File Name: 0914181444_091318-808482-USDA-FSIS-Transition

Transition to Whole Genome Sequencing

September 13, 2018

[START OF TRANSCRIPT]

AT&T Operator, Aasif: Welcome everyone and thank you for joining today's Transition to Whole Genome Sequencing webinar. Please note that all participant lines will be muted until the Q&A portion of the call, and we will provide you with instructions on how to ask a verbal question at that time. However, you are also welcome to submit written questions throughout this presentation. These also will be addressed in the Q&A. To submit a written question, please use the chat panel on the lower right-hand side of your screen, and then choose all panelists from the sent to drop down menu. If at any time you require technical assistance, please send a note to the event producer or call our help desk at 888-796-6118, and with that, I'd like to turn the call over to Buck McKay, Public Affairs Specialist.

Buck McKay: Good afternoon. Thank you all for dialing into today's webinar on the transition to Whole Genome Sequencing. The purpose of this webinar is to further discuss in part the whole genome sequencing, MLST approach, and codes as well as related Whole Genome Sequencing topics. Before we start the webinar, I want to let you know that this webinar will be recorded. We will provide the transcript and the presentation on the FSIS website in the following weeks.

During the webinar, we will take a pause after each slide to address questions pertaining to the slide currently being shown. At the end of the presentation, we will open the floor up to address any other questions you may have, alternating between text and voice questions. It's my pleasure to introduce Dr. Meryl Silverman, Senior Food Technologist Risk, Innovations and Management Staff from our Office of Policy and Program Development, and Dr. Glenn Tillman, Branch Chief of Eastern Laboratory Microbiology and Characterization Branch.

Glenn Tillman: Thank you for that introduction. Today, we're going to be giving you an overview of Whole Genome Sequencing, or WGS, from here on out, and how it compares to Pulsed-Field Gel Electrophoresis, or PFGE. We're going to give you some background on transition from PFGE to WGS for *Listeria monocytogenes* (*Lm*) analyses, as well as the current status of WGS characterization of *Lm* isolates, and how we came to reporting Whole Genome Sequencing results in the quarterly establishment letters. Finally, we'll give you an update on our future plans of transitioning WGS for other pathogens. The

content we are covering today may be familiar to you, but we'll have to update information that we're going to share with you. Also, I would like to mention that if there is a question you need to ask, go ahead and do that before we go to the next slide as Buck mentioned.

So, what is PFGE? PFGE is a laboratory technique used to characterize bacterial pathogens. Enzymes are used to cut DNA (Deoxyribonucleic acid); an electrical field periodically changes direction or pulsed field to separate those DNA fragments on gel matrix. At the end, a photo is taken of the pattern as DNA fragments make when separated.

On this slide, we have a conceptual framework for how the process works. Again, enzymes are used to cut the fragments and that results in DNA fragments of different sizes, then the electrical field separates the DNA fragments on a gel matrix. As you can see here, a photo was taken of the fragments. One limitation of this method is that bands of the same size from two different isolates may not come from the same part of the chromosome. We've also needed to use other tools to provide additional information about pathogens.

This leads us to Whole Genome Sequencing, or WGS, which is a DNA sequencing technology that can be used to help characterize bacterial pathogens. Advances in DNA sequencing technology make WGS a powerful tool to characterize isolates for greater detail and accuracy than PFGE. Whole Genome Sequencing reveals the DNA, the complete DNA makeup of an organism, enabling us to better understand bacteria with greater precision that other technologies do not allow.

To compare these two methods, PFGE only gives information at a cut site based on the banding pattern that depends on the weight of the different pieces. Again, it doesn't tell us whether bands of the same size or two different isolates may come from the same part of the chromosome, whereas, WGS has the ability to give us information at nearly every position in the genome depending on the type of tool.

WGS has a number of uses that benefit FSIS, and its mission to protect public health. These uses include identifying harborage of pathogen in FSIS regulated facilities, tracing human illness outbreak data to regulated food products faster resulting in the avoidance of additional illnesses, then finally the determination of unique genes including virulent and antimicrobial resistance genes. WGS analyses have the potential of streamlining multiple laboratory

methodologies into one analytical stream while providing even more information about a bacterial strain.

FSIS began using PFGE to characterize bacterial isolates in the mid-1990s. FSIS began performing WGS in parallel with PFGE, for *Listeria monocytogenes* starting in fiscal year 2013 and for all pathogens starting at fiscal year 2016. The Center for Disease Control and Prevention, or CDC PulseNet, partners are transitioning away from using PFGE as the primary molecular characterization tool toward using WGS.

CDC PulseNet is a national laboratory headquartered at the CDC consisting of public health and food regulatory laboratories that contribute isolate characterization information for sharing across public health and regulatory partners. In coordination with CDC PulseNet, FSIS suspended PFGE for *Listeria monocytogenes* as of January 15, 2018, and now generates *Lm* characterization through WGS only. It was really that transition by PulseNet that led us to make the change and need to update how we communicate *Lm* characterization results. Although PFGE analyses are no longer needed for *Lm*, the laboratories maintain this capability.

FSIS uses different tools to analyze WGS information, including Multi-locus Sequence Typing, or MLST, a tool that for FSIS results in both the public sequence type and an allele code and high-quality single nucleotide polymorphisms, or hqSNP. FSIS finds value in looking at the results of both MLST and hqSNP tools together, so that each method provides a unique level of specific information. MLST provides enhanced qualitative detail, while hqSNP allows us to quantify the differences between genomes. I'm going to talk more about these tools in the next few slides.

First of all, I'm going to talk about MLST. MLST can generate a pattern name or designation similar to that of a PFGE pattern name based on differences in a predefined set of genes. Again, FSIS analysis results in both the public sequence type and an allele code. The public sequence type which sometimes you'll see referred to as MLST, ST, or pubST compares differences in a small number of genes, for *Lm* that’s seven genes, and generates a sequence type using the publicly available database developed by Jolley & Maiden. An example of a public sequence type is public ST09. The allele code, on the other hand, compares nearly 1,800 genes and generates a numerical code. These codes are generated by the CDC PulseNet. An example of an allele code is LMO 1.0-5.1.1.2.5.1

Because the allele code compares many more genes and that of a sequence type, again over 1,800 genes versus seven for *Listeria*, the

sequence type. The allele code is much more specific than the public sequence type. As you can see depicted in this example, one public sequence type can include many different allele codes. Because of the specificity, FSIS is focusing its communication, for example, in this quarterly establishment letter around the allele code. The allele code is useful for communication, because it gives the name or designation for an isolate in a simpler way to a PFGE pattern name, so it gives us a way to talk about an isolate.

As I mentioned earlier, the allele code generates a numerical code such as the one shown on this slide. We wanted to break down the different parts of this code so to help us understand better how we compare multiple allele codes. The first part of this code, the LMO tells you which pathogen this code represents. In this case, LMO stands for *Listeria monocytogenes*. The next part, the 1.0 is the version of the naming scheme or system. Like a software update, this naming system could be changed in the future, so FSIS' plan, are to report or a result reporting to include version numbers from the start. Each number or set of numbers after the dash in between the period is called a field.

By comparing fields, numbers after the dash from the code, you can actually tell how closely related two isolates are to each other. You notice in the diagram, the number of the alleles that are allowed to be different get progressively smaller for each field from left to right. For the first field, the match to isolates can have up to 71 allelic differences. For the second field to match, that can be up to 51 allelic differences. If the first four fields of allele codes of two isolates match, FSIS considers this one measure that two isolates are closely related. It's possible for two isolates be closely related without being perfect matches. They could be at the 19 allelic differences out of the 1,800 genes compared by MLST for two isolates be considered closely related. Although the more numbers that match, the more closely related the isolates are.

In addition to MLST, FSIS uses another tool to analyze WGS information called high-quality single nucleotide polymorphisms or hqSNP. hqSNP gives a quantitative measure, specifically the number of nucleotide or SNP differences between two isolates. Unlike MLST that compares differences in a predefined set of genes, hqSNP compares differences between the genes as well as part of the genome outside of the genes. Really, we're talking about measuring SNPs, we're talking about counting the number of nucleotide differences between two or more genome sequences to measure genetic differences.

These tools, multi-locus sequence typing, and more specifically, public sequence type in allele code as well as hqSNP, provide greater and greater resolution as we move from one tool to the next. Now, I'm going to turn the presentation over Dr. Meryl Silverman who will continue explaining how FSIS uses and communicates WGS information.

Meryl Silverman: Thanks, and before I continue, I just want to take another second to see if there are any questions up until this point?

Buck McKay: This is Buck McKay. Aasif (AT&T Operator), can you go over with our participants again, the process of asking a question?

AT&T Operator, Aasif: Absolutely. Ladies and gentlemen, at this time, if you'd like to ask a verbal question over the telephone line, please press pound two on your telephone keypad. You will hear a notification when your line is unmuted, and at that time, please then state your name and question. You are also welcome to submit a written question. To do so, please use the chat panel on the lower right-hand side of your screen, and then submit your question to all panelists and we'll just give it a moment.

Meryl Silverman: Okay, so we'll continue on. Now that Glenn has given an overview of WGS and the types of tools FSIS uses to analyze WGS information, I'm going to explain how we use these results to determine isolates are closely related for the purposes of determining potential harborage, reintroduction, or cross contamination. I'm also going to explain how FSIS communicates WGS information. When FSIS laboratories determine whether two isolates are closely related, well with PFGE would have been called indistinguishable. They look at results from all of the different tools Glenn talked about.

They look at whether the public ST matches, whether the allele code matches by at least four fields after the dash, as we just discussed, and then also that there are small number of SNP differences by that high-quality SNP analysis. In addition to that, there also needs to be an epidemiological or contextual link to connect one or more isolate. This could be what we traditionally think of as an epidemiological link, such as traceback information or could be compliance findings, such as that the samples were collected from within the same establishment.

To put this together in an example that we have on the slide, we have two isolates here. They both share the same public ST type of ST09. In this case, all six of the fields in the allele code match, but again, we would be looking for at least the first four fields after the dash to match. There are only two SNP differences between them

which is considered to be small. Since both of the isolates are collected from the same establishment, we have an epidemiological or compliance findings link, and since the isolates were collected two years apart, we would consider this evidence of potential harborage or reintroduction.

As you'll hear me talk about later when we get to the quarterly establishment information letters, if isolates A and B have been collected during the same sampling event, they would have been considered consistent with cross contamination.

FSIS shares comprehensive WGS analyses, including the MLST and hqSNP results we just talked about for *Lm* isolates with FSIS personnel, upon request. We also provide establishments summaries of their data on a quarterly basis through the quarterly establishment information letters. FSIS began nationwide implementation of quarterly establishment information letters about a year ago in August 2017. These letters include all sampling results for an establishment within the 12-month window of the letters, as well as industry averages. FSIS recommends that an establishment consider the information provided in the quarterly establishment letters to evaluate the effectiveness of their overall HACCP (Hazard Analysis and Critical Control Points) system processes and take preventive actions were necessary.

With a transition to WGS, FSIS plans to include the MLST designation and particularly the allele code in place of the PFGE pattern name for *Lm* results in the next quarterly letters, set to issue by September 30. These next letters will contain a summary of an establishment results for the third quarter. FSIS plans to include the MSLT designation for all isolates reported in the 12-month window in the letters. This includes isolates where FSIS may have reported the PFGE pattern name in the previous letter. You will also see on the next slide that while the results include individual results in the 12-month window, for purposes of determining harborage, results are compared against available data in a five-year period.

On this slide, you can see how the *Lm* results will be displayed in the quarterly establishment letters. Each positive sample result will be displayed in a row, and again results for all *Lm* positive in the 12-month window will be reported in the letters. For *Lm*, you can see the third column from the right is where the MLST designation, or allele code, is reported out.

Next, to the MLST designation in that same column, the number of times that particular MLST designation has recurred at the establishment in the last five years will be recorded in the

parentheses. Recurrences counted went up to the first four fields after the dash and the allele code match. One thing to note is that the number in the parentheses does not count or include the first occurrence of an MLST designation. The one in the parentheses in this example next to the circled text means the matching MLST designation has occurred two times at the establishment. This is the same way recurrence was determined and reported for PFGE.

In the next column, next to the MLST designation, a yes or no is reported depending on whether results are consistent with harborage or not. For purposes of the quarterly letter, results are reported to be consistent with harborage or repeated introduction when two or more isolates found in product, food contact, or non-food contact environmental samples that were collected over multiple days, weeks, months, or years have at least the first four fields of the MLST allele code designation match.

In the example on the slide, the product isolate from September 4, 2017 in the second row of the table, has an MLST designation of LMO 1.0-1.2.3.4.6.7 (editor correction from the recording), and the food contact isolate from May 10, 2018, in the first row had an MLST designation of LMO 1.0-1.2.3.4.5.6 (editor correction from the recording). These are simulated codes, but just to show you that potential harborage is indicated by a yes in this example, because the first four fields after the dash, that is the 1.2.3. and 4 in the red circle match, and the isolate were from samples collected over multiple years.

One thing to note is that the first or oldest sample will have a no for harborage, as you can see here. That's because at the time, there was only one positive sample result, so there was no isolate to compare to. Subsequent isolates will be compared, and if there is an MLST match again, based on the first four fields after the dash, a yes will be indicated.

As I explained earlier, FSIS laboratories perform a comprehensive analysis using hqSNP, in addition to MLST, to just further determine whether results are consistent with harborage, repeated introduction, or cross contamination. I do want to note though that FSIS does not collect samples for purposes of conclusively determining whether there's harborage or repeated introduction. If the establishment produces adulterated product under the HACCP regulations, the establishment is required to identify the underlying cause as part of corrective actions.

This slide has the same table, but here we've changed the example to show you when cross contamination would be indicated. For

purposes of the quarterly letter, results are reported to be consistent with cross-contamination when two or more isolates found in product, food contact, and environmental non-food contact samples where they were collected during the same sampling event have at least the first four fields of the MLST allele code designation match.

In this example, a product isolate from June 10, 2018, in the first row had an MLST designation of LMO 1.0-1.2.3.4.5.6, and the food contact isolate from the same day, June 10, 2018, had an MLST designation of LMO 1.0-1.2.3.4.6.7. Potential cross-contamination is indicated in the letter by a yes, because the first four fields after the dash again, that's 1.2.3.4 in the red circle, match. And the isolates were from samples collected during the same sampling event. Again, FSIS laboratories perform that more comprehensive analysis using high-quality SNP in addition to MLST to further determine whether results are consistent with cross contamination.

That is an overview of the WGS analyses FSIS performs for *Lm* characterization, and how WGS information will be shared in the quarterly establishment letters. Although FSIS has transitioned away from PFGE to using only WGS for characterizing *Lm*, PFGE analysis in addition to WGS will continue to be performed for other pathogens, because CDC PulseNet continues this PFGE for these other pathogens. We do anticipate that in FY2019 and continue coordination with CDC PulseNet, that the agency will replace PFGE with WGS as the primary subtyping tool for *Campylobacter*, Shiga toxin-producing *Escherichia coli* (or STEC) and *Salmonella*.

FSIS will inform stakeholders when it transitions to WGS for other pathogens. Until then, we'll continue to report the PFGE results for those other pathogens in the quarterly establishment information letters and share PFGE results.

We're also working on how to share accession numbers with establishments through the PHIS industry access, so that establishments can see the WGS results for other pathogens in the database that is hosted by the National Institutes of Health (NIH) through the National Center for Biotechnology Information, or NCBI. We'll share more information once that's ready.

At this time, we have additional time to take questions. If after the webinar you have further questions, you can always submit a question to askFSIS, selecting “sampling” from the product field, so that they go to the appropriate staff.

AT&T Operator, Aasif: Ladies and gentlemen, at this time, if you'd like to ask a verbal question, please press pound two on your telephone keypad. If you'd like to ask a written question, please use the chat panel on the lower right-hand side of your screen and submit your question to all panelists. At this time, we have not received any questions.

Buck McKay: Thank you, Dr. Tillman and Dr. Silverman.

Meryl Silverman: Yes. We do have a question. This is Meryl Silverman. Are these relatedness criteria specific to *Lm*, like the first four fields in the MLST allele code, and less than or equal to two SNP? And could these change with other species?

Glenn Tillman: Meryl, would you like to take this--

Meryl Silverman: Yes, we'll direct that to Glenn Tillman.

Glenn Tillman: Okay. Thank you. Thank you for the question. Yes, to answer that shortly, yes. These criteria are specific for *Listeria monocytogenes* this time. We do anticipate those values will change as we move into other organisms, specifically *Salmonella* and *Campylobacte*r. Within each serovar, we'll probably see a lot lower SNP differences or allele differences in between [isolates] before determining relatedness. So those criteria will be coming out. When those are released as well, we'll have a criteria for those.

Meryl Silverman: Glenn, could you just clarify that the two SNPs were just from our example. That's not the cut off that used for relatedness.

Glenn Tillman: Sure. Yes, relatedness. The two SNPs was an example Meryl noted in that particular example that was given. Typically, it's that four-digit represents approximately 19 alleles apart.

Meryl Silverman: Okay, and this is Meryl Silverman. We have an additional question. Will the recording of this webinar be posted to the website, if so, when?

Buck McKay: This is Buck McKay. Yes, this presentation will be posted to FSIS' website, and as far as when it will be posted in the weeks to come. Once it is posted and released, it will come up in the *Constituent Update*. Are there any other questions?

AT&T Operator, Aasif: At this time, we have no other questions.

Buck McKay: Okay. In that case, we will conclude this webinar. I'd like to thank Dr. Tillman and Dr. Silverman for presenting and thank all the attendees for calling or logging in. Have a wonderful day.

Glenn Tillman: Thank you.

AT&T Operator, Aasif: Thank you all for joining today's conference. The conference has now concluded, and you may now disconnect.

[END OF TRANSCRIPT]