

# Molecular Epidemiology

Lee W. Riley

## 1. Introduction

Molecular epidemiology is now an established discipline in epidemiology.<sup>(1)</sup> It is the contemporary stage in the evolution of laboratory-based epidemiology that may have begun with the discovery in the late 1800s of ways to differentiate bacterial organisms by pure culture in artificial media.<sup>(2)</sup> Molecular epidemiology uses new molecular biology tools to address questions difficult or not possible to address by old laboratory tools. Just as statistical tools have become indispensable in epidemiological investigations and interpretations of epidemiologic data, molecular biology tools today have come to elucidate epidemiologic features of diseases that cannot be easily characterized by conventional techniques. Applied to infectious diseases, molecular biology methods have also come to challenge our traditional notions about the epidemiology of these diseases and have engendered novel opportunities for their prevention and control. This chapter will (1) review definitions commonly used in molecular epidemiology, (2) present an overview of molecular biology methods used to study infectious disease epidemiology, and (3) describe examples of the types of epidemiologic problems that can be addressed by molecular biology techniques, highlighting new concepts that emerged in the process of applying this approach to study bacterial infectious diseases.

## 2. Definitions

*Molecular epidemiology:* In Chapter 1, epidemiology is defined as “. . .the study of the distribution and determinants of health-related states, conditions, or events in specified

populations and the application of the results of this study to the control of health problems”. Correspondingly, molecular epidemiology can be defined as “the study of the distribution and determinants of health-related states, conditions, or events in specified populations that utilizes molecular biology methods”. In molecular epidemiology of infectious diseases, the word “determinants” is emphasized. Furthermore, as we learn more about the biology of agents of infectious diseases, it is becoming evident that disease distribution in a community of hosts may be determined not only by factors such as the distribution of contaminated vehicles, host risk characteristics, and environmental contexts in which these diseases occur, but also by a pathogen’s genetic composition.

For example, communicable bacterial pathogens that cause diarrhea are almost always transmitted by the fecal–oral route. Colonization of the intestine is an important prerequisite for these pathogens to facilitate their transmission to new hosts. Similarly, the agent of tuberculosis *Mycobacterium tuberculosis*, which is transmitted by the airborne route, must cause active disease in the host’s lungs to assure its transmission to new hosts. An agent of leptospirosis, *Leptospira*, prolongs its survival in a community of hosts by establishing carriage in a mammalian host’s kidney tubules so that it can be excreted in urine to infect other hosts. Mechanisms that underlie these organisms’ ability to infect a specific target organ, which determines the specific pattern of transmission, are genetically determined. Thus, the definition of molecular epidemiology can be expanded to include the following: a study of the genetic factors that determine and regulate an organism’s specific mode of transmission among hosts within an environmental context.

*Differentiating molecular epidemiology from taxonomy and phylogeny:* In February 2009, a bibliographic search by PubMed using the words “molecular epidemiology” generated 27,822 publications that contained these words in the title or the abstract. Most of these publications dealt with

---

Lee W. Riley • University of California – Berkeley, School of Public Health, 140 Warren Hall, Berkeley, California 94720, USA.

the laboratory methods used to subtype microorganisms. They did not discuss the use of these techniques to characterize disease occurrence, distribution, or determinants of disease distribution. Thus, the term “molecular epidemiology” is often used to describe studies that should be more appropriately called “molecular taxonomy” or “phylogeny”. *Taxonomy* is the science of classification of organisms into naturally related groups based on a factor common to each. *Phylogeny* is the study of lines of descent or evolutionary development of an organism.<sup>(1)</sup>

Molecular epidemiology, taxonomy, and phylogeny may employ the same molecular biology laboratory techniques, but each is based on distinct principles. In taxonomy or phylogeny applied to infectious agents, the data that are generated are used to describe properties of organisms and their relationships to each other, based on an inferred model. Molecular epidemiology seeks to identify relationships between the etiologic agent of a disease (organism causing the disease), the host, and the environment in which the etiologic agent and the host reside. Both epidemiology and phylogeny as applied to infectious diseases may describe the distribution of particular genetic attributes of a pathogen over time. However, in phylogeny, the relationships of these genetic attributes over time must be guessed, since past evolutionary events cannot be empirically confirmed. That is, phylogenetic analysis seeks to infer past evolutionary events based on a model from a set of observations made with materials available in the present. Epidemiology focuses on events of the present or the recent past. It uses analytical and empirical study designs to identify factors that predict the relationship of attributes of pathogens to disease distribution in place, person, and time. It attempts to identify factors that determine disease transmission, manifestation, and progression. In molecular epidemiology, the hypotheses that are generated are testable or can be made to be testable empirically when enough study subjects are identified. Finally, epidemiology is always motivated by an opportunity or possibility for intervention and prevention. Thus, an application of a molecular biology technique to type an organism per se does not by itself constitute molecular epidemiology. However, it should be stressed that the disciplines of taxonomy and phylogeny play an important role in the discipline of molecular epidemiology, and this idea is reviewed in detail in a separate reference.<sup>(1)</sup>

Terminologies introduced in this chapter are based on definitions recommended by the European Study Group on Epidemiological Markers<sup>(3)</sup> and the Molecular Typing Working Group of the Society for Healthcare Epidemiology of America.<sup>(4)</sup> Others were adapted from a reference book *Molecular Epidemiology of Infectious Diseases: Principles and Practice*.<sup>(1)</sup> The terminologies below are mostly used in ways to apply to prokaryotic organisms.

*Discriminatory power*: The ability of a laboratory test to generate distinct and discrete units of information from different isolates, usually at a subspecies level.

*Isolate*: A population of microbial cells in pure culture derived from a single colony on an isolation plate and identified to the species level.

*Strain*: An isolate or a group of isolates exhibiting phenotypic and/or genotypic traits belonging to the same lineage, distinct from those of other isolates of the same species.

*Clone*: An isolate or a group of isolates descending from a common precursor strain by non-sexual reproduction exhibiting phenotypic or genotypic traits characterized by a strain typing method to belong to the same group.

*Type, typing information, taxonomic unit, or marker*: A specific and discrete unit of information or character belonging to a strain displayed upon application of a strain typing procedure. (e.g., antibiotic resistance pattern, serotype, electrophoretic banding pattern, allele).

*Genotype*: A genetic characteristic of a cell or an organism according to its entire genome or a specific set of genetic loci (allele).

*Phenotype*: An observable characteristic expressed by a cell or an organism, such as drug resistance, virulence, and morphology.

*Biotype*: A taxonomic unit of a microorganism belonging to one species based on a panel of biochemical tests.

*Serotype*: A taxonomic unit of a microorganism based on antigenic properties of a distinct set of surface structures.

*Pathotype*: A taxonomic unit of a microorganism that describes a pathogenic variant of a species or subspecies of microorganisms that normally colonize a host. Also called *pathovar*.

*Allele*: A nucleotide sequence variant of a gene at a particular locus.

*Single nucleotide polymorphism (SNP)*: Small changes or variation in corresponding nucleotide sequences at a given locus (e.g., corresponding locus TTAGGTCCTTTA in strain 1 and TTTGGTCCTTTA in strain 2 differs by a single nucleotide in the third position).

*Microsatellites*: DNA sequences containing tandem repeats of a short sequence (e.g., CCG repeated 10 times in succession). Most often used in reference to such sequences in eukaryotic cells.

### 3. Methodology

It should be emphasized at the outset that molecular biology methods are no more special than any other methods—laboratory or non-laboratory—used in epidemiology. Curiously, however, when biochemical laboratory

methods are used to characterize organisms in an epidemiologic investigation, we do not use the term “biochemical epidemiology”. For some reason, when molecular biology methods are used in an epidemiologic study, the study is described as “molecular epidemiology”. This probably results from a peculiar human habit of glorifying certain characteristics over others. For instance, it is more fashionable in research universities to call a department doing the same type of research a “molecular biology” instead of a “microbiology” department. Molecular epidemiology is epidemiology. In truth, molecular biology methods are nothing more than one of many tools used in epidemiology to order data into separate units or strata so that they become amenable for different types of analytical procedures.

All laboratory-based methods used to type microorganisms (*typing systems*) fall into *phenotypic* and *genotypic* methods, as described by Maslow et al.,<sup>(5)</sup> Tenover et al.,<sup>(4)</sup> and Lipuma.<sup>(6)</sup> Most of the “conventional” laboratory typing methods fall into phenotypic methods, which are based on the detection of phenotypes or characteristics expressed by an organism. Strain typing methods based on genotypes rely on the analysis of nucleic acid contents and gene sequence polymorphisms (chromosomal DNA, extrachromosomal DNA, and RNA). Nowadays, with routine applications of molecular biology tests, the distinction between “conventional” and molecular biology laboratory techniques may no longer be meaningful. Many standard strain typing procedures today are built upon layers of tests that include both molecular and non-molecular biology techniques.

### 3.1. Phenotyping Methods

The first step in data stratification as applied to bacterial infectious disease agents is differentiating pathogens from normal commensal or saprophytic bacterial organisms. Commensal organisms reside, metabolize, and replicate in a healthy, immunocompetent host without causing disease. They colonize the skin, mouth, nasopharynx, gastrointestinal tract, vaginal mucosa, and others. A saprophyte is a type of microorganism that grows normally on dead matter, and hence is found in the environment. This initial step separates organisms into species. Microorganism characteristics used to define species include growth, morphologic, biochemical, serologic, functional, and physiologic properties.

**3.1.1. Differentiation by Growth and Morphologic Characteristics.** Differences in nutritional requirements of bacterial organisms determine their growth characteristics. They can be classified according to whether they grow at all or the way they grow on solid agar culture plates or in liquid media. Morphologically, they may be differentiated by their colony shape, color, texture, smell, or whether they “swarm”

on the plate. They are also differentiated by the way they stain (e.g., Gram stain, acid fast stain) as observed under a light microscope, or their shape and size as observed under an electron microscope.

**3.1.2. Typing Based on Biochemical Characteristics.** Biochemical tests used to differentiate bacteria are based on the organism’s metabolic activities. Strains differentiated by a panel of biochemical tests group into *biotypes*, and the tests are thus referred to as biotyping tests. The discriminatory power of a biotyping test depends on the number of biochemical tests used to subtype strains—the larger the number of biochemical characteristics analyzed in a pathogen, the greater the discriminatory power of the test. The tests depend on an organism’s metabolic activities, which are sometimes subject to minor changes in growth condition (e.g., nutrient concentration, pH, and temperature) used, and hence the information that is generated may not always be reliable for epidemiologic studies. In developed countries, these tests have come to rely on kits and automated systems, facilitating rapid turnover. However, the large number of tests often needed to subtype a strain renders biochemical classification of pathogens expensive.

**3.1.3. Typing by Serologic Characteristics.** Another frequently used phenotyping method is based on a characteristic antigen-specific serologic (antibody) response induced by an invading organism in a mammalian host. Bacterial antigens that induce antibody response used for serologic classification include proteins and polysaccharides. For example, the species *Escherichia coli* characterized biochemically can be further subtyped into a “serogroup”, based on the organism’s characteristic somatic antigen (O-polysaccharide). A numerical designation is assigned to an *E. coli* strain carrying a characteristic O-polysaccharide that induces an antibody response (e.g., O157). The same strain, if further subtyped based on its flagellar protein (H-antigen), is classified as a “serotype”. It will then be given an O:H designation (e.g., O157:H7). The genus *Salmonella* is biochemically separated into species and subspecies. They are further subtyped serologically according to their O and H antigens. Unlike *E. coli* serotypes, *Salmonella enterica* subsp. *enterica* (subspecies I) serotypes are given a descriptive name rather than a numerical designation, e.g., *S. enterica* serotype (or serovar) Typhimurium or *S. enterica* serotype Enteritidis. Nearly 1,500 serotypes of *S. enterica* subspecies I are currently recognized.<sup>(7)</sup> The specific serogrouping or serotyping assays include bacterial agglutination test with a panel of antigen-specific antisera, enzyme-linked immunosorbent assay (ELISA), latex particle agglutination, Western blot (or immunoblot) analysis, or a modification of these tests.

One limitation of serotyping is that different strains from the same species or even strains from different species

could carry cross-reacting antigens, which can result in false-positive tests. On the other hand, strains that do not express typing antigens are classified as non-typeable (e.g., non-typeable *Haemophilus influenzae*). Serologic tests, on the most part, are simple to perform, but they require maintaining a stock of typing polyclonal or monoclonal antisera. For example, a large number (~200) of antisera need to be used to serotype *Salmonella*. Thus, serotyping could be expensive and time consuming, and is usually confined to reference laboratories.

**3.1.4. Typing by Functional or Physiologic Characteristics.** Other ways to differentiate bacterial isolates involve examination of their response to specific manipulations. They include identifying differences in susceptibility to antimicrobial agents (antibiograms), lysis by or susceptibility to bacteriophages (phage typing), patterns of association with tissue culture cells (cell culture assays), toxigenicity, survival under in vitro or in vivo stress, and metabolic enzyme expressions (multilocus enzyme electrophoresis, or MLEE). These are reviewed in detail elsewhere.<sup>(1)</sup>

## 3.2. Genotyping Methods

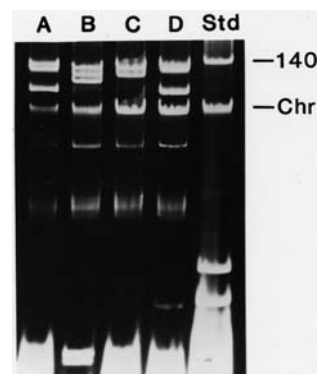
Genotyping or nucleic acid-based typing methods are ultimately based on the nucleic acid sequence differences in the chromosomal or extrachromosomal nucleic acid contents. As opposed to hundreds of ways to phenotype bacteria, all genotyping tests can be grouped into a variant of just three basic analytical procedures—(a) hybridization, (b) gel electrophoresis, and (c) nucleic acid sequencing. This allows for the use of common equipment and standard reagents to analyze many different types of infectious agents. Furthermore, genotypic characterizations of pathogens facilitate standard information storage and data analyses, interpretation, and communication, which are all amenable to computer-assisted manipulations.

In bacteria, genotyping analyses are based on comparison of two types of DNA: the genome and the extrachromosomal DNA. Typing systems based on the analysis of the genome fall into those designed to compare differences based on nucleotide changes that occur in parts of the genome (microdiversity) or the entire genome (macrodiversity). These typing methods, of course, had to be developed in place of direct whole genome DNA sequence analysis, which, despite recent advances in the sequencing technology, remains cumbersome and expensive for bacterial organisms (although no longer the case with viruses). Further detailed reviews on this topic are mentioned at the end of this chapter (Suggested Reading).

**3.2.1. Extrachromosomal DNA Genotyping Methods.** For bacteria plasmids are the extrachromosomal DNA elements used to type strains. Plasmids are covalently closed circular DNA that carry genes that may encode characteristics such as antibiotic resistance, virulence factors, or other functions that provide selective advantage to the organism but are not essential for its growth or survival. In fungi or protozoas, extrachromosomal elements used for typing may include mitochondria<sup>(8)</sup> and kinetoplasts (schizodeme analysis).<sup>(9)</sup>

Plasmid-based typing is an example of an electrophoresis-based strain typing procedure. It compares differences in molecular weight (MW) and the number of different plasmids a bacterium may carry. Nucleic acids and proteins are negatively charged. Plasmids, thus, can be resolved electrophoretically in an agarose gel matrix according to MW. That is, the separation of these molecules depends on their drag through the gel matrix as they migrate toward the positive charge pole. Large MW plasmids migrate more slowly in the gel matrix than low MW plasmids. Thus, depending on the number of different MW plasmids, each strain represented by a single electrophoretic lane will generate a banding pattern, a “fingerprint”, or a profile, that can then be compared across multiple lanes (Figure 1).

In one early study designed to validate the discriminatory power of plasmid profile analysis of bacterial pathogens, Holmberg et al. compared *S. Typhimurium* isolates from 20 well-documented outbreaks by phage types, antimicrobial resistance patterns, and plasmid profiles.<sup>(10)</sup> They found that among nine outbreaks in which non-outbreak isolates were also examined, the plasmid profile analysis was able



**Figure 1.** Agarose gel electrophoresis of plasmids (plasmid profile) extracted from drug-resistant *Salmonella enterica* serotype Typhimurium isolates from children with diarrhea in Brazil.<sup>(75)</sup> Lanes A–D show distinct banding patterns. All isolates carried a 140 MDa plasmid as well as lower MW plasmids. “Chr” denotes chromosomal DNA. MW marker standard is shown in lane marked “Std”.

to distinguish outbreak-related isolates from non-outbreak-related isolates more frequently than the other methods (phage typing and antibiogram), suggesting that the plasmid profile analysis was more discriminatory than the other methods used to type strains at that time.

One drawback of the technique is that the procedure depends on the presence of at least one plasmid in a bacterium. If isolates carry only one or two plasmids, the patterns generated will not be sufficiently discriminating enough to distinguish the strains as related or unrelated. Isolates that contain only one or two plasmids of identical MW need to have the plasmids digested with endonucleases (enzymes that cut DNA molecules at a specific sequence site; see details under “Restriction Endonuclease Analysis or REA”) to generate smaller fragments that can then be electrophoretically resolved. Another potential problem with this method is that plasmids on prolonged storage of a bacterium can be spontaneously lost, which can affect reproducibility of the typing method. Although plasmid profiles are relatively easy to perform and interpret, it does require a laboratory equipped for basic molecular biology techniques, including DNA extraction, restriction endonuclease digestion, electrophoresis, and photodocumentation of the results.

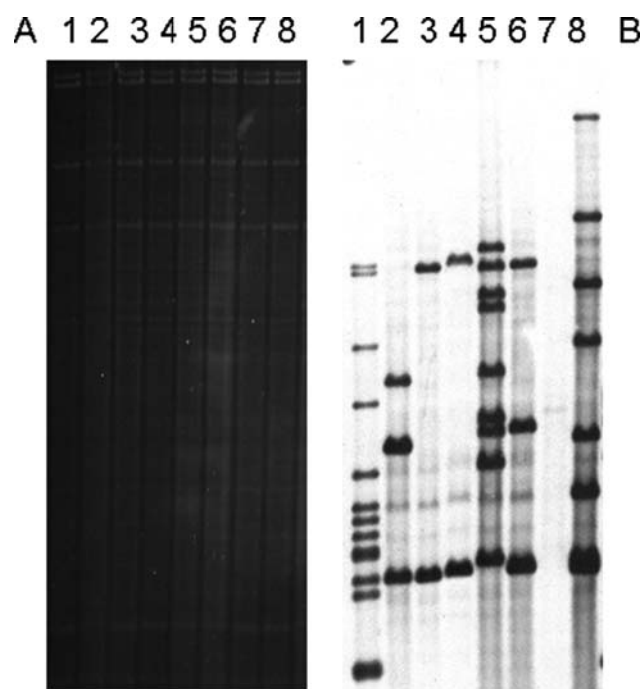
**3.2.2. Genome-Based Typing Methods.** In genome-based typing methods the entire genomic content of a microorganism is compared among different strains (macrodiversity analysis). There are three basic genome-based typing methods used in epidemiologic investigations, which are based on (1) allelic differences that occur within restriction endonuclease recognition sites in a genome, (2) comparison of multiple nucleotide sequences of whole genomes (called comparative genomics), and (3) whole genome hybridization patterns (see below for the description of hybridization) obtained from arrays of thousands of distinct short DNA fragments representing the whole genome fixed onto a glass slide or nitrocellulose filter (“gene chip” or microarray). The latter two methods have only recently begun to be applied to epidemiology.

**3.2.3. Restriction Endonuclease Analysis (REA).** Bacterial organisms express enzymes called restriction endonucleases that recognize unique sequences of several nucleotides in length and cut the DNA molecule at a specific position. They are called restriction enzymes because bacteria use these enzymes to eliminate (restrict) foreign DNA accidentally introduced into them. Thus one enzyme may generate distinct sets of DNA fragments from two different strains if the strains have a sequence difference at the corresponding restriction sites. Comparison of electrophoretic patterns based on the number and differences in MW of cut DNA fragments is called restriction endonuclease analysis

(REA) or restriction fragment length polymorphism (RFLP) analysis.<sup>(11–13)</sup>

Enzymes that cut DNA molecules at recognition sites that occur frequently throughout a microorganism’s chromosome are colloquially referred to as “frequent cutters”. Electrophoresis of DNA fragments generated from such enzymes will produce thousands of overlapping bands upon gel electrophoresis, and the resulting patterns become difficult to interpret (Figure 2A). Two different patterns ways have been developed to overcome this limitation, as outlined below.

**3.2.4. Southern Blot Hybridization.** Southern blot hybridization analysis, named after its developer,<sup>(13)</sup> is a type of analysis that facilitates detection of a particular DNA fragment of interest among hundreds of other DNA fragments generated by REA. DNA fragments separated in an agarose gel are first transferred onto a piece of nitrocellulose or nylon membrane. The membrane is then exposed to a piece of



**Figure 2.** Restriction fragment length polymorphism (RFLP) analysis with Southern blot hybridization of eight clinical isolates of *M. tuberculosis*. Panel (A) shows *M. tuberculosis* DNA digested with a restriction endonuclease PvuII. Note the smear pattern in each lane, representing thousands of DNA fragments resolved by electrophoresis in an agarose gel. However, when the DNA fragments of the gel are transferred onto a piece of nylon membrane and hybridized with a probe directed at *M. tuberculosis*’s IS6110 insertion element, only the DNA fragments that contain this element becomes visualized (B). Thus, the patterns generated in the membrane in panel B become interpretable and can be compared among strains.

single-stranded DNA (probe) labeled with a molecule that facilitates visual detection of a selected target DNA fragment. The probe will specifically bind (hybridize) to its complementary DNA sequence embedded in the membrane. For example, a nucleic acid sequence or probe *ATCGGCGATG* will hybridize to a target containing the sequence *TAGC-CGCTAC*. Thus, this probe-based detection of target DNA fragments is called hybridization. If the organism of interest harbors more than one copy of the target DNA sequence in different locations in the chromosome, the membrane will show multiple bands (banding pattern) in a single lane, which can be visually detected. Of course, the number of such visualizable bands will be few and will not be obscured by the other fragments in the membrane, which cannot be visualized because they do not hybridize to the probe. This allows for comparison and interpretation of the band patterns (Figure 2B).

This method has been applied to subtype a wide variety of bacterial organisms, and is the current standard method used to type *M. tuberculosis*, in which the organism's repetitive DNA element or insertion element called *IS6110* is used as the target for a hybridization probe (called *IS6110* RFLP).<sup>(14)</sup> One limitation of this approach is that the DNA sequence selected as a target for a hybridization probe must be present in the chromosome in high copy numbers (usually >5) for the technique to be sufficiently discriminating. Clinical isolates of *M. tuberculosis* completely lacking or having small copy numbers of *IS6110* have been found.<sup>(14)</sup> In such situations, secondary typing methods must be applied. Another drawback of the RFLP analysis followed by hybridization is that it requires a specialized reference or research laboratory to perform.

When the hybridization typing method uses ribosomal genes (*rrn*) as a target, it is called ribotyping. Ribosomal RNA sequences are highly conserved, especially within members of the same genus. This property can be used, then, to type a wide range of organisms related at the species level, but by the same token, the discriminatory power of ribotyping is low. One study found that *E. coli* isolates obtained from widely separate geographic areas shared identical ribotypes.<sup>(15)</sup> Thus, for *E. coli*, ribotyping would not be discriminating enough for local epidemiologic studies or regional surveillance, although it may be useful for studying evolutionary relationships.

### 3.2.5. Pulsed Field Gel Electrophoresis (PFGE).

The second way developed to overcome the problem of uninterpretable electrophoretic patterns generated by RFLP is to use endonucleases that recognize rare restriction site sequences ("rare cutters"). Such enzymes will produce very large DNA fragments, which, when resolved electrophoretically, will generate interpretable patterns. However,

electrophoresis in agarose gels of large pieces of DNA introduces two additional technical constraints. Linear pieces of DNA  $\geq 25,000$  bases (25 kilobases or kb) are difficult to resolve by the conventional agarose gel electrophoresis method. They get stuck during migration across the gel matrix. Thus, PFGE was developed in which the orientation of the electrical field across the gel is changed (pulsed) periodically.<sup>(16)</sup> This allows the DNA fragments to move in different directions, releasing them from the stuck position. The DNA fragment will thus take a path of least resistance along its gel migration toward the positive charge pole.

The second technical obstacle that had to be overcome was the tendency for large DNA fragments to undergo spontaneous autodigestion. This problem was eventually solved by enzymatically lysing and digesting the cell wall and proteins of the test bacterial organism embedded in an agarose gel plug. This releases the bacterial DNA directly into the plug. The released DNA can then be digested within the plug with a restriction enzyme. This prevents autodigestion.

The major advantage of PFGE is that virtually all genome-containing organisms are typeable by this method. In general, genetic changes detected by PFGE reflect genetic events in an organism that occurred in the recent past. Hence, it is a highly informative technique with wide epidemiologic appeal. PFGE is the method of choice used by a national foodborne disease surveillance system called PulseNet, which is collaboration among Centers for Disease Control and Prevention (CDC), Association of Public Health Laboratories (APHL), and the food regulatory agencies, including the United States Department of Agriculture (USDA) and the Food and Drug Administration (FDA). Foodborne pathogens included in the PulseNet surveillance currently include *Salmonella* Typhimurium, *E. coli* O157:H7, *Listeria monocytogenes*, *Campylobacter jejuni*, and *Shigella sonnei*<sup>(17)</sup> ([www.cdc.gov/pulsenet](http://www.cdc.gov/pulsenet)). PulseNet has contributed to detecting, investigating, and controlling multistate outbreaks of *E. coli* O157:H7,<sup>(18)</sup> *Salmonella* Agona infections,<sup>(19)</sup> and listeriosis.<sup>(20)</sup> The network will expand to PulseNet Canada, PulseNet Europe, PulseNet Asia Pacific, and PulseNet South America.

One major limitation of PFGE is that it is technically demanding, requiring an expensive electrophoresis apparatus. Thus, this typing system is currently used only by reference or research laboratories.

### 3.2.6. Whole-Genome Sequence Comparison.

The reduced cost and ability to rapidly sequence large pieces of DNA engendered a new discipline called comparative genomics.<sup>(21, 22)</sup> With viruses, whose genomes are considerably smaller than those of bacteria, comparative genomics is already applied to directly genotype clinical isolates. That is, hundreds of genomes of viral strains belonging to one species

can now be compared rapidly to each other and applied to address epidemiologic questions. For example, it was the comparative genomics analyses that enabled investigators to conclude that the agent of severe acute respiratory syndrome (SARS) belonged to a group of coronaviruses.<sup>(23)</sup>

With bacteria, because of the large genome, such comparisons are still not possible for a large number of strains. That is, hundreds of genomes of a single bacterial species cannot be compared because they have not been sequenced. However, new information obtained from comparison of a limited number of whole genome sequence databases is used to develop alternative genotyping methods. For example, at least eight different clinical isolates of *Staphylococcus aureus* have now undergone whole genome sequence analysis. Multiple alignment comparisons reveal several regions of difference. One such region called RD13 containing a cluster of *Staphylococcus* exotoxin (*set*) genes has been recently exploited to develop a new genotyping method called exotoxin sequence typing.<sup>(24)</sup>

Whole genome sequence comparison of *M. tuberculosis* H37Rv, CDC1551, strain 210, and *Mycobacterium bovis* AF2122/97 led to the identification of 212 *single nucleotide polymorphism* (SNP) markers.<sup>(25)</sup> More recently, an analysis of these SNP markers in a global collection of 294 *M. tuberculosis* strains identified six phylogenetically distinct SNP cluster groups (SCGs). SCGs were strongly associated with geographic origin of these isolates.<sup>(26)</sup>

**3.2.7. Microarray Comparisons.** Another way to compare whole genomes is based on hybridization patterns generated from thousands of short pieces of DNA (probes) of known sequence representing the genome, arranged on a glass slide or nitrocellulose membrane, often referred to as microarray or “gene chip”. Microarrays are a type of DNA hybridization matrix typing method described previously. The hybridization technology is not new, and the microarray technology is just a recent variant of the technology made possible by miniaturization techniques developed in the semiconductor industry.<sup>(27–29)</sup>

Most microarray-based applications to infectious disease currently focus on taxonomic differentiation or diagnosis, and not on epidemiologic investigations. In one study, a whole genome DNA microarray analysis of 13 vaccine strains of *M. bovis*-BCG found 16 regions deleted in BCG strains relative to the sequenced *M. tuberculosis* H37Rv genome.<sup>(30)</sup> Kato-Maeda et al. found 25 deleted regions among 19 clinical *M. tuberculosis* isolates by a whole genome microarray approach.<sup>(31)</sup> Here, they reported that there was an inverse relationship between the proportion of deleted regions in these strains and the percentage of patients with cavitory disease infected with these strains. The discovery that *M. tuberculosis* contained such regions

of deletion led some investigators to propose a genotyping scheme called “deligotyping”.<sup>(32)</sup> In this method, clinical isolates of *M. tuberculosis* are compared according to differences in the regions of deletions.

Microarrays facilitate comparison of a large number of strains more rapidly than the whole genome sequence comparisons, but they are dependent on the number and types of DNA fragments that are included in the array, and hence will not be as discriminatory as the latter. Furthermore, they require specialized laboratories equipped with a scanner to read the microarray and expensive software to interpret the hybridization patterns on the chip. Thus, this technology is not likely to become readily available in the near future in most places to conduct epidemiologic investigations.

**3.2.8. PCR-Based Genotyping Methods.** All of the genotyping methods described above are based on DNA extracted from bacteria that are cultivated in vitro. Culturing bacteria serves two important purposes—(1) makes available large amounts of DNA for the genotype analysis and (2) assures that the genetic material being analyzed represents a distinct bacterial strain, obtained from pure culture.

For bacterial organisms that grow very slowly (*M. tuberculosis*) or cannot be cultivated artificially (*Treponema pallidum*), the above genotyping methods would be difficult to apply. Improper storage of bacterial isolates may also cause them to die, in which case genotyping by any of the above methods will not be possible. Polymerase chain reaction (PCR)-based tests can overcome these limitations. The PCR technology is a powerful and simple molecular biology tool, which is now applied in a variety of ways to genotype microorganisms in taxonomic as well as epidemiologic studies. It should be noted that PCR is only one of several methods called nucleic acid amplification tests (NAAT) used to make multiple copies of a target DNA or RNA of an organism or cells in vitro without culturing the organism or cells. PCR is perhaps the most frequently used NAAT and the most common method applied to subtype microorganisms. There are many excellent books that detail the procedure and are listed at the end of this chapter (Suggested Reading).

**3.2.9. Overview of PCR.** There are a number of frequently used terms that the reader should become familiar with in the discussion of PCR tests. Those familiar with the procedure may skip this section.

PCR involves multiple cycles of a three-step process: (1) denaturation of double-stranded DNA into single strands, (2) annealing of oligonucleotide primers to target DNA sequences called a template, and (3) primer extension along the template sequence from the primer binding site. These steps are carried out in a plastic microcentrifuge tube containing a reaction mixture of buffers, nucleotides, primers, heat-stable polymerase, and a sample DNA template. DNA

when heated to a sufficiently high temperature will separate into single strands. Separation into single strands of a double-stranded DNA is called denaturation. Thus, in the first step, the sample DNA (template) is usually heated to 95–100°C. Primers are synthetic single strands of DNA (oligonucleotides) of variable lengths designed to bind (anneal) to specific complementary target sites in a DNA template. They can be custom-made commercially. Primers bind to the target sequences when the temperature of the reaction mixture is decreased from a denaturation temperature to an annealing temperature range (~40–60°C). At about 70–75°C (primer extension temperature), a thermostable enzyme called DNA polymerase will add nucleotides (extension) complementary to the unpaired DNA strand onto the annealed primer. In this way, a DNA segment located between the two annealed primers will be copied. This three-step cycle involves change in temperature only—denaturation, annealing, and extension temperature—which is facilitated automatically by a programmable heating device called thermocycler. Depending on the time applied to each of the steps in the cycle, the type of thermocycler used, and quality of the polymerase, a single copy of DNA can be multiplied to more than a billion copies after 30 cycles in less than 3 h. The amplified product can then be visualized in an electrophoretic gel under UV illumination after staining with ethidium bromide.

For PCR tests, the cultivation of organisms is usually needed not so much to make enough nucleic acid material available for analyses, but to assure that the product being analyzed represents a single or pure strain of an organism. For bacteria, if a clinical specimen is obtained from a normally sterile body site such as blood, cerebrospinal fluid, or deep tissue, and it is obtained in a sterile fashion, PCR assay may be performed directly on the specimen itself. The primer

sequences could be designed to bind to sequences that are specific to a target organism (e.g., IS6110 of *M. tuberculosis*). In this way, DNA targets in an organism infecting non-sterile sites (sputum, urine, stool) could be amplified. Blood, urine, and other tissues may contain products inhibitory to PCR, but if the specimen contains a sufficiently large number of the infectious agent of interest, the desired target DNA could be amplified and detected. If a stored organism becomes contaminated with other organisms, or lose their viability, a PCR test can still amplify DNA from that organism, as long as the DNA material is intact. If the primers are highly species specific, a PCR test can distinguish the target sequences from DNA derived from contaminating organisms, thus precluding culturing the target organism from a specimen.

**3.2.10. PCR-Based Strain Typing Methods.** PCR-based genotyping methods that compare electrophoretic bands are classified into two general categories: (1) those based on MW polymorphism of a single amplified product and (2) those that display band patterns (fingerprints) from multiple amplified products. The former method is often used to type organisms to predict species,<sup>(33–35)</sup> serogroups, serotypes,<sup>(34, 36, 37)</sup> or *pathotypes*.<sup>(37)</sup> The major advantage is that this PCR-based method is considerably simpler than the conventional methods used to group them into species, serotypes, or pathotypes, but its discriminatory power is limited to differentiating organisms belonging to a large taxonomic unit. It does not discriminate beyond these taxonomic units.

More discriminating PCR-based typing methods require generation of fingerprints. These latter techniques can be divided into three major groups (Table 1): (1) those that rely on sequence polymorphism in the whole genome,

**Table 1. PCR-Based Strain Typing Methods Used in Epidemiologic Studies**

General category of methods	Specific techniques
Methods that rely on sequence polymorphism in the whole genome	Arbitrarily primed PCR (AP-PCR) or randomly amplified polymorphic DNA (RAPD) PCR
Methods based on heterogeneity within endonuclease restriction sites	PCR-restriction enzyme analysis (REA) or PCR-restriction fragment length polymorphism (RFLP) analysis Amplified fragment length polymorphism (AFLP) analysis Restriction-site-specific PCR (RSS-PCR) analysis Amplified ribosomal DNA restriction analysis (ARDRA)
Methods based on interspersed DNA repetitive elements	Extragenic palindromic (REP) assay Enterobacterial repetitive intergenic consensus (ERIC) sequence assay Mosaic repetitive element PCR (BOX-PCR) Insertion element PCR assays (IS6110, IS1245, IS900) Short interspersed DNA repetitive element assay (PGRS, VNTR, MPTR, ETR)
Methods that compare allelic differences in selected target genes by sequencing PCR products	Direct repeat (DR) PCR or spoligotyping Multilocus sequence typing (MLST), single nucleotide polymorphism (SNP) analysis



(2) those based on heterogeneity within known restriction endonuclease recognition sites, and (3) those that take advantage of repetitive elements interspersed in the target genome.

One relatively new category of PCR-based typing methods is based on comparison of sequences of PCR products of selected gene targets instead of comparison of band patterns by electrophoresis. One example of such a method is called multilocus sequence typing (MLST), discussed below. All of the above methods are described in detail in reference (1). Methods that are most frequently used in epidemiologic investigations are described below.

**3.2.11. PCR Typing Methods Based on Sequence Polymorphism in the Whole Genome.** Arbitrarily primed PCR (AP-PCR), also called randomly amplified polymorphic DNA (RAPD), are PCR techniques that rely on arbitrarily or randomly designed sequences of primers that anneal specifically to DNA templates in a target organism.<sup>(38–40)</sup> If a set of these randomly designed primers anneal to target DNA sequences, the intervening segments that are proximal enough to these annealing sites can be amplified and generate products of different MW. Such products will display a band pattern in an electrophoresis gel for each strain. The advantage of this method is that the target nucleotide sequence does not have to be known. The primer sequences are determined empirically until a set that works is found. The final choice of the primer design is based on the generation of reproducible and interpretable band patterns. The major drawback of this test, however, is its reproducibility. Because this technique is based on random sequences of the primers, there is always a risk that contaminating DNA may get amplified. To minimize these nonspecific amplification, the following precautions are taken: (1) DNA templates should be from organisms recovered in pure culture, (2) include a sample in which no template is added (negative control), (3) standardize template DNA concentration in all samples to be tested, and (4) if possible, test all the desired samples in one batch.

**3.2.12. PCR-Restriction Enzyme Analysis.** The same principles behind REA, RFLP, and PFGE analyses of genomes of pathogens can be applied to PCR products. A PCR product, if it is sufficiently large, 1,000–2,000 base pairs or bp, can be digested with an endonuclease, and the generated DNA fragments can be resolved by gel electrophoresis and visualized. If such amplified products contain nucleotide changes at the restriction sites, varied electrophoretic band patterns will result from the digested PCR products. One drawback of this PCR-REA (or PCR-RFLP analysis) method is that part of the target sequence to be amplified has to be known ahead of time.

**3.2.13. PCR Typing Methods Based on DNA Repetitive Elements.** Repetitive DNA sequence elements are found in both prokaryotic and eukaryotic chromosomes. A variety of repetitive DNA elements have been identified in bacterial pathogens, which have been exploited to develop PCR-based strain typing assays. All of these procedures use the same approach, in which primers are designed to anneal to target repetitive sequences in the outward direction (outward primers) near the ends of these repetitive elements. The newly synthesized DNA strand extends away into the non-repetitive sequences. Thus, sequences located between the elements are amplified. Because these elements are interspersed in the chromosome at different locations between species or strains, the amplified products will be of variable number and MW. The size of the amplified products will depend on the distance between the repetitive elements and the quality of the polymerase, and other PCR conditions used. Of course, longer the distance between the repetitive elements, harder it is to amplify such a region.

Table 1 lists some of the major bacterial repetitive DNA elements that have been used for strain typing. One frequently used repetitive element genotyping method is called enterobacterial repetitive intergenic consensus (ERIC) sequence PCR.<sup>(41, 42)</sup> ERIC sequences are dispersed throughout the bacterial chromosome, especially of Gram-negative bacteria, and are located in extragenic regions.<sup>(41, 43)</sup> The exact function of ERIC sequences is not known.

Another repetitive element used for typing by PCR is called BOX, first reported in Gram-positive bacterial organisms.<sup>(44)</sup> BOX elements include mosaic combinations of three subunits, boxA (59 bp), boxB (45 bp), and boxC (50 bp),<sup>(44)</sup> which are all found in *Streptococcus pneumoniae*. Gram-negative organisms such as *E. coli* and *S. typhimurium* contain boxA-like sequences, but not boxB or boxC sequences. BOX elements have been widely applied to characterize the epidemiology of pneumococcal diseases.<sup>(45–48)</sup>

**3.2.14. Insertion Sequence (IS)-Based PCR.** One potential disadvantage of PCR tests based on cross-species, widely distributed repetitive elements like ERIC and BOX, is the possibility for a nonspecific amplification when DNA from one organism is contaminated with DNA from another. In some bacterial species, this potential problem can be overcome by techniques that target repetitive elements that are unique to the species.

Bacterial organisms carry certain genetic elements that “jump”, translocate, or transpose to new locations in the chromosome. Transposable elements that carry no genetic information except that needed by the elements to insert themselves into a new site are called simple transposons or insertion sequences (IS). Some insertion sequences are

found across species while others are species specific. If the copy number of an IS element in a bacterial strain is high enough and if it is randomly distributed in the chromosome, DNA sequences between these elements can be amplified, as described in repetitive element PCR assays above.

One example of a widely targeted IS element used to type a bacterial organism is IS6110 of *M. tuberculosis*. IS6110 is an IS unique to the members of the *M. tuberculosis* complex (*M. tuberculosis*, *M. bovis*, *M. microti*, and *M. africanum*). IS6110 has been used as a target as well as a probe for RFLP-Southern blot fingerprint analysis of *M. tuberculosis* strains. Ross and Dwyer reported a PCR-based method designed to amplify sequences between IS6110 elements to generate DNA fingerprint patterns.<sup>(49)</sup>

Of course, for IS6110-PCR-based typing method to work, the IS element has to be present in large numbers in *M. tuberculosis*. Thus, methods that take advantage of other different interspersed short repetitive DNA elements present in *M. tuberculosis* have been developed. There are at least four such short repetitive elements used to perform PCR-based genotyping of *M. tuberculosis*—(1) the polymorphic GC-rich tandem repeat sequence (PGRS), (2) the major polymorphic tandem repeat (MPTR) loci, (3) the exact tandem repeat (ETR) loci, and (4) the direct repeat (DR) elements located at a single locus in *M. tuberculosis* complex strains (Table 1).

PGRS elements together with IS6110 were exploited for a PCR-based typing of *M. tuberculosis* strains by a method called double-repetitive-element (DRE) PCR.<sup>(50)</sup> The greater number of repetitive elements represented by PGRS and IS6110 increases the chance of amplifying sequences between these elements. When the DRE-PCR method was compared to the standardized IS6110 RFLP analysis of *M. tuberculosis* strains during a TB outbreak in a prison in Havana, Cuba, the two methods identified the same clusters of TB patients who were clearly linked epidemiologically in the prison, which validated the PCR-based test for epidemiologic application.<sup>(51)</sup>

Like most PCR-based genotyping tests, the DRE-PCR test suffers from low reproducibility. However, because of its simplicity, the DRE-PCR assay may be useful for screening a large number of isolates prior to using a more labor-intensive but discriminating test such as IS6110-RFLP.

The variable numbers of tandem repeats (VNTR) found in the MPTR and ETR loci in *M. tuberculosis* have also been exploited to type this pathogen.<sup>(52–54)</sup> In *M. tuberculosis*, the MPTR contains 15-bp repeats with a single consensus sequence and high sequence variability between these repeats.<sup>(55, 56)</sup> Each ETR locus contains large tandem repeats with identical repeat DNA sequences and each locus has a unique repeat sequence.<sup>(53)</sup> Thus, the number of these repeat

units at each MPTR or ETR locus can be predicted from the size of the PCR products, which becomes the basis for VNTR genotyping. VNTR genotyping has been applied to a variety of bacterial organisms.<sup>(57–59)</sup> A recent variant of this method applied to type *M. tuberculosis* is called mycobacterial interspersed repetitive unit (MIRU) VNTR, which is based on *microsatellite*-like sequences scattered throughout the *M. tuberculosis* chromosome.<sup>(60, 61)</sup> By using 12 of these MIRU loci, several investigators have demonstrated that the resolution achieved by this typing method is comparable to that of IS6110-RFLP typing method.<sup>(62–64)</sup>

Another PCR-based method used to type *M. tuberculosis* is based on a region called direct repeat (DR) locus. This locus contains multiple copies of a repeat 36 bp sequence interspersed with unique spacer sequences which are 34–41 bp long.<sup>(55, 65)</sup> A typing method based on this variation in the spacer sequences is called spoligotyping (spacer oligotyping).<sup>(66)</sup> Oligonucleotides representing the spacers are synthetically made and covalently attached to a piece of membrane. This membrane then is used as a target for hybridization against PCR products obtained from amplification of the spacer sequences located between the DRs. PCR products used as probes are generated from the total genome of a strain of *M. tuberculosis* by one set of primers designed to amplify spaces between the DRs in the test strain. If the membrane contains an oligonucleotide representing a spacer sequence also present in the test strain, an amplified product will hybridize to it. This will generate a hybridization matrix, not unlike the microarray discussed earlier. Thus, unlike the previously described PCR-based typing methods, the patterns that are generated and compared are not electrophoretic fingerprints.

One important advantage of spoligotyping over the standardized IS6110-based method<sup>(14)</sup> is that it reliably differentiates *M. tuberculosis* from *M. bovis*, two members of the *M. tuberculosis* complex.<sup>(66)</sup> This is because of the low copy number of IS6110 in *M. bovis*. Discrimination of clinical isolates of *M. tuberculosis* by spoligotyping followed by DRE-PCR performed in French West Indies has been found to be almost as discriminating as the reference IS6110-RFLP method.<sup>(67)</sup> The spoligotyping membrane is commercially available. Thus, any laboratory capable of performing PCR should be able to characterize *M. tuberculosis* isolates by spoligotyping.

### 3.2.15. Multilocus Sequence Typing (MLST).

Many other PCR-based genotyping methods for bacterial pathogens have been described. They all use variations of the general theme described above (Table 1). One relatively new method called multilocus sequence typing (MLST) combines PCR assay with sequencing to assign a sequence type designation to a strain. Standardized MLST analysis

for most bacterial organisms is based on sequencing a PCR-amplified region from each of a set of unlinked housekeeping genes. For example, with *S. aureus*, these genes include seven genes: *arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi*, and *yqiL*.<sup>(68,69)</sup> Each of the amplified segments from these genes shows potential allelic variation in different *S. aureus* isolates. An MLST clone that is designated as a particular sequence type (ST) is defined as a strain belonging to a set of isolates with identical sequences at each of the seven loci. A methicillin-resistant *S. aureus* (MRSA) strain designated ST8 (also called PFGE type USA300) has recently been reported to cause widespread community outbreaks of severe skin and deep tissue infections in different regions of the United States.<sup>(70, 71)</sup> Because this is a sequence-based method, it is highly reproducible across different laboratories, and thus standardized protocols have been developed for several bacterial organisms. Protocols as well as ST designations for these organisms are maintained and accessible at <http://www.mlst.net>.

#### 4. Epidemiologic Problems Addressed by Genotyping Techniques

At least six common epidemiologic concerns related to bacterial infectious diseases are addressed by genotyping tests. They are (1) characterizing distribution and dynamics of disease transmission in geographically widespread areas; (2) identifying and quantifying risk in sporadic occurrence of infectious diseases; (3) stratifying data to refine epidemiologic study designs; (4) distinguishing pathovars from non-pathovars; (5) addressing hospital and institutional infectious disease problems; and (6) identifying genetic determinants of disease transmission. These are discussed separately below.

##### 4.1. Studying Dynamics of Disease Transmission in Geographically Widespread Areas

Much of the literature that applies molecular biology tools to genotype bacterial pathogens addresses this epidemiologic issue—characterizing geographic distribution of clones or clonal groups. Of course, this is an important component of epidemiology, but unfortunately these types of studies are somewhat limited in their ability to provide good understanding of the transmission dynamics of an infectious disease that yields practical solutions to disease prevention and control. Very often, these studies use convenient samples, in which available collection of microorganisms is genotyped first and questions are asked later. A typical

question that is asked afterward is “What do we do with the strain type data?” When no other types of data are collected (i.e., demographic characteristics of the human subjects from which the isolates are obtained, exposure measures, clinical features, etc.) the analysis of the isolates is limited to just describing the number of clones, the proportion that each clone contributes to the study collection, speculations about which clone descended from which, and their geographic distribution. Unfortunately, these types of data contribute little to public health.

For molecular epidemiology to be truly relevant, the right questions need to be asked first before any strains are genotyped. Tracking a strain from place to place and over time is one application of molecular biology techniques, but such an application is not sufficient to address important and relevant questions about disease transmission. In infectious disease epidemiology, researchers and practitioners ask the following types of questions<sup>(1)</sup>: How does an organism get introduced into a community, and how and why does it spread? What are the risk factors for infection and transmission in a community and how is the infection maintained in a community of hosts? How and why does a particular strain suddenly emerge to become the predominant strain in a community, and why does it disappear? How is drug-resistant infectious disease prevalence in a community affected by fluctuations in clonal distribution over time in that community? How is the prevalence of spectrum of clinical manifestations (mortality, severe disease, asymptomatic infection) of an infectious disease affected by clonal distribution of a pathogen in a community? Do unique biologic characteristics of a member of a bacterial species contribute to enhanced transmission and hence increased prevalence of that member in a community or institution? How do we use the knowledge we gain about the dynamics of disease transmission to implement effective disease control programs? How do we measure the effect and impact of the intervention made based on the knowledge acquired from the use of a particular laboratory test? Once these questions are asked, one can then decide how to collect the isolates, which collection of isolates should be genotyped, what comparison isolates (controls) should be collected, and which of the many strain typing techniques should be used.

##### 4.2. Identifying and Quantitating Risk in Sporadic Occurrence of Infectious Diseases

When a disease occurs as part of a recognized outbreak, it is possible to identify vehicles or risk factors for that illness. This is because in an outbreak, an analytical study

(case-control, cross-sectional, or cohort) can be designed to implicate a source or reservoir of an infectious agent responsible for the outbreak. If a risk factor or a vehicle of an outbreak is identified, the number of cases in that outbreak attributable to that risk factor (attributable risk fraction) can be calculated. In an outbreak, this would be 100% minus the background disease prevalence for that place. Unfortunately, in places like the United States, it is estimated from surveillance reports that less than 20% of reported cases of foodborne diseases, such as salmonellosis, come from recognized outbreaks.<sup>(72)</sup> The others are reported from sporadic cases. It is extremely difficult, if not impossible, to identify risk or estimate attributable risk fraction for sporadic diseases. The ability to genotype bacterial isolates from sporadic disease cases provides an opportunity to identify risks and calculate attributable fractions for those risks. One early example of such a study follows.

In late summer of 1981, case-control studies of two different salmonellosis outbreaks in New Jersey and Pennsylvania caused by *S. Newport* and *S. Typhimurium* implicated precooked roast beef produced at a meat processing plant in Philadelphia as the vehicle.<sup>(73)</sup> The distribution of the contaminated beef, of course, was not limited to just these two outbreak settings. It was distributed statewide in both New Jersey and Pennsylvania. The question that was asked was, “what proportion of the salmonellosis cases in these two states during this period that were not part of the two recognized outbreaks were caused by consumption of the contaminated precooked roast beef?” That is, the question asked for the proportion of all sporadic cases of salmonellosis in the two states attributable to the contaminated precooked roast beef.

None of the *S. typhimurium* isolates contained any plasmids, so this technique could not be applied to further subtype this *Salmonella* serotype. *S. newport* isolates from the two outbreaks and an isolate from the implicated meat product, however, contained identical 3.7 and 3.4 MDa plasmids.<sup>(74)</sup> *S. newport* isolates from persons in the two states not involved in the outbreaks were obtained from the respective state health department laboratories from the summer of 1981. In addition, *S. Newport* isolates from the same states before and after this summer were from other states during the same time period (geographic control), was obtained and analyzed. Telephone interviews of the people who got sick during the summer of 1981 revealed that a significantly greater proportion of those infected with a strain containing the 3.7 and 3.4 MDa plasmids ate precooked roast beef shortly before they developed diarrhea.<sup>(74)</sup> In the summer of 1981, nearly 45% of the *S. Newport* isolates from sporadic cases of salmonellosis in New Jersey and Pennsylvania belonged to this strain. Thus, the fraction of all sporadic

*S. Newport* diarrhea cases during this period that could be attributed to this single contaminated product was estimated to be 45%.

This study demonstrated that the incidence of diseases-like salmonellosis in a widespread geographic region could be rapidly and greatly affected by the introduction of a single contaminated food product. In fact, even after recall of the contaminated product, *S. Newport* strains associated with the summer outbreaks were isolated from two persons. It turns out these two cases occurred in siblings of an older child who had acquired the infection during the outbreak period after ingesting the beef.<sup>(74)</sup> Thus, salmonellosis in a community could be maintained by person-to-person transmission after introduction into that community by a contaminated product. The plasmid profile analysis, thus, helped not only to estimate an attributable fraction of a risk factor but also to describe the dynamics of community transmission of salmonellosis that would have been difficult, if not impossible, to do by conventional epidemiologic methods.

#### 4.3. Stratifying Data to Refine Epidemiologic Study Designs

Stratification of data is one of the most common analytical procedures used in epidemiology to reduce bias after data had already been collected. It helps to determine more precisely an association of an attribute with disease or outcome, controlling for the effects of a larger stratum to which that attribute belongs.<sup>(1)</sup> Because molecular strain typing methods can group data related to infectious agents by subtypes, the strain subtype information itself can be used to develop a new case definition, which may comprise one stratum.

For example, in a chest clinic in a large city, patients from whom *M. tuberculosis* is isolated would all be grouped as culture-positive tuberculosis (TB) patients. If one were to do a study to identify risk factors for TB, the initial case definition of TB may be patients from whom *M. tuberculosis* is isolated. Here, an analytical study to identify risk factor for TB is likely to identify host-related factors, such as immunosuppression, old age, diabetes, and poverty. However, if the *M. tuberculosis* isolates from different patients can be tested for drug susceptibility, then these TB patients may be reclassified, according to the drug resistance of their isolates. Such TB patients may then be regrouped into two new strata—those with multidrug-resistant TB (MDRTB) and those with drug-susceptible TB. Each patient stratum (MDRTB versus drug-susceptible TB) can then be used to define cases or controls or comparison group in a case-control analysis to identify a new risk factor (e.g., risk for MDRTB). Such an analysis will find risk factors that are more specific than the

risk factors identified for TB above. They may include, for example, previous history of TB treatment.

Then, if the MDRTB *M. tuberculosis* isolates from above are genotyped and the collection of these isolates are found to be comprised of one clone responsible for 40% of that collection, one can then look for risk factors for infection with that clone. Persons infected with that clone may share a common exposure, such as being exposed to a co-worker with TB at an office. The fact that 40% of them belonged to a cluster also suggests that the people infected with this clone were recently exposed and that they did not result from an infection that occurred many years ago.

One reported study illustrates this idea.<sup>(75)</sup> In the 1970s and 1980s in São Paulo, Brazil, multidrug-resistant strains of *Salmonella* serotype Typhimurium were a common cause of childhood diarrhea as well as meningitis.<sup>(76)</sup> The sources of these *Salmonella* organisms were unknown. In a multi-clinic study of acute diarrhea in children less than 6 years of age in the city in the early 1980s, *Salmonella* were isolated from 40 (11%) of 358 children with diarrhea, 22 of which were *S. Typhimurium*; all were multidrug resistant. A case-control study comparing children infected with *S. Typhimurium* versus other serotypes of *Salmonella* found recent hospitalization to be a risk factor for *S. Typhimurium* infection. In the study, children who came to the study clinics with diarrhea developed diarrhea at home. Children with history of recent hospitalization were previously hospitalized for illnesses such as pneumonia and acute respiratory infections that required antimicrobial drug treatment. Because it was known at that time that recent exposure to penicillins or antimicrobial agents was a risk factor for acquisition of multidrug-resistant salmonellosis,<sup>(77)</sup> the investigators were faced with the question of whether the children with *S. Typhimurium* salmonellosis got infected at home or during the previous hospitalization.

The investigators examined plasmid profile of the *S. Typhimurium* isolates. They identified four distinct plasmid profiles that accounted for 47% of the 22 isolates (Figure 1). Each profile was found to cluster in time and in a particular hospital in the city. The others spread over a 2-year period in a typical endemic pattern of occurrence. Thus, 47% of the *S. Typhimurium* infections could be traced to particular hospitals in the city.<sup>(77)</sup> Furthermore, the study showed that what appeared to be an endemic pattern of disease occurrence in a community was actually composed of at least four outbreaks that occurred in hospitals, unrecognized until the isolates were genotyped.

A similar observation was made by a more recent study using PFGE to subtype *S. Typhimurium* isolates from “sporadic” cases of salmonellosis in Minnesota between 1994 and 1998.<sup>(78)</sup> Of 174 distinct PFGE patterns found among 958

isolates, the investigators detected four large outbreaks in the state that had not been previously recognized.<sup>(78)</sup> They also identified many multidrug-resistant strains that had been previously described among *S. Typhimurium* isolates obtained by CDC in the 1970s–1980s.

Thus, in the above two examples of salmonellosis, strain typing analyses (plasmid profile or the PFGE strain typing) helped to “unmask” several discrete outbreaks from a background of so-called sporadic or “endemic” cases of salmonellosis in a city and state. Endemic salmonellosis appears to be comprised of many small outbreaks. This is one of the major discoveries about infectious disease epidemiology made in the last 25 years by the application of molecular biology tools. In fact, it challenges the traditional view about infectious disease occurrence and raises a rather provocative suggestion that there is no such thing as endemic salmonellosis; all cases of salmonellosis may actually be part of outbreaks that go undetected because of absence of strain type information.

Thus, as illustrated above, different laboratory tests identified distinct strata, each of which then could be used to define a new group (i.e., new case definition). Then each stratum could be compared against others in a case-control study design to identify a risk factor for one group. Genotyping data provide the most refined stratum, and if the sample size does not get too small, an analysis can be performed to identify a specific risk factor for that stratum. More refined the stratum, more specifically the risk factor for that stratum is identified. Furthermore, stratification by genotyping methods identifies risk factors related to the pathogen itself—specific site of transmission of a specific clone in the example of salmonellosis in São Paulo, Brazil. Thus, genotyping creates an opportunity to do an analytical study that generates a new hypothesis—refines a study design.

Most importantly, from a public health perspective, a study design that can identify risks for a refined stratum can also identify a more focused intervention option. Stratification of cases, for instance, by antibiograms of the case isolates usually identifies host-related factors as risk factors—e.g., previous exposure to antibiotics. This knowledge, while helpful, does not provide a focused intervention option, and often, as we all know, such knowledge does not contribute to long-lasting effective interventions. In the example of salmonellosis in Brazil, however, further stratification of the *Salmonella* isolates by plasmid profile analysis identified a more specific risk factor related to the isolates themselves. The knowledge that many cases of multidrug-resistant salmonellosis were occurring in specific hospitals in the city brought attention to the need to identify factors within the hospitals that contribute to *Salmonella* transmission. Today, nosocomial transmission of multidrug-resistant

salmonellosis in São Paulo is rare. Thus, this ability to create an opportunity to identify specific intervention options is another major contribution to infectious disease epidemiology made by molecular epidemiology. Many examples of such applications are described elsewhere.<sup>(1)</sup>

#### 4.4. Distinguishing Pathovars from Nonpathovars

*E. coli* can cause diarrhea, urinary tract infection (UTI), meningitis, septicemia, septic shock, and wound infections. Yet there are *E. coli* strains that colonize our intestine without ever causing disease. Among *E. coli* strains that cause diarrhea, there are several distinct groups, and they differ from those that cause UTI, meningitis, septicemia, and other diseases. Thus, being able to distinguish *E. coli* strains associated with a particular type of disease, and separating them from non-pathogenic variants of *E. coli*, is an important activity in epidemiologic investigation. This way of classifying strains is called virulence or pathogen typing. *E. coli* strains that cause disease are the best examples of human *pathotype* or *pathovar*.

While traditional methods are also available to separate *E. coli* into distinct virulence types, molecular biology techniques have greatly simplified this task. In particular, *E. coli* strains associated with diarrhea have come to be classified into distinct groups (enteropathogenic *E. coli*, enterotoxigenic *E. coli*, enteroinvasive *E. coli*, enterohemorrhagic *E. coli*, shigatoxin-producing *E. coli*, enteroaggregative *E. coli*, diffuse-adherent *E. coli*), based on a unique set of factors that define the group.<sup>(79)</sup> However, it should be noted that what made possible the attribution of these virulence factors to these groups of *E. coli* is that many of these groups have been recognized to cause outbreaks, either at the community level or in institutional settings.

In obvious outbreaks, it is not necessary to genotype the implicated pathogen to claim that there was an outbreak. In fact, it is the other way around. Outbreaks are used to validate a genotyping test itself. If a particular genotyping method shows that a strain maker is indistinguishable in all the isolates from an outbreak, but that it differs from comparison isolates obtained from sources unrelated to that outbreak, that method is said to be validated for epidemiologic investigation of that organism. That is, the method can be used to study apparent sporadic diseases caused by that organism.

However, there is another group of *E. coli* that causes extraintestinal infections (sometimes referred to as ExPEC, or extraintestinal pathogenic *E. coli*), which are not as obvious as to why they cause disease.<sup>(80)</sup> That is, it is not clear whether these *E. coli* strains are members of the normal intestinal flora that cause disease only when they happen to enter a sterile niche, such as the blood stream, bladder,

CSF, or deep tissue. Unlike the groups of *E. coli* that cause diarrhea, ExPECs do not carry any virulence factors that clearly distinguish them from those strains isolated from the intestine. Unlike diarrheagenic *E. coli*, ExPECs are not usually recognized to be associated with obvious outbreaks. This makes attribution of a clone of ExPEC to disease more difficult.

Thus, with ExPECs, the application of genotyping methods to address their epidemiology uses an approach distinct from that used to study diarrheagenic *E. coli*. A recent set of studies of community-acquired urinary tract infections (UTI) suggest that outbreaks of UTI do occur in community settings.<sup>(81–84)</sup> Such observation was used by investigators to postulate that some of the multidrug-resistant uropathogenic *E. coli* strains causing UTI in women may have food animal as a reservoir.<sup>(81, 85–87)</sup>

In diseases such as UTI where an outbreak occurrence is not obvious, a genotyping method itself may provide a clue that an outbreak may have occurred in a community; this observation generates a new hypothesis about disease transmission. The hypothesis, then, can be tested by conducting a case-control or cross-sectional study. For example, with UTI, a study can be performed to see if there is any association between a predominant clonal group of uropathogenic *E. coli* identified by a strain typing method and a risk factor. If so, this would provide evidence that there was indeed an outbreak. Here, the ultimate confirmation of the hypothesis will be made by showing that the removal of that risk factor leads to amelioration of the recognized problem.

Same questions arise in distinguishing organisms that also happen to colonize non-sterile niches of the human host or occur as saprophytes in the environment. Examples of the latter include *S. pneumoniae*, *Helicobacter pylori*, *S. aureus*, and Gram-negative bacteria other than *E. coli*. Approaches similar to those used to study ExPECs can be applied for these organisms.

#### 4.5. Addressing Hospital and Institutional Infectious Disease Problems

Hospital-associated or nosocomial infections are defined elsewhere in this book (See Chapter 26). In developed countries, organisms that are most frequently associated with nosocomial infections are usually members of the “normal” flora in the non-sterile niche of a human host, which include the skin, nasopharynx, and colon, or saprophytic organisms in the hospital environment. The most common such agents of hospital infections include coagulase-negative *Staphylococci*, *S. aureus*, *Enterococcus spp.*, *Clostridium difficile*, *E. coli*, *Enterobacter spp.*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Acinetobacter spp.*, *Proteus spp.*,

*Stenotrophomonas maltophilia*, *Serratia* spp., and *Candida* spp. Non-commensal or non-saprophytic pathogens that cause hospital infections include those transmitted by blood products (e.g., HIV, HBV, HCV, *Yersinia enterocolitica*) and respiratory pathogens (including severe acute respiratory syndrome [SARS]-associated coronavirus). In developing countries infections caused by *M. tuberculosis* and *Salmonella* introduced into a hospital by an index case, staff, visitors, or contaminated food are still common.

Genotyping bacterial organisms isolated in hospital or other institutional settings have become routine in many places, especially in developed countries. In fact, these techniques have become an indispensable part of hospital infection control activities.<sup>(88)</sup> While probe-based methods are most frequently used to detect drug-resistant organisms to augment clinical management of patients, many hospitals use genotyping tests to support nosocomial infection control practices. Teaching hospitals use genotyping methods to study nosocomial infection epidemiology.

Many of the infectious disease epidemiologic concerns in institutional settings are distinct from those related to field or community settings. While obvious outbreaks do occur in hospitals, very often, clusters of infectious diseases observed in hospitals are difficult to characterize as outbreaks. This is because patients in hospitals have host-related factors (immunosuppression, prolonged antimicrobial drug exposure, underlying illness) that predispose them to infection with “normal flora”, saprophytes, or recognized pathogens. The distinction among colonization, contamination, and infection becomes blurred and is a frequent topic of discussion among hospital staff. One important question that arises in these settings, therefore, is, “do these organisms have biologic characteristics that enable them to cause disease or do they cause disease only because of host susceptibility factors?”

Another major area of concern in hospital infections, of course, is the drug-resistant infections. Very often, when risk factor studies for drug-resistant infections in hospitals are carried out, they find host-related factors—antimicrobial drug exposure, old age, underlying medical conditions, prolonged hospital stay, etc. Conventional epidemiologic studies rarely identify factors that can be attributed directly to the organisms themselves—sources or reservoir, transmissibility, and their biologic characteristics that determine disease production versus asymptomatic colonization. Thus, in hospital epidemiology, another important question is, “do antimicrobial drug-resistant microorganisms predominate in hospital settings because they are selected by the use of antimicrobial agents, or do they predominate because of distinct biologic characteristics possessed by a limited number of clonal groups that happen to be drug resistant?” Molecular

epidemiologic methods are beginning to address these questions.

Given that normal flora and saprophytes found in hospital settings are comprised of a large number of strain types, one would expect drug selection pressure to generate a large number of drug-resistant strains to cause hospital infections. One of the most intriguing observations made recently about drug-resistant bacterial hospital infections, however, is that, despite strong antimicrobial drug selection pressures, most of these infections are actually caused by a limited set of bacterial strains. Most studies found that drug-resistant strains that are isolated from hospital invasive disease sources are comprised of a limited number of strains. In New York, a 2-year survey of vancomycin-resistant *Enterococcus faecium* (VRE) infections at a teaching hospital found 54 different PFGE subtypes among 182 isolates. One-third of these were accounted for by just two PFGE subtypes.<sup>(89)</sup> In Tennessee, a similar survey in different hospitals over a 3-year period showed that among 34 different PFGE types, one was responsible for 61% of 106 VRE isolates obtained from seven different hospitals.<sup>(90)</sup> In Tennessee, the investigators found that the introduction of the predominant PFGE types at different institutions was caused by new patients transferred to these institutions. Thus, these studies suggest that a clonal strain carried by a transfer patient, once introduced into a hospital setting, can rapidly become predominant in that hospital, and that such a mode of transmission is more important than the selective pressures of antimicrobial drugs in establishing a large number of drug-resistant nosocomial infections. Without the genotype data, such a conclusion would be difficult to make. This type of observation also helps to streamline nosocomial infection control practices. In this case, the recommendation based on the observation is to screen transfer patients for VRE before they are sent to a particular ward in the recipient hospital.

The observations described above with VRE apply to other hospital infections, including those caused by methicillin-sensitive (MSSA) or methicillin-resistant *S. aureus* (MRSA), *E. coli*, and other Gram-negative bacteria. It is beyond the scope of this chapter to discuss each of these agents of hospital infection, and the reader is encouraged to refer to reference (1) for further reading.

#### 4.6. Identifying Genetic Determinants of Disease Transmission

The next stage in the evolution of laboratory-based epidemiology is the methodology to identify genetic basis for disease transmission. This, of course, applies to both the host and the pathogen. Host factors that enhance the

likelihood of development of a clinical manifestation after a bacterial infection could also render such hosts possible new sources of transmission to other hosts. Communities with a high prevalence of persons with HIV infection or AIDS facilitate greater likelihood of transmission of tuberculosis, just because of the greater prevalence of cases of tuberculosis. Thus, host genetic factors that have a similar effect on infection could contribute to enhanced transmission of a disease. Studies that examine host genetic factors that affect disease transmission are just beginning.<sup>(91)</sup> Of course, in addition to host factors, environment can greatly influence infectious disease distribution, including temperature, weather, season, urbanization, industrialization, and socioeconomic conditions. Such influences on disease transmission and epidemiology of infectious diseases require a separate chapter in itself. In this chapter, only pathogen genetic factors that facilitate transmission and therefore disease distribution are dealt with.

As mentioned earlier, one of the common molecular epidemiologic exercises involves attempts to explain over-representation of clones, clonal groups, or clonal complex strains in a community or institutional settings. Many of these can be explained on the basis of epidemiology, such as point-source contamination of a widely distributed food product (as in salmonellosis) or airborne transmission from a single index case (as in tuberculosis). However, we also know of many situations in which there is no obvious explanation for the predominance of one particular strain of a microorganism. Once epidemiologic explanation for the predominance is ruled out, possible biologic bases for this observation needs to be examined.

All pathogens have developed their own distinct strategies to transmit themselves to new hosts within a context of an environmental setting. Studies to date that have attempted to identify a pathogen's specific genetic determinants of transmission have used two different approaches: (1) those that attempt to identify a type of infection in a host that enhances opportunities for transmission to new hosts and (2) those that attempt to identify a unique niche or reservoir that becomes a major source of transmission to human or other animal hosts.

Tuberculosis can arise from an infection that occurred in the remote past (reactivation disease) or from a recent new infection (rapidly progressive disease). As mentioned earlier, over-representation of a particular strain of *M. tuberculosis* in a community could result from recent transmission from a highly infectious index case (e.g., someone with a cavitary lung lesion working in an enclosed space with an opportunity for contact with many others). In a study conducted in the early 1990s in New York City, one drug-susceptible strain of *M. tuberculosis* called "C strain" was found to be responsible for nearly 10% of all newly diagnosed cases

of tuberculosis.<sup>(92)</sup> A series of case-control studies found C strain to be significantly associated with injection drug use. However, there was no obvious epidemiologic relationship among the injection drug users infected with this strain, nor was there any epidemiologic relationship among non-users infected with the same strain. This prompted investigators to look for a biologic basis for the predominance of C strain in New York City.

Ultimately, C strain but none of the other *M. tuberculosis* strains identified in the city was found to be resistant to reactive nitrogen intermediates (RNI) and hydrogen peroxide in vitro.<sup>(92)</sup> These are effector molecules that *M. tuberculosis* is likely to encounter inside macrophages in a host after infection. It was hypothesized that since injection drug users, who are normally a high-risk group for tuberculosis, regularly inject themselves with a variety of antigens that stimulate their macrophages to express RNI and hydrogen peroxide, such a practice will eventually select for RNI and hydrogen peroxide-resistant strain of *M. tuberculosis* to emerge in that risk group. If such a strain infects a non-user, it will cause a rapidly progressive disease because such a host does not carry macrophages making RNI and hydrogen peroxide, which would otherwise control the initial infection with *M. tuberculosis*. That is, infection/disease ratio increases when one is infected with such a strain. Thus, more people will develop rapidly progressive disease in a community where such a strain is circulating. This would explain the predominance of C strain in New York City. Hence, in this case, a human behavior appears to have selected for a new strain of *M. tuberculosis* that eventually affected the epidemiology of tuberculosis in New York City.

C strain was further studied in the laboratory and ultimately, a gene called *noxR1* was identified that when transferred into RNI-susceptible *E. coli* or *M. smegmatis* strains rendered these hosts resistant to RNI.<sup>(93)</sup> Subsequently, additional genes that exhibited RNI and hydrogen peroxide resistance were identified,<sup>(94, 95)</sup> which opened a new area of tuberculosis pathogenesis research—study of the mechanisms of resistance to natural antimicrobial effector molecules elaborated by the human host against *M. tuberculosis*. These genes may not completely explain the mode of transmission and reason for the predominance of *M. tuberculosis* strain in a community, but these are the earliest studies that applied molecular epidemiologic methods to identify a particular clone of a clinical isolate to study its biologic basis for its distribution and occurrence in a community. Here, the over-representation in New York City of an *M. tuberculosis* strain was attributed to this strain's enhanced ability to cause rapidly progressive disease after a new infection. A mechanism by which an *M. tuberculosis* strain can cause rapidly progressive disease was sought; the strain's resistance to macrophage effector molecules was found to be



possibly related to this mechanism. Genes that mediate this phenotype were identified.

The second type of approach used to identify a genetic basis for disease transmission focused on salmonellosis caused by *S. enterica* serotype Enteritidis. *S. Enteritidis* is the most common cause of egg-associated salmonellosis in developed countries. In a study of henhouses in the United States, 18–25% of captured mice were shown to be infected with *S. Enteritidis*, while *S. Typhimurium* was cultured from only 1.5% from the same henhouses.<sup>(96)</sup> In the same study, chicken eggs contaminated with *S. Enteritidis* were found in 9 (43%) of the 21 henhouses.<sup>(96)</sup> Unlike the other *Salmonella* serotypes that infect the surface of egg shells, *S. Enteritidis* infect eggs internally.<sup>(97)</sup> *S. Enteritidis* multiply inside the egg if left at room temperature. Thus, eggs serve as an “amplifier” of transmission to human hosts, which explain the unique epidemiology of salmonellosis caused by this pathogen. Thus, this second approach to identify a genetic basis for disease transmission attempted to look for a biologic mechanism for the establishment of a reservoir for a pathogen that then serves as a major source for human infection. The question, therefore, is, what enables *S. Enteritidis* and not the other *Salmonella* serotypes to survive in the egg?

This question was addressed by Lu et al., who examined multiple clinical isolates of *S. Enteritidis* for susceptibility to chicken egg albumen. They first found that most clinical isolates of *S. Enteritidis* were relatively more resistant to egg albumen than most clinical isolates of *S. Typhimurium*.<sup>(98)</sup> One *S. Enteritidis* strain that was highly resistant was analyzed further. Lu et al. eventually identified a gene called *yafD* that when transferred into an egg-susceptible strain of *S. Typhimurium* rendered it resistant.<sup>(98)</sup> Disruption of *yafD* in the original resistant *S. Enteritidis* strain made it more susceptible to egg albumen. The mechanism of resistance to albumen associated with *yafD* is not yet known.

Thus, once again in this example, a molecular epidemiologic approach first identified a clinical strain of *S. Enteritidis*, which when tested further for a distinct phenotype led to the identification of a gene associated with that phenotype. These types of studies should yield new understanding of how infectious disease transmission occurs and should provide novel approaches to disease control and prevention. These are just beginning to be applied in this new phase of molecular epidemiology research.

## 5. Summary

Molecular epidemiology should not be regarded as just another technique or a tool, but as an established discipline in epidemiology that can advance our knowledge about

infectious diseases in ways that were not previously possible. Molecular epidemiology has begun to challenge old notions and reveal new paradigms concerning the epidemiology of infectious diseases. This chapter introduced new concepts that emerged within this discipline and a sample of its applications. The literature is vast and molecular epidemiologic methods are now a major component of most laboratory-based infectious disease epidemiologic studies. However, it is a constantly evolving discipline and should continue to provide new insights into elucidating epidemiology of infectious diseases in the future.

## 6. References

1. Riley LW. Molecular epidemiology of infectious diseases: Principles and practices. Washington, DC: American Society for Microbiology Press, 2004.
2. Hitchens AP, L.C. Leikind. The introduction of agar-agar into bacteriology. *J. Bacteriol.* 1939;37:485–493.
3. Struelens MJ, and the Members of the European Study Group on Epidemiological Markers (ESGEM). Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. *Clin. Microbiol. Infection.* 1996;2:2–11.
4. Tenover FC, R.D. Arbeit, R.V. Goering, and the Molecular Typing Working Group of the Society for Healthcare Epidemiology of America. How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. *Infect. Control. Hosp. Epidemiol.* 1997;18:426–439.
5. Maslow JN, M.E. Mulligan, and R.D. Arbeit. Molecular epidemiology: the application of contemporary techniques to typing bacteria. *Clin. Infect. Dis.* 1993;17:153–162.
6. Lipuma JJ. Molecular tools for epidemiologic study of infectious diseases. *Pediatr. Infect. Dis. J.* 1998;17:667–675.
7. Bopp CA, F.W. Brenner, P.I. Fields, J.G. Wells, and N.A. Strockbine. *Escherichia, Shigella, and Salmonella, In Manual of Clinical Microbiology.* Murray, P.R., E.J. Baron, J.H. Jorgensen, M.A. Pfaller, and R.H. Tenover (eds). 8 ed. Washington, DC: ASM Press, 2003 (Manual of Clinical Microbiology).
8. Varma A, and K.J. Kwong-Chung. Restriction fragment polymorphism in mitochondrial DNA of *Cryptococcus neoformans*. *J. Gen. Microbiol.* 1989;135:3353–3362.
9. Morel C, E. Chiari, E.P. Camargo, D.M. Mattei, A.J. Romanha, and L. Simpson. Strains and clones of *Trypanosoma cruzi* can be characterized by pattern of restriction endonuclease products of kinetoplast DNA minicircles. *Proc. Natl. Acad. Sci. USA* 1980;77:6810–6814.
10. Holmberg SD, I.K. Wachsmuth, F.W. Hickman-Brenner, and M.L. Cohen. Comparison of plasmid profile analysis, phage typing, and antimicrobial susceptibility testing in characterizing *Salmonella typhimurium* isolates from outbreaks. *J. Clin. Microbiol.* 1984;19:100–104.
11. Drahos D, J. Brackin, and G. Barry. Bacterial strain identification by comparative analysis of chromosomal DNA restriction patterns. *Phytopathology* 1985;75:1381.
12. Lazo GR, R. Roffey, and D.W. Gabriel. Pathovars of *Xanthomonas campestris* are distinguishable by restriction fragment length polymorphism. *Int. J. Syst. Bacteriol.* 1987;37:214–221.

13. Southern EM. Analysis of restriction fragment patterns from complex deoxyribonucleic acid species. *Biochem. Soc. Symp.* 1979;44:37–41.
14. van Embden JD, M.D. Cave, J.T. Crawford, J.W. Dale, K.D. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, T.M. Shinnick, and P.M. Small. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J. Clin. Microbiol.* 1993;31:406–409.
15. Maslow JN, T. Whittam, R.A. Wilson, M.E. Mulligan, C. Gilks, K.S. Adams, and R.D. Arbeit. Clonal relationship among bloodstream isolates of *Escherichia coli*. *Infect. Immun.* 1995;63:2409–2417.
16. Schwartz DC, and C.R. Cantor. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* 1984;37:67–75.
17. Swaminathan B, T.J. Barrett, S.B. Hunter, R.V. Tauxe and the CDC PulseNet Task Force. PulseNet: The molecular subtyping network for foodborne bacterial disease surveillance, United States. *Emerg. Infect. Dis.* 2001;7:382–389.
18. Cody SH, M.K. Glynn, J.A. Farrar, K.L. Cairns, P.M. Griffin, J. Kobayashi, M. Fyfe, R. Hoffman, A.S. King, J.H. Lewis, B. Swaminathan, R.G. Bryant, and D.J. Vugia. An outbreak of *Escherichia coli* O157:H7 infection from unpasteurized commercial apple juice. *Ann. Intern. Med.* 1999;130:202–209.
19. Centers for Disease Control and Prevention. Multistate outbreak of *Salmonella* serotype Agona infections linked to toasted oats cereal—United States, April–May, 1998. *Morb. Mortal. Wkly. Rep.* 1998;47:462–464.
20. Centers for Disease Control and Prevention. Multistate outbreak of listeriosis—United States, 1998. *Morb. Mortal. Wkly. Rep.* 1998;47:1085–1086.
21. Fitzgerald JR, and J.M. Musser. Evolutionary genomics of pathogenic bacteria. *Trends Microbiol.* 2001;9:547–553
22. van de Vijver MJ, Y.D. He, L.J. van't Veer, H. Dai, A.A. Hart, D.W. Voskuil, G.J. Schreiber, J.L. Peterse, C. Roberts, M.J. Marton, M. Parrish, D. Atsma, A. Witteveen, A. Glas, L. Delahaye, T. van der Velde, H. Bartelink, S. Rodenhuis, E.T. Rutgers, S.H. Friend, and R. Bernards. A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* 2002;347:1999–2009.
23. Rota PA, M.S. Oberste, S.S. Monroe, W.A. Nix, R. Campagnoli, J.P. Icenogle, S. Peñaranda, B. Bankamp, K. Maher, M. Chen, S. Tong, A. Tamin, L. Lowe, M. Frace, J.L. DeRisi, Q. Chen, D. Wang, D.D. Erdman, T.C.T. Peret, C. Burns, T.G. Ksiazek, P.E. Rollin, A. Sanchez, S. Liffick, B. Holloway, J. Limor, K. McCaustland, M. Olsen-Rasmussen, R. Fouchier, S. Günther, A.D.M.E. Osterhaus, C. Drost, M.A. Pallansch, L.J. Anderson, and W.J. Bellini. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 2003;300:1394–1399.
24. Fitzgerald J.R. SDR, E. Ruotsalainen, T.J. Tripp, M. Liu, R. Cole, P. Kuusela, P.M. Schlievert, A. Jarvinen, J.M. Musser. Genome diversification in *Staphylococcus aureus*: Molecular evolution of a highly variable chromosomal region encoding the Staphylococcal exotoxin-like family of proteins. *Infect Immun.* 2003;71:2827–2838.
25. Alland D. TSW, M.B. Murray, M.D. Cave, M.H. Hazbon, K. Dix, M. Kokoris, A. Duesterhoeft, J.A. Eisen, C.M. Fraser, R.D. Fleischmann. Modeling bacterial evolution with comparative-genome-based marker systems: application to *Mycobacterium tuberculosis* evolution and pathogenesis. *J. Bacteriol.* 2003;185:3392–3399.
26. Filliol I, A.S. Motiwala, M. Cavatore, W. Qi, M.H. Hazbon, M. Bobadilla del Valle, J. Fyfe, L. Garcia-Garcia, N. Rastogi, C. Sola, T. Zozio, M.I. Guerrero, C.I. Leon, J. Crabtree, S. Angiuoli, K.D. Eisenach, R. Durmaz, M.L. Joloba, A. Rendon, J. Sifuentes-Osornio, A. Ponce de Leon, M.D. Cave, R. Fleischmann, T.S. Whittam, D. Alland. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J. Bacteriol.* 2006;188:759–772.
27. Fodor SP, J.L. Read, M.C. Pirrung, L. Stryer, A.T. Lu, and D. Solas. Light-directed, spatially addressable parallel chemical synthesis. *Science* 1991;251:767–773.
28. Lipshutz RJ, S.P. Fodor, T.R. Gingeras, and D.J. Lockhart. High density synthetic oligonucleotide arrays. *Nat. Genet* 1999;21:20–24
29. Lockhart DJ, and E.A. Winzeler. Genomics, gene expression, and DNA arrays. *Nature* 2000;405:827–836
30. Behr MA, M.A. Wilson, W.P. Gill, H. Salamon, G.K. Schoolnik, S. Rane, and P.M. Small. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 1999;284:1520–1523.
31. Kato-Maeda M, J.T. Rhee, T.R. Gingeras, H. Salamon, J. Drenkow, N. Smittipat, and P.M. Small. Comparing genomes within the species *Mycobacterium tuberculosis*. *Genome Res.* 2001;11:547–554.
32. Goguet de la Salmoniere YO, C.C. Kim, A.G. Tsolaki, A.S. Pym, M.S. Siegrist, P.M. Small. High-throughput method for detecting genomic-deletion polymorphisms. *J. Clin. Microbiol.* 2004;42:2913–2918.
33. Belli A, D. Farcia, X. Palacios, B. Rodriguez, S. Valle, E. Videia, E. Tinoco, F. Marin, and E. Harris. Widespread atypical cutaneous leishmaniasis caused by *Leishmania chagasi* in Nicaragua. *Am. J. Trop. Med. Hyg* 1999;6:380–385.
34. Harris E, T.G. Roberts, L. Smith, J. Selle, L.D. Kramer, S. Valle, E. Sandoval, and A. Balmaseda. Typing of dengue viruses in clinical specimens and mosquitoes by single-tube multiplex reverse transcriptase PCR. *J. Clin. Microbiol.* 1998;36:2634–2639.
35. Noyes H, M. Chance, C. Ponce, E. Ponce, and R. Maingo. Leishmania chagasi: genotypically similar parasites from Honduras cause both visceral and cutaneous leishmaniasis in humans. *Exp. Parasitol.* 1997;85:264–273.
36. Taha M-K. Simultaneous approach for nonculture PCR-based identification and serogroup prediction of *Neisseria meningitidis*. *J. Clin. Microbiol.* 2000;38:855–857.
37. Tornieporth NG, J. John, K. Salgado, P. Jesus, E. Latham, C. Melo, S. Gunzburg, and L.W. Riley. Differentiation of pathogenic *E. coli* in Brazilian children by polymerase chain reaction. *J. Clin. Microbiol.* 1995;33:1371–1374.
38. Berg DE, N.S. Akopyants, and D. Kersulyte. Fingerprinting microbial genomes using the RAPD and AP-PCR method. *Methods Mol. Cell. Biol.* 1994;5:13–24.
39. Welsh JaMM. Genomic fingerprints produced by PCR with consensus tRNA gene primers. *Nucleic. Acids Res.* 1991;19:861–866.
40. Williams JGK, A.R. Kubelik, K.J. Livak, J.A. Rafalski, and S.V. Tingey. DNA polymorphisms amplified by arbitrary primers are useful as genetic-markers. *Nucleic. Acids Res.* 1990;18:6531–6535.
41. Hulton CSJ, C.F. Higgins, and P.M. Sharp. ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium*, and other enterobacteria. *Molec. Microbiol.* 1991;5:825–834.
42. Versalovic J. TK, and J.R. Lupski. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic. Acids Res.* 1991;19:6823–6831.
43. Stern MJ, G.F.L. Ames, N.H. Smith, E.C. Robinson, and C.F. Higgins. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* 1984;37:1015–1026.
44. Martin BO, M. Humbert, E. Camara, J. Guenzi, T. Walker, P. Mitchell, M. Andrew, G. Prudhomme, R. Alloing, D. Hakenbeck,

- D.A. Morrison, G.J. Boulnois, and J.-P. Claverys. A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. *Nucleic Acids Res.* 1992;20:3479–3483.
45. Ko AI, J.N. Reis, S.J. Coppola, E.L. Gouveia, R.M. Pinheiro, T.S. Lobo, R.M. Pinheiro, K. Salgado, C.M.R. Dourado, J. Tavares-Neto, H. Rocha, M.G. Reis, W.D. Johnson, Jr. and L.W. Riley. Clonally related penicillin-nonsusceptible *Streptococcus pneumoniae* serotype 14 from cases of meningitis in Salvador, Brazil. *Clin. Infect. Dis.* 2000;30:78–86.
  46. Lalitha MK, M. Baarnhielm, E. Khilstrom, and G. Kronvall. Epidemiological typing of *Streptococcus pneumoniae* from various sources in Sweden and India using Box A PCR fingerprinting. *APMIS* 1999;107:389–394.
  47. Rodriguez-Barradas MC, R.A. Tharapel, J.E. Groover, K.P. Giron, C.E. Lacke, E.D. Houston, R.J. Hamill, M.C. Stenhoff, and D.M. Musher. Colonization by *Streptococcus pneumoniae* among human immunodeficiency virus-infected adults: prevalence of antibiotic resistance, impact of immunization, and characterization by polymerase chain reaction with BOX primers of isolates from persistent *S. pneumoniae* carriers. *J. Infect. Dis.* 1997;175:590–597.
  48. van Belkum A, M. Sluijter, R. de Groot, H. Verbrugh, and P.W. Hermans. Novel BOX repeat PCR assay for high-resolution typing of *Streptococcus pneumoniae* strains. *J. Clin. Microbiol.* 1996;34:1176–1179.
  49. Ross BCaBD. Rapid, simple method for typing isolates of *Mycobacterium tuberculosis* by using the polymerase chain reaction. *J. Clin. Microbiol.* 1993;31:5329–334.
  50. Friedman CR, M.Y. Stoeckle, W.D. Johnson, and L.W. Riley. Double repetitive element PCR method for subtyping *Mycobacterium tuberculosis* clinical isolates. *J. Clin. Microbiol.* 1995;33:1383–1384.
  51. Montoro E, J. Valdivia, and S.C. Leao. Molecular fingerprinting of *Mycobacterium tuberculosis* isolates obtained in Havana, Cuba, by IS6110 restriction fragment length polymorphism analysis and by the double repetitive element PCR method. *J. Clin. Microbiol.* 1998;36:3099–3102.
  52. Frothingham R. Differentiation of strains in *Mycobacterium tuberculosis* complex by DNA sequence polymorphisms, including rapid identification of *M. bovis* BCG. *J. Clin. Microbiol.* 1995;33:840–844.
  53. Frothingham R, and W.A. Meeker-O'Connell. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology* 1998;144:1189–1196.
  54. Goyal M, D. Young, Y. Zhang, P.A. Jenkins, and R.J. Shaw. PCR amplification of variable sequence upstream of *katG* gene to subdivide strains of *Mycobacterium tuberculosis* complex. *J. Clin. Microbiol.* 1994;32:3070–3071.
  55. Hermans PWM, D. van Soolingen, and J.D.A. van Embden. Characterization of a major polymorphic tandem repeat in *Mycobacterium tuberculosis* and its potential use in the epidemiology of *Mycobacterium kansasii* and *Mycobacterium goodii*. *J. Bacteriol.* 1992;174:4157–4165.
  56. Shinnick TM. A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. *Nucleic Acids Res.* 1990;18:6503–6508.
  57. Pourcel C, Y. Vidgop, F. Ramière, G. Vergnaud, and C. Tram. Characterization of a tandem repeat polymorphism in *Legionella pneumophila* and its use for genotyping. *J. Clin. Microbiol.* 2003;41:1819–1826.
  58. Sabat A, J. Krzyszton-Russjan, W. Strzalka, R. Filipek, K. Kosowska, W. Hryniewicz, J. Travis, and J. Potempa. New method for typing *Staphylococcus aureus* strains: multiple-locus variable-number tandem repeat analysis of polymorphism and genetic relationships of clinical isolates. *J. Clin. Microbiol.* 2003;41:1801–1804.
  59. Keim P, L.B. Price, A.M. Klevytska, K.L. Smith, J.M. Schupp, R. Okinaka, P.J. Jackson, and M.E. Hugh-Jones. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J. Bacteriol.* 2000;182:2928–2936.
  60. Magdalena J, A. Vachée, P. Supply, and C. Locht. Identification of a new DNA region specific for members of *Mycobacterium tuberculosis* complex. *J. Clin. Microbiol.* 1998;36:937–943.
  61. Supply P, J. Magdalena, S. Himpens, and C. Locht. Identification of novel intergenic repetitive units in a mycobacterial two component system operon. *Mol. Microbiol.* 1997;26:991–1003.
  62. Hawkey PM, E.G. Smith, J.T. Evans, P. Monk, G. Bryan, H.H. Mohamed, M. Bardhan, and R.N. Pugh. 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.* 2003;41:3514–3520.
  63. Mazars E, S. Lesjean, A.-L. Banuls, M. Gilbert, V. Vincent, B. Gicquel, M. Tibayrenc, C. Locht, and P. Supply. High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc. Natl. Acad. Sci., USA* 2001;98:1901–1906.
  64. Supply P, S. Lesjean, E. Savine, K. Kremer, D. van Soolingen, and C. Locht. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *J. Clin. Microbiol.* 2001;39:3563–3571.
  65. Groenen PMA, A.E. Bunschoten, D. van Soolingen, and J.D.A. van Embden. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*: application for strain differentiation by a novel method. *Mol. Microbiol.* 1993;105:1057–1065.
  66. Kamerbeek J, L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuliper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal, and J.D.A. van Embden. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* 1997;35:907–914.
  67. Sola C, L. Horgen, J. Maisetti, A. Devallois, K.S. Goh, and N. Rastogi. Spoligotyping followed by double-repetitive-element PCR as rapid alternative to IS6110 fingerprinting for epidemiologic studies of tuberculosis. *J. Clin. Microbiol.* 1998;36:1122–1124.
  68. Maiden MCJ, J.S. Bygraves, E. Feil, G. Morelli, J.E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D.A. Caugant, I.M. Feavers, M. Achtman, and B.G. Spratt. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic organisms. *Proc. Natl. Acad. Sci. USA* 1998;95:3140–3145.
  69. Enright MC, N.P. Day NP, C.E. Davies, S.J. Peacock, and B.G. Spratt. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J. Clin. Microbiol.* 2000;38:1008–1015.
  70. Miller LG, F. Perdreau-Remington, G. Rieg, S. Mehdi, J. Perlroth, A.S. Bayer, A.W. Tang, T.O. Phung, B. Spellberg. Necrotizing fasciitis caused by community-associated methicillin-resistant *Staphylococcus aureus* in Los Angeles. *N. Engl. J. Med.* 2005;352:1445–1453.
  71. Pan ES, B.A. Diep, E.D. Charlebois, C. Auerswald, H.A. Carleton, G.F. Sensabaugh, F. Perdreau-Remington. Population dynamics of nasal strains of methicillin-resistant *Staphylococcus aureus*—and their relation to community-associated disease activity. *J. Infect. Dis.* 2005;192:811–818.
  72. Cohen ML, R.V. Tauxe. Drug-resistant *Salmonella* infections in the United States: an epidemiologic perspective. *Science* 1986;234:964–969.

73. Centers for Disease Control and Prevention. Multistate outbreaks of salmonellosis associated with precooked roast beef. *Morb. Mortal. Wkly. Rep.* 1981;30:391–392.
74. Riley LW, G.T. DiFerdinando, T.M. DeMelfi, and M.L. Cohen. Evaluation of isolated cases of salmonellosis by plasmid profile analysis: introduction and transmission of a bacterial clone by precooked roast beef. *J. Infect. Dis.* 1983;148:12–17.
75. Riley LW, B.S.O. Ceballos, L.R. Trabulsi, M.R.F. Toledo, and P.A. Blake. The significance of hospitals as reservoirs for endemic multiresistant *Salmonella typhimurium* causing infection in urban Brazilian children. *J. Infect. Dis.* 1984;150:236–241.
76. Pessoa GVA, K. Irino, C.E.A. Melles, CT. Calzada, M. Raskin, and E. Kano. Ocorrência de bacterias enteropatogenicas em São Paulo no Septenio 1970–1976. II. O surto epidemico de *Salmonella typhimurium* em São Paulo. *Rev. Inst. Adolfo Lutz.* 1978;38:107–127.
77. Riley LW, M.L. Cohen, J.E. Seals, M.J. Blaser, K.A. Birkness, N.T. Hargrett, S.M. Martin, and R.A. Feldman. Importance of host factors in human salmonellosis caused by multiresistant strains of *Salmonella*. *J. Infect. Dis.* 1984;149:878–883.
78. Bender JB, C.W. Hedberg, D.J. Boxrud, J.M. Besser, J.H. Wicklund, K.E. Smith, and M.T. Osterholm. Use of molecular subtyping in surveillance for *Salmonella enterica* serotype *Typhimurium*. *N. Engl. J. Med.* 2001;344:189–195.
79. Nataro JP, and J.B. Kaper. Diarrheagenic *Escherichia coli*. *Clinical Microbiol. Rev.* 1998;11:142–201.
80. Russo TA, and J.R. Johnson. Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: ExPEC. *J. Infect. Dis.* 2000;181:1753–1754.
81. Manges AR, J.R. Johnson, B. Foxman, T.T. OBryan, K.R. Fullerton, and L.W. Riley. Widespread distribution of urinary tract infections caused by a multidrug resistant *Escherichia coli* clonal group. *New Eng. J. Med.* 2001;345:1007–1013.
82. Prats G, F. Navarro, B. Mirelis, D. Dalmau, N. Margall, P. Coll, A. Stell, and J.R. Johnson. *Escherichia coli* serotype O15:K52:H1 as an uropathogenic clone. *J. Clin. Microbiol.* 2000;38:201–209.
83. Dalmau D, F. Navarro, B. Mirelis, J. Blanco, J. Garau, and G. Prats. *Escherichia coli* bacteraemia: serotype O15:K52:H1 as a urinary pathogen. *J. Hosp. Infect.* 1996;34:233–234.
84. Olesen B, H.J. Kolmos, F. Orskov, and I. Orskov. Cluster of multiresistant *Escherichia coli* O78:H10 in Greater Copenhagen. *Scand. J. Infect. Dis.* 1994;26:406–410.
85. Ramchandani M, A.R. Manges, C. DebRoy, S.P. Smith, J.R. Johnson, L.W. Riley. Possible animal origin of human-associated, multidrug-resistant, uropathogenic *Escherichia coli*. *Clin. Infect. Dis.* 2005;40:251–257.
86. Johnson JR, M.A. Kuskowski, K. Smith, T.T. O'Bryan, S. Tatini. Antimicrobial-resistant and extraintestinal pathogenic *Escherichia coli* in retail foods. *J. Infect. Dis.* 2005;191:1040–1049.
87. Johnson JR, P. Delavari, T.T. O'Bryan, K.E. Smith, S. Tatini. Contamination of retail foods, particularly turkey, from community markets (Minnesota, 1999–2000) with antimicrobial-resistant and extraintestinal pathogenic *Escherichia coli*. *Foodborne Pathog. Dis.* 2005;2:38–49.
88. Hacek DM, T. Suriano, G.A. Noskin, J. Kruszynski, B. Reisberg, and L.R. Peterson. Medical and economic benefit of a comprehensive infection control program that includes routine determination of microbial clonality. *Am. J. Clin. Pathol.* 1999;111:647–654.
89. Mato R, H. de Lencastre, R.B. Roberts, and A. Tomasz. Multiplicity of genetic backgrounds among vancomycin-resistant *Enterococcus faecium* isolates recovered from an outbreak in a New York City hospital. *Microbial. Drug Res.* 1996;2:309–317.
90. D'Agata EMC, L. Hanjing, C. Gouldin, and Y.W. Tang. Clinical and molecular characterization of vancomycin-resistant *Enterococcus faecium* strains during establishment of endemicity. *Clin. Infect. Dis.* 2001;33:511–516.
91. Harrison LH, and D.E. Griffin *Infectious Diseases*. p. 301–339. *In: Molecular Epidemiology: Principles and Practices*, R.E. Hurst (ed). New York: Academic Press, Inc., 1993.
92. Friedman CR, G.C. Quinn, B.N. Kreiswirth, D.C. Perlman, N. Salomon, N. Schluger, M. Lutfey, J. Berger, N. Poltoratskaia, and L.W. Riley. Widespread dissemination of a drug-susceptible strain of *Mycobacterium tuberculosis*. *J. Infect. Dis.* 1997;176:478–484.
93. Ehrh S, M.U. Shiloh, J. Ruan, M. Choi, S. Gunzburg, C. Nathan, Q.W. Xie, and L.W. Riley. A novel antioxidant gene from *M. tuberculosis*. *J. Exp. Med.* 1997;186:1885–1896.
94. Ruan J, G. St. John, S. Ehrh, L.W. Riley, and C. Nathan. *noxR3*, a novel gene from *Mycobacterium tuberculosis*, protects *Salmonella Typhimurium* from nitrosative and oxidative stress. *Infect. Immun.* 1999;67:3276–3283.
95. Master SS, B. Springer, P. Sander, E.C. Boettger, V. Deretic, and G.S. Timmins. Oxidative stress response genes in *Mycobacterium tuberculosis*: role of *ahpC* in resistance to peroxyntirite and stage-specific survival in macrophages. *Microbiology.* 2002;148:3139–3144.
96. Guard-Petter J, DJH, M.M. Rahaman, and R.W. Carlson. On-farm monitoring of mouse-invasive *Salmonella enterica* Serovar *Enteritidis* and a model for its association with the production of contaminated eggs. *Appl. Env. Microbiol.* 1997;63:1588–1593.
97. Gast RD, and C.W. Beard. Production of *Salmonella Enteritidis*-contaminated eggs by experimentally infected hens. *Avian Dis.* 1990;34:438–446.
98. Lu S, P.B. Killoran, and L.W. Riley. Association of *Salmonella enterica* serovar *Enteritidis* YafD with chicken egg albumen resistance. *Infect. Immun.* 2003;71:6734–6741.

## 7. Suggested Reading

### 7.1. Reviews of Strain Typing Methods

- Lipuma, J.J. Molecular tools for epidemiologic study of infectious diseases. *Pediatr. Infect. Dis. J.* 1998;17:667–675.
- Maslow, J.N., and M.E. Mulligan. Epidemiologic typing systems. *Infect. Cont. Hosp. Epidemiol.* 1996;17:595–604.
- Thompson, R.C.A., C.C. Constantine, and U.M. Morgan. Overview and significance of molecular methods: what role for molecular epidemiology? *Parasitology* 1998;117:S161–S175.
- Tompkins, L.S. The use of molecular methods in infectious diseases. *N. Engl. J. Med.* 1992;327:1290–1297.
- van Belkum, A., M. Struelens, A. De Visser, H. Verbrugh, and M. Tibayrenc. Role of genomic typing in taxonomy, evolutionary genetics, and microbial epidemiology. *Clin. Microbiol. Rev.* 2001;14:547–560.
- Vnanechoutte, M. DNA fingerprinting techniques for microorganisms. A proposal for classification and nomenclature. *Molec. Biotechnol.* 1996;6:115–142.

## 7.2. PCR Techniques

- Bartlett, J. and D. Stirling. PCR Protocols 2nd ed. Methods in Molecular Biology, 226. Totowa, New Jersey: Humana Press, 2003.
- Dieffenbach, C.W. and D.S. Dveksler. PCR Primer: A laboratory manual. New York: Cold Spring Harbor Laboratory Press, 1995.
- Harris, E. A low-cost approach to PCR. New York: Oxford University Press, 1998.
- White, B.A. PCR Protocols: Current Methods and Applications. Methods in Molecular Biology, 15. Totowa, New Jersey: Humana Press, 1993.