Molecular Fingerprinting of *Mycobacterium tuberculosis*: How Can It Help the Clinician?

Marcel A. Behr and Peter M. Small

In just a few years, molecular fingerprinting of *Mycobacterium tuberculosis* has provided clinicians with significant insight into the epidemiology of tuberculosis. This methodology has allowed for a new understanding of the extent of new transmission of tuberculosis among residents of various communities and within institutions. It has also allowed for differentiation between episodes of reinfection and relapse, a task hitherto almost impossible to accomplish. In addition, molecular fingerprinting has allowed assessment of situations where laboratory cross-contamination is suspected. Thus, this technology has in many ways made clinicians reexamine many of their long-held beliefs regarding tuberculosis. In this report, Drs. Behr and Small provide a lucid description of molecular fingerprinting of *M. tuberculosis*, its current uses, and its future potential value.

—Wafaa El-Sadr
Chair, Tuberculosis Committee,
Infectious Diseases Society of America

In the 5 years since the development of DNA-based fingerprinting of *Mycobacterium tuberculosis* [1, 2], nearly 100 articles have been published on the use of this methodology in various epidemiological and clinical studies. In the public health domain, coordinated efforts involving conventional and molecular methods have shown foci of ongoing transmission in cities [3, 4]. The demonstration of transmission of *M. tuberculosis* in nosocomial settings [5–7] and congregate living facilities [8] and among persons at high risk such as the homeless [9, 10] and those who are HIV infected [6, 8] has been especially important. Fingerprinting in the context of geographic studies has shown the acquisition of *M. tuberculosis* of Tunisian or Ethiopian genotypes by Dutch persons who resided in Tunisia or Ethiopia [11], as well as spread of the organism between Greenland and Denmark [12].

Instances of low-level transcontinental transmission have been observed within the United States [13], while certain strains of *M. tuberculosis* have been traced through interstate spread [14]. The results of laboratory studies based on molecular fingerprinting have highlighted the significant problem of laboratory cross-contamination, with its attendant sequelae of infection-control costs and unnecessary therapy [15]. Observational studies have demonstrated exogenous reinfection of patients receiving antituberculous therapy [16] and the development of drug resistant tuberculosis in patients coinfected with HIV [17].

Although fingerprinting of *M. tuberculosis* is primarily performed in research facilities, fingerprinting is now available to all clinicians in the United States through the Centers for Disease Control and Prevention, and requests by clinicians for fingerprinting are becoming more frequent. In the present review we will briefly outline the techniques currently used in determining the molecular epidemiology of tuberculosis, describe some of the limitations of these techniques, and describe examples of where such techniques can assist clinicians in the treatment of patients with tuberculosis. For a more detailed bibliography on fingerprinting of *M. tuberculosis*, or for a listing of related activities in the field, the reader is referred to our World Wide Web home page and relevant links at: http://molepi.stanford.edu.

**Methods**

The most widely used methodology for fingerprinting *M. tuberculosis* exploits restriction fragment length polymorphism (RFLP) of chromosomal DNA. A simplified illustration of the procedure is shown schematically in figure 1.

Simple endonuclease digestion of whole DNA, followed by electrophoretic separation (known as Restriction Enzyme Analysis or REA), allows comparisons of small numbers of strains. However, the generation of many bands renders the gels difficult to interpret and makes comparison of many gels nearly impossible. To simplify analysis, it is possible to perform
although PGRS is used in many laboratories for isolates with The limitations of this methodology include the existence of repeat (DR) sequences for identification. The selection of such others accept limited changes such as single band additions, deletions, or direct transmission. Conversely, it may be that genetic change has occasionally been detected over time in strains with very low numbers of copies or no copies of IS\textsubscript{6110}.

Figure 1 shows the IS\textsubscript{6110} blotting of two similar isolates of M. tuberculosis. 1. The genomic DNA of M. tuberculosis is shown as a closed circle; shaded ovals indicate IS\textsubscript{6110}, an insertion sequence whose variability in copy number and genomic location is the basis for genotyping. 2. M. tuberculosis DNA is cleaved into many fragments with a restriction endonuclease (for the purpose of illustration, relatively few fragments are shown here, while in reality there are thousands); a subset of fragments contain IS\textsubscript{6110}. 3. DNA fragments are separated electrophoretically according to size (in actuality, the thousands of fragments would appear nearly confluent, with little visible separation). 4. Hybridization with probe for IS\textsubscript{6110} and chemoluminescent detection reveals those fragments that contain IS\textsubscript{6110}; this pattern constitutes the DNA fingerprint unique to each strain. Loss of restriction site indicated by * leads to changes seen with isolate #2. During electrophoresis, two fragments have been replaced by one larger (slower) fragment, while the final IS\textsubscript{6110} blot has a single shift.

Southern blotting of electrophoretically-separated DNA and hybridize this DNA with probes to determine the presence and size of fragments containing specific DNA sequences. Repetitive elements called insertion sequences (ISs) are present in variable sites and copy numbers, allowing comparison of the number and size of fragments containing an IS. The most commonly used IS is IS\textsubscript{6110}, which is found throughout the M. tuberculosis complex, usually in 5–20 copies at various positions in the chromosome. Bacteria with fewer copies, including Mycobacterium bovis and M. bovis BCG, often cannot be discriminated with use of IS\textsubscript{6110} alone and require secondary typing procedures such as blotting for polymorphic GC (guanine-cytosine)-rich sequences (PGRS) or the use of IS\textsubscript{1081} or direct repeat (DR) sequences for identification. The selection of such secondary typing techniques is somewhat controversial, and although PGRS is used in many laboratories for isolates with fewer than six copies of IS\textsubscript{6110}, at this time there is little consensus as to the relative utility of the various secondary typing schemes.

After blots are developed, the patterns obtained may be visually compared; however, if large numbers of patterns are generated, they are scanned into a computer database. This permits comparisons with previous fingerprints by using a combination of several human readers and computer assistance. The criteria for considering patterns to be “matched” vary between applications, with some requiring identity and others permitting some variability. Differences between patterns may represent genetic differences or artifacts that result from varying gel conditions and hybridization efficiency. The two principal types of genetic difference that can be observed involve change in the number of hybridization fragments (representing the number of IS\textsubscript{6110} copies in an isolate) and their size (representing the length of the restriction fragments containing IS\textsubscript{6110}). Since a single mutation can modify the distribution of restriction sites, it is possible through one molecular event to alter the number and size of bands. Artifactual changes may result when two bands are so close together that they appear as one wide band or when a band is so faint because of poor hybridization that it is imperceptible.

Figure 2 demonstrates the IS\textsubscript{6110} blot of several isolates of M. tuberculosis, illustrating the typical image obtained and the representative differences between isolates. Because differences in restriction enzymes and electrophoresis and hybridization techniques can complicate comparisons, a standard methodology with use of IS\textsubscript{6110} was agreed upon by international consensus to permit interlaboratory strain comparisons [19].

Issues of interpretation, however, are less concretely resolved. It appears clear that the detection of isolates with identical fingerprints containing many copies of IS\textsubscript{6110} indicates transmission of M. tuberculosis [5]. Likewise, markedly different patterns are accepted to indicate unrelated strains. However, when two isolates have similar but not identical patterns, there are two possible interpretations: it is possible that the two isolates diverged before the respective patients became infected and that the similarity is therefore a manifestation of shared history without direct transmission. Conversely, it may be that genetic change of the organism occurred during the process of transmission, and thus isolates with simple differences could represent temporally proximate transmission. In support of this possibility, subtle changes have occasionally been detected over time in strains isolated from chronically infected patients [20].

The rapidity of pattern change (the molecular clock) is unclear but appears to vary according to the technique used (e.g., probing for IS\textsubscript{6110} vs. PGRS); the impression is that IS\textsubscript{6110} is less stable [18]. Consequently, the epidemiological interpretation of similar fingerprints remains controversial; some investigators define transmission on the basis of identical fingerprints [3, 4], while others accept limited changes such as single band additions, deletions, and shifts [5, 10].

The limitations of this methodology include the existence of strains with very low numbers of copies or no copies of IS\textsubscript{6110}
Figure 2. IS (insertion sequence) 6110-based DNA fingerprinting of 12 Mycobacterium tuberculosis complex isolates. The sizes of the restriction fragments are given on the left (in kilobases). Lanes 1 and 12 demonstrate M. tuberculosis reference strain H37Rv; lanes 2 and 3 demonstrate a single (faint) band addition (+); and lanes 4 and 5 demonstrate a single band shift (\(^+\)); lanes 6 and 7 demonstrate shift of a doublet (\(++\)), which could have been mistaken for a single wide band. Lane 8 has 10 bands, sufficient for epidemiological analysis, while lanes 9–11 have two or fewer bands, necessitating secondary typing. Lane 9, two-banded M. tuberculosis; lane 10, M. tuberculosis with no copies of IS6110; lane 11, one-banded pattern of M. bovis bacillus Calmette-Guerin (Glaxo).

Indications

The probability of obtaining useful information with fingerprinting of M. tuberculosis, as with any laboratory test, is a function of the \(a priori\) suspicion of a positive result. In the context of a tight geographic and temporal cluster, fingerprints will clearly demonstrate whether cases were part of the epidemic. When isolates of less-defined epidemiology are studied, similarities may be observed [24], but the implication of these similarities in ongoing transmission is less certain.

At present, the indications for M. tuberculosis fingerprinting include the situations described below, but as new questions arise and the technology improves, other scenarios will likely present themselves.

**Laboratory cross-contamination.** Episodes of laboratory cross-contamination by M. tuberculosis have been delineated with the assistance of IS6110 genotyping, both in the context of traditional mycobacteriology and radiometric systems [15]. It is believed that the sampling needles used in automated machines are the source of contamination; however, it has also been demonstrated that during conventional decontamination of specimens, there can be splashing and transfer of organisms from the mucolytic solution [25]. When a patient is reported to be culture positive for M. tuberculosis but the clinical data are not consistent with the diagnosis, the clinician may suspect laboratory cross-contamination from another patient’s sample. The first step in evaluating this possibility is to determine whether the patient’s acid-fast smear was negative and whether any other M. tuberculosis isolates were processed during the same period at the same laboratory; if other isolates were processed, they can be typed to determine whether there was possible in vitro cross-contamination.

Fingerprinting would also be useful in the event that an unexpected number of cultures are found to be positive for...
**M. tuberculosis.** In such a case, it would be reasonable to perform fingerprinting as a quality control measure in the absence of clinical suspicion of a false-positive diagnosis. The results should be evaluated in the context of the clinical findings, since matched patterns could be due to a true outbreak or cross-contamination (a pseudo-outbreak).

**Exogenous reinfection.** A patient previously treated for *M. tuberculosis* infection who presents with a second episode of tuberculosis may have relapsed because of inadequate therapy or may represent exogenous reinfection. IS6110 blotting of the past and present isolates will determine whether a new strain is responsible for the second presentation. Therapy for the second infection will be governed by the results of antimicrobial susceptibility studies, but knowledge that a patient has been newly infected should alert clinicians and public health officials to reevaluate the case (e.g., for the presence of risk factors such as HIV infection) and to perform an additional source investigation.

Furthermore, given that tuberculosis treatment trials have generally been performed in high-prevalence areas, it is likely that ongoing transmission at these sites occurred, and some cases classified as relapse may have been reinfection. Investigators conducting current efficacy trials use genotyping of subsequent isolates to differentiate relapse from reinfection [26].

**Investigating the source of infection.** Where a transmission link is epidemiologically suspected (e.g., among health care workers [27]), it may be possible to obtain the suspected source patient’s culture and perform genotyping of both isolates to demonstrate transmission. In settings where an ongoing surveillance project exists (e.g., San Francisco), the fingerprint of the new isolate can be immediately matched to an existent fingerprint database that contains all local isolates of *M. tuberculosis*. In this manner, it may be possible to determine if the acquisition was due to occupational exposure or unrelated transmission.

When an unusual number of cases of tuberculosis occurs over time (e.g., during an institutional outbreak), it is possible to do fingerprinting early on in the investigation to determine whether this cluster represents temporal coincidence or a cluster due to transmission. In this situation, if fingerprinting demonstrates different strains, the cases are not due to transmission, and there is no need for further epidemiological evaluation.

**Changes in antimicrobial susceptibility.** When multiple isolates of *M. tuberculosis* with varying antimicrobial susceptibilities are recovered from a single patient, acquired drug resistance, reinfection, or a laboratory error may be involved. The acquisition of drug resistance, especially in HIV-positive patients, has been demonstrated to occur without change in a strain’s fingerprint [17]. Conversely, a patient may be reinfected with a new strain [16] or harbor more than one strain (author’s unpublished data), which could also result in a variable antibiogram. Clinicians faced with the issue of acquired drug resistance should consider therapeutic issues such as drug malabsorption and the use of directly observed therapy. New infection, on the other hand, would indicate the need for a source investigation.

**Diagnosis of disease due to bacillus Calmette-Guérin (BCG) and other atypical isolates.** Intravesical BCG has been used since 1976 for the treatment of stage I bladder cancer, and although this treatment is generally well tolerated, dissemination of BCG has been implicated in reports of granulomatous hepatitis, interstitial pneumonitis, osteitis, and other infections [28]. In countries where childhood vaccination is routine, disseminated BCG infection has been described in children with immunosuppression due to severe combined immunodeficiency, chronic granulomatous disease, complete DiGeorge syndrome, and AIDS [29]. Since commercial probes used in diagnostic microbiology laboratories are unable to identify *M. tuberculosis* complex isolates to the species level, IS1081 blotting of an isolate, typically along with the BCG strain of interest, has been performed to definitively identify the isolate as BCG [30].

Diagnosis of *M. bovis* infections has typically not been as difficult; however, situations arise where an isolate has features that are not typical of either *M. bovis* or *M. tuberculosis* [31]. Fingerprinting with use of IS6110, as well as other probes, has helped clarify the identity of such strains, showing whether they were acquired from cows or from wild animals [32], or even occupationally, as in the case of seal handlers [33].

**Conclusion**

A century since Koch’s discovery of the tubercle bacillus and a half century since the advent of effective chemotherapy, the white plague has not abated; this disease still causes an estimated 3 million deaths per year. Molecular typing of *M. tuberculosis* is a new, exciting, and powerful tool in the armamentarium of tuberculosis epidemiology whose ultimate contribution is still unclear. It is conceivable that the ability to identify and track individual clones of *M. tuberculosis* may ultimately provide important insights into pathogenesis and novel disease-control strategies. However, for now, clinical applications of this tool are relatively limited, and genotyping of clinical isolates outside a defined research protocol should be considered only in select instances. These instances include suspected laboratory cross-contamination, differentiation of relapse from exogenous reinfection, clarification of the cause of a change in drug susceptibility, recognition of disseminated BCG disease, and corroboration of suspected transmission.

The treatment and control of tuberculosis is a dynamically changing endeavor, and thus the full utility of fingerprinting of *M. tuberculosis* isolates, integrated with conventional methods, remains to be seen. With the ongoing contribution of clinicians and the specific scenarios they present, the fingerprinting of *M. tuberculosis*, together with other clinical information, may realize a greater role in the clinical management of tuberculosis. In addition, the public health role of this tool in gauging the efficiency of control programs is an exciting prospect whose contribution remains to be determined.
References