Review article

Genotyping and identification of mycobacteria by fingerprinting techniques

Azar Dokht Khosravi1, Sepideh Seghatoleslami2

1Department of Medical Microbiology, School of Medicine, and Infectious and Tropical Diseases Research Center, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran
2Department of Medical Microbiology, School of Medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

Received: July 2009 Accepted: September 2009

Abstract

Tuberculosis (TB) remains a worldwide healthcare concern and has been characterized as an epidemic by World Health Organization (WHO). In the last few decades, DNA fingerprint techniques have become available to study the transmission of TB and other mycobacterial infections. The main gold standard typing technique is restriction fragment length polymorphism, which is widely used for molecular epidemiology purposes. However, other recently introduced techniques such as spoligotyping and MIRU-VNTR are also widely used. This review briefly summarizes the recent achievements in TB epidemiology associated with the introduction of molecular typing methods for Mycobacterium tuberculosis complex and non-M. tuberculosis complex mycobacteria. Besides, in this article, the application of molecular methods in the study of epidemiological aspects and diagnosis of mycobacterial diseases are highlighted.

Keywords: Mycobacteria, Mycobacterium tuberculosis, Genotyping, Fingerprinting

Introduction

Tuberculosis (TB) remains a worldwide healthcare concern and has been characterized as an epidemic by World Health Organization (WHO). It is estimated that one third of the world's population has been infected with Mycobacterium tuberculosis (MTB) and that three million people will die of the disease per year between now and 2010. The distribution of TB in different geographic regions is characterized by the prevalence of different MTB strains with varied virulence and drug resistance. Both environmental and host factors are responsible for the transmission and prevalence of different MTB strains [1]. A crucial aspect of any TB control program is the ability to determine where transmission occurs in order to prevent further spread of infection and prevent active disease by identifying newly infected people and providing them with preventive therapy. Genetic fingerprinting of MTB has vastly improved our ability to observe patterns of transmission in populations. It has helped to establish transmission links between individuals and to demonstrate instances in which related people were infected with unrelated strains [2]. Several
molecular techniques can be used for genetic fingerprinting as a tool for studying molecular epidemiology of tuberculosis. Besides, these methods could also be applied for other purposes such as identification and differentiation of MTB and non-tuberculous mycobacteria (NTM) to species level and study of drug susceptibility pattern among MTB isolates. In this paper, the main genotyping methods and their application in the study of different aspects of mycobacteria are reviewed.

Genotyping of mycobacteria for epidemiological purposes

The molecular typing of MTB has greatly improved the knowledge and control of TB by allowing the detection of unsuspected transmission, the identification of false-positive cultures, and the distinction between re-infection and relapse [3]. Molecular markers provide an important method for detecting TB transmission that has been followed by rapid progression to active disease [4]. The understanding of TB transmission dynamics has been greatly enhanced by the development of various DNA typing methods [5,6]. Genotyping can also be used to evaluate an outbreak of TB. If epidemiologic data suggest the occurrence of an outbreak, 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 [3].

There is growing evidence that the genetic diversity of MTB may have important clinical consequences. Nearly 50 years ago, Mitchison and others compared the consequences of infecting guinea pigs with isolates of MTB from either British or Indian patients with pulmonary TB. They reported that the British isolates were more virulent: they caused more-severe and more-widespread disease and killed a higher proportion of animals. However, the investigators were not able to characterize the genetic diversity of the infecting isolates, and they could not find any association between virulence in guinea pigs and the severity and outcome of disease in humans. A further understanding of the relationship between mycobacterial genotype and clinical phenotype came with the advent of mycobacterial genotyping [7].

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 three percent of patients from whom MTB is apparently isolated in clinical laboratories do not have TB; 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 MTB is isolated from a specimen but the clinical findings do not suggest the presence of TB, genotyping of the isolate and other MTB strains handled concurrently in the laboratory can strongly suggest the occurrence of cross contamination and lead to the discontinuation of anti-tuberculosis medications [3]. There is broad variability in the genotypes of MTB isolates from patients with epidemiologically unrelated TB, whereas the genotypes of isolates from patients who were infected by a common source are identical. Therefore, clustered cases of TB, 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.

Molecular epidemiologic studies have shown that the dynamics of the transmission
of TB 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 TB [3]. DNA polymorphism associated with insertion elements and other repetitive DNA elements has been exploited to differentiate clinical MTB isolates for epidemiological studies, based on the assumption that strains with identical DNA fingerprints are epidemiologically linked. Strains with non-identical DNA fingerprints can be grouped into clades or genogroups when they partially share polymorphic sites, such as IS6110 insertion sites and spacer sequences in the polymorphic direct repeat (DR) region of the MTB genome [9].

Genotyping methods

Numerous repeat sequences have been identified in the genome of the M. tuberculosis complex (MTC), including transposable elements [9], trinucleotide repeats [10], variable number tandem repeats, mycobacterial interspersed repetitive units [11], and the DR region [12]. The DR region is one of the most extensively studied loci and consists of direct repeat sequences (36 bp) interspersed with unique spacer sequences (34 to 41 bp), which together are termed direct variable repeat (DVR) sequences. The DR region has evolved through the deletion of DVR sequences by homologous recombination, single nucleotide mutations, and the integration of IS6110 elements [13]. These events are believed to be unidirectional and to occur over time, making the DR region an informative locus for studying the evolution and epidemiology of the MTC [14-17].

RFLP

The standard approach to genotyping MTB isolates is restriction-fragment-length polymorphism (RFLP) analysis of the distribution of the insertion sequence IS6110 in different strains, and large databases of IS6110-based genotypes are available. This method is based on differences in the IS6110 copy numbers per strain, ranging from 0 to about 25, and variability in the chromosomal positions of these IS6110 insertion sequences [18]. 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 [12].

Genetic fingerprinting analysis can only be used in patients with culture confirmed TB, not in patients with a latent infection. IS6110-based genotyping requires sub culturing the isolates for several weeks to obtain sufficient DNA [18]. For fingerprinting analysis, the DNA from the bacterial culture is first extracted and then purified. The DNA is then treated with the restriction enzyme PvuII; this recognizes a specific six-nucleotide palindromic sequence in the DNA and cleaves it at each occurrence of the sequence. The resulting restriction fragments are separated electrophoretically based on their size, transferred onto a nylon membrane, and hybridized with an IS6110 probe. This probe then specifically marks those fragments, which contain the repetitive element IS6110 (Fig. 1). The number of resulting bands corresponds to the number of IS6110 copies, and the localization of the bands reflects the molecular weight of the fragments containing the IS6110 copies [19,20].
Unfortunately, a proportion of MTB isolates contain no, or only a few, copies of IS6110, and this proportion differ significantly by geographical area. Strain typing based on a low copy number of IS6110 is not sufficiently discriminatory, so it has been decided to apply an additional genetic typing method when MTB isolates contain fewer than five copies of IS6110. There are several genetic markers that can be used in addition to IS6110 RFLP typing, one of these relatively simple methods is spoligotyping. Most laboratories routinely apply the polymorphic GC-rich sequence (PGRS) RFLP typing as a supplementary typing method [2]. Typing methods should preferably be rapid, reproducible, easy to perform, inexpensive and directly applicable to clinical material. None of the currently used typing methods meets all these criteria. Furthermore, the degree of discrimination and stability should be appropriate to the research question to be addressed. For outbreak management, other genetic markers are needed than for determination of the evolutionary divergence between microorganisms. Although there is no perfect method, the application of IS6110-based RFLP typing and other molecular typing methods has contributed significantly to our understanding of transmission of mycobacteria. It has also contributed most frequently used DNA fingerprint methods for mycobacteria currently available, and to review the improved understanding of the epidemiology of mycobacterial diseases gained by the application of DNA fingerprinting [18].

**REA**

Another main target for DNA fingerprinting purposes is *hsp65* sequences, which are highly conserved within a species and thus can be used for taxonomic studies. *hsp65* is a promising target because the method based on this target has been widely used since 1992 and the technique employs only two restricted enzymes for broad identification of mycobacterial species [21]. PCR based restriction enzyme analysis (PRA) of the *hsp65* or PCR-RFLP, is one of the simple and rational methods developed by Plikaytis *et al.* and later modified by Telenti *et al.* [22] for rapid identification of the most mycobacterial isolates. In this method, DNA is cleaved with particular restricted enzymes, the resulting DNA restriction fragments are separated on agarose gels and analyzed by eye [21-24]. Later on, repetitive DNA elements were cloned that could be used as probes to visualize only those restriction fragments that contain the DNA sequence complementary to the probe: RFLP typing [21]. Telenti *et al.* [22] demonstrated that a 439-bp portion of the *hsp65* gene could be used for PRA and showed the patterns for most slowly growing mycobacteria and
selected rapid growing mycobacteria (RGM). Steingrube et al. provided the most detailed PRA study to date of the RGM. They reported the PRA patterns from the 439-bp Telenti segment of the \textit{hsp65} gene for 129 clinical and reference strains of RGM belonging to 10 taxonomic groups [25].

In later studies, the method used for differentiation of NTM and the results found the method as a fast and reliable discriminating technique except reported variation for a few NTMs [23]. Besides, the technique was able to fully identify MTB isolates [24], and discriminate them from NTM strains [26]. Vaneechoutte et al. devised another system of enzymatic amplification and restriction analysis using the entire 16S rRNA gene sequence [27]. They studied 18 different species of mycobacterium including strains of \textit{M. fortuitum} and \textit{M. chelonae}. They used different restriction enzymes (\textit{CfoI}, \textit{Mbol}, and \textit{RsaI}) from Telenti et al. [22] and Steingrube et al. (both of whom used \textit{BstEII} and \textit{HaeIII}) and called their method ARDRA (amplified rDNA restriction analysis). Other target sequences have been studied for the identification of mycobacteria by using PRA or sequencing. These include the 32-kDa protein gene, the internal transcribed spacer of the 16S-23S rRNA gene, the superoxide dismutase gene, and the DNA J gene. However, to date, only selected slowly growing species have been extensively studied using these gene sequences and only one or two isolates of RGM have been tested. Preliminary data suggest that for the RGM, most of these gene sequences are much more variable and perhaps less useful for species identification than is the \textit{hsp65} gene [28].

**Spoligotyping**

Molecular typing, targeting different molecular markers, has been utilized recently for the identification of potential sources of infection and epidemiological investigations of tuberculosis. RFLP fingerprinting, with insertion element IS6110 as a probe, has good discriminatory power and is frequently used as a method of choice to differentiate strains of MTB isolates [29]. However, absence or low copy number of IS6110 element in significant number of MTB strains, as reported elsewhere, further limit its usefulness. The introduction of new PCR-based typing methods, spoligotyping, MIRU-VNTR and DRE-PCR typing has allowed simultaneous detection and epidemiologic typing of MTB [30]. Spoligotyping was developed as a tool to provide information on the structure of the DR region in individual MTB strains and in different members of the MTC. Spoligotyping is basically a reverse hybridization technique based on polymorphism in the DR locus of the mycobacterial chromosome [30].

The direct-repeat locus in MTB 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). 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 MTB 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. This method is simple and the results can be obtained from a \textit{M. tuberculosis} culture within one day. Thus, the clinical usefulness of spoligotyping is determined by its rapidity, both in detecting causative bacteria...
and in providing epidemiologic information on strain identities [31]. However, spoligotyping has less power to discriminate among MTB strains than does IS6110-based genotyping [3].

The simplicity of this method has allowed for the establishment of an international spoligotype database, which describes 39,295 entries from 122 countries. Alignment of the spoligotype patterns has allowed authors to group isolates according to similarity to create clades or strain families [32]. In addition, distinctive spoligotype patterns have been linked to defined species of the MTB complex, although, these evolutionary relationships have not been extensively tested with other genotyping methods [33,34].

MIRU-VNTR

The gold standard method for MTB genotyping is IS6110 fingerprinting. However, this method is laborious and requires weeks of MTB culturing, which limits the possibilities to use typing prospectively for more efficient TB control. Furthermore, comparison of the fingerprints from large data sets requires highly standardized experimental and computerized procedures [35]. These problems complicate the determination of clustering rates in population-based studies and the exchange of data. The most promising PCR-based methods are based on PCR amplification of multiple loci containing mycobacterial interspersed repetitive-unit–variable-number tandem repeats (MIRU-VNTR) offers a potential solution to these drawbacks. This method is highly reproducible and yields portable results. It is much faster than IS6110-RFLP typing, since it can be directly applied to crude extracts from colonies or early culture pellets. Among the different described sets of MIRU-VNTR loci [36], a system based on 12 loci, is now the most widely used in clinical mycobacteriology and for local outbreak investigation [37]. Genotypes based on these 12 loci are highly stable among epidemiologically linked isolates but sufficiently diverse to generate a resolution approaching that of IS6110-RFLP in some sets of isolates from low-incidence settings or from diverse geographic origins. A population-based study indicated that the use of this 12-locus-based MIRU-VNTR typing combined with spoligotyping as a first-line approach provides adequate discrimination in most cases for large-scale genotyping of MTB in the United States [38].

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 database. A Web site has been set up so that a worldwide database of MIRU patterns can be created. MIRU genotyping is technically simpler than IS6110-based genotyping and can be applied directly to MTB cultures without DNA purification. It may replace IS6110-based genotyping in the future, particularly if the evaluation of additional loci increases the discriminatory power [3,39].

Pulsed-field gel electrophoresis

Standard gel electrophoresis techniques for separation of DNA molecules provided huge advantages for molecular biology research. However, many limitations existed with the standard protocol in that it was unable to separate very large molecules of DNA effectively. It also requires the use of specific DNA probes and labeling kits [40]. DNA molecules larger than 15-20kb migrating through a gel will essentially move together in a size-independent manner. At Columbia University in 1984, Schwartz and Cantor [41] developed a variation on the standard protocol by
introducing an alternating voltage gradient to better the resolution of larger molecules. This technique became known as Pulsed Field Gel Electrophoresis (PFGE). The development of PFGE expanded the range of resolution for DNA fragments by as much as two orders of magnitude.

This is an alternative method of simplifying restriction digests involves the use of infrequent-cutting restriction endonucleases. These enzymes, which have on average fewer than 10 recognition sites per 106 bp, cut the whole chromosomal DNA into a few large fragments. These large restriction fragments (LRFs) cannot be separated by conventional agarose electrophoresis but can be separated by pulsed-field gel electrophoresis (PFGE). PFGE can cleanly separate DNA fragments as large as 1.2 megabases, so it can separate species, MTB, M. bovis, M. africanum, and M. microti [42]. Pulsed-field gel electrophoresis (PFGE) has been widely used to type various microorganisms in both outbreak and technically challenging. Biosafety considerations and the unique cell wall composition of the organism have led to the development of protocols that are highly complex and difficult to reproduce. Little has been done to develop standardized methods for analysis of MTB PFGE patterns. The lack of standardized methods to generate, store, and compare PFGE patterns has also limited the use of PFGE for population-based molecular epidemiologic studies of MTB [43].

**Genotyping for investigation drug resistance**

Until about a decade ago, the only markers available to study the epidemiology of TB were drug susceptibility profiles and phage types. The use of either method had serious limitations. The drug susceptibility profile of MTB strains is a highly unstable feature, because strains frequently gain resistance to anti-tuberculous drugs during treatment. The predictive value of phage typing to link tuberculosis cases is also limited, because only a few phage types can be distinguished amongst MTB isolates. In most areas, one phage type predominates amongst MTB isolates; related and unrelated cases cannot be distinguished on this basis [18]. 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 anti-tuberculosis therapy, the patient was re-infected with a different MTB strain, or cross-contamination is suspected. Genotyping can be used to distinguish the patient isolates from others. If the original organism developed resistance, the cause could be non-adherence to therapy or reduced concentrations of anti-tuberculosis drugs as a result of malabsorption or drug interactions. If the cause was re-infection, however, public health authorities should attempt to identify the source [44].

**Genotyping for clinical identification purposes**

The cultivable members of the genus mycobacterium can be distinguished into two groupings: the MTC and the non-MTC (atypical) or NTM [18]. While TB due to MTC strain is the most common mycobacterial infection in developing countries, many NTM are also of medical relevance, particularly for immunocompromised patients. The NTM infections occur more frequently in developed countries where the incidence of TB is low [45]. There is a need for rapid diagnosis since increased morbidity and mortality is associated with NTM infections. As few clinical and radiological findings differentiate NTM infections from TB, microbiological identification to the species level is necessary. The distinction between species has not only epidemiological implications but is also
relevant for the management of patients in
to appropriate treatment, isolation, and contact tracing. Antibiotic treatment may vary according to the species encountered [46]. The traditional diagnosis of mycobacterial infections from sputum and other clinical samples in the mycobacterial laboratory is based primarily on demonstrating the presence of the acid-fast bacilli (AFB) in the smear, followed by a positive culture and the testing of the physiological/biochemical identification of the isolate. This approach has a number of defects, including that it is time-consuming, has low sensitivity, and poor discrimination between closely related NTM species [45].

Although molecular methods have not been able to replace culture for the detection of Mycobacteria in clinical specimens, their application combined with cultivation has accelerated the laboratory diagnosis of mycobacterial infections. Recently, PCR or PCR-linked methods have been used for rapid detection and differentiation of MTC and NTM. Multiplex PCR targeting of many different genes simultaneously has been used to detect and identify MTC and NTM in routine diagnostic laboratories. The DNA sequencing of mycobacterial gene targets such as 16S rRNA, rpoB, hsp65, secA, and 16S-23S internal transcribed spacer (ITS) region genes have been potential for species-specific identification of nearly all mycobacterium spp. [47]. Goyal et al. [48] in 2008 preferred the PRA techniques because these are simple and cost effective. This study was undertaken to evaluate a new rapid method to identify the mycobacterial isolates at species level by gene amplification restriction analysis using primers encoding 16S-23S rRNA internal transcribed spacer (ITS) region and flanking parts of the 16S as well as 23S rRNA gene. Restriction was carried out with restriction enzyme Hha I. Khosravi et al. [49] used PRA technique for investigating the possible variation in MTB strains, however their findings revealed little or no genetic heterogeneity among the strains [24].

DNA fingerprinting techniques are of interest for identifying the subtypes/strains, which would be more commonly associated with disease, and also to investigate hospital acquired infections. Several DNA fingerprinting techniques have been tried to investigate the diversity in NTM. Techniques like pulsed field gel electrophoresis, random amplified polymorphic DNA (RAPD)- arbitrary PCR, rRNA gene polymorphism, typing using different insertion/repeat elements, plasmid typing and single gene polymorphism have been successfully used for molecular typing of NTM. Insertion sequence based RFLP methods have been described to be useful for characterization of M. hemophilum, M. avium as well as M. kansasii. With the use of these methods, certain RFLP types of M. avium have been shown to be closely linked with disease in Europe and Africa [50]. The PCR-RFLP technique was successfully applied for identification of MTB and NTM strains in Iran and the technique proved to have good discriminatory power for differentiation between these two groups of mycobacteria [26].

Among targets for molecular techniques, use of the genus-specific 16S rRNA and the rpoB gene of mycobacteria for the identification of NTM has also been documented [51]. The 16S rRNA gene analysis is the most promising molecular method. The 16S rRNA gene is conserved in all species of mycobacterium genus. Therefore, this gene can be amplified by PCR in all species of mycobacterium and the species can be identified by species-specific probes, by PCR primers, or by sequencing the hypervariable region of the 16S rRNA amplicon. Identification of all NTM species may require several species-specific primers and repeated experimentation. In addition, other genes,
such as those encoding the conserved and nonconserved regions in 65-kDa heat shock proteins of mycobacteria, have been found to be suitable for the identification of NTM [51, 52].

References
19) Seidler A, Nienhaus A, Diel R. The transmission of tuberculosis in the light of

20) Suffys PN, de Araujo ME, Degrave WM. The changing face of the epidemiology of tuberculosis due to molecular strain typing- a review. *Mem Inst Oswaldo Cruz.* 1997; 92: 297-316.


36) Blackwood KS, Wolfe JN, Kabani AM. Application of mycobacterial interspersed repetitive unit typing to Manitoba tuberculosis cases: can restriction fragment

Address for correspondence:
Azar Dokht Khosravi, Department of Medical Microbiology, School of Medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran
Tel: +98611 3367543-9; Fax: +98611 3332036
Email: kchosraviz@yahoo.com