MOLECULAR EPIDEMIOLOGY OF TUBERCULOSIS

Zhenhua Yang

Department of Epidemiology, University of Michigan, School of Public Health, 109 S. Observatory Street, Ann Arbor, MI 48109

TABLE OF CONTENTS

1. ABSTRACT

Recent development of molecular techniques in typing pathogenic organisms has changed the way of epidemiologic study of infectious disease. Here I review the recent achievements and challenges in molecular epidemiology of tuberculosis.

2. INTRODUCTION

Tuberculosis (TB) remains the leading infectious cause of global mortality. The emergence of drug resistant TB in recent years and its deadly synergy with HIV highlight the need for a better understanding of the dynamics of TB epidemiology (1-5). Recent advances in molecular techniques have enabled the development of a variety of genotyping methods for differentiation of clinical isolates of *Mycobacterium tuberculosis*, the causative agent of TB disease (6-19). The emerging field of molecular epidemiology, which is the integration of the molecular techniques to track specific strains of pathogen with conventional epidemiological approaches to understanding the distribution of disease in populations, has provided new insights regarding the epidemiology of TB (20-22).

3. GENOTYPING METHODS

During the last 10 to 15 years, a variety of genotyping methods that utilize the insertion sequence (IS) and the short repetitive sequence associated DNA polymorphisms in *M. tuberculosis* complex strains have been established (6-19). In principle, these methods can be categorized into two groups: 1) Southern blot hybridization-based RFLP analysis and 2) PCR-based typing. As discussed in this review, each genotyping method has its advantages and disadvantages, in terms of technical demand, turnaround time, discriminatory power, reproducibility, and complexity of patterns generated. Most researchers, therefore, usually use a combination of multiple methods to differentiate *M. tuberculosis* strains for the purpose of epidemiological investigation.

3.1. Genetic markers

M. tuberculosis complex strains include five genetically related subspecies: *M. tuberculosis*, *M. africanum*, *M. bovis*, *M. microti*, and *M. canetti* (23, 24). Studies have demonstrated a high degree of evolutionary conservation of *M. tuberculosis* complex strains (25-32). The characteristic features of *M. tuberculosis* include its slow growth, dormancy, complex cell envelop, intracellular pathogenesis, and genetic homogeneity (33). Despite this genetic homogeneity in the *M. tuberculosis* complex, however, the high degree of DNA polymorphism associated with repetitive DNA sequences, including ISs and short repetitive DNA sequences, provides the basis for molecular typing methods to differentiate these strains (8, 14, 18, 34-44).

3.1.1. Insertion sequences

Prior to the completion of genomic DNA sequencing of *M. tuberculosis* H37Rv, four ISs, IS*6110* (6, 35), *-1081* (36), *-1547* (42), and the IS-like element (39), had been identified in *M. tuberculosis* complex strains. Analysis of the complete genomic DNA sequence of strain H37Rv disclosed one to three copies of 25 unknown ISs, two prophages, and a novel type of repetitive sequence, the REP13E12 family (44, 45). Of all the insertion sequences found in *M. tuberculosis* complex strains to date, IS*6110* is the one that has been most extensively used in *M. tuberculosis* strain typing for epidemiological investigation because of its apparent mobility and its high copy number in most strains. In contrast, IS*1081* is almost invariably present in only five to seven copies per genome and shows limited DNA polymorphism (36, 46). IS1547 and the IS-like element are present in one or two copies per genome (39, 42, 44), thus limiting their value in discriminating between strains (42). The presence of more recently identified ISs was investigated in the five species of the *M. tuberculosis* complex, including 32 *M. tuberculosis* strains, and four of these (IS*1532*, *-1533*, *-1534*, and *-1561*') were absent from some of the strains tested (45). The usefulness of these new ISs in genotyping is limited by the low copy number of these elements.

3.1.2. Short repetitive DNA sequences

Besides IS elements, five types of short repetitive DNA associated with some degree of genetic diversity have been identified in *M. tuberculosis* complex strains. Three of these, the polymorphic GC-rich tandem repeat sequence (PGRS) (8, 41), a repeat of the triplet GTG (14), and the major polymorphic tandem repeat (MPTR) (38), are present in multiple genomic clusters composed of many non perfect repeats. The genome sequence of *M. tuberculosis* H37Rv shows that PGRSs and MPTRs are often part of the so-called Pro-Glu and Pro-Pro-Glu multigene families, respectively. It has been postulated that the DNA polymorphism associated with these repetitive elements may result in antigenic variation (44). In addition, six exact tandem repeat (ETR) loci containing tandem repeats of identical DNA sequences have been identified (40, 43). The remaining type of the short repetitive DNA found in *M. tuberculosis* complex strains is the direct repeat (DR) elements, which are 36 base pair (bp) long and present at a single genomic locus (37). The polymorphism associated with this locus is driven by the 35-41 bp unique DNA spacer sequences that intersperse the DRs. (18). Although all types of the short repetitive DNA have been used for the strain differentiation of *M. tuberculosis*, the PGRS RFLP analysis has been most widely used in molecular epidemiology studies.

3.2. Restriction fragment length polymorphism analysis
3.2.1. IS *6110* RFLP analysis

The IS*6110* RFLP analysis, also called IS*6110* DNA fingerprinting, is the most standardized and widely applied genotyping method for *M. tuberculosis* (12). This method is based on the variability of genomic location and the copy number of this 1,355 base pair (bp) element in *M. tuberculosis* complex strains. Sequence analysis showed that IS*6110* is related to the IS*3* family of insertion elements, which was initially discovered in gram-negative bacteria. IS*6110* was initially identified in *M. tuberculosis* (35) and subsequently found to be distributed through the *M. tuberculosis* complex, with copy numbers ranging from 0 to 25 per strain (6, 47). The genomic restriction fragments of tubercle bacilli carrying IS*6110* are therefore highly polymorphic.

The IS*6110* RFLP patterns visualized by Southern blot of *M. tuberculosis* genomic DNA restricted with endonuclease *Pvu*II, using peroxidase-labeled IS*6110* as a probe, are relatively simple and amenable to computerized analyses and comparison. Recommendations for a standardized method of strain typing for *M. tuberculosis* were formulated in 1993 (12), and in virtually all published studies, investigators have complied with these recommendations. Thus, fingerprints generated in different laboratories theoretically can be compared, thus allowing the investigation of the prevalence of various types or genotypes in different regions and the trace of interregional transmission of *M. tuberculosis* (48-51). In the standard IS*6110* RFLP analysis, the probe used is a PCR product complementary to the sequence on the right side of the *Pvu*II site within the IS element, designated as IS*6110*-3' probe.

Numerous investigations of TB outbreaks have demonstrated that epidemiologically related isolates show identical IS*6110* RFLP patterns (52-54). Some *M. tuberculosis* clinical isolates, however, contain a low copy number (0-5) of IS*6110* and the variability in genomic location is rather limited among the low-copy strains. For these strains, the IS*6110* RFLP typing is therefore not discriminatory. Secondary typing with other genetic markers is required when *M. tuberculosis* strains contain no or few copies of IS*6110*, such as seen in a significant number of the *M. tuberculosis* isolates from Asia (55-58) and *M. bovis* strains from cattle (59). Furthermore, although in most studies isolates with identical IS*6110* fingerprints comprising > 5 hybridizing bands have been considered to be from the same strain, a recent study found that 53% (10 of 19) of the IS*6110*-3' identified interstate high-copy-number clusters in the CDC National Genotyping and Surveillance Network project during the time period between 1996 and 1997 were subdivided by one or more secondary typing methods, including the PGRS and IS*6110* left-hand probe (IS*6110*-5') secondary fingerprinting and the spoligotyping based on the unique DNA spacer sequences that intersperse the DRs (60). In contrast, clustering of the control group comprising isolates from a single state in the Untied States was virtually identical by all methods. These results demonstrate the importance of secondary typing for investigating the relatedness of clinical isolates of *M. tuberculosis* obtained from widely separated geographic regions. The study also suggests that a combination of IS*6110*-3' and IS*6110*-5' fingerprinting will increase the reliability of IS*6110* fingerprinting in studying the clonal relationship of *M. tuberculosis* isolates and that IS*6110*-5' fingerprinting might serve as a more appropriate primary typing method for a genotyping network that compares isolates from different geographic regions.

3.2.2. PGRS RFLP analysis

The secondary typing that has been used in most molecular epidemiology studies to date is the PGRS RFLP analysis using recombinant plasmid pTBN12 that carries an insert of PGRS as a probe (8, 17, 61-63). The PGRS RFLP analysis is therefore often referred to as pTBN12 fingerprinting and is based on the DNA polymorphism associated with the PGRS. The PGRS occurs at 61 loci of the genome of *M. tuberculosis* H37Rv, where each one has a common 9-bp consensus repeat (GCCGCCGTT) (44). In *M. tuberculosis* the number of PGRS and their distribution vary from strain to strain. The procedure of PGRS RFLP analysis includes digestion of genomic DNA with endonuclease *Alu*I, separation of the restriction fragments on 0.7 % agarose gel, and Southern blot hybridization with peroxidase-labeled pTBN12 probe (17, 61).

To validate the use of PGRS RFLP analysis in molecular epidemiological study of TB, Yang and colleagues fingerprinted *M. tuberculosis* isolates from 67 patients in five states in the United States and in Spain with both IS*6110* and pTBN12 (61). Epidemiological links among the 67 patients were evaluated by patient interview and/or review of medical records. The 67 isolates had five IS*6110* fingerprint patterns with two to five copies of IS*6110* and 18 PGRS RFLP patterns, of which 10 were shared by more than one isolate. Epidemiological links were consistently found among patients whose isolates had identical PGRS RFLP patterns, whereas no links were found among patients whose isolates had unique PGRS RFLP patterns. The study suggests that PGRS RFLP analysis is a useful tool to identify epidemiologically linked TB patients whose isolates have identical IS*6110* fingerprints containing fewer than six bands. Furthermore, in a population-based study, Burman and colleagues found that the use of secondary PGRS typing decreases clustering and improves the correlation between the transmission links predicated by genotyping and epidemiological investigation (63).

3.3. PCR-based typing

Although DNA fingerprinting using IS*6110* and PGRS as probes have proven to be useful tools for molecular epidemiological study of TB, they are technically demanding and time consuming because the methods require a sufficient amount of extracted DNA. To obtain the required amount of genomic DNA, one needs to culture the organisms before DNA extraction, and the minimum turnaround time for RFLP analysis is at least three to four weeks. These disadvantages promoted the development of less expensive and less laborious PCR-based genotyping methods, which include mixed-linker PCR fingerprinting (9), IS*6110* inverse PCR (19), IS*6110* ampliprinting (11), double repetitive element PCR (DRE-PCR) (15), spoligotyping (18), arbitrarily primed PCR (APPCR) (10), VNTR and MIRU-VNTR typing (43, 64). Reviewed below are those that have proven to be useful for genotyping of *M. tuberculosis* isolates, in terms of their reproducibility, specificity, and discriminatory power.

3.3.1. Mixed-linker PCR fingerprinting

Mixed-linker PCR fingerprinting is a rapid typing method based on specific amplification of genomic RFLP fragments containing IS*6110* (9). The procedure of mixed-linker PCR fingerprinting includes four steps. First, the *M. tuberculosis* genomic DNA is restricted with *Hha*I, then a synthesized double-stranded oligonucleotides linker that contains uracil instead of thymidine on one of its two strands is ligated to the ends of the restriction fragments, followed by a treatment with uracil N-glycosylase (UNG) to eliminate the uracil-containing oligonucleotides. Finally the target sequence is amplified using one primer specific for IS*6110* and a second primer complementary to a linker ligated to the *Hha*I restricted genomic DNA. The main advantage of mixed-linker PCR fingerprinting is its independence from mycobacterial growth, which is required for traditional IS*6110* fingerprinting. A mixed-linker PCR fingerprint can be obtained directly from a single colony of the primary isolate without further culture, indicating that single colony mixed-linker PCR fingerprinting may facilitate the investigation of the frequency of simultaneous infections with more than one strain of *M. tuberculosis*. Furthermore, this method can also be used to genotype non-viable tubercle bacilli, enabling outbreak investigation when the original isolates fail to grow due to their inappropriate storage.

3.3.2. Spoligotyping

Spoligotyping is a relatively simple, rapid subtyping method. This method is based on the DNA polymorphism present at one particular genomic region of *M. tuberculosis* complex strains (18). The DR region contains a variable number of direct repeat (DR) of 36 bp that are interspersed by non-repetitive DNA named spacers, each 35 to 41 bp long. Based on the knowledge of the DNA sequences of the spacers present in the DR locus of *M. tuberculosis* strain H37Rv and *M. bovis* BCG vaccine strain P3, 43 synthetic oligonucleotides representing each of the unique spacer sequences have been designed and applied in lines on a DNA membrane. In order to examine the presence of these 43 spacers in the DR region of a particular strain, the whole locus of that strain is amplified by PCR, using two inversely oriented primers complementary to the sequence of the DRs. By using such primers, DNA in between DRs next to each other and in between DRs more distantly positioned is amplified. The PCR products of multiple sizes are applied on the membrane, in rows that overlay and are perpendicular to the rows with the synthetic oligonucleotides. Because one of the PCR primers is labeled with a biotin label, through a streptavidine-peroxidase conjugate and a substrate, the hybridization on the synthetic oligonucleotides can be detected by chemilumminescence. It is observed that the order of "spacers" is about the same among different strains. However, insertions or deletions of spacers and DRs do occur. The spoligotyping pattern therefore reflects the presence or absence of the 43 known spacer sequences. Spoligotyping data can be entered as a word document allowing a rapid comparison of large number patterns in a word processor. For strains containing < 5 copies of IS*6110*, spoligotyping shows higher discriminating power than IS*6110* fingerprinting; it is, however, less discriminating for strains that have > 5 copies of IS*6110*.

3.3.3. VNTR and MIRU-VNTR typing

Variable number of tandem repeats (VNTR) typing of *M. tuberculosis* is a PCR-based typing method

introduced by Frothingham and Meeker-O'Connell (43). This method is based on the variability in the number of exact tandem repeat (ETR) at five loci of *M. tuberculosis* genome. Each locus contains a unique sequence 53 to 79 bp in length that is repeated exactly. The VNTR loci are amplified by PCR by using specific primers complementary to the flanking regions. The number of tandem repeat units is determined by estimating the size of the PCR product on agarose gels. The results are expressed as a five-digit allele profile in which each digit represented the number of copies at a particular locus. Analysis of multiple isolates of particular reference strains demonstrates that all the alleles are reproducible and stable (43).

MIRU-VNTR typing is a new PCR-based typing method introduced recently by Supply and colleagues (65). This method is based on the variable-number tandem repeats of mycobacterial interspersed repetitive units (MIRU-VNTR), including both ETR and MPTR. MIRUs are 40-100 bp DNA elements often found as tandem repeats and dispersed in intergenic regions of the *M. tuberculosis* complex genomes. The *M. tuberculosis* H37Rv genome contains 41 MIRU loci. After PCR and sequence analyses of these loci in 31 *M. tuberculosis* complex strains, 12 of them were found to display variations in tandem repeat copy numbers and, in most cases, sequence variations between repeat units as well. Of the 12 variable loci, only one was found to vary among genealogically distant BCG substrains, suggesting that these interspersed bacterial minisatellite-like structures evolve slowly in mycobacterial populations. In MIRU-VNTR typing, each isolate is typed based on the number of copies of repeated units at the 12 variable loci scattered throughout the genome. The number of repeated units can be determined by the size of the fragment produced by amplification of the entire locus. Recently, an automated MIRU-VNTR typing system was established by Supply and colleagues (66). This system combines analysis of multiplex PCRs on a fluorescence-based DNA analyzer with computerized automation of the genotyping. To fully exploit the portability of this typing system, a website was set up for the analysis of *M. tuberculosis* MIRU-VNTR genotypes via the Internet (66). MIRU-VNTRs can be used for analysis of the global genetic diversity of *M. tuberculosis* complex strains at different levels of evolutionary divergence.

3.4. Comparison of genotyping methods recipient

As described above, each typing method has its own advantages and disadvantages. The reproducibility, discriminatory power, and specificity of currently known typing methods for *M. tuberculosis* isolate was systematically compared through an international collaborative study involving 12 laboratories in five different countries in 1999 (67). In this study, 90 *M. tuberculosis* complex strains, originating from 38 countries, were tested in five RFLP typing methods and in seven PCR-based assays that target one or more repetitive DNA elements. The strain typing and the DNA fingerprint analyses were performed in the laboratories that were most experienced in the respective method. By including blinded duplicate samples, the intralaboratory reproducibility was

assessed. The specificity of the various methods was tested by inclusion of 10 non-*M. tuberculosis* complex strains.

The study found that the most highly reproducible techniques were the RFLP-based methods, mixed-linker PCR, VNTR typing, and spoligotyping. The reproducibility of these methods, ranging from 94 to 100 % (P > 0.1). The 90 strains were best discriminated by IS*6110* RFLP typing, yielding 84 different banding patterns. The mixed-linker PCR, PGRS RFLP, spoligotyping, and VNTR typing each generated 81, 70, 61, and 56 patterns respectively. In this study, IS*1081* and DR, and (GTG)$_5$ RFLP were also evaluated and found to have very limited discriminatory power. The specificity analysis shows that Southern blot–based RFLP analysis surpasses the PCR-based techniques. In this study, the methods that allowed subtyping of the two low-copy-number IS*6110* clusters of two and five *M. bovis* strains, in decreasing order of discrimination, were VNTR typing, PGRS RFLP or DR RFLP, spoligotyping, (GTG)$_5$ RFLP, and IS*1081* RFLP. Furthermore, traditional methods to reveal polymorphisms among bacterial populations were found not to be useful for *M. tuberculosis* complex strain typing, because of the unusually low structural gene variation among *M. tuberculosis* isolates, even when they originate from very diverse geographic regions (67). The conclusion drawn from the study was for epidemiological investigations, IS*6110* RFLP or mixed-linker PCR is the methods of choice for strain differentiation. When less strain discrimination is required, VNTR typing and spoligotyping are reproducible alternatives.

The greater discriminating power of PGRS RFLP analysis in typing low-copy-number isolates as compared with spoligotyping was also found in another study (68). In this study, a total of 88 isolates (100% of the isolates with fewer than six copies of IS*6110* isolated in Arkansas during 1996 and 1997) were included. Among the 88 isolates, 34 different IS*6110* patterns were observed, 10 of which were shared by more than one isolate, involving a total of 64 isolates. The 64 isolates were subdivided into 13 clusters (containing 37 isolates) and 27 unique isolates based on a combination of IS*6110* and PGRS RFLP analysis and into 11 clusters (containing 51 isolates) and 13 unique isolates based on a combination of IS*6110* fingerprinting and spoligotyping. Identical spoligotypes were found among isolates having different IS*6110* patterns, as well as among isolates showing different PGRS RFLP patterns. The clustering rate was 73, 58, and 42%, respectively, for IS*6110* fingerprinting alone, IS*6110* fingerprinting and spoligotyping combined, and IS*6110* and PGRS RFLP combined fingerprinting.

Comparison between IS*6110* fingerprinting and MIRU-VNTR typing on 44 isolates from Paris with no known epidemiological links found that MIRU-VNTR provided a resolution comparable to that of IS*6110* RFLP (65). Furthermore, MIRU-VNTR typing appears to surpass ETR-VNTR typing in regard to discriminatory power and reproducibility, most likely because more loci are compared in MIRU-VNTR typing. Recently, another study that included a set of 180 *M. tuberculosis* and *M. bovis*

isolates having low copy numbers of IS6110 found that MIRU-VNTR produced more distinct patterns (80 patterns) than did IS*6110* RFLP (58 patterns.), and spoligotyping identified 59 patterns. In this study, no single method defined all unique isolates, and the combination of all three typing methods generated 112 distinct patterns, identifying 90 unique isolates and 90 isolates in 22 clusters. The results indicate that typing with multiple methods is required to attain maximum specificity (69).

In regard to analysis of genotyping results, IS*6110* RFLP patterns are relatively easy to compare, because they generally show hybridizing bands of equal intensity, as most IS*6110*-containing restriction fragments carry an intact part of an IS*6110* copy of equal size. In contrast, RFLP typing using repetitive DNA as a probe results in hybridizing bands of variable intensity. Therefore, it has been a common practice to analyze such patterns by eye. Mixed-linker PCR, VNTR typing, and spoligotyping are the methods of choice for reproducible PCR-based typing of *M. tuberculosis*. Moreover, VNTR, MIRU-VNTR typing and spoligotyping have the advantage that their results can be fully expressed in a simple, digital format. Thus, inter- and intralaboratory comparisons should be equally reproducible.

4. MOLECULAR EPIDEMIOLOGICAL STUDIES

4.1. Epidemiological applications of molecular techniques

The disease burden of TB is influenced by the risk of an individual in the community being infected with tubercle bacilli in a given time period, the risk of disease following such infection, and the risk of disease occurring long after the original infection owing to the reactivation of latent infection (70). Epidemiological studies have sought to measure these risks and to identify factors that modify them, particularly those factors that can be changed through specific intervention measures. Prior to the development of molecular typing techniques for *M. tuberculosis* complex strains, much of our understanding of the dynamics of TB in population has been derived indirectly by inference from descriptive epidemiology data (71-73). Our ability to determine transmission links, to define the human and environmental factors associated with risks for acquiring infection with *M. tuberculosis*, and to answer the longstanding question regarding the relative contribution of disease due to the reactivation of latent infection and disease due to progression of currently acquired infection to the burden of TB in populations is limited by technology. The development of molecular typing techniques for *M. tuberculosis* has revolutionized the epidemiology of TB and offered new opportunities for understanding transmission dynamics and control of TB (20-22).

During the last decade, molecular epidemiological study of TB has evolved from traditional outbreak investigation (74-78), detection of epidemiologically unsuspected transmission of TB (79), confirmation of epidemiologically suspected transmission of *M. tuberculosis* (47, 52), identification and investigation of nosocomial or institutional transmission of *M.*

tuberculosis (74, 75, 78, 80), and identification of laboratory cross-contamination (81, 82) to systematic population-based studies that address specific epidemiological questions or test hypotheses, such as the risk for infection, accelerated progression, and community transmission; relative contribution of recent transmission/rapid progression and reactivation of remote infection to the disease burden in populations; and specificity and sensitivity of contact investigation (53, 83-90). Molecular epidemiology has helped to determine the extent of TB transmission between immigrants and non-immigrants in developed countries (84, 91, 92), and it has allowed us to monitor the spread of drug resistant TB and to study the etiology of MDR emergence (93-95). Yet, it has assisted to detect the occurrence of reinfection in both immunocompetent and immunocompromised patients (96-101). Furthermore, large-scale non-population-based investigations of genotypes of *M. tuberculosis* in different geographic regions have allowed the identification of strains of *M. tuberculosis* that have specific properties, such as infectivity, transmissibility, and virulence (32, 50, 102-104). The type of study has also led to the recognition of well-defined genotype families within the *M. tuberculosis* complex (48, 95). Strains belonging to these genotype families possibly share particular phenotypic properties, such as antigens and virulence factors, which may be expressed as distinct manifestations in the pathology and the epidemiology of TB.

4.2. Approach for systematic molecular epidemiology study

The premise on which the approach for systematic molecular epidemiology has been developed is that TB cases clustered by identical genotypes are epidemiologically related, or in other words, caused by recent transmitted infection while cases whose isolates show unique genotypes are due to the reactivation of remote infection. In these systematic molecular epidemiological studies, investigators have applied IS6110 typing and appropriate secondary typing methods to quantify the isolates in clusters versus isolates having unique genotypes, thereafter to quantify the extent of ongoing transmission of *M. tuberculosis* and to determine the relative contributions of recent and remote infection to the burden of TB disease in communities. In conjunction with molecular typing information, conventional epidemiological data are used to identify host-, environment-, and strain-specific risk factors for recent transmission of the disease (53, 62, 63, 83-90).

The approach that has been used for systematic molecular epidemiological study typically include the following steps: 1) enrolling a cohort of TB patients from a specific population at risk for the disease over an extended period of time; sampling strategies have ranged from complete ascertainment of all available cases in a hospital-(54, 105), city- (53, 83), or country- (84, 106, 107) to convenient sampling of a small proportion of the available cases (48, 79, 85, 108, 109); 2) collecting *M. tuberculosis* isolates from the study patients and genotyping the isolates obtained; 3) collecting conventional epidemiological data including demographic information, life-style variables, clinical characteristics, and contact information of the study

patients through routine patient data collection at the time of diagnosis according to the Centers for Disease Control and Prevention's "Report of a Verified Case of Tuberculosis", routine investigation of contacts to a case of pulmonary TB, and interview of clustered patients and their contacts by a research nurse using a standardized questionnaire; 4) categorizing study patients as either clustered or unique based on the genotypes of their isolates; and 5) assessing risk factors for recent TB transmission based on the covariates associated with clustered cases.

While there is no gold standard for cluster definition, in most epidemiological studies a cluster is defined as two or more isolates having identical IS6110 fingerprints with > 5 hybridizing bands, or similar IS6110 fingerprint patterns (with only one band difference) with > 5 hybridizing bands and an identical pTBN12 pattern, or identical IS6110 fingerprints with $\leq$ 5 hybridizing bands and identical PGRS patterns (17, 62, 63). In these systematic studies, the proportion of cases that cluster within the cohort is used to infer the relative proportions of clustered and unclustered cases in the community from which the cohort was drawn. Two different methods have been used to estimate the proportion of clustered cases. The first method, usually referred to as the "n" method, uses the number of all cases that fall into clusters as the estimator of clustered case (54). The second method, the "n minus one" method, assumes that one case per cluster is a case of reactivation TB and thus removes one case per cluster from the counts of "clustered" cases (110). Choosing "n" versus "n minus one" method is based on the specific epidemiological question being addressed; "n" method is used when the objective of the study is to estimate the number of people involved in transmission chains and "n minus one" method if the investigators want to ascertain the number of people with primary versus reactivated disease. (111).

4.3. Interpretation of molecular epidemiological studies

While the integration of modern molecular techniques into traditional epidemiology offers more and more opportunities to understand the epidemiology of TB, considerable caution should be exercised in conducting and interpreting these studies (110-114). Groups of clinical isolates may be identical for reasons other than recent transmission, depending, for example, on the stability of the marker and the number of strains in the community over time. Cases actually due to recent transmission may not be seen as clustered if they are new immigrants to the population or if not all cases in the population are included in the study, as in most studies where a convenient or random sample has been used. If cases that share identical genotype are not included in the study because the sample is too small, clustered cases will be misclassified as unique and the resulting proportion of clustered cases will be underestimated. This, in turn, results in underestimation of the extent of recent transmission and overestimation of the extent of reactivation TB, as well as biased estimation of the effects of risk factors for transmission.

Using stochastic simulation models based on real and hypothetical populations, Glynn and colleagues demonstrated the influence of incomplete sampling on the estimates of clustering obtained. They showed that as the sampling fraction increases, the proportion of isolates identified as clustered also increases and the variance of the estimated proportion of clustered case decreases; the biased estimate of proportion of clustered cases is a function of sampling fraction and the underlying cluster distribution (112). Recently, Murray reported an individual-based micro-simulation of TB transmission that demonstrated that multiple host-related factors contribute to wide variation in cluster distributions even when all strains of *M. tuberculosis* in the community are assumed to be equally transmissible. These host factors include interventions such as chemotherapy, vaccination and chemoprophylaxis, HIV prevalence, the age structure of the population, and the prevalence of latent TB infection (113). Using a Monte Carlo simulation model, Murray further assessed the magnitude of bias incurred by sampling strategies commonly used in the molecular epidemiology of TB and the impact of risk factors for recent TB transmission resulted from sampling bias. Depending on the underlying distribution of cluster sizes, the error involved in overestimating the proportion of unique TB isolates in a sample may be sizeable, even when up to 70% of the complete data is sampled. The odds ratios for risk factors for clustering are markedly underestimated in the transmission scenarios in which there are lower proportions of unique cases or in which smaller clusters predominate. This bias is especially marked when odds ratios are high; in the worst-case scenario, an odds ratio of 10 could be estimated as 1.58 when only 10% of the isolates are sampled. This finding implicates that, in areas with high TB prevalence and low TB detecting and reporting rates, the bias expected in these studies would be so extreme that the findings would be useful only as lower bounds for the proportion of recently transmitted cases and for risk factors for recent transmission. Nevertheless, lower bounds may be informative in situations in which undetected transmission is incorrectly attributed to reactivation disease alone or when a new risk factor for transmission is identified (114). Using data from a published investigation of the epidemiology of TB conducted from 1990 to 1993 among tuberculosis patients in New York City, New York, Murray and Alland further showed how selecting different measures of disease frequency, comparison groups, and sampling strategies may impact the results and interpretability of the study. They demonstrated that current study design and analysis of the molecular epidemiology of tuberculosis do not consistently yield interpretable and comparable results, especially when small sampling fractions have been used (111). In addition, the amount of clustering seen will also depend on the duration of the study. The proportion of clustered isolates increases with the increase of duration of observation (115).

Furthermore, as current studies of risk factors for recent transmission compare the probability of exposure in clustered cases to the probability of exposure in unique cases, they do not identify factors that put uninfected people at risk of primary disease. This question can be addressed by using an appropriate comparison series to determine the distribution of the exposure in the population that gave rise to the cases. Although the true distribution of cluster sizes

cannot be observed in the absence of complete sampling, epidemic models may elucidate factors that contribute to these distributions and help investigators arrive at prior expectations of cluster distributions in the specific transmission scenarios under study (114).

5. PERSPECTIVES

The exponential development of molecular techniques for strain typing of *M. tuberculosis* complex and the increased integration of these techniques with conventional epidemiology approaches over the last decade have enabled researchers to study many important but previously unanswered questions in the epidemiology of TB. Findings from the molecular epidemiological studies of TB have increased our knowledge about the etiology, pathogenesis, and epidemiology of the disease. A better understanding of the disease dynamics in populations has led to the strengthening of TB control. Nevertheless, some methodological problems in the molecular epidemiology of TB remain to be solved. In order to be maximally informative, studies should involve a high proportion of all cases in a population and be conducted in conjunction with conventional epidemiological investigations. The interpretation of molecular epidemiological study results should be made with the consideration of the study setting, the proportion of cases included, the incidence of TB in the population, the duration of study, and the study patients' HIV status, social and ethnic group, and immigration status.

6. ACKNOWLEDGEMENT

7. REFERENCES

1.Dye, C., S. Scheele, P. Dolin, V. Pathania, R.C. Raviglione: Global burden of tuberculosis - estimated incidence, prevalence, and mortality by country. *JAMA* 282, 677-686 (1999)

2.Ganguly, N.K., V.M. Katoch: MDR: Its impact on treatment of AIDS and TB. Multi-Drug Resistance in Emerging and Re-Emerging Diseases 15-22 (2000)

3.Cegielski, J.P., D.P. Chin, M.A. Espinal, T.R. Frieden, R.R. Cruz, E.A. Talbot, D.E.C. Weil, R. Zaleskis, M.C. Raviglione: The global tuberculosis situation - progress and problems in the 20th century, prospects for the 21st century. *Infect Dis Clin N America* 16, 1 (2002)

4.Dye, C., M.A. Espinal, C.J. Watt, C. Mbiaga, B.G. Williams: Worldwide incidence of multidrug-resistant tuberculosis. *J Infect Dis* 185, 1197-1202 (2002)

5.Dye, C., B.G. Williams, M.A. Espinal, M.C. Raviglione: Erasing the world's slow stain: Strategies to beat multidrug-resistant tuberculosis. *Science* 295, 2042-2046 (2002)

6.Cave, M.D., K.D. Eisenach, P.F. McDermott, J.H. Bates, J.T. Crawford: IS*6110* - conservation of sequence in the *Mycobacterium tuberculosis* complex and its utilization in DNA fingerprinting. *Mol Cell Endocrinol* 5, 73-80 (1991)

7.Mazurek, G.H., M.D. Cave, K.D. Eisenach, R.J. Wallace, J.H. Bates, J.T. Crawford: Chromosomal DNA fingerprint patterns produced with IS*6110* as strain-specific markers for epidemiologic-study of tuberculosis. *J Clin Microbiol* 29, 2030-2033 (1991)

8.Ross, B.C., K. Raios, K. Jackson, B. Dwyer: Molecular cloning of a highly repeated DNA element from *Mycobacterium tuberculosis* and its use as an epidemiological tool. *J Clin Microbiol* 30, 942-946 (1992)

9.Haas, W.H., W.R. Butler, C.L. Woodley, J.T. Crawford: Mixed-linker polymerase chain-reaction - a new method for rapid fingerprinting of isolates of the *Mycobacterium tuberculosis* complex. *J Clin Microbial* 31, 1293-1298 (1993)

10.Palittapongarnpim, P., S. Chomyc, A. Fanning, D. Kunimoto: DNA fragment length polymorphism analysis of *Mycobacterium tuberculosis* isolates by arbitrarily primed polymerase chain reaction. *J Infect Dis* 16, 975-978 (1993)

11.Plikaytis, B.B., J.T. Crawford, C.L. Woodley, W.R. Butler, K.D. Eisenach, M.D. Cave,T.M. Shinnick: Rapid, amplification-based fingerprinting of *Mycobacterium tuberculosis*. *J Gen Micobiology* 139, 1537-1542 (1993)

12.van Embden, J., M. Cave, J. Crawford, J. Dale, K. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, T. Shinnich,P. Small: Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: Recommendation for a standardized methodology. *J Clin Microbiol* 31, 406-409 (1993)

13.van Soolingen, D., P.E.W. de Haas, J. Haagsma, T. Eger, P.W.M. Hermans, V. Ritacco, A. Alito, J.D.A. van Embden: Use of various genetic markers in differentiation of mycobacterium bovis strains from animals and humans and for studying epidemiology of bovine tuberculosis. *J Clin Microbiol* 32, 2425-2433 (1994)

14.Wiid, I.J.F., C. Werely, N. Beyers, P. Donald, P.D. Vanhelden: Oligonucleotide (GTG)5 as a marker for mycobacterium- tuberculosis strain identification. *J Clin Microbiol* 32, 1318-1321 (1994)

15.Friedman, C.R., M.Y. Stoeckle, W.D. Johnson Jr, L.W. Riley: Double repetitive element PCR method for subtyping *Mycobacterium tuberculosis* clinical isolates. *J Clin Microbiol* 33, 1064-1069 (1995)

16.Frothingham, R.: Differentiation of strains in *Mycobacterium tuberculosis* complex by DNA-sequence polymorphisms, including rapid identification of *Mycobacterium bovis* BCG. *J Clin Microbiol* 33, 840-844 (1995)

17.Chaves, F., Z.H. Yang, H. ElHajj, M. Alonso, W.J. Burman, K.D. Eisenach, F. Dronda, J.H. Bates, M.D. Cave: Usefulness of the secondary probe pTBN12 in DNA fingerprinting of *Mycobacterium tuberculosis*. *J Clin Microbiol* 34, 1118-1123 (1996)

18.Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal,J. van Embden: Rapid detection and simultaneous strain differentiation of *Mycobacterium tuberculosis* for diagnosis and tuberculosis control. *J Clin Microbiol* 35, 907-914 (1997)

19.Otal, I., S. Samper, M.P. Asensio, M.A. Victoria, M.C. Rubio, R. Gomez-Lus, C. Martin: Use of a PCR method based on IS*6110* polymorphism for typing *Mycobacterium tuberculosis* strains from BACTEC cultures. *J Clin Microbiol* 35, 273-277 (1997)

20.Hopewell, P.C., Small, P.M.: Applications of molecular epidemiology to the prevention, control, and study of

tuberculosis. In: Tuberculosis. Rom, W.N., Garay, S.M., Ed., Little, Brown New York, pp. 113-127 (1996).

21.van Soolingen, D.: Molecular epidemiology of tuberculosis and other mycobacterial infections: Main methodologies and achievements. *Intern Med* 249, 1-26 (2001)

22.Murray, M., E. Nardell: Molecular epidemiology of tuberculosis: Achievements and challenges to current knowledge. *Bull WHO* 80, 477-482 (2002)

23.Wayne, L.G.: Mycobacterial speciation. Marcel Dekker, New York (1984).

24.van Soolingen, D., T. Hoogenboezem, P.E.W. de Haas, P.W.M. Hermans, M.A. Koedam, K.S. Teppema, P.J. Brennan, G.S. Besra, F. Portaels, J. Top, L.M. Schouls,J.D.A. van Embden: A novel pathogen taxon of the *Mycobacterium tuberculosis* complex, canettii: Characterization of an exceptional isolate from Africa. *Int J Syst Bacteriol* 47, 1236-1245 (1997)

25.Collins, D.M., G.W. de Lisle: DNA restriction endonuclease analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovic* BCG. *J Gen Micobiology* 130, 1019-1021 (1984)

26.Imeada, T.: Deoxyribonucleic acid relatedness among selected strains of *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium bovis* BCG, *Mycobacterium microti*, and *Mycobacterium africanum*. *Int J Syst Bacteriol* 35, 147-150 (1985)

27.Eisenach, K.D., J.T. Crawford,J.H. Bates: Genetic relatedness among strains of the *Mycobacterium tuberculosis* complex - analysis of restriction fragment heterogeneity using cloned DNA probes. *Am Rev Respir Dis* 133, 1065-1068 (1986)

28.Kirschner, P., B. Springer, U. Vogel, A. Meier, A. Wrede, M. Kiekenbeck, F.C. Bange,E.C. Bottger: Genotypic identification of mycobacteria by nucleic acid sequence determination: Report of a 2-year experience in a clinical laboratory. Tuber Lung Dis 79, 1730-1734 (1993)

29.Frothingham, R., H.G. Hills, K.H. Wilson: Extensive DNA sequence conservation throughout the *Mycobacterium tuberculosis* complex. *J Clin Microbiol* 32, 1639-1642 (1994)

30.Kapur, V., L.L. Li, S. Iordanescu, M.R. Hamrick, A. Wanger, B.N. Kreiswirth, J.M. Musser: Characterization by automated DNA-sequencing of mutations in the gene (rpoB) encoding the RNA-polymerase beta-subunit in rifampin-resistant *Mycobacterium tuberculosis* strains from New York City and Texas. *J Clin Microbiol* 32, 1095-1098 (1994)

31.Feizabadi, M.M., I.D. Robertson, D.V. Cousins, D.J. Hampson: Genomic analysis of *Mycobacterium bovis* and other members of the *Mycobacterium tuberculosis* complex by isoenzyme analysis and pulsed-field gel electrophoresis. *J Clin Microbiol* 34, 1136-1142 (1996)

32.Sreevatsan, S., X. Pan, K.E. Stockbauer, N.D. Connell, B.N. Kreiswirth, T.S. Whittam, J.M. Musser: Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci* 94, 9869-9874 (1997)

33.Wheeler, P.R.R., C.: Metabolistm of *Mycobacterium tuberculosis*. In: Tuberculosis: Pathogensis, Protection, and Control. Bloom, B.R., Ed., ASM Press, Washington DC, pp. 353-385 (1994).

34.Eisenach, K.D., J.T. Crawford, J.H. Bates: Repetitive DNA-sequences as probes for *Mycobacterium tuberculosis*. *J Clin Microbiol* 26, 2240-2245 (1988)

35.Thierry, D., M.D. Cave, K.D. Eisenach, J.T. Crawford, J.H. Bates, B. Gicquel,J.L. Guesdon: IS*6110*, an IS-like element of *Mycobacterium tuberculosis* complex. *Nucleic Acids Res* 18, 188-188 (1990)

36.Collins, D.M., D.M. Stephens: Identification of an insertion-sequence, IS*1081* in *Mycobacterium bovis*. *FEMS Microbiol Lett* 83, 11-16 (1991)

37.Hermans, P.W.M., D. van Soolingen, E.M. Bik, P.E.W. de Haas, J.W. Dale, J.D.A. van Embden: The insertion element IS*987* from *Mycobacterium bovis* BCG is located in a hot spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. *Infect Immun* 59, 2695-2705 (1991)

38.Hermans, P.W.M., D. van Soolingen, J. D. A. van Embden: Characterization of a major polymorphic tandem repeat in *Mycobacterium tuberculosis* and its potential use in the epidemiology of *Mycobacterium kansasii* and *Mycobacterium gordonea*. *J Bacteriol* 174, 4157-4165 (1992)

39.Mariani, F., E. Piccolella, V. Colizzi, R. Rappuoli, R. Gross: Characterization of an IS-like element from *Mycobacterium tuberculosis*. *J Gen Microbiol* 139, 1767-1772 (1993)

40.Goyal, M., D. Young, Y. Zhang, P.A. Jenkins, R.J. Shaw: PCR amplification of variable sequence upstream of katG gene to subdivide strains of *Mycobacterium tuberculosis* complex. *J Clin Microbiol* 32, 3070-3071 (1994)

41.van Embden, J.D.A., L.M. Schouls, D. van Soolingen: Molecular techniques: Applications in epidemiolgic studies. Thoen, C.O., Steele, J.H., Eds:, Iowa State University Press Ames, Iowa, pp. 15-27 (1995).

42.Fang, Z., N. Morrison, B. Watt, C. Doig, K.J. Forbes: IS*6110* transposition and evolutionary scenario of the direct repeat locus in a group of closely related *Mycobacterium tuberculosis* strains. *J Bacteriol* 180, 2102-2109 (1998)

43.Frothingham, R., Meeker-O'Connell: Genetic diversity in *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology* 144, 1189-1196 (1998)

44.Cole, S., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. Gordon, K. Eiglmeier, S. Gas, C.E. Barry, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. Quail, M. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. Sulston, K. Taylor, S. Whitehead, B.G. Barrell: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537-544 (1998)

45.Gordon, S.V., B. Heym, J. Parkhill, B. Barrell, S.T. Cole: New insertion sequences and a novel repeated sequence in the genome of *Mycobacterium tuberculosis*. *Microbiology* 145, 881-892 (1999)

46.van Soolingen, D., P.W.M. Hermans, P.E.W. de Haas, J.D.A. van Embden: Insertion element IS*1081*-associated restriction fragment length polymorphism in

Mycobacterium tuberculosis complex species: A reliable tool to recognize *Mycobacterium bovis* BCG. *J Clin Microbiol* 30, 1772-1777 (1992)

47. Hermans, P.W.M., A.R.J. Schuitema, D. Vansoolingen, C. Verstynen, E.M. Bik, J.E.R. Thole, A.H.J. Kolk, J.D.A. Vanembden: Specific detection of *Mycobacterium tuberculosis* complex strains by polymerase chain-reaction. *J Clin Microbiol* 28, 1204-1213 (1990)

48. van Soolingen, D., L. Quian, P.E.W. de Haas, J.T. Douglas, H. Traore, F. Portaels, H.Z. Qing, D. Enkhasaikan, P. Nymadawa, J.D.A. van Embden: Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. *J Clin Microbiol* 33, 3234-3238 (1995)

49. Samper, S., C. Martin, A. Pinedo, A. Rivero, J. Blazquez, F. Baquero, D. van Soolingen, J. D. A. van Embden: Transmission between HIV-infected patients of multidrug- resistant tuberculosis caused by *Mycobacterium bovis*. *AIDS* 11, 1237-1242 (1997)

50. Yang, Z.H., P.F. Barnes, F. Chaves, K.D. Eisenach, S.E. Weis, J.H. Bates, M.D. Cave: Diversity of DNA fingerprints of *Mycobacterium tuberculosis* isolates in the United States. *J Clin Microbiol* 36, 1003-1007 (1998)

51. Park, Y.K., G.H. Bai, S.J. Kim: Restriction fragment length polymorphism analysis of *Mycobacterium tuberculosis* isolated from countries in the western pacific region. *J Clin Microbiol* 38, 191-197 (2000)

52. van Soolingen, D., P.W.M. Hermans, P.E.W. Dehaas, D.R. Soll, J.D.A. Vanembden: Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains - evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *J Clin Microbiol* 29, 2578-2586 (1991)

53. Small, P.M., P.C. Hopewell, S.P. Singh, A. Paz, J. Parsonnet, D.C. Ruston, G.F. Schecter, C.L. Daley, G.K. Schoolnik: The epidemiology of tuberculosis in San-Francisco - a population-based study using conventional and molecular methods. *New Engl J Med* 330, 1703-1709 (1994)

54. Alland, D., G.E. Kalkut, A.R. Moss, R.A. McAdam, J.A. Hahn, W. Bosworth, E. Drucker, B.R. Bloom: Transmission of tuberculosis in New York City - an analysis by DNA-fingerprinting and conventional epidemiologic methods. *New Engl J Med* 330, 1710-1716 (1994)

55. Dale, J.W., T.H. Tang, S. Wall, Z.F. Zainuddin, B. Plikaytis: Conservation of IS*6110* sequence in strains of *Mycobacterium tuberculosis* with single and multiple copies. *Tubercle and Lung Disease* 78, 225-227 (1997)

56. Diaz, R., K. Kremer, P.E.W. de Haas, R.I. Gomez, A. Marrero, J.A. Valdivia, J.D.A. van Embden, D. van Soolingen: Molecular epidemiology of tuberculosis in Cuba outside of Havana, July 1994 - June 1995: Utility of spoligotyping versus IS*6110* restriction fragment length polymorphism. *Intl Tuber and Lung Dis* 2, 743-750 (1998)

57. Bauer, J., A.B. Andersen, K. Kremer, H. Miorner: Usefulness of spoligotyping to discriminate IS6110 low-copy-number *Mycobacterium tuberculosis* complex strains cultured in Denmark. *J Clin Microbiol* 37, 2602-2606 (1999)

58. Rasolofo-Razanamparany, V., H. Ramarokoto, G. Auregan, B. Gicquel S. Chanteau: A combination of two genetic markers is sufficient for restriction fragment length polymorphism typing of *Mycobacterium tuberculosis* complex in areas with a high incidence of tuberculosis. *J Clin Microbiol* 39, 1530-1535 (2001)

59. Cousins, D.V., R.A. Skuce, R.R. Kazwala, J.D.A. van Embden: Towards a standardized approach to DNA fingerprinting of *Mycobacterium bovis*. *Int J Tuber and Lung Dis* 2, 471-478 (1998)

60. Yang, Z.H., J.H. Bates, K.D. Eisenach, M.D. Cave: Secondary typing of *Mycobacterium tuberculosis* isolates with matching IS*6110* fingerprints from different geographic regions of the united states. *J Clin Microbiol* 39, 1691-1695 (2001)

61. Yang, Z.H., F. Chaves, P.F. Barnes, W.J. Burman, J. Koehler, K.D. Eisenach, J.H. Bates, M.D. Cave: Evaluation of method for secondary DNA typing of *Mycobacterium tuberculosis* with pTBN12 in epidemiologic study of tuberculosis. *J Clin Microbiol* 34, 3044-3048 (1996)

62. Barnes, P., Z.H. Yang, S. Preston-Martin, J. Pogoda, B. Jones, M. Otaya, K. Eisenach, L. Knowles, S. Harvey, M. Cave: Patterns of tuberculosis transmission in central los angeles. *JAMA* 278, 1159-1163 (1997)

63. Burman, W.J., R.R. Reves, A.P. Hawkes, C.A. Rietmeijer, Z.H. Yang, H. ElHajj, J.H. Bates, M.D. Cave: DNA fingerprinting with two probes decreases clustering of *Mycobacterium tuberculosis*. *Am J Respir Crit Care Med* 155, 1140-1146 (1997)

64. Supply, P., J. Magdalena, S. Himpens, C. Locht: Identification of novel intergenic repetitive units in a mycobacterial two component system operon. *Mol Microbiol* 26, 991-1003 (1997)

65. Mazars, E., S. Lesjean, A.L. Banuls, M. Gilbert, V. Vincent, B. Gicquel, M. Tibayrenc, C. Locht, P. Supply: High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc Natl Acad Sci* 98, 1901-1906 (2001)

66. Supply, P., S. Lesjean, E. Savine, K. Kremer, D. van Soolingen, C. Locht: Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *J Clin Microbiol* 39, 3563-3571 (2001)

67. Kremer, K., D. van Soolingen, R. Frothingham, W.H. Haas, P.W.M. Hermans, C. Martin, P. Palittapongarnpim, B.B. Plikaytis, L.W. Riley, M.A. Yakrus, J.M. Musser, J.D.A. van Embden: Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: Interlaboratory study of discriminatory power and reproducibility. *J Clin Microbiol* 37, 2607-2618 (1999)

68. Yang, Z.H., K. Ijaz, J.H. Bates, K.D. Eisenach, M.D. Cave: Spoligotyping and polymorphic GC-rich repetitive sequence fingerprinting of *Mycobacterium tuberculosis* strains having few copies of IS*6110*. *J Clin Microbial* 38, 3572-3576 (2000)

69. Cowan, L.S., L. Mosher, L. Diem, J.P. Massey, J.T. Crawford: Variable-number-tandem repeat typing of *Mycobacterium tuberculosis* isolates with low copy

numbers of IS*6110* by using mycobacterial interspersed repetitive units. *J Clin Microbiol* 40, 1592-1602 (2002)

70.Smith, P.G., A.R. Moss: Epidemiology of tuberculosis. In: Tuberculosis: Pathogenesis, Protection, and Control. Bloom, B.R., Ed., ASM Press Washington DC, pp. 47-60 (1994).

71.Comstock, G.W.: Community research in tuberculosis, Muscogee County. *Public Health Rep* 79, 1045-1046 (1964)

72.Horwitz, O., E. Wilbek, P.A. Erickson: Longitudinal studies on the risk of tuberculosis in the general population of a low-prevalence area. *Bull WHO* 41, 95-113 (1969)

73.Comstock, G.W.: Frost revisited: The modern epidemiology of tuberculosis. *J Epidemiol* 101, 363-382 (1975)

74.Edlin, B.R., J.I. Tokars, M.H. Grieco, J.T. Crawford, J. Williams, E.M. Sordillo, K.R. Ong, J.O. Kilburn, S.W. Dooley, K.G. Castro, W.R. Jarvis, S.D. Holmberg: An outbreak of multidrug-resistant tuberculosis among hospitalized-patients with the acquired immunodeficiency syndrome. *New Engl J Med* 326, 1514-1521 (1992)

75.Fischl, M.A., R.B. Uttamchandani, G.L. Daikos, R.B. Poblete, J.N. Moreno, R.R. Reyes, A.M. Boota, L.M. Thompson, T.J. Cleary, S.H. Lai: An outbreak of tuberculosis caused by multiple drug resistant tubercle bacilli among patients with HIV infection. *Ann Inter Med* 117, 177-183 (1992)

76.Daley, C.L., P.M. Small, G.F. Schecter, G.K. Schoolnik, R.A. McAdam, W.R. Jacobs, P.C. Hopewell: An outbreak of tuberculosis with accelerated progression among persons infected with the human immunodeficiency virus - an analysis using restriction fragment length polymorphisms. *New Engl J Med* 326, 231-235 (1992)

77.Godfreyfaussett, P., P.R. Mortimer, P.A. Jenkins, N.G. Stoker: Evidence of transmission of tuberculosis by DNA fingerprinting. *Br Med J* 305, 221-223 (1992)

78.Coronado, V.G., C.M. Beck-Sague, M.D. Hutton, B.J. Davis, P. Nicholas, C. Villareal, C.L. Woodley, J.O. Kilburn, J.T. Crawford, T.R. Frieden, R.L. Sinkowitz, W.R. Jarvis: Transmission of multidrug-resistant *Mycobacterium tuberculosis* among persons with human immunodeficiency virus infection in an urban hospital: Epidemiologic and restriction fragment length polymorphism analysis. *J Infect Dis* 168, 1052-1055 (1993)

79.Tabet, S.R., G.M. Goldbaum, T.M. Hooton, K.D. Eisenach, M.D. Cave, C.M. Nolan: Restriction fragment length polymorphism analysis detecting a community based tuberculosis outbreak among persons infected with human immunodeficiency virus. *J Infect Dis* 169, 189-192 (1994)

80.Pearson, M.L., J.A. Jereb, T.R. Frieden, J.T. Crawford, B.J. Davis, S.W. Dooley, W.R. Jarvis: Nosocomial transmission of multidrug resistant mycobacterium tuberculosis - a risk to patients and health care workers. *Ann Inter Med* 117, 191-196 (1992)

81.Small, P.M., N.B. McClenny, S.P. Singh, G.K. Schoolnik, L.S. Tompkins, P.A. Mickelsen: Molecular strain typing of *Mycobacterium tuberculosis* to confirm cross-contamination in the mycobacteriology laboratory and modification of procedures to minimize occurrence of false-positive cultures. *J Clin Microbiol* 31, 1677-1682 (1993)

82.Burman, W.J., M.L. Wilson, R.R. Reves: Specimen contamination in mycobacteriology laboratory detected by pseudo-outbreak of multidrug-resistant tuberculosis: Analysis by routine epidemiology and confirmation by molecular technique. *J Clin Microbiol* 34, 3257-3257 (1996)

83.Genewein, A., A. Telenti, C. Bernasconi, C. Mordasini, S. Weiss, A.M. Maurer, H.L. Rieder, K. Schopfer,T. Bodmer: Molecular approach to identifying route of transmission of tuberculosis in the community. *Lancet* 342, 841-844 (1993)

84.Yang, Z.H., P.E.W. Dehaas, C.H. Wachmann, D. Vansoolingen, J.D.A. Vanembden, A.B. Andersen: Molecular epidemiology of tuberculosis in Denmark in 1992. *J Clin Microbiol* 33, 2077-2081 (1995)

85.PinedaGarcia, L., A. Ferrera, S.E. Hoffner: DNA fingerprinting of *Mycobacterium tuberculosis* strains from patients with pulmonary tuberculosis in Honduras. *J Clin Microbiol* 35, 2393-2397 (1997)

86.Braden, C.R., G.L. Templeton, M.D. Cave, S. Valway, I.M. Onorato, K.G. Castro, D. Moers, Z.H. Yang, W.W. Stead, J.H. Bates: Interpretation of restriction fragment length polymorphism analysis of *Mycobacterium tuberculosis* isolates from a state with a large rural population. *J Infect Dis* 175, 1446-1452 (1997)

87.Bishai, W.R., N.M.H. Graham, S. Harrington, D.S. Pope, N. Hooper, J. Astemborski, L. Sheely, D. Vlahov, G.E. Glass, R.E. Chaisson: Molecular and geographic patterns of tuberculosis transmission after 15 years of directly observed therapy. *JAMA* 280, 1679-1684 (1998)

88.Jasmer, R.M., J.A. Hahn, P.M. Small, C.L. Daley, M.A. Behr, A.R. Moss, J.M. Creasman, G.F. Schecter, E.A. Paz, P.C. Hopewell: A molecular epidemiologic analysis of tuberculosis trends in San Francisco, 1991-1997. *Ann Inter Med* 130, 971-978 (1999)

89.Borgdorff, M.W., M.A. Behr, N.J.D. Nagelkerke, P.C. Hopewell, P.M. Small: Transmission of tuberculosis in San Francisco and its association with immigration and ethnicity. *Int J Tuber and Lung Dis* 4, 287-294 (2000)

90.Chin, D.P., C.M. Crane, M.Y. Diul, S.J. Sun, R. Agraz, S. Taylor, E. Desmond, F. Wise: Spread of *Mycobacterium tuberculosis* in a community implementing recommended elements of tuberculosis control. *JAMA* 283, 2968-2974 (2000)

91.Lillebaek, T., A.B. Andersen, J. Bauer, A. Dirksen, S. Glismann, P. de Haas, A. Kok-Jensen: Risk of *Mycobacterium tuberculosis* transmission in a low-incidence country due to immigration from high-incidence areas. *J Clin Microbiol* 39, 855-861 (2001)

92.Chin, D.P., K. Deriemer, P.M. Small, A. Ponce De Leon, R. Steinhart, G.F. Schecter, C.L. Daley, A.R. Moss, E.A. Paz, R.M. Jasmer, C.B. Agasino, P.C. Hopewell: Differences in contributing factors to tuberculosis incidence in U.S. Born and foreign-born persons. *Am J Respir Crit Care Med* 158, 1797-1803 (1998)

93.Shafer, R.W., P.M. Small, C. Larkin, S.P. Singh, P. Kelly, M.F. Sierra, G. Schoolnik, K.D. Chirgwin: Temporal trends and transmission patterns during the emergence of multidrug-resistant tuberculosis in New York City - a molecular epidemiologic assessment. *J Infect Dis* 171, 170-176 (1995)

94.Moss, A.R., D. Alland, E. Telzak, D. Hewlett, V. Sharp, P. Chiliade, V. LaBombardi, D. Kabus, B. Hanna, L. Palumbo, K. Brudney, A. Weltman, K. Stoeckle, K. Chirgwin, M. Simberkoff, S. Moghazeh, W. Eisner, M. Lutfey, B.N. Kreiswirth: A city-wide outbreak of a multiple-drug-resistant strain of *Mycobacterium tuberculosis* in New York. *Int J of Tuber and Lung Dis* 1, 115-121 (1997)

95.Bifani, P.J., B. Mathema, Z.Y. Liu, S.L. Moghazeh, B. Shopsin, B. Tempalski, J. Driscoll, R. Frothingham, J.M. Musser, P. Alcabes, B.N. Kreiswirth: Identification of a W variant outbreak of *Mycobacterium tuberculosis* via population-based molecular epidemiology. *JAMA* 282, 2321-2327 (1999)

96.Small, P.M., A. Moss: Molecular epidemiology and the new tuberculosis. *Infect Agents Dis* 2, 132-138 (1993)

97.Shafer, R.W., B.R. Edlin: Tuberculosis in patients infected with human immunodeficiency virus: Perspective on the past decade. *Clin Infect Dis* 22, 683-704 (1996)

98.Van Rie, A., R. Warren, M. Richardson, T.C. Victor, R.P. Gie, D.A. Enarson, N. Beyers, P.D. van Helden: Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. *New Engl J Med* 341, 1174-1179 (1999)

99.Caminero, J.A., M.J. Pena, M.I. Campos-Herrero, J.C. Rodriguez, O. Afonso, C. Martin, J.M. Pavon, M.J. Torres, M. Burgos, P. Cabrera, P.M. Small,D.A. Enarson: Exogenous reinfection with tuberculosis on a European island with a moderate incidence of disease. *Am J Respir Crit Care Med* 163, 717-720 (2001)

100.Sonnenberg, P., J. Murray, J.R. Glynn, S. Shearer, B. Kambashi, P. Godfrey-Faussett: HIV-1 and recurrence, relapse, and reinfection of tuberculosis after cure: A cohort study in south African mineworkers. *Lancet* 358, 1687-1693 (2001)

101.du Plessis, D.G., R. Warren, M. Richardson, J.J. Joubert, P.D. van Helden: Demonstration of reinfection and reactivation in HIV-negative autopsied cases of secondary tuberculosis: Multilesional genotyping of *Mycobacterium tuberculosis* utilizing IS*6110* and other repetitive element-based DNA fingerprinting. *Tuberculosis* 81, 211-220 (2001)

102.Rhee, J.T., A.S. Piatek, P.M. Small, L.M. Harris, S.V. Chaparro, F.R. Kramer, D. Alland: Molecular epidemiologic evaluation of transmissibility and virulence of *Mycobacterium tuberculosis*. *J Clin Microbiol* 37, 1764-1770 (1999)

103.Zhang, M., J. Gong, Z. Yang, B. Samten, M.D. Cave, D.G. Russell, P.F. Barnes: Enhanced capacity of a widespread strain of *Mycobacterium tuberculosis* to grow in human macrophages. *Tuber and Lung Dis* 80, (2000)

104.Kato-Maeda, M., P.J. Bifani, B.N. Kreiswirth,P.M. Small: The nature and consequence of genetic variability within *Mycobacterium tuberculosis*. *J Clin Invest* 107, 533-537 (2001)

105.Yang, Z.H., A. Rendon, A. Flores, R. Medina, K. Ijaz, J. Llaca, K.D. Eisenach, J.H. Bates, A. Villarreal, M.D. Cave: A clinic-based molecular epidemiologic study of tuberculosis in Monterrey, Mexico. *Int J Tuber and Lung Dis* 5, 313-320 (2001)

106.Borgdorff, M.W., N. Nagelkerke, D. van Soolingen, P.E.W. de Haas, J. Veen, J.D.A. van Embden: Analysis of tuberculosis transmission between nationalities in the Netherlands in the period 1993-1995 using DNA fingerprinting. *Am J Epidemiol* 147, 187-195 (1998)

107.Bauer, J., Z.H. Yang, S. Poulsen,A.B. Andersen: Results from 5 years of nationwide DNA fingerprinting of *Mycobacterium tuberculosis* complex isolates in a country with a low incidence of m-tuberculosis infection. *J Clin Microbiol* 36, 305-308 (1998)

108.Torrea, G., G. Levee, P. Grimont, C. Martin, S. Chanteau, B. Gicquel: Chromosomal DNA-fingerprinting analysis using the insertion- sequence IS*6110* and the repetitive element DR as strain-specific markers for epidemiologic study of tuberculosis in French Polynesia. *J Clin Microbiol* 33, 1899-1904 (1995)

109.Barnes, P.F., H. ElHajj, S. PrestonMartin, M.D. Cave, B.E. Jones, M. Otaya, J. Pogoda, K.D. Eisenach: Transmission of tuberculosis among the urban homeless. *JAMA* 275, 305-307 (1996)

110.Glynn, J.R., J. Bauer, A.S. de Boer, M.W. Borgdorff, P.E.M. Fine, P. Godfrey-Faussett, E. Vynnycky: Interpreting DNA fingerprint clusters of *Mycobacterium tuberculosis*. *Int J Tuber and Lung Dis* 3, 1055-1060 (1999)

111.Murray, M., D. Alland: Methodological problems in the molecular epidemiology of tuberculosis. *Am J Epidemiol* 155, 565-571 (2002)

112.Glynn, J.R., E. Vynnycky, P.E.M. Fine: Influence of sampling on estimates of clustering and recent transmission of *Mycobacterium tuberculosis* derived from DNA fingerprinting techniques. *Am J Epidemiol* 149, 366-371 (1999)

113.Murray, M.: Determinants of cluster distribution in the molecular epidemiology of tuberculosis. *Proc Natl Acad Sci* 99, 1538-1543 (2002)

114.Murray, M.: Sampling bias in the molecular epidemiology of tuberculosis. *Emerg Infect Dis* 8, 363-369 (2002)

115.Vynnycky, E., N. Nagelkerke, M.W. Borgdorff, D. Van Soolingen, J.D.A. Van Embden, P.E.M. Fine: The effect of age and study duration on the relationship between 'clustering' of DNA fingerprint patterns and the proportion of tuberculosis disease attributable to recent transmission. *Epidemiol Infec* 126, 43-62 (2001)

Sending correspondence to: Zhenhua Yang, Epidemiology Department, University of Michigan, School of Public Health, 109 S. Observatory Street, Ann Arbor, MI 48109-2029, Tel: 734-763-4296, Fax: 734-764-3192, E-mail: zhenhua@umich.edu