APEC WGS Webinar Transcript – All Sector Video

Welcome and thank you so much for joining us. My name is Kelly McCormick and I'm an international policy analyst with the U.S. Food and Drug Administration Center for Food Safety and Applied Nutrition. This webinar was designed to introduce a project which will be conducted over the next several years under the auspices of the Asia Pacific Economic Cooperation's Food Safety Cooperation Forum. The project will explore whole genome sequencing and focus on building laboratory capacity for environmental testing of foodborne pathogens in the APEC region. Our first step toward implementing this project is to attempt to peak interest in the subject matter through a series of sector-specific webinars. This recording which includes the presentations from all three events will provide a basic background on whole genome sequencing by Dr. Eric Stevens also from the USFDA. Afterwards I'll go on to give some background on APEC itself and to introduce the multi-year WGS work stream. And the last part of our webinar consists of perspectives provided by those sector-specific experts. Regulators including Dr. Mark Allard of the USFDA, Dr. Crystal Southern of the US Department of Agriculture's Food Safety and Inspection Service and Dr. Kathy Carrillo of the Canadian Food Inspection Agency. Academics including Dr. David Erickson from the Joint Institute for Food Safety and Applied Nutrition, Dr. Edward Dudley from Pennsylvania State University and Dr. James Herrick from James Madison University. As well as industry experts including Kelly Hoon from Illumina and doctors Ramen Kaksar and Adam Allred from Clear Labs. They'll all explain why they and their organizations have found WGS to be advantageous. They'll share how they employ the technology and they'll provide insights into how they intend to utilize WGS technologies in the future. The full recordings of the three webinars along with this merged version including the presentations for all three sectors will be available on the SPS courses platform. With the account that you created to register for these webinars you'll be able to go in and access those and other resources that we compile at any time and with that we'll get started now with Dr. Eric Stevens.

Hello everybody, my name is Eric Stevens. I am an international policy analyst with the United States Food and Drug Administration and I'm here to talk to you today about whole genome sequencing, providing overview of the technology as it's applied to food safety, especially in the context of laboratory capacity building with an emphasis on environmental testing for foodborne pathogens. This presentation is intended to be a very high level overview for how whole genome sequencing works, and to start in its most basic form, sequencing simply means to determine the number and order of nucleotides that make up a given fragment of DNA. This DNA fragment can be very short, a few nucleotides, or it can encompass millions of nucleotides and make up a complete genome. For our purposes sequencing DNA means to discover the order of the four nucleotides and they are adenine, thymine, guanine and cytosine, otherwise known as a, t, g and c.

This slide provides a good visual example of the various steps for whole genome sequencing because it's important to understand that using current sequencing technologies, we're not yet able to sequence a complete genome of an organism in a single run or a read, meaning a single string full of a's, g's, t's and c's so we have to do is take the genome that we would like to sequence, in our case it's a bacterial genome from a bacterial pathogen which is three to five million base pairs in length in a single circular genome. We take that organism and we grow it up in culture so we get many, many copies and then we extract the multiple copies of the genome and we cut them up into much shorter fragments of DNA. These fragments can range anywhere from 200 nucleotides in length to a few thousand. These shorter fragments are going to then be sequenced using whatever sequencing method that you choose from that you are choosing and then we input those sequenced DNA fragments into a computer algorithm that's going to then reassemble the hundreds of thousands to millions of sequences to reconstruct the genome sequence of the organism that we're trying to sequence. It's also important to be aware of that when we say whole genome sequencing, we're typically aiming to get 98 to 99 of the genome reconstructed because there's always going to be small portions of the genome that are difficult to reconstruct so 98 to 99 percent of a complete genome sequence available is pretty good and this is the general framework that we'll be using going forward.

One of the first classical uses of whole genome sequencing for foodborne safety, it was in the context of responding to foodborne outbreaks and using the sequenced genome information to help infer bacterial relatedness especially between isolates that came from clinical samples with isolates that came from food or environmental samples.

To try and answer that question of whether the genomes match, we use something that's called a phylogenetic tree as shown in this figure, and a phylogenetic tree is a visual representation of an evolutionary and statistical method to say, based on sequence data which isolates are more closely related, genetically speaking. The closer to isolates are the fewer nucleotide differences or snips, single nucleotide polymorphisms, you would expect to see. So in this example we see that the red clinical isolates are very closely related to these blue environmental isolates from flour, and we can kind of build the picture to say that it's likely that the contaminated product or flour is somehow responsible or linked to these clinical illnesses and it can better direct our public health response.

To be able to use whole genome sequencing in that manner it's really important to be able to have the sequence data from clinical samples and food and environmental samples together so that we can constantly and continually compare the genomes of newly sequenced isolates to see where they fall within daily or updating phylogenetic trees. One such mechanism for this is put together through FDA's genome tracker project and pulse nets, or CDC's PulseNet network, which aims for clinical samples, while FDA aims for food environmental samples so that our public health professionals can better use the whole genome sequencing data for its maximum benefit of linking the clinical isolates back to the food or environmental samples.

This slide expands upon how the network operates, whereas you have FDA's GenomeTrakr on one side and CDC's PulseNet on the other side, each of which work with different agencies within the U.S. as well as state and local health agencies, academic labs, agricultural labs, and even other international labs to help gather and obtain the food and environmental isolates that will feed through FDA's GenomeTrakr network, as well as some of the clinical isolates that will feed through CDC's PulseNet. Now both of these networks will then upload their sequence data to NCBI's Pathogen Detection website which is going to contain all the whole genome sequencing data that has been generated. It's important to note too that these data are also linked to submissions from other international counterparts such as Public Health England, Argentina, and every day NCBI's pathogen detection will put out a phylogenetic tree based on species for all of the data that they have, and this allows public health researchers a very easy quick and free method to use whole genome sequencing for its maximum benefit for public health and food safety.

Pathogen Detection Browser is free and publicly available so that anybody in the world at any time can go in and find any of the isolates that have been uploaded. So if you wanted to search for your favorite foodborne pathogen or you wanted to search by country or if you wanted to search by specific food product, all that information can be ascertained by using Pathogen Detection Browser. This is a partnership that the various U.S. public health agencies have with the National Institute of Health, Health and CBI.

In addition to the sequence data, it's also important that other type of information related to the sequence is uploaded as well and all this data is linked together. This other information aside from sequencing data is what we call metadata and it helps to put into context where that sequence information came from. So where was it collected, by whom was it collected, did it come from an environmental swab, so from a facility, did it come from a clinical patient, which state did it come from, which country did it come from, did it come from a specific food product. All that information can really help to put that genomic sequence into context and helps us tell a little bit more of the story of how that DNA sequence is related to others in the database.

I like this slide for a number of reasons. The most striking is that you can see how quickly we've gone from having very few genomes that are publicly available in this database at NCBI, which again is a combination of GenomeTrakr and PulseNet data. So in 2013 you can see we had very few sequences available and at the end of 2019 we had just passed a little over 400,000. Where we are today in 2020, we have well over 550,000 complete genomic sequences for foodborne pathogens that are spread across salmonella, listeria, e-coli, shigella, campylobacter, vibrio and a few other foodborne pathogens. And again all this data is publicly available and I have to say we've come such a far way since 1997 when we've sequenced e-coli for the first time at a cost of millions of dollars through an international collaboration of scientists, and so to be able to have over 550,000 foodborne pathogen genomes available to the public for use for food safety and public health I think is nothing short of tremendous.

It's important to talk about what the resource costs are when talking about how best to use whole genome sequencing for food safety. The majority of the cost associated with it is with the storage of the sequence data itself as well as the data analysis. Other costs, such as just doing the sequencing as well as the network administration or management, really don't compare to the expense of doing the data analysis or providing for the data analysis and storing the data. This is one of the reasons that FDA and CDC went with NCBI Pathogen Detection Browser because NCBI had already been storing and analyzing sequence data that came out of the Human Genome Project, and when talking about the costs of storing human genome sequences versus the cost of storing and analyzing bacterial sequences, there's about a thousand fold difference because the length of the human genome is so much greater than a bacterial genome. So really we were able to be cost effective by going with something that had already been established, thereby saving money to better sequence and better manage the GenomeTrakr and PulseNet networks.

Of course, whole genome sequencing has many roles in food safety aside from linking clinical isolates to food or environmental isolates and trying to help find potential sources of contamination, as well as in helping to define the scope of contamination and the number of illnesses. It's also able to be used to monitor the effectiveness of cleaning and sanitation within facilities, so if facilities find that they have a recurring problem despite cleaning and sanitization, they can make that determination because they see the same sequence, the same bacterial sequence, time and time again. It can also be used to provide a piece of the information used in regulatory action, as well as trying to understand what that root cause of the contamination was, so how did the contamination event occur.

When we talk about some of those maybe root cause analysis, one easy place to start thinking about how they can begin is with environmental surveys. So in this example we're seeing a farm on the eastern shore of Maryland and so you have the Atlantic ocean, the Chesapeake Bay and you can see that there are multiple opportunities or locations to sample and then sequence any positive isolates, and so you can try and paint that picture if you have contaminated produce you can try and match through a phylogenetic tree on the sequences to see if the contamination is related to a creek or a sediment, the streams, any type of specific patches. Maybe it came from seagull droppings or horse droppings. We can use whole genome sequencing as part of an environmental survey to build that picture for how the contamination is taking place and so if we are better able to understand how that contamination is occurring in the first place then we can be better able to prevent such a contamination from occurring.

Whole genome sequencing as I mentioned earlier, can be used to kind of monitor the effectiveness of cleaning and sanitation within a firm and it can also be used to identify that harborage or persistence. So whether you're seeing the same pathogen sequence over and over again, meaning that there's some niche in the facility that is continuing to contaminate despite cleaning or sanitization, and so it signals that there's a specific problem as opposed to you're seeing different sequences of pathogens come through and it's not the same one over and over again. This can also help to provide input or information on where the contamination is coming from if it's prior to a processing facility. So in this case you can use sequencing and compare the genomes of the sequenced isolates to see which of the fields match the final product and whether there's a problem within the processing facility to see if there's an increase in the contamination rate as well.

Whole genome sequencing as I've said is not only to be used by public health professionals to respond to outbreaks, but more and more it's likely that whole genome sequencing is probably going to have a bigger role in helping to prevent foodborne contamination from occurring and thus used as part of helping to inform better agricultural practices. So on a farm, seeing where sources of contamination are most likely happening in ways that we can reduce their impact, as well as looking at the environment within a processing or manufacturing plant, and certainly as you know, international trade and agriculture increases and more and more we are getting what we eat - the food and ingredients that we use - are coming from more and more other countries as international trade increases, whole genome sequencing can really help to spot issues kind of in real time and possibly even before they occur so that we're better able to address these things to not only improving the health of the population in a single country but on a global level for helping the worldwide population.

And aside from specific uses of whole genome sequencing that are either responding or preventing a known foodborne outbreak, whole genome sequencing is also being used as just good old fashioned, just research. So in this example and something that I really think you're going to be seeing a lot more of in the coming years, is using what's called kind of meta genomics or not only sequencing a single pathogen that is present in a sample, but sequencing all of the various types of bacterial or even fungal or viral sequences that are associated with a sample. So in one study, researchers at FDA looked at kind of the flora, the microbiological flora of tomatoes depending on whether they were from the east coast in Maryland, Virginia, and North Carolina versus tomatoes found on the west coast, and they were able to see that different species make up different proportions of the microbiome depending on whether the tomatoes are on the east coast or the west coast. And this could have implications in ways that we might not instantly or immediately recognize. For instance there could be something to do with a microbiological flora which could be if it's disrupted, it could allow more pathogen, pathogenic bacteria to proliferate so if we're able to better understand just the relationship of microflora and kind of what a healthy microflora looks like for tomato, we may be able to use other methods to help improve or reduce the contamination of foodborne bacteria.

One, another example and I believe that this is my final slide, is the use of whole genome sequencing to help monitor antimicrobial resistance, so all of those sequences that have been uploaded through to NCBI's Pathogen Detection Browser. I include the sequence information when looked at through data analysis that public health researchers could identify or find, antimicrobial resistance genes that are associated with those foodborne pathogens. This allows a kind of real time monitoring of the spread of antimicrobial resistance just in the food population and it can help us get a better understanding of that one health relationship which is really between the environment, animals, and human health, as a way to tackle this important issue. I realize that this was a very short presentation and I've only touched upon a few of the myriad uses of whole gem sequencing technology for food safety, some of which you're going to be hearing in subsequent presentations and I would be happy to answer any questions that you might have about any of the specifics and I look forward to answering any questions and hope that you found this enjoyable and that you find the coming presentations enjoyable and informative as well, so thank you very much and it's been a pleasure.

Good morning and good evening. My name is Kelly McCormick and I'm an international policy analyst with the U.S. Food and Drug Administration Center for Food Safety and Applied Nutrition. We recognize that many of you joining tonight or today are unfamiliar with this project and many of you are also unfamiliar with the Asia Pacific Economic Cooperation itself, as well as the work conducted under the auspices of the Food Safety Cooperation Forum's Partnership Training Institute Network. As such, I'll talk to you a bit about what APEC is, why there's a need for food safety work in the APEC region, give some background on the formation of the FSCF and the PTIN and mention some of the work that's been done so far. Then we'll dive a bit more deeply into the APEC WGS Project.

APEC was created as a regional economic forum in 1989 with the goal of supporting sustainable economic growth and prosperity in the region. APEC has an interest in championing free and open trade and investment, promoting and accelerating regional economic integration, and encouraging economic and technical cooperation. So overall APEC has a significant trade focus. APEC has a variety of sector-specific committees and subgroups and has engagement from the executive through the technical levels of government. It is important to note that even though APEC decisions are non-binding decisions and the impact of capacity building efforts often carry over into other international forum which are binding. So the forum has 21 members referred to as member economies. APEC is hosted by a different member economy every year. This year's host is Malaysia and next year is New Zealand. Obviously this has not been a typical year nor is next year expected to be, so many events which would generally have taken place in person on the margins of the formal meetings are being, and will be, implemented virtually.

As far as its reach goes APEC's 21-member economies account for a significant amount of the world's trade. Almost 40 of the world's population live in APEC economies and those economies account for 60 of the world's total GDP and almost 50 percent of the world's trade, so it's a really it's a pretty important region. Because of this large volume of trade and agricultural products, assuring food safety presents a very real and growing challenge in the APEC region and beyond. The overall level of food trade is rising globally and this includes increasing amounts of trade between developed and developing economies. So as you know, companies source food ingredients from markets across the world and export globally as well. This means that one incident can now spread quickly across multiple markets worldwide. So this presents challenges to assure the safety of food being imported from numerous economies and to manufacturers who acquire inputs and ingredients from diverse sources. Adding to this challenge is the fact that different economies have varying levels of adherence to international agreements and standards and in some cases there's a lack of understanding or available resources to meet those standards. In addition, consumer preferences and expectations are evolving and public opinion can also influence policies.

So recognizing the impact of those food safety challenges on the APEC region, a collective mandate was set by APEC leaders to improve practices and standards. This mandate addresses that food safety and internationally harmonized food standards are key factors for improving public health and safety, and facilitating trade and food for APEC economies. And so APEC's FSCF was established by the subcommittee on standards and conformance, that's the SCSC, in 2007. It's co-chaired by Australia and China and formally meets every other year. The FSCF was founded to improve and strengthen food safety information sharing and to identify, prioritize and coordinate capacity building efforts across the APEC region. It focuses on facilitating a food safety capacity building and regulatory dialogue to encourage and align use of science-based international standards to Codex standards and encourage adherence to WTO, SPS and TBT obligations. So to achieve the purpose of the FSCF, an outreach arm was created in 2008 named the Partnership Training Institute Networks (PTIN). It's administered by the United States and was specifically designed to address the need for public-private partnerships to maximize leveraging of expertise and resources. So the goal being to engage food industry scientific societies in academia and innovative ways to strengthen capacity and enhance and streamline technical assistance efforts in food safety.

And it meets every two years alongside the FSCF meeting. The goal of the FSCF PTI and work is that by building stronger food safety systems along with an improved understanding of regulatory compliance, there will ultimately be fewer food safety incidents and trade disputes. In 2010, stakeholders involved in the PTI ETIN collectively established a strategic plan articulating and identifying five priority areas of work. Those being risk analysis, supply chain management, food safety incident management, laboratory capacity, and regulatory systems. These priority areas were seen as having the biggest impact on food safety and trade.

PTIN food safety activities to date primarily fall within these five broad priority areas including work in risk analysis, export certification, aquaculture, MRLs, and food safety modernization, etc. So activity is designed to strengthen capacity in these areas have been endorsed by the PTIN Steering Group and have been implemented using a combination of APEC's secretariat controlled funds, private sector contributions, and individual economy government expenditures. So as mentioned, the FSCF PTIN works collaboratively with a number of private sector and academic institutions. So for example, partners like GIFSAN, Michigan State University, Abbott Waters Corporation, consumer brands, the SOAP association, 3M and many others have provided speakers, training materials, use of facilities and technical input.

Here you can see a list of some of the milestones achieved through PTIN efforts. So streamlining and reduction of export certs, harmonization of MRLs, etc. As far as lessons learned go, the PTIN provides many facilitation and improvement opportunities. It's a great venue for cross-sectoral collaboration. There's much crossover of APEC work to other fora. So for example, the export certification work in in CCFIX, MRL guidance being referenced in WTO SPS committee work, the work on good regulatory practice, and public comments shared at the WTO SPS transparency workshop. Those types of things. So additionally and, sometimes most importantly, it has an excellent multiplier effect. So conducting this work in the forum ensures that results and lessons learned get shared, project designs get replicated, technical documents generated get distributed, etc. throughout an entire diverse region regardless of whether the training efforts themselves only reach a single economy or a small amount of representatives from the area. And so now we come to the reason for convening this webinar. So earlier this year the U.S. proposed via concept note a multi-year project on whole genome sequencing, laboratory capacity building for environmental testing of foodborne pathogens. The effort was co-sponsored by Australia, Chile, Thailand, and the Philippines and was fully endorsed by the FSCF PTIN in late June. So first off, you know, why do we want to focus on WGS and environmental sampling in the PTIN? So essentially we see that the cost of WGS technology is decreasing, making it increasingly accessible and it offers new ways to determine sources of contamination and environmental sampling and subsequent WGS, you know, including analysis of product samples, can help clarify contamination pathways and help to inform better agricultural practices that reduce overall contamination. So to do this we proposed a work stream designed to introduce WGS technologies and the resulting data analysis and data sharing, in order to enhance and optimize pathogen detection techniques that can improve traceback and overall food safety management. So a primary component of the project will be its focus on strengthening capacity of APEC region laboratories' environmental sampling methods for foodborne pathogens. So we believe that the maximum impact on global health can be achieved, you know, through the increased understanding of the environment through sampling and monitoring through the application of WGS and the sharing and access to both the sequence data and its metadata, you know, in so far as it doesn't infringe on into intellectual property. So we feel that this project can improve public health conditions allowing APEC economies to respond more quickly to foodborne outbreaks and to understand the ways in which foodborne contamination occurs, resulting in decreased loss of life and productivity due to foodborne illness thus preserving economic livelihoods and contributing to food security.

So the project will introduce WGS data analysis methods and tiered approaches to environmental food safety testing, including the discussion of when they're appropriate to implement. It will encourage awareness of and partnership in global data sharing networks. It'll provide insight into appropriate and resource effective mitigation strategies, and it'll address records practices through the implementation of a series of activities.

The activities will focus on specific areas that need supplemental emphasis or groundwork accomplished to mitigate risk to APEC consumers while protecting domestic and regional public health and economic interest. We will draw from previously successful APEC FSCF PTIN project models, such as that of the lab competency strengthening work that was conducted roughly between 2010 and 2016. So there's 11 eligible APEC economies who will be supported by the project. Non-travel eligible APEC economies will also be made aware of project webinars and trainings, and we'll be encouraged to participate on a self-funded basis, space permitting. Initial funding for the project will be provided by the U.S. Food and Drug Administration, the U.S. Department of Agriculture's Foreign Agricultural Service, and the U.S. Agency for International Development.

So we're currently in phase one. Welcome. We're convening these webinars to promote awareness of the project principles to policymakers, researchers, instructors, and laboratory managers from the targeted APEC economies. These will hopefully help increase awareness of project objectives, as well as provide background information on methodology and technology to be exhibited during various phases of the project. These webinars will also be used to help participating APEC economies understand the ideal participant profile for phases two and three, which will help to ensure that the best suited participants are recruited for the workshops and trainings.

Phase two is a workshop for food safety experts and policy makers. This was to be held to formally kick off the project in New Zealand in May 2021 during the senior official meeting twos FSCF PTIN food safety suite of meetings. Realistically this will now be a virtual event but regardless, APEC economy representatives attending will consist of a mix of technical and policy staff. The workshop will introduce WGS sequencing technology and data analysis methodologies and the science behind them, and it'll discuss resource implications, cost, personnel requirements, setting the stage for the phase three sub-regional laboratory trainings. The event will provide an overview of current guidelines for implementing and using WGS, such as the GenomeTrakr network event. Content will also delve into the application of WGS for environmental sampling to better understand the natural variation of foodborne organisms in the environment. It'll look into preventive controls on farms, including resident transient strains and virulence. A primary focus of the workshop will stress the importance of incorporating new technologies intelligently and efficiently, and emphasis will be placed on identifying current projects, networks, and informational resources in and or available to APEC economies. So draft agenda topics are pasted in really tiny print on this slide courtesy of our FDA WGA superstar Dr. Eric Stevens. So phase three we have sub-regional laboratory trainings. APEC economy laboratory technicians will participate in one of three trainings. Trainings will consist of a classroom-style introduction to microbial criteria, tiered testing approaches, and WGS methodology, and the resulting data analysis necessary to interpret and act on WGS results. Participants will then move to in-laboratory training to employ the concepts discussed during the first day of the event. Commodities will be selected based on regional relevance. An innovative methodology will be implemented. The training will focus on, I'm sorry, will conclude with another classroom style series of sessions focused on mitigation strategies, data sharing, and records keeping. Training participants should be laboratory staff who are not only active technicians but also who oversee or have the opportunity and authority to influence additional laboratory employees. A mixture of two to three funding permitting individuals from government, academic, and private sector laboratories will be selected from each economy. And phase four we have our Teach-Forward by Training participants. So the participants from phase three events will pass on a variation of the training received to individuals in their economy. They'll be required to submit reports on the trainings that they provide and standout individuals will be asked to participate in phase five workshop at the APEC's senior officials meeting in May 2023.

And finally phase five will be a workshop held to culminate the project. Economy representatives will present case studies of trainings and Teach-Forward initiatives. They will provide analysis strategies employed. They'll share preliminary evaluation results, etc. APEC economy representatives will consist of again, a mix of technical and policy staff, and will use the forum to discuss the integration of innovative methods at the economy level as well as to develop ideas for future capacity building efforts related to the work stream. So that is it for your background on the APEC FSCF PTIN and the multi-year work stream on WGS and environmental testing for foodborne pathogens. So please do get in touch if you're interested in engaging. Next up will be a series of perspectives provided by the sector specific experts. Thanks so much. Hello, my name is Mark Allard. I'm a research microbiologist at the Center for Food Safety and Applied Nutrition at the U.S. Food and Drug Administration and today my talk is on applications of whole genome sequencing at the FDA's foods program. A typical outbreak investigation starts when, it begins when they see a cluster of illnesses then the epidemiology tries to identify the suspected food vehicle, and then with that a positive sample and/or traced back through investigations will link the illnesses to that particular vehicle and then the last step is that public health and regulatory actions may occur depending on what's found. Genomics is directly introduced and is a critical component in linking clinical samples to the food and environmental samples. FDA shares their genomic data through the NCBI direct submission for sharing WGS data globally. The clinical isolates coming from CDC and Department of Health labs and also with support from APHL go through PulseNet, a CDC program and the food from FDA comes through GenomeTrakr and from the USDA Food Inspection Services, the FSIS. All the data goes to NCBI in the U.S. but it might go to the European database or the database of Japan and then FDA also supports GalaxyTrakr, which is a place where validated data analysis pipelines are shared with all of our domestic and global partners.

The metadata that describes the genomic data is as important as the genomic data itself. This includes attributions on the isolates such as the species, the organism, and the strain identifier; the collection attributes, which is when and where the isolate was collected; and other source attribution like whether it's a human, a clinical, or a food or environmental isolate and then there can be additional identifiers of the isolate or, and the whole bio project will have its own place to reside it and see. In finding a source of an outbreak the first step which is, if contaminated food enters the food supply, people get may get sick very fast and we can identify the clusters of illnesses. What's slower is determining what food made them sick and then even slower still is where did the food originate and what was the root cause of the outbreak. Contamination for foods that have a relatively shorter shelf life like leafy greens, genomics has some weaknesses in that a contamination occurs, contaminated product goes into the food supply and we see a cluster of illnesses relatively quickly, but then the public health laboratory's reporting system kicks in on determining what those cases mean and then linking it to a common exposure, and then after EPI comes traced back which is actually going back to the food, and by the time FDA often learns about a contamination event, the all the food is already gone from the food supply. But for foods that have longer shelf life the genomics can intervene in a way that still makes a difference and gets it off the shelf. The primary roles for whole genome sequencing and investigations is that the genomic connections and linkages point to a potential source of contamination. Genomics also helps define the scope of the contaminant and illness, and also whether effective preventative controls and the effectiveness of cleaning and sanitization. And genomics provides a piece of the information used in the regulatory action and in determining the root cause, and hopefully then sorting out a preventative control. In our case study of contamination in enoki mushrooms I'm going to briefly describe the outbreak, and this is the domestic observations that there's 36 cases in 17 states spread out over a four-year period from 2016 to 2019, and this included two fetal losses and four deaths. But essentially over all this time they never determined this food source that was contaminated, although epidemiological signal was available for some kind of Asian style foods, including fresh produce. Along with the domestic observations there were also international linkages to this outbreak and that included both environmental and illnesses in both Canada and Australia which occurred over this three-year time frame. Collaborations and the genomic data sharing from some of our international partners helped FDA identify sampling of mushroom varieties and we sampled all imported mushrooms into the U.S. from foreign manufacturers. And in collecting those samples only enough enoki mushrooms sampled were positive for listeria and in fact, genomic evidence and product labeling and trace back confirmed that the 21 positive samples collected in four different countries between 2017 and 2019 were the product of one manufacturer in the Republic of Korea. An interesting sideline for this case study was the importance of uploading old food and environmental isolates into the database because they can be the trigger or the signal that helps you link clinical illness to a particular food or environmental sample. And it was some of these older isolates that helped in identifying the source or directing the investigation toward enoki mushrooms. When all of the genomic data is combined in a whole genome sequencing phylogeny, which you can find at the NCBI Pathogen Detection website, what you can see is there's, they're fully interconnected of a common contamination across multiple countries. The black arrows represent Canada and Australia clinical and environmental samples. The red highlight the South Korean environmental and food isolates, and then the blue are the U.S. clinical and environmental samples. And what you can see is they're interspersed of a common contaminant being shipped and contaminating cases in multiple international locations. In conclusion I'd like to say that whole genome sequencing is an important link in the food safety chain. The most heavily sequenced pathogens illnesses in the U.S. are decreasing faster relative to other pathogens, so genomics makes a difference. Outbreaks of heavily sequenced pathogens are getting smaller, potentially due to both detection of smaller clusters and faster response for more precise outbreak investigation. Even considering the additional cost, the program likely paid for itself in early implementation and will certainly do so as the program matures. Investments in genomics matter, they reduce illnesses. In addition to direct effects on public health, the program increases accountability as firms that are identified clean up. It also increases effectiveness and efficiency of compliance and enforcement and facilitates root cause, which assessed our analysis of risk assessment and risk management. And we, we have reason to believe we would see similar successes in other countries and regions. Thank you for listening to this talk and if you have any questions please feel free to submit them. Thank you very much.

Hello, I'm Dr. Crystal Southern. Thank you for inviting me to give remarks on behalf of the United States Department of Agriculture Food Safety and Inspection Service. Today I'm going to give a brief presentation on why whole genome sequencing is beneficial to the food safety and inspection service and how we plan to use whole genome sequencing in the future.

Food Safety and Inspection Service or FSIS employees are located throughout the United States at federally inspected slaughter and processing establishments, District and regional offices, FSIS laboratories, and at headquarters in Washington D.C. We work together to accomplish our mission of protecting public health by ensuring the safety of meat, poultry, and egg products.

Whole genome sequencing is a laboratory method that reveals the genetic makeup of an organism. Whole genome sequencing supports FSIS's mission because it strengthens our approach to keeping food safe by improving the ability to compare relatedness between food, environmental, and clinical isolates with a high discriminatory power. Whole genome sequencing consolidates several previously independent lab analyses, such as serotype and antimicrobial resistance, into a single data stream. Therefore we are able to obtain more information from one sample using less resources than with previous technologies. And because whole genome sequencing has been deployed with substantial coordination across a large network of federal, state and local public health laboratories, we can ensure that FSIS's surveillance protocols and science initiatives align with our public health and regulatory partners. As with PulseNet or other international efforts, this is key to having comparative international dialogue.

In this timeline you can see that implementing whole genome sequencing has been a multi-year effort. FSIS began transitioning from performing pulse-filled gel electrophoresis to whole genome sequencing as the primary subtyping tool in 2013. And by 2019 we successfully transitioned also typing activities to whole genome sequencing. In fiscal year 2020 we used whole genome sequencing to characterize about 14,000 food isolates.

FSIS currently uses whole genome sequencing data to identify linkages between human illness and FSIS regulated products during surveillance and investigation of foodborne outbreaks. We also use whole genome sequencing to more accurately discern pathogen and emerging microbiological trends over time, and to infer phenotypic traits from genes such as serotype and antimicrobial resistance. As part of FSIS's collaborative efforts, whole genome sequencing data and limited metadata are stored and shared in a publicly accessible database hosted by the National Center for Biotechnology Information at the National Institutes for Health. This fosters collaboration and allows the data to be amenable for analysis using bioinformatic tools by scientists around the world.

During foodborne outbreak investigations, whole genome sequencing helps us to rule out or rule in clusters and hence better focus our efforts and resources. For example, in 2018 along with the Centers for Disease Control and state and local health departments, FSIS investigated an outbreak of Salmonella Newport. With pulse-filled gel electrophoresis, the outbreak was identified as a single cluster. Whole genome sequencing, however, helped to identify that there were in fact three separate clades which are genetically related groups and not a single cluster. In addition, one clade was multi-drug resistant. So here we were able to clearly demonstrate the discriminatory power of whole genome sequencing. We were also able to determine that this outbreak was closely related to a previous outbreak associated with beef and travel to Mexico. As a result of this investigation, more than 12 million pounds of beef products were recalled.

And here is an example of how FSIS uses whole genome sequencing to support our inspection and verification processes. It is widely known that listeria monocytogenes has the ability to survive well in food processing environments. It can persist in difficult to clean areas, and from there, may contaminate ready-to-eat foods. FSIS uses whole genome sequencing to assess if positive listeria samples from a single establishment are an indication of recent contamination or environmental harborage or recurrence. In this first example, there are two samples from one establishment where at least the first four fields of the whole genome sequencing allele code match. An allele code is a standardized name applied to isolates that explains genetic relations. In this example we also make the condition that this is the first time we're seeing this listeria allele code in this establishment. Because the samples were collected on the same day, FSIS interprets this to be a result of cross-contamination.

In the second example, we have an additional positive sample from the same establishment, where again at least the first four fields of the whole genome sequencing a leopold match. However, because the sample was taken a little more than a year later than the previous samples, FSIS interprets this to be harbridge and the establishment is required to take corrective actions with respect to the food context surface.

The application of whole genome sequencing and food safety is evolving. In previous years FSIS focused on implementation, upgrading infrastructure and applying whole genome sequencing to routine activities where we previously used pulse field gel electrophoresis. Our vision is that we can use whole genome sequencing to form a foundation that will support projects that promote innovation and public health policies that are both science-based and data-driven. We now have a vast amount of data available for analysis and seek to find answers to the question - what more can we achieve? Using the tool in the toolbox analogy, FSIS is exploring which tool or combination of tools are right to get us to where we can use whole genome sequencing to inform risk reduction policies, develop targeted interventions, further improve outbreak response, or rank higher risk pathogens. Whole genome sequencing is an important component of FSIS's public health mission and helps us to understand the relationship between clinical, food, and environmental isolates. We are looking at ways to use whole genome see... Hello everyone I'm really happy to be here to tell you a little bit about how genomics has impacted the work that we do in food microbiology at the Canadian Food Inspection Agency. So for context, the CFIA is Canada's largest science-based regulator responsible for mitigating risks in the food supply. So every year about 30,000 samples are tested by the CFIA food microbiology laboratories which are located across the country. The organisms we look for are listeria, salmonella, and e coli, and action is taken as soon as we identify these. So food recalls would generally be done long before sequence data is available. This technology comes into play once you have isolated bacteria from foods. So in about 2014, our laboratories started sequencing all of the bacterial isolates recovered from CFIA food testing programs and this was done alongside the standard methods that were being used at the time. And we started looking at how the results of our WGS- based predictions compared to the traditional methods. As of 2018, we've eliminated many of these tests because results can be predicted from the sequencing data. So moving to the sequencing technologies, though, has provided a number of very important advantages. So first you can maintain custody of samples. In the past, food testing labs across the country were actually shipping all of their isolates to one and maybe two different locations for doing further testing, and you might have tests being done at four different labs. So not only is this expensive in terms of shipping and human resource requirements, but it takes quite a lot of time and there's always potential for errors as strains are handled by many different people in the process of generating results. We can derive results for most tests from sequence data and data is is a lot easier to ship than strains.

And one of the advantages that you'll likely hear about a lot is that the high resolution of the WGS-based typing provides much more confidence in cases where we're trying to match different types of bacteria. And you can imagine that you have two to five million data points in a whole genome sequence based analysis compared to some of the earlier methods, which give you 200 to 3,000 data points.

And these are just a couple of examples of why this matters. This is an investigation we were doing in 2016 to try to find the source of a cluster of listeriosis cases. So we compared the clinical samples to all of the strains that we had that were the same type. And we found some close matches from meat and cheese over here in green, but these isolates had more than 40 differences out of maybe the 3 million data, you know, the bases in the genome. So we could exclude these from further investigation. When the source was finally identified, the isolates from the implicated milk product were identical to the clinical isolates. So this technology helped us to both exclude several potential sources and then confirm matches once we found them.

And here's an example of an investigation of salmonella recovered from a sprouted flaxseed product in 2014. So this product was made by sprouting seeds. The sprouts were dried and ground up.

And you know, the salmonella was detected in the powder. So in a follow-up investigation in 2016, the labs were trying to find out if the salmonella was coming in from the seeds. When they tested the seeds directly, salmonella was not recovered but they could find it once they sprouted the seeds. And experiments done about a year after the initial investigation and in a different lab than the one that initially isolated the salmonella, the isolates recovered in the, you know, a couple of years later were identical to the initial salmonella isolates, zero to one SNPs, showing us that it was very likely that the contamination was coming in from the seeds. So third, with the availability of the sequence data we can start to look for genes which could make the organism more dangerous. E coli is a really good example of where we use this because most e coli is not harmful but certain types can bind to cells in the gut and produce toxins. We can use sequence data to rank e coli to try to predict how dangerous they are likely to be. And here's an example of where we use this. And this was an outbreak investigation in 2017 associated with a type of e coli O121:H19 and flour. When we were testing the flour we found a number of other toxin-producing e coli that weren't associated with the outbreak. The non-outbreak strains were determined to be in a lower risk category based on gene content and none of these other types of e coli were ever found in the clinical cases.

We can also look for genes that give the bacteria the ability to survive in a food production environment. For example, the type of listeria that we find the most often in foods are more likely to have genes that give them the ability to resist disinfectants. Listeria with these genes could potentially be more dangerous because they're harder to eliminate in a food production facility, leading to an increased risk of exposure to these types of listeria. We can also look at things like antimicrobial resistance. And this is an analysis of antimicrobial resistance and e coli recovered from different products. We see more resistance in some meats relative to others. While this doesn't currently impact regulatory action, we can contribute data to the Canadian AMR surveillance programs for virtually no cost, adding value to the data generation that we're already doing.

And that leads us to the final advantage. Data sharing is really easy with whole genome sequence data. And data can be compared even if they are generated with very different platforms. In January, we submitted about 1400 listeria sequences to the global public repository at NCBI, the Pathogen Detection Database, and this was a really interesting experience for us as this provided some context for some of the food isolates in our collection. And we could identify isolates that clustered with clinicals in the U.S. or other countries that had not previously been associated with Canadian outbreaks. And this is an example of how the submission led to the resolution of a global foodborne listeriosis outbreak linked to enoki mushrooms. The story was just published by the FDA and it's a really great example of the value of data sharing for resolving outbreaks associated with globally distributed foods. Whole genome sequencing has been a really fast transition for the government and particularly for food microbiology. Sequencing is giving us relatively quick and reliable answers and I think that the sequence data we are generating now really has long term value, particularly as we get better at interpreting it. At this point we're trying to digitize some of our really valuable historical collections. One of the things that we've come across is the lack of standardization of information collected with the samples, which really impacts their usefulness, but we are working with public health partners on improving this and if we do this right we can really future-proof information and make it easier to put data together for larger scale analysis. When we get there we can start to use it for monitoring trends over time and maybe identify emerging threats much earlier than we do now. As the sequencing technology gets cheaper, we can expect to see much more data coming into the ecosystem as costs go down and more labs are able to do the analyses. Data sharing is still a bit of a challenge for us but I think success stories like the enoki mushrooms will increase comfort level for governments so that we can get to the point of sharing data in real time, and ultimately reduce the number of illnesses related to...

Greetings, my name is David Erickson. I work with Dr. Meng at the Joint Institute for Food Safety and Applied Nutrition. JIFSAN works closely with the Food and Drug Administration and acts as a bridge to engage with academics and research scientists in the United States and around the world, and increasingly using WGS to ask questions related to food safety and foodborne pathogens. So in this presentation I'll give you an overview of a number of the projects that we're doing and how we use WGS.

In my experience, WGS tends to be used in two different contexts, just the definition - whole genome sequencing - is the most familiar and so what that means and what the cartoon at the bottom seeks to represent is the idea that we take a sample and that sample may be from many different kinds of environments but include the diversity of organisms. We might not care that much about the commensal background environment and we want to look more closely and specifically at foodborne pathogens, and so as we go from the diversity of microbes to the flask we're enriching for a specific desired pathogen. Once we have that and we reduce it to a single clone, we sequence it and capture data representing the entire genome. This allows us to assemble the genome, annotate it, and describe it. And there's many ways that we can describe it and there's many ways that we can use that described information. So the phylogeny in the right side of the slide describes work from a graduate student here Hee Jin Kwon, who used the sequenced genomes of a series of listeria to look at the role that mobile elements, particularly phage, can be used for source attribution and description of very fine scale gene flow and microevolution. And so she was able to use a phylogeny that was established by SNPs, also using whole genome sequencing, and map on the presence or absence of different kinds of phage and describe and confirm that they show a very fine scale resolution that allows us to distinguish between and identify different food processing plants. And this then allows us to understand movement and gene flow between them.

The second way WGS is often used and described Is Whole Genome Shotgun. So unlike the first slide where we were trying to take a complex sample and reduce it down to a single clone that can be sequenced, here in our little cartoon we start with a sample collected from the environment. It's diverse, but we maintain that diversity and we seek to collect genetic data from all of the constituents. So essentially we're describing the diversity that's in the sample, and this is often termed metagenomics. And so the example that I give looks at metagenomic reconstruction of different waste water facilities and the types of materials produced in the wastewater facilities, and they can describe not only the taxonomic, what they say is metagenomic, but also the resiststome - the diversity of resistance genes that are present in these different environments and look for correlations there. And so I'll give an example later where we are doing the same thing. So another example from a student here at JIFSAN looked at the relationship between isolates of e coli collected from cattle with clinical isolates that are known to cause human disease. And so by capturing those isolates, sequencing them, assembling the genomes, that allows us to create a phylogeny where we look at the relationship between these e coli isolates recovered from cattle clinical isolates and then map on virulence and AMR determinants to look at the relationship between these isolates and the genes that induce pathogenicity. An interesting aspect of the study is that the student found novel serotypes with different combinations of novel combinations of virulence and AMR genes. These novel serotypes would have not been able to be identified correctly or classified if we had used traditional pcr-based methods. So by sequencing the whole genome, getting a proper assembly allowed us to annotate and correctly identify these novel serogroups. Another project that involves both a mixture of these two different kinds of WGS whole genome sequencing, and also to a certain extent, shotgun sequencing for metagenomics. So JIFSAN works with a diversity of different faculty and so we're a bridge between researchers throughout the United States and also in other parts of the world with a globally dependent food network, we're essentially all in this together. So the objective of this study is to look at surface water environments and to link them and compare them with clinical isolates. This allows us to get a sense of where pathogens that enter the food chain are coming from. The current project a lot of the sequencing is done at JIFSAN but much of it is done with collaborators in Brazil, Chile and Mexico and so this is a big project with hundreds of isolates, where we're looking at surface water contamination. And so initial analysis of this where we're able to create a phylogeny that includes surface water samples and clinical isolates shows that there's a lot of diversity in the clinical isolates that are captured. There's a high degree of local clonality so that things from the same environment tend to be clonally related to one another but not exclusively, and some of the strains are widely distributed. This is not exclusively the whole genome sequencing because increasingly we're going to be looking at the greater diversity of those water samples that are collected to look at how the commensal background environments may relate to the maintenance of pathogens and their introduction into the food chain. Last thing that we're working on is a project where it's metagenomics and we're using whole genome shotgun from samples that are collected along the path of food processing. So the cartoon at the bottom Farm to Fork looks at taking samples throughout the chain of processing and examining how these environments change, how the microbial environments associated with each one of these, and so the representation there is as we go from farm to transport to processing, we can compare changes in taxonomy, functional gene diversity, resistome diversity. This is with a colleague who's over in the the College of Veterinary Medicine and he works primarily with chicken so we'll be examining the broiler chain and we'll be looking at the complete diversity of all these characters to get a sense of where interventions are possible and effective. Okay, so first of all wanted to thank the organizers of this session for the invite to come and speak to you for a couple minutes today about how we're using whole genome sequencing in the state of Pennsylvania, particularly at the Pennsylvania State University. My name is Dr. Edward Dudley. I'm a professor of food science at Penn State and also the director of a large culture collection called the E coli Reference Center. So up to this point you've certainly heard about the U.S. Food and Drug Administration's GenomeTrakr program that has this goal of implementing whole genome sequencing for public health across the United States and internationally. So GenomeTrakr within the state of Pennsylvania, which is in the northeast part of the United States, is set up as a consortia between the large university, being the Pennsylvania State University, and the Pennsylvania Department of Health. So Penn State as I mentioned is a larger university located approximately the middle of the state with about 40,000 undergraduates and 6,000 graduate students. And as part of a consortia of universities that are called land-grant institutions that were first set up in the 1860s with the goal of promoting agriculture and engineering education across the United States. Our partners, the Pennsylvania Department of Health are primarily located in Harrisburg, Pennsylvania which is the capital of the state, and the laboratories are in Exton, Pennsylvania which is just outside of Philadelphia. And together for the last four years we've been using whole genome sequencing to address questions of common interest, including what pathogens are currently in foods that are being transmitted to humans within the state, whether there are new pathogens that are being transmitted into Pennsylvania by human or wildlife movement, and what I'll focus most on today is what antibiotic resistance genes are circulating within and outside of the state. So the first story deals with the organism shigella. So shigella is a global problem; about 100 million people per year become ill through infection with this organism. And within the United States there's about 77,000 cases you know per year, by a group that we call the drug resistant shigella. Now these are labeled by our centers for disease control as a serious threat, primarily because as is shown in the graph in the bottom right, I'm sorry the bottom left, there's an increasing prevalence of antibiotic resistance that's seen with the shigella, particularly the frontline antibiotics such as ciprofloxacin and azithromycin and particularly in the last, you know, two to three years there's been a dramatic increase in resistance. So to initiate this study we focused in on a particular shigella species called shigella sonnei, had a collection from the Pennsylvania Department of Health from individuals who both had just recently traveled internationally and those who hadn't left the the United States in any recent time. So we use whole genome sequencing to evaluate these organisms with a couple of primary questions in mind. First being, are the isolates that are circulating within the state of Pennsylvania similar to those that are circulating worldwide? And additionally, are the individuals who become sick from international travel carrying with them strains that are distinct from those that they would have acquired within the United States? So as I mentioned, we sequence the genomes of all these and we can do an analysis that's kind of shown in the upper right part of this slide. We build what's called a phylogenetic tree, which is just simply more of kind of a family tree to assess how related are these individual bacterial isolates to one another. Now global populations of shigella sonnei can be divided into one of five related clusters that are referred to as lineage one through lineage five. So one of the first analyses that we're showing in this grouping is that most of the isolates that we obtained from patients in Pennsylvania, group into lineage two or lineage three, which is noteworthy because these are the lineages that are most commonly seen in worldwide collections of clinical isolates. In the bottom left we can show that from the genome sequences we can make predictions of what antibiotics these organisms are resistant to. So on the vertical axis are some individual isolates that we looked at and going left to right in the colored boxes are various antibiotics that we predict these isolates are resistant to. And really just trying to communicate these isolates are resistant to numerous antibiotics, again, including several ones that are currently front-line drugs including some quinolones that are shown on the right-hand side in green. So the take-home message of this study was that the shigella isolates that are circulating within Pennsylvania are very similar to those that are circulating globally and that we don't see any genetic difference between those that are acquired through international travel versus are acquired domestically. So really that communicates to our Department of Health that treatments that are currently used for patients infected with shigella sonnei in other parts of the world should be effective within the state of Pennsylvania. The second story I'm going to tell you deals with the organism e coli, which is a common bacteria that's found in the intestinal tract of mammals including ourselves, and for various farm animals including cattle. So in this study, students went out to various farms and collected e coli, both from young animals (from calves as) well as from from cows. There were two different types of farms that they looked at. Some were beef cattle farms; so these are farms in which the animals are raised for their meat. Otherwise, the other types of farms were dairy, where the animals were raised for their milk. Animals were given different feeding regiments but noteworthy for the calves, they received milk through one of three different mechanisms: either direct nursing from their mothers, receiving waste milk which was often coming from from cows that were currently being treated for various infections, or bulk tank milk which was a collection of milk from various animals. So, isolated these e coli, generate the genome sequences, and again predicted from the sequences what antibiotics these organisms may be resistant to. And what we found pretty strikingly in this study is that the e coli from the young animals, from the calves that had received milk, that were coming from animals that were currently being treated for various infections with antibiotics, had a much greater prevalence of antibiotic resistance than e coli from any of the other categories of either cattle cows or calves in the study. Additionally, we can use the whole genome sequence to make predictions about the risk that these certain e coli may pose to the human population. So again, we've got the various farms, the feeding regiments, the different isolates that we analyze in this table. And a couple of different analyses that we can do, such as sequence typing, serotyping, and pathotyping - which is just another way of saying that we can make some predictions for the genome sequence of whether these isolates fall into different categories that are known to pose a threat to human populations. And basically from these isolates that are shown in this table, they carried many of the genes or other genetic characteristics that are typical of e coli that cause disease in human beings. So again, whole genome sequencing not only allows us to tell how related our various bacterial isolates to one another make predictions about antibiotic resistance, but also all it provides to us some evaluation of risk assessment that isolates may serve to human populations. So from a standpoint of a university, I think one of the biggest impacts of whole genome sequencing is not only the ability to work together with our department of health to promote projects that help the health of our populations, but particularly to educate the next generation. Whole genome sequencing is a tool that is absolutely critical for the next generation of microbiologists and food safety professionals to be well-versed in. Pointing out a picture from my lab from a few years ago, highlighting this woman here, Rebecca Abelman, who currently works for the Maryland Department of Health, she's the one who led the shigella study that I led this off with. Andrea Kiefer, who's right here, who did another study that I won't have time to present, looking at the genomes of salmonella that are isolated from retail meats around the state of Pennsylvania. And Sydney Major Wicks who led this study down here, which would involve the genome sequencing of over a thousand e coli that were isolated from diverse environments. So again, whole genome sequencing provides kind of a really cutting-edge tool that's really exciting for students to engage in, allows them to to learn this technology while also helping with important public health questions. So with that I wanted to thank the the current lab, particularly Yessi Phew who is the postdoc who's leading the genome tracker program in my lab at the moment. And Aaron Naraki who is assisting with with some of this as well. And Sharon Diaz Miranda, who also has a pretty exciting project, again, that I won't have time to go into. And particularly over here, Mikusha Mikanatha, who's a lead epidemiologist at the Pennsylvania Department of Health who's been absolutely instrumental in terms of integrating many of these projects in with the needs of our Department of Health. And lastly, the funding agencies including the the U.S. Food and Drug Administration for providing the funding to run GenomeTrakr within Pennsylvania.

Hi, my name is Jim Herrick and I'm very happy to be speaking with you today. I'm going to give you a brief overview of a laboratory course sequence that we've developed in environmental pathogen surveillance and pathogenomics. This course sequence is made up of two modules: a wet lab module and a genomics module. It requires some basic background in microbiology lab techniques, but no previous knowledge of microbial bioinformatics. The details and some of the results will be appearing in two papers that will be published in the next few months. So just as a little bit of background. The university where I teach, James Madison University, is located in Virginia in the eastern United States. Particularly, we're in the middle of the Shenandoah Valley, which is west of Washington DC, an area famous in U.S. history but also well known for its agriculture, especially animal husbandry. So here we see the number of animals in just four of the central counties in the Valley in 2012. A vast number of animals and as a result, a vast number of manure being produced. As expected, much of this manure finds its way into rivers and streams. On the right you can see red waterways which are impaired due to agriculture as measured by e coli counts. Preliminary work in our research lab showed that we could readily isolate salmonella and more recently, klebsiella pneumoniae, from these waterways. So we have developed protocols for enrichment and isolation from environmental and manure samples. This has been incorporated into a course, a pure laboratory course taught to junior and senior college students each semester. This model has been implemented at other universities including Kansas State University and Longwood University.

The genomics module was recently taught in a three-day summer workshop at JMU for faculty students and other researchers from the eastern U.S. This was sponsored by our own Center for Genome and Metagenome Studies. So we have developed detailed protocols for each of the modules here. You see some of the bioinformatics protocols, and this is one of the particular protocols in bioinformatics and then a wet lab protocol.

When the students collect their stream samples they also collect metadata using their phones. For example, time and date, and latitude and longitude, and various abiotic site characteristics. And this is compiled using the app epicolect5. This is the current salmonella isolation and identification workflow. We begin typically with stream sediment because total bacterial counts and sediments are typically much higher then in the water column, and salmonella and other fecal bacteria have been shown to persist so we believe that sediment populations represent a more long-term reservoir. It's a little more challenging than with water in some ways but this protocol is an effective one.

Once the students have isolated and identified their isolates, they're sent off to the Virginia Public Health Laboratory, the Division of Consolidated Laboratory Services, where they sequence these for us as they're very interested in tracking strains isolated from the environment. When we receive the reads back from sequencing run, the students are then taught how to do quality control, how to assemble, how to visualize their genome assemblies, how to annotate their genomes, and then they also do downstream analyses, including serotyping, antimicrobial resistance gene identification, identification of pathogenicity islands and plasmids, etc. There are a lot of excellent easy to use computational tools now available for studying microbial genomics without any previous knowledge of bioinformatics or scripting or coding. The main bioinformatics platform we use is Galaxy, specifically an implementation of Galaxy known as GalaxyTrakr, developed by the FDA. This was specifically designed for public health and pathogenomics, so the tools are limited to those useful in these fields which makes it quite tractable for students. There are more tools available in Galaxy itself, in Galaxy Europe and Galaxy Pasteur, and since these have an identical interface it's an easy transition to using these tools from GalaxyTrakr.

In addition to short read Illumina Sequencing, we also often use Nanopore Sequencing for isolates of particular interest. And Nanopore Sequencing is much easier to use than Illumina. It's suitable even for beginners, as you see my students shown here, and it's also faster and cheaper overall, at least in our hands, than Illumina. Long-read Nanopore data can be combined with short-read Illumina data to get nearly complete genome assemblies using a hybrid approach. We can see the results of this on the left is a visualization of assemblies from short reads only. You see they're somewhat scrambled, whereas on the right we see the result of a hybrid assembly with short and long reads. And this results in often near complete assemblies as you can see, of chromosomes and often of plasmids, for which this is especially useful since plasmids are a special interest of our research lab. We can visualize phytogenetic relationships, serotypes sources In our case we had a broad diversity of serotypes that we isolated and cordino multilocus sequence types. A very useful tool for us is a Microreact. Microreact is an interactive web-based application from the Wellcome Sanger Institute that allows us to look at the distribution of our isolates across both geographic space, as you see here on the left, and a temporal space, at the time in which they're collected, on the bottom. So we can look at everything from our sites; there's the isolates, isolated from specific sites, their serotype placement on a phytogenic tree, and then actually see when they were isolated. So for example, well we can also look at metadata like multi-locus sequence type and endometric resistance phenotypes. So for example, here we found this identical salmonella brender up in two locations and they were actually isolated in the same stream over the course of a year. One can also use phylogenetics to look at possible sources of strain. So here a student found that our typhomerium isolates clustered most closely with typhumerian from poultry.

Our lab has a particular interest in plasmids and we've developed a very easy, efficient laboratory method for isolating large self-transmissible plasmids. These were isolated by students in the class. You can see here on the top, the large plasmids, these run from about 50 to 200 kb and we have a particular interest in our lab in transmissible antibiotic resistance.

This is a very large self-transmissible plasmid from a student isolate which one of my research lab students assembled using a hybrid approach. It has an amazing number of antimicrobial resistance genes, metal resistance genes, of transposons IS elements, and even has a virulence gene on it. So plasmids such as this have great potential for mobilizing resistance genes to and from other pathogens and to other native bacteria in the stream where they're found. The isolates in their genomes become then raw material for my research lab where we study environmental human pathogens and their mobile genetic elements. And these are some of the topics that we are addressing currently. So I just want to thank all who have contributed to our studies, and especially those who support our sequencing efforts at the FDA and the Virginia DCLS. Thank you.

Thank you to the Food Safety Cooperation Forum, the Partnership Training Institute Network for the invitation to present in the awareness webinars. And this webinar being whole genome sequencing laboratory capacity building for environmental food safety testing. Wow that's a mouthful! My name is Kelly Hoon and I'm a microbiology and infectious disease specialist for Illumina, and it's my pleasure to be here today. So let's get started. At a very basic level, genomics is a tool that changes lives. So what do I mean by that? You know thirty percent of children with rare disease die before the age of five today. And children whose genetic abnormalities are not diagnosed or treated early enough sometimes suffer irreversible damage. In the future, we believe every child will get sequenced at birth. Those with genetic diseases can be treated within 24 hours, ensuring the best quality of life. For the immune system, in the past five years there's been an amazing progress in cancer therapies but still only 20 to 30 percent of cancers respond to immunotherapy today. In the future we'll understand enough about the immune response to create combination therapies that will effectively stimulate the immune system. T-cell therapies and cancer vaccines will be developed that encourage targeted and personalized therapies. For the microbiome, the microorganisms that live within us or among us outnumber human cells three to one and have been implicated in Parkinson's Disease, anxiety, and mental health. In the future, the microbiome will be regularly assessed, like a mammogram or an x-ray, to understand the health of your gut bacteria. Consumer sequencing, another area not only are we able to understand genetically where our ancestry is from but in the future, consumer sequencing will also be able to tell us about our lifestyle or our diet or needs to change either one of those. Food and safety, it's hard to even know where to start here, but with over 800 million people in the world that are food insecure we estimate that global food production will need to double by 2050 to feed a population of 9 billion people. Now it cannot just be a matter of supply, it has to be safe as well, and this is why we're here today.

By now, I believe Eric Stevens has already spoken about Illumina technology and the science behind it. So I've been asked to share a little information about the company, the products, and the people that are here to support you.

So who are we? Illumina is a global leader in genomics helping to solve the toughest challenges in human condition, and inspire hope for people around the world. Our dedication is to make our solutions increasingly simple, more accessible, and always reliable. This has broken down barriers and fostered the growth of an entire genomics ecosystem. We're headquartered in San Diego, California in the United States, but we have offices all over the world with over 7800 employees and just celebrating our 20th anniversary back in 2018. At Illumina, we manufacture the microarrays and sequencing products and the instruments that they're read on, and our mission is to enable the power of the genome to improve human health.

Simply put, we're here to help, with offices located around the world. We can be reached by phone during normal business hours, and technical support online 24 hours a day seven days a week. More specifically, for Asia Pacific and Japan I've included here contacts for various countries across Asia Pacific and Japan. We sell and support directly with Illumina employees and in conjunction with our channel partners. You can see here in more detail by region who those contacts are, and I've included a link in the top left corner for more details on our website with phone numbers and email addresses for each contact in each country. Our customers represent an active install base of more than 13,000 sequencers at 6300 organizations in 90 countries, and they'll drive the next chapter of genomics and they'll do this together. Over 5000 luminous systems are connected to our BaseSpace cloud, creating a global community of more than 28,000 active users and 3700 developers who have built over 8600 custom sequencing workflows on BaseSpace. Our customers are generating sequence data at an unprecedented rate and scale, and in 2018 over a hundred petabytes of data were generated across our systems - a record of sequencing data generated in one year.

Let me introduce you to our portfolio, supplying solutions for every application and every lab. They all have the latest innovation of ease of use while creating the highest accuracy and next generation sequencing. Starting on the left, the iSeq 100 system is the smallest and most affordable of our system and essentially is a one foot by one foot by one foot cube. Moving into our MiniSeq and MiSeq systems, these have similar footprints and the number of reads, but the MiSeq offers the longest read chemistry of all of our platforms. As you likely saw in the previous slide, the highest number of systems are placed with this system as well. As you scale in needs the NextSeq system generates a higher amount of data to allow for more samples and applications. And lastly the NovaSeq 6000 which generates flexibility and scale from metagenomics, human genomic sequencing, and population scale applications. Here is what the workflow looks like from sample prep to analysis starting with sample preparation on the left, this is the extraction of nucleic acids from your sample. For example, this could be an environmental swab or a cultured isolate. Library preparation, which takes this nucleic acid and adds adapters that are recognized by the sequencer. From there, sequencing is performed. All of our systems are easy to load with touchscreen technology and RFID implementation to track reagent lots and expirations for each run. While the system's running, data analysis will start automatically. Primary analysis is done on the instruments and this is turning the signal into base calls like t, c, a and g. Secondary analysis, assembling of the reads or comparing to a genome. And then interpretation in food safety, this is the part that links the sample back to the product or the clinical case.

In simple terms, this is essentially what you need to start up a new lab. Minimal laboratory setup or equipment is really required for whole genome sequencing and that's what we're talking about today, for the sequencing of of microbial contaminants within a sample. And so you have a centrifuge or a picafuse, a thermocycler. This is an image of our Illumina benchtop sequencer, this is the MiSeq system that you saw in one of the previous slides. Then you've got the sequencing regents themselves, general pipets for the lab, and a magnet stand. And I've also included here a link to our support page for the library prep that's specific to this application. On that support page it does have an equipment list for recommendation but I would also suggest that you contact your Illumina representative, that can walk you through that as well.

At this point, Mark Allard from the Center for Food Safety & Applied Nutrition from the USFDA has already spoken about how they use this technology to understand or identify contaminants or foodborne pathogens. I'd like to follow that up with applications for manufacturing and give you an example of how that was implemented.

How is next generation sequencing impacting food safety testing? From outbreak detection, and this is where the regulatory agencies work with the data to understand key questions of the outbreak. What is it and are these samples related? It provides nucleotide level resolution for strain and serotype identification. This data is then used to understand if it is related to another sample like a clinical infection, or supplies for the product, or if it has never been seen before. In outbreak management, the idea is to avoid a contaminant altogether by testing proactively the environment, the supply chain, etc. and find potential causes before they actually happen.

It can also be used for additional analysis or what I like to refer to as the applied space, to apply this to standard practices for protocol development for say, sterilization, for example. Understanding organisms pathogenicity or spoilage or even resistance to cleaning protocols. Another capability is speciation for adulteration. So for example, when the product says it's albacore tuna, is it really albacore tuna? There are a lot of examples of where next generation sequencing is being used in food manufacturing, but here's a great example of where a plant in Poland used it to assess its protocols and product safety. They took both the culture-dependent and culture-free approach to sample the product, in this case vacuum-packed ham, over time to detect the organisms and assess if their existing protocols and storage conditions were sufficient for product safety. They followed these products in their existing refrigerated storage conditions and saline solution from the time of packaging and sealing of the vacuum packed ham, and did various time course samplings to understand, was 30 days the appropriate life of this product? You can see here where we are able to identify organisms and shifts in the organisms' population that could be contributed to the smell, appearance, texture changes, and the product over time.

Using this approach, this group was able to detect specific spoilage organisms in cooked ham, including microorganisms that were culturable in appropriate media as well as microorganisms that are hard to culture or unculturable under laboratory conditions. The microbial evaluation of foods is critical during the development and implementation of new technological solutions, and packaging techniques for livestock based products and concluded that they may be employed for predicting shelf life as well.

Where we've seen a large amount of success, and and why we're here today, is to the application of whole genome sequencing to identify the entire genome of a sample and to understand what is going on. The most successful model has been the evolution and collaborative effort between the FDA, the Food and Drug Administration, and the CDC, the Center for Disease Control, with the GenomeTrakr project and the PulseNet network to trace and evaluate products through manufacturing in the U.S. And then in public health outbreaks to link clinical outcomes back to potential contaminated products that patients may have eaten. The idea here is really to prevent this earlier from even getting into the food supply. So Mark Allard has actually shown this in his talk provided earlier, but this is the whole idea, is to use these experiences, share these procedures to develop similar models in other regions to prevent infection and improve products manufactured in different regions.

With that, I will stop and share with you a few resources. Some I've provided earlier in the presentation, as well as a few additional resources that you can use in the future. when you have this deck in your possession. There's a large amount of information here, if it's protocols or videos or trainings, all of which are located on our website. Our support page is an amazing resource for information so I'd encourage you to explore. Also included here is a link to our list of channel partners and contacts in country, so you have that information as well. With that, I would like to thank you for joining us today to learn more about Illumina and the people here to support you.

Hi, my name is Adam Allred. Thank you for the opportunity to speak to you today. I am a bioinformatician at Clear Labs, and I'm going to speak to you about the past, present, and future of whole genome sequencing at Clear Labs.

How do we use WGS at Clear Labs? First, we use it to inform development for microbial detection and molecular stereotyping assays. We also use it to confirm the true species or serotype of archived microbial specimens. And finally, we confirm species and serotype of microbes in cases where there are discrepant results between different tests using WGS. What technologies do we use for WGS at Clear Labs? We use both Illumina iSeq and Oxford Nanopore GridiON for WGS. Our main applications are microbial genomics of bacteria and viruses including salmonella, listeria, and SARS-CoV-2 and I'm going to explain later on in the presentation how that potentially relates to food production.

One of the reasons we like Illumina iSeq is because it offers a high accuracy and high throughput. The reasons why we like Oxford Nanopore is that it offers a smaller footprint and greater speed. And in fact, you can see on the right we have the Illumina iSeq on top and on the bottom is our Clear Labs end-to-end liquid handling solution, including sequencing. And because of the small footprint of the GridiON it actually sits inside that robot there and that's our Clear Lab system for WGS.

So we use WGS to inform assay development as I mentioned. With enough representative genomes of a given species and serotype, we are able to find patterns in the genomes that can be exploited to distinguish that species or serotype from all others. And we use WGS to supplement genomes in public databases where we can identify unique signature regions in these genomes and develop targeted sequencing assays using that information. What this looks like is first we have a signature identification patient step where we have sorted our genomes into two different partitions, either inclusion genomes or exclusion genomes. And from this we have custom scripts that can identify signature regions that are present in all inclusion genomes and simultaneously not present in all exclusion genomes. The next step is to use primer design software to design primers against that signature region.

We have some in silico checks that we do to check for specificity of the primers to make sure that we're amplifying everything that we want to amplify from the primers in the inclusion set, and also that we are not amplifying what we do not want to amplify in the exclusion set. And finally we can iterate on this and develop primers, specific pairs of primers, for a number of different species or serotypes and we can combine that in a targeted sequencing assay and that's what's depicted in this final step.

Another way that we use WGS at Clear Labs is to confirm the serotype of archived specimens. We've noticed that archive specimens may have a serotype label that is incorrect and we routinely source salmonella stereotypes, for example, from academic institutions with highly regarded bacterial isolate collections, to serve as positive controls for our development of our assays. And yet we find that some of these samples are actually mislabeled such that they give us misleading results, because we get a result from our essay that is discrepant from the label. But when we actually do WGS and subsequent SeqSero analysis, and I'll talk a little bit about SeqSero in a moment, we find that the serotypes of these specimens are actually different than what their label was. And so WGS is the gold standard that helps us resolve that, and here are four samples that we sequenced in-house that we found that were different than their collection label was. And SeqSero is the tool that we use to do that, and what SeqSero does is it extracts the anagenic regions out of the genomes and can then give you the serotype of the salmonella genome based on that. So it's very useful for these kinds of situations. We have also used WGS at Clear Labs to resolve discrepancies in listeria detection and speciation. We recently sourced listeria samples from a food producer and these samples were subjected to three different tests and the three different tests gave different results. So this again is another place where WGS can help us resolve discrepancies.

In this case, qPCR was run by the customer on these samples and showed positive for listeria. We ran a MOX assay and the observation was typical for all three samples. API biochemical assay actually showed listeria ivanovi.

And our clear safety listeria test showed negative for these. When we did the WGS we discovered through phylogenetic analysis that these samples actually belong to listeria sensu lato species, so they were outside of the typical six listeria sensu stricto species which are the target of our clear safety listeria assay, and so that explains why they were negative for that assay. So WGS was able to show in this case that the samples belonged to listeria sensu lato and this underscores the importance of WGS as the gold standard for listeria speciation.

One of the things I wanted to touch on was how we plan to use WGS in the future. One of the ways that we continue to use WGS is to supplement genomes from public databases in order to enhance MLST-like molecular serotyping. MLST for those unfamiliar stands for multi-locus sequence typing and I will discuss this further in a subsequent side. A more recent focus for us is to use WGS to track SARS-CoV-2 in food production facilities, and I will discuss this further as well. Regarding MLST or multi-locus sequence typing, WGS can be used to identify unique combinations of alleles that are specific to a particular serotype. So the way this works is we sequence genomes and we find regions that when the alleles of those regions are combined they give a unique fingerprint for that serotype. And an advantage of an MLST approach, relative to other molecular stereotyping approaches based on KW for example, is specificity. The KW targets, which are the antigenic regions of the bacteria, are present in other Enterobacteriaceae outside of salmonella, and so we can gain increased specificity through this MLST approach. And clear safety molecular serotyping allows for expansion of serotype support as we sequence additional genomes and add those alleles and those combinations of alleles that are specific to their stereotypes to the database.

A new area of interest for us is using WGS to track SARS-CoV-2 infection and I want to illustrate an example of how this can be useful. Using Nanopore sequencing technology, we recently produced whole genomes for 20 clinical samples. And these samples were analyzed along with publicly available SARS-CoV-2 genomes to produce a phylogenetic tree, shown here. The 20 samples fell into 13 distinct clusters which are indicated by the purple in the tree.

So let's zoom in on one particular cluster in this tree and see what we can learn from this analysis. The purple outlined genomes were from a particular state which we're calling State C and were sequenced at clear labs. Other genomes from that state that were acquired from public databases are in blue. The earliest observed case was on March 13th in State A.

The tree shows that the virus was likely transmitted then from State A to State B and from State A to State C.

We observed two mutations which occurred along this branch that is pointed to. A C to T non-synonymous mutation in the nsp7 protein, and a G to A energenic mutation between orf9 and orph10.

We can also see that it's likely the virus traveled then out of the country to Argentina and England since we see positives from those locations.

So now having gotten a flavor for how WGS and phylogenetic analysis can help us track the flow of SARS-CoV-2 infection, I wanted to discuss a possible scenario that could face a food production facility. So let's consider the case where two workers in the same food production facility test positive for SARS-CoV-2. Because of that, there might be a threat of shutting down the facility. But when we use WGS and phylogenetic analysis we can actually see whether the shutdown would be a productive outcome or not. And we can use the shape of the phylogenetic tree to help us decide that. So if for example, the tree looks like this, then we know that the sample from worker A and the sample from worker B are phylogenetically linked and the conclusion from this is that the positives are likely attributable to an outbreak at the facility, which may mean a outcome that is different than if these two samples weren't linked.

On the other hand, if we do the WGS of the two samples and the phylogenetic analysis reveals a tree that looks like this, where the samples are not phylogenetically linked, in fact they are more closely related to samples outside of the facility, then the conclusion is different. The conclusion is that the positives are attributable to independent infections and it does not look likely that this was an outbreak situation at the facility itself. And perhaps with this information we can avoid shutting down a food production facility. In conclusion, WGS can help supplement public databases in order to expand speciation and serotyping efforts of existing detection platforms. WGS is useful as a gold standard to resolve discrepant assay calls or correct mislabelings. And WGS is useful for virus source tracking and may help prevent unnecessary shutdowns of food production facilities.

That's the end of the presentation today. I'd like to thank the organizers for the opportunity to be able to speak with you.