms present in a metagenome). You can imagine shotgun metagenomics as process, where all books from your library are shredded, mixed up, and then you will assemble the small shredded pieces to find text overlap and piecing the cover of all books together to reassemble each of your favorite books. *Shotgun metagenomics* approximates this analogy.

Metagenome and *metagenomics* are often used interchangeably. Where *metagenome* is the total collection of all genetic material from a given samples. *Metagenomics* is the method to obtain a *metagenome* that utilizes a shotgun sequencing approach to sequence all these genetic material at once.

Shotgun sequencing and *shotgun metagenomics* are also used interchangeably. Shotgun sequencing is a technique where you fragment large DNA strands into small pieces and sequence all small fragments. Now, if you apply such techniques to sequence a *metagenome,* than we call it *shotgun metagenomics*.

GS/DM: What can metagenomics data tell us that a Targeted 16S (Bacterial) or ITS (Fungal) Microbiome does not?
NH: Actually, the differences are quite profound! With 16S or ITS, you typically only get taxonomic composition of bacterial and fungal community at a genus level. In contrast, metagenomics gives you more precise taxonomic composition of both bacteria and fungi at species, sub-species or strain-level accuracy. Shotgun metagenomics also captures viruses and eukaryotes, and provides information on antimicrobial resistance and virulence genes in parallel, which is not possible with targeted 16S, ITS, etc. Because shotgun metagenomics sequences the entire genetic content of a sample, it offers the opportunity to characterize functional and metabolic potential of a given community. For example, new species, enzymes, or pathways can be discovered. Furthermore, genomic linkages between function and phylogeny for uncultured organisms can be elucidated. Additionally, shotgun metagenomics has high specificity of detection and unbiased measurement of organism abundance. It is more "representative" of the natural community than targeted 16S or ITS sequencing, which can disrupt the natural composition due to PCR amplification bias.

GS/DM: Your Company's innovative science and technology uses metagenomics to determine microbiomes of a host of samples and niches. How has the technology of CosmosID changed the whole field of metagenomics?
NH: In short; we offer a solution to the data analysis bottleneck created by shotgun metagenomics. Recent advances in next-generation sequencing (NGS) technologies (i.e., significantly increased throughput and rapid turnaround time), have made it possible to sequence complex biological samples deep enough so that metagenomics can be used to identify and characterize all pathogens and commensals in any given sample, thereby establishing community profiling for decision-making. While metagenomics offers many

the time and expense of bioinformatics to piece together complex unassembled reads and lack of comprehensive databases containing curated and validated genomes of all possible organisms of interest was a significant bottleneck for its applications in many critical areas including food quality and safety. Realizing such critical need, CosmosID has developed an integrated platform with novel bioinformatics tools and expertly curated genome databases to facilitate ultra-fast metagenomic analysis with high specificity required for applications like clinical diagnostics and food safety. We converted complex bioinformatics data analysis processes into a simple interface that is user-friendly so that microbiologists, even without any bioinformatics skill, can do their own analysis and obtain rapid, reliable, high-resolution identification and accurate quantitation of microbial populations. Furthermore, characterization of the microbial population attributes, such as antibiotic resistance and virulence is possible with this tool. To our knowledge, the CosmosID platform is the only metagenomics solution that can profile pathogens at sub-species and strain level with highly accuracy. The CosmosID platform even provides comparative statistical analysis across many different datasets based on distinct cohorts and/or other associated metadata. With this technology scientists can now better explore biodiversity, understand microbial interactions and community interplay in various biological processes, and monitor pathogen and antibiotic resistance transmission. When this pool of data is combined, source trace back and outbreak control activities become easier and quicker to manage.

GS/DM: Of course the microbial world is not only composed of bacteria. Metagenomics captures sequences of all DNA-based life forms in a sample. Could this tool also be used for detecting foodborne viruses and protozoal pathogens?
NH: Oh Yes! The beauty of metagenomics that it's a one time universal assay; therefore, your detection capability is not restricted to bacteria and fungi only, metagenomics can also detect both viruses and other non-microbial eukaryotes simultaneously from the same assay.

GS/DM: What if a virus in a sample is an RNA virus?
NH: That's a very good question. To capture the RNA viruses, you need to take a slightly different approach, called microbial transcriptomics, where you isolate RNA from a given sample and do RNA sequencing. Interestingly, this approach not only ensures detection of RNA viruses, but also detects all DNA-based life forms. However, this method is relatively more expensive compared to DNA sequencing-based methods.

GS/DM: What if a fermented food producer is interested in bacteriophage that might impact their process; would this technique help them?
NH: Yes, using metagenomics, we can readily identify, detect, and track bacteriophages of interest from fermented food products and manufacturing environments. In fact, our curated virus database already includes large number of bacteriophages, in addition to DNA and RNA viruses. Furthermore, if a fermented food producer encounters bacteriophages that are not listed in our database; we can easily update our database to include those bacteriophage genomes to enable tracking of these specific bacteriophages in the fermented food manufacturing environment. Our ability to confidently detect phages is another key differentiator of the CosmosID solution, which offers invaluable information particularly when a phage-specific organism appears in a food or environmental sample at low abundance.

GS/DM: You indicated that the length of a genome obtained in shotgun sequencing is significantly longer than that obtained by a targeted amplicon sequencing method such as obtained in a Targeted 16S microbiome. Why is that important?
NH: Yes, that's true. Typically bacterial genomes comprise about 4,000 genes but when you are using targeted sequencing like 16S, you are essentially sequencing only a single gene, such as the 16S gene, or even part of a gene (i.e., variable regions). You are then using the differences you observed in those sequences, which represent only about 0.025% of a genome, to infer identity of a bacterium. Extrapolation of a single gene polymorphism may not be accurate and often looses resolution in detecting closely related organisms. In shotgun sequencing, instead of a single gene, you are sequencing the entire genome and leveraging the sequence information of all or many of these 4,000 genes to investigate the complete genetic make-up of an organism. Shotgun sequencing offers high precision and accuracy in detection provides information related to other genes including antibiotic resistance and virulence, and can even reconstruct an organism's full length genome to understand evolution, pathogenesis, and clonal transmission.

GS/DM: Do you envision a time when metagenomics sequencing will become a solution or a "one-stop shop" for microbial diagnostics in food?
NH: Yes! When you use metagenomics, you are using an unbiased, culture-free, universal method that gives you the opportunity to investigate complete genetic make-up of a food sample, offering insights into the associated microbial community - their diversity, composition, functional, metabolic and virulence potential. You can also focus on your pathogen of interest to understand its source in a food production environment. Because the food genome is also sequenced, insights on authenticity and economic fraud can also be made. Thus, metagenomics offer a tantalizing new tool for supplier verification activities.

GS/DM: Would metagenomics analysis be useful in the monitoring of antimicrobial resistance genes in food production environments or even from bacteria on food itself?
NH: Yes, using metagenomics you can readily detect and track antibiotic resistance genes in food production and monitor their transmission. CosmosID has developed a comprehensive antibiotic resistance database that contains resistance gene sequences from all major classes of antibiotics; therefore, you can probe metagenomic or individual bacterial sequence data against that database to profile the antibiotic resistance genes carried by the community as a whole (community resistome) or by the individual bacteria.

GS/DM: CosmosID's Technology is intimately tied to curated databases. What is a curated database and what if a specific life-form is not included in that database? For example, you mentioned this earlier in a statement about bacteriophage.
NH: Yes, CosmosID solution is tied to curated databases. By curated database we mean a database whose content is expertly checked and screened for a variety of common sequence errors, contaminations, miss-assemblies, and taxa misclassifications that may otherwise affect accuracy of metagenomic detection. Examples of errors include nematode sequences being submitted at NCBI as a bacterial sequence and human sequence reads assembled into microbial genomes including foodborne pathogens. We do extensive cleaning of genomes before they are incorporated into our databases; however, to get high resolution and accurate results, it takes more than just "clean genomic sequences". For example, we have built whole genome

relatedness trees and due to such phylogenetic organization of our database, we are able to identify an organism down to the sub-species or strain level. Additionally, if a specific life-form is not represented in the database, we can identify the organism to it's nearest phylogenetic neighbor, and statistics of such identification will be indicative of it being a novel organism. Furthermore our database is incredibly flexible. We can readily incorporate new genomes and gene sequences into the existing databases and utilize the modified databases to probe metagenome datasets. This allows us to develop custom databases for specific clients by incorporating any novel and/or proprietary genomes they may have.

GS/DM: Food fraud is rapidly becoming a topic of keen interest in the food production world. You mentioned earlier that CosmosID technology can address this problem. How would it do that?
NH: Yes! As I mentioned earlier that metagenomics is a universal assay, therefore, as long as you have a qualified database to represent the plant and animal species of your interest and sequencing depth are tailored appropriately, you can also detect plant and animal species using metagenomics. In fact, one of our current databases includes about 1500 plant and animal reference genomes and using this database we can detect those plant and animal species readily. What we may not be able to address is if identified species are intentionally added or unintentionally present due to incidental agricultural or transport carryover.

GS/DM: Some practical questions: First, how many target organisms must be in a sample to be detected in a *metagenome?*
NH: Very interesting question. Basically, it is possible to detect as low as 1 to 10 cells of an organism in a specimen (per gram or mL) by metagenomics as long as you sequence your sample deep enough. In general, when using targeted 16S sequencing at a sequencing depth of ~20,000, for a taxonomic unit to show up as a discrete grouping (e.g. genus or species) and appear as a piece of a pie chart diagram requires approximately 103-104 gene copies or roughly 103-104 cells. For metagenomic shotgun sequencing Now, how deep* a sample needs to be sequenced to attain such lower level of detection depends on the sample type, sample size, diversity, and richness of underlying microbes associated with it. For example, soil is a very complex sample type with enormous diversity and richness of both microbes and non-microbial eukaryotes, whereas a fermented food product may be considered as relatively simple sample type as its expected microbial diversity and richness would be much lower. Therefore, to detect a particular organism, say *E. coli,* at 1 to 10 cells level one will need to sequence a high diversity sample like soil is much deeper than what you need for a low diversity sample such as a fermented food product. Additionally, use of pathogen-specific enrichment (Ottesen et al., 2016) prior DNA isolation and/or use of larger sample size should be considered to increase detection sensitivity.

(Authors note: we will explore sampling and sample sizes further in a future installment of Food Genomics, - GS/DM)

GS/DM: If DNA from dead or lysed bacteria or other organisms is detected, what does that mean in terms of food safety or quality?

NH: This is a critical point in food safety or quality point of view. The method is so sensitive that it is likely to detect DNA from dead or lysed organisms, and if detected, we expect their abundance and coverage to be very low as well. However, we are mindful that even low-level presence of some food pathogens can have important food safety or quality implications, therefore, our recommendations is to use microbiological culture to confirm viability when food pathogens are detected, especially at low level abundance or coverage. Use of RNA isolation and sequencing instead of DNA, pretreatment of samples with a membrane-impermeable reagent like propidium monoazide (PMA) which enables selective amplification of DNA from viable cells, or carrying out multiple metagenomic analyses to measure quantitative changes over time could be used to circumvent this problem.

GS/DM: Will a shotgun approach go beyond the level of the genus or species taxonomic classification?
NH: Oh, yes! Shotgun metagenomics approach not only detects organisms beyond genus or species, it readily detects organisms at sub-species and strain level. In many cases it also can detect multiple strains of a species if present in a given sample. We frequently observe detection of multiple strains of a species in fermented food or probiotic samples. In our recent studies on *Listeria monocytogenes* in the Blue Bell ice-cream outbreak (Ottensen et al., 2016) and *Aeromonas hydrophila* in necrotizing fasciitis (Ponnusamy et el., 2016), we have demonstrated such capability.

GS/DM: For someone to use this technology or your technology specifically, what type of sample would they have to obtain? How much sample is needed? Could it be preserved in ethanol or should it be frozen? If so at what temperature do you recommend? What is the average TAT (turn around time) for metagenomics analysis?
NH: We always prefer to use fresh food samples for nucleic acid extraction. Typically 100 to 250 mg of solid or 1 to 2 mL of liquid food samples yields adequate nucleic acid for next generation sequencing. However, when it is not possible to use fresh samples, freezing and storage at −80 °C should be considered. Ethanol preservation can also be used, and in that case we suggest using > 95% ethanol to allow for more rapid penetration of cellular membranes and deactivation of DNases. However, it is important to remember that ethanol preserved samples often yields lower DNA yield. While we have expedited the turnaround time for sequence data analysis to be done in minutes, the current turnaround times for the entire workflow (sample to report) are 1 to 2 weeks. However, depending on the batch size of the samples and use of integrated workflow turnaround time is rapidly approaching 48 to 96 hours.

GS/DM: Analytical costs are an obvious consideration for the food industry. Could you share some general ranges of costs for this technology? Do you see prices for metagenomics analysis dropping as sequencing and processor costs fall?
NH: Pricing is dependent on a few factors, such as type of sample, batch size, sample prep difficulty, genome complexity and data analysis package desired. Generally, it ranges from $300 to $1000/sample. We expect the cost of sequencing to drop significantly in the next few years. Coupled with workflow automation and automated user friendly bioinformatics that we have already developed, I expect this method to be more cost competitive in the coming years.

GS/DM: Finally, Dr. Hasan, would you give us your view on how metagenomics will impact food production as a routine tool in the food microbiologist's toolkit?
NH: Long term, as costs come down and the sequencing and analytics capabilities become more portable and cost effective, I think you'll see metagenomics become the main diagnostic tool for both routine monitoring through the entire food supply chain, food fraud, outbreak trace back and spoilage investigations. In terms of timing, I think we are a few years out for mass adoption, but it will still be less than a decade, and the technology is really coming up quickly! Even today, as you can see from the *Listeria* study (Ottesen et al., 2016); it is possible to fast track Listeria outbreak investigations from weeks to few days using metagenomics. As you can imagine, this sort of technology, can and will revolutionize the microbiology of food and food fermentation with lot of new insights enabling critical improvements in manufacturing efficiency, reducing food recalls, wastage, improving shelf life and ultimately providing safer foods, while reducing total costs across the entire food supply chain.

Thank you very much Dr. Hasan for sharing your expert knowledge and insights through this interview and article.

We encourage our readers to please comment or ask questions in the section below. Also, if you have specific ideas for Food Genomic topics, please communicate those to Doug Marshall, Greg Siragusa or the editors of Food Safety Tech eMagazine.
Stay tuned for another Food Genomics installment in Food Safety Tech.

** Read Depth (or Sequencing Depth) is the number of times a sequence is determined or read for a single sample or the average number of times that a particular nucleotide is represented in a collection of random raw sequences. A single read can have errors so multiple reads are desired for data quality. Sequencing depth can range from thousands to millions.*

References

Schloss, Patrick D., and Jo Handelsman. "Status of the microbial census." *Microbiology and Molecular Biology Reviews* 68.4 (2004): 686-691.

Hugenholtz, Philip, Sean D. Hooper, and Nikos C. Kyrpides. "Focus: Synergistetes." *Environmental microbiology* 11.6 (2009): 1327-1329.

Curtis, T. P., W. T. Sloan, and J. W. Scannell. 2002. Estimating prokaryotic diversity and its limits. *Proc. Natl. Acad. Sci.* USA 99:10494-10499.

Dykhuizen, D. E. 1998. Santa Rosalia revisited: Why are there so many species of bacteria? *Antonie Leeuwenhaek* 73:25-33.

Ottesen, Andrea, et al. "Enrichment dynamics of Listeria monocytogenes and the associated microbiome from naturally contaminated ice cream linked to a listeriosis outbreak." *BMC microbiology* 16.1 (2016): 275.

Ponnusamy, Duraisamy, et al. "Cross-talk among flesh-eating Aeromonas hydrophila strains in mixed infection leading to necrotizing fasciitis." *Proceedings of the National Academy of Sciences* 113.3 (2016): 722-727.

Resources

More references can be obtained by emailing the authors or editors.

Food SafetyTech

Food Genomics

June 9, 2017

To Be or Not to Be: Choosing the Best Indicator using Microbiomes

By Douglas Marshall, Ph.D., Gregory Siragusa, Ph.D.

Now we have a tool for sighting more meaningful indicators, microbiomes.

Whenever an order is placed for an aerobic plate count, lactic acid bacteria count, Enterobacteriaceae count, coliform count, fecal coliform count, Escherichia coli count, or yeast and mold count, it involves ordering an indicator test. So, ordering an indicator test is only routine and frequent activity. In fact, many business-to-business transactions are partially dictated by the outcome of indicator tests in the form of purchase specifications. Raw material producers and ingredient manufacturers are required to deliver products that meet the expectations of the buyer. Should such product exceed the predefined specifications, the expected transaction becomes nullified. Finished food product manufacturers also must meet specifications set by retailers and food service buyers. Some regulatory jurisdictions and public health agencies also are in this game, offering regulatory specifications for targeted groups of indicator microbes, such as EPA water quality specifications or FDA zero tolerance of pathogens in ready-to-eat foods. Here we argue that microbiomes can be a valuable tool to help choose and validate the best indicator(s) that may be used under a variety of circumstances.

The microbial indicator premise is that the presence or population size of a single indicator microbe or groups of microbes has some respective correlation with the presence of or population of either undesirable microbes (spoilers or pathogens) or desirable microbes (starter cultures or probiotics). On the public health side, indicator presence or populations can be used to define risk of an adverse public health outcome. Indicators also have utility in assessing process effectiveness, such as presence or populations of spore formers after a heat process. Sanitation efficacy can be judged by the amount of an appropriate indicator, such as residual ATP on a surface or presence of Listeria spp. in a floor drain. Culture houses (i.e., starter cultures, probiotics) and companies that manufacture fermented foods can do routine QC testing for the amount of metabolic byproducts (CO2 or acid development) as an indicator of microbial activity and also measure culture population levels.

Our customers frequently ask us, “Which is the best indicator for my ingredients, process, and products?” Of course they are looking for a very simple answer but the reality is, we must know many details about the ingredients, product, process, and intended use before we can offer a best guess. Clearly best guesses, even by esteemed experts, can lead to inappropriate indicator choices. At worse, standard industry practice informed by years of use may not offer appropriate scientific validation of the use of chosen indicators.

Another customer question we frequently hear: "My product is not reaching intended shelf life, but indicator counts show it should be fine. What is causing product performance failure?" In this scenario, the chosen indicator(s) may not allow for cultivation of the offending microbe, resulting in an "all clear" test result. Each indicator test (see Table I) will grow only the microbes able to multiply on the selected medium, at the selected incubation temperature, for the selected incubation time, and under the selected incubation atmosphere.[1] Differences in media brand, or even slight deviations in media nutrition, media selective agents, temperature, time or atmosphere will have dramatic implications on what ultimately grows. What is telling is that there are many microbes in the sample that may not be cultivable at all, yet they may contribute to product performance failures. Wouldn't it be nice if you could run one test and get a good snapshot of the all microbiota present in the specimen?

A further issue is that some microbes, which may be perfectly able to grow under a certain set of conditions, might be outgrown by other competitors. Therefore, they may not contribute to the countable population. If they are found as a minor population, the odds of identifying them from a plate count are remote. Such microbes may in fact contribute to product failure and yet never be detected by an indicator assay.

An example of well-publicized historical misuse of indicators is the application of fecal coliform counts as indicators of fecal contamination of some dry leafy food products, such as tea leaves. For decades, periodic popular press exposé articles about food service iced tea with high fecal coliform counts have appeared in the news. The respective author's dramatic conclusion is that such teas are contaminated with feces and a threat to public health. However, in reality, when the bacteria associated with these high counts were actually identified, they were determined to be natural constituents of dry tea leaves and had no association with animal feces.[2]

An unconventional hypothetical indicator example seems worthy here. If you are a manufacturer of a dried ready-to-eat product or ingredient and your hazard analysis has identified Salmonella as a reasonably foreseeable environmental hazard, most will choose coliform, fecal coliform, *E. coli*, and/or Enterobacteriaceae counts as potential indicators. We're sure this sounds familiar so you're feeling pretty good about now—well, read on, please. What may be less obvious is the potential usefulness of a yeast and mold count for this purpose, because low-level moisture intrusion may lead to growth of these fungal groups and also may lead to enhanced survival/growth of Salmonella. Therefore, one may find the best indicator by looking for an indicator of moisture control problems rather than an indicator of potential fecal contamination.

Finally, verification screening of all raw materials, ingredients, processes, environmental locations, and products using traditional microbiology tests can quickly become expensive if you are looking at all the potential indicators shown in the table. By first running a single microbiome on a specimen, the predominant microbes and their relative proportional populations will be determined. This knowledge can be used to develop appropriate targeted verification screening for indicators that you now know are relevant to the specimen. Furthermore, the impact of changes in suppliers, processes, or product formulation can be measured using microbiomes to again gain confidence that appropriate indicators are still being used.

We hope this installment of Food Genomics triggers the reader to rethink the indicators they are using and ask the following questions:

- Why are we using our chosen indicators?
- Are our indicators telling us what we really need to know?
- Are there better indicators for my supplier verification program?
- Are there better indicators for my process verification program?
- Are there better indicators for my environmental monitoring program?
- Are there better indicators that more accurately predict product shelf life?

Indicator Test	Uses	Microbiome
Aerobic Mesophilic Plate Count	Estimate population of microbes able to grow at 35°C with air. Overall food quality indicator, shelf life/spoilage predictor	Names of predominant microbes and relative proportions of each constituting the aerobic mesophilic population
Anaerobic Mesophilic Plate Count	Estimate populations of microbes able to grow at 35°C without oxygen. Shelf life/spoilage predictor of vacuum packaged or modified atmosphere packaged foods.	Names of predominant microbes and relative populations of each constituting the anaerobic mesophilic population
Standard Plate Count	Similar to APC but used for dairy products. Estimate population of microbes able to grow at 30°C with air. Overall milk quality indicator, shelf life/spoilage predictor	Names of predominant microbes and relative populations of each constituting the aerobic mesophilic population
Psychrotrophic Plate Count	Estimate population of microbes able to grow at refrigerated temperatures (incubation temperature can vary from 5° to 15°C) with air. Shelf life/ spoilage predictor	Names of predominant microbes and relative populations of each constituting the aerobic psychrotrophic population
Anaerobic Psychro-trophic Plate Count	Estimate population of microbes able to grow at refrigerated temperatures without oxygen. Refrigerated shelf life/spoilage predictor of vacuum or modified atmosphere packaged foods	Names of predominant microbes and relative proportions of each constituting the anaerobic psychrotrophic population
Aerobic Thermophilic Plate Count	Estimate population of bacterial spores able to grow at high storage temperatures (incubation temperature can vary but usually >45°C) in air or survive a thermal process. Indicator of process failure. Spoilage indicator of improperly hot held foods	Names of predominant spores and relative proportions of each constituting the aerobic thermophilic population
Aerobic Mesophilic Spore	Estimate population of bacterial spores able to grow at 35°C with air. May indicate possible *Bacillus cereus* risk.	Names of predominant spores and relative proportions of each constituting the aerobic mesophilic spore population
Anaerobic Mesophilic Spore Count	Estimate population of bacterial spores able to grow at 35°C without oxygen. Potential shelf life/spoilage indicator of vacuum or modified atmosphere packaged foods. May indicate possible *Clostridium* botulinum risk.	Names of predominant microbes and relative proportions of each constituting the anaerobic psychrotrophic population
Aerobic Psychrophilic Spore Count	Estimate population of bacterial spores able to grow at refrigeration temperature with air. Spoilage indicator of refrigerated foods	Names of predominant spores and relative proportions of each constituting the aerobic psychrotrophic spore population
Anaerobic Psychrophilic Spore Count	Estimate population of bacterial spores able to grow at refrigeration temperature without oxygen. Potential shelf life/spoilage indicator of refrigerated vacuum or modified atmosphere packaged foods. May indicate possible nonproteolytic Clostridium botulinum risk	Names of predominant spores and relative proportions of each constituting the anaerobic psychrotrophic spore population
Aerobic Thermophilic Spore Count	Estimate population of bacterial spores able to grow at high temperature in air. Spoilage indicator of heat processed foods.	Names of predominant spores and relative proportions of each constituting the aerobic thermophilic spore population

Anaerobic Thermophilic Spore Count	Estimate population of bacterial spores able to grow at high temperature without oxygen. Spoilage indicator of heat processed, vacuum or modified atmosphere packaged foods	Names of predominant spores and relative proportions of each constituting the anaerobic thermophilic spore population
Thermoduric Plate Count	Estimate population of microbes able to survive a pasteurization process. Used as a shelf life/spoilage predictor	Names of predominant microbes and relative proportions surviving a thermal process
Lactic Acid Bacteria Count	Estimate population of bacteria able to produce lactic acid during growth. Indicator of fermentation success or spoilage failure	Names of predominant microbes and relative proportions that produce lactic acid
Proteolytic Plate Count	Estimate population of microorganisms that produce protease enzymes. Indicator of putrefactive spoilage potential	Names of predominant microbes and relative proportions that produce proteases
Lipolytic Plate Count	Estimate population of microorganisms that produce lipase enzymes. Indicator of lipid hydrolytic rancidity spoilage potential	Names of predominant microbes and relative proportions that produce lipases
Saccharolytic Plate Count	Estimate population of microorganisms that produce amylase enzymes. Indicator of starch hydrolysis spoilage potential	Names of predominant microbes and relative proportions that produce amylases
Pectinolytic Plate Count	Estimate population of microorganisms that produce pectinase enzymes. Indicator of pectin hydrolysis spoilage potential	Names of predominant microbes and relative proportions that produce pectinases
Aciduric Plate Count	Estimate population of microorganisms able to grow in a high acid/low pH food. Indicator of spoilage potential	Names of predominant microbes and relative proportions surviving an a high acid product
Aciduric Flat Sour Sporeformer Count	Estimate population of bacterial spores able to tolerate high acid foods and produce acid without gas production. Indicator of high-acid canned food spoilage potential	Names of predominant bacterial spores and relative proportions that grow in a high-acid canned food
Thermophilic Flat Sour Spore Former Count	Estimate population of bacterial spores able to grow at high temperature and produce acid. Indicator of canned food spoilage potential	Names of predominant bacterial spores and relative proportions that grow and produce acid in a canned food
Sulfide Sporeformer Count	Estimate populations of bacterial spores that produce sulfur aroma compounds. Indicator of canned food spoilage potential	Names of predominant bacterial spores that produce sulfur compounds
Halophilic Plate Count	Estimate population of microorganisms able to grow at high salt concentrations. Indicator of microbes that can spoil low water activity foods	Names of predominant microbes and relative proportions that grow in a high-salt food
Osmophilic Plate Count	Estimate population of microorganisms able to grow at high sugar concentrations. Indicator of microbes that can spoil low water activity foods	Names of predominant microbes and relative proportions that grow in a high-sugar food

Table I. Common microbial indicator tests and the use of microbiomes for validation of effectiveness.[1]

References

1. Salfinger, Y, and M.L. Tortorello. (2015). Compendium of Methods for the Microbiological Examination of Foods, 5th Ed. American Public Health Association, Washington, D.C.

2. Zhao, T., M.R.S. Clavero, M. Doyle, and L.R. Beuchat. (1997). Health relevance of the presence of fecal coliforms in iced tea and leaf tea. J. Food Prot. 60(3):215-218.

Food Genomics
September 1, 2017

GenomeTrakr: What Do You Know and What Should You Know?

By Douglas Marshall, Ph.D., Gregory Siragusa, Ph.D.

FDA's Eric Brown, Ph.D. and Marc Allard, Ph.D. explain in detail how FDA's program for characterizing foodborne pathogens works in a two part column.

This month we are happy to welcome our guest co-authors and interviewees Eric Brown, Ph.D. and Marc Allard, Ph.D. of CFSAN as we explore the FDA's GenomeTrakr program in a two-part Food Genomics column. Many of our readers have heard of GenomeTrakr, but are likely to have several questions regarding its core purpose and how it will impact food producers and processors in the United States and globally. In Part I we explore some technical aspects of the topic followed by Part II dealing with practical questions.

Part I: The basics of GenomeTrakr

Greg Siragusa/Doug Marshall: Thank you Dr. Allard and Dr. Brown for joining us in our monthly series, Food Genomics, to inform our readers about GenomeTrakr. Will you begin by telling us about yourselves and your team?

Eric Brown/Marc Allard: Hello, I am Eric, the director of the Division of Microbiology at the U.S. Food and Drug Administration at the Center for Food Safety and Applied Nutrition. Our team is made up of two branches, one that specializes in developing and validating methods for getting foodborne pathogens out of many different food matrices and the other branch conducts numerous tests to subtype and characterized foodborne pathogens. The GenomeTrakr program is in the subtyping branch as Whole Genome Sequencing (WGS) is the ultimate genomic subtyping tool for characterizing a foodborne pathogen at the DNA level.

Hello, my name is Marc, I am a senior biomedical research services officer and a senior advisor in Eric's division. We are part of the group that conceived, evaluated and deployed the GenomeTrakr database and network.

Siragusa/Marshall: Drs. Allard and Brown, imagine yourself with a group of food safety professionals ranging from vice president for food safety to director, manager and technologists. Would you please give us the 'elevator speech' on GenomeTrakr?

Brown/Allard: GenomeTrakr is the first of its kind distributed network for rapidly characterizing bacterial foodborne pathogens using whole genome sequences (WGS). This genomic data can help FDA with many applications, including trace-back to determine the root cause of an outbreak as well providing one work-flow for rapidly characterizing all of the pathogens for which the agency has responsibility. These same methods are also very helpful for antimicrobial resistance monitoring and characterization.

Siragusa/Marshall: From the FDA website, GenomeTrakr is described as "a distributed network of labs to utilize whole genome sequencing for pathogen identification." We of course have very time-proven methods of microbial identification and subtyping, so why do we need GenomeTrakr for identification and subtyping of microorganisms?

Brown/Allard: If all you want to know is species identification then you are correct, there are existing methods to do this. For some applications you need full characterization through subtyping (i.e., Below the level of species to the actual strain) with WGS. WGS of pathogens provides all of the genetic information about an organism as well as any mobile elements such as phages and plasmids that may be associated with these foodborne pathogens. The GenomeTrak network and database compiles a large amount of new genetic or DNA sequence data to more fully characterize foodborne pathogens.

GenomeTrakr and WGS are a means to track bacteria based on knowing the sequence of all DNA that comprises that specific bacterium's genome. It can be called the "ultimate identifier" in that it will show relationships at a very deep level of accuracy.

Siragusa/Marshall: Is it an accurate statement that GenomeTrakr can be considered the new iteration of PulseNet and Pulse field gel electrophoresis (PFGE)? Will PulseNet and PFGE disappear, or will PulseNet and GenomeTrkr merge into a single entity?

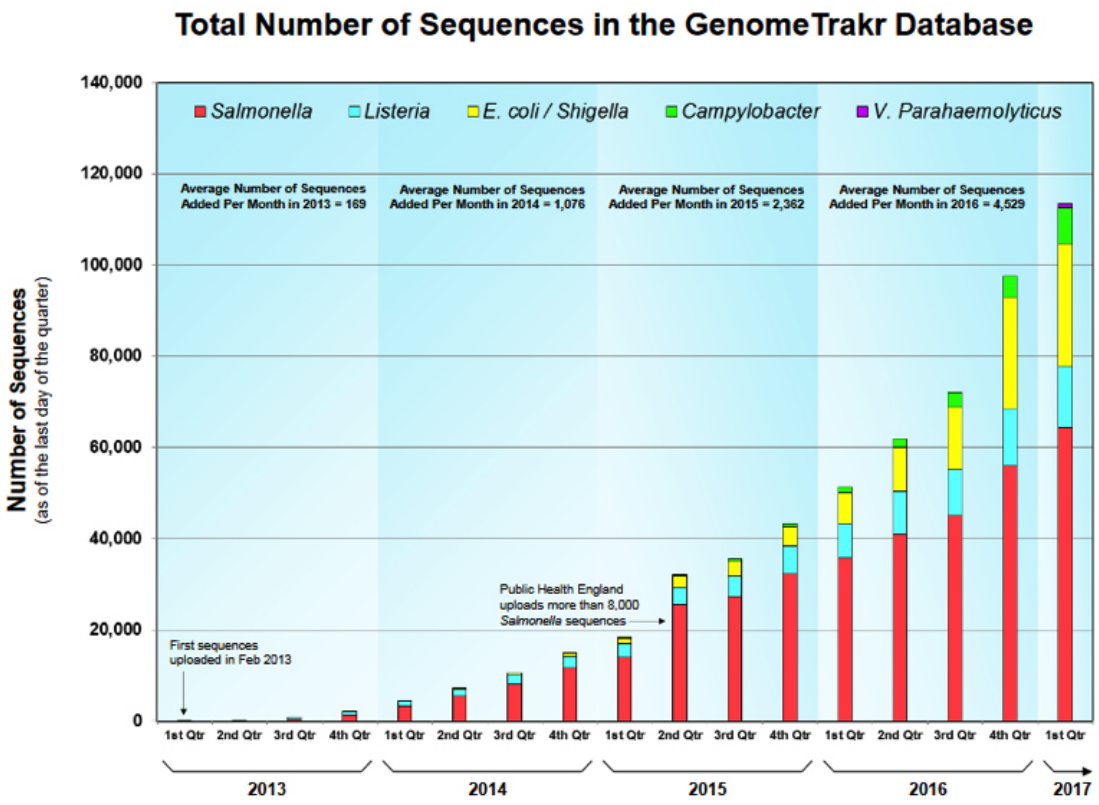
Brown/Allard: PulseNet is a network of public health labs run by the CDC, with USDA and FDA as active participants. The network is alive and well and will continue subtyping pathogens for public health. The current and historical subtyping tool used by PulseNet for more than 20 years is PFGE. It is expected that CDC, USDA and FDA's PFGE data collection will be replaced by WGS data and methods. That transformation has already begun. GenomeTrakr is a network of public health labs run by the FDA to support FDA public health and regulatory activities using WGS methods. Starting in 2012, this network is relatively new and is focused currently on using WGS for trace back to support outbreak investigations and FDA regulatory actions. CDC PulseNet has used WGS data on Listeria and collects draft genomes (i.e., unfinished versions of a final genome are used for quicker assembly) of other foodborne pathogens as well, and USDA's FSIS has used WGS for the pathogens found on the foods that they regulate. All of the data from GenomeTrakr and Pulsenet are shared at the NCBI Pathogen Detection website *(see Figure 1 on the next page).*

Siragusa/Marshall: Does an organism have to be classified to the species level before submitting to GenomeTrakr?

Brown/Allard: Yes, species-level identification is part of the minimal metadata (all of the descriptors related to a sample such as geographic origin, lot number, sources, ingredients etc.) required to deposit data in the GenomeTrakr database. This allows initial QA/QC metrics to determine if the new genome is labeled properly.

Siragusa/Marshall: After an isolate is identified to the species level, would you describe to the reader what the basic process is going from an isolated and speciated bacterial colony on an agar plate to a usable whole genome sequence deposited in the GenomeTrakr database?

(Figure 1)



Brown/Allard: The FDA has a branch of scientists who specialize in ways to isolate foodborne pathogens from food. The detailed methods used ultimately end up in the Bacteriological Analytical Manual (BAM) of approved and validated methods. Once a pathogen is in pure culture then DNA is extracted from the bacterial cells. The DNA is then put into a DNA sequencing library, which modifies the DNA to properly attach and run sequencing reactions depending on the specific sequencing vendor used. The sequence data is downloaded from the sequencing equipment and then uploaded to the National Center for Biotechnology Information (NCBI) Pathogen Detection website. The database is publicly open to allow anyone with foodborne pathogens to upload their data and compare their sequences to what is available in the database.

Siragusa/Marshall: Suppose a specific sequence type of a foodborne bacterial pathogen is found and identified from a processing plant but that the plant has never had a positive assay result for that pathogen in any of its history of product production and ultimate consumption. If an outbreak occurred somewhere in the world and that same specific sequence type were identified as the causative agent, would a company be in anyway liable? Could one even make an association between the two isolates with the same sequence type isolated at great distances from open another?

Brown/Allard: The genetic evidence from WGS supports the hypothesis that the two isolates shared a recent common ancestor. If, for example, the isolate from the processing plant and the outbreak sample where genetically identical across the entire genome, the prediction is that the two samples are connected in some way that is currently not understood. The genetic matches guide the FDA and help point investigations to study the possible connections. This might

include additional inspection of the processing plant as well as linking this to the typical epidemiological exposure data. Sometimes due to the indirect nature of how pathogens circulate through the farm to fork continuum and the complex methods of trade, no connection is made. \More commonly, these investigative leads from genetic matches help the FDA establish direct links between the two bacterial isolates through a shared ingredient, shared processing, distribution or packaging process. The genetic information and cluster helps the FDA discover new ways that the pathogens are moving from farm to fork. We are unaware of any example where identical genomes somehow independently arose and were unrelated. This is counter to molecular evolutionary theory anyway. Genetic identity equals genetic relatedness and the closer two isolates are genetically to each other, the more recent that they shared a common ancestor. With regard to liability, this is a topic beyond the scope of our group, but genomic data does not by itself prove a direct linkage and that is why additional investigations must follow any close matches.

Siragusa/Marshall: We know that SNPs (Single Nucleotide Polymorphisms or single base pair differences in the same location in a genome) are commonly used to distinguish clonality of bacteria with highly similar genomes. Are there criteria used by GenomeTrakr bioinformaticists that are set to help define what is similar, different or the same?

Brown/Allard: As the database grows with more examples of diverse serotypes or kinds of foodborne pathogens, the FDA WGS group is observing common patterns that can be used as guidance to define what is same or different. For example, closely related for Salmonella and E. coli are usually in the five or fewer SNPs, and closely related for Listeria is 20 or fewer SNPs using the current FDA validated bioinformatics pipeline. These values are not set in stone but should be considered more like guidance for what FDA and GenomeTrakr have observed already from earlier case studies that have already been collected and examined. Often, a greater number (e.g., 21-50) of SNP differences have been observed between strains isolated in some outbreaks. Any close match might support or direct an outbreak investigation if there is evidence that suggests that a particular outbreak looks most closely like an early case from a specific geographic location. WGS data helps investigators focus their efforts toward and international verses domestic exposure or possible country of origin. Even more divergent WGS linkages, when SNPs are greater than 50-100, often connect to different foods or different geographic locations that would lead investigators away from the source of an outbreak as the data provides both inclusivity as well as exclusivity.

When two strains have more than 50–100 SNPs, different food or geographic sources of those trains can be incorrectly linked resulting in investigators pursuing an incorrect source.

Siragusa/Marshall: Can SNPs be identified from different agar-plate clones of the same strain\ (i.e., Different colonies on the same plate)?

Brown/Allard: Since understanding the natural genetic variation present in foodborne pathogens is the basis to understanding relatedness, the FDA conducted validation experiments on growing then sequencing colonies from the same plate, colonies from frozen inocula, thawing and plating, as well as running the same DNAs on different instruments and with different sequencing technicians. The FDA's work with Salmonella enterica Montevideo sequencing as well as ongoing proficiency testing among laboratories shows that the same isolate most often has no differences, although some samples have 1-2 SNP differences. Genetic differences observed in isolates collected by FDA inspectors all related to a common outbreak generally have more genetic differences, and this appears to be dependent on the nature of the facility and the length of time that the foodborne pathogen has been resident in the facility and the selective pressure to which the pathogen was exposed to in a range from 0–5 SNPs different.

Siragusa/Marshall: Regarding the use of WGS to track strains in a particular processing plant, is it possible that within that closed microenvironment that strains will evolve sufficiently so that it becomes unique to that source?

Brown/Allard: Yes, we have discovered multiple examples of strains that have evolved in a unique way that they appear to be specific to that source. Hospitals use the same practice to understand hospital-acquired infections and the routes of transmission within a hospitals intensive care unit or surgery. Food industry laboratories as well as FDA investigators could use WGS data in a similar way to determine the root cause of the contamination by combining WGS data with inspection and surveillance. The FDA Office of Compliance uses WGS as one piece of evidence to ask the question: Have we seen this pathogen before?

Siragusa/Marshall: The number of sequences in the GenomeTrakr database is approaching 120,000 (~4,000 per month are added). Are the sequences in the GenomeTrakr database all generated by GenomeTrakr Network labs?

Brown/Allard: The sequences labeled as GenomeTrakr isolates at the NCBI biosample and bioproject databases are the WGS efforts supported by the U.S. FDA and USDA FSIS. GenomeTrakr is a label identifying the FDA, USDA FSIS and collaborative partner's efforts to sequence food and environmental isolates. Additional laboratories, independent and beyond formal membership in the GT network, upload WGS data to the NCBI pathogen detection ebsite of which GenomeTrakr is one part. CDC shares WGS data on primarily clinical PulseNet

isolates and USDA FSIS shares WGS foodborne pathogens for foods that they regulate. Numerous international public health laboratories also upload WGS data to NCBI. The NCBI pathogen detection website includes all publicly released WGS data for the species that they are analyzing, and this might include additional research or public health data. The point of contact for who ubmitted the data is listed in the biosample data sheet, an example of which can be seen here.

Siragusa/Marshall: Once sequences are deposited into the GenomeTrakr database, are they also part of GenBank?

Brown/Allard: The majority of the GenomeTrakr database is part of the NCBI SRA (sequence read archive) database, which is a less finished version of the data in GenBank. GenBank data is assembled and annotated, which takes more time and analysis to complete. Once automated oftware is optimized and validated, NCBI likely will place all of the GenomeTrakr data into GenBank. Currently, only the published WGS data from GenomeTrakr is available in GenBank. All of the GenomeTrakr data is available in SRA both at GenomeTrakr bioprojects and in the NCBI pathogen detection website.

Readers, look for the Part II of this column where we continue our exploration with Drs. Brown nd Allard and ask some general questions about the logistics surrounding GenomeTrakr. As always, please contact either Greg Siragusa or Doug Marshall with comments, questions or ideas for future Food Genomics columns.

About the Interviewees

Marc W. Allard. Ph.D.
Marc Allard, Ph.D. is a senior biomedical research services officer specializing in both phylogenetic analysis as well as the biochemical laboratory methods that generate the genetic information in the GenomeTrakr database, which is part of the NCBI Pathogen Detection website. Allard joined the Division of Microbiology in FDA's Office of Regulatory Science in 2008 where he uses Whole Genome Sequencing of foodborne pathogens to identify and characterize outbreaks of bacterial strains, particularly Salmonella, E. coli, and Listeria. He obtained a B.A. from the University of Vermont, an M.S. from Texas A&M University and his Ph.D. in biology in from Harvard University. Allard was the Louis Weintraub Associate Professor of Biology at George Washington University for 14 years from 1994 to 2008. He is a Fellow of the American Academy of Microbiology.

Eric W. Brown, Ph.D.
Eric W. Brown, Ph.D. currently serves as director of the Division of Microbiology in the Office of Regulatory Science. He oversees a group of 50 researchers and support scientists engaged in a multi-parameter research program to develop and apply microbiological and molecular genetic strategies for detecting, identifying, and differentiating bacterial foodborne pathogens such as Salmonella and shiga-toxin producing E. coli. Brown received his Ph.D. in microbial genetics from The Genetics Program in the Department of Biological Sciences at The George Washington University. He has conducted research in microbial evolution and microbial ecology as a research fellow in the National Cancer Institute, the U.S. Department of Agriculture, and as a tenure-track Professor of Microbiology at Loyola University of Chicago. Brown came to the Food and Drug Administration in 1999 and has since carried out numerous experiments relating to the detection, identification, and discrimination of foodborne pathogens.

Food Genomics
October 4, 2017

Part II: Logistics of GenomeTrakr

By Douglas Marshall, Ph.D., Gregory Siragusa, Ph.D.

FDA answers more general questions surrounding the program.

This month in Food Genomics we asked FDA scientists Drs. Marc Allard and Eric Brown to help the readers of FoodSafety Tech understand the process used by GenomeTrakr. In part two we cover some logistical and more general questions.

Greg Siragusa/Douglas Marshall: Why should a food producer or processor submit its own pathogen isolates to GenomeTrakr? Are there any legal liabilities incurred by doing so?

Eric Brown/Marc Allard: The database is available publicly for any outside laboratory to be able to rapidly compare their new WGS data to all of the data in the database. The data is all publicly available so food industry members should carefully consider the strengths and weaknesses of sharing data. The main reason for sharing data is that if any matches arise then this would be immediately known for an investigation and corrective action. With knowledge, companies can better understand their risk and exposure to occasional contamination events.

S**iragusa/Marshall:** Are there private third-party providers who will perform the same method of sequence analysis for private companies that GenomeTrakr uses in the FDA?

Brown/Allard: Yes, as all of the FDA methods of data collection and analysis are fully transparent and publicly available, any expert third-party provider could easily set up and reproduce the GenomeTrakr methods. Third-party support may be an excellent mechanism for food industry partners that wish to examine the pathogens they have found connected to their products but do not wish to maintain an active WGS laboratory. An internet and reference search will uncover these private third-party providers, as this is a growing market with a diversity of services provided. The FDA works closely with the Institute for Food Safety and Health (IFSH) to share information that may be valuable to their industry partners.

Siragusa/Marshall: Will the FDA perform analysis of isolates for private parties and the sequence not made publicly available?

Brown/Allard: No. While we will sequence relevant strains from many different sources, as a matter of protocol we will submit all of these data to the GenomeTrakr database. That is, currently, the FDA sequences and uploads all available genomic strain data. All data are made publicly available through the GenomeTrakr and NCBI pathogen detection website. The metadata describing each isolate only includes species, date, state location and a general food description which could include the type of food (e.g., an egg) and/or the type of sample (e.g., environmental swab, surface water, sediment, etc.) as well as production date, pH, fat content and water activity. No trade or industry brand names are made publicly available, and the location is ambiguous down to the state level to allow for anonymity of specific farm names or processing centers. An example of metadata in the GenomeTrakr database might include Salmonella, from Washington State in spinach from 2015.

Siragusa/Marshall: Is the CDC tied into GenomeTrakr and if so, how?

Brown/Allard: CDC labels their clinical WGS data as PulseNet with the data uploaded to the NCBI Pathogen Detection website. USDA FSIS also uploads the isolates that they have collected and sequenced from foods that they regulate. All of this WGS data is housed in a centralized repository at NCBI Pathogen Detection website where NCBI conducts rapid analysis for QA/QC. The NCBI posts a daily tree for all species that recently have been uploaded. This way all of the data collected by these federal laboratories and their state and international partners are made publicly available for direct comparison. Numerous other international and academic laboratories also provide data to the NCBI centralized database. When isolates cluster together and appear to be closely related, the FDA works with CDC and USDA FSIS through the normal channels. The great benefit of combining food, environmental and clinical isolate genomes in a common database cannot be overstated.

Siragusa/Marshall: In the event of an outbreak, is it possible to obtain WGS's from using a shotgun metagenome (a microbial and organismic profile obtain by sequencing all of the DNA in a sample, not just bacterial analysis of an enrichment thereby precluding isolation? (Refer to lossary; see Table 1)

Brown/Allard: Yes, preliminary research has documented the potential to obtain WGS data from cultural enrichments, saving the time it takes for full pure culture isolation, which potentially could provide time savings of two to five days depending on the pathogen. Having well characterized draft genomes such as those in the GenomeTrakr database will support rapid characterization from metagenomes after cultural enrichment. A future goal for the FDA is to transform and expand GenomeTrakr into metaGenomeTrakr to support either pure culture or enriched shotgun metagenomic samples.

Siragusa/Marshall: Is there any way that associated metadata tied to a strain (and hence its sequence) can be unmasked through legal action?

Brown/Allard: FDA protects confidential metadata collected during inspection just as it has always done with PFGE data. WGS data is protected at the same level as other types of subtyping information.

Siragusa/Marshall: Is the GenomeTrakr database associated with the GMI (Global Microbial Identifier)?

Brown/Allard: The GMI is a consortium of like-minded public health scientists who wish to collaborate to create a harmonized global system of DNA genome databases that is publicly available to promote a one-health approach. The GenomeTrakr is one of the databases that make up this larger effort that includes some data from members of the GMI.

Siragusa/Marshall: This column is meant to keep food safety professionals abreast of the latest knowledge, technology and uses of genomics for food safety and quality. Tell us your vision of how or which changes in technology (sequencing chemistry, bioinformatics, etc.) will be coming down the pike and how it might impact GenomeTrakr?

Brown/Allard: New technology has been constantly improving in WGS and in sequencing for the last 20 years, and there is no sign of this slowing down. Improvements continue to accrue in chemistry, equipment and software analysis. Likely future improvements will include more turnkey solutions for WGS from sample to report. This includes both DNA extraction and library preparation for sequencing, as well as data analysis pipelines (the system of analyzing the actual sequence data) that provide rapid, accurate and simple language results. Smaller mobile WGS devices are starting to show feasibility that would bring the lab to the samples and decrease the time to an answer (See: https://nanoporetech.com/products/minion) Metagenomics approaches appear to be maturing so that technology improvements are moving this out of a research phase and into direct applications. Currently MISeq (a commonly used workhorse nucleic acid sequencer made by the Illumina Co.) outputs are on the order of 300 base pair read lengths of nucleotides (i.e. A's, T's. C's G's), long read sequencing technologies, upwards of 1,500 base pairs may make analysis much easier so that more assembled and completed finished genomes are available in the databases. Cloud-based solutions of data analysis pipelines may provide simple solutions, giving wider access to rapid, validated data analysis and results. FDA researchers are working on all of these aspects of improvements in WGS technology as well as expanding the network to more global partners.

Siragusa/Marshall: Sequences deposited into GenBank (as part of GenomeTrakr) are accessible to anyone anywhere. Does this essentially usher in a whole new chapter in food microbiology especially at the pre-harvest level?

Brown/Allard: Yes, having well characterized reference genomes provided by GenomeTrakr partners will support microbial ecology and metagenomics studies. Metagenomics or microbiomes describing which species are present and what they may be doing in the ecology is providing new knowledge in all aspects of the farm to fork continuum. As the costs for these services decrease, we are seeing an increase in use to answer questions that have been impossible or extremely difficult in the past.

Siragusa/Marshall: GenomeTrakr is not a project per se; rather it is a program. How is it funded and will it continue on stable fiscal footing for the foreseeable future?

Brown/Allard: GenomeTrakr started as a research project in the Office of Regulatory Science in CFSAN, but much of this data collection is no longer research. Today, and for some time in the future, WGS at the FDA is collected as fully validated regulatory data to support outbreak and compliance investigations. As such, the FDA is in transition of moving WGS into a phase for more stable regulatory support. Research and development for future applications and technology exploration will always be a part of the FDA portfolio, although typically at lower funding levels than the regulatory offices. Public health funding is generally protected as everyone wants safe food.

Siragusa/Marshall: Are there any restrictions of isolate source? For instance, can isolates from poultry flocks or even wild birds be deposited?

Brown/Allard: The GenomeTrakr and NCBI pathogen detection databases are open to the public and thus there are no restrictions as long as the minimal metadata and QA and QC metrics are met. Current GenomeTrakr WGS foodborne pathogen data includes samples from

both poultry and wild birds, as well as turtles, snakes and frogs. Members interested in what is in the database can go to the NCBI Pathogen Detection website and filter on simple words like avian, bird, gull, chicken, wheat, avocado, etc. An example is as follows for a snake.

Siragusa/Marshall: If a company deposits an isolate, will it have access to the GenomeTrakr derived sequence exclusively or at least initially for some period before that information becomes public?

Brown/Allard: No, currently the FDA does not hold WGS data. All data collected by the FDA is uploaded and released publicly at the GenomeTrakr bioprojects and at NCBI pathogen detection website with no delays. If companies wish to hold data then they need to look to third-party solutions for their needs. The reason that GenomeTrakr has been so successful is due to the real-time nature of the released information and that it is globally available.

Siragusa/Marshall: Will the extensive data obtained on Salmonella be fuel to finally develop a sequence based serotyping surrogate for the genus?

Brown/Allard: Yes, the FDA already collaborates with University of Georgia investigators using a program called SeqSero that rapidly identifies serotype from the draft WGS. Other software tools are also being built to go from sequence to serotype.

Siragusa/Marshall: Currently there are a few foodborne pathogens that comprise the GenomeTrakr program's database. How easy would it be to expand the scope of microorganisms under the GenomeTrakr umbrella? For instance, would a private company be able to use the GenomeTrakr resources to build their own private database? Will the realm ever be expanded to mycotoxigenic fungi or foodborne viruses?

Brown/Allard: Yes, it is relatively easy to take any of the data that is publicly available in GenomeTrakr and download the data to build a private database to add value or to provide additional tools to a private user group. This is already happening with GenomeTrakr data. In addition, the NCBI mechanisms are not species specific, and so any private group can build a bioproject of draft WGS for any species that they wish. To use the NCBI bioproject tools the data would have to be publicly released after a year. Similarly, the NCBI pathogen detection website, which does QA/QC and builds a new phylogenetic tree every time there is new data, is currently open to any human pathogen. The main constraint to building a new species database is that the people who want to build such a database would need to speak to NCBI representatives to convince them that enough data was going to be uploaded and released publicly and that there would be enough use to justify NCBIs efforts to validate and test these new pipelines. Industry should also ask about non-pathogens such as spoilage organisms, although this may be out of the scope of NCBI. The FDA plans to expand to foodborne virus such as Hepatitis A, although fungi may have lower priority. The FDA is willing to work with industry to identify and populate mutually beneficial databases that the industry will use to improve food safety. For example, the FDA has suggested that there may be value of typing foodborne pathogens that have shown known resistance to cleaners and sterilizers with the goal of understanding how these pathogens are avoiding the preventative controls that are put in place to keep them out of the food supply. Better characterization of the genes responsible for resistance may lead to rapid PCR tools to understand and screen for the ability of pathogens to persist in the environment.

Siragusa/Marshall: We would like to end with the following question; why should food producers and processors embrace GenomeTrakr and become a part of it (i.e., Why should a company send in its pathogen isolates)?

Brown/Allard: The FDA understands that most food producers will not likely send their strains to the FDA or make WGS data they have generated publicly available due to a concern for legal liability risks. This then highlights the importance of third-party members providing assistance to the food industry to utilize the power of WGS data for understanding their own processing facilities and supply chains. The FDA is committed to support industry adoption of these WGS methods to improve food safety, which can be accomplished by data sharing, methods validation and data interpretation and education.

Siragusa/Marshall: Thank you Drs. Allard and Brown. That was most informative and we appreciate your sharing this knowledge.

Readers, we hope this interview will be useful to your work and knowledge base. Obtaining whole genome sequences of organisms is a routine service offered by many labs worldwide for reasonable costs. At the current time, there are no standardized methods, but we are informed that it is possible at some time in the future that an official method of analysis (OMA) will be formulated, similar to our other cultural tools, making WGS a tool available on a routine rather than special project basis with more comparable results. For more information on the GenomeTrakr Network of labs visit the GenomeTrakr Network.

As always, please contact either Greg Siragusa or Doug Marshall with comments, questions or ideas for future Food Genomics columns.

Resources

1. Allard, M. W., Strain, E., Melka, D., Bunning, K., Musser, S. M., Brown, E. W., & Timme, R. (2016). Practical Value of Food Pathogen Traceability through Building a Whole-Genome Sequencing Network and Database. Journal of Clinical Microbiology, 54(8), 1975–1983. https://doi.org/10.1128/JCM.00081-16

2. Burall, L. S., Grim, C. J., Mammel, M. K., & Datta, A. R. (2016). Whole Genome Sequence Analysis Using JSpecies Tool Establishes Clonal Relationships between Listeria monocytogenes Strains from Epidemiologically Unrelated Listeriosis Outbreaks. PloS One, 11(3), e0150797. https://doi.org/10.1371/journal.pone.0150797

3. Nutrition, C. for F. S. and A. (n.d.-a). Whole Genome Sequencing (WGS) Program – GenomeTrakr Fast Facts [WebContent]. Retrieved May 4, 2017, from https://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/ucm403550.htm
4. Nutrition, C. for F. S. and A. (n.d.-b). Whole Genome Sequencing (WGS) Program – Proactive Applications of Whole Genome Sequencing Technology [WebContent]. Retrieved May 4, 2017, from https://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/ucm422077.htm

5. Zhang, S., Yin, Y., Jones, M. B., Zhang, Z., Deatherage Kaiser, B. L., Dinsmore, B. A., Deng, X. (2015). Salmonella serotype determination utilizing high-throughput genome sequencing data. Journal of Clinical Microbiology, 53(5), 1685–1692. https://doi.org/10.1128/JCM.00323-15

About the Interviewees

Marc W. Allard. Ph.D.
Marc Allard, Ph.D. is a senior biomedical research services officer specializing in both phylogenetic analysis as well as the biochemical laboratory methods that generate the genetic information in the GenomeTrakr database, which is part of the NCBI Pathogen Detection website. Allard joined the Division of Microbiology in FDA's Office of Regulatory Science in 2008 where he uses Whole Genome Sequencing of foodborne pathogens to identify and characterize outbreaks of bacterial strains, particularly Salmonella, E. coli, and Listeria. He obtained a B.A. from the University of Vermont, an M.S. from Texas A&M University and his Ph.D. in biology in from Harvard University. Allard was the Louis Weintraub Associate Professor of Biology at George Washington University for 14 years from 1994 to 2008. He is a Fellow of the American Academy of Microbiology.

Eric W. Brown, Ph.D.
Eric W. Brown, Ph.D. currently serves as director of the Division of Microbiology in the Office of Regulatory Science. He oversees a group of 50 researchers and support scientists engaged in a multi-parameter research program to develop and apply microbiological and molecular genetic strategies for detecting, identifying, and differentiating bacterial foodborne pathogens such as Salmonella and shiga-toxin producing E. coli. Brown received his Ph.D. in microbial genetics from The Genetics Program in the Department of Biological Sciences at The George Washington University. He has conducted research in microbial evolution and microbial ecology as a research fellow in the National Cancer Institute, the U.S. Department of Agriculture, and as a tenure-track Professor of Microbiology at Loyola University of Chicago. Brown came to the Food and Drug Administration in 1999 and has since carried out numerous experiments relating to the detection, identification, and discrimination of foodborne pathogens.

ABOUT THE AUTHORS

Sanjay K. Singh
Technical Marketing and Product Manager
Eurofins MWG Operon LLC

Sanjay K. Singh is the technical marketing and product manager at Eurofins MWG Operon LLC.

Douglas Marshall
Chief Scientific Officer
Eurofins Microbiology Laboratories

Douglas Marshall is chief scientific officer with Eurofins Microbiology Laboratories, Inc., a division of the global life sciences company Eurofins Scientific. He also is cofounder and director of the Food Safety Institute, LLC, an integrated consulting and analytical services company affiliated with Eurofins. He currently holds adjunct professor positions with Colorado State University and Florida State College. His former positions include associate dean and professor of public health at the College of Natural and Health Sciences, University of Northern Colorado; adjunct professor with the Colorado School of Public Health; professor of food science, nutrition, and health promotion at Mississippi State University; assistant professor of food science at Louisiana State University; contributing editor for the peer-reviewed scientific journal Food Microbiology, and four consecutive terms on the editorial board of the Journal of Food Protection. He is a frequent consultant to NIH, WHO, FAO, USDA and other government agencies and private companies. His research and expertise has been featured in popular press venues such as Consumer's Reports, Fine Cooking, USA Today, Fitness, Health, Men's Health, Chemtech, Nature Science Updates, and ASM Journal Highlights. Marshall is a frequently invited speaker and a prolific book chapter writer. With more than 250 publications and more than 150 invited presentations, his scientific research and outreach interests focus on improving

the microbiological quality and safety of foods. Among these was the completion of the four-volume Handbook of Food Science, Technology, and Engineering, which he co-edited. He has been the recipient of a number of awards for his scholarly efforts including the Mississippi Chemical Corporation Award of Excellence for Outstanding Work and the International Association for Food Protection Educator Award. He is a fellow of the Institute of Food Technologists, where he has previously served as chair of two divisions and two regional sections, member of the board of directors, an Inaugural member and chair of the International Food Science Certification Commission, and a founding member of the Global Traceability Center.

Gregory Siragusa
Senior Principal Scientist
Eurofins Microbiology Laboratories

Gregory Siragusa is senior principal scientist with Eurofins Microbiology Laboratories, Inc., a division of the global life sciences company Eurofins Scientific. He has held positions with Danisco/DuPont and the USDA. He has been a reviewer for the Journal of Food Protection and Applied Environmental Microbiology.

Siragusa's research spans fields of microbiology focusing on foods. He has been a speaker on the topics of food genomics and antibiotic alternatives. He obtained the B.S./M.S. in microbiology from Louisiana State University and a Ph.D. from the University of Arkansas. He has authored more than 100 peer-reviewed papers, chapters and abstracts. His latest activities focus on applications of genomics to food microbiology.

Nur A. Hasan
VP of Research and Development
CosmosID, Inc.
nur.hasan@cosmosid.com

Food Genomics
December 13, 2017

What's in a name? - Probiotic Analysis and Genomics

By Gregory Siragusa, Ph.D., Douglas Marshall, Ph.D.

In the era of newer probiotic organisms, asking the right questions is crucial.

In short, in the world of regulatory and probiotic microbiology the "name" is critical. Whether you are a probiotics manufacturer, blender, or user, we are all likely aware that usage and sales of probiotic strains of bacteria and yeasts is burgeoning. Estimates of sales growth are impressive, with $24 billion and $5 billion USD in human and animal markets respectively projected by the year 2024.("Global Probiotics Market to Reach $24 Billion by 2017," 2013) ("Probiotics in Animal Feed Market Worth 5.07 Billion USD by 2022 | Markets Insider,") On a more personal level, prove it to yourself by visiting your local grocery store and take a trip down the aisle where probiotic supplements are displayed. Read content labels and see if you recognize the microbe names. There are many probiotic organisms' names you are likely familiar with, most of which are lactic acid bacteria (LAB). However, we are quickly entering an age of more novel or even new probiotic organisms that may be unfamiliar to you. Some of these are not always as easy to culture as the LAB.(Treven, 2015) On the same labels you may see claims of viability and cell population declarations (usually in CFU's or colony forming units). Also notice that many probiotics are retailed in dry form, while others are marketed as

liquids. As food safety scientists and practitioners, questions are probably popping out of your head as to how probiotic species and populations are verified and how these various preparations survive expected shelf life.

Most will agree that before anyone starts consuming pills or eating foods with billions of viable bacteria, it is obviously a prudent idea that the manufacturer has the means to assure safety and quality. The details and scope of probiotic safety and microbial analysis are much too complex and broad to deal with in a few pages. For more details we direct the reader to two key publications. ("Guidelines for Probiotics in food," n.d.) (Huys et al., 2013)

Identity, viability, and populations are attributes largely measurable by methods that rely on culture, phenotypic analysis, genomics, and combinations thereof. Here we will share a primer on genomic methods for probiotic analysis starting with a very basic aspect critical to all of microbiology – taxonomy and asking the "right" questions.

'Why not just do plate count?' and Revisiting Taxonomy

The process of identifying or classifying organisms, also known as the science of taxonomy or systematics, has often been given less than stellar treatment among the community of microbiologists. We are frustrated with the tendency of taxonomists to change genus or species names just as people learned the old existing names. But every dog will have its day and for microbial systematists that day has arrived since application of genomic tools to the taxonomist toolbox has coincided with growth of the probiotics industry. Practically speaking for the probiotic microbiologist, there is a lot more to a name than just nomenclature. Microbial taxonomy, and specifically bacterial taxonomy, becomes vitally important as more and more products are produced and as regulations increase in scope. Bacterial nomenclature is an ever

changing field, but at least naming has become more centralized with its own website (Parte, A.C., 2017).

For the probiotic manufacturer, some important questions require answers: “Is it ‘my’ strain?”, “What’s in the mixture?”, “Is the label accurate?”, and “Are they alive?”. So why are we addressing this topic as a subject for *Food Genomics*? Confronted with the shear variety of bacterial names it is easy to see why and how genomic tools offer a solution to this complexity. We now have tools that augment, complement, or even in some cases, replace cultural microbiology as a means to classify, identify, and analyze probiotics (see table 1) (Davis, 2014).

“Why not just perform a plate count?”. Obviously plate counts have a pivotal role in the analytical microbiology of probiotics and will likely remain a gold standard for enumeration of viable counts. In fact the unit of viable cell counts, colony forming units or CFU, is the recommendation for use to verify probiotic populations (Daniels, n.d.). Unfortunately, most plate counts do not name or identify the microbes we count as colonies. Occasionally, with precise selective and differential media the identity of the colonies growing on the plate can be reliably called, but misidentification is common. Other tools, such as PCR, can be useful in counting genes from an organism’s DNA. Quantitative PCR (qPCR) and flow cytometry both rely on a probe specific for a species or even a strain in order to estimate cell numbers, including viable cell counts (Bunthof & Abee, 2002) (Sattler, Mohnl, & Klose, 2014) (Kramer, Obermajer, Bogovic Matijasić, Rogelj, & Kmetec, 2009) (Davis, 2014).

So how will modern genomics help you with the analysis of your probiotics? The following are some questions, examples, and comments that illustrate genomic applications for probiotic analysis that you should be familiar with. These methods, whether sequencing-based, PCR-based, or flow cytometry-based, all require some form of sequence determination, detection, and analysis.

Table 1: Genomic Tools for Probiotic Analysis

General Method	Application Notes
Targeted Microbiome	• Genus/Species level resolution • Bacterial or fungal profiles (16S/ITS gene) • Suitable for multi-strain probiotic products • Unknown or QA analysis
Shotgun Metagenome	• Species/Possible Strain level resolution • Bacterial and fungal profiles in same assay • Well suited for multi-strain probiotic products
Qualitative Microarray	• Species/Possible Strain level resolution • Non-quantitative, descriptive only • Well suited for multi-strain probiotic products
PCR	• Species/Strain, probe designed specificity • Qualitative
qPCR	• Species/strain probe designed specificity • Quantitative against CFU standard curve • Can be designed to detect viability
Flow Cytometry	• Probe designed specificity • Quantitative against CFU standard curve • Viable, injured, dead cell detection possible • High throughput

First, let's begin by clarifying some terminology. As it pertains to taxonomy or identification, it is important to specify to what level of identification is required. The time-honored system of Linnaean classification (See Figure 1) is still our basis for microbial taxonomy or identification.

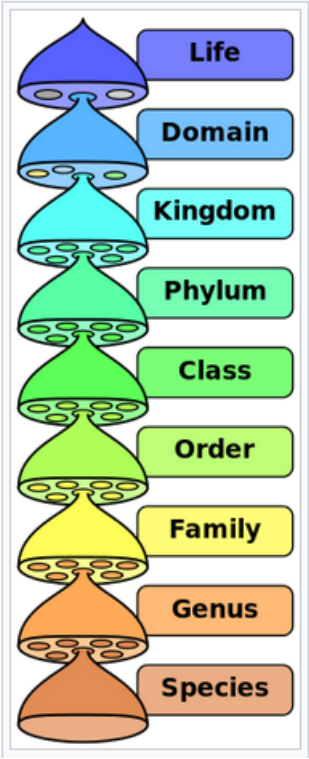


Figure 1. Linnaean classification hierarchical categories or ranks. (Source: https://simple.wikipedia.org/wiki/Biological_classification)

As we will see, in most cases for bacterial identification, our good friend the 16S gene sequence analysis, while not for all cases, is normally well suited to obtain a species name (Hug et al., 2016). What about use of the term "strain"? In some cases the term strain has a dual meaning. It can refer just to any organism added to a product. For example, "This is a three strain product", meaning it contains three different species of probiotic microorganisms. However, the term "strain" is frequently misused as synonomous with species. In the taxonomic sense, a strain represents a subtype or subspecies ranking. Technically, if correctly classified, a strain is but a different member of the same species group. Sounds picky doesn't it? Perhaps, but it is always necessary to be able to assure any label claims and at this point in time we are still using the Linnaean system of nomenclature for nomenclature. Strain or subspecies names might be based solely on extrinsic

factors like the name of the person who isolated the strain, or the number of the colony picked from a culture plate, or the source, a place, the date, etc. It also might be based on intrinsic biological traits, e.g. a specific trait or single base pair difference in an important gene of the species or even distinguishing a pathogen from its non-pathogemnic cousin (Natarajan & Madasamy, 2014).

In an example (Figure 2), we see that there are 130 strains deposited in culture collections that are all speciated as *Lactobacillus acidophoilus*. In order to get to the "strain" level of classification one must have a unque identifier that is a signature for any particular strain. That identifier may be as simple as knowing the source or much more complex as in knowing some signature phenotypic trait (colony or cellular morpholopgy, or some unique biochemical reaction for instance) or even more complex by a known DNA sequence difference as small as a single nucleotide difference (called a SNP or single nucleotide polymorphism).

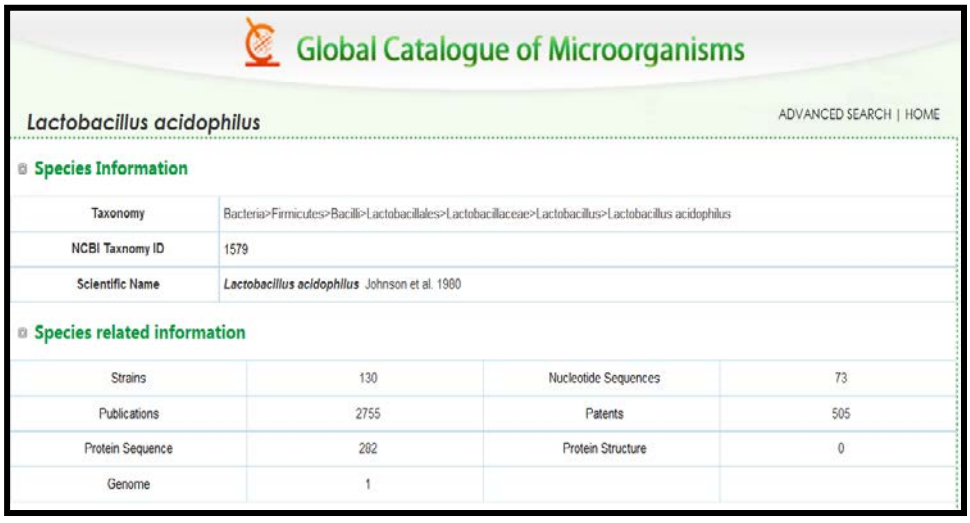


Figure 2. Taxonomic classification and ancillary information of *Lactobacillus acidophilus*. Note there are 130 strains deposited into culture collections and all are classified as *L. acidophilus*. Source: (http://gcm.wfcc.info/speciesPage.jsp?strain_name=Lactobacillus%20acidophilus).

Is it possible that several probiotic product producers are using the same strain of a particular probiotic? Yes, and in fact it is likely that strains available from just a few probiotic manufacturers are used throughout the probiotic industry under different labels and

commercialized names. When asking for analysis, before using the term "strain" ask yourself if you understand exactly what it is you need. If it is a species name, then ask for species. If indeed you need to know if it is the exact same strain that was purchased from a manufacturer, then inquire to the manufacturer whether they already have an assay (such as a PCR-based test or a flow cytometry probe) specific for their strain or if they have had the strain sequenced and possess a whole genome sequence. If the former, an appropriate PCR assay can be used to detect to the strain level. If the later, the whole genome sequence data could be used to develop a PCR assay (which normally means contracting with a genomics lab for PCR primer and assay design) that can distinguish a strain-identifying marker(Huys et al., 2013) (Treven, 2015). Alternatively, the sequence data can be used to develop a flow cytometry probe for single strain compositions of probiotics(Bunthof & Abee, 2002) (Kramer et al., 2009).

What About Mixed Probiotic Products?

Let's move on to the really difficult determinations, i.e. analyzing probiotic mixtures of multiple species (and strains). Again, it is not impossible that selective/differential agar media can separate and quantify to some level of resolution; but what that level is, other than the common descriptive or functional group name (e.g. some type of lactic acid bacterium, or some type of bifidobacteria) is unknown. If we assume the concentrated product is labelled with a genus and species name, can we get to the level of species or strain using culture? Yes, but, not directly. This is not only a long and tedious task, but, a highly costly one. Assume we picked some representative number of colonies, say 5 to 10, from a selective/differential agar plate (that we chose because of the stated composition of the probiotic product), and streak for isolation and then proceed to have each colony individually identified by either a phenotypic or genomic method. For this relatively simple exercise we are already amassing a bill of around $500 to

$1,000, and that is for just a single selective/differential agar plate! If a probiotic mixture requires 2 to 4 different media, then we jump to a conservative estimate of $1,000 to $2,000 to get a name. Even then, all of this relies on a microorganism's ability to form a colony under the growth conditions used as well as the lab technician subjectively picking a representative sampling of the different visible colony types. Although there are some amazingly specific selective/differential media available, these rarely can reliably get to the species level within the complex taxonomy of the Lactobacillaceae or Bifidobacteriaceae families. What culture can clearly do better than almost any other method is to enumerate colony forming units when the identity of the bacterium in a single-organism product is known and it is cultivable. Indeed, the standard recommended by the IPA (International Probiotics Association) is the colony forming unit or CFU. To complicate matters further, probiotic marketing managers engage in "more is better" product differentiation arms race. Where once you could market a probiotic with a single strain, it is now fashionable to add multiple strains to a product mix and to upsell products based on probiotic dose. Products with multiple probiotic strains and larger declared probiotic populations command a greater price than products with fewer strains and smaller populations. We dare you to try and verify such claims. It is easy to throw up one's hands and ask "this is all a big mess, what can we do?".

Fear not, take a deep breath, sit back and relax, because this is where you must agree that it is a great time to be a microbiologist. Now we are equipped with not only good cultural tests, but with the tools of genomics, specifically using the microbiome technique, we can determine the makeup of a mixture of probiotics without the need to culture, hope things grow, pick, and then identify. Recall from the first article in this series that a microbiome is the identification of a community of bacteria in a single sample (see:

https://foodsafetytech.com/column/microbiomes-move-standard-plate-count-one-step-forward/

and/or https://foodsafetytech.com/column/metagenomes-and-their-utility/) . We can perform a DNA extraction on the mixture then using a 16S Targeted Bacterial Microbiome resolve the bacterial mixture into a genus-level profile, or if the 16S sequence reads are of sufficient length, we can identify to the species level as well as determine the relative proportions of each taxon. If the probiotic mixture contains both bacteria and yeasts or other fungi we can perform two different targeted Microbiomes (bacterial and fungal) or rely on a metagenome-based microbiome to determine sequences and identify all microorganisms with signature DNA in that sample. If relative proportions of the constituents are not needed, we have an option to use a microbial microarray, which will give species names and in some cases even a strain or subtype. There are pros and cons to each method, but cost and very often time is saved with these approaches.

Finally we would like to address issues of genomic methods to distinguish viability. As stated in a previous column, PCR-based methods detect DNA (or RNA) sequences and the quantity or copy numbers of sequences in a sample. Recall that PCR is similar in function to a copy machine in that it spits out billions of copies of a single gene sequence, with a final copy number that is proportional to the starting number. However, DNA in the sample could have been derived from dead (non-viable) cells. Because PCR counts genes, such dead cell gene DNA can be detected resulting in a false identification. Is this a deal-breaker for using genomic methods for microbial analysis of probiotics? No not at all. There are assays that can use viability dyes (e.g. propidium monoazide) to distinguish live from dead bacteria(Fittipaldi, Nocker, & Codony, 2012). These dye assays can function well in both a PCR format or flow cytometry. Flow cytometry is currently receiving widespread attention because of assay speed

and apparent ability to distinguish live, dead and sub-lethally injured cells (Bunthof & Abee, 2002) (Kramer et al., 2009) (ISAPP, 2016).

Knowledge is Power

Knowledge of your probiotic strain(s) is critical for quality assurance and regulatory compliance. Perhaps the ultimate in genomic characterization is to determine the whole genome sequence (WGS) of your organism(s)(Li, Gu, & Zhou, 2016) (Khatri, Tomar, Ganesan, Prasad, & Subramanian, 2017) (Natarajan & Madasamy, 2014). This sequence provides many potential uses; [a] sequence sites for designing strain-specific PCR or flow cytometry primers and probes, [b] product mix identifications for regulatory compliance, [c] assurance of strain ID for customers, [d] presence of single nucleotide polymorphisms for intellectual property protection,[e] quality control assays for maintaining stock and production strain purity, [f] assurance that strains are non-pathogenic subtypes of species that might have pathogenic strains within the same species, and [g] assessing the presence of functional genes that confer beneficial (digestion aids, gut survival) or detrimental (toxins or virulence factors, antibiotic resistance) characteristics to a strain.

It must be stated that this science is still very much in development, and which methods might become more standard across the industry is still a wide open and rapidly evolving question. For example, there are issues with the accuracy and strain-resolution capability of these methods. One group reported that probiotic compositions often do not meet the stated label claims (Morovic, Hibberd, Zabel, Barrangou, & Stahl, 2016). These authors pointed out that of 52 different probiotic products tested, one-third were below label claim for cell counts (CFU) before expiration. They further demonstrated by using a multiplexed-PCR assay (a PCR assay that detects more than a single target in a single assay) that 58% of the sample set had incorrect

labels, including: [a] the name of the organism due to wrong taxonomy, [b] undeclared species were found that were not on the label, or [c] species on the label were missing from the product. Indeed, the science of subtyping, which is not an easy task in part due to the very complex family structures of many microbial groups, as well as their less-than-hardy constitution, needs much improvement. For example, one report demonstrated that only a single example out of sixteen *Bifidobacterium* spp. probiotic preparations correctly matched the label claims for viability and species or subspecies type (Lewis et al., 2016).

The combination of culturing with some form of genomic analysis is a powerful tool that probiotic manufacturers should exploit and we hope this article has helped by sharing some basic information. We have covered a smattering of genomic methods, these are by no means the only ones, but we have selected tools that are currently gaining favor for their precision and accuracy over other methods. We also see that genomic tools are not only useful for probiotic producers and users, but, are becoming the standard for probiotic analysis by regulators. Regulatory science and agencies currently rely upon the level of detail provided by genomic methods to help support their activities.

What is in the near future of this field? Look for suggestions that other genes besides the 16S gene will be used to discern and resolve complicated taxonomic groups. Look for multiplexed PCR assays coupled with viability indicators to become in more widespread use. Finally, look for rapid whole genome sequencing tools for developing probes specific for not only species but your own specific strains. Such tools will become significantly less costly and more rapid than current reality. The future in this arena is quite bright.

In this installment of Food Genomics we explored the basis and some uses of genomic tools for probiotic analysis. We hope the reader is now more prepared to ask the right questions

and understand which tools are available for which job. As always, please email us with your comments and questions.

References

Bunthof, C. J., & Abee, T. (2002). Development of a Flow Cytometric Method To Analyze Subpopulations of Bacteria in Probiotic Products and Dairy Starters. *Applied and Environmental Microbiology*, *68*(6), 2934–2942. https://doi.org/10.1128/AEM.68.6.2934-2942.2002

Daniels, S. (n.d.). IPA petitions FDA on requiring CFUs for probiotic labeling. Retrieved November 15, 2017, from https://www.foodnavigator-usa.com/Article/2016/11/22/IPA-petitions-FDA-on-requiring-CFUs-for-probiotic-labeling

Davis, C. (2014). Enumeration of probiotic strains: Review of culture-dependent and alternative techniques to quantify viable bacteria. *Journal of Microbiological Methods*, *103*(Supplement C), 9–17. https://doi.org/10.1016/j.mimet.2014.04.012

Fittipaldi, M., Nocker, A., & Codony, F. (2012). Progress in understanding preferential detection of live cells using viability dyes in combination with DNA amplification. *Journal of Microbiological Methods*, *91*(2), 276–289. https://doi.org/10.1016/j.mimet.2012.08.007

Global Probiotics Market to Reach $24 Billion by 2017. (2013, December 16). Retrieved from https://www.naturalproductsinsider.com/news/2013/12/global-probiotics-market-to-reach-24-billion-by-2.aspx

Guidelines for Probiotics in food. (n.d.). Retrieved November 15, 2017, from http://internationalprobiotics.org/resources/guidelines/

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., … Banfield, J. F. (2016). A new view of the tree of life. *Nature Microbiology*, *1*(5), 16048. https://doi.org/10.1038/nmicrobiol.2016.48

Huys, G., Botteldoorn, N., Delvigne, F., Vuyst, L. D., Heyndrickx, M., Pot, B., … Daube, G. (2013). Microbial characterization of probiotics–Advisory report of the Working Group "8651 Probiotics" of the Belgian Superior Health Council (SHC). *Molecular Nutrition & Food Research*, *57*(8), 1479–1504. https://doi.org/10.1002/mnfr.201300065

ISAPP. (2016, January 4). Flow Cytometry for Probiotic Enumeration | ISAPP. Retrieved from https://isappscience.org/flow-cytometry-for-probiotic-enumeration/

Khatri, I., Tomar, R., Ganesan, K., Prasad, G. S., & Subramanian, S. (2017). Complete genome sequence and comparative genomics of the probiotic yeast Saccharomyces boulardii. *Scientific Reports*, *7*(1), 371. https://doi.org/10.1038/s41598-017-00414-2

Kramer, M., Obermajer, N., Bogovic Matijasić, B., Rogelj, I., & Kmetec, V. (2009). Quantification of live and dead probiotic bacteria in lyophilised product by real-time PCR and by flow cytometry. *Applied Microbiology and Biotechnology*, *84*(6), 1137–1147. https://doi.org/10.1007/s00253-009-2068-7

Lewis, Z. T., Shani, G., Masarweh, C. F., Popovic, M., Frese, S. A., Sela, D. A., … Mills, D. A. (2016). Validating bifidobacterial species and subspecies identity in commercial probiotic products. *Pediatric Research*, *79*(3), 445–452. https://doi.org/10.1038/pr.2015.244

Li, P., Gu, Q., & Zhou, Q. (2016). Complete genome sequence of Lactobacillus plantarum LZ206, a potential probiotic strain with antimicrobial activity against food-borne pathogenic microorganisms. *Journal of Biotechnology*, *238*, 52–55. https://doi.org/10.1016/j.jbiotec.2016.09.012

LPSN - List of Prokaryotic names with Standing in .html. (n.d.). Retrieved from http://www.bacterio.net/

Morovic, W., Hibberd, A. A., Zabel, B., Barrangou, R., & Stahl, B. (2016). Genotyping by PCR and High-Throughput Sequencing of Commercial Probiotic Products Reveals Composition Biases. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.01747

Natarajan, P., & Madasamy, P. (2014). First Complete Genome Sequence of a Probiotic Enterococcus faecium Strain T-110 and Its Comparative Genome Analysis with Pathogenic and Non-Pathogenic Enterococcus faecium Genomes. *Journal of Genetics and Genomics*, *42*. https://doi.org/10.1016/j.jgg.2014.07.002

Parte, A.C. (n.d.). LPSN - List of Prokaryotic names with Standing in Nomenclature. Retrieved from http://www.bacterio.net/index.html

Probiotics in Animal Feed Market Worth 5.07 Billion USD by 2022 | Markets Insider. Retrieved November 15, 2017, from http://markets.businessinsider.com/news/stocks/Probiotics-in-Animal-Feed-Market-Worth-5-07-Billion-USD-by-2022-1002272597

Sattler, V. A., Mohnl, M., & Klose, V. (2014). Development of a strain-specific real-time PCR assay for enumeration of a probiotic Lactobacillus reuteri in chicken feed and intestine. *PloS One*, *9*(2), e90208. https://doi.org/10.1371/journal.pone.0090208

Treven, P. (2015). Strategies to develop strain-specific PCR based assays for probiotics. *Beneficial Microbes*, *6*(6), 887–898. https://doi.org/10.3920/BM2015.0009

Resources

More references can be obtained by emailing the authors or editors.

Authors

Gregory Siragusa, Senior Principal Scientist, Eurofins Microbiology Laboratories, gregorysiragusa@eurofinsus.com.

Douglas Marshall, Chief Scientific Officer, Eurofins Microbiology Laboratories,

douglasmarshall@ eurofinsus.com.