The Changing Face of the Epidemiology of Tuberculosis due to Molecular Strain Typing - A Review

Philip N Suffys†, Marcelo E Ivens de Araujo, Wim M Degraw

Laboratório de Biologia Molecular e Diagnóstico de Doenças Infecciosas, Departamento de Bioquímica e Biologia Molecular, Instituto Oswaldo Cruz, Av. Brasil 4365, 21045-900 Rio de Janeiro, RJ, Brasil

About one third of the world population is infected with tubercle bacilli, causing eight million new cases of tuberculosis (TB) and three million deaths each year. After years of lack of interest in the disease, World Health Organization recently declared TB a global emergency and it is clear that there is need for more efficient national TB programs and newly defined research priorities. A more complete epidemiology of tuberculosis will lead to a better identification of index cases and to a more efficient treatment of the disease. Recently, new molecular tools became available for the identification of strains of Mycobacterium tuberculosis (M. tuberculosis), allowing a better recognition of transmission routes of defined strains. Both a standardized restriction-fragment-length-polymorphism-based methodology for epidemiological studies on a large scale and deoxyribonucleic acids (DNA) amplification-based methods that allow rapid detection of outbreaks with multidrug-resistant (MDR) strains, often characterized by high mortality rates, have been developed. This review comments on the existing methods of DNA-based recognition of M. tuberculosis strains and their peculiarities. It also summarizes literature data on the application of molecular fingerprinting for detection of outbreaks of M. tuberculosis, for identification of index cases, for study of interaction between TB and infection with the human immunodeficiency virus, for analysis of the behavior of MDR strains, for a better understanding of risk factors for transmission of TB within communities and for population-based studies of TB transmission within and between countries.

Key words: tuberculosis - molecular epidemiology - fingerprinting - restriction fragment length polymorphism

This research received financial support from Faperj, CNPq, PAPES and CABBIO.
†Corresponding author. Fax: +55-21-270.9997
Received 24 July 1996
Accepted 21 October 1996

Upon opening the World Wide Web (Internet) version of the World Health Organization (WHO) Report on the Tuberculosis (TB) Epidemic, three major headlines stand out: TB will kill 30 million people this decade; someone is infected with TB every second and TB drugs may become useless. This dramatic introduction is followed by: The TB epidemic is growing larger and more dangerous each year. If the TB epidemic continues to be neglected, future generations will remember this decade as the time when humanity allowed deadly bacilli that travel through the air to become drug-resistant and incurable throughout the world. In 1993, WHO declared TB as a global emergency since it is globally the leading cause of death associated with infectious diseases. Tuberculosis is especially predominant in developing countries where it causes 26% of all avoidable deaths and it is expected that during the next ten years, the number of TB cases will still increase substantially.

This is partly due to the interaction between TB and the human immunodeficiency virus (HIV) epidemic; by the end of the century, TB is likely to be the leading cause of death among HIV-positive persons. Furthermore, due to poorly managed TB control programs, multidrug-resistant (MDR) TB is increasing all over the world. Some of MDR-TB is incurable and several reports mention cases of HIV-positive patients who deceased as a consequence of infection with MDR-TB within a couple of weeks.

What needs to be done to prevent further increase of TB and to revert this situation? Control of TB can be focused from various angles: (i) by preventing infection of the uninfected; (ii) by improving the detection of the infected; (iii) by better prevention of development of the disease among the infected; (iv) by improving diagnosis of active TB, and (v) by improving its treatment. One of the most efficient ways to fight TB is by stopping it at the source: uncontrolled spread of TB is caused by sick and infectious TB patients; properly treated patients do not disseminate bacilli. Therefore, WHO’s strategy to prevent further worsening of TB is through proper treatment of patients, basically through directly-observed short-course treatment by implementation of cost-effective TB control strategies. One of the major priorities for the success of a con-
trol program should therefore be identification and treatment of acid-fast bacillus smear positive cases and this can be achieved better by understanding the factors that foster transmission of Mycobacterium tuberculosis. Transmission of TB is influenced by a large number of risk factors and transmission dynamics are therefore different in distinct geographic regions. Furthermore, due to the increased mobility of the world population, globalization of TB has been intensified and caused an even more complex pattern of TB spreading and epidemiology. Because of political, social and ethical reasons, it is difficult to prevent the migration of infected individuals and an optimal control strategy has therefore to be developed, taking into consideration the social and biologic factors that foster TB transmission.

Traditional epidemiology continues to be the cornerstone for the understanding of TB spreading but the dynamics of the latter will only be better understood upon study of the spreading of individual strains of M. tuberculosis within a certain population. Until recently, detection of person to person transfer of M. tuberculosis within a certain population. Until recently, detection of person to person transfer of M. tuberculosis clones was a difficult, if not impossible task; only subgroups of strains could be identified through phage typing and occurrence of some special drug-resistance profiles. Only upon the recent development of molecular methods based on recognition of sequences of deoxyribonucleic acids (DNA) and their rearrangements, it has been possible to fingerprint single strains of M. tuberculosis. This has resulted in a better detection of outbreaks of particular MDR strains of M. tuberculosis and one can now prove the spillover of these strains from high-risk groups to the general population. Also, TB-spreading in particular geographic regions can now be evaluated and the construction of international networks on DNA fingerprinting of M. tuberculosis strains should allow epidemiology of TB on a global scale, hopefully improving the combat of the “Captain of all of these Men of Death”.

**SOME GENERAL FEATURES OF TB**

TB is a bacterial disease caused mostly by infection with M. tuberculosis, an organism belonging to the M. tuberculosis-complex which includes M. tuberculosis, M. bovis, M. bovis-BCG, M. microti and M. africanum. M. tuberculosis is a facultative intracellular parasite, generally invading macrophages, and although the disease affects mostly the lung, virtually every tissue or organ can be infected. Mycobacteria can be detected in clinical samples after acid-fast staining and microscopic examination. Tuberculosis is acquired through airborne transmission of droplet nuclei and risk of infection increases with their concentration and with time of exposure (Bass et al. 1990). Once an individual has been infected, he remains infected for a long time, possibly progressing to active disease, sometimes years after the initial infection, in about 10% of the infected cases when no immunosuppressive disorders are present. Conditions such as physical and emotional stress or immunosuppression upon HIV-infection increase the chance to develop active TB. Upon infection with M. tuberculosis, most individuals develop some degree of delayed-type hypersensitivity to tuberculin, providing as such a measure for infection status. However, this measure varies with age and ethnic and geographic origin of the population under study, parameters that also seems to determine the efficiency of protection of BCG vaccines against TB (Fine 1989). Age also influences the risk for developing progressive disease after primary infection, being highest among young children, while TB in adults generally appears many years after infection. Development of progressive disease can be due to reactivating a latent form of the disease or to reinfection; the attribution of each of the two mechanisms depends on the risk of infection within the community and the immune status of the individual (Sutherland 1976).

Diagnosis of TB can sometimes be a problem: clinical specimens are not always easily obtained; culture is slow; microscopic analysis and serological tests are insensitive or lack specificity and recent techniques based on DNA amplification are still fraught with technological problems. Therefore, diagnosis is mostly based on clinical examination and without receiving a proper treatment, 50% of the patients die within five years of diagnosis (Grzybowski et al. 1978). Upon the discovery by Koch that TB was caused by infection with a bacillus, some advances in therapy were made but the major breakthrough of TB control came in 1944, with the use of streptomycin, later followed by isoniazid and rifamicins, shortening chemotherapeutic treatment significantly. Several treatment regimens for drug-susceptible TB yield cure rates of more than 95% (Sbarbaro 1995) but due to the emergence of MDR-strains, we are now confronted with a new era where epidemics of untreatable TB are encountered. Prevention of TB is a better control measure than its treatment and improvement of socio-economic conditions is considered one of the most effective preventive measures. Vaccination is also considered but due to the variability of protection by the BCG-vaccine, an alternative preventive immunotherapeutic agent is needed (Grange & Stanford 1994). Prophylaxis is an alternative but because of problems with compliance and toxicity, this can only be applied in quite specific situations such as for HIV-positives in an endemic region for TB (Israel 1993).
CURRENT SITUATION OF TB

Worldwide - Tuberculosis is the world’s major cause of death associated with a single infectious disease in adults. It has been estimated that approximately 1.7 billion people were infected with \textit{M. tuberculosis} until 1990, with the great majority of infected persons residing in developing countries (for data on TB infection and disease see Murray et al. 1990, Kochi 1991, Dolin et al. 1994). The estimated number of TB cases and deaths in different geographic areas in 1990 is presented in Table I. That year, about 7.5 million cases of TB occurred all over the world, about 95% in third world countries. Due to the improvement of the standard of living in industrialized countries and to the introduction of chemotherapy, there has been a gradual decline in the number of TB cases during the twentieth century. The rate of this decline varies in different parts of the world and strongly depends on the economic situation in a country. In some countries, the rate of decline has reversed (Table II). Surprisingly, some of the countries known to have very high standards of living such as Denmark and Switzerland now present an increase in the number of TB cases (Bloom 1992).

In developing countries - The death rate in Europe and USA has decreased from 300 (around 1850) to 1-2 per 100 000, while that of developing countries has declined much less: from 100 to 50 per 100 000 (Kochi 1994). Due to the less effective application of chemotherapy and the large increase of the population, the annual number of TB deaths in developing countries has reached 2.7 million and there appears to be no tendency for a significant decrease of the disease. Lack of reliable statistics in many countries hampers a clear picture, but current estimates show that the largest number of cases occur in Asia while the highest incidence is seen in sub-Saharan Africa (Grzybowski 1991). Many factors contribute to the size of the TB problem but the most important ones are: (i) the stage of the epidemic reached within each country; (ii) the efficacy of the tuberculosis program and (iii) the prevalence of HIV infection. In many industrialized countries, the TB epidemic reached its heights at the end of the eigh-

### Table I

<table>
<thead>
<tr>
<th>Region</th>
<th>Tuberculosis incidence</th>
<th>Tuberculosis mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases</td>
<td>Rate(^b)</td>
</tr>
<tr>
<td>Southeastern Asia</td>
<td>3 106 000</td>
<td>237</td>
</tr>
<tr>
<td>Western Pacific(^c)</td>
<td>1 839 000</td>
<td>136</td>
</tr>
<tr>
<td>Africa</td>
<td>992 000</td>
<td>191</td>
</tr>
<tr>
<td>Eastern Mediterranean</td>
<td>641 000</td>
<td>165</td>
</tr>
<tr>
<td>Americas(^d)</td>
<td>569 000</td>
<td>127</td>
</tr>
<tr>
<td>Eastern Europe</td>
<td>194 000</td>
<td>47</td>
</tr>
<tr>
<td>Industrialized countries(^e)</td>
<td>196 000</td>
<td>23</td>
</tr>
<tr>
<td>All regions</td>
<td>7 537 000</td>
<td>143</td>
</tr>
</tbody>
</table>

\(a\): as published by Raviglione et al. 1995; \(b\): incidence and mortality rate per 100 000 population; \(c\): all countries of the region except Australia, Japan and New Zealand; \(d\): all countries of the region except Canada and United States; \(e\): Western Europe, Australia, Canada, Japan, New Zealand and United States.

### Table II

<table>
<thead>
<tr>
<th>Areas of the world resource</th>
<th>Annual risk of infection</th>
<th>Health availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>I Industrialized</td>
<td>0.1-0.01, &gt;10</td>
<td>Excellent</td>
</tr>
<tr>
<td>II Middle-income in Latin America, West Asia and North America</td>
<td>0.5-1.5, 5-10</td>
<td>Good</td>
</tr>
<tr>
<td>III Middle-income in East and Southeast Asia</td>
<td>1.0-2.5, &lt;5</td>
<td>Good</td>
</tr>
<tr>
<td>IV Sub-Saharan and Indian subcontinent</td>
<td>1.0-2.5, 0-3</td>
<td>Poor</td>
</tr>
</tbody>
</table>

\(a\): as published by Kochi et al. 1991.
teenth century while in many developing countries, this occurred much later. The stage of an epidemic is reflected by an age-specific incidence of morbidity and mortality and in many developing countries, TB is mainly a disease of young adults. Chemotherapy, if properly used, can reduce enormously TB in the community but because of the fragile structure of treatment programs in many countries, although saving many lives, TB cases are not completely cured and patients remain infectious for a much longer time. Another important consequence of poor treatment is development of drug resistance: in many developing countries, primary drug resistance appears to be 25% while secondary resistance sometimes reaches 75% (Grzybowski 1991). HIV infection is quite common in developing countries; it is widespread in Sub-Saharan Africa and rapidly increasing in Asia. HIV infection has a disastrous effect on TB taking it back to the heights of the epidemic (Schultzer et al. 1992).

In Brazil - The implementation of a national program against TB in Brazil is complex due to its large geographical area, the size of its population, its federal structure, its extreme variability in population density and economic development, high levels of internal migration, a large number of the population living in marginal conditions, and the reorganization and deterioration of the health sector. The situation is worst in the north and northeast of the country, regions that are characterized by a high poverty. Passive case finding is realized in about 4,000 Health Centers, coordinated by 23 State Reference Centers, 7 Regional Reference Centers and 1 National Reference Center, screening about 400,000 individuals yearly. According to the Ministry of Health, in 1992, 86,000 new cases of TB were reported, representing about 75% of the real total number of new cases, being 80% with a pulmonary form and with about 62% of these confirmed by microscopy. Curing rate is around 80% but every year more than 5,000 patients die. The national incidence rate is 66/100,000 but this number is much higher in Brazil’s largest cities: São Paulo, the country’s leader in number of cases, and Rio de Janeiro, with an incidence of 165/100,000 (SUS, Informe Epidemiológico, novembro 1992, Boletim de Pneumologia Sanitária, Número Especial, 1993). In these cities, the number of reported cases, the mortality rate and the number of patients abandoning treatment are increasing and some recent studies demonstrated the increase in primary resistance both in HIV-negative and -positive TB patients (Pinto et al. 1995, Fadinho et al. 1995) and transmission of MDR-TB to contacts (Kritski et al. 1995). In the large cities, there is a strong association between TB and HIV-seropositivity and TB is the second most important opportunistic disease among individuals infected with HIV. In 1988, 1.5% of TB patients in Rio de Janeiro were HIV-positive; this number has increased to 11% (Kritski et al. 1993).

**STRAIN TYPING BY NON-DNA BASED METHODS**

Members of the *M. tuberculosis*-complex are differentiated through evaluation of biochemical features and their different susceptibility to antibiotics (Collins et al. 1982). However, classical biochemical tests, DNA-DNA hybridization, restriction endonuclease cleavage analysis (Baess 1979, Imaeda 1985) and DNA sequencing (Frothingham et al. 1994) all suggest a very close relationship between these organisms. Although some authors consider members of the *M. tuberculosis*-complex as variants of the same species, their epidemiology is different and the great importance of *M. tuberculosis* as a pathogen has induced most investigators to consider *M. tuberculosis* as a separate species. Before molecular techniques were available, the most applied method for differentiation of strains of the TB-complex and of *M. tuberculosis* strains was phage typing (Bates & Fitzhugh 1967). Although this method is cumbersome and its differentiating potential is small due to the limited number of mycobacteriophage types, the technique has shown some utility for typing *M. tuberculosis* strains from distinct geographic areas (Jones Jr & Woody 1983), for typing of cultures of *M. tuberculosis* from outbreaks (Snider et al. 1984), for detection of laboratory cross-contamination (Jones Jr 1988) and for confirmation of TB transmission in a church (Dutt et al. 1994). Another way of typing isolates of *M. tuberculosis* is through testing of antibiotic sensitivity but, due to the limited number of resistance patterns, this method is only interesting for tracing the spread of strains with unusual patterns of drug resistance. Alternative methods for typing of *M. tuberculosis* isolates have been reported, each with considerable limitations. Whole-cell protein typing electrophoresis demonstrated informative and isolate-specific banding patterns (Miller & Want 1992) but replicates of the same strain did not always cluster; furthermore, the method suffers from lack of reproducibility and requires polyacrylamide-gel electrophoresis. Serotyping cannot differentiate between infection with different strains of *M. tuberculosis* (Grange & Laszlo 1990) and although some biochemical differences were observed between different isolates (Hoffner et al. 1993), reproducibility and limited strain variation is a problem.
MOLECULAR METHODS FOR FINGERPRINTING OF M. TUBERCULOSIS

The first reports on differentiation of strains of the *M. tuberculosis*-complex and strains of *M. tuberculosis* using nucleic acid-based technology were based on strain-specific differences and frequencies of certain DNA sequences in the chromosomal DNA, as demonstrated upon digestion of genomic DNA with specific restriction enzymes and analysis of the generated patterns after separation of the DNA fragments on agarose gel (Collins & Lisle 1984, Patel et al. 1986). The advantage of this kind of analysis is that it is technically simple: no hybridization step with defined probes is needed. However, interpretation of the results is a somehow intricate task because the large number of fragments generate a complex pattern and only a small number of different restriction fragment length polymorphism (RFLP) types was observed. Upon using less-cutting restriction enzymes that generate fragments with relative high molecular weight and separation of these fragments under special conditions in pulsed field gel electrophoresis (PFGE), one can simplify RFLP analysis. Because of the necessity to digest intact genomic DNA and because of the lysis-resistant cell wall of mycobacteria, culture conditions and DNA extraction procedures are quite demanding. Strains of *M. tuberculosis* and other mycobacteria have been differentiated with PFGE (Varnerot et al. 1992, Zhang et al. 1992) but the main limitation of the technique is that the small polymorphism characteristic for different strains will not always produce different patterns. DNA polymorphism can also be demonstrated through hybridization of digested nucleic acids with genomic DNA or cloned fragments (Fig. 1). Several strategies for differentiation of strains of *M. tuberculosis* through hybridization pattern analysis were described. One of these, using a total DNA probe and four-base restriction enzymes, could differentiate between a small number of strains of *M. tuberculosis* analyzed (Ross et al. 1991) but the use of the complete genome as a probe usually results in considerable background and affects the interpretation of the results. An alternative is the use of repetitive DNA sequences as probes. Some organisms have been typed by ribotyping, a technique using part of the coding sequence for rRNA as a probe; however mycobacteria are characterized by the pres-

![Fig. 1: detection of specific DNA molecules by gel-transfer hybridization.](image)
ence of only one or two copies of rDNA (Bercovier et al. 1986), which limits the polymorphism between strains belonging to this genus. At the end of the eighties, some groups used cloned repetitive DNA from \textit{M. tuberculosis} as a probe (Eisenach et al. 1986, 1988) and one of them could actually differentiate all individual strains of \textit{M. tuberculosis} analyzed (Zainuddin & Dale 1989). Very recent, the use of an oligonucleotide probe (GTG)$_5$, that recognizes polymorphic loci was shown to differentiate between several strains of \textit{M. tuberculosis} and could have some use when other, more specific probes are not informative (Wiid et al. 1994).

**CHARACTERIZATION OF REPETITIVE SEQUENCES WITHIN THE GENOME OF \textit{M. TUBERCULOSIS} AND THEIR USE FOR DETECTION OF STRAIN POLYMORPHISM**

Repetitive elements and insertion sequences are frequently used as target sequences for differentiation between bacterial strains. Several of these elements have been characterized in the TB-complex and some of them are considered as mobile genetic elements (Poulet & Cole 1995a, b, Dale 1995). Table III summarizes the repetitive sequences encountered in members of the TB-complex and in some additional mycobacterial species. For use of repetitive sequences in epidemiologic studies, the position of the sequence in the bacterial genome must be sufficiently stable but RFLP between different strains must be present.

The repeated DNA sequence with potential for epidemiological applications that was first characterized is an insertion sequence belonging to the enterobacterial IS3 family (McAdam et al. 1990). This sequence was detected by Southern analysis (Eisenach et al. 1988), shows homology with a plasmid isolated from \textit{M. fortuitum} (Zainuddin & Dale 1989) and, depending on the organism in which it was characterized is called IS6110 or IS986 (\textit{M. tuberculosis}) or IS987 (\textit{M. bovis} BCG; Eisenach et al. 1990, Hermans et al. 1990b, Thierry et al. 1990a, b). It is a 1 361 base pair long sequence that was detected exclusively in members of the TB-complex and differences of only a few nucleotides have been detected between the sequenced copies. The sequence is flanked by two 28-base pair repeats and has two open reading frames (ORF) that show homology with genes coding for putative transposases of other elements of the IS3 family, which are typical features of mobile elements. Although transposition of IS6110 has not yet been experimentally demonstrated in \textit{M. tuberculosis}, mobility of IS986 has been observed in \textit{M. smegmatis} (Fomukong & Dale 1993). The number of copies of IS6110 present in the genome is species- and strain-dependent (Table III); most strains of \textit{M. tuberculosis} carry between 8 to 15 copies in different positions of the genome, although several single-copy strains of \textit{M. tuberculosis} have been reported. It appears that the position of IS6110 in some BCG strains and in single-copy \textit{M. tuberculosis} strains is quite conserved and

**TABLE III**

<table>
<thead>
<tr>
<th>Repeated sequence</th>
<th>Host range</th>
<th>Copy number</th>
<th>Polymorphism</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS6110 (IS986, IS 987)</td>
<td>\textit{M. tuberculosis}</td>
<td>0 - 20</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>\textit{M. africanum}</td>
<td>0 - 20</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>\textit{M. bovis}</td>
<td>1 - 20</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>\textit{M. bovis}-BCG</td>
<td>1 - 2</td>
<td>No</td>
</tr>
<tr>
<td>IS1081</td>
<td>\textit{M. tuberculosis}</td>
<td>5 - 6</td>
<td>Small</td>
</tr>
<tr>
<td></td>
<td>\textit{M. africanum}</td>
<td>5 - 6</td>
<td>Small</td>
</tr>
<tr>
<td></td>
<td>\textit{M. bovis}</td>
<td>5 - 6</td>
<td>Small</td>
</tr>
<tr>
<td></td>
<td>\textit{M. bovis}-BCG</td>
<td>5 - 6</td>
<td>Small</td>
</tr>
<tr>
<td>DR cluster</td>
<td>\textit{M. tuberculosis}</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>\textit{M. africanum}</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>\textit{M. bovis}</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>\textit{M. bovis}-BCG</td>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>MPTR$^b$</td>
<td>Tuberculosis complex</td>
<td>+/- 80</td>
<td>Small</td>
</tr>
<tr>
<td>PGRS$^c$</td>
<td>Tuberculosis complex</td>
<td>26 - 30</td>
<td>Yes</td>
</tr>
</tbody>
</table>

$^a$: modified from Poulet et al. 1995b; $^b$: also described in \textit{M. kansasii} and \textit{M. gordonae}; $^c$: also described in \textit{M. kansasii}, \textit{M. gordonae}, \textit{M. marinum}, \textit{M. microti}, \textit{M. gastri} and \textit{M. szulgai}. 

located within a so-called “hot spot integration region”; multi-copy strains nearly always carry an IS6110 within this site and few strains carry none or two copies of IS6110 in this region (Hermans et al. 1991). This “hot spot” is also characterized by the presence of virtually identical direct repeats (DR) separated by spacer DNA which, on their turn, serve as epidemiologic markers.

Collins and Stephens (1991) identified a 1 324 bp insertion sequence called IS1081, typical for the TB-complex and with homology to IS256 from Staphylococcus aureus. Similar to IS6110, IS1081 is flanked by 15 bp inverted repeats and has an ORF with homology to a putative transposase. However, its mobility has never been demonstrated and its low transposition activity results in a very limited degree of polymorphism (Van Soolingen et al. 1992, 1993). Also, the copy-number of IS1081 is lower than that of IS6110, limiting its use in epidemiological studies to differentiation of M. bovis-BCG from the other members of the TB-complex (Van Soolingen et al. 1992).

Several short repeated sequences have been identified in the genome of members of the TB-complex. The hot spot of integration of IS6110 contains a variable number DR of 36 bp, separated by unique spacer sequences of 35 to 41 bp. Both the number of DR (ranging from 10-50 copies) and the presence of determined spacer sequences varies from strain to strain, allowing the development of several fingerprinting methods (Van Soolingen et al. 1995). Strain typing on the basis of variations in the DR cluster has been performed through Southern analysis (Van Soolingen et al. 1993, Sahadevan et al. 1995) and two different PCR-based methods based on variations within the DR cluster have been developed.

Another repetitive element called “major polymorphic tandem repeat” (MPTR) was characterized as a nonperfect repeat of 10 bp separated by 5 bp spacers. Although a copy number of around 80 was initially estimated (Hermans et al. 1992), recent data indicate the presence of less copies (Poulet & Cole 1994). The presence of MPTR, in contrast to the aforementioned repetitive elements, is not restricted to organisms belonging to the TB-complex (Table III) and it was postulated that these elements could be involved in genetic rearrangements (Hermans et al. 1990a). The MPTR demonstrated limited RFLP but a PCR-based assay, detecting the variability in the distance between IS6110 and MPTR seems to be more promising. Other mycobacteria demonstrate larger MPTR-based RFLP and the use of this sequence as an epidemiologic marker for some pathogenic species has been suggested (Hermans et al. 1992).

The most abundant repetitive element in the TB-complex is a polymorphic GC-rich repetitive sequence (PGRS) that appears to have some similarity with the MPTR on the basis of its sequence and host range. Several copies of this element have been cloned and sequenced independently (De Wit et al. 1990, Ross et al. 1992, Doran et al. 1993, Poulet & Cole 1994) and PGRS consists of many tandem repeats of a 9 bp GC-rich consensus sequence. The repetitive sequence is present in at least 26 sites of the M. tuberculosis chromosome (Poulet & Cole 1995a) and was detected in some other mycobacterial species not belonging to the TB-complex (Table III). Contrary to the MPTR, PGRS is quite polymorphic and has been used in some epidemiologic studies of both human and bovine TB (Ross et al. 1992, Doran et al. 1993, Cousins et al. 1993, Dwyer et al. 1993, Van Soolingen et al. 1993). Taking advantage of the PGRS polymorphism, some PCR-based typing methods use this repetitive element as one of the target sequences.

Recently, a GC-rich repetitive sequence that could eventually be used as a target for RFLP-based fingerprinting was characterized in M. tuberculosis and some other mycobacteria. However, preliminary hybridization results on a small number of M. tuberculosis isolates indicate that the position of the sequence within the genome of M. tuberculosis is probably too stable for its use as a marker (Verma et al. 1995).

**USE OF THE INSERTION SEQUENCE IS6110 AS A TOOL FOR TYPING OF M. TUBERCULOSIS: DEVELOPMENT OF A STANDARDIZED PROTOCOL**

**IS6110: occurrence, characteristics and some limitations** - Even before IS6110 was sequenced and completely characterized, its value for epidemiological studies of TB had already been suggested since unrelated strains of M. tuberculosis showed different RFLP patterns after hybridization with cloned fragments containing the insertion sequence (Eisenach et al. 1988, Zainuddin & Dale 1989), an observation that was later confirmed by many other groups. Initial studies on stability, polymorphism and copy number of IS6110 demonstrated the important epidemiological value of this system. Hermans et al. (1990b) showed that in vivo passage of M. tuberculosis strains in guinea pigs during two months did not change the DNA pattern of the parasite and the same group demonstrated that neither long-term serial passage in culture or in macrophages, nor selection of drug-resistance in vitro or in vivo resulted in change of the RFLP pattern (Van Soolingen et al. 1991). However, both reports showed that IS6110 is suffi-
ciently mobile to differentiate between epidemiologically unrelated strains; yet strains from an outbreak or from household contacts show identical fingerprints. Also, isolation of several isolates from the same individual over a period of one or several years yielded the same DNA pattern (Van Soolingen et al. 1991, Otal et al. 1991, Cave et al. 1994), and the RFLP pattern of strains isolated from patients before and after development of resistance in vivo (Godfrey-Faussett et al. 1993) also remained the same. Several studies showed the presence of 5 to 15 copies per genome using strains from African and European origin but the number of copies could be related with their geographic origin, as suggested by the higher copy number found in isolates from Hong Kong (Das et al. 1993). Van Soolingen et al. (1993) reported that many strains from the Far East contain a single IS6110 element as was also the case for about 30% of the strains derived from India; in another study, 40% of the strains from Madras were reported to have either no or a single copy of IS6110 (Das et al. 1995). Within a total of 41 strains from patients of Vietnamese origin, five strains had a single IS6110 copy while four strains did not contain the sequence (Yuen et al. 1993). Several single-copy strains were also characterized from Malaysia, Tanzania and Oman (Fomukong et al. 1994). Interestingly, many of the low copy number strains isolated in Tanzania (Yang et al. 1995b) and Denmark (Yang et al. 1995a), were obtained from patients originating from Asia. As in most of the cases where only one IS6110 copy is present, the insertion sequence is localized on a 1.5-kb PvuII fragment, and in this case, many of the epidemiologically unrelated M. tuberculosis strains give the same RFLP pattern. It was also observed that some unrelated M. tuberculosis strains carrying two IS6110 elements at identically sized PvuII fragments could be distinguished by other typing methods, demonstrating that additional typing techniques have to be considered when dealing with low IS6110 copy-number strains (Van Soolingen et al. 1993).

A standardized protocol for RFLP typing of M. tuberculosis strains - In the last couple of years, several molecular methods for typing M. tuberculosis strains have been developed. Although using the same target sequence in their hybridization system, some laboratories used different restriction enzymes or different parts of a specific sequence as a probe to generate RFLP patterns, making it impossible to compare fingerprints from different laboratories. Even when adopting the same methodology, variations in experimental conditions and in laboratory practices generated RFLP-patterns that are difficult to reproduce. Therefore, a standardized method for RFLP typing of M. tuberculosis strains has been developed (Van Embden et al. 1993a) consisting basically of extraction and purification of genomic DNA of M. tuberculosis from a well-grown culture; digestion of nucleic acids with PvuII; separation of DNA fragments under well-defined conditions on agarose gel, transfer of the DNA to a nylon membrane and hybridization with a fragment of the IS6110 sequence generated by PCR (Fig. 1). Importantly, a specific amount of internal molecular weight marker is mixed with each digested DNA sample and after hybridization with the IS6110 probe, a second hybridization with internal marker is performed. Besides these internal markers, an external marker consisting of PvuII-digested DNA from a reference M. tuberculosis strain (Mt14323) is included in each gel. This method has several advantages since many parameters influencing the quantity and position of the bands of the RFLP pattern can be compensated for; moreover, IS6110 has an internal PvuII site and the probe hybridizes with one side only, generating bands with comparable intensity, rendering analysis more simple.

Computer analysis - In cases of TB outbreaks, generally a relative small number of DNA patterns need to be compared and analysis of strain relatedness can be done visually without the need to consider internal markers. However, more and more DNA patterns are being generated, using the same approach, in different laboratories all over the world and, depending on the objectives of the study, the amount of strains that have to be compared varies considerably. For comparison of a large number of strains, computer analysis allowing normalization of banding patterns is a necessity. One of the earliest examples of application of computer-assisted DNA pattern recognition was for assessment of strain relatedness of clinical isolates of Candida albicans. Both band position and -intensity were considered, allowing the calculation of strain similarity values and grouping of related strains through the use of dendograms after pair-wise comparison of patterns (Schmid et al. 1990, Fig. 2). Applied to mycobacterial typing, this kind of analysis was further elaborated by the inclusion of a two-band internal size standard which allowed the construction of a library of normalized DNA-patterns of known mycobacterial samples and identification of a limited number of unknown samples, including M. tuberculosis, through the creation of a probability matrix (Plikaytis et al. 1992). Due to the characteristics of the IS6110-RFLP method, differences in intensities of bands do not need to be considered; furthermore, when a large number of bands are
present, similarity coefficients become more reliable (Godfrey-Faussett et al. 1993). Recently, software packages such as GelCompar (Applied Maths) and BioImage (Millipore Systems) have been constructed for the storage and analysis of a large amount of normalized DNA patterns for visualization of strain relation and comparison of a certain DNA pattern with a database.

APPLICATIONS OF MOLECULAR STRAIN TYPING AS AN ADDITIONAL TOOL FOR EPIDEMIOLOGY OF TUBERCULOSIS: A REVIEW OF THE LITERATURE

Outbreaks and detection of index cases - By means of traditional identification methods such as tuberculin skin test conversion of contacts with a suspected index case and their follow up, spread of new infections with TB in institutional settings was well recognized before the use of fingerprinting (King & Geis 1977, Stead et al. 1985, Nardell et al. 1986). This methodology has specific limitations and the use of molecular strain identification furnishes more convincing evidence for whether a single strain is being transmitted in a certain setting. Although identity of fingerprints from two strains does not prove that one individual has passed the strain to the other, combination of traditional and molecular investigation can lead to the index case. The first evaluation of IS6110-RFLP for epidemiological purposes was published by Hermans et al. (1990b), comparing 15 strains from various regional hospitals and laboratories. Nine strains were found to have the same fingerprint and all originated from patients that had been treated by the same physician. Pioneer work on the use of fingerprinting for tracing outbreaks detected an epidemic among homeless people in Amsterdam, identified two persons contaminated during bronchoscopy on the same day in the same clinic, demonstrated that a woman contaminated her husband and showed transmission within certain families in Czechoslovakia (Van Soolingen et al. 1991), identified sources of infection in a pub and a discotheque (Van Embden et al. 1993b) and demonstrated infection by a neighbor (Godfrey-Faussett et al. 1992a). The importance of transmission of TB within shelters for the homeless was also demonstrated in a study with TB patients from Australia (Dwyer et al. 1993) and by means of contact investigation by IS6110-RFLP, it was shown that TB was actively transmitted in a neighborhood bar in Minneapolis, where an index case infected 41 of his contacts (Kline et al. 1995). These studies clearly demonstrate the value of RFLP analysis in situations where traditional contact tracing alone would not be able to detect clusters of TB; it also shows that single, highly infectious cases can have a severe influence on TB transmission within a community and, as a consequence, on TB programs. Tuberculosis has been known to be endemic in correctional facilities for many years (Stead 1978) and active transmission of TB in a New York jail has been confirmed by phage typing and RFLP-analysis (Pelletier et al. 1993). Most outbreaks of TB have been observed in hospitals and patient care settings and many of these involve recent transmission of TB to patients with immunodepression, as demonstrated in a renal transplant unit (Jereb et al. 1993), in housing facilities or care centers for HIV-infected persons (Daley et al. 1992, Kent et al. 1994) and in hospital facilities (Edlin et al. 1992, Coronado et al. 1993, Beck-Sagué et al. 1992, Wenger et al. 1995). In 1990 through 1992, the Centers for Disease Control (CDC) investigated seven outbreaks of MDR-TB involving over 200 persons with most cases occurring among persons with HIV infection (CDC unpublished data) and one of these outbreaks involved a M. tuberculosis strain resistant to seven drugs causing extremely high mortality rates. Not only HIV-infected patients but also health care workers such as nurses, house keepers and labora-
tory workers (HIV-infected or not) developed MDR-TB during an outbreak (Jereb et al. 1995). Using an alternative typing technique by PFGE, it was shown that even upon short exposure to an index case, high rates of tuberculosis infection and disease in health care workers can occur (Griffith et al. 1995).

**Community transmission** - By the use of fingerprinting, one can nowadays better identify risk factors for transmission of TB within a certain community. A study using IS610-RFLP recently showed that extensive transmission of TB is taking place in Europe in the same social settings as those reported in the USA (Genewein et al. 1993); more than one quarter of the patients in Berne (Switzerland) were linked to other patients and mainly associated with outbreaks within families or with drug addicts and homeless persons. Importantly, transmission was not limited to these high risk groups as there was a “spill over” to the general population. Other studies on transmission of TB in large urban settings through cluster analysis (Alland et al. 1994, Small et al. 1994, Friedman et al. 1995a, Sepkowitz et al. 1995) demonstrated that besides drug resistance, young age, race, social status and being health care worker are risk factors.

**Influence of HIV on TB transmission** - Infection with HIV and development of AIDS has a serious impact on the global epidemiology of TB, due to the increase of the risk for developing TB. Impaired immunity allows latent *M. tuberculosis* to proliferate and evolve to active disease; also, whether or not previously infected with TB, HIV-infected persons have less control on new infections. Although reactivation of TB was thought to be the major cause of disease among AIDS patients, few data were available to determine the relative contribution of reactivation and recent (re-)infection. During the last couple of years, molecular typing methods have demonstrated the important contribution of recent transmission of *M. tuberculosis* for rapid development of TB in HIV-seropositive individuals. First, as stated earlier, several outbreaks with drug susceptible or MDR-TB strains have been observed among HIV-infected patients. Also, typing of representative *M. tuberculosis* strains in community based studies has revealed that persons with HIV have a considerable higher risk of developing a recently transmitted TB (Alland et al. 1994, Tabet et al. 1994, Shafer et al. 1995, Sepkowitz et al. 1995). Comparing strains from HIV-seropositive patients with those from AIDS patients, increased probability of belonging to a recent transmission was found to be associated with advanced immunodepression (Small et al. 1994). It has been suggested that strains infecting AIDS patients could have some peculiar characteristics but a recent study done with TB patients from Tanzania, comparing fingerprints of *M. tuberculosis* strains isolated from HIV-seropositive and -seronegative cases showed no evidence for particular clones playing a dominant role in a HIV-positive population (Yang et al. 1995b).

**Laboratory cross-contamination** - Laboratory cross-contamination is suspected when cultures yield positive for *M. tuberculosis* in the absence of clinical signs, and this can be the result of specimen mix-up, of contamination of specimens or reagents with environmental mycobacteria or because of transfer from bacteria from one sample to another. Until recently, cross-contamination in the laboratory was demonstrated on the basis of antimicrobial susceptibility patterns, by some special biochemical characteristics and by phage typing, all of them typing methods with little differentiating power (Jones Jr 1988, Smith & Vance 1991). Thanks to molecular typing, it has been possible to identify cross-contamination and the strain-typing results were used in decisions regarding patient care and for identifying problematic laboratory procedures (Small et al. 1993a, b).

**Laboratory acquired infection** - Laboratory workers have a higher risk for developing TB than the general population and at least two clear cases of laboratory contamination have been supported by the use of RFLP-IS6110. In the first study, it was shown that a laboratory worker got infected with the reference strain “Erdman” with which he had been working (Mazurek et al. 1991). A second study revealed infection of a technician with a clinical isolate of *M. tuberculosis* during preparation of inocula for identification after scratching his hand on the edge of a safety cabinet (Peerbooms et al. 1995). These kind of studies allow a better knowledge on laboratory safety conditions and a better intervention in case of an accident.

**Mixed infections** - Although it is generally believed that patients are only infected with a single strain of *M. tuberculosis* at a certain time, some data on drug resistance testing suggest mixed-strain infection and in these cases, strains isolated from different sites of the body or clonal cultures can be analyzed. Godfrey-Faussett et al. (1992b) showed a high proportion of paired isolates from individual patients suggesting that most patients have an unique isolate of *M. tuberculosis* at any one time. However, some HIV-infected patients showed paired isolated with unrelated DNA patterns, demonstrating infection with multiple strains during active TB. Mixed-strain infection was also demonstrated in a patient from whom a drug sensitive and an MDR-strain with completely different
RFLP-patterns was isolated, after exclusion of acquired resistance (Theisen et al. 1995). These kinds of studies suggest that repeated resistance testing in patients with delayed response to therapy is of value.

Influence of drug resistance on fingerprint patterns - During the last couple of years, many reports on outbreaks of MDR-TB (caused by strains resistant to at least rifampicin and isoniazide) have been reported in the United States. Two important questions arise when dealing with strain typing and drug resistance: (i) does a change in drug resistance induce a change in the isolate's fingerprint? and (ii) are strains with specific drug resistance patterns correlated with particular fingerprint patterns? To answer the first question, one has to consider the stability of IS6110-generated DNA patterns in time and it has been shown in various studies that RFLP-patterns do not change rapidly in vivo which means that no major DNA rearrangements occur during a 2- to 3-year period (Otal et al. 1991). No changes in fingerprint were found in a large number of sequential isolates from the same patients before and after acquired antibiotic resistance (Godfrey-Faussett et al. 1993), suggesting that occurrence of mutations causing resistance instead of a selection of a resistant sub-population. Reinfection is the most probable mechanism underlying the fingerprint pattern change during development of drug resistance, as suggested by study of TB-patients in Hong Kong (Yuen et al. 1995) and of HIV-infected patients (Small et al. 1993b).

In relation to the second question, many studies on fingerprinting of a large number of M. tuberculosis strains from specific populations looked for a correlation between strain clustering and drug resistance profile, so far without any evidence that a certain resistance is associated with a particular fingerprint type (e.g. Thierry et al. 1993). If any clustering is found among drug-resistant and MDR strains, it is because of the higher chance of these strains having been transmitted recently (Shafer et al. 1995). However, it is still possible that drug resistant strains have some peculiar behavior that influences their chance for clustering.

Population/geographic studies on transmission within and between countries - In an early report on the existence of region-specific strain relationships, it was suggested that this phenomenon is more frequent in high prevalence African countries than in countries where prevalence is low such as The Netherlands (Van Soolingen et al. 1991). Godfrey-Faussett et al. (1992a, b), demonstrated separate clustering between strains isolated from English TB patients from the Midlands and from the Indian subcontinent (1992a) and between strains from patients from Malawi and Kenya (1992b), confirming sufficient heterogeneity of strains from high incidence countries to allow transmission studies with molecular typing. Since then, a limited number of studies on RFLP analysis of strains from specific geographic regions from several countries have been published, including countries with high, medium and low prevalence in TB. Takahashi et al. (1993) analyzed strains isolated in 18 Japanese sanatoria and observed a certain similarity in their banding patterns, possibly associated with a relative high prevalence of TB in Japan. Although his samples seemed to be chosen randomly, strains were included from several small outbreaks which could bias towards a higher degree of strain similarity. In an effort to identify the importance of single cases for the spread of TB, no clustering was found among strains isolated from 31 patients in Austria, a low-incidence country, unless those derived from a married couple (Vogetseder et al. 1994). A large scale study was performed on 201 strains from Tunesia, a country in which TB is endemic, where apparently unrelated strains were collected from four district hospitals in the northern part of the country (Chevrel-Dellagi et al. 1993). This study indicates that in some regions, microepidemics still occur while population heterogeneity in the other regions could bias to strain difference. Hermans et al. (1995) observed that 62% of the Tunisian isolates belonged to three genetically related groupings and they explained these regional differences by the fact that transmission rates of TB in Tunesia differ greatly from province to province. The influence of geographic isolation on DNA polymorphism of M. tuberculosis strains was also demonstrated with strains from Greenland (Yang et al. 1994, 1995a). All 1992 M. tuberculosis isolates from that country were typed, showing a high degree of region-specific similarity and leading the authors to suggest that geographical separation may lead to evolution of distinct bacterial clones. When these strains were compared with strains collected in Denmark during the same period, clusters in Danish strains was traced to immigrants from Greenland, to Danish citizens who had contacts with immigrant or had travel or working activities in Greenland or to certain risk groups in the capital, demonstrating transmission of TB between the two countries. Upon analysis of a large number of strains from the capital of Tanzania, once more strain clustering was related with the geographic origin of the isolates. Because Dar es Salaam is a large city and a trade center with a large number of immigrants, the degree of polymorphism of strains isolated from that city was comparable to that of cities such as New York and San Francisco (Yang et al. 1995b). A further study with strains...
from Tanzania showed that a certain cluster could only be found in patients from the Kilimanjaro region, while another was found in patients belonging to the Masai tribe (Gillespie et al. 1995). Isolates from the People’s Republic of China and Mongolia belonged to a distinct group of strains which share more than two-thirds of their bands and have a very particular spoligotyping pattern, the so-called “Beijing family” (Van Soolingen et al. 1995). The authors suggest that strains have diverged from a common ancestor and that “Beijing family” related strains could have spread more aggressively in the South-Asian region. Finally, considerable differences in strain clustering between a high- (Ethiopia), a medium- (Tunisia) and a low-incidence (The Netherlands) country was demonstrated (Hermans et al. 1995). In this study, both clustering of identical and of similar patterns was considered and it was shown that transmission rate is not necessarily directly correlated with homogeneity of the strain population; other factors such as migration, regional differences within a country and BCG-vaccination could all have an influence on population dynamics of M. tuberculo-
sis. It was once more confirmed that in high incidence countries, M. tuberculosis strains have a tendency to belong to a limited number of families. Also, some patterns specific for Ethiopia or Tunisia were encountered in The Netherlands, allowing tracing of TB across national borders or continents.

**EPIDEMIOLOGICAL STUDIES USING ADDITIONAL MOLECULAR MARKERS AND AMPLIFICATION BASED METHODS**

**Molecular markers different from IS6110**

Because a significant number of strains show a particularly low copy number or even lack IS6110, and unrelated low-copy number strains often generate the same fingerprint (see above), Van Soolingen et al. (1993) analyzed the differentiating power of other genetic markers on a large number of strains, including strains that showed the same or two band RFLP pattern after IS6110-RFLP. They showed that DR and PGRS-based fingerprinting can differentiate low-IS6110 copy number strains but that both IS6110 and PGRS are the best markers known for differentiating M. tuberculosis strains; IS6110 still has more discriminatory power when a large number of copies are present. The same conclusion was drawn from a study comparing both typing methods on strains derived from patients before and after developing drug resistance (Yuen et al. 1995) and in a study showing transmission of TB among the homeless in Australia (Dwyer et al. 1993). The DR region is also sufficiently polymorphic to allow differentiation of low IS6110 copy number strains while both IS1081 and MPTR are too stable to allow significant differentiation. The larger discriminative power of IS6110 as compared to DR has also been shown for strains from Denmark and again, IS1081 was found little discriminatory (Yang et al. 1994).

**Amplification based methods**

One of the major limitations of RFLP-based typing systems is the requirement of a well grown culture for DNA extraction. The time-lag between isolation of M. tuberculosis from a clinical sample and the growth of a mycobacterial culture is often too long, sometimes resulting in the use of non optimized treatment protocols and, as a consequence, spread of eventual MDR bacilli. Therefore, quicker typing techniques have been developed and most of these are based on enzymatic amplification of a determined DNA sequence of the M. tuberculosis genome, including often IS6110. One of the biggest advantages of PCR-based typing methods, although not always elaborated in the studies presenting these systems, is the possibility of typing M. tuberculosis strains directly in clinical samples, increasing substantially the speed of identification of the organism. Moreover, this kind of analysis can be fairly easily automated and could be used when only non-viable isolates are available or when isolates can not be restricted. Most of the systems described however lack reproducibility or have less discriminating power than IS6110-RFLP (Table IV). One of these systems is called “amplityping” and is based on the use of oligonucleotide primers, hybridizing with the ends of IS6110 and generating a PCR reaction directed away from the insertion sequence, generating amplification products through primer annealing with neighboring IS6110 and/or random priming (Ross & Dwyer 1993). Isolates that clustered with IS6110 also clustered by amplityping but the method is not suited for comparison of a large number of strains and should therefore preferentially be used for investigation of, for example, strains from a suspected outbreak. Also, the amplityping results demonstrated non-specific amplification and lack of reproducibility, which could only be resolved upon inclusion of a hybridization step, an event that strongly decreased the number of bands and as a consequence, the discriminatory power of the technique (Yuen et al. 1995). Another gene amplification method for typing M. tuberculosis strains is based on the detection of differences in the distance between IS6110 and MPTR through unilateral nested PCR and hybridization analysis. Once more, a correlation was found with the results obtained by traditional fingerprinting and although this method was also used directly on sputum samples, its main
drawback is the limited number and size of the generated PCR products, decreasing the information on strain relatedness in comparison with the standardized method (Plikaytis et al. 1993). Amplification of the spacer region between the genes coding for 16S and 23S rRNA and posterior digestion of the amplicon with restriction enzymes was also performed for differentiation of M. tuberculosis strains (Abed et al. 1995a). A better discrimination using this amplification system was obtained upon random amplified polymorphic DNA analysis (RAPD) of the amplified product (Abed et al. 1995b). Although RAPD performed with genomic DNA generates patterns containing too many bands for simple analysis, RAPD performed with the amplified 16S-23S spacer region generates patterns that can be analyzed easily and seem to have a high discriminatory power. Unfortunately, both the reproducibility and the final discriminative power of the RAPD-based method was found limited (Frothingam 1995, Glennon & Smith 1995). A recent study tested the use of a large number of arbitrary primers on total M. tuberculosis genomic DNA in the hope of finding a reproducible RAPD-based typing method and, interestingly, the most useful primers were homologous to a linker ligated to the genomic DNA digested with a restriction enzyme. In contrast to most PCR-based typing methods, mixed-linker polymerase chain reaction (ML-PCR) typing sometimes generates more bands than traditional fingerprinting and can be applied directly on smear-positive clinical specimens. The reproducibility and applicability of the method using it directly in clinical samples for screening of TB-patients still has to be proven but the PCR-based RFLP method was used on TB-cultures derived from renal transplant patients to demonstrate an outbreak of TB in a transplant unit (Jereb et al. 1993) and to demonstrate nosocomial transmission of M. tuberculosis in an outbreak among health care workers (Zaza et al. 1995). Another elegant method for strain typing of M. tuberculosis strains is based on detection of DNA polymorphism in the DR cluster and is called "direct variable repeat polymer chain reaction" (DVR-PCR, Groenen et al. 1993). The method is based on outward amplification of IS6110 and the DR, generating as such a strain-specific banding pattern upon hybridization with a DR-probe. One of the main advantages of the method is that the degree of DNA polymorphism is not reflected by size variation in banding patterns but rather by the presence or absence of certain bands belonging to a multimer family, simplifying comparison and analysis. The method has a good differentiating power when a limited number of strains is tested but the stability of the DR region is higher than that of IS6110 and several strains differentiated by RFLP analysis of Alul-digested chromosomal DNA hybridized with DR were found identical by DRV-PCR. Another method that seems to have a comparable differentiating power as traditional fingerprinting is

<table>
<thead>
<tr>
<th>Method</th>
<th>Target sequence</th>
<th>Polymorphism</th>
<th>No. of bands</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplityping</td>
<td>IS6110</td>
<td>Similar</td>
<td>High</td>
<td>Low reproducibility</td>
</tr>
<tr>
<td>IS6110-ampliprinting</td>
<td>IS6110/MPTR</td>
<td>Similar</td>
<td>Low</td>
<td>Small PCR products</td>
</tr>
<tr>
<td>RAPD analysis</td>
<td>16S-23S spacer</td>
<td>High</td>
<td>Low</td>
<td>Low reproducibility</td>
</tr>
<tr>
<td>Mixed-linker PCR</td>
<td>IS6110/linker</td>
<td>High</td>
<td>High</td>
<td>None</td>
</tr>
<tr>
<td>Direct variable repeat PCR</td>
<td>IS6110/DR</td>
<td>Similar</td>
<td>High</td>
<td>Complex PCR</td>
</tr>
<tr>
<td>Double repetitive element PCR</td>
<td>IS6110/PGRS</td>
<td>Similar</td>
<td>Low</td>
<td>None</td>
</tr>
<tr>
<td>Spoligotyping</td>
<td>DR/spacers</td>
<td>Lower</td>
<td>High</td>
<td>None</td>
</tr>
</tbody>
</table>

*a*: as compared to the standardized IS6110 fingerprinting method; *b*: major polymorphic tandem repeat; *c*: polymerase chain reaction; *d*: random amplified polymorphic DNA analysis; *e*: direct repeats; *f*: polymorphic GC-rich repetitive sequence.
“double-repetitive-element PCR” (DRE-PCR), based on outward amplification of IS6110 and PGRS, generating a banding polymorphism because of the differences in distance between these elements (Friedman et al. 1995b). The positive predictive value of this method as compared to the traditional system was found to be 96% and the DNA patterns obtained through DRE-PCR seem to be sufficiently stable to use the method for epidemiology. Finally, an amplification-based technique called spoligotyping is based on amplification of the DR-region and subsequent differential hybridization of the amplified products with membrane bound oligonucleotides, complementary to the variable spacer regions localized between the DRs (Van Soolingen et al. 1995). The presence of these spacer sequences is variable in different M. tuberculosis strains and is visualized by a spot on a fixed site of the membrane, which simplifies to a large extend visual comparison and analysis of a large number of M. tuberculosis strains, allowing data to be stored and compared with a simple software. The differentiating power of spoligotyping is less than traditional fingerprinting when high IS6110 copy number strains are analyzed; in the case of strains with few IS6110 elements, discrimination is better achieved with spoligotyping. This method can distinguish easily between M. tuberculosis and M. bovis, can be used with culture positive clinical samples and has been used for demonstration of the clonal nature of the so-called “Beijing family” from the People’s Republic of China and Mongolia (Van Soolingen et al. 1995).

MOLECULAR TYPING OF BRAZILIAN ISOLATES OF M. TUBERCULOSIS

We are currently typing strains of M. tuberculosis using the standardized IS6110-RFLP method. For this, DNA is extracted (Santos et al. 1992) from a Lowenstein-Jensen culture of tubercle bacilli, followed by analysis of nucleic acids by RFLP as recommended by a standardized protocol (Van Embden et al. 1993a). IS6110-specific banding patterns are generated and normalized by a computer program (GelCompar, Applied Maths, Kortrijk, Belgium) through superposition of IS6110 banding patterns with the internal markers. Fingerprints are then analyzed and stored in a database. An example of DNA fingerprints of M. tuberculosis strains isolated from patients from Rio de Janeiro is shown in Fig. 3A. Fingerprints from Brazil, just like from many parts of the world mostly contain between 5 and 15 bands and are highly polymorphic. For comparison, fingerprints generated after hybridization with part of IS1081 is shown in Fig. 3B, confirming the reduced discriminative power of this system. As a quality control for reproducibility of the RFLP method, a reference strain called Mt14323 is included in each gel. Upon normalization of the banding patterns by computer using the internal markers (Fig. 3C), the reproducibility of the fingerprinting can be determined through comparison of the fingerprints of the reference strains from different gels (Fig. 4). As an illustration, Mt14323 patterns from different gels with or without normalization were compared, demonstrating both intra-laboratory re-
producibility and necessity for normalization and use of internal markers. This made it possible to compare fingerprints generated in our laboratory with fingerprints from strains from Rio Grande do Sul that were analyzed in The Netherlands (Suffys et al. 1996) and with a fingerprint database generated under the EC Concerted Action on the Epidemiology of Tuberculosis (unpublished). Furthermore, our preliminary results indicate a higher rate of clustering in strains from AIDS patients from a hospital at the Oswaldo Cruz Institute (Rio de Janeiro), as compared with strains from TB patients without AIDS from Rio de Janeiro (Ivens de Araujo et al. 1995, Ivens de Araujo et al. manuscript in preparation). Although highly reproducible and of great use in specific epidemiologic studies, the standardized protocol is labor-intensive and the time lag between isolation of the strain and generation of its fingerprint is long. We therefore started using the PCR-based technique called spoligotyping and are currently comparing fingerprints generated with both techniques.

CONCLUSIONS

The WHO estimates that, if the effectiveness of TB control does not improve substantially, the number of TB cases will reach 10 million in the year 2000, due to a combination of demographic factors, population movements, the expanding HIV epidemic and increasing drug resistance (Kochi 1994). Some drastic control measures should be taken both to protect individuals from TB and to interrupt the cycle of transmission. Improvement of socio-economic conditions in developing countries should have a drastic effect in reducing the disease while a better case finding and treatment will reduce considerably the risk of transmission (Rodrigues & Smith 1990). Research priorities should be focused on improved epidemiological studies, on better prevention and diagnosis, on development of more efficient chemotherapeutic regimens, on better program design and on understanding of the interaction of HIV and TB (Murray et al. 1990).

Molecular strain typing can be used as an additional tool in epidemiological investigations in order to gain a better understanding of factors that influence TB transmission, for identification of risk factors of TB transmission in a community and for evaluation of regional control programs, permitting as such a rational design of more adequate control measures. Wenger et al. (1995) showed that transmission of MDR-TB can be stopped when CDC guidelines are fully implemented, and that molecular fingerprinting has been shown to be useful for decision making within control programs. The methodology used for molecular typing will strongly depend on the questions one needs to answer; a standardized technique should be used for comparison of strains between laboratories, regions, countries and continents and for the understanding of global migration of *M. tuberculosis* (Small 1995). Creation of national and international networks for fingerprinting of TB could stimulate the use and spreading of the “know how” of the technique. The creation of a European Community-supported network of laboratories from Europe is an example of such an effort and more than 6 000 fingerprints are now available for comparison (Herre Heersma, personal communication). An effort for the construction of a network of Latin-American and European laboratories is done under supervision of the Latin-American Network for TB (RELACTB), supported by the United Nations University. In New York, a proper network has been created (TBNetwork) with the construction of its own TB fingerprint database. If one needs more rapid identification of strain spreading, an amplification-based fingerprinting method can be used. Although several of these methods have been developed, their applicability still has to be proven by comparison with the standardized method on a
large scale; preliminary results with some of these methods seem promising and, hopefully, it will be possible in the near future to routinely type *M. tuberculosis* directly in clinical samples. Finally, fingerprinting could also help resolve questions on pathogenesis of *M. tuberculosis*, on the relation between drug- and MDR-resistance and virulence, on influence of BCG on TB, on interactions between *M. tuberculosis* and its host under specific circumstances and on the relation between population genetics and susceptibility to TB.

**REFERENCES**


Edlin BR, Tokars JJ, Grie MH, Crawford JT, Will-


Plaiykitis BD, Plaiykitis BB, Shinnick TM 1992. Computer-assisted pattern recognition model for the iden-


