American Journal of
EPIDEMIOLOGY

Volume 153
Number 12
June 15, 2001

Sponsored by the Society for Epidemiologic Research
Published by Oxford University Press

COMMENTARY

Molecular Epidemiology: Focus on Infection

Betsy Foxman[1] and Lee Riley[2]

Molecular biology techniques have become increasingly integrated into the practice of infectious disease epidemiology. The term "molecular epidemiology" routinely appears in the titles of articles that use molecular strain-typing ("fingerprinting") techniques—regardless of whether there is any epidemiologic application. What distinguishes molecular epidemiology is both the "molecular," the use of the techniques of molecular biology, and the "epidemiology," the study of the distribution and determinants of disease occurrence in human populations. The authors review various definitions of molecular epidemiology. They then comment on the range of molecular techniques available and present some examples of the benefits and challenges of applying these techniques to infectious agents and their affected host using tuberculosis and urinary tract infection as examples. They close with some thoughts about training future epidemiologists to best take advantage of the new opportunities that arise from integrating epidemiologic methods with modern molecular biology. *Am J Epidemiol* 2001;153: 1135–41.

communicable diseases; epidemiology, molecular; tuberculosis; urinary tract infections

Over the past two decades, there has been a proliferation of subspecialties among epidemiologists. Perhaps none of these subspecialties has been received with more controversy than "molecular epidemiology," as the term "molecular" describes neither a disease category nor a substantive area (1) but in jargonese refers to characteristics based on nucleic acid- or amino acid-based content. The issue is further confused by the independent emergence of the term molecular epidemiology during the 1970s and early 1980s in three separate substantive areas: cancer epidemiology, environmental epidemiology, and infectious disease epidemiology. In many epidemiologic textbooks, molecular epidemiology has been defined almost exclusively in terms of biomarkers (2), ignoring the many applications in both genetic and infectious disease epidemiology.

Received for publication June 21, 2000, and accepted for publication October 5, 2000.

Abbreviation: IS*6110*, insertion sequence 6110.

[1] Department of Epidemiology and Center for Molecular and Clinical Epidemiology of Infectious Diseases, School of Public Health, University of Michigan, Ann Arbor, MI.

[2] Divisions of Infectious Diseases and Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, CA.

Reprint requests to Dr. Betsy Foxman, Department of Epidemiology, School of Public Health, University of Michigan, 109 Observatory Street, Ann Arbor, MI 48109-2029 (bfoxman@umich.edu).

WHAT EXACTLY IS MOLECULAR EPIDEMIOLOGY?

Many different definitions of molecular epidemiology have been published (table 1); all mention the use of molecular tools, but not all explicitly mention epidemiology. This is unfortunate, as molecular epidemiology is not just molecular taxonomy, phylogeny, or population genetics but the application of these techniques to epidemiologic problems.

Molecular taxonomy, phylogeny, population genetics, and molecular epidemiology may use the same laboratory techniques, but each follows distinct principles. In phylogeny/taxonomy, the data are generated to describe properties and characteristics of organisms. Population genetics often intersects with epidemiology: both use population approaches to describe the distribution of characteristics of interest and analyze data to identify the determinants of that distribution. Epidemiology attempts to identify factors that determine disease distribution in time and place, as well as factors that determine disease transmission, manifestation, and progression. Further, epidemiology is always motivated by an opportunity or possibility for intervention and prevention.

What distinguishes molecular epidemiology is both the "molecular," the use of the techniques of molecular biology to characterize nucleic acid- or amino acid-based content, and the "epidemiology," the study of the distribution and determinants of disease occurrence in human populations.

TABLE 1. A snapshot of various definitions of molecular epidemiology

Author(s) and ref. no.	Reference	Definition
Higginson J (37)	Am J Pathol 1977;86:460–84	"the application of sophisticated techniques to the epidemiologic study of biological material" (p. 463)
Schulte PA (2)	In: Schulte PA, Perera FP, eds. San Diego, CA: Academic Press, 1993:3–44	"molecular epidemiology is the use of biologic markers or biologic measurements in epidemiologic research" (p. 13)
Tompkins LS (38)	In: Miller VL, Kaper JB, Portnoy DA, et al, eds. Washington, DC: American Society for Microbiology, 1994:63–73	"the application of molecular biology to the study of infectious disease epidemiology" (p. 65)
McMichael AJ (39)	Am J Epidemiol 1994;140:1–11	"using molecular biomarkers in epidemiology" (p. 5)
Groopman JD, Kensler TW, Links JM (40)	Toxicol Lett 1995;82-83:763–9	"molecular epidemiologic research involves the identification of relations between previous exposure to some putative causative agent and subsequent biological effects in a cluster of individuals in populations" (p. 763)
Hall A (41)	Trop Med Int Health 1996;1:407–8	"the analysis of nucleic acids and proteins in the study of health and disease determinants in human populations" (p. 407)
Shpilberg O, Dorman JS, Ferrell RE, et al. (42)	J Clin Epidemiol 1997;50:633–8	"molecular epidemiology uses molecular techniques to define disease and its pre-clinical states, to quantify exposure and its early biological effect, and to identify the presence of susceptibility genes" (p. 633)
Levin BR, Lipsitch M, Bonhoeffer S (43)	Science 1999;283:806–9	"the practical goals of molecular epidemiology are to identify the microparasites responsible for infectious diseases and determine their physical sources, their biological relationships, and their route of transmission and those of the genes responsible for their virulence, vaccine-relevant antigens and drug resistance" (p. 806)

Molecular techniques may be applied to the measurement of host or agent factors and of exposures. When applied to studies of disease, the resulting enhanced measurement increases our ability to more reliably detect associations. Molecular techniques help to stratify and to refine data by providing more sensitive and specific measurements, which facilitate epidemiologic activities, including disease surveillance, outbreak investigations, identifying transmission patterns and risk factors among apparently disparate cases, characterizing host-pathogen interactions, detecting uncultivatable organisms, providing clues for possible infectious causes of cancer and other chronic diseases, and providing better understanding of disease pathogenesis at the molecular level.

MOLECULAR TECHNIQUES

Molecular techniques do not substitute for conventional methods. They address epidemiologic problems that cannot be approached or would be more labor intensive, expensive, and/or time consuming to address by conventional techniques. Today's molecular technique can become tomorrow's conventional diagnostic tool or even consigned to the wastebasket. For example, plasmid profile analysis was a mainstay of molecular fingerprinting just a short while ago and now has been almost entirely replaced by other techniques.

Acknowledging that any list of molecular techniques will be outdated from the time it is published, we present in table 2 techniques that have been applied in epidemiologic studies of infectious disease. They fall into two large categories: identification and fingerprinting (strain typing). Rather than describe the techniques themselves in detail, we describe how the application of some of these techniques has increased our understanding of the epidemiology of two important infectious agents: *Mycobacterium tuberculosis*, which causes tuberculosis, and uropathogenic *Escherichia coli*, which causes urinary tract infection. Tuberculosis is the most common infectious cause of deaths in adults worldwide (3), and urinary tract infection is one of the most common bacterial infections, affecting half of all women (4) and one seventh of all men at least once during their lifetime (5). We will use these pathogens to illustrate the distinct approaches and principles that must be considered when conducting epidemiologic investigations using molecular techniques.

TABLE 2. Applications of molecular techniques in epidemiologic studies and available techniques as of this writing

Applications	Method	Technique
Identification	Conventional	Culture Enzyme-linked immunosorbent assay (ELISA) Enzyme immunosorbent assay (EIA) Monoclonal antibodies
	Nucleic acid based	DNA hybridization for known genes Direct sequencing of one or more regions Multilocus sequence typing (MLST)
	PCR* based	Amplification of a single target specific to a pathogen Ligase chain reaction (LCR)
	Protein based	Western blot or immunoblotting
Fingerprinting	Conventional	Serotype Antibiotic susceptibilities
	Nucleic acid based	Plasmid profiles Restriction fragment length polymorphism (RFLP) Pulsed field gel electrophoresis (PFGE) Segmented RNA gel electrophoresis Ribosomal RNA gel electrophoresis Direct sequencing of one or more regions Multilocus sequence typing (MLST)
	PCR based	Amplification of a single target specific to a pathogen Targeting known repetitive sequences (enterobacterial repetitive intergenic consensus sequences (ERIC), repetitive extragenic palindromic sequences (REP), double repetitive element (DRE), BOX, insertional sequence (IS), polymorphic guanine/cytosine-rich repetitive sequences (PGRS)) Random primers (randomly amplified polymorphic DNA (RAPD), arbitrary primed PCR (AP-PCR)) Restriction endonuclease of a single amplified product Amplified fragment length polymorphism (AFLP)
	Protein based	Multilocus enzyme electrophoresis (MLEE)
	Gene expression	Reverse transcriptase PCR Microarray technologies

* PCR, polymerase chain reaction.

MOLECULAR EPIDEMIOLOGY OF TUBERCULOSIS

Tuberculosis is one infectious disease in which molecular biology techniques have yielded novel information that would have been difficult if not impossible to obtain by conventional laboratory methods. Several molecular techniques are currently used to subtype *M. tuberculosis*, the etiologic agent of tuberculosis. The method that has come to be accepted as standard is based on a repetitive DNA element called insertion sequence 6110 (IS*6110*) (6). The strains are typed according to electrophoretic banding patterns generated by strain differences in location in the chromosome and copy numbers of this repetitive DNA element. In strains with fewer than six copies of IS*6110*, one or more of several secondary typing methods are used. The discriminatory power and reproducibility of these secondary typing methods have been compared and recently reported (7).

The IS*6110*-based strain-typing method has been used for a variety of tuberculosis epidemiologic investigations, including the following: 1) confirmation of an outbreak in institutional settings, 2) identifying an outbreak in what appears to be sporadic cases of tuberculosis, 3) identifying risk factors for recent infections or rapidly progressive disease, 4) tracking geographic spread of *M. tuberculosis* clones of public health importance, and 5) evaluating laboratory cross-contamination with *M. tuberculosis*.

Confirmation of an outbreak in institutional settings

One of the first epidemiologic applications of the IS*6110*-based typing method was performed in a study of a tuberculosis outbreak in a housing facility for human immunodeficiency virus-infected persons in San Francisco, California (8). In this situation, it was clear that there was an outbreak of tuberculosis even before the *M. tuberculosis* isolates were typed. The importance of this study was that the IS*6110*-based typing method itself was validated in a recognized outbreak setting. All patients suspected to be part of an institutional outbreak were infected with strains that had similar

IS*6110* patterns, while those unrelated to this cluster of tuberculosis cases did not have these patterns. Therefore, the relatedness of the *M. tuberculosis* isolates in an outbreak setting was not determined by the similarity of the electrophoretic banding patterns generated by the IS*6110* typing method but by the knowledge that the patients with tuberculosis were part of a recognized outbreak. Outbreaks provide one of the best opportunities to validate a new molecular strain-typing method for epidemiologic application. Once validated, the technique can then be applied to other epidemiologic investigations, as outlined below.

Identifying an outbreak in what appears to be sporadic cases of tuberculosis

Tuberculosis in most communities occurs in a typical endemic pattern, without obvious clustering in time or place. In New York City in the early 1990s, there were several institutional outbreaks of multidrug-resistant tuberculosis that were traced to a single strain with a characteristic IS*6110* banding pattern designated the "W" strain (9). This multidrug-resistant strain had an atypical antimicrobial resistance pattern, one of which was resistance to kanamycin. Hence, resistance to kanamycin served as a relatively easy identifiable and tractable marker for multidrug resistance for this particular strain of *M. tuberculosis* in New York City. However, the relation of drug-susceptible tuberculosis cases is more difficult to establish epidemiologically. Analysis of *M. tuberculosis* isolates by the IS*6110* typing method in a number of studies helped to identify cases of tuberculosis belonging to clusters that were not found using conventional contact-tracing methods (10–14).

Identifying risk factors for recent infection or rapidly progressive disease

Tuberculosis results from either reactivation of an infection acquired in the remote past or from rapid progression from an infection acquired recently. It has come to be generally accepted that cases of tuberculosis caused by *M. tuberculosis* strains with identical IS*6110* banding patterns isolated from two or more persons (cluster patterns) represent recent exogenous infection, while those infected with strains with patterns that are not observed among any other clinical isolate (unique patterns) from that community are considered to represent endogenous reactivation disease (10, 12, 15). It should be cautioned, however, that the interpretation of IS*6110* cluster patterns as representing cases of tuberculosis arising from recent infection may not always be appropriate. In highly stable populations, this assumption may not be valid, as was shown in a study in Arkansas, a state with largely rural populations (16). A large proportion of newly diagnosed tuberculosis cases were found to have isolates with similar IS*6110* patterns, and many of these cases were not epidemiologically linked. Therefore, additional factors must be taken into consideration to make the assumption that cluster patterns represent recent infections. These include information such as the mobility of the study population, time of arrival and duration of residence of those who develop tuberculosis in the geographic place of study, differences in the mean age of cases of tuberculosis infected with cluster pattern strains versus those infected with unique pattern strains, and other epidemiologically plausible characteristics supportive of this assumption. Again, it is the epidemiologic information that validates the technique, as well as the interpretation based on the information provided by the technique.

The ability to differentiate the proportion of tuberculosis cases in a community due to recent infection versus reactivation has major implications in the assessment of a tuberculosis control program in that community. New cases of tuberculosis that arise from recent infections reflect rates of current active transmissions in that community: the higher the incidence, the poorer the tuberculosis control program. Hence, obtaining information about the incidence of tuberculosis due to recent infection, regardless of whether this occurs in human immunodeficiency virus-infected or uninfected populations, is an important component of tuberculosis control efforts. The IS*6110*-based strain-typing method helps to obtain such information.

Once this type of information is obtained, conventional analytical epidemiologic methods are used to identify risk factors of tuberculosis due to recent infection. Such studies are usually performed using a case-control design, where cases are defined as those who develop tuberculosis from recent infection and controls are those patients who develop reactivation tuberculosis. The ability to further stratify tuberculosis patients by their *M. tuberculosis* isolates based on the IS*6110* patterns provides an opportunity to refine case-control analyses. The conventional laboratory methods are often not discriminating enough to provide the data stratification needed to conduct such an analysis. Case-control studies made possible by the IS*6110* typing analysis of *M. tuberculosis* isolates have identified a variety of risk factors associated with recent infection. Many of these factors are shared among tuberculosis patients in different communities (young age, ethnic minorities, homelessness, acquired immunodeficiency syndrome), but some are specific to a particular community (10, 12, 14, 15, 17–19) .

Tracking geographic spread of *M. tuberculosis* clones of public health importance

The comparison of the standardized IS*6110* pattern databases from different geographic areas makes it possible to track *M. tuberculosis* strains considered to be of major public health importance. It also helps to evaluate the ability of certain strains of *M. tuberculosis* to spread in a community. For example, a search of databases was performed to track the spread of the highly multidrug-resistant strain of *M. tuberculosis* from New York City designated as the "W" strain (9). This strain has been identified from tuberculosis patients in Florida, Nevada, Georgia, Colorado, and France. The patients in Denver and France were previous residents of New York City. In Denver, possible secondary transmissions were documented, where the index case spread the infection to at least two household contacts and several health care facility workers. The so-called "W" strains in New York City belong to a clade of *M. tuberculosis* called

the Beijing family (20). Members of this clade have been recently shown to be responsible for large epidemics of multidrug-resistant tuberculosis in Russia.

Evaluating laboratory cross-contamination with *M. tuberculosis*

One of the common problems that hospital epidemiologists must deal with is to assess whether a cluster of infections that arises in a nosocomial setting represents a true outbreak or a pseudo outbreak due to laboratory contamination. This is an epidemiologic issue that is not easily addressed by conventional epidemiologic methods. The application of the IS*6110*-based analysis has proven to be quite useful in confirming cross-contamination in clinical mycobacteriology laboratories. In two reports, investigations of clusters of *M. tuberculosis* cultures associated with patients without clinical suspicion of tuberculosis led to a conclusion that laboratory contamination had occurred (21, 22).

Summary

The molecular epidemiologic approach to studying tuberculosis epidemiology has identified several new observations that could not have been obtained by conventional epidemiologic or laboratory approaches. We now know that, in places such as the United States, new cases of tuberculosis resulting from recent infections are more common than previously believed, and that in persons with acquired immunodeficiency syndrome, greater than 50 percent of them develop the disease from new infections. The ability to distinguish tuberculosis due to recent infection from reactivation tuberculosis helps to assess the current rates of active transmission of tuberculosis in a community and may guide appropriate tuberculosis control efforts in such a community. The application of molecular epidemiologic methods to the study of tuberculosis can be generalized to other infectious diseases that clearly occur as outbreaks.

MOLECULAR EPIDEMIOLOGY OF *E. COLI* URINARY TRACT INFECTION

Urinary tract infection is a common, frequently recurring condition that affects almost 11 million women annually (4) and is a major cause of nosocomial infection (23). In contrast with tuberculosis, urinary tract infection is not usually thought of as a disease that occurs in outbreaks. We are aware of only two reports of *E. coli* urinary tract infection outbreaks outside hospital settings (24, 25). Both outbreaks were detected because of the unique phenotype of the agent combined with a particularly severe clinical presentation. Probably many such outbreaks occur, but they are not easily detected. Detection is difficult, because the background rate of urinary tract infection is so high, and because many different organisms can cause urinary tract infection. Further, the bacteria that most commonly cause urinary tract infection, *E. coli*, a usual bowel inhabitant, are extremely heterogeneous. For multiagent syndromes such as urinary tract infection or diseases caused by heterogeneous species like *E. coli*, molecular methods are essential for understanding the epidemiology.

The types of questions that are addressed by molecular techniques for diseases like urinary tract infection are different from those for tuberculosis. Questions pertinent for urinary tract infection include the following. 1) Can any *E. coli* cause urinary tract infection (i.e., is urinary tract infection a result of fecal contamination)? 2) In the absence of obvious outbreaks, how are uropathogenic *E. coli* transmitted between persons? 3) Does the great variety of uropathogenic *E. coli* reflect multiple pathogenic mechanisms, the horizontal gene transfer into different strains of *E. coli*, or both?

Can any *E. coli* cause urinary tract infection?

E. coli (a normal bowel inhabitant) cause a number of human diseases and are the most common cause of urinary tract infection (26). Although a single species when classified by biochemical tests, *E. coli* are quite heterogeneous. The genome ranges from 4.5 to 5.5 megabases (27, 28), with sequence homology as low as 70 percent. In comparison, human and chimpanzee genomes have 98 percent sequence homology. *E. coli* cause the vast majority of urinary tract infections. Given the heterogeneity of *E. coli*, and that *E. coli* are a normal bowel inhabitant, it is reasonable to wonder if any bacteria can cause urinary tract infection in an otherwise healthy individual. This question cannot be answered without molecular techniques.

To address this question, fecal *E. coli* from normal, healthy individuals have been compared with *E. coli* isolated from persons with kidney and bladder infections, by a variety of different molecular methods (29, 30). All comparisons showed that uropathogenic *E. coli* share certain characteristics and that these characteristics differentiate them from fecal *E. coli*, implying that urinary tract infection in normal, healthy individuals is not merely the result of fecal contamination.

Transmission of uropathogenic *E. coli*

Because *E. coli* are found in the bowel flora, it has been presumed that uropathogenic *E. coli* are transmitted by the fecal-oral route. Epidemiologic evidence suggests that, at least in sexually active young women, both the frequency of vaginal intercourse and the duration of sexual partnership are important predictors of urinary tract infection acquisition. While this does not preclude fecal-oral transmission, it does suggest that sexual transmission may be possible. This observation led us to search for urethral colonization with *E. coli* in the male sex partners of women with urinary tract infection (31). Urinary, vaginal, and fecal *E. coli* isolates from 19 women with urinary tract infection were compared with *E. coli* found in random initial voids from their most recent male sex partner. *E. coli* were isolated from four of 19 male sex partners. In each case, the *E. coli* isolated from the male were identical by pulsed-field gel electrophoresis and bacterial virulence profile to the urinary *E. coli* from his sex partner, suggesting that sexual activity may lead to transmission. Without

molecular techniques, in this case pulsed-field gel electrophoresis, we would have been unable to determine whether the *E. coli* isolated from males were identical to those causing urinary tract infection in their sex partner.

Diversity of uropathogenic *E. coli*

The genomes of bacteria can be quite plastic. In addition to adding and deleting genes on the chromosome, bacteria can change phenotype through the gain or loss of plasmids. Antibiotic resistance, for example, can be transferred via plasmid and has been demonstrated to be transmitted across bacterial species. It may be that at least some uropathogenic characteristics are acquired in this fashion. In this scenario there could be heterogeneity in the organism structure despite using a common method of pathogenesis. An outbreak would follow the path of the plasmid (horizontal and vertical gene transfer) rather than the path of a particular bacterial clone. The identification of uropathogenic factors and their mode of transmission between pathogens would greatly assist our understanding of urinary tract infection epidemiology and pathogenesis and our ability to prevent disease via vaccination or other strategies. These types of studies require epidemiologic methods to collect appropriate sample isolates from well-defined populations and to make appropriate inferences about the findings based on laboratory analyses.

As an example, we studied a collection of first-episode urinary tract infection isolates from epidemiologically well-characterized women and grouped them by the presence or absence of nine putative virulence genes. Although the numbers were small, we detected a time-space cluster in one such grouping. Pulsed-field gel electrophoresis analysis showed an apparent clonal group. Only a small proportion of fecal *E coli* with the same combination of virulence genes had the same pulsed-field gel electrophoresis pattern. We reasoned that, for genomic subtraction, choosing one isolate from the apparent cluster and the second isolate from the fecal isolates would increase our possibility of detecting urinary tract infection virulence factors (32). This experiment resulted in the identification of 37 DNA sequences physically located all over the genome of the chosen urinary tract infection isolate.

Summary

The application of molecular techniques to the study of heterogeneous organisms enhances epidemiologic studies by improving our ability to subclassify these organisms into meaningful groups. This facilitates the detection of disease outbreaks that may otherwise be undetected and allows the epidemiologist to identify risk factors of outbreaks, sporadic cases, or both. In the case of tuberculosis, epidemiologic information helped to validate the molecular techniques, which were, in turn, applied to further characterize the epidemiology of tuberculosis. In the case of *E. coli*, both epidemiologic information and molecular laboratory information have to be analyzed simultaneously to characterize the epidemiology of urinary tract infection. With *E. coli*, collections resulting from population-based epidemiologic studies can assist the molecular biologist in identifying groups for studying pathogenesis and in making inferences about the potential role of newly identified genes. These results, in turn, can be used to further characterize the epidemiology.

A LOOK TO THE FUTURE

We touched here upon the applications of molecular techniques to the study of the epidemiology of infectious agents, focusing on the pathogens themselves. We did not touch on the other side of the host-pathogen relation, the role of the host in disease susceptibility and resistance. Recent studies have revealed an association of a number of candidate genes with tuberculosis, although such genes comprise a small attributable fraction of all those who develop tuberculosis (33). There is also some suggestion that chronic recurring urinary tract infection (four or more episodes in a 12-month period) may have a genetic component: women who are nonsecretors of blood group antigens are more likely to have recurring urinary tract infections (34), and they are also more susceptible than secretors to colonization with *E. coli* with uropathogenic potential (35). Studies of host genetic susceptibility to infection are indeed part of the molecular epidemiology discipline and are reviewed in more detail elsewhere (36).

There is a clear need to train individuals who are capable of developing new theories and methods to address the questions that will arise from the interface of molecular biology and epidemiology. Most of the molecular techniques listed in table 2 can be mastered in a short period of time. The more time-consuming part of this type of training is the epidemiology. By definition, molecular epidemiology training requires practical application of both the laboratory and epidemiologic techniques to address a real-world infectious disease problem. The molecular epidemiologist will need to interface with clinicians, statisticians, epidemiologists, molecular biologists, computer scientists, engineers, and practitioners in the new field of bioinformatics and computational biology. Although we will never conquer infectious diseases, we can certainly learn to live in greater harmony with them. This may perhaps be our most effective intervention strategy. Discovering how to do so will be the great challenge for molecular epidemiologists of the future.

REFERENCES

1. Ambrose CB, Kadlubar FF. Toward an integrated approach to molecular epidemiology. Am J Epidemiol 1997;146:912–18.
2. Schulte PA, Perera FP, eds. Molecular epidemiology: principles and practices. San Diego, CA: Academic Press, 1993.
3. Raviglione MC, Snider DE, Kochi A. Global epidemiology of tuberculosis. JAMA 1995;273:220–6.
4. Foxman B, Barlow R, d'Arcy H, et al. Urinary tract infection: self-reported incidence and associated costs. Ann Epidemol 2000;10:509–15.
5. Zielske JV, Lohr KN, Brook RH, et al. Conceptualization and measurement of physiologic health for adults. Vol 16. Urinary

tract infection. Arlington, VA: The Rand Corporation, 1981. (Document no. R2262/16-HHS).
6. van Embden JD, Cave MD, Crawford JT, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
7. Kremer K, van Sooligen D, Frothingham R, et al. Comparison of methods based on different molecular epidemiological markers for typing of *M. tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol 1999;37:2607–18.
8. Daley CL, Small PM, Schecter GF, et al. An outbreak of tuberculosis with accelerated progression among persons infected with the human immunodeficiency virus. N Engl J Med 1992; 326:231–5.
9. Bifani PJ, Plikaytis BB, Kapur V, et al. Origin and interstate spread of a New York City multidrug-resistant *Mycobacterium tuberculosis* clone family. JAMA 1996;275:452–7.
10. Small PM, Hopewell PC, Singh SP, et al. The epidemiology of tuberculosis in San Francisco: a population-based study using conventional and molecular methods. N Engl J Med 1994;330: 1703–9.
11. Tabet SR, Goldbaum GM, Hooton TM, et al. Restriction fragment length polymorphism analysis detecting a community-based tuberculosis outbreak among persons infected with human immunodeficiency virus. J Infect Dis 1994;169:189–92.
12. Friedman CR, Stoeckle MY, Kreiswirth BN, et al. Transmission of multidrug-resistant tuberculosis in a large urban setting. Am J Respir Crit Care Med 1995;152:355–9.
13. Barnes PF, Yang Z, Preston-Martin S, et al. Patterns of tuberculosis transmission in Central Los Angeles. JAMA 1997;278: 1159–63.
14. Van Deutekom H, Gerritsen JJ, van Soolingen D, et al. A molecular epidemiological approach to studying the transmision of tuberculosis in Amsterdam. Clin Infect Dis 1997;25:1071–7.
15. Alland D, Kalkut GE, Moss AR, et al. Transmission of tuberculosis in New York City: an analysis by DNA fingerprinting and conventional epidemiologic methods. N Engl J Med 1994; 330:1710–16.
16. Braden CR, Templeton GL, Cave DM, et al. Interpretation of restriction fragment length polymorphism analysis of *Mycobacterium tuberculosis* isolates from a state with a large rural population. J Infect Dis 1997;175:1446–52.
17. Friedman CR, Quinn GC, Kreiswirth BN, et al. Widespread dissemination of a drug-susceptible strain of *Mycobacterium tuberculosis*. J Infect Dis 1997;176:478–84.
18. Tornieporth NG, Ptachewich Y, Poltoratskaia N, et al. Tuberculosis among foreign-born persons in New York City, 1992–94: implications for tuberculosis control. Int J Tuberc Lung Dis 1997;1:528–35.
19. Bauer J, Yang Z, Poulsen S, et al. Results from 5 years of nationwide DNA fingerprinting of *Mycobacterium tuberculosis* complex isolates in a country with a low incidence of *M. tuberculosis* infection. J Clin Microbiol 1998;36:305–8.
20. Van Soolingen D, Qian L, De Haas PE, et al. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J Clin Microbiol 1995;33:3234–8.
21. Small PM, McClenny NB, Singh SP, et al. Molecular strain typing of *Mycobacterium tuberculosis* to confirm cross-contamination in the mycobacteriology laboratory and modification of procedures to minimize occurrence of false-positive cultures. J Clin Microbiol 1993;31:1677–82.
22. Dunlap NE, Harris RH, Benjamin WH, et al. Laboratory contamination with *Mycobacterium tuberculosis* cultures. Am J Crit Care Med 1995;152:1702–4.
23. Sedor J, Mulholland SG. Hospital-acquired urinary tract infections associated with the indwelling catheter. Urol Clin North Am 1999;26:821–8.
24. Riley PA, Threlfall EJ, Cheasty T, et al. Occurrence of Fime plasmids in multiply antimicrobial-resistant *Escherichia coli* isolated from urinary tract infection. Epidemiol Infect 1993; 110:459–68.
25. Olesen B, Komos HJ, Orskov F, et al. Cluster of multiresistant *Escherichia coli* O78:H10 in Greater Copenhagen. Scand J Infect Dis 1994;26:406–10.
26. Kunin CM. Detection, prevention and management of urinary tract infections. 5th ed. Baltimore, MD: Williams & Wilkins, 1997.
27. Arthur M, Johnson CE, Rubin RH, et al. Distribution of chromosome length variation in natural isolates of *Escherichia coli*. Mol Biol Evol 1998;15:6–16.
28. Blattner FR, Plunkett G, Bloch CA, et al. The complete genome sequence of *Escherichia coli* K-12. Science 1997;277: 1453–74.
29. Johnson JR. Virulence factors in *Escherichia coli* urinary tract infection. Clin Microbiol Rev 1991;4:80–128.
30. Caugnant DA, Levin BR, Lidin-Janson G, et al. Genetic diversity and relationships among strains of *Escherichia coli* in the intestine and those causing urinary tract infections. Prog Allergy 1983;33:203–27.
31. Foxman B, Zhang L, Tallman P, et al. Transmission of uropathogens between sex partners. J Infect Dis 1997;175: 989–92.
32. Zhang L, Foxman B, Manning SD, et al. Molecular epidemiologic approaches to urinary tract infection gene discovery in uropathogenic *Escherichia coli*. Infect Immun 2000;68:2009–15.
33. Abel L, Casanova JL. Genetic predisposition to clinical tuberculosis: bridging the gap between simple and complex inheritance. Am J Hum Genet 2000;67:274–7.
34. Hooton TM, Roberts PL, Stamm WE. Effects of recent sexual activity and use of a diaphragm on the vaginal microflora. Clin Infect Dis 1994;19:274–8.
35. Stapleton A, Hooton TM, Fennell C, et al. Effect of secretor status on vaginal and rectal colonization with fimbriated *Escherichia coli* in women with and without recurrent urinary tract infection. J Infect Dis 1995;171:717–20.
36. Harrison LH, Griffin DE. Infectious diseases. In: Schulte PA, Perera FP, eds. Molecular epidemiology: principles and practices. San Diego, CA: Academic Press, 1993:301–39.
37. Higginson J. The role of the pathologist in environmental medicine and public health. Am J Pathol 1977;86:460–84.
38. Tompkins LS. Molecular epidemiology: development and application of molecular methods to solve infectious disease mysteries. In: Miller VL, Kaper JB, Portnoy DA, et al, eds. Molecular genetics of bacterial pathogenesis: a tribute to Stanley Falkow. Part 1. Retrospective look at early advances. Washington, DC: American Society for Microbiology, 1994: 63–73.
39. McMichael AJ. Invited commentary—"molecular epidemiology": new pathway or new travelling companion? Am J Epidemiol 1994;140:1–11.
40. Groopman JD, Kensler TW, Links JM. Molecular epidemiology and human risk monitoring. Toxicol Lett 1995;82-83:763–9.
41. Hall A. What is molecular epidemiology? (Editorial). Trop Med Int Health 1996;1:407–8.
42. Shpilberg O, Dorman JS, Ferrell RE, et al. The next stage: molecular epidemiology? J Clin Epidemiol 1997;50:633–8.
43. Levin BR, Lipsitch M, Bonhoeffer S. Population biology, evolution, and infectious disease: convergence and synthesis. Science 1999;283:806–9.