HHS Public Access

Author manuscript

Food Prot Trends. Author manuscript; available in PMC 2024 March 08.

Published in final edited form as:

Food Prot Trends. 2019 November ; 36(6): 482–485.

From Outbreak Catastrophes to Clades of Concern: How Whole Genome Sequencing Can Change the Food Safety Landscape

Robert V Tauxe, MD, MPH [Director]

Division of Foodborne, Waterborne, and Environmental Diseases, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia

Attending the IAFP conference has been an annual highlight for the past decade, and a time for learning and sharing for more and more of my colleagues at CDC. It is an honor and a pleasure to join you here now, and to honor the formative contributions of John H Silliker to the field of food safety.

When I first came to CDC to start a two year fellowship in field epidemiology and joined the small excited group that just 6 months before had identified *E. coli* O157 as a human pathogen, there were many questions in the air. One of the most interesting and persistent was around the first efforts towards molecular epidemiology. How could the new laboratory tools of molecular subtyping help solve epidemiological problems? Back then we tried plasmid profiles and satisfied ourselves that they could sometimes help us solve outbreaks. When a staff position was offered at CDC, I was fortunate to be able to start a career working in that same group for what has become several exciting decades.

Many of the things we do in public health at CDC start with surveillance. Surveillance means that when someone gets sick and sees a doctor, who diagnoses an infection that is on the list of diseases that should be reported, the public health department is informed. That is the start of a lot of what we do. We investigate and control outbreaks, to stop them and prevent future illnesses, with a lot of partners. We use the information we gather to drive illness prevention policy. We support our state and local public health partners to do what they do better, and our federal regulatory agencies and other partners to fulfill their primary roles in food safety. The group that I joined innovated particularly by applying advanced technologies to make our public health processes and decision making better.

Public health is a collaborative process with a lot of partners. Caregivers and clinical microbiologists make the diagnoses and report them. Members of the public, ill and well, provide critical information to the investigators. Local and state health department staff work to investigate and solve local outbreaks, conducting interviews, subtyping bacterial strains, and conducting traceback and control for events confined to that state. CDC is the national public health agency, conducting the epidemiologic investigations that are launched when a multi-state outbreak is detected, and supporting state and local health departments in what they do. FDA and USDA/FSIS, the federal regulatory agencies, engage in the investigations to trace suspect foods to sources, assess production and processing facilities, and implement control measures. More and more the food industry itself is a partner, assisting with traceback, assessment, recalls and prevention. Our arena of public health is a unified cycle. It starts with surveillance which may detect a problem, then investigation that

reveals the nature of the problem and may indicate how an immediate action can control it and keep more from getting sick. The investigation may also indicate that research is needed to better understand how to prevent something similar from happening again. Academic partners are critical to that better understanding. Once control and prevention measures are taken, we close the loop as surveillance tells us that whether the frequency of illness has decreased, a sign that the new prevention steps are effective.

About a decade after I started at CDC, the Hubble telescope was launched, and in 1995, with improved optics, a remarkable experiment was conducted, called the Deep Field Survey. The scope was pointed towards the darkest part of the sky, with the fewest stars, and the camera hatch was opened for a 10 days exposure. The photographic image that resulted stunned astronomers, as it revealed vast numbers of distant galaxies and star clusters that had previously been invisible. The tools you use define what you can see. This is true for microbiology as well as astronomy, of course. That same year, we launched the first pilot test of the concept of routine molecular subtyping of bacteria to detect clusters of related infection, using a then state-of-the art method called pulsed field gel electrophoresis (PFGE). And also in 1995, we had our first Hubble moment with PFGE. That year, the Minnesota Department of Health was the first to try subtyping all of their *E. coli* O157:H7 strains. The 183 confirmed cases started to be reported in April, rose to a peak in July and August, and then tailed off into the winter low. Though no discrete outbreaks were apparent in the general pattern, PFGE subtyping revealed seven discrete clusters easily separated from the background of sporadic cases, each a distinct entity that could be investigated separately. With this promising observation, the following year, we launched PulseNet formally, starting in Minnesota and three other states, and then rapidly growing as more states joined the network. We started with *E. coli* O157, then expanded to include *Listeria monocytogenes*, *Salmonella* and other enteric bacteria, and the initial effort in 4 pilot states expanded to include all 50 states, and several large cities. The central premise was that if we found a cluster of infections caused by strains of *E. coli* O157 that were indistinguishable by the PFGE methods, they are likely to share a source; if we interviewed those patients, we might very well be able to identify what it was, something we wouldn't have known otherwise. The PFGE subtyping enhanced signal detection against the background noise of sporadic cases. PulseNet is now over 20 years old, with participating laboratories testing over 50,000 bacteria a year, using a standard subtyping method and storing the information into a common database, that all participants can review and use. FDA and USDA/FSIS participated from the beginning, testing isolates from foods and animals and putting the results into the same database so that they can be compared immediately with the clinical isolates.

As PulseNet grew, the number of multistate foodborne outbreaks reported in our outbreak surveillance surged, from 2–3 per year in the decades preceding the launch, to 23 such outbreaks in 2010. By linking cases together across the subtyping network, we could now detect and investigate a dispersed cluster spread across several states, which meant we were able to stop an ongoing outbreak, identify food safety gaps early in the production of a food that was then distributed across the country, and drive improvements in prevention across the system. This depended on finding that signal by subtyping strains, investigating those clusters by interviewing patients that have bacteria with the same subtype, and linking that

information to food source traceback information and to any non-human isolates that may be relevant.

PulseNet and other major improvements in surveillance and in prevention followed in the wake of the large and devastating outbreak of *E. coli* O157 infections in that was linked to ground beef in 1993. This was a time of great improvement in our food safety systems, including modernization of inspection of meat and poultry with a new focus on reducing pathogen contamination, and great attention to the safety of processed meats. Also in 1996, we at CDC with 10 state partners, FDA, and USDA/FSIS began a comprehensive surveillance program called FoodNet, to track the frequency of several infections often spread by food. Between 1996 and the mid 2000's, FoodNet documented major reductions of between 43 to 49% in infections caused by *E. coli* O157, *Listeria monocytogenes*, and *Campylobacter* though *Salmonella* infections remained essentially unchanged. Since the mid-2000s, FoodNet has not seen further substantial reductions in these infections, which makes it unlikely that we will reach the desired 25% reductions from 2007 that are part of the Healthy People 2020 goals, unless something changes.

What we decided we could do at CDC was to make surveillance and investigations even more powerful by applying whole genome sequencing for routine subtyping in PulseNet. The cost of sequencing had dropped dramatically, and the technology could now be applied in public health laboratories. This looked like it would have several benefits. Perhaps this method would let us detect, investigate and control outbreaks we may have been missing and identify other emerging problems that need to be addressed. We could also use our surveillance to attribute illnesses to specific food categories and target interventions. New tools applied to the sequence data could give us the ability to predict important strain characteristics directly from sequence, including serotype, antimicrobial resistance and Shiga toxin type, making that information available early in an investigation. Finally, building a large database of sequences is a critical step to developing the metagenomic tools for public health that we will need to have in the future.

How do we determine that a food is the source of an outbreak of illness? We assemble three main types of information to make that judgement. First, we use epidemiologic evidence, showing that the persons who are infected with the same strain report consuming a particular food at a far higher frequency than would be expected. Second, traceback information can help to show that the suspect food item eaten in several locations came from the same source, suggesting where contamination might have been introduced. Third, microbiological assessment of that food, facility or other source may show that the same pathogens is found in that food, facility or farm. Having information of all three types is ideal, two types of information can together be enough to take action.

These types of evidence are evolving and are in transition. For example, the new culture independent clinical diagnostic tools mean that some infections that were formerly rarely reported are now being identified more frequently, such as those caused by *Cyclospora*, *Vibrio*, *Yersinia* or enterotoxigenic *E. coli*. Traceback is improving, using digital supply chain systems that industry is developing to identify sources more quickly and definitively. Traceability is accountability, a critical part of the whole food safety system. And then there

is the transition from the over 20 year old PFGE method for subtyping to whole genome sequencing for routine subtyping in PulseNet.

Starting in 2013, we started a collaborative pilot effort to apply WGS to infections with *Listeria monocytogenes*, the cause of rare but severe infections in the elderly, immunocompromised or pregnant individuals. We at CDC tested all the isolates from clinical sources, approximately 800 per year, while colleagues at FDA and USDA/FSIS tested isolates from foods, food animals and other sources, with a new partner, the National Center for Bioinformatic Information serving as a publicly available repository for DNA sequence information. After three first years, the results were striking. The number of clusters under investigation increased, the average size of those clusters decreased, and the number for of clusters that were solved, i.e. for which the source was determined jumped three-fold. When PGFE itself was introduced, the number of listeriosis outbreaks detected and solved had increased from 2 per decade to 2 per year, and now with WGS, the number increased again to 6 per year. These outbreaks were traced to familiar sources like soft cheeses and sprouts, and to new and previously unsuspected risks, like caramel apples, stone fruits and ice cream, leading to new prevention efforts in those sectors. During this pilot effort, we saw that events were not always a group of highly related cases clustered tightly in space and time. Our appreciation for what an outbreak is began to broaden.

For example, one large outbreak detected in 2015 was caused by *Listeria monocytogenes* strains with five different PFGE patterns but shown by WGS to be highly related. Had we been relying on PFGE to detect the outbreak and guide its investigation, we likely would not have realized it was a cluster at all, and not launched an investigation. The final count was 30 human infections in residents of 10 states, of which 28 were hospitalized and 3 died, along with one fetal loss. Of those interviewed, 75% reported eating a soft cheese of Middle Eastern, Mediterranean or Hispanic style, that was traced back to one cheese plant in California, where the same strain was found. This led to a voluntary recall of the cheeses from that plant. Interestingly, review of WGS found that human infections from closely related strains were identified as far back as 2010, growing slowly in number to 2015, and matching a strain isolated at the same plant in 2010. This outbreak may thus have been the result of in-plant harborage lasting for years, that had now been detected and stopped.

The experience was not limited to *Listeria*. In 2015, the Minnesota state public health laboratory began WGS typing of *Salmonella* serotype Enteritidis (SE), a serotype that has been difficult to resolve with PFGE, which provides only a few main subtypes. That summer, within months of starting to type SE, they identified two separate outbreaks related to two similar products that were frozen stuffed breaded raw chicken products. Each was caused by two different PFGE types, but WGS showed that the strains in each outbreak had closely related sequences, and matching strains were found in the two brands of product. Again, relying on PFGE probably would have meant that these outbreaks would not have been detected at all. Though the number of cases identified in Minnesota was small, five in one outbreak and seven in the other, the actual outbreaks were probably much larger; both implicated products were distributed to many states, and both were recalled nationwide. We don’t really know how long these outbreaks had been going on. Problems like this have been identified before, but these two outbreaks in one summer underlined the ongoing risk of this

particular product type. Efforts had been made to improve the consumer instructions on the packaging, so it was interesting that many people who became ill told the investigators that they read and followed the directions for cooking, and some used a meat thermometer to be sure it was cooked thoroughly. The event raised concern that the tactile clues provided by a dry, breaded and browned product may override instructions about desired food safety behaviors always needed after handling raw chicken products.

The following year, a second state health department began evaluating WGS for *Salmonella* Enteritidis. In 2016 Tennessee found a small cluster of 6 closely related infections, who all ate at a restaurant where steak with raw egg Bearnaise sauce was the exposure they had in common. The eggs came from a small local farm, with fewer than 3000 hens; sampling on the farm did not yield SE. A month later a second outbreak occurred due to a strain of SE that was 3 single nucleotide polymorphisms different from the first outbreak cluster (the technical term I use is “DNI or darned near identical”). These ill persons had eaten at a second restaurant, making raw egg mayo using eggs from the same small farm. This time reinvestigation of the farm identified SE in the chicken litter, and more specific advice and control measures were instituted. Again, WGS helped to find and link these small outbreaks and point to a gap in our food safety system, as the 2010 egg regulation applies only to flocks that have at least 3000 hens.

This showed us that subtyping based on genetic relationships can give us better separation of truly related infections from those that are not improving our definitions of case vs non-case. While PFGE could only say “same” vs “different”, WGS provides a scale of genetic difference. Some very common PFGE types can be divided into meaningful subtypes, as may often be the case with *Salmonella* Enteritidis. Some WGS clusters of closely related strains may bring together a variety of PFGE patterns, that appear different because variable plasmid content masks a fundamental genetic similarity. Applied across human, food, animal and environmental isolates, it can suggest and confirm relationships with more confidence than PFGE did.

Starting in 2017, we began preparing the transition to WGS in PulseNet. With support from several funding sources including the Initiative to Combat Antibiotic Resistant Bacteria to support better surveillance for antibiotic resistance, and the Advanced Molecular Detection program we began to expand our information backbone, and to equip, train and certify state health department laboratories in this new method. We partnered closely with the USDA/FSIS, and with the FDA’s Genome TrakR program that was equipping food laboratories in many states with WGS. In 2018, many states began sequencing their *Listeria* strains and on July 15, 2019, we made the transition to WGS as the standard PulseNet method for *E. coli* O157 and *Salmonella*. As we saw in the *Listeria* pilot program, we now expect to find more clusters of all these pathogens that are truly genetically related, leading to more successful epidemiological investigations, and defining more targets for improved prevention. Anticipating the need for more epidemiological investigators in state health departments, we have also been supporting more capacity for investigation in many state health departments around the country, so that there are staff to handle the growing number of clusters of illness identified.

It is important to understand that just showing that clinical strains have very similar sequences to a strain from a food, or that two strains separated by several years are closely related genetically does not by itself prove that they are from the same source. We still need to investigate what foods the patients ate, looking for unexpected similarities in exposures, and tracing suspect foods back to their sources. There are many dotted lines that need to be filled in, and investigations are not always successful. The genetic similarity certainly provides a strong clue, or hypothesis, to the investigators, but those investigations are critical to solving the outbreak.

It is important to be alert for multiclonal outbreaks. Not every outbreak is caused by a single strain or subtype. A product may be heavily contaminated with many different pathogens. For example, last year we investigated an outbreak of at least 199 infections linked to consuming kratom, the leaves of a South East Asian shrub in the coffee family, used for its opioid-like effects. People buy it at smoke or vape shops in many states, or online. The samples tested by FDA yielded *Salmonella* with 85 different WGS profiles, and strains in 6 serotypes matched strains from ill persons, who had been using kratom. There seem to be major flaws in kratom processing, that somehow are leading to extraordinarily high contamination with many strains in this imported product.

It is also important to remember that it is not always food. We have every year many infections with enteric pathogens that are related to backyard chickens, pet reptiles, and other animal exposures. When we approach a cluster genetically related infections, we must consider a range of hypotheses beyond the things that they may have eaten.

This greater confidence in taxonomic genetic relationships provided by WGS means that the definition of what an outbreak is seems to be broadening. We can see beyond the classic cluster focused tightly in place and time, to find broader events caused by taxonomically related groups, that last for years rather than just weeks or months, that may signal a long-lasting harborage site somewhere in the production system. We can discern the spread of a group of related strains that emerges in a food animal species or a produce growing area, that can manifest as repeated or sustained events affecting many herds, flocks or fields at once. To describe these broader groups, I borrow a taxonomic term, “clade”, derived from the Greek word *klados* for “branch”, that geneticists already use to describe broad groups of related organisms with more precision.

Here are two examples of recurring or emerging “clades” of concern. In the spring of 2018, a large outbreak of *E. coli* O157 infections was traced to Romaine lettuce. This outbreak was the largest O157 outbreak in a decade, with 210 confirmed cases in 36 states, leading to 96 hospitalizations and 5 deaths. It started in one state as a local cluster of infections related to several outlets of a restaurant chain, and then was quickly recognized across the country. Thanks to rapid investigation by public health epidemiologists in many states, and great efforts to trace suspect lettuce back, it was linked to Romaine lettuce from the Yuma growing area. One noteworthy investigation occurred in Nome, Alaska, where an outbreak in the local jail was linked to whole head Romaine, that could be traced to a single farm; while products implicated in other clusters were mixed chopped and bagged Romaine from multiple farms. In the end the strongest tracebacks led to approximately 23 farms,

spread out across 50 miles, separated by many apparently unaffected farms. This suggested a widespread though patchy contamination event occurred. It ended after repeated warnings and the end of harvest. WGS brought together 22 different PFGE patterns among the cases into two main clades, both of which we found in the freely flowing water in the irrigation canal that supplied the area. The largest clade had caused two outbreaks the preceding year, one among persons swimming in a small lake in California, 500 miles away, and another in a group who became ill after exposure to a salad bar in the mid-West. This recurrent clade needs heightened surveillance, new prevention measures and research into where and how it persists in the environment. We are following this clade of concern throughout 2019, ready to rapidly investigate any clusters that may occur. In April this year, the Leafy Green Marketing Agreement of California introduced a new requirement for farmers to sanitize surface water before they are sprayed on leafy greens for irrigation or aerial application in the three weeks before harvest, an important step forward in prevention.

A second example is an emerging clade of *Salmonella* Infantis, whose multiple PFGE types are unified by WGS. This clade is resistant to or has decreased susceptibility to 10 antimicrobials, including ceftriaxone and ciprofloxacin, the two first line agents for treating severe salmonellosis. The first strains appeared in 2012, among travelers returning from Peru. The first non-travel associated case was seen in 2014, and the numbers increased rapidly in 2017–2018. The first isolate found by FSIS in chickens was in 2013, and those also increased rapidly in 2017. After one single PFGE type within this clade surged in 2018, we focused investigation on that one type; 89% of those interviewed had eaten or prepared raw chicken products, of many different types, brands, and purchased from many different stores. The same PFGE pattern was found in strains from chickens at slaughter and chicken meat from many different processors. This multi-drug resistant strain has appeared in chickens from many farms, upstream from the processing sites. Although the PFGE type that increased in 2018 has declined, the broader emerging clade still persists in humans and chickens. The emergence of this clade was discussed with the chicken industry, along with the need for industry wide approaches to prevention. Again, more detailed preharvest investigations and interventions are needed to improve prevention.

In broad terms, these “clades of concern” revealed by WGS are groups of closely related strains that cause repeated outbreaks, persist in specific reservoirs or geographic areas, and emerge in people consuming specific commodities. They may be multidrug resistant and difficult to treat, which adds to the concern. To prevent these infections, we collectively need sustained surveillance, better understanding of their persistence, and investigation of the root causes leading to contamination and spread. Commodity-specific prevention strategies are going to be critical, whether it be treating the water that is sprayed onto leaves, vaccinating food animals, or other measures. I am confident that prevention can be improved, if attention is focused on the problem at hand.

To help target the more general interventions, by developing general estimates of the fraction of foodborne infections that can be attributed to individual food commodities, an interagency work group called the Interagency Food Safety Analytics Consortium (IFSAC) has been analyzing the surveillance information available. IFSAC members have constructed a model based on foodborne outbreaks reported over the last 18 years, giving more weight to the

most recent 5 years, and parsing the number of illnesses across 17 major food categories, for four major pathogens. The most recent summary is based on outbreaks from 1988 – 2016, and the plan is to update this annually. Thus, the 2016 IFSAC estimates for *Salmonella* attribute 19% of foodborne salmonellosis to seeded vegetables, like tomatoes, cucumbers and peppers, 13% to chicken, 11% to pork, and 10% to fruits, and 7–9% to other produce, eggs, and beef. The 2016 estimates for foodborne *E. coli* O157 infections differ, attributing 43% to vegetable row crops like leafy greens, 30% to beef, and 8% to dairy products. These estimates are helping guide the priorities to prevention efforts for these and other pathogens.

Now, whole genome sequencing offers new tools to help improve the targeting of foodborne disease prevention efforts more generally. As I have shown you, we can define broad clades of concern among human pathogens that are emerging, persisting or recurring. We can examine the frequency of those same clades in isolates from meats, poultry and food animals and in the more limited libraries of isolates from other foods and environmental sources. IFSAC is now taking up the effort of refining the attribution estimates by clade within the main pathogens. With international collaboration, we can be alert for the emergence of clades in other countries and find them quickly when they appear here.

Because microbiology never stays still, new tools for clinical laboratories are changing diagnostic practices in human medicine. Since 2015, the use of rapid multi-pathogen diagnostic panels has been increasing, providing proprietary PCR-based detection of many pathogens within a few hours. The result is that more people are being tested for more pathogens, including some that could not be routinely detected before, and fewer are culture-confirmed. This means that we now need to account in our surveillance for the changes in diagnostic testing taking place in clinical laboratories. These tests do not yield a living bacterial isolate, unless the specimen that was positive in the test is then cultured for that organism. Access to the isolate is necessary for PulseNet subtyping, so if a clinical laboratory decides not to perform these “reflex” cultures, they may send the positive specimen to the public health laboratories to be cultured there. This challenge of the culture-independent diagnostic test is part of the changing landscape of food safety.

So how do we in public health stay on the cutting edge as microbiological and diagnostic methods march forward? As big an advance as it is, whole genome sequencing still requires an isolate, and still can take upwards of two weeks to turn results around. In the future, public health is going to need more advanced molecular methods for direct use on clinical specimens to get the same sequence information in hours. Clinical researchers are exploring metagenomic methods now, that would provide identification and sequences for all organisms represented in the sample, and work has begun at CDC as well. The rapidly expanding WGS databases of pathogens are a bridge to this post-isolate future. Five to ten years from now, we may see diagnosis and pathogen sequencing using metagenomic methods providing results in close to real time and moving from the laboratory to the patient’s bedside.

As we contemplate how to make food safer in the 21st century, I hope that you will agree that whole genome sequence-based surveillance is an important evolutionary step forward. This approach offers more precise subtyping, which combined with better capacity for

patient interviews and faster tracebacks will mean that more food safety gaps can be found and corrected. The methods allow us to track antibiotic resistance, serotype and virulence profiles in close to real time as well. These methods can be applied to many other infections as well as to enteric bacteria. They are also helping us to define broader clades of concern that emerge, recur and persist, offering new targets for more prevention earlier in the food supply chain. These methods can help us refine our estimates of attribution to better target broad prevention strategies, and they provide a bridge to the future when public health will have culture-independent tools providing sequence information rapidly. We anticipate that in the longer term, the effect will be to help industry, regulators, researchers and consumers drive down the number of outbreaks and the incidence of foodborne infections. Thanks to all of you for the roles that you play in food safety, and for staying to the end of the conference.