

Rethinking Tuberculosis Epidemiology: the Utility of Molecular Methods

Working within a classic TB control framework while using newer molecular techniques, epidemiologists are reevaluating *M. tuberculosis*

Barun Mathema and Barry N. Kreiswirth

Although effective chemotherapy for treating disease caused by *Mycobacterium tuberculosis* infections has been available since the early 1950s, the morbidity and mortality attributed to this pathogen remain alarmingly high, especially in poor countries and in those lacking a good tuberculosis (TB) control infrastructure. Other factors, including the ongoing HIV/AIDS pandemic, complicate efforts to control TB.

Outcome differences among individuals who are infected with *M. tuberculosis* were pivotal in formulating basic hypotheses regarding the pathogenesis and transmission of this bacillus. For example, approximately one-third of those exposed to these bacilli show evidence of infection during the course of their lives, but, of those infected, only 10% ever become ill or pass the pathogen on to others. Individuals also vary in terms of disease severity, duration, and type of symptoms that they develop. It is commonly believed that these differences are due to host-environmental factors and host-pathogen interactions.

Identifying TB Cases Is Key to Controlling This Disease

A key principle underlying TB control is to identify cases and manage them efficiently. Understanding how TB is transmitted comes from descriptive epidemiology and from animal studies, particularly environmental experiments with guinea pigs. From historical studies, many

TB experts believed that once individuals were infected with *M. tuberculosis*, they were protected against a secondary *M. tuberculosis* infection. Thus, any relapse in symptoms among individuals whose TB was considered cured was thought to involve "reactive" disease from endogenous infections.

The tuberculin skin test remains the only practical means for identifying individuals recently infected with tuberculosis. However, because of the high prevalence of individuals latently infected with tuberculosis and because the Calmette-Guérin (BCG) vaccine is widely administered in developing countries, this screening test often fails due to its low specificity. In fact, where tuberculin skin testing is used routinely, such as in most low-incidence TB countries, a positive reading cannot identify the source, duration of infection, or the drug susceptibility of the strain, without having substantial

epidemiological evidence also available. A definitive diagnosis of tuberculosis depends on recovering *M. tuberculosis* by culturing methods, a procedure that is prohibitively expensive in most high-incidence countries.

Because TB is transmitted primarily as aerosolized bacilli from an infected to an uninfected individual, identifying recent cases of TB critically depends on an ability to differentiate clinical specimens. Subspeciating isolates can help to elucidate transmission pathways and also sets the stage for addressing important questions about virulence and pathogenesis. Various

Preventing new cases of TB critically depends on an ability to differentiate clinical specimens

Barun Mathema is a Research Epidemiologist and Barry N. Kreiswirth is the Director of the Tuberculosis Center, Public Health Research Institute, Newark, N.J.

methods were used to differentiate clinical isolates, with some of the earlier methods relying on phenotypic characteristics such as mycobacterial growth rates, colony morphology, susceptibility to bacterial phages, and antibiotic susceptibility profiles.

More recently, researchers identified genetic elements that can be used to discriminate among clinical isolates of *M. tuberculosis* by means of molecular methods, including PCR, Southern blot hybridization, and DNA sequencing. These high-resolution methods provide genetic evidence to address two basic epidemiological questions: are isolates from localized cases the same or different, and are strains causing disease in one geographic area related to those circulating in other regions?

IS6110 DNA Fingerprint Analysis

DNA sequence alignment and comparison of the completed *M. tuberculosis* genomes reveal remarkable sequence conservation in this species. However, numerous repetitive elements are scattered throughout these highly conserved genomes, providing a useful means for differentiating these and other clinical isolates.

Currently, the most robust and widespread genotyping technique for subspeciating *M. tuberculosis* is based on a particular insertion sequence, IS6110, which is a member of the IS3 family of transposable elements and is unique to the *M. tuberculosis* complex. Although some "hot spots" have been noted, IS6110 is more or less randomly distributed within the chromosome, and its copy number ranges from rare clones that lack an insertion to those with 26 copies. This insertion element moves by a replicative-transposition process that remains stable over time, making it useful for epidemiological investigations.

To make it relatively easy to compare results between labs, researchers have standardized IS6110 analysis, using a Southern blot hybridization procedure with a common restriction enzyme, *Pvu*II, a common DNA probe, and a common molecular weight standard. *Pvu*II cleaves IS6110 at only one asymmetric site, and the common DNA probe being used is directed against IS6110 sequences to the "right" of the *Pvu*II cleavage site. Hence, each hybridizing band is generally representative of a single IS6110 copy.

IS6110-based DNA fingerprint (or restriction fragment length polymorphism) analysis of *M. tuberculosis* is now frequently used for studying the molecular epidemiology of this bacterial pathogen, and more than 60,000 isolates have been genotyped with this method worldwide. Although IS6110 is reliable for evaluating TB transmission dynamics, the method does have certain limitations. For instance, isolates with six or fewer copies of IS6110 cannot be accurately discriminated on the basis of the hybridization patterns. That is, isolates with the identical hybridization pattern may not be clonal when only six or fewer bands comigrate by gel electrophoresis. An IS6110 band on a blot indicates only the presence and size of the insertion and does not provide the chromosomal location. Moreover, this insertion sequence cannot be used to distinguish among isolates within the *M. tuberculosis* complex, defined as *M. tuberculosis*, *M. bovis*, *M. africanum*, and *M. microti*. Additional genotyping tools have been developed to distinguish members of the *M. tuberculosis* complex.

Additional Genotyping Methods Serve Specialized Needs

Researchers have evaluated some of these additional genotyping markers for subspeciating *M. tuberculosis* by comparing them to the IS6110-based typing system. These secondary DNA target regions vary greatly, and each on its own is generally inferior in terms of discriminatory power to IS6110-based analysis. Most commonly used secondary methods include spacer oligonucleotide typing, called "spoligotyping," and polymorphic GC-rich repetitive sequence typing (PGRS). Both techniques are used to supplement IS6110 DNA fingerprinting analysis. More recently, PCR methods directed at multiple targets have also been employed. These include genomic deletion analysis and mycobacterial interspersed repetitive units typing (MIRU).

Analysis of DNA sequence has also served to provide reliable secondary markers, including regions or genes known to confer clinical drug resistance. Drug resistance in *M. tuberculosis* correlates with genetic alterations in specific housekeeping genes within the chromosome. In contrast to many other bacterial species, the high genetic conservation found in *M. tuberculosis* provides a background in which the DNA

sequence remains largely unchanged among susceptible isolates. Hence, mutations in these characterized genes nearly always correlate with drug resistance.

As most changes in target genes are point mutations, DNA sequencing provides a robust method for evaluating drug resistance genotypes of clinical isolates and another means of subspeciating this pathogen. Target gene sequencing is being used increasingly to evaluate transmission dynamics of mono-, poly-, and multidrug-resistant organisms. Thus, in addition to providing a precise molecular basis for resistance, analysis of sequence changes can also provide a means for tracking the spread of drug-resistant strains and for differentiating between primary and acquired resistance.

As with IS6110-based fingerprinting, secondary methods have limitations that need to be noted. Moreover, choosing appropriate methods to address a particular question will depend on the background strain population (presence of endemic clones), the hypothesis being tested, and the characteristics of the isolates themselves, such as those with six or fewer copies of IS6110 or ones that are grouped as drug resistant.

Genotyping Used for Controlling Tuberculosis, Following MDR

Investigators are only beginning to apply *M. tuberculosis* genotyping analysis to the purpose of controlling and preventing TB. Although only a decade old, genotypic analysis of clinical *M. tuberculosis* isolates is accumulating rapidly, providing information for an extensive and varied number of diverse geographic regions. Incorporating these molecular methods into traditional TB epidemiology and surveillance activities is proving crucial for investigators who are evaluating laboratory cross-contamination, confirming suspected outbreaks, tracking unsuspected transmissions, discerning between endogenous versus exogenous disease, and demonstrating exogenous reinfections in immunocompetent and immunocompromised patients. Genotyping has also been used to evaluate transmission dynamics in large population-based studies in defined geographic regions.

The resurgence of TB in the United States during the late 1980s was complicated by three main factors: a dismantled TB control system,

the HIV/AIDS epidemic, and the spread of drug-resistant organisms. New York City, for example, reported over 3,800 cases at the peak of the epidemic. Between 1990 and 1993, 357 patients were reported as being infected with a multi-drug-resistant (MDR) strain that could not be effectively treated with otherwise standard antibiotics, such as isoniazid, rifampin, streptomycin, ethambutol, pyrazinamide, and often kanamycin. Moreover, this strain was spreading into other individuals at local hospitals and in prisons.

IS6110 DNA fingerprinting analysis of 253 MDR patient isolates identified a distinct 18-band IS6110 hybridization pattern, designated strain W. These strain W cases were predominantly HIV positive (86%), and the mortality rate exceeded 85%. Further molecular analysis revealed a number of unusual markers. First, strain W, like all members of the W-Beijing family strains, has a high IS6110 band count, identical spoligotype, and a unique IS6110 insertion in the origin of replication (*ori*). The W strain also contains a rare dinucleotide change in codon 315 of the catalase-peroxidase gene, *katG*, correlating with isoniazid resistance, and two copies of IS6110 in the head-to-tail orientation in the specific DNA locus initially designated NTF. Finally, all W strains share an identical array of mutations in five drug resistance target genes associated with first-line antitubercular resistance.

The molecular and epidemiological data strongly support the hypothesis that the MDR strain W outbreak arose from a clone that developed comprehensive first-line drug resistance before being disseminated in the community. This hypothesis received further support when investigators identified five fluoroquinolone-resistant W strains with the expected array of mutations in their first-line, drug resistance gene targets plus five unique mutations in the *gyrA* loci. These observations demonstrate that fluoroquinolone resistance was acquired at least five independent times.

Since the strain W multi-institutional TB outbreaks in the early 1990s, investigators have characterized 11 MDR strains with variant IS6110 hybridization patterns. All of the strain W variants contain *katG* dinucleotide mutation, an IS6110 insertion in *ori*, two IS6110 insertions in the NTF region, and the same array of genotypic changes in the same five drug resistance

genes. The 11 variants display additional traits conferring drug resistance or other changes involving IS6110 elements. Besides containing such genetic changes, strain W and progenies have spread to 10 other states. Strain W thus is responsible for causing perhaps the most dramatic multidrug-resistant outbreak of TB so far, accounting for almost 25% of all MDR cases reported in the United States during the 43-month outbreak in New York City.

Paradigm Shift: Endogenous vs. Exogenous

Natural immunity against TB infection and disease has been debated ever since the administration of the BCG vaccine. The purpose behind administering BCG, an attenuated form of *M. bovis*, is to induce the host immune system to recognize and protect against uncontrolled replication and dissemination of the *M. tuberculosis* bacilli.

However, the efficacy of the BCG vaccine remains in doubt. Results from various studies are inconsistent and, in some instances, contradictory. Although BCG vaccination is widely believed to prevent extrapulmonary TB, it apparently has little to no effect in controlling pulmonary TB. Moreover, the ongoing high incidence and prevalence of both pulmonary and extrapulmonary TB in regions where BCG is administered raise further questions regarding the overall effectiveness of this vaccine. The extent of immunity against TB from early exposure to *M. tuberculosis* or from prior bouts with this disease is also far from being established.

Assessing the impact of either endogenous reactivation or exogenous reinfection on postprimary tuberculosis has immense implications for those developing vaccines with which to control TB. Many clinical TB experts believe that most recurrent cases of TB are due to endogenous reactivation of *M. tuberculosis*. Meanwhile, evidence suggests that reinfection also occurs on a limited basis, mainly in immunocompromised individuals. However, which host factors influence this phenomenon is not known.

Clinical, epidemiological, and microbiological data typically cannot determine whether recurrent TB is due to reactivation or reinfection. In contrast, molecular typing provides a robust approach for distinguishing whether such cases

of recurrent disease are attributable to relapse or reinfection. Indeed, molecular typing methods offer a means for capturing and characterizing the precise history of *M. tuberculosis* infections within a patient.

Paul van Helden of the University Stellenbosch Medical School in Tygenberg, South Africa, and his collaborators undertook the first population-based study to determine the extent to which recurrent TB is due to exogenous reinfections. Van Helden and his collaborators tested *M. tuberculosis* isolates from some 700 cases over a six-year period in an area of high TB burden in South Africa. All isolates were genotyped using IS6110 DNA fingerprint analysis. Sixteen patients with documented cure had recurrent TB, and fifteen of them were not infected with HIV.

When the initial and recurrent isolates from those patients were compared using DNA fingerprinting methods, the secondary infection in 12 of the 16 cases (75%) was due to exogenous reinfection by strains that differed from the initially infecting strains. The high rate of tuberculosis in that community is believed to account for its high proportion of exogenous reinfection. Another study by José Caminero from the University General Hospital “Dr. Negrín,” Las Palmas de Gran Canaria in Spain found a 44% rate of exogenous reinfections in recurrent cases of TB. The lower rate of exogenous reinfection in this study of TB patients in Gran Canaria Island, Spain, apparently reflects the moderate TB burden in this area.

The IS6110-based genotyping results provide useful evidence for reevaluating some important dogmas in TB epidemiology. The finding that exogenous reinfection accounts for a high proportion of TB in some regions has important implications for local control and prevention activities, including vaccine development. If this phenomenon were valid for most settings, it would lend additional credence to control strategies such as directly observed therapy, in which health care personnel systematically verify that TB patients take their antibiotics. As these hypotheses are further tested, those responsible for implementing global TB control strategies must be vigilant in their general efforts to reduce transmission of this pathogen because those efforts will also help reduce recurrent cases of TB caused by exogenous reinfections.

Population Studies: Elucidating Transmission Dynamics

M. tuberculosis causes disease with a varied latency period, ranging from as little as several months to a lifetime. Individuals can become symptomatic for different reasons, most commonly host immunosuppression, but also because of nutritional status, occupational risk, and stress. Hence, TB control programs for almost half a century have typically relied on conventional “shoe-leather” investigations, including interviews and tuberculin testing. Although contact investigations are useful, they are limited by the epidemiological setting in question. For instance, this approach has been effective in identifying links in low-incident environments or in narrowly defined outbreaks.

However, due to the disease’s variable latency and the pathogen’s airborne route of transmission, the thread of epidemiological links can quickly become obscure. For example, population-based, molecular-epidemiologic studies of TB within population groups in San Francisco and Baltimore could identify only 10–25% of the patient links within molecular-designated clusters.

These studies and others suggest that one of the main benefits in conducting molecular epidemiological studies on a population-based scale is to identify high-risk areas and groups. Molecular epidemiological investigations were crucial in confirming outbreaks in institutional facilities, such as that involving the W strain of *M. tuberculosis* in New York City, as well as in determining transmission patterns in social settings such as bars or churches. Several such studies suggest that casual contact is sufficient for *M. tuberculosis* to be transmitted between two individuals.

Contact investigations often do not identify the setting of transmission because of the transient nature of an individual frequenting any specific location. In contrast, population-based molecular epidemiology studies may reveal more about the setting where people are becoming infected with *M. tuberculosis* and thus could provide valuable information about localized TB transmission dynamics and ultimately could help those developing specific prevention and control activities for specific settings.

Population-based, molecular-epidemiological studies that use multiple genotyping methods

enable investigators not only to recognize outbreaks or epidemic situations in defined areas but also can provide insights when outbreaks are not even suspected. For example, in a recent localized, population-based study, we identified a large but previously unrecognized TB case cluster in northeastern New Jersey. We used several analytic tools to evaluate 43 isolates of *M. tuberculosis*, which were unlinked by traditional epidemiology, and then assigned them to a subgroup. Although most of these cases could not be linked by standard contact-tracing methods, the infected individuals shared demographic characteristics such as age, race/ethnicity, HIV serology, birthplace, and county of residence. We also used genetic markers to group isolates with similar, but unidentical IS6110 DNA fingerprint patterns into one epidemiologically relevant cluster, providing a model for recognizing the microevolution of a clone during an ongoing outbreak. Without molecular studies, this cluster would probably not have been detected within the generally high background of reported TB cases in that region of the state.

Applying Molecular Techniques to Population Studies May Improve TB Control Programs

Population studies that use both molecular techniques and conventional epidemiological approaches are enabling investigators to better address two key aspects of community transmission dynamics. First, molecular studies can improve the accuracy of contact investigations that, when used alone for determining clusters, are imprecise, highly interpretative, and otherwise limited. Second, the level of molecular clustering in a population-based sample is proportional to the TB cases attributable to a recently circulating strain. Hence, routinely collected genotyping data may uncover previously unrecognized outbreaks or sites where the pathogen is being transmitted. Properly applying population-based studies that use both surveillance and molecular techniques thus may lead to the development of more effective TB control programs.

As new molecular methods continue to be developed, investigators specializing in TB epidemiology can apply those tools for use in routine protocols for controlling this disease. Already, the integration of molecular methods is

aiding investigators to raise questions about key dogmas underlying *M. tuberculosis* epidemiology. The global burden of TB adds new urgency to these efforts, highlighting the value in developing more rapid and robust methods for genotyping this class of pathogens. Perhaps, with the *M. tuberculosis* genomic sequence now determined, additional methods that provide genome-wide analyses of synonymous single nucleotide polymorphisms will soon become more widely available as a way of genotyping clinical isolates.

Amid the possibilities surrounding these powerful methods, one must bear in mind that inferences drawn from molecular epidemiologic studies of TB can only be as precise and valid as the inherent limitations of the analytic tools themselves. Nevertheless, these recent applications of molecular techniques to address important questions within the classic TB control framework are making us reevaluate several of the basic biological and epidemiological properties inherent to this extensively studied, yet often enigmatic pathogen.

SUGGESTED READING

- Bifani, P. J., B. Mathema, N. E. Kurepina, and B. N. Kreiswirth. 2002. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. *Trends Microbiol.* 10:45–52.
- Bifani, P. J., B. Mathema, Z. Liu, S. L. Moghazeh, B. Shopsis, B. Tempalski, J. Driscoll, R. Frothingham, J. M. Musser, P. Alcabes, and B. N. Kreiswirth. 1999. Identification of a W variant outbreak of *Mycobacterium tuberculosis* via population-based molecular epidemiology. *JAMA* 282:2321–2327.
- Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole. 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci. USA* 99:3684–3689.
- Comstock, G. W. 1975. Frost revisited: the modern epidemiology of tuberculosis. *Am. J. Epidemiol.* 101:363–382.
- Frieden, T. R., L. F. Sherman, K. L. Maw, P. I. Fujiwara, J. T. Crawford, B. Niviv, V. Sharp, D. Hewlett, Jr., K. Brudney, D. Alland, and B. N. Kreiswirth. 1996. A multi-institutional outbreak of highly drug-resistant tuberculosis: epidemiology and clinical outcomes. *JAMA* 276:1229–1235.
- Kato-Maeda, M., P. J. Bifani, B. N. Kreiswirth, and P. M. Small. 2001. The nature and consequence of genetic variability within *Mycobacterium tuberculosis*. *J. Clin. Invest.* 107:533–537.
- Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M. Musser, and J. D. van Embden. 1999. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. *J. Clin. Microbiol.* 37:2607–2618.
- van Embden, J. D., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, and T. M. Shinnick. 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J. Clin. Microbiol.* 31:406–409.
- van Rie, A., R. M. Warren, N. Beyers, R. P. Gie, C. N. Classen, M. Richardson, S. L. Sampson, T. C. Victor, and P. D. van Helden. 1999. Transmission of a multidrug-resistant *Mycobacterium tuberculosis* strain resembling “Strain W” among noninstitutionalized, human immunodeficiency virus-seronegative patients. *J. Infect. Dis.* 180:1608–1615.
- Vynnycky, E., and P. E. Fine. 2000. Lifetime risks, incubation period, and serial interval of tuberculosis. *Am. J. Epidemiol.* 152:247–263.