Molecular Epidemiology of Tuberculosis

Peter F. Barnes, M.D., and M. Donald Cave, Ph.D.

The ability to discern the molecular “fingerprint” (genotype) of Mycobacterium tuberculosis isolates has revolutionized our understanding of the transmission of tuberculosis. In this review, we summarize the main methods of determining the genotype and discuss the relevance of genotyping to the control and understanding of the pathogenesis of tuberculosis.

The standard approach to genotyping M. tuberculosis isolates is restriction-fragment–length polymorphism (RFLP) analysis of the distribution of the insertion sequence IS6110 in different strains, and large data bases of IS6110-based genotypes are available (Fig. 1). Isolates from patients infected with epidemiologically unrelated strains of tuberculosis have different RFLP patterns, whereas those from patients with epidemiologically linked strains generally have identical RFLP patterns. Strains with fewer than six IS6110 insertion sites have a limited degree of polymorphism, and supplementary methods of genotyping are used in these cases. IS6110-based genotyping requires subculturing the isolates for several weeks to obtain sufficient DNA.

The genome of M. tuberculosis contains many mycobacterial interspersed repeat units (MIRUs), some containing identical repeat units and others containing repeats that vary slightly in sequence and length. MIRU genotyping categorizes the number and size of the repeats in each of 12 independent MIRUs, with the use of a polymerase-chain-reaction (PCR) assay, followed by gel electrophoresis (Fig. 1). Two to eight alleles are at each of the 12 loci, yielding approximately 20 million possible combinations of alleles. The discriminatory power of MIRU genotyping is almost as great as that of IS6110-based genotyping. Unlike IS6110-based genotyping, MIRU analysis can be automated and can thus be used to evaluate large numbers of strains, yielding intrinsically digital results that can be easily catalogued on a computer data base. A Web site has been set up so that a worldwide data base of MIRU patterns can be created. MIRU genotyping is technically simpler than IS6110-based genotyping and can be applied directly to M. tuberculosis cultures without DNA purification. It may replace IS6110-based genotyping in the future, particularly if the evaluation of additional loci increases the discriminatory power.

The direct-repeat locus in M. tuberculosis contains 10 to 50 copies of a 36-bp direct repeat, which are separated from one another by spacers that have different sequences. However, the spacer sequences between any two specific direct repeats are conserved among strains. Because strains differ in terms of the presence or absence of specific spacers, the pattern of spacers in a strain can be used for genotyping (spacer oligonucleotide typing, or “spoligotyping”) (Fig. 2). Spoligotyping has two advantages over IS6110-based genotyping. First, because small amounts of DNA are required, it can be performed on clinical samples or on strains of M. tuberculosis shortly after their inoculation into liquid culture. Second, the results of spoligotyping, which are expressed as positive or negative for each spacer, can be expressed in a digital format.
Figure 1. Chromosome of Mycobacterium tuberculosis Hypothetical Strain X and Genotyping of M. bovis Bacille Calmette–Guérin (BCG), the M. tuberculosis Laboratory Strain H37Rv, and Strain X on the Basis of IS6110 Insertion Sequences and Mycobacterial Interspersed Repetitive Units (MIRUs).

The top left-hand panel shows the chromosome of hypothetical strain X, as shown by the arrows. The top right-hand panel shows the results of IS6110-based genotyping. Mycobacterial DNA is digested with the restriction enzyme PvuII. The IS6110 probe hybridizes to IS6110 DNA to the right of the PvuII site in IS6110. The size of each hybridizing fragment depends on the distance from this site to the next PvuII site in adjacent DNA (fragments a through f), as reflected by gel electrophoresis of the DNA fragments of BCG, H37Rv, and X. The horizontal lines to the right of the electrophoretic strip indicate the extent of the distribution of fragments in the gel, including PvuII fragments that contain no IS6110. The three bottom panels show the results of MIRU-based genotyping. MIRUs contain repeat units, and MIRU analysis involves the use of polymerase-chain-reaction (PCR) amplification and gel electrophoresis to categorize the number and size of repeats in 12 independent loci, each of which has a unique repeated sequence. The sizes of molecular-weight markers (M) and PCR products for the loci A, B, C, and D in BCG, H37Rv, and X are shown. The specific sizes of the various MIRUs in each strain result in a distinctive fingerprint for the strain.
ligotyping has less power to discriminate among M. tuberculosis strains than does IS6110-based genotyping.\(^7\)

**Figure 2. Spoligotyping.**

The direct-repeat (DR) locus is a chromosomal region that contains 10 to 50 copies of a 36-bp direct repeat, separated by spacer DNA with various sequences, each of which is 37 to 41 bp. A copy of IS6110 is inserted within a 36-bp direct repeat in the middle of the DR locus in most strains. *Mycobacterium tuberculosis* strains have the same overall arrangement of spacers but differ in terms of the presence or absence of specific spacers. Spacer oligonucleotide typing (spoligotyping) involves polymerase-chain-reaction (PCR) amplification of the DR locus, followed by hybridization of the labeled PCR products to a membrane that contains covalently-bound oligonucleotides corresponding to each of 43 spacers. Individual strains have positive or negative signals for each spacer. The top section shows the 43 direct repeats (rectangles) and spacers (horizontal lines) used in spoligotyping. The middle section shows the products of PCR amplification of spacers 1 through 6 of *M. bovis* bacille Calmette–Guérin (BCG), *M. tuberculosis* strain H37Rv, and *M. tuberculosis* hypothetical strain X, with the use of primers (white and black arrowheads) at each end of the DR locus. The bottom section shows the spoligotypes of the three strains.

Genotyping of isolates from patients is useful in several situations. The results can be used to confirm the occurrence of cross-contamination in the laboratory. Approximately 3 percent of patients from whom *M. tuberculosis* is apparently isolated in clinical laboratories do not have tuberculosis; the positive cultures are due to cross-contamination. The occurrence of cross-contamination is most likely when acid-fast smears are negative and only one specimen is culture-positive.\(^8\) When *M. tuberculosis* is isolated from a specimen but the clinical
findings do not suggest the presence of tuberculosis, genotyping of the isolate and other M. tuberculosis strains handled concurrently in the laboratory can strongly suggest the occurrence of cross-contamination and lead to the discontinuation of antituberculosis medications.

Genotyping permits the evaluation of isolates with different patterns of drug susceptibility. Such an evaluation may be helpful in cases in which the original organism developed drug resistance during or after antituberculosis therapy, the patient was reinfeasured with a different M. tuberculosis strain, or cross-contamination is suspected. Genotyping the isolates from the patient and other isolates processed at the same time can distinguish among these possibilities. If the original organism developed resistance, the cause could be nonadherence to therapy or reduced concentrations of antituberculosis drugs as a result of malabsorption or drug interactions. If the cause was reinfection, however, public health authorities should attempt to identify the source.

Genotyping can be used to evaluate second episodes of tuberculosis. Genotyping of isolates from both episodes will determine whether the second episode is due to relapse or reinfection. In the case of a relapse, the susceptibility of the original isolate can be used to guide treatment, and reasons for treatment failure can be evaluated. If reinfection is documented, a source should be sought. In antituberculosis-treatment trials, it is important to determine whether recurrent tuberculosis is due to relapse or reinfection, since only the former represents treatment failure.

Genotyping can also be used to evaluate an outbreak. If epidemiologic data suggest the occurrence of an outbreak of tuberculosis, genotyping of the isolates, in combination with an epidemiologic investigation, can help determine whether an outbreak has occurred or whether there is a coincidental occurrence of a large number of cases. This strategy can delineate the extent of the outbreak and guide public health measures to reduce disease transmission.

Isolates in the United States can be genotyped by regional laboratories that are funded by the Centers for Disease Control and Prevention. Specimens must be transferred from local to regional laboratories for Disease Control and Prevention. Specimens by regional laboratories that are funded by the Centers for Disease Control and Prevention. Specimens by regional laboratories that are funded by the Centers for Disease Control and Prevention. Specimens by regional laboratories that are funded by the Centers for Disease Control and Prevention. Specimens by regional laboratories that are funded by the Centers for Disease Control and Prevention. Specimens by regional laboratories that are funded by the Centers for Disease Control and Prevention. Specimens by regional laboratories that are funded by the Centers for Disease Control and Prevention. Specimens by regional laboratories that are funded by the Centers for Disease Control and Prevention. Specimens.

There are two caveats to this concept. First, genotyping data must be interpreted together with epidemiologic information, which is usually obtained by interviewing patients, preferably when tuberculosis is diagnosed. For example, in many cities, patients whose M. tuberculosis isolates have the same genotype have all been recently infected, whereas if genotyping is performed in only a minority of cases.

There is broad variability in the genotypes of M. tuberculosis isolates from patients with epidemiologically unrelated tuberculosis, whereas the genotypes of isolates from patients who were infected by a common source are identical. Therefore, clustered cases of tuberculosis, defined as those in which the isolates have identical or closely related genotypes, have usually been transmitted recently. In contrast, cases in which the isolates have distinctive genotypes generally represent a reactivation of infection acquired in the distant past. There are two caveats to this concept. First, genotyping data must be interpreted together with epidemiologic information, which is usually obtained by interviewing patients, preferably when tuberculosis is diagnosed. For example, in many cities, patients whose M. tuberculosis isolates have the same genotype have all been recently infected, whereas if genotyping is performed in only a minority of cases. Second, clustered cases will be missed if genotyping is performed over a short period. Some clustered cases will be overlooked if genotyping is performed over a short period, because tuberculosis will not yet have developed in some infected persons.

Molecular epidemiologic studies have shown that the dynamics of the transmission of tuberculosis vary greatly geographically. Where homelessness is common, shelters are often the foci of tuberculosis transmission. In other locations, health care facilities and bars have been important sites of transmission. Therefore, local efforts to identify high-risk populations and transmission sites are crucial for the effective control of tuberculosis.

Ninety percent of cases of tuberculosis in industrialized nations were once believed to result from a reactivation of infection acquired in the distant past. However, population-based genotyping indicates that recent transmission causes 20 to 50 per-
Continuing transmission results from two major factors. First, there is a substantial rate of transmission among casual contacts in workplaces and other social settings, and patients with tuberculosis whose isolates have the same genotype often have limited contact. Second, in most cases, transmission probably precedes antituberculosis therapy. Therefore, even after many years of universal, directly observed therapy and high rates of completion of treatment, studies have shown that one third of cases of tuberculosis are still due to recent transmission. Furthermore, patients with tuberculosis generally seek medical care long after symptoms develop, contributing to the spread of disease despite the existence of excellent tuberculosis-control programs.

Genotyping has been used to study the transmission dynamics of tuberculosis both in developed countries and in developing nations. However, only in the former, where more resources are available, have the results been used to guide tuberculosis-control efforts.

**Use of Genotyping by Tuberculosis-Control Programs**

Genotyping has been used to study the transmission dynamics of tuberculosis both in developed countries and in developing nations. However, only in the former, where more resources are available, have the results been used to guide tuberculosis-control efforts.

**Identification of Groups at Increased Risk for Tuberculosis**

Molecular epidemiologic studies have identified many unsuspected community outbreaks of tuberculosis, resulting in focused public health interventions. For example, in Los Angeles, homeless shelters were identified as major sites of tuberculosis transmission, and screening homeless persons identified many cases of tuberculosis and helped decrease the rate of tuberculosis. Chest radiography and the examination of sputum smears for acid-fast bacilli are excellent screening tools for highly selected groups, such as homeless or incarcerated persons, particularly when a few transmission sites contribute heavily to the disease-related morbidity in a population. Screening high-risk groups for latent tuberculosis infection is a less attractive approach because people in such groups rarely complete therapy.

Subpopulations with high rates of transmission of tuberculosis have been identified through the use of analytic methods that estimate the transmission index, defined as the number of cases of tuberculosis generated by a single source case. For example, in San Francisco, the transmission index was seven times as high among U.S.-born blacks as among U.S.-born whites or Hispanics. This difference was not due to an increased prevalence of human immunodeficiency virus infection or of acid-fast–positive sputum smears among blacks but may have been related to delays in seeking medical attention or to social factors such as overcrowding that facilitate the transmission of diseases. Regardless of the reasons, focusing public health resources on subpopulations in which transmission indexes are highest should reduce the rate of transmission.

**Improving Investigations of Contacts**

Molecular epidemiology has shown that the links between patients infected with M. tuberculosis strains of the same genotype are often not identified by routine public health investigations. Concerted efforts have been made to improve contact investigations, resulting in a three-to-fourfold increase in the median number of contacts identified per new case of tuberculosis in one city. It is imperative to ask patients with tuberculosis about their contacts outside the home and workplace and to identify locations that patients frequent. Persons at these locations can then be screened for tuberculosis infection and disease, since they may have been infected by the source patient despite having had only minimal contact.

**Evaluation of Tuberculosis Programs**

Population-based genotyping can be used to evaluate the efficacy of tuberculosis-control efforts. From 1991 to 1997, the incidence of recently transmitted cases of tuberculosis in San Francisco, as measured by the number of clustered cases in the population, decreased substantially, suggesting that the intensification of tuberculosis-control measures during this period reduced the spread of disease.

**Future Uses of Genotyping**

To date, genotyping has been used by a limited number of tuberculosis-control programs, usually in collaboration with researchers. Barriers to the widespread use of genotyping are the complexity of IS6110-based genotyping, the long period required to obtain results, and the cost. If these obstacles can be overcome, prospective, population-based, surveillance genotyping can be incorporated into rou-
tine tuberculosis-control activities to identify unsuspected outbreaks. Surveillance genotyping requires a commitment of resources; close communication between laboratories and public health workers; the availability of reproducible, relatively simple, rapid genotyping methods; and access to simple methods to compare the genotypes of large numbers of isolates. MIRU analysis is the most promising candidate for surveillance genotyping and could identify clustered cases several days after an acid-fast–positive specimen or a positive culture for M. tuberculosis is obtained. Patients with clustered cases could be interviewed soon after receiving a diagnosis, and sites of tuberculosis transmission could be identified and subjected to intensive interventions. Evaluation of social networks may be useful to identify the places and types of persons associated with the highest risk of tuberculosis transmission.29

### RELEVANCE OF GENOTYPING TO THE PATHOGENESIS OF TUBERCULOSIS

#### MICROBIAL FACTORS THAT FAVOR TRANSMISSION

The traditional view is that M. tuberculosis strains are equally virulent. However, population-based genotyping has demonstrated that a small percentage of strains cause a disproportionately large number of cases,10,11,30 implying that some strains are spread more effectively than others. A widespread strain in New York City was more resistant to reactive nitrogen intermediates than were other M. tuberculosis strains, perhaps favoring its survival in macrophages.31 The CDC1551 strain caused a large outbreak,40 and lipids extracted from this strain stimulated monocytes to produce extremely high concentrations of inflammatory cytokines,32 perhaps accounting for the unusually large size of the tuberculin skin-test responses in persons infected with this strain.20 In another study, mice infected with one clinical M. tuberculosis strain failed to mount an appropriate immune response and died more rapidly than mice infected with another strain.33 These studies suggest that the various strains of M. tuberculosis have distinct interactions with the host and may differ in their transmission potential.

The most widespread M. tuberculosis strains are those of the Beijing family, which has caused outbreaks throughout the world34-36 and constitutes the dominant family of strains in multiple locations in Asia and North America.35,37-40 The members of the Beijing family have an IS6110 insertion in the M. tuberculosis chromosomal origin of replication, a specific spoligotype pattern, and other characteristic DNA sequences.40 The Beijing family includes the W family strains, which have caused large outbreaks of multidrug-resistant and drug-susceptible tuberculosis in the northeastern United States,40 and the 210 strain, which is widespread in the southwest and south central United States.11,14,34 These strains share 16 IS6110 insertion sites,41 and IS6110 may alter the expression of adjacent genes to favor the spread of these strains in the population.42

The Beijing strains may be widely distributed because they were introduced into multiple locations before other strains were and thus have had more time to spread in these communities. However, we believe that the transmission potential of these strains is likely to be enhanced, as compared with that of other strains, because the former are more readily aerosolized, can establish infection more effectively, or progress more frequently from infection to disease. This hypothesis is supported by three findings. First, Beijing isolates were more common in young patients than in older patients in Vietnam, suggesting that these strains spread recently.39 Second, after the introduction of one Beijing isolate to an island, it spread rapidly and was the cause of 27 percent of the cases of tuberculosis within three years.36 Third, the 210 strain, a member of the Beijing family, grows more rapidly in macrophages than do other strains, suggesting that it avoids host defenses more effectively.43

#### DRUG RESISTANCE AND DISEASE TRANSMISSION

Isoniazid-resistant strains of M. tuberculosis are less virulent than drug-susceptible isolates in guinea pigs,44 but studies based on tuberculin skin testing of the contacts of patients with isoniazid-resistant tuberculosis and the contacts of those with drug-susceptible tuberculosis yielded conflicting results.45,46 Molecular epidemiologic studies in the Netherlands and in San Francisco showed that isoniazid-resistant isolates were 30 to 80 percent less likely to be clustered than drug-susceptible organisms,30,47 and multidrug-resistant isolates in Mexico and South Africa were 70 to 80 percent less likely to be clustered than drug-susceptible strains.48,49 These studies suggest that drug-resistant isolates cause fewer secondary cases than drug-susceptible organisms, perhaps because some drug-resistance genes reduce the virulence of mycobacteria.
Even if drug-resistant strains have a reduced potential for transmission, infection-control precautions should be maintained, since patients with drug-resistant tuberculosis are likely to be infectious for long periods, and since the public health consequences of drug-resistant tuberculosis are more serious than those of drug-susceptible disease. Furthermore, multidrug-resistant tuberculosis has caused large outbreaks, and the transmission potential of individual strains cannot be determined.

**THE FREQUENCY OF REINFECTION**

When immunocompetent persons become infected with *M. tuberculosis* or become symptomatic, they are believed to be resistant to infection with a second strain. Therefore, isoniazid is not recommended for immunocompetent persons with a previously positive tuberculin skin test who are exposed to a new case of tuberculosis. Similarly, recurrent tuberculosis has been treated as a reactivation of disease that was due to the original organism. These assumptions have been challenged by molecular epidemiologic studies in South Africa and Europe, which found that exogenous reinfection caused second episodes of tuberculosis in 16 to 75 percent of patients. The percentage of cases due to reinfection was greater in populations with a high incidence of tuberculosis, which have an increased risk of exposure to new infections.

These findings have important implications with respect to the nature of protective immunity against tuberculosis. The failure of natural infection to protect against reinfection may partially explain the relative ineffectiveness of vaccination with *M. bovis* bacille Calmette–Guérin and implies that to be effective, vaccination must do more than mimic natural infection. Frequent reinfection would necessitate the treatment of persons with recent exposure to tuberculosis, regardless of the results of prior tuberculosis tests. Similarly, patients with second episodes of tuberculosis should be treated with a regimen that reflects current patterns of drug resistance in the community rather than that of the patient’s original strain, particularly in places where the prevalence of tuberculosis is high.

Supported by a grant from the National Institutes of Health (AI44935), an Interagency Agreement with the Department of Veterans Affairs and the Centers for Disease Control and Prevention (FED10318), the Cain Foundation for Infectious Disease Research, and the Centers for Pulmonary and Infectious Disease Control. Dr. Barnes holds the Margaret E. Byers Cain Chair for Tuberculosis Research.

The University of Arkansas Board of Trustees holds four patents for molecular tests for diagnosing tuberculosis; Dr. Cave receives a fraction of the related licensing fees.

REFERENCES

17. Kline SE, Hedemark LL, Davies SE. Out-