CURATED COLLECTION

Advances in Molecular Epidemiology of Infectious Diseases

Microbiology Spectrum
American Society for Microbiology Press

Laboratory Methods in Molecular Epidemiology: Bacterial Infections*

LEE W. RILEY[1]

[1]Division of Infectious Diseases and Vaccinology, School of Public Health, University of California, Berkeley, CA 94720

ABSTRACT In infectious disease epidemiology, the laboratory plays a critical role in diagnosis, outbreak investigations, surveillance, and characterizing biologic properties of microbes associated with their transmissibility, resistance to anti-infectives, and pathogenesis. The laboratory can inform and refine epidemiologic study design and data analyses. In public health, the laboratory functions to assess effect of an intervention. In addition to research laboratories, the new-generation molecular microbiology technology has been adapted into clinical and public health laboratories to simplify, accelerate, and make precise detection and identification of infectious disease pathogens. This technology is also being applied to subtype microbes to conduct investigations that advance our knowledge of epidemiology of old and emerging infectious diseases. Because of the recent explosive progress in molecular microbiology technology and the vast amount of data generated from the applications of this technology, this *Microbiology Spectrum* Curated Collection: Advances in Molecular Epidemiology of Infectious Diseases describes these methods separately for bacteria, viruses, and parasites. This review discusses past and current advancements made in laboratory methods used to conduct epidemiologic studies of bacterial infections. It describes methods used to subtype bacterial organisms based on molecular microbiology techniques, following a discussion on what is meant by bacterial "species" and "clones." Discussions on past and new genotyping tests applied to epidemiologic investigations focus on tests that compare electrophoretic band patterns, hybridization matrices, and nucleic acid sequences. Applications of these genotyping tests to address epidemiologic issues are detailed elsewhere in other reviews of this series. *This article is part of a curated collection.

INTRODUCTION

Molecular epidemiologic investigations must adhere to standard epidemiologic principles to minimize information bias in the application of microbiology protocols and guidelines. The first step in a laboratory-based epidemiologic investigation or surveillance of infectious diseases is, of course, diagnosis. Until recently, the diagnosis of a suspected bacterial infection and identification of its etiologic agent largely relied on isolating the bacteria by pure culture in an artificial medium. With increasing accessibility and decreasing cost of genome sequencing, discussions have emerged about culture-independent diagnostic tests (CIDTs) (1, 2). Nevertheless, with a few exceptions for which no artificial medium to culture bacteria exists (e.g., *Treponema pallidum*) or the culture procedure is very cumbersome (e.g., *Chlamydia* spp.), the ability to isolate bacteria in pure culture provides several advantages that CIDTs still lack. These include the ability of culture-based tests to (i) determine clinically relevant or meaningful infectious inoculum size, (ii) perform quantitative drug susceptibility tests (e.g., MIC determination), (iii) genotype a pathogen of interest from a nonsterile clinical specimen, (iv) make isolates available to specialized laboratories

Received: 22 March 2018, **Accepted:** 4 June 2018, **Published:** 2 November 2018

Editor: Michael Sadowsky, BioTechnology Institute, University of Minnesota, St. Paul, MN

Citation: Riley LW. 2018. Laboratory methods in molecular epidemiology: bacterial infections. *Microbiol Spectrum* 6(6):AME-0004-2018. doi:10.1128/microbiolspec.AME-0004-2018.

Correspondence: Lee W. Riley, lwriley@berkeley.edu

Curated Collection: Advances in Molecular Epidemiology of Infectious Diseases.

for reference testing (e.g., WHO Collaborating Center-directed quality assurance and standardization tests), and (v) facilitate the use of archived isolates for future analyses (e.g., for developing and validating new diagnostic tests, identifying new drug targets, designing new vaccines, and conducting animal experiments and human volunteer studies). For these reasons, now and in the foreseeable future, epidemiologic investigations of bacterial infectious diseases cannot be totally dependent on methods that exclusively rely on CIDTs (3).

Microbiologic diagnosis of infections must also consider the type of infection site. With bacterial infections of certain body sites considered sterile (blood, peritoneal or pleural cavity, brain, cerebrospinal fluid, bone marrow, and subcutaneous tissue), a laboratory-based epidemiologic investigation usually begins with isolation of the etiologic agent in pure culture. Of course, we now know that sites previously considered sterile may not be, as such sites often contain a variety of viruses (and perhaps even uncultivatable bacteria) that can be detected by nucleic acid amplification methods. Nevertheless, such sites should not consistently contain live, cultivatable bacterial organisms and if they are recovered by culture, they would be presumed to be pathogens in most situations.

With bacterial infections that occur in a nonsterile niche (e.g., gastrointestinal tract, vaginal canal, upper respiratory airways, and skin), the identification of a pathogen requires an additional step of separating the pathogen from normal commensal or saprophytic organisms present at the same site. Commensal bacteria (normal flora) metabolize nutrients and replicate at the site of their colonization. They reside in a healthy, immunocompetent host without causing disease in that host. Saprophytes, on the other hand, are microorganisms that gain nutrients from dead organic matter. They are thus found in the environment outside of live animals or plants but can transiently colonize the human host. Both commensal and saprophytic bacteria can interfere with the detection of a bacterial pathogen if the latter causes an infection at the same site as colonized by commensal organisms. Thus, a diagnostic test for an infectious disease at such sites must be able to selectively isolate and separate pathogens from these other microbes.

In an epidemiologic investigation of bacterial infections, other issues, such as specimen collection, handling, transport, storage, and selective decontamination, are also critical in ensuring reliable diagnostic test results and minimizing misclassification bias at the time of data analyses. These procedures are described in detail elsewhere (4).

Molecular epidemiology analyses begin only after a diagnosis is made and confirmed. The confirmed or tentatively confirmed diagnosis, then, becomes part of a case definition in an epidemiologic investigation. The subsequent molecular microbiology analyses all deal with comparison of subtypes of isolated organisms implicated in infections of human or nonhuman animal populations. The remainder of this review discusses these subtyping methods applied to bacteria.

TAXONOMIC ANALYSES OF BACTERIA

Taxonomic classification of microbes forms the foundation of microbiology and infectious disease diagnosis. In both sterile- and nonsterile-site infections, the isolated bacterial pathogen is first identified as belonging to a distinct taxonomic unit: species. Thus, species identification constitutes the first step in data stratification of bacteria isolated during an epidemiologic investigation. But what is a bacterial or prokaryotic "species"? Reference books, such as the current *Bergey's Manual of Systematic Bacteriology*, classify bacterial organisms based on their phenotypic and 16S rRNA sequences (5). However, what constitutes "species" requires more extensive discussions on how microbes are categorized into new taxonomic ranks. This issue is reviewed in detail elsewhere (6–8).

Simply stated, species classification of prokaryotes is dependent on the laboratory technique used to subtype the bacterial organism. For example, the number of what are considered *Salmonella* species has changed over the years, depending on whether the bacteria are classified biochemically (Ewing's classification), by serotyping according to their O polysaccharide, flagellar antigen (H), and capsular types (K) (Kauffmann-White scheme), by their 16S rRNA sequences, or by percent DNA-DNA hybridization (9–11). The Centers for Disease Control and Prevention (CDC) uses the nomenclature proposed by the WHO Collaborating Center, which currently divides the genus *Salmonella* into 2 species, each of which contains multiple serotypes (9). The two species are *Salmonella enterica* and *S. bongori* (9).

Currently, the most widely used method of assigning new taxonomic ranks to prokaryotic organisms is based on 16S rRNA gene sequences. As of 30 September 2016, 3,356,809 rRNA sequences have been catalogued (http://rdp.cme.msu.edu/). However, whole-genome sequence (WGS) analyses have revealed that many bacterial organisms with identical or nearly identical 16S

rRNA gene sequences have extensive differences in other regions of their genomes, which suggests that the 16S rRNA gene provides limited resolution for designating prokaryotic organisms at the species level (8, 12, 13).

A reader interested in this bacterial classification conundrum is referred to excellent discussions reviewed elsewhere (12, 14, 15). With the "big data" generated from WGS analysis of bacterial organisms, what is considered species will continue to evolve. In fact, the genome sequences generated from a large number of bacterial strains have engendered a new taxonomic term—"pan-genome"—which is defined as a sum of all genes shared among multiple strains belonging to a "species" (core genes) plus genes found in only one strain (unique genes) or two or more but not all strains (dispensable genes) (16–19). Species such as *Escherichia coli* are considered to have an "open" pan-genome because each new genome that is sequenced reveals new unique genes. In fact, it was predicted that *E. coli* genomic diversity represents an open pan-genome that contains a reservoir of more than 13,000 genes (19) or even larger, as more recent analyses suggest (20, 21). It is possible that the term species applied to bacteria may become obsolete in the near future.

It should be emphasized that nucleic acid sequence-based taxonomic classification of bacteria is only one of many bacterial classification methods. Which classification method to use depends on the intended application of the method. For application to molecular epidemiology, subtyping bacterial organisms to assess strain relatedness will increasingly become dependent on nucleic acid sequence-based methods. This idea will be explored further in a later review on phylogenetic concepts and tools applied to epidemiologic investigations of infectious diseases.

BACTERIAL SUBTYPING METHODS FOR APPLICATION TO EPIDEMIOLOGY

One important analytical task of an epidemiologic investigation is data stratification. In the example discussed above, the classification of *Salmonella* into just 2 species is not very informative or helpful for surveillance or epidemiologic investigations of salmonellosis. The National Notifiable Diseases Surveillance System (https://wwwn.cdc.gov/nndss/) and the FoodNet surveillance system (https://www.cdc.gov/foodnet/index.html) are based on classification of *Salmonella* by serotypes. This is possible because the *Salmonella* subtyping scheme (Kauffmann-White) stratifies *Salmonella* into more than 2,400 distinct serotypes. Serotype information is used to monitor relative incidence of salmonellosis outbreaks and trends caused by different serotypes, which makes possible identification of specific risk factors associated with each serotype. For example, a large proportion of salmonellosis caused by *Salmonella enterica* serotype Enteritidis and *S. enterica* serotype Heidelberg in the United States is associated with contaminated poultry products, while most cases caused by *S. enterica* serotype Typhimurium and *S. enterica* serotype Newport are associated with a wide variety of meat products. Identification of distinct risk factors facilitates focused intervention. Without the serotype information, it would be difficult for such public health interventions to be made. Thus, epidemiologic investigations of bacterial infections require strain subtype information beyond the level of species.

Bacterial "Clones" as a Taxonomic Unit of Epidemiologic Investigation

One important aim in most molecular epidemiologic investigations of bacterial infections is to identify clones or clonal complexes responsible for an outbreak, epidemic, or pandemic or for sporadic infections. That is, clones are the ultimate subtype unit of analysis we seek to conduct molecular epidemiologic studies. We look for clones or clonal lineages associated with a clinical outcome, and then we investigate risk factors for infection with or modes of transmission of such lineages. But what is a bacterial "clone" or "clonal lineage" (22)?

The standard definition of a clone is any isolate or a group of isolates descending from a common precursor strain by nonsexual reproduction. Because of recombination events and horizontal gene transfers, many bacterial species do not exhibit strict clonal reproduction. Certainly, the term clone is not applicable to characterize eukaryotic pathogens that reproduce sexually (discussed in a separate review). The term bacterial clones is loosely used in much of the literature to refer to a group of bacterial strains belonging to a species that are indistinguishable by a genotyping test used. If evidence can be empirically provided that a variant of a genotype descended from an ancestral genotype, or a consensus definition is proposed to characterize variant genotypes to be related, then strains belonging to such a cluster of genotypes are considered members of a clonal complex or clonal group.

The problem arises when WGS comparison is used to characterize evolutionary relatedness of bacterial strains. Strains considered to belong to the same clonal group or clonal complex by multilocus sequence typing (MLST) or pulsed-field gel electrophoresis (PFGE) (or

any of the other analytical tests described below) may, by WGS, be shown to exhibit greater polymorphism. At this time, we do not have standardized criteria (e.g., the number or type of genome-wide nucleotide substitutions) that can be used to include a group of bacterial strains in or exclude them from a clonal group or complex.

However, the above concern mainly pertains to molecular evolution or phylogenetic investigations. In molecular epidemiology, this question is important but addressable. Ultimately, what constitutes clones or a clonal complex is determined empirically. It is determined by an epidemiologic study that identifies a risk factor to be statistically significantly associated with what is defined as a clone or a group of clones. If indeed what is called a clone or a group of clones is eliminated by an intervention based on removal of that risk factor, then it does not really matter what truly constitutes a clone or clonal complex. What is important is to determine if strains in question or strains designated as clones are epidemiologically linked.

The relationships among clonal lineages may be phylogenetically represented (based on some evolutionary model or nucleic acid sequence similarity indices). The objective of molecular epidemiology is to determine whether the evolutionary or mathematical representation has any epidemiologic significance. Such a determination must be made empirically. As in the above example, if a risk factor for infections associated with a clonal lineage is eliminated after an intervention based on an epidemiologic investigation that relied on genotyping data, then the evolutionary or mathematical representation of a clonal lineage relationship can be said to exhibit an epidemiologic relationship. The phylogenetic classification is only one of many tools used to generate data points that can be included in an epidemiologic analysis to identify risks and modes of transmission.

The determination of epidemiologic relatedness of clones or groups of clonal lineages is best made by well-defined, geographically and temporally restricted outbreak investigations. One of the major reasons for investigating an outbreak is to help validate new genotyping tests. In a clearly defined bacterial disease outbreak, a new strain typing test can be validated with the isolates obtained from cases (as defined in the outbreak investigation) and compared to those isolates belonging to the same species from sources that are clearly distinct from the outbreak. The test may show all the isolates from the outbreak to share a distinct and indistinguishable genotypic marker and nonoutbreak isolates to have different markers. If so, the test can be said to be validated for epidemiologic investigation purposes. It is also possible that the outbreak isolates show a range of similar but not identical genotypic markers. Then, the test can accommodate such a degree of variation (clonal complex) to be applied in the outbreak risk factor analysis and subsequent epidemiologic investigations. This notion is extensively discussed in a previous book on molecular epidemiology in a section that discusses how to validate new genotyping tests for epidemiologic investigations (23).

With WGS comparisons, the same approach may be used to determine a threshold of the number and type of genome-wide nucleotide substitutions needed to characterize strains to belong to epidemiologically related lineages or clonal complexes. In a study of hospital outbreaks of methicillin-resistant *Staphylococcus aureus* (MRSA), WGS analyses of the MRSA isolates implicated in the same outbreaks were found to have many variations, including single nucleotide polymorphisms, small genetic rearrangements, and the presence or absence of mobile genetic elements encoding resistance and virulence genes (24). With some patients, multiple isolates from the same patient showed polymorphism—referred to as "clouds" of related isolates (24). The information derived from WGS may be considered too discriminatory, which complicates analyses of genetic relatedness of the strains. However, the fact that they are part of the same outbreak (especially if the outbreak is investigated carefully) provides the evidence that all the isolates from cases in the outbreak are epidemiologically linked. The outbreak itself provides the evidence for strain relatedness independent of the WGS data. Thus, the range of variation in WGS accepted as a criterion for epidemiologic relatedness can be validated against an outbreak. Once determined, this range in genomic variation can be used to infer relatedness of isolates from other outbreaks or from what appear to be sporadic cases of the same disease caused by the same bacterial species.

Of course, in prolonged hospital MRSA or Gram-negative bacterial infection outbreaks, the number of genome-wide variations in the isolates is likely to be higher than in isolates obtained from acute outbreaks. In such a situation, the determination of a threshold based on number or type of genome-wide nucleotide substitutions to distinguish outbreak from nonoutbreak strains becomes more complicated. The degree of genome-wide changes across strains can also vary with bacterial species that undergo frequent recombination or horizontal gene transfers. With bacterial organisms that

cause chronic infections (e.g., *Mycobacterium tuberculosis*), it would also be difficult to identify a range of genomic variation that can be accepted with confidence to distinguish lineages of epidemiologic interest from others. Epidemics of *M. tuberculosis* infections do occur, but they may go on for many months or even years. How does one determine the difference in the number of genome-wide nucleotide substitutions needed to show relatedness of *M. tuberculosis* isolates from people living in the same community who develop tuberculosis spanning across many years? Finally, when occurrence of an outbreak or an epidemic is uncertain (e.g., community-acquired urinary tract infection), WGS information may not be usable to show epidemiologic relatedness of isolated bacterial strains. In such situations, other empirical approaches need to be used. These approaches will be discussed in subsequent reviews of specific bacterial infections.

Phenotyping Tests for Bacteria

Bacterial subtyping methods are divided into phenotyping and genotyping methods (25–27). Phenotyping methods include those that compare bacterial strain characteristics expressed by their genes. Such characteristics include growth behavior in an artificial medium, morphologic characteristics, biochemical properties, serologic characteristics, and functional or physiologic properties. Specific examples of each of these characteristics are shown in Table 1 and are described in detail in standard clinical microbiology textbooks and are not discussed further here (4).

Genotyping Tests for Bacteria

Genotyping methods rely on manipulation and analysis of microbial nucleic acid contents. In bacteria, these nucleic acid contents include chromosomal DNA, extrachromosomal DNA (plasmids and phages), and RNA. In general, genotyping tests provide a higher resolution of strain stratification than phenotyping methods. That is, genotyping tests are usually more discriminating than phenotyping tests (although this depends on the genotyping test used). As described above, phenotyping tests encompass a wide variety of laboratory procedures that rely on several different types of equipment and analytical instruments, reagents, and laboratory skills. Standardization, exchange or comparison across different laboratories, or comparison over time in the same laboratory of some of these phenotypic data can be complex and labor-intensive. Maintenance of these requirements can also be costly.

Today, there is a large repertoire of genotyping tests used in epidemiologic investigations, reviewed elsewhere (28) (Table 2). In contrast to phenotyping tests, however, most of these genotyping tests represent just a refinement and variation of only 3 general categories of molecular biology procedures: (i) electrophoresis of nucleic acid fragments, (ii) hybridization, and (iii) nucleic acid sequencing. Consequently, these tests can all be performed with a commonly shared and limited set of equipment or analytical instruments, reagents, and skill sets, and they can be applied to all forms of nucleic acid-based infectious agents (bacteria, viruses, fungi, protozoa, and helminths). Most of the steps in nearly all of these tests have been automated, and data generated from these procedures can be stored digitally, which facilitates data management standardization, portability, and electronic exchange across same and different laboratories.

However, one new development that complicates some of the new-generation genotyping tests is the advanced skill set needed to analyze the vast amounts of data generated from these tests. A separate review will address this new development. Below we describe these genotyping tests, especially those useful for epidemiologic investigations of bacterial infections, and discuss their advantages and limitations.

TABLE 1 Phenotypic characterization of bacterial organisms

Characteristic	Examples of assays
Growth behavior	Nutrient requirement; growth in the presence or absence of oxygen; requirement of light or darkness; temperature; slow or rapid growth
Morphology	Colony color, shape, smell; staining patterns (Gram stain, acid-fast stain); shape observed under microscope (rod, coccus, spirochete)
Bichemical properties	Fermentation, gas production, substrate utilization, mass spectrometry profile
Serologic properties	Recognition by antibodies to O polysaccharide (O antigen), flagellar protein (H antigen), capsular polysaccharide (K antigen)
Functional or physiological properties	Virulence factors (cell attachment, invasion, cytotoxicity, toxin production, pathogen-associated molecular patterns, disease production in an animal model), susceptibility to antimicrobial drugs, host effector molecules (antibodies, reactive oxygen and nitrogen intermediates, antimicrobial peptides, granzymes), phages, and other stresses; host subversion and evasion mechanisms; infectious inoculum dose

TABLE 2 Genotyping tests used to conduct epidemiologic investigations of bacterial infections

Genotyping test	Application(s)	Comments
Electrophoresis-based tests		
Plasmid profile analysis	Foodborne and health care-associated infections caused by bacterial species that harbor plasmids	Mostly of historical interest
AP-PCR; RAPD	Used to genotype microbes without prior knowledge of nucleic acid sequences	Prone to nonspecific DNA amplification
PCR-REA or PCR-RFLP	Used to compare long sequences (>1,000 bp) within a genome	Need prior knowledge of nucleic acid sequence
RSS-PCR	Used to subtype strains by pathotypes or serotypes	Not as discriminating as other tests
AFLP	Targets a whole genome	Results affected by ligation efficiency and PCR conditions
REP-PCR	Used to subtype Gram-negative or positive bacteria containing palindromes	Unreliable batch-to-batch reproducibility
ERIC-PCR	Used to genotype Gram-negative bacterial species that carry ERIC DNA (e.g., *E. coli*)	Unreliable batch-to-batch reproducibility
BOX (boxA, boxB, boxC) PCR	Used to genotype Gram-positive bacterial species (e.g., pneumococcus)	Unreliable batch-to-batch reproducibility
PGRS	Used to genotype *M. tuberculosis* only	Not very discriminating
MIRU-VNTR	Used to genotype *M. tuberculosis*-complex only	Strain discrimination depends on the number of repetitive elements amplified
MLVA	Used to genotype any bacterial organism containing VNTR sequences	Highly discriminatory but labor-intensive
PFGE	Widely used test to genotype bacterial species; prior knowledge of DNA sequence not needed	Reference standard for many species; labor-intensive; lab-to-lab comparison difficult
Hybridization arrays		
IS*6110*-RFLP (RFLP-hybridization)	Used to genotype *M. tuberculosis* complex	Reference standard for *M. tuberculosis*
Spoligotyping	Used to genotype bacteria containing DR sequences or CRISPR sequence (e.g., *M. tuberculosis*)	Not as discriminating as other tests for *M. tuberculosis*
Sequence-based tests		
MLST	Used to type any bacteria	Not as discriminating as NGS tests
NGS	Used to type any bacteria	Most discriminating genotyping test

Genotyping tests for bacteria fall into two general categories: those based on analysis of extrachromosomal DNA elements and those based on analysis of the whole genome. Analyses of parts of a genome to represent relatedness of strains compare strain microdiversity, while analyses of the entire genome compare strain macrodiversity. The higher speed and decreasing cost of next-generation sequencing (NGS) technology is making comparison of strain macrodiversity increasingly accessible. Furthermore, information gathered from genomic sequence analyses contributes to the development of new strain microdiversity genotyping as well as phenotyping tests with higher precision. Thus, for applications in phylogenetics and molecular epidemiology, both categories of genotyping tests will remain important.

Extrachromosomal DNA Genotyping Tests

Plasmids comprise the main extrachromosomal DNA of bacteria. Although subtyping bacteria by resolving plasmids of different molecular weights (MW) by agarose gel electrophoresis (AGE) is mostly of historical interest, plasmids still play an important role as targets of new molecular biology and epidemiology tools. They are discussed later under "PCR-based genotyping tests" (Table 2).

Plasmid profiles, "plasmid fingerprints," or plasmid band patterns are generated from multiple plasmids resolved by AGE, and these patterns based on plasmid MW differences are compared to infer strain relatedness. In instances where strains carry only one or two plasmids, the plasmids can be digested with a restriction endonuclease and their DNA fragments separated by gel electrophoresis to provide greater resolution.

The first epidemiologic application of plasmid profile analysis was done to investigate hospital outbreaks. Schaberg et al. resolved plasmids of drug-resistant *Pseudomonas aeruginosa* and *Serratia marcescens* isolates by AGE to demonstrate 7 distinct hospital outbreaks in the United States caused by these Gram-negative bacteria (29). Each outbreak was shown to have been caused by strains with distinct AGE patterns, which differed across hospitals (29). These outbreaks, as well as other outbreak investigations by the CDC (30), helped to establish plasmid profile analysis as an epidemiologic tool.

Subsequent applications of AGE of plasmids have contributed to a wide variety of epidemiologic investigations of infectious diseases in health care and community settings (31). These investigations have included (i) identifying specific vehicles of infection in community setting outbreaks (32–35), (ii) demonstrating food animals as a source of drug-resistant *Salmonella* organisms causing human salmonellosis (34–39), (iii) estimating the attributable fraction of a contaminated food product as a cause of sporadic cases of salmonellosis in a community (36), (iv) unmasking multiple small outbreaks from what appear to be endemic cases of hospital infections (40), (v) characterizing outbreaks of *Salmonella* infections among food animals (41), and (vi) conducting surveillance of drug-resistant enteric bacterial pathogens regionally (37, 42, 43), to name a few examples.

Most epidemiologic investigations of community-acquired infections cited above that examined plasmid profiles involved enteric pathogens. This is because most of these Gram-negative enteric pathogens (*Salmonella*, *E. coli*, *Shigella*, and *Campylobacter*) carry plasmids, and AGE analysis of plasmids is a relatively simple procedure. Of course, this technique cannot be applied to bacteria that do not harbor plasmids. The instability of some plasmids during storage of bacterial isolates could also affect results. Finally, plasmid profiles based on isolates obtained from geographically distant sources or sources separated by long time periods cannot be reliably assessed for strain relatedness.

Another plasmid-based bacterial subtyping method takes advantage of a specific gene of interest (e.g., virulence gene) in a plasmid by PCR. Such an assay will generate only one band on AGE, and the MW difference of the band between strains is compared to distinguish subtypes. One of the earliest applications of this approach was used to differentiate pathogenic variants (pathovars) of *E. coli* (44). Plasmids of intestinal pathogenic *E. coli* (IPEC) carry distinct virulence genes that can be targeted by PCR to differentiate IPEC strains into enteropathogenic *E. coli* (EPEC), enteroinvasive *E. coli* (EIEC), and enterotoxigenic *E. coli* (ETEC) groups (44, 45). These plasmid genes include the bundle-forming pilus gene (*bfpA*) of EPEC (46), invasion plasmid antigen H gene (*ipaH*) of EIEC (47), and heat-labile and -stable toxin genes of ETEC (48, 49). Primer pairs were designed to amplify a segment of these genes to generate products of different MW; this MW difference was used to differentiate the *E. coli* strains into distinct IPEC groups.

More discriminating electrophoresis-based and nucleic acid sequence-based genotyping tests have largely replaced plasmid AGE analysis of bacterial pathogens as a tool to conduct epidemiologic studies. However, as other reviews of this series will show, detailed nucleic acid sequence analysis of plasmids themselves has become a new and an important approach to characterize pathogenesis and bacterial strain fitness associated with their transmissibility and disseminative capacity. Thus, plasmids remain an important extrachromosomal DNA target of epidemiologic investigations.

Genome-Based Genotyping Tests

Genomic DNA-based genotyping tests can be separated into four general categories: analysis of (i) electrophoretic band patterns generated from PCR-amplified DNA products, (ii) electrophoretic band patterns of restriction enzyme-digested genomic DNA, (iii) nucleic acid hybridization arrays on a membrane substrate, and (iv) nucleic acid sequences (Table 2).

PCR-Based Genotyping Tests

PCRs can be used to rapidly subtype a wide variety of bacterial species for epidemiologic investigations. All of the PCR-based tests are designed to generate "fingerprint" patterns of PCR-amplified products resolved by AGE. These patterns are then compared to infer strain relatedness. Some PCRs are custom-designed to genotype bacterial strains belonging to one species, while others are universally designed to genotype any bacterial species. The major advantages of PCR-based genotyping tests are their speed, ease, low cost, and versatility. The major disadvantage is their limited reproducibility. These features are discussed below for each type of PCR-based genotyping test.

Arbitrarily primed PCR and randomly amplified polymorphic DNA

Arbitrarily primed PCR (AP-PCR) and randomly amplified polymorphic DNA (RAPD) do not depend on prior knowledge of target nucleic acid sequences (50–55). Thus, they can be used to type strains belonging to any bacterial species (or, for that matter, any organism containing a DNA genome). Both tests generate multiple amplified DNA products by subjecting genomic DNA to PCR by short (8- to 12-base) oligonucleotide primers of arbitrary sequences. In its first reported application to genotype 24 bacterial strains belonging to 5 *Staphylococcus* species, AP-PCR used a single primer under two cycles of low-stringency annealing temperature followed by many cycles of the high-stringency condition to generate a discrete set of amplicons resolved by AGE to yield distinct band patterns (54).

Oligonucleotide primer sequences for RAPD/AP-PCR are determined empirically by trial and error, and the final choice of the primer design is based on reproducibility of interpretable band patterns by gel electrophoresis of the amplified products. To ensure reproducible results useful for epidemiologic studies, it is important to first validate the test with a large collection of oligonucleotides examined against a well-characterized panel of bacterial strains belonging to one species. Reproducibility is affected by differences in the concentration of salts in the PCR buffer; annealing temperature; primer concentration, sequence and length; cycling condition; template DNA concentration; and quality and age of polymerase.

Another limitation of RAPD/AP-PCR is that the template DNA must be prepared from bacteria recovered from pure culture that does not include any contaminating DNA from other sources. Since recombinant polymerase, such as *Taq* polymerase, is produced in *E. coli*, even a commercially obtained enzyme may be contaminated with *E. coli* DNA, which could get amplified by arbitrary sequence primers. One way to minimize nonspecific amplification is to use larger primer pairs (18 to 24 bases each). A variant of RAPD called large-primer RAPD (LP-RAPD) has been shown to generate reproducible fingerprints under different PCR conditions (56).

A PCR sample without any template DNA can be included as a negative control to rule out contaminating DNA in the PCR reagents, but this will not eliminate contamination in the test DNA sample itself. Even with these caveats and precautions, batch-to-batch variability of electrophoretic gel banding patterns is inevitable. Band patterns obtained by RAPD/AP-PCR from the same set of strains on separate days during one week may change when they are run several weeks apart (50). Thus, for epidemiologic applications, such as long-term disease surveillance, these genotyping tests would not be very useful. However, for acute outbreak investigations or as a screening test prior to performing more labor-intensive or expensive genotyping tests, RAPD/AP-PCR tests are highly useful.

Electrophoretic band pattern analysis of restriction enzyme-digested PCR products

A PCR product in the range of 1,000 to 2,000 bp can be digested with a restriction endonuclease to generate DNA fragments that can be resolved by agarose or acrylamide gel electrophoresis. Such a procedure is called PCR-restriction enzyme analysis (PCR-REA) or PCR-restriction fragment length polymorphism (PCR-RFLP) (57, 58). When a target of PCR is the ribosomal DNA (rDNA) sequences, the technique is referred to as amplified rDNA restriction analysis (ARDRA) (59, 60). In most microorganisms, because of the conserved nature of rDNA sequences within a genus, the epidemiologic utility ARDRA is limited. It may be used to rapidly identify bacterial strains belonging to a genus that otherwise may involve labor-intensive phenotyping procedures to characterize, such as *Mycobacterium* spp. in the bloodstream of patients with AIDS (61). Unlike RAPD/AP-PCR, both PCR-REA and ARDRA require prior knowledge of the target sequences to be amplified. Such knowledge minimizes nonspecific amplification of contaminating DNA.

Restriction site-specific PCR

Another PCR-based genotyping test that relies on mutations that occur at an endonuclease restriction site is called restriction site-specific PCR (RSS-PCR) (62). In this method, no endonucleases are actually used. Rather, RSS-PCR is based on PCR amplification of spaces between restriction enzyme recognition sequences that occur frequently in a target gene to generate electrophoretic fingerprint patterns by AGE. The restriction recognition sites that are targeted are in the genes that exhibit high sequence diversity (e.g., outer membrane protein genes). If these recognition sites contain mutations, especially at the 3′ end of the primer-binding site, the mismatch will preclude amplification of spaces between these sites. Differences in these mismatches among different strains will generate varied band patterns on AGE.

The first application of RSS-PCR was used to distinguish *E. coli* O157:H7 from other intestinal pathogenic *E. coli* strains (62). The primers were based on the restriction enzyme recognition sequences of *Bbv*I and *Taq*II found in the *chuA* gene of *E. coli* O157:H7 strain EDL933 (GenBank accession no. U67920), which encodes an outer membrane heme receptor protein involved in iron transport and is specific to *E. coli* (63). This gene contains four *Bbv*I recognition sites and three *Taq*II recognition sites. RSS-PCR may not be as discriminating as the other PCR-based genotyping tests, but it is useful for typing bacteria to a subspecies level that may correspond to pathotypes or serotypes.

Amplified fragment length polymorphism genomic analysis

RSS-PCR depends on allelic differences in the targets to generate DNA fragments resolved by AGE. There are other PCR-based tests based on differences in tar-

get alleles. They are collectively referred to as amplified fragment length polymorphism genomic analysis (AFLP) (64–66). Unlike PCR-REA and PCR-RFLP, which compare banding patterns of DNA fragments representing a small region of a genome, DNA fragments produced by AFLP represent total genomic DNA. First, the genomic DNA is digested with endonucleases, usually with a combination of two types of enzymes. The DNA fragments are then ligated to oligonucleotide adapters. The adapter sequences serve as targets for primers to anneal, and the restriction fragments of various lengths are selectively amplified. The amplified products are then resolved by electrophoresis, which generates band patterns. These band patterns are then compared to determine genetic relatedness.

AFLP does not require prior knowledge of the target DNA sequences. In general, information generated from AFLP matches closely with subtyping assignment made by other genotyping methods (64). However, the consistency of gel band patterns produced by this procedure may be affected by technical variability, such as sample-to-sample variation in the amount of DNA extracted from bacteria, quality of endonucleases, and efficiency of ligation of the adapters.

Broad-spectrum repetitive DNA element PCR typing methods

Stretches of identical nucleic acid sequences that occur multiple times in the genome are called repetitive elements. The genomes of both prokaryotic and eukaryotic microbes contain many such elements that can be exploited to genotype them. Bacterial genomes contain several types of such repetitive elements—interspersed repetitive sequences comprised of short, noncoding oligonucleotide sequences (~15 to several hundred base pairs) that occur in clusters at different locations in a genome, rRNA genes, and insertion sequences (IS) that jump from one location to another in a genome (67, 68). The last two repetitive elements are large coding sequences that occur in low copy numbers. Some of the interspersed repeat sequences are conserved across several bacterial species or even genera, while others are restricted to one species or a limited number of species in a genus. The former are described as broad-spectrum repetitive DNA elements, while the latter are referred to as species-specific repetitive DNA elements.

All of the subtyping tests based on repetitive elements are designed to amplify sequences located between these elements. These subtyping tests are collectively referred to as repetitive element PCR (rep-PCR). The repeat sequences are targeted as annealing sites for outwardly oriented complementary oligonucleotide primers which amplify spaces located between these repeats (Fig. 1). Thus, only one set of PCR primers are needed to generate many amplicons of different sizes that can be resolved and visualized by AGE for each bacterial strain. The genetic distance between strains is inferred from an assumption that indistinguishable or similar band patterns represent occurrences of these repetitive elements in the same or similar locations in a genome.

Major broad-spectrum bacterial interspersed repetitive elements include repetitive palindromic elements (REP) or palindromic units (PU), enterobacterial repetitive intergenic consensus (ERIC) sequences, and BOX sequences, reviewed elsewhere (67–73) (Table 2). REP are comprised of a 38-bp palindromic consensus sequence, first described for two members of the family *Enterobacteriaceae*, *E. coli* and *Salmonella* Typhimurium (70, 71, 74). Between 500 and 1,000 REP sequence copies are distributed as clusters in the *E. coli* genome (70), which makes possible the generation of multiple amplicons by PCR. REP sequences have been reported for other bacterial phyla as well (75). REP or PU clusters have been found to contain other conserved repetitive elements in a mosaic pattern, which are called bacterial interspersed mosaic elements (76).

ERIC or intergenic repeat units are 124 to 127 bp long; *E coli* has about 30 to 50 copies, and *S.* Typhimurium has about 150 copies (72, 77). Although they were initially reported to occur in enteric bacterial species, PCR and hybridization studies have demonstrated variants of these sequences throughout the eubacterial kingdom (67).

BOX sequences belong to a class of mosaic repetitive elements first discovered in the intergenic regions of *Streptococcus pneumoniae* (78). They are mosaic elements because they form clusters comprised of different combinations of the same subunit repetitive sequences. The *S. pneumoniae* genome contains about 25 BOX repeat units, comprised of various combinations of three subunits: boxA (59 bp), boxB (45 bp), and boxC (50 bp) (78). They are widely distributed among Gram-positive bacteria. Later studies have shown *E. coli* and *S.* Typhimurium to contain boxA-like sequences but not boxB or boxC (79).

These broad-spectrum repetitive elements have been targeted to genotype a wide variety of bacterial genera by rep-PCR, including MRSA (80), *Legionella* spp. (81), and *Xanthomonas* and *Pseudomonas* pathovars (82). While this could be an advantage, one important caveat is that the primers designed to target these repetitive elements could potentially anneal nonspecifically to DNA sequences in the same way RAPD primers do.

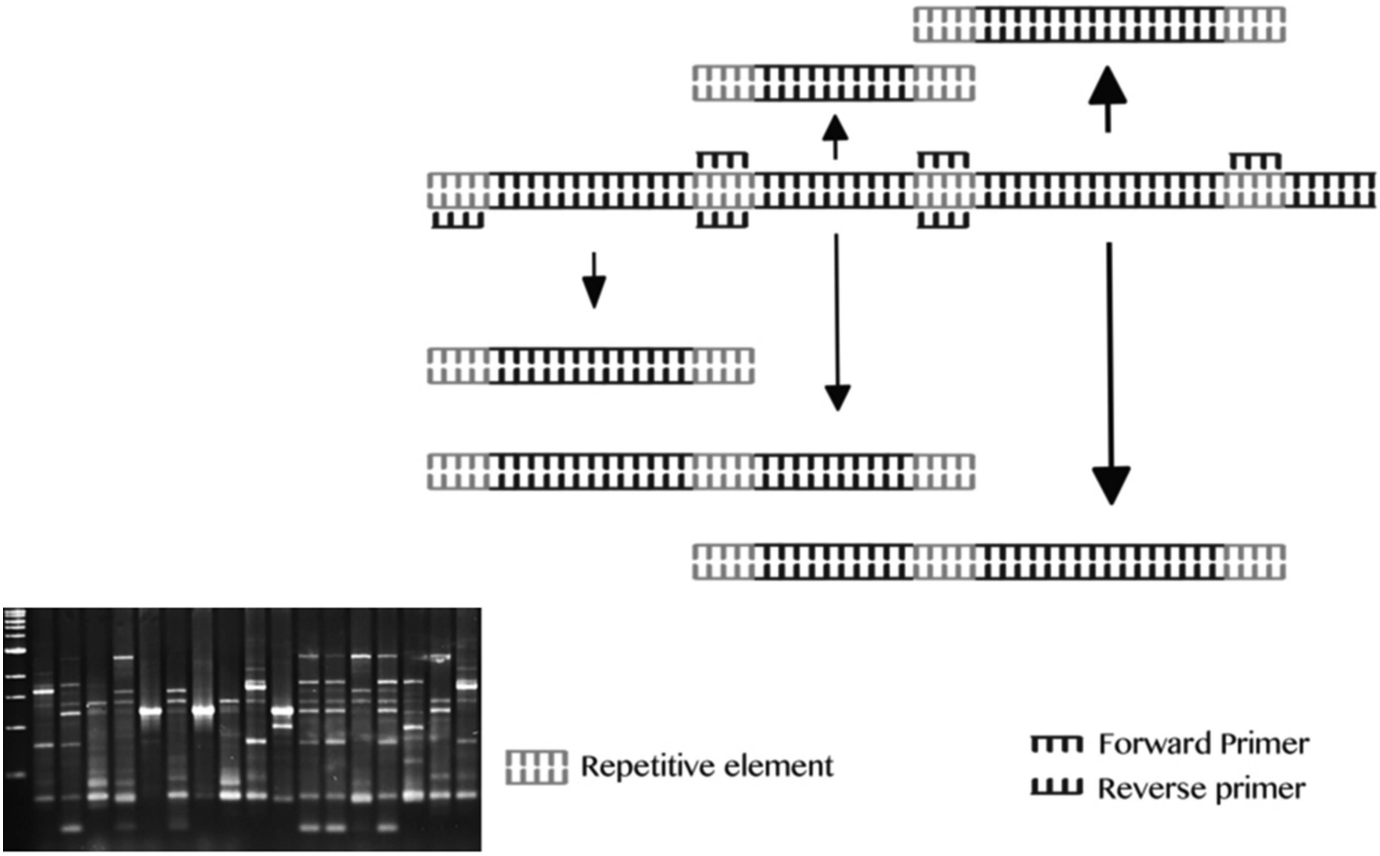

FIGURE 1 Repetitive element PCR (rep-PCR). The repeat double-stranded DNA sequences (light grey segments) are targeted as annealing sites for outwardly oriented complementary oligonucleotide primers, which amplify spaces located between these repeats (indicated as dark segments). Only one set of primers is needed to amplify the different segments, which generates amplicons of different sizes, shown at the end of each arrow, that can then be resolved by AGE for comparison of band patterns. (Illustrated by Paolo Harris Paz.) (Photo inset) AGE of band patterns generated by ERIC2 PCR analysis (a type of rep-PCR) of *E. coli* strains isolated from patients with community-acquired urinary tract infection. (Photo by Reina Yamaji.)

In one study, when DNA samples from a broad range of organisms were tested by ERIC-PCR under standard annealing temperatures, complex fingerprint patterns were generated not only from eubacteria but also from bacteriophages and eukaryotes (83). When a higher annealing temperature was used, bands were generated only from enterobacterial organisms. Thus, contaminating DNA could potentially affect results, as observed with RAPD-PCR.

Species-specific repetitive DNA elements

The repetitive elements described above are conserved across several bacterial species and genera. There are other repetitive elements restricted to one or a small set of bacterial species that can be targeted for genotyping strains belonging to those species. In fact, WGS analyses have shown that a large proportion of prokaryotic organisms carry distinct species-specific repetitive elements. A 2012 report by Koressaar and Remm stated that of 613 prokaryotic species, 97% contained at least one species-specific repetitive element (84). A computational method to identify such repeats in microbes has been developed; it can be used to design species-specific PCR primers for these repeats (85) (http://bioinfo.ut.ee/multimprimer3/). Thus, it is possible to develop new species-specific rep-PCR tests to genotype strains across a wide range of bacterial pathogens.

One such group of repetitive elements often targeted to conduct epidemiologic investigation of tuberculosis includes the polymorphic GC-rich tandem repeat sequence (PGRS) (86), the mycobacterial interspersed repetitive units (MIRU) (87–89), and the direct repeat

(DR) elements (90–92). While these repetitive elements are not necessarily restricted to the species *M. tuberculosis*, they occur only in a very small subset of mycobacterial species, which makes it possible for them to be highly useful to genotype *M. tuberculosis*. Genotyping tests based on these repetitive units and their applications to study epidemiology of tuberculosis are discussed further in a separate review.

M. tuberculosis and other bacterial organisms, and even eukaryotic organisms, contain clusters of repeat elements called the variable numbers of tandem repeats (VNTR) (87, 88, 93–95). Mycobacterial organisms carry human microsatellite-like VNTR sequences MIRU, which are 40- to 100-bp DNA elements dispersed in intergenic regions of the chromosome (87, 88). There are 41 distinct MIRU loci in the chromosome of *M. tuberculosis* H37Rv (87). These loci have been targeted to develop a genotyping test called MIRU-VNTR (87, 89). Each locus may vary in the copy number of the repetitive elements. This copy number polymorphism determines the number of alleles per locus, which is the basis for strain differentiation. The copy numbers at each VNTR locus are predicted from their PCR product size in a gel. By targeting multiple VNTR sites by PCR, PCR products of various sizes can be generated and resolved electrophoretically in separate lanes of an agarose gel. Each set of lanes in a gel will thus show a VNTR allelic profile per strain. The copy numbers predicted from the PCR amplicon size may be used to numerically code each band in each lane (94). Thus, VNTR profiles can be digitalized and compared across laboratories.

With *M. tuberculosis*, early studies have suggested that the resolution achieved by a MIRU-VNTR that targets 12 loci is comparable to that of IS*6110*-RFLP (described below) (87, 88, 96), but later studies have shown that a 24-locus MIRU-VNTR test has greater discriminatory power that may even exceed that of IS*6110*-RFLP when combined with spoligotyping (described below) (89, 97, 98). Of course, this higher discriminatory power is achieved with a higher cost since 24 different PCR tests need to be performed for each *M. tuberculosis* strain.

In fact, MIRU-VNTR is a type of test collectively recognized as multilocus VNTR analysis (MLVA) (99). MLVA is applied to genotype a wide variety of bacterial species (100–107). It is used by public health laboratories of many countries to complement their PFGE-based surveillance system for enteric pathogens (108, 109). In Denmark, MLVA is used along with PFGE for both surveillance and outbreak investigations of *S.* Typhimurium. MLVA was found to be more discriminating than PFGE (110). In a study that genotyped Shiga toxin-producing *E. coli* O157:H7, MLVA was shown to be equally as sensitive as but more specific than PFGE (111). It should be noted that the development of many of the MLVA was made possible because of the genome sequence information provided by WGS analyses of these organisms.

Another group of repeat DNA sequences used for genotyping includes mobile elements, such as transposons or IS. IS are transposable elements that only carry genetic information needed to insert themselves into a new site. Some IS are shared by multiple species, while others are species specific. If such sequences occur in many copies distributed across a genome, spaces between such sequences can be amplified by PCR with outward-oriented primers annealed to the ends of IS, as described above for interspersed repetitive elements.

For example, *M. tuberculosis* carries IS*6110*, and Ross and Dwyer developed a PCR-based method to amplify sequences between IS*6110* to generate DNA fingerprint patterns (112). Of course, such a method requires *M. tuberculosis* to carry enough copies of this IS located relatively close to each other so that spaces in between can be amplified. By targeting both PGRS (present in >30 copies in *M. tuberculosis*) and IS*6110*, another rep-PCR method called double-repetitive element PCR was developed to rapidly genotype *M. tuberculosis* (113).

Limitations of PCR-based genotyping tests

Despite the increased availability of sequence-based tests, PCR-based genotyping tests for bacteria remain widely used because of their simplicity, speed, versatility, high throughput, and low cost. These tests can be used in outbreak settings to rapidly assess infectious disease spread, identify risks, and even trace vehicles or sources of such outbreaks. They are also frequently used as a screening test to help reduce the number of isolates that need to be tested by more discriminating but labor-intensive or expensive procedures.

However, one limitation of the PCR-based tests discussed above is their dependence on electrophoretic band patterns for analysis. The reproducibility of the band patterns is dependent on many technical parameters related to the PCR procedure itself, such as the type of thermocycler used, the amount of DNA template, primer composition and concentration, annealing temperature, quality of polymerase, buffer salt concentration, physical quality of agarose gel, and other PCR-related conditions. These factors can cause electrophoretic band intensity to vary from one lane to

another or between different gels, which complicates pattern interpretation to determine similarity distances. Patterns from the same strains can also vary from day to day even when the PCR procedure is executed by the same person in the same laboratory. In addition, DNA contamination can cause misinterpretation of electrophoretic band patterns. As described earlier, DNA contamination can occur inadvertently or be inherent in the PCR procedure if the recombinant polymerase used in the test contains *E. coli* DNA. PCR genotyping tests that are not dependent on knowledge of target sequences are particularly prone to nonspecific amplification of contaminating DNA sequences. Thus, all PCR-based genotyping tests must take into consideration protocols designed to minimize these sources of electrophoretic band pattern misinterpretation (114, 115) (http://www.protocol-online.org/biology-forums-2/electrophoresis.html).

Tests Based on Electrophoresis of Restriction Enzyme-Digested Genomic DNA

Restriction endonuclease digestion of genomic DNA may generate thousands of fragments that, when resolved by AGE, will not be interpretable. Two approaches have been developed to overcome this problem of interpretation of restriction enzyme-digested fingerprint patterns. One is to use an endonuclease that targets a rare recognition sequence in a genome, which would generate a small number of large DNA fragments that can be resolved in a gel; this approach is the basis for the PFGE genotyping procedure discussed below. The other is an approach designed to detect only the DNA fragments of interest among other fragments by a target-specific DNA probe, discussed under hybridization-based tests below.

Pulsed-field gel electrophoresis

DNA fragments are negatively charged. Conventional AGE separates DNA fragments in an electrical field applied in only one direction (Fig. 2). A bacterial genome cut with an endonuclease that recognizes rare restriction sequences will produce a small number (usually, less than 20) of large pieces of DNA, which, when resolved in a gel, would generate fingerprint patterns amenable to similarity analysis. To do so, however, requires overcoming two technical challenges. Conventional AGE cannot separate pieces of DNA of >50,000 bp due to the average pore size of the agarose gel matrix. Additionally, large DNA fragments undergo spontaneous autodigestion after treatment with a "rare-cutter" restriction enzyme, which would preclude consistent preparation of such fragments for electrophoresis.

The first challenge was overcome by applying electrical field in alternately pulsed, perpendicularly oriented directions (116). An improved method applies an electrical field along a contour of a hexagonal array of electrodes powered to generate two alternating electric field vectors (clamped homogeneous electric fields) (117). The electrical field pulsed in different directions across the gel allows linear pieces of DNA to "snake through" the path of least resistance along its migration in the agarose gel matrix (Fig. 2).

The second problem was solved by embedding bacterial cells in an agarose gel plug to enzymatically lyse and digest the bacterial cell wall and proteins inside the gel. This releases chromosomal DNA into the plug. The DNA fragments produced from the digested chromosome inside the plug with a rare-cutter restriction enzyme will be protected from autodigestion (118, 119). The plug can then be loaded directly into the well of an electrophoresis gel.

PFGE fingerprints represent changes that occur at endonuclease recognition sites (point mutation leading to loss or gain of a restriction site, insertion of DNA into a restriction site, or deletion of DNA from a fragment). Therefore, such information can be used to infer strain genetic relatedness. Tenover et al. proposed a set of guidelines for interpreting DNA restriction patterns generated by PFGE (Tenover's criteria) (26, 120). The guidelines are based on correlations of bacterial strain PFGE typing results with epidemiologic data obtained from dozens of outbreak investigations (120). Strains are considered genetically "indistinguishable" if their PFGE patterns show the same number and size of DNA fragments. They are considered "closely related" if the PFGE patterns differ from a reference strain (e.g., outbreak strain) by changes consistent with a single genetic event (a point mutation, insertion, or deletion at a restriction site). Such patterns will differ from the reference strain pattern by 2 or 3 bands. If the PFGE patterns differ from a reference strain pattern by 4 to 6 bands, they are considered to have undergone two independent genetic events and therefore be "possibly related." If the strains have >6 band differences compared to a reference strain pattern, they are considered "different" and genetically unrelated (26, 120).

It should be noted that Tenover's criteria apply only to the interpretation of PFGE patterns and not to electrophoretic patterns generated from other genotyping tests described here. The number of genetic events cannot be inferred from electrophoretic band patterns based on plasmid, repetitive DNA element, or AP-PCR/RAPD genotyping tests. Another important point is that the

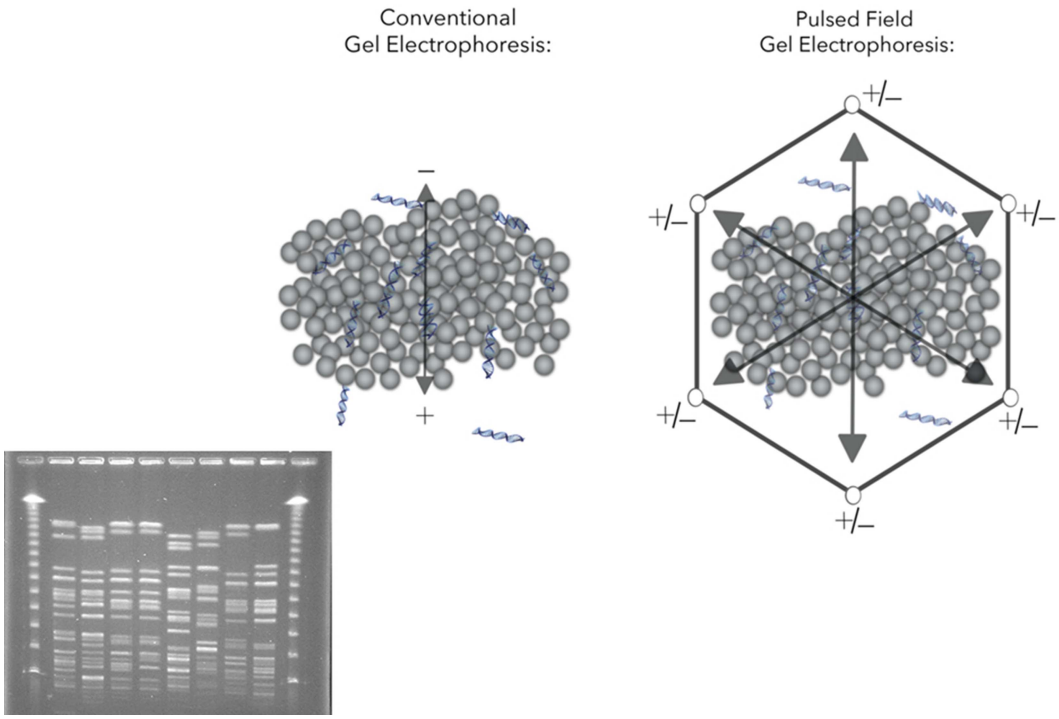

FIGURE 2 Conventional gel electrophoresis (left) and pulsed-field gel electrophoresis (PFGE) (right). In conventional gel electrophoresis, negatively charged DNA fragments migrate in one direction towards the positively charged electrode. In PFGE, the electrical field is applied along a contour of a hexagonal array of electrodes powered to generate two alternating electric field vectors, which allows linear pieces of DNA to "snake through" the path of least resistance along its migration in the agarose gel matrix. This allows large DNA fragments to be resolved to generate band patterns. (Illustrated by Paolo Harris Paz.) (Photo inset) PFGE of extraintestinal pathogenic *E. coli* isolates. (Photo by Meena Ramchandani.)

PFGE patterns of a set of test isolates must be compared to a reference strain pattern to satisfy Tenover's criteria (121). Patterns differing by 1 to 6 bands among isolates from unknown or uncharacterized sources cannot be concluded to represent closely related or possibly related strains. Strains that can be used as a reference include (i) a well-characterized epidemic or outbreak strain or (ii), if the outbreak occurrence is not certain, a predominant strain in limited institutional (e.g., hospital), geographic (e.g., community), and temporal settings identified by a surveillance system.

PFGE can be applied to genotype any cultivatable bacterial organisms without prior knowledge of their nucleic acid sequences. In general, it is more discriminating than other electrophoresis-based tests. A recent study comparing WGS to PFGE demonstrated discordance between the two tests in typing strains of vancomycin-resistant *Enterococcus faecium* (VRE), MRSA, and *Acinetobacter baumannii* (122). The level and direction of discordance varied according to bacterial species. Of 64 isolates of VRE shown to be indistinguishable by PFGE, 9 (14%) were found to be nonclonal by WGS; all 18 VRE isolates distinguished by PFGE were also distinguished (nonclonal) by WGS. Of 20 MRSA isolates found to be indistinguishable by PFGE, 15 (75%) were nonclonal by WGS; all 35 PFGE-distinguished isolates were concordant with WGS. Of 6 *A. baumannii* isolates deemed indistinguishable by PFGE, 2 (33%) were nonclonal by WGS and 28 isolates were concordant by both tests (122). All isolates of VRE and MRSA found by PFGE to be closely or possibly related (by Tenover's criteria [26]) were found to be nonclonal by WGS, but 16 (23%) of 71 *A. baumannii* isolates shown to be clonal by WGS were found to be either closely or possibly related by PFGE (differed by 1 to 6 bands) (26). In general, if the PFGE patterns are shown to represent unrelated strains (>6 band difference), WGS will also show them to be nonclonal.

PFGE has been in use for many years by a national foodborne disease surveillance system called PulseNet-

USA (or PulseNet-Central), launched in 1996 as a collaboration among the CDC, Association of Public Health Laboratories, and food regulatory agencies, including the U.S. Department of Agriculture (USDA) and the Food and Drug Administration (FDA) (https://www.cdc.gov/pulsenet/). By 2001 PulseNet had become a nationwide surveillance system covering all 50 state public health laboratories. It currently includes 83 laboratories in seven regions of the United States. Pathogens typed by PulseNet-USA include *Salmonella* spp., Shiga toxin-producing *E. coli*, *Listeria monocytogenes*, *Campylobacter jejuni*, *Cronobacter* (formerly *Enterobacter sakazakii*), *Shigella* spp., *Vibrio cholerae*, and *Vibrio parahaemolyticus*. This surveillance system identifies about 1,500 clusters of foodborne disease and approximately 30 multistate or national outbreaks each year. CDC's PulseNet Central identifies and follows, on average, 280 clusters of foodborne disease each year.

PulseNet has expanded internationally over the last 20 years. PulseNet International now includes 83 countries in Africa, the Asia Pacific, Canada, Europe, Latin America and the Caribbean, the Middle East, and the United States.

What has been the public health impact of PulseNet? One of the most effective uses of this surveillance system has been its ability to precisely pinpoint contaminated food products, which then triggers product recall before they are consumed by the public. It has had a major impact, particularly on interstate-distributed food products. The CDC estimates that more than 1 billion pounds of contaminated food have been recalled since the inception of this surveillance system (https://www.cdc.gov/pulsenet/). Such recalls have prevented many local and nationwide foodborne disease outbreaks. It has also been used to identify pathogens included in the surveillance system linked to nonfood sources, such as pets, pet food, petting zoos, reptiles, and recreational bodies of water. Information obtained from such findings improves our understanding of the epidemiology of these pathogens, which creates an opportunity to devise new intervention and preventive strategies not considered previously.

A recent economic impact study found that nearly 270,000 illnesses from *Salmonella*, 9,500 illnesses from *E. coli*, and 60 illnesses due to *Listeria monocytogenes* are avoided annually because of this PFGE-based surveillance system (123). This translates to annual savings of medical and productivity costs of $507 million. Annual costs to run the PulseNet by public health agencies are $7.3 million (123).

The PulseNet surveillance system is the first foodborne-disease surveillance system in the world based on genotypic stratification of bacterial pathogens. It provides a proof of concept that a genotype-based surveillance system can make major new contributions to public health control of human illnesses. Despite its proven utility, however, it has its limitations. Reliance on analysis and comparison of electrophoretic band patterns requires strict adherence to protocols by all participating laboratories. It requires a specialized reference laboratory to ensure standardization of the procedures and maintenance of the PFGE patterns and their associated meta-database. Reproducibility remains a concern outside of reference laboratories when PFGE is applied to type certain bacterial species (124–126). The increasing use and application of WGS is beginning to either complement or replace PFGE tests for surveillance purposes. The CDC and other national and regional public health laboratories have begun this process and will ultimately replace PFGE with WGS for their genotype-based foodborne (as well as other infectious disease)-pathogen surveillance systems.

Restriction fragment length polymorphism-hybridization analysis

Another approach to facilitate ready interpretation of restriction enzyme-digested DNA fragments on a gel is based on targeting only a sequence of interest among the many DNA fragments. A probe complementary to a sequence within a DNA fragment of interest will bind (hybridize) to its target sequence, which is referred to as Southern blot hybridization, named after its developer (127). Southern blot hybridization requires target DNA fragments to be bound to a piece of nitrocellulose or nylon membrane. The DNA fragments can be transferred to the membrane directly from a gel containing the resolved fragments or be blotted onto the membrane into an array with a dot or slot blot apparatus. The DNA probe is conjugated to a reporter that facilitates visual detection of a selected target DNA fragment. A fingerprint or a matrix pattern is thus revealed among thousands of invisible DNA fragments if there is more than one copy of the target sequence in the membrane. Such patterns can then be compared across strains.

This approach has been widely used to genotype *M. tuberculosis* and until the advent of WGS analysis remained the reference standard for molecular epidemiologic investigations of tuberculosis (128). This approach is called IS*6110*-RFLP analysis, which combines REA of the *M. tuberculosis* genome and Southern blot hybridization targeting IS*6110* (mentioned above).

M. tuberculosis contains variable copy numbers of IS*6110* distributed in different locations of the chromosome (129–132). The genome is digested with the enzyme *Pvu*II, which generates hundreds of DNA fragments. IS*6110* contains one *Pvu*II recognition sequence and hence is cut only once. A probe is hybridized to the IS*6110* fragment (128). Thus, depending on the copy number and location of IS*6110* in the genome, a fingerprint representing an *M. tuberculosis* strain can be produced and compared across strains. Applications of IS*6110*-RFLP analysis to investigate epidemiology of tuberculosis are discussed in a separate review.

Nucleic Acid Hybridization Array

While the function of most bacterial interspersed repetitive elements described earlier is not well established, in *M. tuberculosis*, the DR site that occurs as a single cluster is actually a CRISPR (clustered regularly interspaced short palindromic repeats) locus (133–135). CRISPRs serve as a type of bacterial immune system designed to sense and eliminate incoming DNA, such as phage and plasmid DNA (134, 136). In *M. tuberculosis*, this locus serves as the target of a genotyping test called spacer oligonucleotide typing (spoligotyping).

Spoligotyping is an example of a genotyping method based on patterns visualized on a hybridization array. It is used to genotype *M. tuberculosis* (91). Synthetic oligonucleotides representing 43 spacers in the CRISPR locus of reference strain *M. tuberculosis* H37Rv and *Mycobacterium bovis* BCG are covalently linked to a membrane in an arrayed matrix. These oligonucleotides bound to the membrane serve as probes (reverse blot hybridization). The DR sequence and spacers between DRs in extracted *M. tuberculosis* DNA are amplified by PCR with primers that anneal to the DR sequence. One of the primers is linked to biotin. The amplified products are then hybridized to the oligonucleotides on the membrane. The spacer sequences vary by strain. Thus, after incubating the membrane with streptavidin conjugated to peroxidase (or any other type of reporter molecule), a hybridization pattern that differs by strain can be visualized (91). Strain polymorphism results from differences in the number and sequences of the spacers. The hybridization membrane containing the oligonucleotides is commercially available. Quality control of the membrane is important, as there may be membrane-to-membrane variation in the amount of synthetic oligonucleotides covalently linked to the membrane.

One important advantage of spoligotyping is the ease of analysis, interpretation, and interlaboratory comparison of the patterns of hybridization of the 43 spacers in the membrane. A rational numerical system that depicts the degree of similarity between the patterns was proposed (137). This step has facilitated creation of international, publicly accessible spoligotype databases (138–140). SpolDB was an early such database, but it has been largely replaced by SITVITWEB, which maintains these numerical representations as spoligotype international types (SITs) of clinical isolates of *M. tuberculosis* (http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE/). An algorithm called SPOTCLUSTER can be accessed online (http://tbinsight.cs.rpi.edu/run_spotclust.html) to assign the lineage of a newly genotyped strain. As of December 2016, the SITVITWEB database contained 2,747 SITs among 58,187 isolates representing 105 countries.

Despite these advantages, spoligotyping has limited epidemiologic utility. It does not have the resolution of IS*6110*-RFLP or the 24-locus MIRU-VNTR genotyping tests (discussed below). Major spoligotype clades include the East African-Indian (EAI), Beijing, Haarlem, Latin American and Mediterranean, Central Asian, European clade of IS*6110* low banders, and X clades (140). Those that do not belong to these clades are called orphan clades. Studies based on *M. tuberculosis* spoligotypes have shown global and regional differences in the prevalence of these clades, but they have not contributed much to our understanding of specific sources, dynamics, or risk factors for tuberculosis transmission. Much is already well known about the global distribution of spoligotypes, and additional such studies are not likely to reveal anything more informative.

Many other bacterial pathogens containing CRISPR sequences can be genotyped as described above (141–147). The publicly accessible database can be queried (CRISPRdb [http://crispr.i2bc.paris-saclay.fr/crispr/]) for CRISPR sequences to develop such a test for any bacterial organism containing these repetitive DNA sequences (141).

Other hybridization-based genotyping tests

The spoligotyping membrane contains only 43 spacer sequences of *M. tuberculosis* CRISPR locus. Miniaturization of hybridization arrays using semiconductor or photolithography technology has made it possible to spot thousands of oligonucleotide probes onto a substrate, which came to be known as microarray or "gene chip" (148–150). An automated arrayer can be used to spot PCR-amplified products onto a glass slide surface to custom-construct a microarray. Before NGS technologies came to be widely used, these microarrays were used to analyze (i) variations based on single nucleotide

polymorphisms in a set of genes of interest (genotyping), (ii) mRNA transcripts (transcriptional profiling), and (iii) whole genomes of different bacterial strains (comparative genome profiling). However, microarrays are limited by their inconsistent detection or quantification of differentially expressed genes, nonspecific cross-hybridization, inability to identify allele polymorphism, and requirement of prior knowledge of nucleic acid sequences to construct the oligonucleotides. For these reasons, microarrays are rapidly being replaced by WGS for genotyping and comparative genome profiling, or by RNA sequence analysis for microarray transcriptional profiling (151–154). Applications of these NGS tools in epidemiologic investigations of infectious disease are discussed in other reviews of this series.

Nucleic Acid Sequence-Based Genotyping Tests

Multilocus sequence typing

The representative and most widely used nucleic acid sequence-based genotyping test is MLST (155–157). MLST is based on comparison of sequence polymorphism within short segments (400 to 600 bp) among a set of housekeeping genes. The target loci are amplified by PCR and then the PCR product is sequenced. Since each of these target genes can have multiple alleles, all possible combinations of these sequences will generate many allele profiles or sequence types (STs). For example, in the MLST scheme for *Staphylococcus aureus*, the genes that are targeted include *arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi*, and *yqiL*. A locus from each of these genes is amplified by PCR and the product is sequenced. The sequence data are then deposited into a web-accessible database, which has an automated submission system that allows the database curators to provide an ST designation (ST number) for the deposited allelic profile data. Each ST assignment is based on allelic profiles of the 7 loci in which each allele sequence is assigned a unique number in order of its discovery. For example, a pandemic lineage of *S. aureus* ST8 (also called USA300 by PFGE) has the allelic profile *arc*-3, *aroE*-3, *glpF*-1, *gmk*-1, *pta*-4, *tpi*-4, and *yqi*L-3. No weighting is given based on the number of nucleotide differences between alleles.

The sequences from all 7 loci can be concatenated to represent an *S. aureus* genotype (Fig. 3). The concatenated sequences from multiple strains can be aligned to construct a dendrogram. Since each of these gene loci could have 18 to 33 alleles per locus, all different combinations of the sequences could generate billions of possible concatenated sequences. The concatenated sequences, of course, are not found as such in the genomes, but the dendrograms based on these concatenated sequences are considered to reflect evolutionary relationships of the genotyped strains. This type of phylogenetic analysis based on comparison of concatenated sequences is called multilocus sequence analysis, and it is most often used to compare sequences of bacterial species in which recombination events do not occur frequently.

MLST data can also be phylogenetically represented based on pairwise comparison of the allelic profiles. A minimal spanning tree algorithm, such as eBURST (http://eburst.mlst.net/default.asp), is designed to identify a putative founding genotype from a group of MLST allelic profiles and link other genotypes predicted to descend from the founding genotype. Of course, it is impossible to determine the true evolutionary descent of strains among a collection of bacterial isolates, but such pairwise comparisons to represent strain relatedness provide a hypothetical model that can be used to infer epidemiologic relationships, especially when such isolates are obtained from cases that are not part of any obvious outbreaks.

MLST is an example of microdiversity analysis since it is based on sequence differences observed in just 7 genes. The question is whether such differences can be used to infer strain phylogenetic relatedness. A large-scale WGS analysis of 312 extraintestinal pathogenic *E. coli* isolates showed *E. coli* MLST clonal structures to match closely with the phylogenomic structure based on WGS (20). Thus, MLST provides high discriminatory power and portability and is not affected by variability in test conditions.

One site (PubMLST) hosted by Oxford University (http://pubmlst.org/) contains MLST schemes for bacterial and fungal species, as well as for bacteriophages and plasmids. Other sites include MLST (http://www.mlst.net/), which has 29 schemes hosted by Imperial College, and EnteroBase (http://enterobase.warwick.ac.uk/), hosted by Warwick Medical School, which curates data for *Enterobacteriaceae* members *E. coli/Shigella*, *Salmonella*, *Yersinia*, and *Moraxella*. Today, MLST data can also be extracted from WGS databases, precluding performing PCRs on the housekeeping gene loci. MLST derived *in silico* from such sources are called whole-genome MLSTs.

Variants of MLST are described that are based on amplifying and sequencing a set of gene loci other than those of housekeeping genes included in the databases mentioned above. One such test, called multivirulence sequence typing, targets three virulence genes (*prfA*,

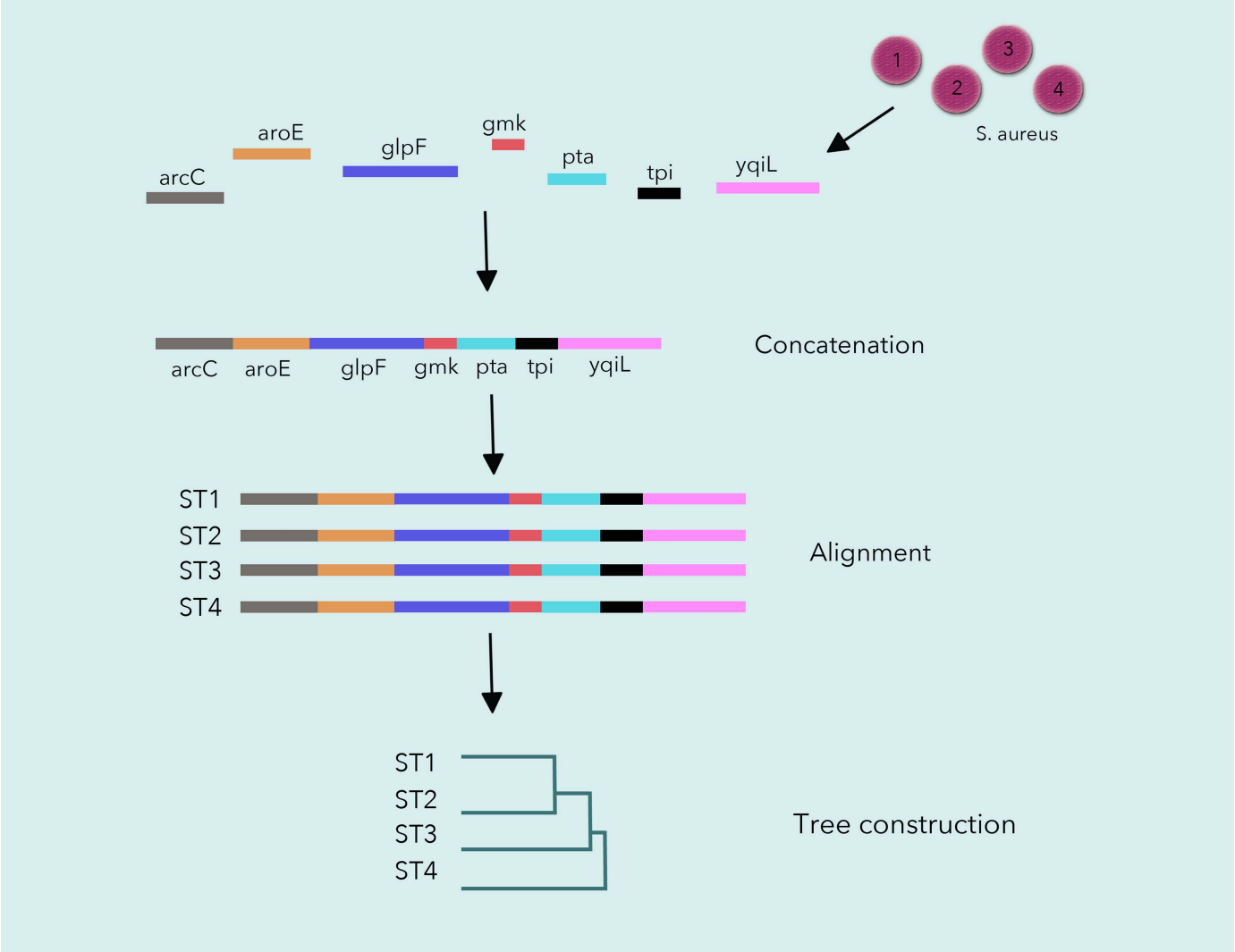

FIGURE 3 Multilocus sequence typing (MLST) of *Staphylococcus aureus*. A 400- to 600-bp region of housekeeping genes (*arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi*, and *yqiL)* is amplified by PCR and the amplified products are sequenced. The sequences of the 7 loci are then concatenated, and all the concatenated sequences representing different genotypes are aligned to generate a tree, which can be used to depict strain relatedness. The analysis of similarity by comparison of concatenated sequence is called multilocus sequence analysis. The allele sequences can be uploaded onto an automated submission system (e.g., http://pubmlst.org/ or http://www.mlst.net/), which allows the database curators to provide an ST designation (ST number). Pairwise comparison of allele profiles can be performed by minimal spanning tree analysis using algorithms such as eBURST (http://eburst.mlst.net/default.asp). (Illustrated by Paolo Harris Paz.)

inlB, and *inlC*) and three virulence-associated genes (*dal*, *lisR*, and *clpP*) to genotype *Listeria monocytogenes* (158). It was shown to be more discriminating than PFGE using *Apa*I restriction enzyme or MLST based on the PubMLST scheme (158).

Comparison of levels of resolution achieved with MLST and PFGE shows mixed results. One study found MLST to have a higher discriminatory ability than PFGE for genotyping extended-spectrum-β-lactamase-producing *E. coli* (159), while another study showed further differentiation by PFGE of uropathogenic *E. coli* isolates belonging to ST69 (160). Applications of MLST in epidemiologic investigations of specific infectious diseases are discussed in other reviews.

Next-generation sequence-based genotyping tests

NGS technology can be divided into three major categories—targeted sequencing, WGS, and metagenomics ("deep sequencing"). Targeted sequencing compares sequences of specific genes (e.g., 16S rDNA sequences) to determine strain relatedness. Targeted sequencing is primarily used in phylogenetics or molecular evolution research to profile phylogenetic or taxonomic diversity of microbes. Information obtained from such analyses could be used for epidemiologic investigations. Early applications of WGS created a discipline called comparative genomics, but today, the large number of genome sequences in publicly accessible databases and the ease, speed, and decreasing cost of WGS have made it possible to apply WGS to epidemiologic investigations of a wide range of infectious diseases. Later reviews will discuss these new applications. Basic concepts and platforms of NGS technology are discussed in a separate review of this series.

One major limitation of the WGS technology as well as most of the genotyping tests described earlier in this review is that they have to be performed on cultured bacterial specimens. Cultured specimens ensure the purity of the DNA to be analyzed. A metagenomics approach may soon provide a way to overcome this limitation. An approach called meta-MLST was recently reported to predict a social network within a university (161). Salivary samples from study subjects were directly examined for *Streptococcus viridans*, a ubiquitous member of the human oral microflora. Seven housekeeping gene loci in *S. viridans* were amplified by PCR directly from salivary samples and were subjected to metagenomics analysis. Each subject's quality-filtered reads were aligned to an index created from nearly 15 million unique reads, and their pairwise comparisons were used to construct a social network tree (161). The tree closely matched the social network determined by interviews (161). Thus, this assay enabled simultaneous comparison of multiple alleles of these gene loci from multiple strains directly from salivary samples obtained from several study subjects.

The applications of these NGS-based tests in epidemiologic investigations are discussed in other reviews of this series on molecular epidemiology of infectious diseases.

CONCLUDING REMARKS

Ever since microbiologists first discovered how to grow bacteria in artificial medium in pure culture, bacterial strain subtyping methods have undergone evolution of their own. As new advancements in technology were made, the resolution of units of analysis became greater. Early technologies were used to differentiate bacterial organisms based on their phenotypic characteristics. With the advent of molecular microbiology, these organisms could be differentiated into genotypes. Comparison of genotypes made possible the establishment of molecular epidemiology as a subdiscipline of epidemiology. NGS technology will drive the field of next-generation molecular epidemiology. However, data generated from NGS technology are not as analytically approachable or friendly as are data generated from early-generation genotyping technologies. A separate review in this series will discuss advancements made in NGS platforms and computational biology and how they may be applied to next-generation epidemiology.

CURATED COLLECTION

Click here to read other articles in this collection.

REFERENCES

1. **Cronquist AB, Mody RK, Atkinson R, Besser J, Tobin D'Angelo M, Hurd S, Robinson T, Nicholson C, Mahon BE.** 2012. Impacts of culture-independent diagnostic practices on public health surveillance for bacterial enteric pathogens. *Clin Infect Dis* **54**(Suppl 5):S432–S439. http://dx.doi.org/10.1093/cid/cis267.

2. **Rogers GB, Carroll MP, Bruce KD.** 2009. Studying bacterial infections through culture-independent approaches. *J Med Microbiol* **58**:1401–1418. http://dx.doi.org/10.1099/jmm.0.013334-0.

3. **Iwamoto M, Huang JY, Cronquist AB, Medus C, Hurd S, Zansky S, Dunn J, Woron AM, Oosmanally N, Griffin PM, Besser J, Henao OL, Centers for Disease Control and Prevention.** 2015. Bacterial enteric infections detected by culture-independent diagnostic tests–FoodNet, United States, 2012–2014. *MMWR Morb Mortal Wkly Rep* **64**:252–257.

4. **Jorgensen JHPM, Carroll KC, Funke G, Landry ML, Richter SS, Warnock DW (ed).** 2015. *Manual of Clinical Microbiology*, 11th ed. ASM Press, Washington, DC.

5. **Whitman WGM, Kämpfer P, Busse H-J, Trujillo M, Ludwig W, Suzuki K, Parte A.** 2012. *Bergey's Manual of Systematic Bacteriology*, 2nd ed. Springer-Verlag, New York, NY.

6. **Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky MI, Moore LH, Moore WEC, Murray RGE, Stackerbrandt E, Starr MP, Truper HG.** 1987. Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *Int J Syst Evol Microbiol* **37**:463–464. http://dx.doi.org/10.1099/00207713-37-4-463.

7. **Goebel BM, Stackebrandt E.** 1994. Cultural and phylogenetic analysis of mixed microbial populations found in natural and commercial bioleaching environments. *Appl Environ Microbiol* **60**:1614–1621.

8. **Konstantinidis KT, Tiedje JM.** 2007. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr Opin Microbiol* **10**:504–509. http://dx.doi.org/10.1016/j.mib.2007.08.006.

9. **Brenner FW, Villar RG, Angulo FJ, Tauxe R, Swaminathan B.** 2000. *Salmonella* nomenclature. *J Clin Microbiol* **38**:2465–2467.

10. **Edwards PR, Kauffmann F.** 1952. A simplification of the Kauffmann-White schema. *Am J Clin Pathol* **22**(7_ts):692–697. http://dx.doi.org/10.1093/ajcp/22.7_ts.692.

11. **Crosa JH, Brenner DJ, Ewing WH, Falkow S.** 1973. Molecular relationships among the Salmonelleae. *J Bacteriol* **115**:307–315.

12. Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* **102:**2567–2572. http://dx.doi.org/10.1073/pnas.0409727102.

13. Welch RA, Burland V, Plunkett G, III, Redford P, Roesch P, Rasko D, Buckles EL, Liou SR, Boutin A, Hackett J, Stroud D, Mayhew GF, Rose DJ, Zhou S, Schwartz DC, Perna NT, Mobley HL, Donnenberg MS, Blattner FR. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* **99:**17020–17024. http://dx.doi.org/10.1073/pnas.252529799.

14. Janda JM, Abbott SL. 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol* **45:**2761–2764. http://dx.doi.org/10.1128/JCM.01228-07.

15. Schleifer KH. 2009. Classification of Bacteria and Archaea: past, present and future. *Syst Appl Microbiol* **32:**533–542. http://dx.doi.org/10.1016/j.syapm.2009.09.002.

16. Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. 2005. The microbial pan-genome. *Curr Opin Genet Dev* **15:**589–594. http://dx.doi.org/10.1016/j.gde.2005.09.006.

17. Tettelin H, Riley D, Cattuto C, Medini D. 2008. Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol* **11:**472–477. http://dx.doi.org/10.1016/j.mib.2008.09.006.

18. Rouli L, Merhej V, Fournier PE, Raoult D. 2015. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect* **7:**72–85. http://dx.doi.org/10.1016/j.nmni.2015.06.005.

19. Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, Gajer P, Crabtree J, Sebaihia M, Thomson NR, Chaudhuri R, Henderson IR, Sperandio V, Ravel J. 2008. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J Bacteriol* **190:**6881–6893. http://dx.doi.org/10.1128/JB.00619-08.

20. Salipante SJ, Roach DJ, Kitzman JO, Snyder MW, Stackhouse B, Butler-Wu SM, Lee C, Cookson BT, Shendure J. 2015. Large-scale genomic sequencing of extraintestinal pathogenic *Escherichia coli* strains. *Genome Res* **25:**119–128. http://dx.doi.org/10.1101/gr.180190.114.

21. Kaas RS, Friis C, Ussery DW, Aarestrup FM. 2012. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. *BMC Genomics* **13:**577. http://dx.doi.org/10.1186/1471-2164-13-577.

22. Spratt BG. 2004. Exploring the concept of clonality in bacteria. *Methods Mol Biol* **266:**323–352.

23. Riley LW. 2004. *Molecular Epidemiology of Infectious Diseases: Principles and Practices*. ASM Press, Washington, DC. http://dx.doi.org/10.1128/9781555817688.

24. Lindsay JA. 2014. Evolution of *Staphylococcus aureus* and MRSA during outbreaks. *Infect Genet Evol* **21:**548–553. http://dx.doi.org/10.1016/j.meegid.2013.04.017.

25. Maslow JN, Mulligan ME, Arbeit RD. 1993. Molecular epidemiology: application of contemporary techniques to the typing of microorganisms. *Clin Infect Dis* **17:**153–162; quiz, 163–164.

26. Tenover FC, Arbeit RD, Goering RV, Molecular Typing Working Group of the Society for Healthcare Epidemiology of America. 1997. How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. *Infect Control Hosp Epidemiol* **18:**426–439. http://dx.doi.org/10.2307/30141252.

27. Lipuma JJ. 1998. Molecular tools for epidemiologic study of infectious diseases. *Pediatr Infect Dis J* **17:**667–675. http://dx.doi.org/10.1097/00006454-199808000-00002.

28. Sabat AJ, Budimir A, Nashev D, Sá-Leão R, van Dijl J, Laurent F, Grundmann H, Friedrich AW, ESCMID Study Group of Epidemiological Markers. 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro Surveill* **18:**20380. http://dx.doi.org/10.2807/ese.18.04.20380-en.

29. Schaberg DR, Tompkins LS, Falkow S. 1981. Use of agarose gel electrophoresis of plasmid deoxyribonucleic acid to fingerprint gram-negative bacilli. *J Clin Microbiol* **13:**1105–1108.

30. Holmberg SD, Wachsmuth IK, Hickman-Brenner FW, Cohen ML. 1984. Comparison of plasmid profile analysis, phage typing, and antimicrobial susceptibility testing in characterizing *Salmonella typhimurium* isolates from outbreaks. *J Clin Microbiol* **19:**100–104.

31. Farrar WE, Jr. 1983. Molecular analysis of plasmids in epidemiologic investigation. *J Infect Dis* **148:**1–6. http://dx.doi.org/10.1093/infdis/148.1.1.

32. Taylor DN, Wachsmuth IK, Shangkuan YH, Schmidt EV, Barrett TJ, Schrader JS, Scherach CS, McGee HB, Feldman RA, Brenner DJ. 1982. Salmonellosis associated with marijuana: a multistate outbreak traced by plasmid fingerprinting. *N Engl J Med* **306:**1249–1253. http://dx.doi.org/10.1056/NEJM198205273062101.

33. Kapperud G, Gustavsen S, Hellesnes I, Hansen AH, Lassen J, Hirn J, Jahkola M, Montenegro MA, Helmuth R. 1990. Outbreak of *Salmonella typhimurium* infection traced to contaminated chocolate and caused by a strain lacking the 60-megadalton virulence plasmid. *J Clin Microbiol* **28:**2597–2601.

34. Spika JS, Waterman SH, Hoo GWS, St. Louis ME, Pacer RE, James SM, Bissett ML, Mayer LW, Chiu JY, Hall B, Greene K, Potter ME, Cohen ML, Blake PA. 1987. Chloramphenicol-resistant Salmonella newport traced through hamburger to dairy farms. A major persisting source of human salmonellosis in California. *N Engl J Med* **316:**565–570. http://dx.doi.org/10.1056/NEJM198703053161001.

35. Lin FY, Morris JG, Jr, Trump D, Tilghman D, Wood PK, Jackman N, Israel E, Libonati JP. 1988. Investigation of an outbreak of Salmonella enteritidis gastroenteritis associated with consumption of eggs in a restaurant chain in Maryland. *Am J Epidemiol* **128:**839–844. http://dx.doi.org/10.1093/oxfordjournals.aje.a115037.

36. Riley LW, DiFerdinando GT, Jr, DeMelfi TM, Cohen ML. 1983. Evaluation of isolated cases of salmonellosis by plasmid profile analysis: introduction and transmission of a bacterial clone by precooked roast beef. *J Infect Dis* **148:**12–17. http://dx.doi.org/10.1093/infdis/148.1.12.

37. Threlfall EJ, Hampton MD, Chart H, Rowe B. 1994. Use of plasmid profile typing for surveillance of Salmonella enteritidis phage type 4 from humans, poultry and eggs. *Epidemiol Infect* **112:**25–31. http://dx.doi.org/10.1017/S0950268800057381.

38. Dorn CR, Silapanuntakul R, Angrick EJ, Shipman LD. 1993. Plasmid analysis of Salmonella enteritidis isolated from human gastroenteritis cases and from epidemiologically associated poultry flocks. *Epidemiol Infect* **111:**239–243. http://dx.doi.org/10.1017/S0950268800056946.

39. O'Brien TF, Hopkins JD, Gilleece ES, Medeiros AA, Kent RL, Blackburn BO, Holmes MB, Reardon JP, Vergeront JM, Schell WL, Christenson E, Bissett ML, Morse EV. 1982. Molecular epidemiology of antibiotic resistance in salmonella from animals and human beings in the United States. *N Engl J Med* **307:**1–6. http://dx.doi.org/10.1056/NEJM198207013070101.

40. Riley LW, Ceballos BS, Trabulsi LR, Fernandes de Toledo MR, Blake PA. 1984. The significance of hospitals as reservoirs for endemic multiresistant Salmonella typhimurium causing infection in urban Brazilian children. *J Infect Dis* **150:**236–241. http://dx.doi.org/10.1093/infdis/150.2.236.

41. Nakamura M, Sato S, Ohya T, Suzuki S, Ikeda S. 1986. Plasmid profile analysis in epidemiological studies of animal *Salmonella typhimurium* infection in Japan. *J Clin Microbiol* **23:**360–365.

42. Wray C, McLaren IM, Jones YE. 1998. The epidemiology of Salmonella typhimurium in cattle: plasmid profile analysis of definitive phage type (DT) 204c. *J Med Microbiol* **47:**483–487. http://dx.doi.org/10.1099/00222615-47-6-483.

43. Ostroff SM, Griffin PM, Tauxe RV, Shipman LD, Greene KD, Wells JG, Lewis JH, Blake PA, Kobayashi JM. 1990. A statewide outbreak of *Escherichia coli* O157:H7 infections in Washington State. *Am*

J Epidemiol **132:**239–247. http://dx.doi.org/10.1093/oxfordjournals.aje.a115653.

44. Tornieporth NG, John J, Salgado K, de Jesus P, Latham E, Melo MC, Gunzburg ST, Riley LW. 1995. Differentiation of pathogenic *Escherichia coli* strains in Brazilian children by PCR. *J Clin Microbiol* **33:**1371–1374.

45. Gunzburg ST, Tornieporth NG, Riley LW. 1995. Identification of enteropathogenic *Escherichia coli* by PCR-based detection of the bundle-forming pilus gene. *J Clin Microbiol* **33:**1375–1377.

46. Girón JA, Donnenberg MS, Martin WC, Jarvis KG, Kaper JB. 1993. Distribution of the bundle-forming pilus structural gene (bfpA) among enteropathogenic *Escherichia coli*. *J Infect Dis* **168:**1037–1041. http://dx.doi.org/10.1093/infdis/168.4.1037.

47. Sethabutr O, Venkatesan M, Murphy GS, Eampokalap B, Hoge CW, Echeverria P. 1993. Detection of shigellae and enteroinvasive *Escherichia coli* by amplification of the invasion plasmid antigen H DNA sequence in patients with dysentery. *J Infect Dis* **167:**458–461. http://dx.doi.org/10.1093/infdis/167.2.458.

48. Moseley SL, Hardy JW, Hug MI, Echeverria P, Falkow S. 1983. Isolation and nucleotide sequence determination of a gene encoding a heat-stable enterotoxin of *Escherichia coli*. *Infect Immun* **39:**1167–1174.

49. Blomén ILS, Stenström TA, Norberg R. 1993. Identification of enterotoxigenic *Escherichia coli* isolates: a comparison of PCR, DNA hybridization, ELISA and bioassays. *J Microbiol Methods* **17:**181–191. http://dx.doi.org/10.1016/0167-7012(93)90045-J.

50. Berg DEAN, Kersulyte D. 1994. Fingerprinting microbial genomes using the RAPD or AP-PCR method. *Methods Mol Cell Biol* **5:**13–24.

51. Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* **18:**6531–6535. http://dx.doi.org/10.1093/nar/18.22.6531.

52. del Tufo JPTS, Tingey SV. 1994. RAPD assay. A novel technique for genetic diagnostics. *Methods Mol Biol* **28:**237–241.

53. Lin AW, Usera MA, Barrett TJ, Goldsby RA. 1996. Application of random amplified polymorphic DNA analysis to differentiate strains of *Salmonella enteritidis*. *J Clin Microbiol* **34:**870–876.

54. Welsh J, McClelland M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res* **18:**7213–7218. http://dx.doi.org/10.1093/nar/18.24.7213.

55. Berg DE, Akopyants NS, Kersulyte D. 1994. Fingerprinting microbial genomes using the RAPD or AP-PCR method. *Methods Mol Cell Biol* **5:**13–24.

56. Gillings M, Holley M. 1997. Amplification of anonymous DNA fragments using pairs of long primers generates reproducible DNA fingerprints that are sensitive to genetic variation. *Electrophoresis* **18:**1512–1518. http://dx.doi.org/10.1002/elps.1150180904.

57. Goh SH, Byrne SK, Zhang JL, Chow AW. 1992. Molecular typing of *Staphylococcus aureus* on the basis of coagulase gene polymorphisms. *J Clin Microbiol* **30:**1642–1645.

58. Sayada C, Denamur E, Orfila J, Catalan F, Elion J. 1991. Rapid genotyping of the *Chlamydia trachomatis* major outer membrane protein by the polymerase chain reaction. *FEMS Microbiol Lett* **67:**73–78. http://dx.doi.org/10.1111/j.1574-6968.1991.tb04392.x.

59. Heyndrickx M, Vauterin L, Vandamme P, Kersters K, DeVos P. 1996. Applicability of combined amplified ribosomal DNA restriction analysis (ARDRA) patterns in bacterial phylogeny and taxonomy. *J Microbiol Methods* **26:**247–259. http://dx.doi.org/10.1016/0167-7012(96)00916-5.

60. Massol-Deya AA, Odelson DA, Hickey RF, Tiedje JM. 1995. *Bacterial Community Fingerprinting of Amplified 16S and 16-23S Ribosomal DNA Gene Sequences and Restriction Endonuclease Analysis (ARDRA)*. Kluwer, Dordrecht, The Netherlands. http://dx.doi.org/10.1007/978-94-011-0351-0_20.

61. De Baere T, de Mendonça R, Claeys G, Verschraegen G, Mijs W, Verhelst R, Rottiers S, Van Simaey L, De Ganck C, Vaneechoutte M. 2002. Evaluation of amplified rDNA restriction analysis (ARDRA) for the identification of cultured mycobacteria in a diagnostic laboratory. *BMC Microbiol* **2:**4. http://dx.doi.org/10.1186/1471-2180-2-4.

62. Kimura R, Mandrell RE, Galland JC, Hyatt D, Riley LW. 2000. Restriction-site-specific PCR as a rapid test to detect enterohemorrhagic *Escherichia coli* O157:H7 strains in environmental samples. *Appl Environ Microbiol* **66:**2513–2519. http://dx.doi.org/10.1128/AEM.66.6.2513-2519.2000.

63. Torres AG, Payne SM. 1997. Haem iron-transport system in enterohaemorrhagic *Escherichia coli* O157:H7. *Mol Microbiol* **23:**825–833. http://dx.doi.org/10.1046/j.1365-2958.1997.2641628.x.

64. Janssen P, Coopman R, Huys G, Swings J, Bleeker M, Vos P, Zabeau M, Kersters K. 1996. Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. *Microbiology* **142:**1881–1893. http://dx.doi.org/10.1099/13500872-142-7-1881.

65. Huys G, Coopman R, Janssen P, Kersters K. 1996. High-resolution genotypic analysis of the genus *Aeromonas* by AFLP fingerprinting. *Int J Syst Bacteriol* **46:**572–580. http://dx.doi.org/10.1099/00207713-46-2-572.

66. Vos P, Hogers R, Bleeker M, Reijans M, Lee T, Hornes M, Friters A, Pot J, Paleman J, Kuiper M, Zabeau M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* **23:**4407–4414. http://dx.doi.org/10.1093/nar/23.21.4407.

67. Versalovic J, Koeuth T, Lupski JR. 1991. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Res* **19:**6823–6831. http://dx.doi.org/10.1093/nar/19.24.6823.

68. Versalovic JSM, de Bruijn FJ, Lupski JR. 1994. Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. *Methods Mol Cell Biol* **5:**25–40.

69. Lupski JR, Weinstock GM. 1992. Short, interspersed repetitive DNA sequences in prokaryotic genomes. *J Bacteriol* **174:**4525–4529. http://dx.doi.org/10.1128/jb.174.14.4525-4529.1992.

70. Stern MJ, Ames GF, Smith NH, Robinson EC, Higgins CF. 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* **37:**1015–1026. http://dx.doi.org/10.1016/0092-8674(84)90436-7.

71. Gilson E, Clément JM, Brutlag D, Hofnung M. 1984. A family of dispersed repetitive extragenic palindromic DNA sequences in E. coli. *EMBO J* **3:**1417–1421.

72. Hulton CS, Higgins CF, Sharp PM. 1991. ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. *Mol Microbiol* **5:**825–834. http://dx.doi.org/10.1111/j.1365-2958.1991.tb00755.x.

73. Sharples GJ, Lloyd RG. 1990. A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. *Nucleic Acids Res* **18:**6503–6508. http://dx.doi.org/10.1093/nar/18.22.6503.

74. Higgins CF, Ames GF, Barnes WM, Clement JM, Hofnung M. 1982. A novel intercistronic regulatory element of prokaryotic operons. *Nature* **298:**760–762. http://dx.doi.org/10.1038/298760a0.

75. Woese CR. 1987. Bacterial evolution. *Microbiol Rev* **51:**221–271.

76. Higgins CF, McLaren RS, Newbury SF. 1988. Repetitive extragenic palindromic sequences, mRNA stability and gene expression: evolution by gene conversion? A review. *Gene* **72:**3–14. http://dx.doi.org/10.1016/0378-1119(88)90122-9.

77. Shyamala V, Schneider E, Ames GF. 1990. Tandem chromosomal duplications: role of REP sequences in the recombination event at the join-point. *EMBO J* **9:**939–946.

78. Martin B, Humbert O, Camara M, Guenzi E, Walker J, Mitchell T, Andrew P, Prudhomme M, Alloing G, Hakenbeck R, Morrison DA, Boulnois GJ, Claverys J-P. 1992. A highly conserved repeated DNA ele-

ment located in the chromosome of *Streptococcus pneumoniae*. *Nucleic Acids Res* **20:**3479–3483. http://dx.doi.org/10.1093/nar/20.13.3479.

79. Koeuth T, Versalovic J, Lupski JR. 1995. Differential subsequence conservation of interspersed repetitive *Streptococcus pneumoniae* BOX elements in diverse bacteria. *Genome Res* **5:**408–418. http://dx.doi.org/10.1101/gr.5.4.408.

80. Del Vecchio VG, Petroziello JM, Gress MJ, McCleskey FK, Melcher GP, Crouch HK, Lupski JR. 1995. Molecular genotyping of methicillin-resistant *Staphylococcus aureus* via fluorophore-enhanced repetitive-sequence PCR. *J Clin Microbiol* **33:**2141–2144.

81. Georghiou PR, Doggett AM, Kielhofner MA, Stout JE, Watson DA, Lupski JR, Hamill RJ. 1994. Molecular fingerprinting of *Legionella* species by repetitive element PCR. *J Clin Microbiol* **32:**2989–2994.

82. Louws FJ, Fulbright DW, Stephens CT, de Bruijn FJ. 1994. Specific genomic fingerprints of phytopathogenic *Xanthomonas* and *Pseudomonas* pathovars and strains generated with repetitive sequences and PCR. *Appl Environ Microbiol* **60:**2286–2295.

83. Gillings M, Holley M. 1997. Repetitive element PCR fingerprinting (rep-PCR) using enterobacterial repetitive intergenic consensus (ERIC) primers is not necessarily directed at ERIC elements. *Lett Appl Microbiol* **25:**17–21. http://dx.doi.org/10.1046/j.1472-765X.1997.00162.x.

84. Koressaar T, Remm M. 2012. Characterization of species-specific repeats in 613 prokaryotic species. *DNA Res* **19:**219–230.

85. Koressaar T, Jõers K, Remm M. 2009. Automatic identification of species-specific repetitive DNA sequences and their utilization for detecting microbial organisms. *Bioinformatics* **25:**1349–1355. http://dx.doi.org/10.1093/bioinformatics/btp241.

86. Ross BC, Raios K, Jackson K, Dwyer B. 1992. Molecular cloning of a highly repeated DNA element from *Mycobacterium tuberculosis* and its use as an epidemiological tool. *J Clin Microbiol* **30:**942–946.

87. Supply P, Mazars E, Lesjean S, Vincent V, Gicquel B, Locht C. 2000. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Mol Microbiol* **36:**762–771. http://dx.doi.org/10.1046/j.1365-2958.2000.01905.x.

88. Mazars E, Lesjean S, Banuls AL, Gilbert M, Vincent V, Gicquel B, Tibayrenc M, Locht C, Supply P. 2001. High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc Natl Acad Sci U S A* **98:**1901–1906. http://dx.doi.org/10.1073/pnas.98.4.1901.

89. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, Savine E, de Haas P, van Deutekom H, Roring S, Bifani P, Kurepina N, Kreiswirth B, Sola C, Rastogi N, Vatin V, Gutierrez MC, Fauville M, Niemann S, Skuce R, Kremer K, Locht C, van Soolingen D. 2006. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* **44:**4498–4510. http://dx.doi.org/10.1128/JCM.01392-06.

90. van Soolingen D, de Haas PE, Hermans PW, Groenen PM, van Embden JD. 1993. Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. *J Clin Microbiol* **31:**1987–1995.

91. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J. 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* **35:**907–914.

92. Groenen PM, Bunschoten AE, van Soolingen D, van Embden JD. 1993. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol Microbiol* **10:**1057–1065. http://dx.doi.org/10.1111/j.1365-2958.1993.tb00976.x.

93. Frothingham R. 1995. Differentiation of strains in *Mycobacterium tuberculosis* complex by DNA sequence polymorphisms, including rapid identification of *M. bovis* BCG. *J Clin Microbiol* **33:**840–844.

94. Frothingham R, Meeker-O'Connell WA. 1998. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology* **144:**1189–1196. http://dx.doi.org/10.1099/00221287-144-5-1189.

95. van Belkum A. 1999. Short sequence repeats in microbial pathogenesis and evolution. *Cell Mol Life Sci* **56:**729–734. http://dx.doi.org/10.1007/s000180050019.

96. Hawkey PM, Smith EG, Evans JT, Monk P, Bryan G, Mohamed HH, Bardhan M, Pugh RN. 2003. Mycobacterial interspersed repetitive unit typing of *Mycobacterium tuberculosis* compared to IS6110-based restriction fragment length polymorphism analysis for investigation of apparently clustered cases of tuberculosis. *J Clin Microbiol* **41:**3514–3520. http://dx.doi.org/10.1128/JCM.41.8.3514-3520.2003.

97. Christianson S, Wolfe J, Orr P, Karlowsky J, Levett PN, Horsman GB, Thibert L, Tang P, Sharma MK. 2010. Evaluation of 24 locus MIRU-VNTR genotyping of *Mycobacterium tuberculosis* isolates in Canada. *Tuberculosis (Edinb)* **90:**31–38. http://dx.doi.org/10.1016/j.tube.2009.12.003.

98. Valcheva V, Mokrousov I, Narvskaya O, Rastogi N, Markova N. 2008. Utility of new 24-locus variable-number tandem-repeat typing for discriminating *Mycobacterium tuberculosis* clinical isolates collected in Bulgaria. *J Clin Microbiol* **46:**3005–3011. http://dx.doi.org/10.1128/JCM.00437-08.

99. Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, Jackson PJ, Hugh-Jones ME. 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J Bacteriol* **182:**2928–2936. http://dx.doi.org/10.1128/JB.182.10.2928-2936.2000.

100. Ramisse V, Houssu P, Hernandez E, Denoeud F, Hilaire V, Lisanti O, Ramisse F, Cavallo JD, Vergnaud G. 2004. Variable number of tandem repeats in *Salmonella enterica* subsp. *enterica* for typing purposes. *J Clin Microbiol* **42:**5722–5730. http://dx.doi.org/10.1128/JCM.42.12.5722-5730.2004.

101. Le Flèche P, Hauck Y, Onteniente L, Prieur A, Denoeud F, Ramisse V, Sylvestre P, Benson G, Ramisse F, Vergnaud G. 2001. A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. *BMC Microbiol* **1:**2. http://dx.doi.org/10.1186/1471-2180-1-2.

102. Farlow J, Smith KL, Wong J, Abrams M, Lytle M, Keim P. 2001. *Francisella tularensis* strain typing using multiple-locus, variable-number tandem repeat analysis. *J Clin Microbiol* **39:**3186–3192. http://dx.doi.org/10.1128/JCM.39.9.3186-3192.2001.

103. Onteniente L, Brisse S, Tassios PT, Vergnaud G. 2003. Evaluation of the polymorphisms associated with tandem repeats for *Pseudomonas aeruginosa* strain typing. *J Clin Microbiol* **41:**4991–4997. http://dx.doi.org/10.1128/JCM.41.11.4991-4997.2003.

104. Noller AC, McEllistrem MC, Stine OC, Morris JG, Jr, Boxrud DJ, Dixon B, Harrison LH. 2003. Multilocus sequence typing reveals a lack of diversity among *Escherichia coli* O157:H7 isolates that are distinct by pulsed-field gel electrophoresis. *J Clin Microbiol* **41:**675–679. http://dx.doi.org/10.1128/JCM.41.2.675-679.2003.

105. Pourcel C, André-Mazeaud F, Neubauer H, Ramisse F, Vergnaud G. 2004. Tandem repeats analysis for the high resolution phylogenetic analysis of *Yersinia pestis*. *BMC Microbiol* **4:**22. http://dx.doi.org/10.1186/1471-2180-4-22.

106. Pourcel C, Vidgop Y, Ramisse F, Vergnaud G, Tram C. 2003. Characterization of a tandem repeat polymorphism in *Legionella pneumophila* and its use for genotyping. *J Clin Microbiol* **41:**1819–1826. http://dx.doi.org/10.1128/JCM.41.5.1819-1826.2003.

107. Liu Y, Lee MA, Ooi EE, Mavis Y, Tan AL, Quek HH. 2003. Molecular typing of *Salmonella enterica* serovar Typhi isolates from various countries in Asia by a multiplex PCR assay on variable-number tandem repeats. *J Clin Microbiol* **41:**4388–4394. http://dx.doi.org/10.1128/JCM.41.9.4388-4394.2003.

108. Lindstedt BA, Torpdahl M, Vergnaud G, Le Hello S, Weill FX, Tietze E, Malorny B, Prendergast DM, Ní Ghallchoir E, Lista RF, Schouls LM, Söderlund R, Börjesson S, Åkerström S. 2013. Use of multilocus variable-number tandem repeat analysis (MLVA) in eight European countries, 2012. *Euro Surveill* **18:**20385. http://dx.doi.org/10.2807/ese.18.04.20385-en.

109. Hyytiä-Trees E, Smole SC, Fields PA, Swaminathan B, Ribot EM. 2006. Second generation subtyping: a proposed PulseNet protocol for multiple-locus variable-number tandem repeat analysis of Shiga toxin-producing *Escherichia coli* O157 (STEC O157). *Foodborne Pathog Dis* **3:**118–131. http://dx.doi.org/10.1089/fpd.2006.3.118.

110. Torpdahl M, Sørensen G, Lindstedt BA, Nielsen EM. 2007. Tandem repeat analysis for surveillance of human Salmonella Typhimurium infections. *Emerg Infect Dis* **13:**388–395. http://dx.doi.org/10.3201/eid1303.060460.

111. Noller AC, McEllistrem MC, Pacheco AG, Boxrud DJ, Harrison LH. 2003. Multilocus variable-number tandem repeat analysis distinguishes outbreak and sporadic *Escherichia coli* O157:H7 isolates. *J Clin Microbiol* **41:**5389–5397. http://dx.doi.org/10.1128/JCM.41.12.5389-5397.2003.

112. Ross BC, Dwyer B. 1993. Rapid, simple method for typing isolates of *Mycobacterium tuberculosis* by using the polymerase chain reaction. *J Clin Microbiol* **31:**329–334.

113. Friedman CR, Stoeckle MY, Johnson WD, Jr, Riley LW. 1995. Double-repetitive-element PCR method for subtyping *Mycobacterium tuberculosis* clinical isolates. *J Clin Microbiol* **33:**1383–1384.

114. Harris E. 1998. *A Low-Cost Approach to PCR*. Oxford University Press, New York, NY.

115. White BA. 1993. *PCR Protocols: Current Methods and Applications*, vol 15. Humana Press, Totowa, NJ. http://dx.doi.org/10.1385/0896032442.

116. Schwartz DC, Cantor CR. 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37:**67–75. http://dx.doi.org/10.1016/0092-8674(84)90301-5.

117. Chu G, Vollrath D, Davis RW. 1986. Separation of large DNA molecules by contour-clamped homogeneous electric fields. *Science* **234:**1582–1585. http://dx.doi.org/10.1126/science.3538420.

118. Arbeit RD, Arthur M, Dunn R, Kim C, Selander RK, Goldstein R. 1990. Resolution of recent evolutionary divergence among *Escherichia coli* from related lineages: the application of pulsed field electrophoresis to molecular epidemiology. *J Infect Dis* **161:**230–235. http://dx.doi.org/10.1093/infdis/161.2.230.

119. Maslow JN, Arbeit RD. 1993. Application of pulsed field gel electrophoresis to molecular epidemiology. *Clin Infect Dis* **17:**153–162.

120. Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH, Swaminathan B. 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol* **33:**2233–2239.

121. Goering RV. 2010. Pulsed field gel electrophoresis: a review of application and interpretation in the molecular epidemiology of infectious disease. *Infect Genet Evol* **10:**866–875. http://dx.doi.org/10.1016/j.meegid.2010.07.023.

122. Salipante SJ, SenGupta DJ, Cummings LA, Land TA, Hoogestraat DR, Cookson BT. 2015. Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology. *J Clin Microbiol* **53:**1072–1079. http://dx.doi.org/10.1128/JCM.03385-14.

123. Scharff RL, Besser J, Sharp DJ, Jones TF, Peter GS, Hedberg CW. 2016. An economic evaluation of PulseNet: a network for foodborne disease surveillance. *Am J Prev Med* **50**(Suppl 1)**:**S66–S73. http://dx.doi.org/10.1016/j.amepre.2015.09.018.

124. Chung M, de Lencastre H, Matthews P, Tomasz A, Adamsson I, Aires de Sousa M, Camou T, Cocuzza C, Corso A, Couto I, Dominguez A, Gniadkowski M, Goering R, Gomes A, Kikuchi K, Marchese A, Mato R, Melter O, Oliveira D, Palacio R, Sá-Leão R, Santos Sanches I, Song JH, Tassios PT, Villari P, Multilaboratory Project Collaborators. 2000. Molecular typing of methicillin-resistant *Staphylococcus aureus* by pulsed-field gel electrophoresis: comparison of results obtained in a multilaboratory effort using identical protocols and MRSA strains. *Microb Drug Resist* **6:**189–198. http://dx.doi.org/10.1089/mdr.2000.6.189.

125. Cookson BD, Aparicio P, Deplano A, Struelens M, Goering R, Marples R. 1996. Inter-centre comparison of pulsed-field gel electrophoresis for the typing of methicillin-resistant *Staphylococcus aureus*. *J Med Microbiol* **44:**179–184. http://dx.doi.org/10.1099/00222615-44-3-179.

126. Davis MA, Hancock DD, Besser TE, Call DR. 2003. Evaluation of pulsed-field gel electrophoresis as a tool for determining the degree of genetic relatedness between strains of *Escherichia coli* O157:H7. *J Clin Microbiol* **41:**1843–1849. http://dx.doi.org/10.1128/JCM.41.5.1843-1849.2003.

127. Southern EM. 1979. Analysis of restriction-fragment patterns from complex deoxyribonucleic acid species. *Biochem Soc Symp* **44:**37–41.

128. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, Hermans P, Martin C, McAdam R, Shinnick TM. 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* **31:**406–409.

129. Thierry D, Brisson-Noël A, Vincent-Lévy-Frébault V, Nguyen S, Guesdon JL, Gicquel B. 1990. Characterization of a *Mycobacterium tuberculosis* insertion sequence, IS6110, and its application in diagnosis. *J Clin Microbiol* **28:**2668–2673.

130. Thierry D, Cave MD, Eisenach KD, Crawford JT, Bates JH, Gicquel B, Guesdon JL. 1990. IS6110, an IS-like element of *Mycobacterium tuberculosis* complex. *Nucleic Acids Res* **18:**188. http://dx.doi.org/10.1093/nar/18.1.188.

131. Zainuddin ZF, Dale JW. 1989. Polymorphic repetitive DNA sequences in *Mycobacterium tuberculosis* detected with a gene probe from a *Mycobacterium fortuitum* plasmid. *J Gen Microbiol* **135:**2347–2355.

132. Eisenach KD, Crawford JT, Bates JH. 1988. Repetitive DNA sequences as probes for *Mycobacterium tuberculosis*. *J Clin Microbiol* **26:**2240–2245.

133. Jansen R, van Embden JD, Gaastra W, Schouls LM. 2002. Identification of a novel family of sequence repeats among prokaryotes. *OMICS* **6:**23–33. http://dx.doi.org/10.1089/15362310252780816.

134. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315:**1709–1712. http://dx.doi.org/10.1126/science.1138140.

135. Nakata A, Amemura M, Makino K. 1989. Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K-12 chromosome. *J Bacteriol* **171:**3553–3556. http://dx.doi.org/10.1128/jb.171.6.3553-3556.1989.

136. Mojica FJ, Díez-Villaseñor C, Soria E, Juez G. 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol* **36:**244–246. http://dx.doi.org/10.1046/j.1365-2958.2000.01838.x.

137. Dale JW, Brittain D, Cataldi AA, Cousins D, Crawford JT, Driscoll J, Heersma H, Lillebaek T, Quitugua T, Rastogi N, Skuce RA, Sola C, Van Soolingen D, Vincent V. 2001. Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: recommendations for standardised nomenclature. *Int J Tuberc Lung Dis* **5:**216–219.

138. Sola C, Filliol I, Gutierrez MC, Mokrousov I, Vincent V, Rastogi N. 2001. Spoligotype database of *Mycobacterium tuberculosis*: biogeographic distribution of shared types and epidemiologic and phylogenetic perspectives. *Emerg Infect Dis* **7:**390–396. http://dx.doi.org/10.3201/10.3201/eid0703.0107304.

139. Demay C, Liens B, Burguière T, Hill V, Couvin D, Millet J, Mokrousov I, Sola C, Zozio T, Rastogi N. 2012. SITVITWEB—a publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infect Genet Evol* **12:**755–766. http://dx.doi.org/10.1016/j.meegid.2012.02.004.

140. Filliol I, Driscoll JR, van Soolingen D, Kreiswirth BN, Kremer K, Valétudie G, Dang DA, Barlow R, Banerjee D, Bifani PJ, Brudey K, Cataldi A, Cooksey RC, Cousins DV, Dale JW, Dellagostin OA, Drobniewski F, Engelmann G, Ferdinand S, Gascoyne-Binzi D, Gordon M, Gutierrez MC, Haas WH, Heersma H, Kassa-Kelembho E, Ho ML, Makristathis A, Mammina C, Martin G, Moström P, Mokrousov I, Narbonne V, Narvskaya O, Nastasi A, Niobe-Eyangoh SN, Pape JW, Rasolofo-Razanamparany V, Ridell M, Rossetti ML, Stauffer F, Suffys PN, Takiff H, Texier-Maugein J, Vincent V, de Waard JH, Sola C, Rastogi N. 2003. Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. *J Clin Microbiol* **41:**1963–1970. http://dx.doi.org/10.1128/JCM.41.5.1963-1970.2003.

141. Grissa I, Vergnaud G, Pourcel C. 2009. Clustered regularly interspaced short palindromic repeats (CRISPRs) for the genotyping of bacterial pathogens. *Methods Mol Biol* **551:**105–116. http://dx.doi.org/10.1007/978-1-60327-999-4_9.

142. Hoe N, Nakashima K, Grigsby D, Pan X, Dou SJ, Naidich S, Garcia M, Kahn E, Bergmire-Sweat D, Musser JM. 1999. Rapid molecular genetic subtyping of serotype M1 group A *Streptococcus* strains. *Emerg Infect Dis* **5:**254–263. http://dx.doi.org/10.3201/eid0502.990210.

143. Price EP, Smith H, Huygens F, Giffard PM. 2007. High-resolution DNA melt curve analysis of the clustered, regularly interspaced short-palindromic-repeat locus of *Campylobacter jejuni*. *Appl Environ Microbiol* **73:**3431–3436. http://dx.doi.org/10.1128/AEM.02702-06.

144. Pourcel C, Salvignol G, Vergnaud G. 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151:**653–663. http://dx.doi.org/10.1099/mic.0.27437-0.

145. Mokrousov I, Limeschenko E, Vyazovaya A, Narvskaya O. 2007. *Corynebacterium diphtheriae* spoligotyping based on combined use of two CRISPR loci. *Biotechnol J* **2:**901–906. http://dx.doi.org/10.1002/biot.200700035.

146. Delannoy S, Beutin L, Fach P. 2012. Use of clustered regularly interspaced short palindromic repeat sequence polymorphisms for specific detection of enterohemorrhagic *Escherichia coli* strains of serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19, O145:H28, and O157:H7 by real-time PCR. *J Clin Microbiol* **50:**4035–4040. http://dx.doi.org/10.1128/JCM.02097-12.

147. Delannoy S, Beutin L, Burgos Y, Fach P. 2012. Specific detection of enteroaggregative hemorrhagic *Escherichia coli* O104:H4 strains by use of the CRISPR locus as a target for a diagnostic real-time PCR. *J Clin Microbiol* **50:**3485–3492. http://dx.doi.org/10.1128/JCM.01656-12.

148. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. 1991. Light-directed, spatially addressable parallel chemical synthesis. *Science* **251:**767–773. http://dx.doi.org/10.1126/science.1990438.

149. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. 1999. High density synthetic oligonucleotide arrays. *Nat Genet* **21**(Suppl)**:**20–24. http://dx.doi.org/10.1038/4447.

150. Lockhart DJ, Winzeler EA. 2000. Genomics, gene expression and DNA arrays. *Nature* **405:**827–836. http://dx.doi.org/10.1038/35015701.

151. Chu Y, Corey DR. 2012. RNA sequencing: platform selection, experimental design, and data interpretation. *Nucleic Acid Ther* **22:**271–274. http://dx.doi.org/10.1089/nat.2012.0367.

152. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320:**1344–1349. http://dx.doi.org/10.1126/science.1158441.

153. Martin JA, Wang Z. 2011. Next-generation transcriptome assembly. *Nat Rev Genet* **12:**671–682. http://dx.doi.org/10.1038/nrg3068.

154. Ozsolak F, Milos PM. 2011. Transcriptome profiling using single-molecule direct RNA sequencing. *Methods Mol Biol* **733:**51–61. http://dx.doi.org/10.1007/978-1-61779-089-8_4.

155. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95:**3140–3145. http://dx.doi.org/10.1073/pnas.95.6.3140.

156. Maiden MC. 2006. Multilocus sequence typing of bacteria. *Annu Rev Microbiol* **60:**561–588. http://dx.doi.org/10.1146/annurev.micro.59.030804.121325.

157. Enright MC, Spratt BG. 1999. Multilocus sequence typing. *Trends Microbiol* **7:**482–487. http://dx.doi.org/10.1016/S0966-842X(99)01609-1.

158. Zhang W, Jayarao BM, Knabel SJ. 2004. Multi-virulence-locus sequence typing of *Listeria monocytogenes*. *Appl Environ Microbiol* **70:**913–920. http://dx.doi.org/10.1128/AEM.70.2.913-920.2004.

159. Nemoy LL, Kotetishvili M, Tigno J, Keefer-Norris A, Harris AD, Perencevich EN, Johnson JA, Torpey D, Sulakvelidze A, Morris JG, Jr, Stine OC. 2005. Multilocus sequence typing versus pulsed-field gel electrophoresis for characterization of extended-spectrum beta-lactamase-producing *Escherichia coli* isolates. *J Clin Microbiol* **43:**1776–1781. http://dx.doi.org/10.1128/JCM.43.4.1776-1781.2005.

160. Tartof SY, Solberg OD, Manges AR, Riley LW. 2005. Analysis of a uropathogenic *Escherichia coli* clonal group by multilocus sequence typing. *J Clin Microbiol* **43:**5860–5864. http://dx.doi.org/10.1128/JCM.43.12.5860-5864.2005.

161. Francis SS, Plucinski MM, Wallace AD, Riley LW. 2016. Genotyping oral commensal bacteria to predict social contact and structure. *PLoS One* **11:**e0160201. http://dx.doi.org/10.1371/journal.pone.0160201.