DNA fingerprinting of mycobacterium of tuberculosis: bands and links

de Boer, A.S.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
CHAPTER 8

GENERAL DISCUSSION
Chapter 8
INTRODUCTION

This thesis combines studies on the interpretation and the application of RFLP typing of *M. tuberculosis* from an epidemiological perspective.

The first study on the interpretation of RFLP typing addressed the stability of IS6110 RFLP patterns over time. The probability over time that changes in IS6110 RFLP patterns occur was estimated and possible factors associated with instability were studied. The second study on the assessment of RFLP typing included the prevalence, nature and practical interpretation of low-intensity bands in the IS6110 RFLP patterns. In the third study, based on the interpretation of IS6110 RFLP patterns of *M. tuberculosis*, false-positive test results were identified. Combining these test results with laboratory and patient information, risk factors for and consequences of false-positive *M. tuberculosis* cultures were studied.

The application of RFLP typing of *M. tuberculosis* was the basis of the other studies in this thesis. First, exogenous re-infection as a cause of recurrent tuberculosis was studied, adding direct evidence from observational data to the mathematical modelling results which had indicated that exogenous re-infection may cause a considerable proportion of tuberculosis morbidity. Second, the impact of defaulting from treatment on the transmission of *M. tuberculosis* was studied in a case-control study.

This chapter discusses the implications and shortcomings of the studies outlined above. In particular, different aspects of the stability of RFLP patterns of *M. tuberculosis* are addressed. In addition, methodological issues which are important for the epidemiological studies in this thesis are discussed.

THE INTERPRETATION OF IS6110 RFLP PATTERNS OF *M. TUBERCULOSIS*

An estimate of the molecular clock of IS6110 RFLP patterns

The study on the molecular clock of IS6110 RFLP patterns of *M. tuberculosis* based on serial patient isolates (Chapter 2) showed that 4.6% of follow-up isolates had RFLP patterns which differed from the initial isolates. Changes in RFLP patterns were positively associated with extrapulmonary disease and unclustered isolates,
which could suggest that the growing conditions of the bacteria affect the stability of IS6110 RFLP of *M. tuberculosis*. The main implication of this study is that the half-life of IS6110 RFLP of *M. tuberculosis*, 3.2 years, supports the appropriateness of IS6110 typing in epidemiological studies of ongoing transmission of tuberculosis.

A major concern regarding the use of genetic typing methods in epidemiological studies of tuberculosis is that the processing of the *M. tuberculosis* culture may trigger genetic changes in the bacterial population. For instance, before the actual IS6110 RFLP typing takes place, the *M. tuberculosis* isolate that the peripheral laboratory obtained by culturing a patient sample, is recultured to obtain sufficient quantities of DNA for the RFLP typing. If the amount of DNA is not sufficient after the first time the *M. tuberculosis* isolate was recultured, the culturing is repeated, until a sufficient amount of DNA is obtained. This means in practice that the DNA used for the IS6110 RFLP typing is obtained after culturing a patient isolate at least two times. As during a rapid growth phase it is likely that part of the bacterial population undergoes genetic rearrangements, it is possible that the culturing in itself is a trigger for changes to occur. A genetic typing method that directly targets DNA present in the clinical sample, such as spoligotyping, could be a worthwhile alternative for the current IS6110 RFLP typing. Unfortunately, spoligotyping is not sufficiently discriminatory to be the typing method of choice. The new PCR-based typing method which uses the variable-number tandem repeats of mycobacterial interspersed repetitive units (MIRU-VNTR typing) is very promising [1], but it is not as easy to perform as spoligotyping [2]. Furthermore, it will exhibit the same drawbacks of for instance inhibition of the PCR when applied directly to clinical material.

Chapter 2 describes that changed RFLP patterns were more common for extrapulmonary than for pulmonary isolates. However, an experimental study would need to be carried out to determine to what extent adverse bacterial conditions can stimulate transpositional events. It has been shown before that anaerobic conditions stimulate mutations in *M. tuberculosis* in vitro [3]. It has also been suggested that certain *M. tuberculosis* strains are more likely to change than others [4]. An experiment, culturing isolates under anaerobic and aerobic conditions, should be carried out with specific *M. tuberculosis* genotypes, e.g., with the Beijing strain. After a time period of several weeks, the RFLP patterns of *M. tuberculosis* isolates that grew under aerobic and under anaerobic conditions would need to be compared to those of the original strain to see if the anaerobic conditions led to more changed RFLP patterns. Other adverse bacterial conditions that
could be studied are antibiotic pressure and increased levels of immunoglobulins.

The stability of RFLP patterns in relation to genetic heterogeneity

The study on the genetic heterogeneity in *M. tuberculosis* isolates (Chapter 3) showed that low-intensity bands in IS6110 RFLP patterns can represent mixed bacterial populations with slightly different RFLP patterns. It is possible that these low-intensity bands reflect a gradual change of the bacterial population in the human body over time. This would be consistent with the association between low-intensity bands and a higher patient age. It is also possible that these low-intensity bands reflect an IS6110-mediated genetic adaptation of *M. Tuberculosis* to changes in the environmental conditions during the dormant state or reactivation thereafter. This would imply a role of IS6110 as a driving force for evolution.

In this study we assumed that the quality of IS6110 RFLP typing was good, e.g., that DNA was completely and evenly transferred during blotting. We showed that the routine interpretation of RFLP patterns with low-intensity bands is not standardised, resulting in an overestimation of the amount of clustering. We therefore recommend that bands with an intensity of more than 15% of the average intensity are to be considered as normal intensity bands, and bands with an intensity of less than 10% of the average intensity as no bands. Bands with an intensity of 10-15% of the average intensity should be labelled as low-intensity bands, and isolates with such bands ideally need to be fingerprinted again to find the originating patterns. Comparisons with other RFLP patterns would need to be made based on all originating patterns.

In the study on genetic heterogeneity of *M. tuberculosis* it was found that certain DNA fragment sizes are more often represented as low-intensity bands in the IS6110 RFLP pattern than others. That some DNA fragment sizes are not often represented as low-intensity bands could be related to the preferential insertion sites for IS6110 [5-7]. Knowing the complete sequence of the *M. tuberculosis* genome [8], it would be worthwhile to study the genomic regions flanking the IS6110 insertion sites of the DNA fragment sizes that are often represented as low-intensity bands. Such a study could determine whether the variability of these fragments has any functional implications, e.g., whether certain coding regions are interrupted or genes are over- or underexpressed which may explain the adaptation of *M. Tuberculosis* to certain conditions.
Chapter 8

The combined results of the study on the low-intensity bands in RFLP patterns of *M. tuberculosis* and on the stability of the RFLP patterns of *M. tuberculosis* suggest that an original homogenous population of *M. tuberculosis* bacteria can be gradually replaced by mutant bacteria. The mutations in bacterial populations of *M. tuberculosis* result in a different DNA fingerprint pattern for a part of a given *M. tuberculosis* population compared to the predominant pattern of the larger part of that bacterial population. If a part of a population is typed, only the DNA fingerprint pattern of that part of the population is revealed [9].

The stability of RFLP patterns of isolates from different settings

Although it has been suggested that different strains may have a different molecular clock, e.g., that particular multi drug resistant *M. tuberculosis* strains may evolve too fast for IS6110 RFLP typing to be useful to identify links between patients [4], we found that the stability of RFLP patterns of *M. tuberculosis* in data sets on serial patient isolates from other settings was not different from ours. For this ecological analysis we used data sets from The Netherlands (514 patients) [10], Denmark (119 patients), Germany (56 patients) [11], the United States of America (49 patients) [12], Spain (27 patients), Austria (15 patients), and Belgium (9 patients) [13]. The probability of changes in RFLP patterns over time was estimated using exponential survival analysis, assuming a constant rate of change and assuming that changes in the fingerprint pattern occurred between the date of the initial isolate and the date of the follow-up isolate. Although the median time intervals between serial isolates, the relative number of drug resistant and extrapulmonary isolates differed, the estimated half-life of RFLP patterns did not differ between these settings. This suggests that the probability of changes in RFLP patterns over time is not associated with these factors. This is in contrast with the findings in Chapter 2.

The stability of RFLP patterns and number of bands in the patterns

It has been suggested that the number of IS6110 copies (bands) in the IS6110 RFLP pattern of *M. tuberculosis* influences the rate of change of RFLP patterns, i.e. that isolates with a high number of IS6110 copies are more likely to change than isolates with a low copy number RFLP pattern [14]. When data from The Netherlands, Denmark, Germany, the United States, Spain, Austria and Belgium were analysed together [13], it was found that the molecular clock of RFLP patterns did not differ between isolates with a low, medium and high copy number isolates. Also when the number of copies was added as a
Generall discussio n

confounde rr  t o  th e  mode l whic h  estimate s  th e  probabilit y  o f  change s  i n
RFL PP  pattern s  ove r  time , th e  fit  o f  th e  resultin g  mode l di d  no t
improv ee  compare d  t o  th e  origina l mode l without  th e  number  o f
copies .

Is the rate of change constant over time?

Our assumption that the rate of change of RLFP patterns is constant
over time has recently been under debate. It is thought that changed
RLFP patterns constitute part of the overall bacterial population or
that the rate of change is proportional to the growth rate of
mycobacteria. If the latter is true, changes in RFLP patterns result
from changes during the rapid growth of the bacteria, i.e., during the
early stages of the infectious period. Modelling the probability of
change with a smooth non-parametric model in several data sets, a
model which assumes an instantaneous change at time zero, followed
by a zero rate of change, was found to suit the data best. If this
theory is correct, the rate of change during latency may be close to
zero [15], but much higher during the revival of bacteria thereafter,
when adaptation to the environment is most crucial.

Another study also focused on the rate of change related to the
growth rate of mycobacteria. In a study on 349 South African patients
with serial *M. tuberculosis* isolates, the half-life of IS6110 RFLP
patterns observed in this group was 0.6 years for serial isolates
spanning less than 21 days (group 1) and 10.7 years for serial isolates
spanning more than 250 days (group 2) [16]. This was explained for
group 1 as changes early in the disease process, prior to
antituberculosis treatment, influenced by active growth or adaptation
to the new host environment. For group 2 this was explained as
changes occurring when the growth rate was much lower. However,
this study only included data on 14 patients with serial isolates with
discordant IS6110 RFLP patterns. Calculating the half-life of IS6110
RFLP patterns spanning less than 21 days and spanning more than 250
days, the results for Dutch patients with serial *M. tuberculosis* isolates
are 0.3 years for group 1 and 19.1 years for group 2. This implies that
indeed the rate of change of IS6110 RFLP patterns during active
growth of the mycobacteria may be fast, followed by a slower rate of
change. Therefore, when presenting the half-life of IS6110 RFLP, it
should be stratified for time intervals of the studied serial isolates.

The stability of RFLP patterns in transmission chains

Only two studies tried to determine the molecular clock of IS6110
RFLP patterns of *M. tuberculosis* in transmission chains. One study
from Germany concluded that pattern alterations in contacts who
developed tuberculosis were rare [17]. Another study from South
Africa concluded that the minimum rate of appearance of variant strains was calculated to be 0.14 variant case per source-case per year for strains with more than 5 IS6110 elements [18]. This implies that clustering of isolates based on identical RFLP patterns is expected to underestimate transmission. The proposed quantification of the stability of RFLP patterns in transmission chains assumes that the risk of transmission is constant over time. However, this risk depends on the quality of the tuberculosis control programme and the background prevalence. Both may vary over time and will be different in different geographical areas. Therefore, the stability of RFLP patterns in transmission chains needs to be quantified in another way, maybe simply by presenting what proportion of a given study population is infected with a variant strain. If the treatment delay of patients is short, the infectious period of the patient and the period of active growth of the mycobacteria are also short. Therefore, a short treatment delay may not only be associated with a decreased risk of transmission of *M. tuberculosis* [19,20], but also with a slower molecular clock in RFLP patterns in transmission chains. Unfortunately data from The Netherlands or other settings on the stability of IS6110 RFLP patterns in transmission chains are unavailable yet.

**False-positive *M. tuberculosis* cultures**

IS6110 RFLP typing revealed that false-positive *M. tuberculosis* cultures occurred in 2.4% of all cultures which peripheral laboratories sent to the national reference laboratory (Chapter 4). This rate is similar to the 2% recently reported in studies in the United States. One used a laboratory based definition as we did [21], and one used a definition based on reviewing all medical and laboratory records and epidemiological investigations [22]. The rate of false-positive cultures differed greatly between peripheral laboratories in The Netherlands, as was observed in London [23]. Although false-positive test results affected only few patients in our study, they had serious consequences for those patients and their environment. The estimated costs of false laboratory diagnosis of tuberculosis were about $900 per patient. A recent study on three cases in the United States reported an average cost of almost $11,000 [21]. Although the discrepancy between our estimation and the calculation from the United States could be due to the small number of cases in the American study or to the generally higher medical costs in the United States, it is more likely that our estimation may be too low.

The results of our study suggest that the number of false-positive cultures could be further reduced, e.g., by only culturing *M. tuberculosis* in laboratories processing a large number of patient samples per year. Regarding the interpretation of IS6110 RFLP
patterns it is possible that the clustering information is influenced by some undetected false-positive cultures. Because of the reporting system from the national reference laboratory to the peripheral laboratories and the municipal health services, this bias is probably small. This study stresses the importance of DNA fingerprinting without delay at the national reference laboratory to evaluate the test results of the peripheral laboratories, and the importance of rapid feedback of these results to the peripheral laboratories. Unfortunately, the laboratory-based definition we used is not applicable in settings with a high prevalence of tuberculosis, a high throughput of samples per day, and less means to process samples optimally. Still, measures to prevent the occurrence of false-positive cultures are basically applicable in all settings [24,25].

EPIDEMIOLOGICAL INSIGHTS GAINED THROUGH THE USE OF RFLP TYPING

Methodological issues
A meaningful interpretation of DNA fingerprint clusters for epidemiological studies requires at least information on case ascertainment, case definition, geographical area, time period, definition of clustering, and cluster size distribution, and data should be broken down at least by age, sex and immigration status [26]. In this thesis molecular epidemiological studies of tuberculosis are based on the nationwide surveillance of tuberculosis in The Netherlands. For these studies case ascertainment, time windows and the definition of clustering need special attention.

Case ascertainment
If not all tuberculosis patients are included in the DNA fingerprint clusters, e.g., because of difficulties in culturing and fingerprinting, this sampling bias will lead to underestimating recent transmission and the impact of risk factors for recent transmission [27].

In addition, as IS6110 RFLP typing is based on isolates from culture positive patients, case ascertainment could benefit from standardisation of culturing *M. tuberculosis*. In 1998, 916 out of 1341 (68%) of patients were culture positive (NTR). It is likely that peripheral laboratories, which culture *M. tuberculosis*, have different protocols and procedures for culturing *M. tuberculosis*, perhaps resulting in differences in case ascertainment. Because there are as many as 44 peripheral laboratories involved in culture-confirmation of
tuberculosis, this standardisation could probably benefit also from reducing the number of culturing laboratories.

**Time windows**

It has been shown that short time windows may reduce clustering rates dramatically [28]. As all studies in this thesis are based on data from at least four years of nationwide surveillance, we think the bias related to time period is limited.

**Definition of clustering**

By definition, clustering of tuberculosis patients is based on identical fingerprints. Since 1999, fingerprints which differ one IS6110 copy from the cluster-specific fingerprint have also been included, provided conformation on epidemiological linkage between patients is available. As this information is often difficult to obtain, especially among tuberculosis risk groups, a bias is introduced. It is likely that the clustering rates among tuberculosis risk groups are biased downwards compared with patients with well-established contact information. As the clustering rates among tuberculosis risk groups are expected to be relatively high, this bias may not be noticed.

**Exogenous re-infection**

The proportion of recurrent tuberculosis attributable to exogenous re-infection varies greatly in different settings [5,29-33]. These findings are usually based on IS6110 RFLP typing, but it has been reported that spoligotyping can be used as well [34]. However, with spoligotyping the occurrence of re-infection will be underestimated, as this typing technique is less discriminatory than IS6110 RFLP typing. Although the level of laboratory cross contamination and other reasons for false-positive cultures could be a partial explanation for these difference, re-infection seems to be a minor problem in low incidence countries (The Netherlands, Germany and Norway) and a major problem in high incidence countries (sub-populations in South Africa). Interesting anecdotal evidence has been presented on reactivations after many years in Denmark [35] and The Netherlands [36]. However, although the baseline prevalence of tuberculosis mostly determines the risk for re-infection, the proportion of recurrent tuberculosis attributable to exogenous re-infection can be increased in specific settings, where there is a high HIV prevalence and/or an inadequate tuberculosis control programme.

Direct evidence of recurrent tuberculosis due to exogenous re-infection in The Netherlands has been reported for two out of 546 patients with serial isolates (0.4%) [9]. When excluding patients for
whom treatment completion was not certain, none of the six cases of recurrent tuberculosis with cure after the first episode were due to exogenous re-infection. Although the proportion of recurrent tuberculosis due to exogenous re-infection was expected to be low in The Netherlands, the power of this study was so limited that another approach, including historical data, was also used. This study revealed that approximately one out of six new disease episodes among patients with previous tuberculosis infection or disease may be attributable to recent re-infection (Chapter 5). A limitation of this study was that it only used indirect evidence of re-infection. Still, it implies that re-infection needs to be taken into account when studying tuberculosis epidemiology and organising tuberculosis control programmes, even in settings with a low incidence of tuberculosis.

**Defaulting from treatment**

In a case-control study it was found that for Dutch patients defaulting from treatment increased the risk of transmission of *M. tuberculosis* (Chapter 7). In addition, defaulting index cases with a drug-susceptible strain were twice as likely compared to non-defaulters to lead to secondary cases with drug-resistant tuberculosis, although this was not statistically significant. Cases were the first cases of clusters of patients with identical DNA fingerprints of *M. tuberculosis*, so probably linked by ongoing transmission. Controls were patients with an unclustered DNA fingerprint of *M. tuberculosis*. A strength of this study was that a case-control design could be applied, as the study base was the large national surveillance database. A weakness of the study was the definition of a case, as first cases of a cluster may not be the source case of the cluster. Limiting the study to clusters with epidemiologically confirmed patients would theoretically improve the study design. However, the number of such clusters in the national surveillance database in The Netherlands is too small to perform such a study.

It is recommendable to study the effect of defaulting from treatment on transmission of *M. tuberculosis* in other settings, so that our estimate of the impact of defaulting can be validated.

**BENEFITS OF RFLP TYPING FOR TUBERCULOSIS CONTROL IN THE NETHERLANDS**

RFLP typing of *M. tuberculosis* has been an integral part of the nationwide tuberculosis surveillance in The Netherlands since 1993, because it has been expected to yield epidemiological insights and thus support traditional contact investigations. However, the effects
of RFLP typing on tuberculosis control have been mainly indirect. The routine DNA fingerprint surveillance provides information on clustering which helps tuberculosis control staff to evaluate contact investigations and detect unexpected links between patients. In the period from 1997 to 1999 the tuberculosis control staff had expected epidemiological links for 22% of patients. After the results of DNA fingerprinting surveillance had become available, epidemiological links could be established for 47% of patients. The DNA fingerprinting surveillance also showed what groups are difficult to reach with traditional contact investigation [37]. Most patients are identified because of complaints (81% of Dutch patients and 75% of non-Dutch patients). In addition, passive case finding identified 15% of Dutch patients and 5% of non-Dutch patients, and active case finding identified 4% of Dutch patients and 20% of non-Dutch patients [38]. This suggests that non-Dutch patients are more often missed in traditional contact investigations, and that screening identifies a relatively large proportion of non-Dutch patients. The implications of the possibly failing tuberculosis control strategy for risk groups are unknown. A change in the conventional tuberculosis control strategy is the screening of asylum seekers since the beginning of the 1990s. The effects of this screening strategy are now being evaluated. It has been found recently that screening of asylum seekers identified 454 new cases in the period from 1993 to 1998 [39] and that this prevented secondary transmission [40]. If screening would indeed lead to a declining incidence of tuberculosis, the amount of clustered strains would also decrease. In this way trends in clustering rates can help to evaluate tuberculosis programmes, as was shown recently for a sentinel surveillance population in the United States [41].

To strengthen the contact investigation efforts among risk groups, a thorough analysis of how to interrupt transmission in these groups is necessary. Novel epidemiological methods could help to unravel transmission routes. One such method is based on generating a demographic profile of patients and thus searching for contacts of tuberculosis patients across boundaries [42]. In practice this would mean a widening of the traditional contact investigation, both with respect to the kind of people included and with respect to the geographical area under investigation. Another method could be network analysis, in which a network between persons is reconstructed based on interviews, followed by a quantification of the relative importance of different actors (persons and places) playing a role in the transmission chain [43]. In practice this would mean standardised contact investigations, with a quantification of what routes of transmission are most important. This may give clues where to intensify efforts to contact possible secondary patients.
Further intervention studies should focus on the accessibility of health care, both for persons with a latent infection and for patients with active tuberculosis, especially for risk groups which conventionally are hard to reach such as homeless, drug addicts and illegal immigrants. To improve the accessibility new strategies to reach these risk groups need to be developed. Additional measures to improve their situation, e.g., targeted on living conditions, could perhaps be beneficial elements in such strategies. Improving the situation of these risk groups will also in itself help to reduce the incidence of tuberculosis, as low socio-economic status [44-47], crowding [44,45,48,49], poor nutrition [50-52], immunodeficiency and comorbidity [53-55] are associated with an increased incidence of tuberculosis. Such a general approach to fight tuberculosis also proved right in the past: the substantial decline in tuberculosis incidence in the early twentieth century was achieved by improving the socio-economic conditions of the population at large, not by focusing on individuals at risk [56].

FINAL COMMENTS

The Netherlands Tuberculosis Register, and the possibility to link patient information to microbiological test results, has enabled many investigations into the epidemiology of tuberculosis over the years. However, epidemiological investigations on tuberculosis could still gain from securing the completeness and standardisation of patient data over the years, a further implementation of information technology and a direct link between patient data and microbiological test results. To comply with privacy regulations, this link could perhaps be automatically encrypted.

Routine IS6110 RFLP typing of all *M. tuberculosis* isolates in The Netherlands has provided valuable insights in the epidemiology of tuberculosis. Yet, the typing itself needs to be further standardised, addressing the genetic variations expressed in RFLP patterns from related isolates over time, as well as the uniform interpretation of RFLP patterns. If a genetic typing method that directly targets DNA present in the clinical sample would be developed, it is recommended to establish a new national database of DNA fingerprints.

Organised in a modern, multidisciplinary way a national surveillance of tuberculosis, based on both epidemiological and microbiological data, will enhance research on the molecular epidemiology of tuberculosis, and thus contribute to the prevention of tuberculosis morbidity and mortality.
REFERENCES


13. de Boer AS, Borgdorff MW, Kremer K, Bauer J, Niemann S, Small PM et al. The stability of IS6110 RFLP patterns of *Mycobacterium tuberculosis* is not associated with the number of IS6110 copies in strains, but may be


