

Whole-Genome Sequencing Is Taking over Foodborne Disease Surveillance

Public health microbiology is undergoing its biggest change in a generation, replacing traditional methods with whole-genome sequencing

Heather A. Carleton and Peter Gerner-Smith

About 1 in 6 people are sickened with foodborne diseases each year, and for the most part those illness bouts are a nuisance and self-limiting. However, for some vulnerable populations, severe foodborne illnesses can require hospital care and may even lead to death. Once an individual becomes sick enough to visit a physician, he or she typically collects a stool sample to send to a clinical microbiology laboratory for testing and diagnosis. If the clinical laboratory identifies an enteric pathogen, the physician is notified.

Clinical laboratories will usually also send bacterial cultures or samples to local public health laboratories. What goes on there has no direct consequence for patients and their doctors unless a particular patient is suspected to be part of an outbreak. In the public health laboratory, the isolate will be characterized and subtyped in the PulseNet system, the US national network that is the primary early warning laboratory system for foodborne outbreaks. Patients who are involved in outbreaks typically are contacted by an epidemiologist asking about what foods might be responsible for the outbreak.

The PulseNet testing method to screen samples changed very little until recently, when whole-genome sequencing (WGS) began to replace pulsed-field gel electrophoresis (PFGE), which was used almost exclusively for the past 20 years. This network is coordinated by a team of microbiologists at the Centers for Disease Control and Prevention (CDC) in Atlanta, Ga., who work closely with microbiologists in more than 80 local, state, and federal public health and food regulatory laboratories. One of their main tasks is to collect molecular data characterizing foodborne bacteria that infect patients along with demographic data omitting personal identifiers, and then to submit that information to a national database.

Similar data characterizing microorganisms isolated from foods and food production facilities are also submitted to PulseNet. To detect outbreaks, microbiologists at the CDC as well as state and local health departments look for trends in these data. Are a higher number of similar molecular fingerprints being generated in a particular region of the US than at the same time during past years? Has the fingerprint never been seen in the PulseNet database except for the last month or so? By analyzing these types of clues, the PulseNet teams can determine whether the bacteria being isolated are linked to outbreaks, and then they can communicate this information to epidemiologists who lead investigations of suspected outbreaks.

Public Health Microbiology and Molecular Surveillance Workflows

Microbiologists working at public health laboratories typically will identify the genus and species

SUMMARY

- Whole-genome sequencing (WGS) is beginning to replace pulsed-gel electrophoresis (PFGE) for subtyping of foodborne pathogens from stools and other specimens for outbreak surveillance.
- For more than 20 years, PFGE, a molecular fingerprinting technique that can be adapted to determine the subtype of almost any bacteria, was the principal method for detecting and investigating foodborne disease outbreaks in the United States.
- Investigators at PulseNet are working with other investigators in and outside the United States (US) from public health, food regulatory laboratories, and universities to build standard WGS multilocus sequence typing (MLST) databases for analyzing common foodborne pathogens.
- The development of metagenomic sequencing-based tools combining diagnostics and surveillance, a challenging task, could soon enable public health investigators to detect outbreaks much earlier and likely shortly after the first patients begin visiting their physicians.

of the bacteria being tested. However, they may also further characterize the samples they are analyzing, depending on the pathogen in question. For example, they may serotype *Salmonella*, Shiga toxin-producing *Escherichia coli* (STEC), *Shigella*, *Vibrio*, and *Listeria*; characterize virulence determinants carried by diarrheagenic *E. coli* pathotypes (STEC, enteropathogenic *E. coli* [EPEC], enterotoxigenic *E. coli* [ETEC], enteroaggregative *E. coli* [EAEC], enteroinvasive *E. coli* [EIEC]), and *Shigella*; or *Vibrio* spp., and may determine the antimicrobial susceptibilities of isolates.

The assays used to characterize such samples vary in complexity and can include a large spectrum of phenotypic and molecular tests. They include observing growth of the bacteria on different types of media, fermentation and biochemical reaction tests, agglutination with diagnostic antisera, immunofluorescence, cell culture assays, protein electrophoretic assays, and PCR. Each isolate typically is characterized using multiple independent assays. If the isolated pathogen is a *Salmonella*, STEC, *Shigella*, *Listeria*, *Vibrio*, or *Campylobacter*, it likely will also be subtyped with a highly discriminatory DNA fingerprinting method to determine if it could be part of an outbreak.

For more than 20 years, PFGE, a molecular fingerprinting technique, was the principal method for detecting and investigating foodborne disease outbreaks in the United States. PFGE is the only restriction fragment length polymorphism (RFLP) procedure from the 1970s and 1980s still in wide use. It entails the use of restriction enzymes to cut bacterial genomes into 10–30 large pieces (10–500 kb) that are separated in agarose gels that are exposed to alternating, or pulsed, electric fields under conditions that separate the fragments according to their size. The banding pattern provides a characteristic pattern, or “fingerprint,” for each bacterial strain that the microbiologist compares to those for other isolates to detect whether an outbreak is going on.

This method survived so long because it can be adapted to subtype almost any bacteria and standardized to compare results from different laboratories, and proved highly efficient in detecting and investigating outbreaks. Until WGS was introduced, no other subtyping method had all these characteristics. Moreover, when PFGE does not sufficiently discriminate between isolates, the method can be supplemented with other patho-

gen-specific assays. Most recently, multilocus variable number of tandem repeats analysis was used for outbreaks involving pathogens such as *E. coli* O157 and *Salmonella enterica* serovars Typhimurium and Enteritidis.

With PFGE still in wide use, public health laboratories rely on several complex, dated, and therefore expensive pathogen-specific methods that take specialized expertise to perform and interpret. Additionally, the turnaround times for characterizing foodborne pathogens in many public health laboratories range from four days to several weeks or months, depending on the workflow for a particular pathogen. Therefore, a method that could simplify and accelerate such testing is highly desirable, especially if it is cost-efficient and could replace these older technologies. WGS has the potential to do exactly that (Fig. 1).

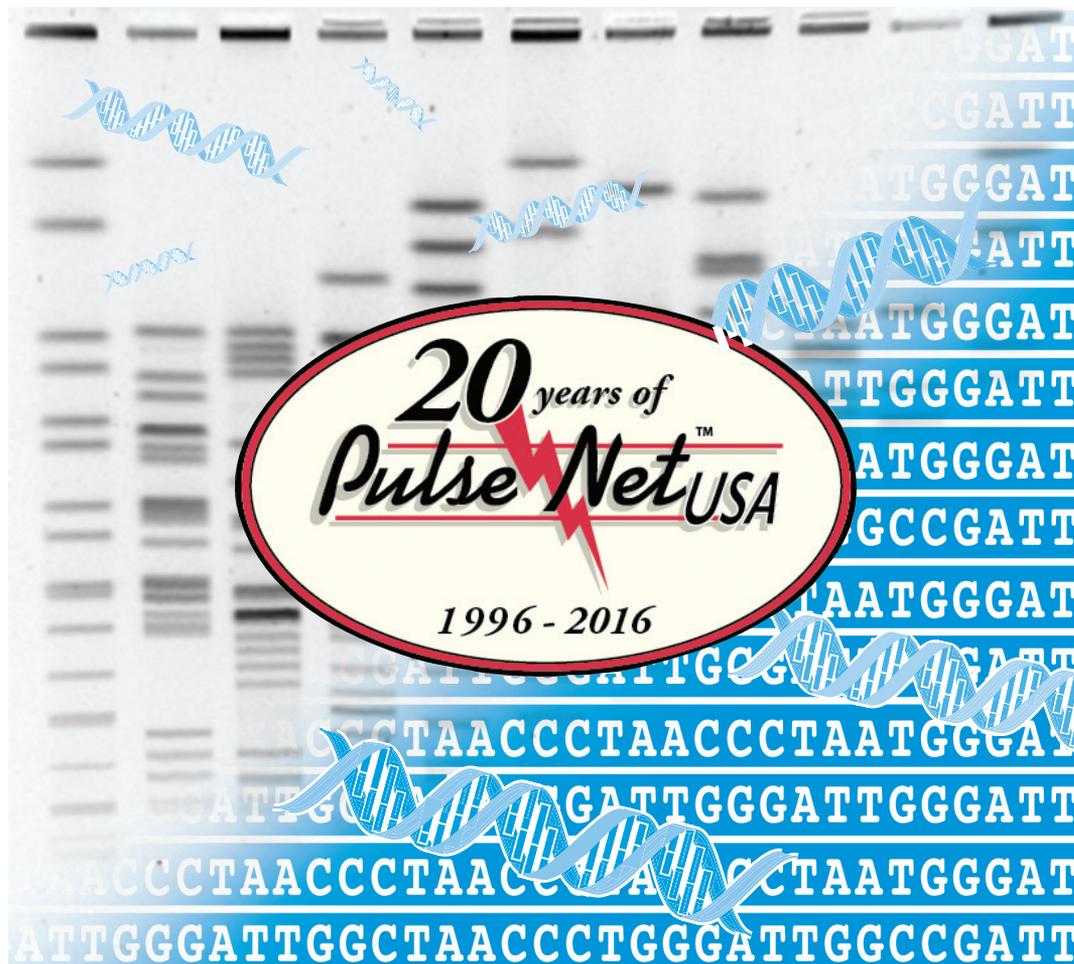
Public Health Microbiology Embracing Whole-Genome Sequencing

Next-generation sequencing (NGS) technology is drastically reducing the cost and time needed to sequence bacterial genomes, making this analytic approach feasible for both reference and subtyping purposes at public health laboratories. Instead of relying on multiple workflows to identify pathogens and their serotypes, virulence factors, antimicrobial resistance factors, and molecular fingerprint on pulse gels, much of this information can be extracted from WGS data. Additionally, NGS can reduce turnaround times to a mere 2–4 days.

The genetics underlying many phenotypic tests of bacterial pathogens are known, and PCR assays already replaced many of these tests. WGS can easily, in turn, replace PCR or those older tests. Indeed, many WGS-based analyses are already freely available to the scientific community and could be made even more useful if applied to public health.

For example, the Center for Genomic Epidemiology at the Danish Technical University (<https://cge.cbs.dtu.dk/services/>) is a particularly good source for several such tools, including tools to detect antimicrobial resistance (ResFinder) and virulence genes in *E. coli* (VirulenceFinder), to determine serotype of *E. coli* (SerotypeFinder), and to characterize plasmids in *Enterobacteriaceae* (PlasmidFinder). The University of Georgia

FIGURE 1



PulseNet, the national molecular subtyping network for foodborne disease surveillance, is replacing older subtyping methods like pulsed field gel electrophoresis (PFGE) with whole-genome sequencing (WGS).

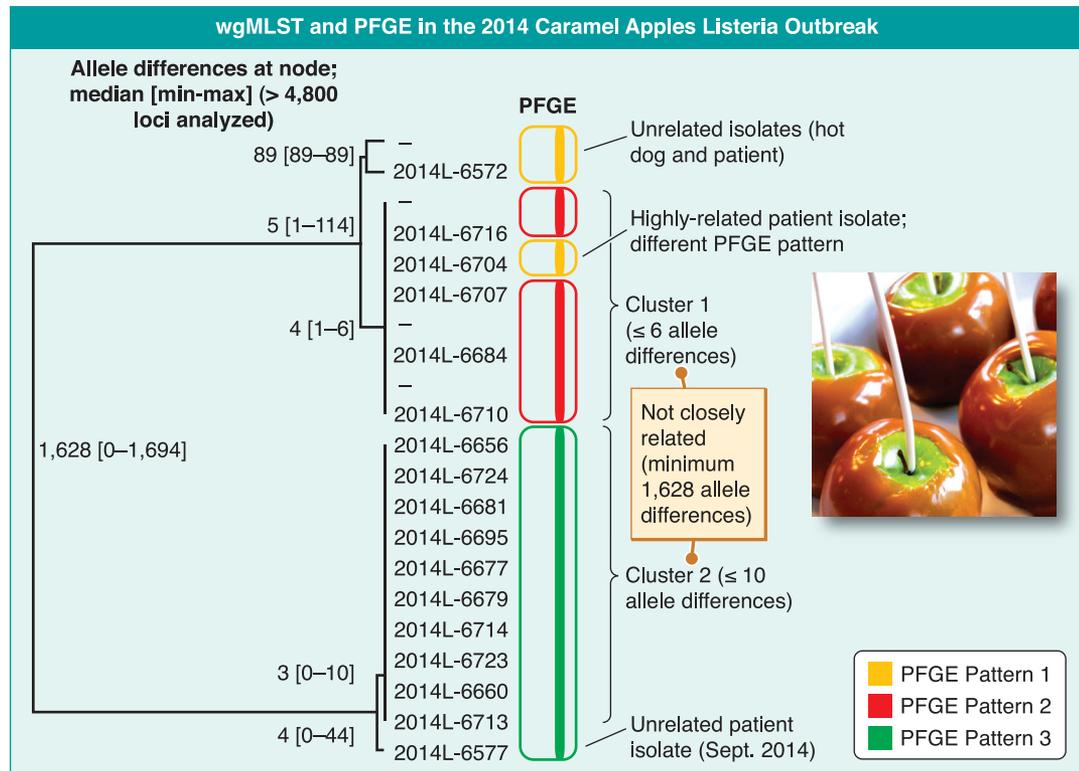
hosts a valuable tool for serotyping of *Salmonella* (<http://www.denglab.info/SeqSero>).

The serotyping tools for *E. coli* and *Salmonella* determine the serotype from genes that encode the O and H antigens from assemblies or raw sequence data, and the serotypes are therefore with few exceptions fully compatible with the existing serotyping schemes and have the advantage of being able to type rough isolates that are untypeable by traditional agglutination tests. However, the drawback of using these Web services is that the user can use only one tool at a time, even though sequences of multiple isolates may be batched. The number of different traditional tests will be replaced by the same number of queries of the tools.

For subtyping, WGS-based approaches provide better resolution to identify relatedness of isolates during an outbreak than almost any other method, including PFGE (Fig. 2). Rather than comparing isolates by their pattern of 15–30 differently sized bands, isolates can be compared across millions of base pairs by doing comparisons of single-nucleotide polymorphisms (SNPs) or gene-by-gene comparisons, such as whole-genome multilocus sequence typing (wgMLST).

Both analytical approaches are being used for investigating outbreaks and seem to be equally useful. Numerous SNP analyses pipelines are available in the public domain but require bioinformatics expertise to perform and generally require a priori knowledge about the isolates being

FIGURE 2



The subtyping utility of WGS. wgMLST and PFGE in the 2014 caramel apples *Listeria* outbreak. Improved resolution of WGS over PFGE for outbreak investigations. The figure shows WGS-based similarity tree of 18 outbreak-related isolates and three isolates unrelated to the outbreak that are indistinguishable by PFGE from the outbreak isolates. Three PFGE patterns are illustrated by the colored bars to the right of the tree. By sequencing, two PFGE patterns (in red and yellow) appear to be related to each other by WGS and outbreak related; the wgMLST differences are listed per branch as median [range]. The branch that contains ≤ 6 alleles difference between isolates (cluster 1) contains 5 patient isolates (denoted by number) and 3 food/environmental isolates (denoted by dash-). This cluster was distinct from the unrelated isolates at the top of the tree by 114 allele differences, though the isolates shared a common PFGE pattern. Cluster 2 (in green) contained 10 isolates associated with the same source as Cluster 1 but a different PFGE pattern. One isolate that was collected during the same time period as cluster 2 and was the same PFGE pattern, but was clearly distinct by WGS and shown to be unrelated to the outbreak. The food and environmental isolates were sequenced by FDA and the FDA sponsored GenomeTrakr network.

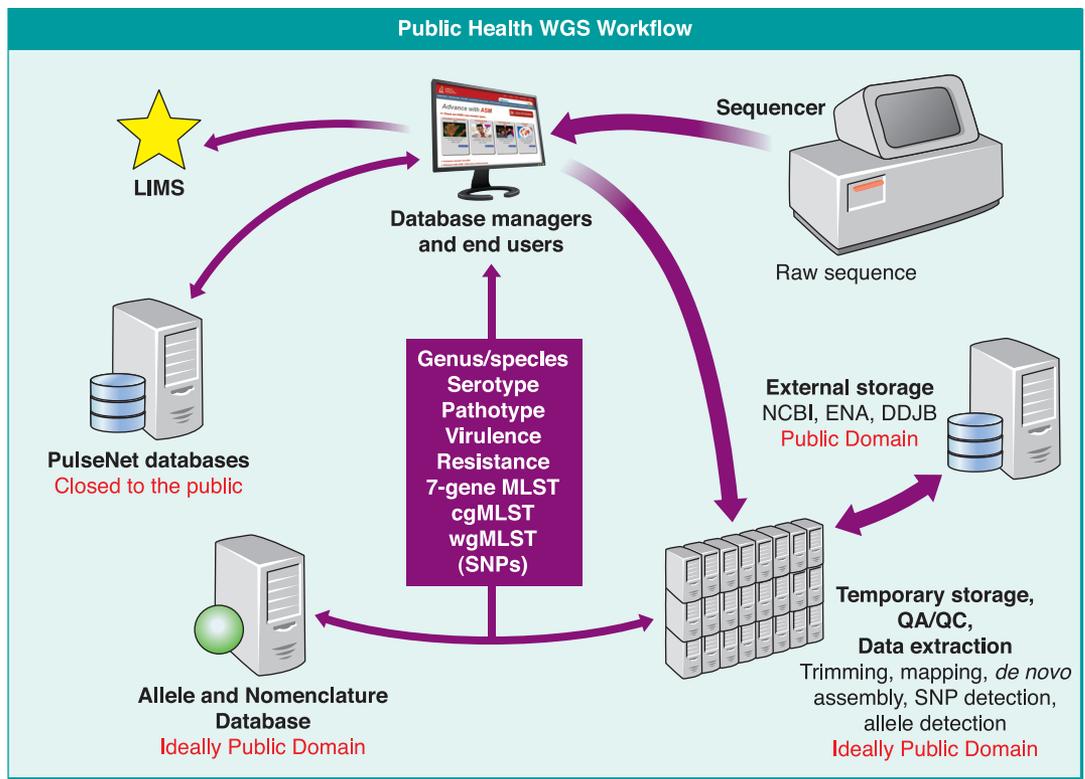
sequenced. This requirement does not hold for wgMLST because it relies on a database of all genes, or loci, generated from multiple, diverse reference genomes. These wgMLST databases are built to provide maximum discrimination for all isolates of a given genus or species. Isolates from different outbreaks of the same species can easily be compared using wgMLST at variance with SNP analysis, which is reliable only in a narrow phylogenetic context. Isolates from different outbreaks that were investigated using different SNP reference strains cannot readily be compared. For these reasons, the gene-by-gene approach is the

clear winner for national and international surveillance of foodborne pathogens.

In the United States alone, more than 60,000 enteric bacterial isolates are analyzed at local, state, and federal public health laboratories each year. Because quick turnaround times are needed to detect and investigate outbreaks, any tool developed to analyze WGS data must have a simple workflow to meet both reference and outbreak surveillance analysis needs, meaning it must be easy to operate for a public health microbiologist with little to no bioinformatics expertise.

Because the use of several different analytical

FIGURE 3



Model for whole-genome sequencing workflow for reference characterization of foodborne pathogens in public health laboratories. Arrows indicate the direction of the data flow and the thickness of the arrows the amount of data exchanged; raw sequence data (thick arrows) versus sequence derived data (thin arrows). All sequences are uploaded to a public repository from a temporary storage site following data cleaning, data extraction, and quality assessment. The alleles are called and the wgMLST profiles named through comparison with an allele and nomenclature database in public domain before being transferred to the end-user and public health database along with other WGS-derived information about species identification, serotype, antimicrobial resistance determinants, etc. In the United States a national database is housed at CDC and local databases in each PulseNet participating laboratory.

tools separately is inefficient, it is far better to include all tools in a single analytical workflow for both reference characterizations and subtyping. Such a system must freely import and export data to other systems, including surveillance databases and laboratory information management systems. To meet the needs of public health laboratories, typically such versatile systems either are built in-house from scratch or by combining different databases with analytical software programs because versatile commercial software packages are scarce or not fully adapted for use with WGS. However, one software package, BioNumerics, which is marketed by Applied Maths of Austin, Tex., includes both database and advanced analytical functionality including

WGS analyses capabilities in its latest edition (v7.5).

Investigators at PulseNet are working with other investigators in and outside the United States (US) from public health, food regulatory laboratories, and universities to build standard wgMLST databases for analyzing common foodborne pathogens (Fig. 3). Tools that identify species on the basis of average nucleotide identity (ANI), serotyping, virulence, and antimicrobial resistance factors are also being incorporated. These analytic packages are being developed in a tiered fashion, beginning with *Listeria monocytogenes*, and it will be made available in all public health laboratories in the US with sequencing capacity by the summer of 2016. Comparable

packages for identifying *Campylobacter*, STEC, and *Clostridium botulinum* are expected to follow later in 2016, and those for *Salmonella* in 2017, other diarrheagenic *E. coli* and *Shigella* and *Vibrio* in 2018, and for *Yersinia enterocolitica* and *Cronobacter* in 2019.

Global partners are also developing quality standards for raw DNA sequences and proficiency testing standards to ensure that anyone using these analytic systems can produce high-quality results that may be compared reliably on national, regional, and global scales. In 2013 PulseNet and other US public health labs began using WGS routinely for surveillance of listeriosis; this approach has led to the detection of more outbreaks, and more outbreaks have been solved. Even though the incidence of foodborne illnesses is not increasing, the technology will likely lead officials to recognize many more outbreaks caused by foodborne pathogens. This anticipated increase in reported outbreaks could prove a challenge for those states whose public health departments are not prepared to take on larger workloads.

The Future of Public Health Microbiology

Clinical laboratories increasingly are relying on multiplexed molecular panels to test stool specimens for enteric pathogens, determining within a matter of hours bacterial, viral, and parasitic pathogens with high sensitivity and specificity. To clinicians, these tests are a huge step forward because they can detect pathogens that previously rarely were looked for, including diarrheagenic *E. coli* in addition to STEC, viruses, and parasites. Thus, they provide actionable results and help with the management of patients.

However, because such tests do not require the culturing of microorganisms, those labs are no longer setting isolates aside for surveillance. To address this issue, measures are being put in place to maintain the flow of isolates from positive tests at clinical or public health laboratories now using those culture-independent testing methods. However, this approach will not be sustainable in the long-term. Hence, new metagenomic sequencing-based pathogen detection and subtyping tools are being developed to characterize stool specimens.

The development of these metagenome sequencing-based surveillance tools is not a trivial task. Any approach needs to be low cost to stay affordable for public health labs to use and must

also provide epidemiologically meaningful subtyping capabilities during disease outbreaks. Meeting this latter need can prove challenging when identifying pathogens in the context of the normal enteric flora, which contains many microorganisms that closely resemble enteric pathogens. Data from WGS of pure cultures will be critical to develop these metagenomics surveillance tests, as well other advances in sequencing and bioinformatics technology.

As these technologies advance, the day will come when individuals with foodborne diseases will visit their physicians, who will perform WGS on stool samples using instruments that directly plug into laptops or smartphones. Once sequencing is completed, the patients and their physicians will know not only which pathogen caused a particular illness but will also have detailed information about what virulence and antimicrobial resistance factors it carries. Finally, those analytic results will also automatically be submitted to local public health authorities to determine whether individual patients are part of larger outbreaks. This way, public health investigators will be able to detect outbreaks soon after the first patients become sick—much faster than the 1–2 weeks required with current culture-based detection technologies.

Heather A. Carleton is Leader of the Bioinformatics Team and Peter Gerner-Smidt is Branch Chief in the Enteric Diseases Laboratory Branch, Centers for Disease Control & Prevention, Atlanta, Ga.

Disclaimers

The findings and conclusions in this feature are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. Use of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention or by the U.S. Department of Health and Human Services.

Suggested Reading

- CDC PulseNet website.** <http://www.cdc.gov/pulsenet>
CDC AMD website. <http://www.cdc.gov/amd/>
European Centre for Disease Prevention and Control. Expert Opinion on the introduction of next-generation typing methods for food- and waterborne diseases in the EU and EEA. Stockholm: ECDC; 2015. http://ecdc.europa.eu/en/publications/_layouts/form/Publication_DispForm.aspx?List=4f55ad51-4aed-4d32-b960-af70113dbb90&ID=1389

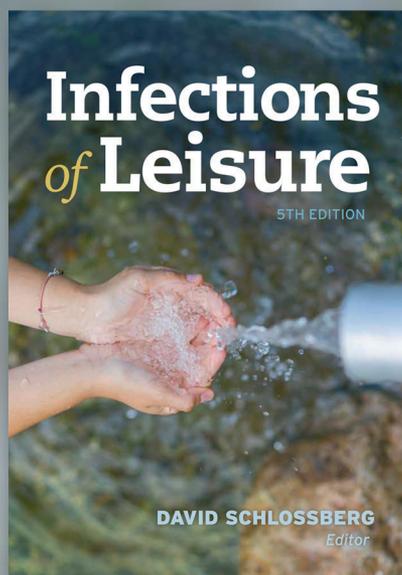
Jackson BR, Tarr C, Strain E, Jackson KA, Conrad A, Carleton H, Katz LS, Stroika S, Gould LH, Mody RK, Silk BJ, Beal J, Chen Y, Timme R, Doyle M, Fields A, Wise M, Tillman G, Defibaugh-Chavez S, Kucerova Z, Sabol A, Roache K, Trees E, Simmons M, Wasilenko J, Kubota K, Pouseele H, Klimke W, Besser J, Brown E, Allard M, Gerner-Smidt P. Implementation of nationwide real-time whole-genome sequencing to enhance listeriosis outbreak detection and investigation. *Clin. Infect. Dis.* 2016 Apr 18. pii: ciw242.

Maiden MC, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, et al. 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nature Rev. Microbiol.* 11:728–736.

Trees E, Rota PA, MacCannell D, Gerner-Smidt P. 2015. Chapter 10: Molecular Epidemiology, p. 131–160. In Jorgensen JH, Pfaller MA, Carroll KC, Funke G, Landry ML, Richter SS, et al. (ed.), *Manual of clinical microbiology*, 11th ed. ASM Press, Washington, DC.

Infections *of* Leisure

5TH EDITION



A DAY AT THE BEACH: DELIGHTFUL, RESTORATIVE, AND POTENTIALLY DANGEROUS.

Editor: David Schlossberg, MD, FACP; Temple University School of Medicine; University of Pennsylvania School of Medicine; Philadelphia Department of Public Health.

“A unique and very readable collection of infectious and non-infectious diseases associated with travel, geography, activity and behavior...a great text that is easy to navigate for microbiologists and clinicians interested in internal medicine, sports medicine, travel medicine and infectious disease.”

David W. Craft PhD, Medical Director of Microbiology,
Penn State Milton S. Hershey Medical Center

LIST PRICE: \$80 | ASM MEMBER PRICE: \$64

Feb 2016. 411 pages. Full-color illustrations,
Index. Paperback or eBook

©2016 ASM Press

When you buy directly from ASMscience.org, you support the society that supports the microbial sciences.

www.ASMscience.org | books@asmusa.org

