

From the Department of Microbiology, Tumor- and Cell Biology
Karolinska Institutet, Stockholm, Sweden and the Swedish
Institute for Communicable Disease Control, Department of
Preparedness, Solna, Sweden

MOLECULAR EPIDEMIOLOGY OF TUBERCULOSIS

Ramona Groenheit



**Karolinska
Institutet**



Solna 2011

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

© Ramona Groenheit, 2011
ISBN 978-91-7457-465-4

Printed by



www.reproprint.se

Gårdsvägen 4, 169 70 Solna

ABSTRACT

Tuberculosis is a global epidemic, with one third of the world's population estimated to be infected, around 9 million new active cases per year and close to 2 million deaths per year. Without adequate chemotherapy tuberculosis may be a mortal disease.

A century ago, the estimated tuberculosis incidence in Sweden was higher than in most high incidence countries of today's sub-Saharan Africa. Today however, the majority of patients with tuberculosis in Sweden are immigrants from countries with a high incidence of tuberculosis. The incidence among the Swedish-born population has continued to decrease while it has increased among the foreign-born. In the West African country Guinea-Bissau, tuberculosis is a common disease and the incidence is believed to be further increased by the epidemic of the human immunodeficiency virus. In countries like Sweden the mortality was dramatically reduced about half a century ago when living conditions improved, public health measures were taken and treatments were made available. These gains are however seriously jeopardized by the now emerging multidrug resistant and extensively drug-resistant tuberculosis.

Different genotypes of *Mycobacterium tuberculosis* complex predominate in different geographical regions of the world and strain-to-strain variations may have important consequences for instance when it comes to transmissibility. Future diagnostics, drugs and vaccines are affected by these strain variations and it is therefore of great importance to establish the whole spectrum of strains of the *M. tuberculosis* complex worldwide. Despite the high prevalence of tuberculosis in Africa, relatively little is known about the *M. tuberculosis* complex genetic diversity in this continent. The studies included in this thesis phylogenetically and epidemiologically characterized *M. tuberculosis* complex isolates obtained from tuberculosis patients in Sweden and Guinea-Bissau using molecular techniques such as Restriction Fragment Length Polymorphism, spacer oligonucleotide typing and 24-loci Mycobacterial Interspersed Repetitive Units-Variable Numbers of Tandem Repeats. The work was performed with the view to understand species and strain diversity as well as transmission patterns.

It was illustrated that the great majority of tuberculosis patients with drug resistant isolates in Sweden were foreign-born and that their strain lineages to a large extent reflected genotypes common in their country of origin. One large outbreak of isoniazid resistant tuberculosis was identified, up to date (October 2011) involving 117 patients, mainly from the Horn of Africa. This outbreak represents one of the largest outbreaks of tuberculosis ever reported in a low incidence country and was an important warning signal to the Swedish authorities. By whole genome sequencing this outbreak strain showed to be exceptionally stable genetically. It was obvious that molecular epidemiological typing is a powerful tool to monitor and identify chains of transmission which could indicate deficiencies in national tuberculosis control programs. It was also discovered that Beijing lineage strains, which elsewhere in the world have caused large outbreaks, have not been able to spread within Sweden in spite of the proximity to high prevalence countries such as Russia and the Baltic countries. When isolates from patients born in Sweden before 1945 were studied, a highly homogenous bacterial population with a domination of the T, Haarlem and Latin-American-Mediterranean lineages was found. It was concluded that evolutionary recent (PGG2/3) strains restricted to Sweden and its immediate neighbours appeared to have caused the epidemic during the first half of the 20th century, while ancestral (PGG1) strains were usually linked to immigrant populations in today's Sweden. Guinea-Bissau was revisited and it was established that the country has the highest prevalence of *M. africanum* recorded in the African continent and that the Guinea-Bissau family of strains demonstrated high phylogeographical specificity for Western Africa.

LIST OF PUBLICATIONS

- I. Ghebremichael S, **Petersson R**, Koivula T, Pennhag A, Romanus V, Berggren I, Petrini B, Hoffner S, Källenius G. Molecular epidemiology of drug-resistant tuberculosis in Sweden. *Microbes and Infection*. 2008 May 10(6):699-705.
- II. Ghebremichael S, **Groenheit R**, Pennhag A, Koivula T, Andersson E, Bruchfeld J, Hoffner S, Romanus V, Källenius G. Drug resistant *Mycobacterium tuberculosis* of the Beijing genotype does not spread in Sweden. *PLoS One*. 2010 May 5(5):e10893.
- III. Sandegren L, **Groenheit R**, Koivula T, Ghebremichael S, Advani A, Castro E, Pennhag A, Hoffner S, Mazurek J, Pawlowski A, Kan B, Bruchfeld J, Meleforts Ö, Källenius G. Genomic stability over 9 years of an isoniazid resistant *Mycobacterium tuberculosis* outbreak strain in Sweden. *PLoS One*. 2011 Jan 6(1):e16647.
- IV. **Groenheit R**, Ghebremichael S, Pennhag A, Jonsson J, Hoffner S, Koivula T, Rastogi N, Källenius G. *Mycobacterium tuberculosis* strains causing the TB epidemic in Sweden a century ago. Manuscript.
- V. **Groenheit R**, Ghebremichael S, Svensson J, Rabna P, Colombatti R, Riccardi F, Couvin D, Hill V, Rastogi N, Koivula T, Källenius G. The Guinea-Bissau family of *Mycobacterium tuberculosis* complex revisited. *PLoS One*. 2011 Apr 6(4):e18601.

CONTENTS

1	INTRODUCTION	1
1.1	HISTORY AND GLOBAL HEALTH BURDEN OF TUBERCULOSIS	1
1.2	TUBERCULOSIS CONTROL	3
1.2.1	Tuberculosis in Sweden	3
1.2.2	Tuberculosis in Guinea-Bissau	5
1.3	MYCOBACTERIUM TUBERCULOSIS COMPLEX	6
1.3.1	<i>Mycobacterium tuberculosis</i>	6
1.3.2	<i>Mycobacterium bovis</i> and <i>Mycobacterium bovis</i> BCG	6
1.3.3	<i>Mycobacterium africanum</i>	7
1.3.4	<i>Mycobacterium canettii</i>	7
1.3.5	<i>Mycobacterium microti</i>	8
1.3.6	<i>Mycobacterium pinnipedii</i>	8
1.3.7	<i>Mycobacterium caprae</i>	8
1.4	LABORATORY DIAGNOSIS OF TUBERCULOSIS	9
1.4.1	Identification of the <i>M. tuberculosis</i> complex	9
1.4.2	Strain identification of the <i>M. tuberculosis</i> complex	10
1.4.3	Lineages of the <i>M. tuberculosis</i> complex	13
1.4.4	Drug resistant tuberculosis	14
2	THE PRESENT INVESTIGATION	15
2.1	PURPOSE OF THE STUDY	15
2.1.1	Specific objectives	15
2.2	MATERIALS AND METHODS	16
2.2.1	Study area	16
2.2.2	Sample processing	16
2.2.3	Chromosomal DNA isolation	17
2.2.4	Spoligotyping	17
2.2.5	RFLP	17
2.2.6	MIRU-VNTR	17
2.2.7	RD polymorphism	18
2.2.8	Rv3135 gene analysis	18
2.2.9	<i>mutT</i> gene polymorphism	18
2.2.10	Genotyping for specific resistance mutations	18
2.2.11	Sanger sequencing of the <i>inhA</i> promotor	19
2.2.12	Whole genome sequencing	19
2.2.13	Growth rate and TNF induction in macrophages	19
2.2.14	Single nucleotide polymorphism in the <i>narGHJI</i> operon	20
2.3	RESULTS AND DISCUSSION	21
2.3.1	Paper I	21
2.3.2	Paper II	23
2.3.3	Paper III	25
2.3.4	Paper IV	27
2.3.5	Paper V	31
2.4	CONCLUDING REMARKS	34
3	ACKNOWLEDGEMENTS	37
4	REFERENCES	38

LIST OF ABBREVIATIONS

AFRI	<i>Mycobacterium africanum</i> lineage
ATP	Adenosintriphosphate
BCG	Bacillus Calmette-Guérin
CAS	Central Asian Strain
CFU	Colony-Forming Units
DNA	Deoxyribonucleic Acid
DOTS	Directly Observed Treatment Short course
DST	Drug Susceptibility Testing
DVR	Direct Variable Repeat
EAI	East-African-Indian
EB	Ethambutol
ELISA	Enzyme-Linked Immunosorbent Assay
H	Haarlem
HIV	Human Immunodeficiency Virus
INH	Isoniazid
IS	Insertion Sequence
LAM	Latin-American-Mediterranean
LJ	Löwenstein-Jensen (egg-medium)
LSP	Large Sequence Polymorphism
M-CSF	Macrophage-Colony Stimulating Factor
MDR	Multidrug Resistant
MIRU	Mycobacterial Interspersed Repetitive Unit
MST	Minimum Spanning Tree
MTC	<i>Mycobacterium tuberculosis</i> complex
PBMC	Peripheral Blood Mononuclear Cell
PCR	Polymerase Chain Reaction
PGG	Principle Genetic Group
PZA	Pyrazinamide
RIF	Rifampicin
RD	Region of Difference
RFLP	Restriction Fragment Length Polymorphism
rRNA	Ribosomal Ribonucleic Acid
SIT	Spoligotype International Type
SNP	Single Nucleotide Polymorphism
SM	Streptomycin
Spoligotyping	Spacer Oligonucleotide Typing
TB	Tuberculosis
TCH	Thiophene-2-Carboxylic Acid Hydrazide
TDR	Totally Drug-Resistant
TNF	Tumor Necrosis Factor
WHO	World Health Organization
VNTR	Variable Numbers of Tandem Repeats
XDR	Extensively Drug-Resistant

1 INTRODUCTION

1.1 HISTORY AND GLOBAL HEALTH BURDEN OF TUBERCULOSIS

Tuberculosis (TB) is not a novel disease. Early estimates suggested that the causative agent of TB, *Mycobacterium tuberculosis*, was 15,000 years old^{1,2}, whereas recent evidence suggests that ancestral mycobacteria may have already affected early hominids in East Africa 3 million years ago³. It is further believed that human TB originated in East Africa and was expanded and diversified during its spread out of Africa⁴.

TB is known to be a disease of antiquity, with spinal disease being found in Egyptian mummies 5000 years ago⁵. Although TB may have been common in populations of the Old and New Worlds several hundred years ago, it was during the Industrial Revolution in the 18th century, that it acquired notoriety as ‘the captain of all these men of death’⁶. By the 19th century, TB was a major cause of death in most European countries⁷. Treatment then consisted of fresh air, a good healthy diet and rest, and from this the concept of the sanatorium emerged. During the 20th century a drastic decline in incidence rate took place in Western Europe and USA as social factors and living conditions improved, public health measures were taken, and treatments were made available⁸. TB somehow lost its visibility and the general feeling may have been that TB was a conquered disease and so the interest in the disease therefore declined. The neglect of TB control coupled with the arrival of the human immunodeficiency virus (HIV) epidemic may explain why sub-Saharan Africa in time experienced increased numbers of TB cases. The dissolution of the former Soviet Union resulting in a collapse of the health service may also be a cause of the increase in number of TB cases in that part of the world⁹. **Figure 1** illustrates worldwide estimated TB incidence rates in 2009.

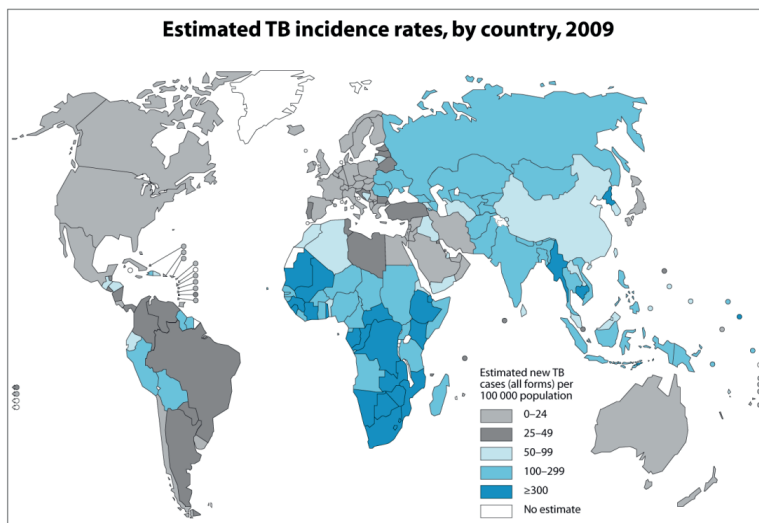


Figure 1 Estimated TB incidence rates by country in 2009. (Source: Global tuberculosis control; WHO report 2010; reproduced with permission from WHO Press.)

In our day, TB is fairly easy to diagnose and when the right combination of medications is made available *and* taken by the patient, the disease can often be cured¹⁰. Despite remarkable achievements TB is still one of the leading causes of death from an infectious disease worldwide. Estimates indicate that one third of the world's population is infected; there are around 9 million new active cases per year and close to 2 million deaths per year¹¹. Although it is a global epidemic, TB predominantly affects the populations of resource-poor countries, where 98% of all TB deaths occur⁸. The great majority (~85%)¹² of the world's total population lives in low- or middle-income countries which is also where 95% of all TB cases occur. These countries are furthermore now being ravaged by the HIV epidemic - the most powerful factor ever known to favour the development of TB¹³.

The rapid emergence of drug-resistant TB has revealed the need for rapid and low-cost methods for detection as well as drug susceptibility testing (DST). The now also emerging multidrug resistant (MDR) and extensively drug-resistant (XDR) strains pose increasing threats to TB control. Although sometimes treatable with alternative drugs, the cost is enormous and accordingly not undertaken in many low-income nations^{8,14}. If XDR-TB patients are treated with inadequate regimens we will probably generate pan-resistant-TB. Clinicians treating XDR-TB patients in South Africa have suggested that we have already entered a stage of totally drug-resistant (TDR) TB¹⁵. TDR-TB strains have previously been reported in Iran where it was found that 95% of the patients with TDR-TB had a previous history of TB¹⁶. The different Variable Numbers of Tandem Repeats (VNTR) profiles dismissed the possibility of recent transmission among the cases observed in this study. In 2010, a World Health Organization (WHO) MDR-TB survey reported the highest rates ever of MDR-TB, with peaks of up to 28% of new TB cases in some settings of the former Soviet Union¹⁷. South Africa is also experiencing an increasing MDR-TB epidemic and it has there been estimated that treatment of MDR-TB cases consumes nearly 70% of the budget allocated to fight the entire TB epidemic in South Africa. Thus, valuable resources are directed away from combating the drug susceptible TB epidemic which has now reached alarming proportions¹⁵. As a result of alarming MDR-TB prevalence rates in Europe the WHO recently presented a consolidated action plan to prevent and combat MDR- and XDR-TB in the European region¹⁸.

Epidemiological studies are important for understanding the transmission of TB and for tracing sources of outbreaks in different settings. Genotyping tools are recommended¹⁰ to elucidate TB risk factors and transmission dynamics. The current study has characterized *M. tuberculosis* complex (MTC) isolates from TB patients in Sweden and Guinea-Bissau with the view to understand species and strain diversity as well as transmission patterns.

1.2 TUBERCULOSIS CONTROL

In 1993, the WHO declared TB a global emergency and launched the Directly Observed Treatment Short course (DOTS) strategy for global TB control. DOTS has been recognized as highly efficient and cost-effective and countries applying this strategy on a wide scale have witnessed remarkable results¹⁹. The programme comprises five components²⁰:

- 1) Sustained political and financial resources for TB control.
- 2) Diagnosis by quality ensured sputum-smear microscopy.
- 3) Standardized short-course anti-TB treatment given under direct observation.
- 4) A regular, uninterrupted supply of high quality anti-TB drugs.
- 5) A standardized recording and reporting system that allows assessment of programme performance.

The global targets set by the WHO are to detect 70% of infectious cases and to cure 85% of those detected. The number of countries adopting DOTS has increased dramatically over the past years. Unfortunately, even though 148 countries by 2002 had accepted the DOTS strategy only 27% of infectious pulmonary TB cases were treated under DOTS programmes¹⁹.

1.2.1 Tuberculosis in Sweden

Today, Sweden is a high-income nation with a low prevalence of TB. However, less than a century ago, the prevalence of TB in the Nordic countries was among the highest in the world. In 1905, the estimated overall TB incidence in Sweden was 890/100,000 population²¹; higher than in most high incidence countries in sub-Saharan Africa today. During the first half of the 20th century immigration to Sweden was limited and occurred mainly from bordering countries. During the following decades immigration increased and in 2010 85% of all TB cases were born abroad, mainly sub-Saharan Africa²², compared to 4.8% in 1951, when most foreign-born came from the other Nordic countries²³.

In 1975, the Bacillus Calmette-Guérin (BCG) vaccination policy in Sweden changed from routine vaccination of all newborns to selective vaccination of groups at higher risk. The TB incidence in Sweden was at that time declining, and one of the main reasons for the change in policy was that an increased frequency of BCG vaccine induced osteomyelitis had been noticed. The now increasing proportion of foreign-born TB patients points to the need to continue selective vaccination of children in families originating from countries with high TB incidence²⁴.

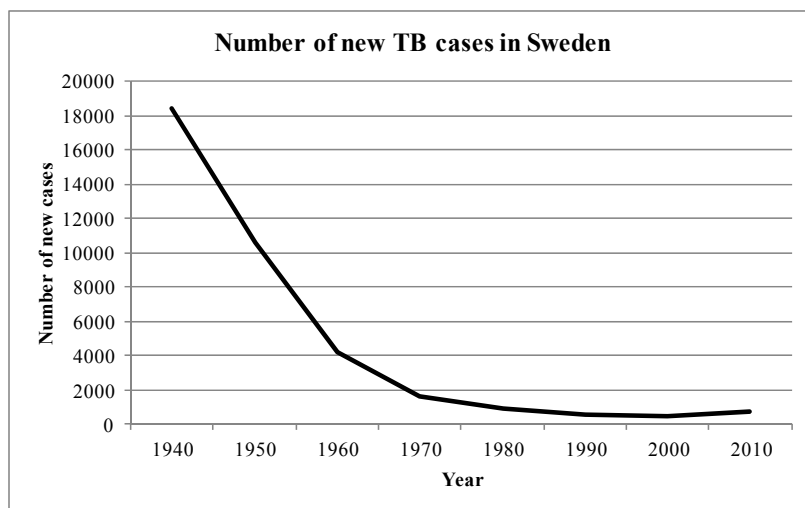


Figure 2 Number of new TB cases in Sweden between the years 1940 and 2010.

Sweden has nowadays one of the lowest TB incidences in the world. Less than a century ago however, the situation was much different as depicted in **Figure 2**. **Figure 3** looks closer at the number of new TB cases in Sweden during the past 21 years. From 1990 to 2003 there was a decrease in incidence that could be ascribed to a decrease among Swedish-born cases. During the past few years however the number and proportion of foreign-born TB patients among all cases have increased. The particularly observed increase of TB in 2005 was only in part attributed by an outbreak at a day nursery, but is still a reminder of the serious consequences of delayed diagnosis^{25,26}.

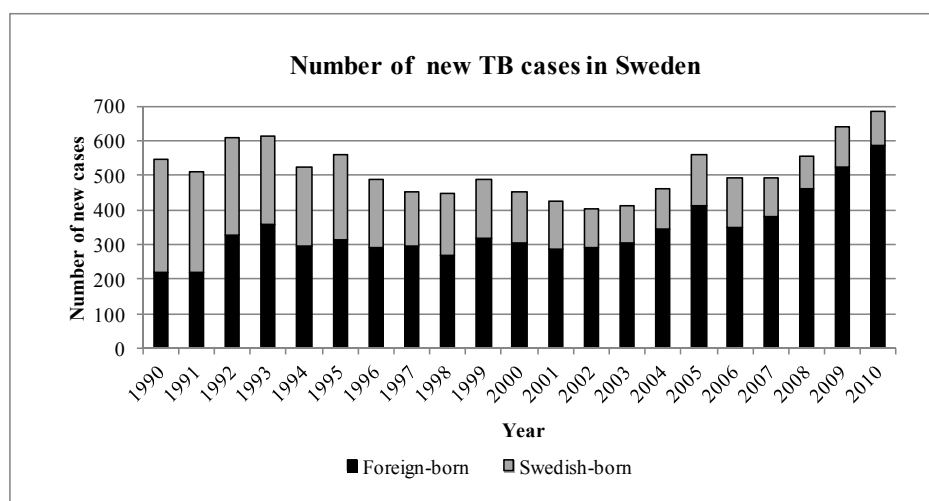


Figure 3 Number of new TB cases in Sweden among foreign-born (black) and Swedish-born (grey) between the years 1990 and 2010.

In 2010, 683 new cases (584 foreign-born and 99 Swedish-born) of TB were reported generating an overall incidence of 7.3 cases per 100,000 population. The increase of 6% compared to 2009 was especially seen among foreign-born patients and highlights the need for physical examinations of individuals entering Sweden. Also, screening for latent TB may be important to evaluate possible prophylactic treatment. The great majority (85%) was foreign-born and several (55%) originated from Africa, in particular Somalia. The high TB incidence among foreign-born patients much reflects the incidence in the countries these patients originate from. TB in the Swedish-born population has over the years decreased and is today rare. Last year the median age of Swedish-born TB patients was 70 years for women and 67 years for men²². Most of these patients were in all likelihood infected in Sweden during the early or middle parts of the 1900s but it is important to remember that domestic spread of infection is occurring both among children and adults. The greatest danger in a country with low incidence of TB is that the diagnosis might be neglected²⁷.

For the last four years an increase of MDR-TB has been noted with for instance 18 new cases in 2010, compared to about four new cases per year in 1994-2006²².

Although the DOTS programme has showed promising results in many countries it is only on exceptional cases basis employed in Sweden^{28,29}.

1.2.2 Tuberculosis in Guinea-Bissau

The republic of Guinea-Bissau is a small West African country situated on the Atlantic coast, south of Senegal and northwest of Guinea Conakry. Guinea-Bissau has an estimated population of 1.6 million inhabitants that face the low life expectancy of 49 years^{30,31}. Since independence from Portugal in 1974, Guinea-Bissau has experienced considerable political and military turmoil. In 1998, the country and especially the capital Bissau suffered a civilian war that after a peace agreement in 1999 demanded a challenging rebuilding of the country.

In Guinea-Bissau TB is a common disease. In 2009, the WHO estimated the TB incidence to be 229 cases per 100,000 population³². The epidemic of HIV infections in Guinea-Bissau is also believed to further increase the incidence of TB. Guinea-Bissau first presented the highest prevalence of HIV-2 in the world. Later, declining prevalence rates of HIV-2 along with increasing prevalence rates of HIV-1 were reported³³.

The national TB control programme in Guinea-Bissau has regrettably been having difficulties in its function, for instance in periods with lack of anti-TB drugs. During the above mentioned war the main service infrastructure, including the reference TB hospital, and the TB laboratory at the Laboratório Nacional de Saúde Pública were destroyed. DOTS has been implemented but the global targets set by the WHO has not been reached.

1.3 MYCOBACTERIUM TUBERCULOSIS COMPLEX

The MTC consists of a highly related group of acid-alcohol-fast bacilli which are human and animal pathogens³⁴. Detection of the different species within the complex has mainly been based on the analysis of phenotypic characteristics such as acid-fastness, colony morphology, growth rate and biochemical tests. Nucleic acid probe technologies, DNA fingerprinting techniques and genotyping methods have nowadays made epidemiological studies and rapid species discrimination more achievable. These techniques have also widened our understanding of phylogenetic relations and of the evolutionary origin of the members of the MTC. Genomically, species of the MTC present near-identical sequence homology, but are most notably differentiated by large sequence polymorphisms (LSPs)³⁵. The MTC includes seven members where *M. tuberculosis* is the primary causative agent of human TB; *M. bovis* explains bovine TB and includes the vaccine strain *M. bovis* BCG; *M. africanum* often causes TB in West Africa³⁶; *M. canettii* appears be linked to the Horn of Africa^{37,38}; *M. microti* infects voles^{39,40}; *M. pinnipedii* causes disease in seals⁴¹ and *M. caprae* in goats⁴².

1.3.1 *Mycobacterium tuberculosis*

This organism, first described by Koch⁴³ in 1882, has proven to be one of the most successful bacterial pathogens in humans. *M. tuberculosis*, the predominant cause of human TB, is mainly spread by small air droplets containing the bacteria. Pulmonary TB may be caused by inhalation of these small droplets. *M. tuberculosis* survives and multiplies within macrophages and the infection can persist in a latent phase for several years. The bacillus is a slowly growing, lipid-rich and acid-fast bacterial rod that requires about 18 hours to undergo one cycle of replication⁴⁴. It was previously speculated that *M. tuberculosis* evolved from *M. bovis* by specific adaptation of an animal pathogen to a human host⁴⁵. Conversely, comparative genomics later uncovered several variable genomic regions in the members of MTC. Based on the presence or absence of a *M. tuberculosis* specific deletion (TbD1), *M. tuberculosis* strains can be divided into ancestral and modern strains. In addition, successive loss of DNA, reflected by region of difference (RD) 9 and other subsequent deletions, has been identified for an evolutionary lineage represented by *M. africanum*, *M. microti*, and *M. bovis* that diverged from the progenitor of the present *M. tuberculosis* strains before TbD1 occurred. *M. canettii* and ancestral *M. tuberculosis* strains lack none of these deleted regions and therefore seem to be direct descendants of an ancestral tubercle bacilli. Consequently, the common ancestor of the tubercle bacilli could well have been a human pathogen already⁴⁶.

1.3.2 *Mycobacterium bovis* and *Mycobacterium bovis* BCG

The bovine tubercle bacillus *M. bovis* causes disease among a wide range of wild and domestic animals but also in humans where the disease cannot be distinguished from an infection with *M. tuberculosis*. In industrialized countries, animal TB control and elimination programs, together with milk pasteurization, have drastically reduced the incidence of disease caused by *M. bovis* in both cattle and humans. In resource-poor countries however, animal TB is widely distributed, control measures are not applied or are applied sporadically, and pasteurization is rarely practiced⁴⁷. The world wide used live attenuated vaccine strain *M. bovis* BCG has its origin in a virulent *M. bovis* strain

that lost its virulence in 1921 after 230 *in vitro* passages⁴⁸. The BCG vaccine is the most widely used vaccine worldwide despite showing highly variable efficacy⁴⁹. BCG is believed to efficiently protect children against the more serious forms of TB⁵⁰⁻⁵² but estimates of protection against adult pulmonary TB range from 0–80% and the lowest levels of protection have been found in countries with the highest incidence of TB⁵³. This variability has been attributed to many different factors, for instance strain variation in BCG preparations, genetic, nutritional or socioeconomic differences between populations, poor cold-chain maintenance, or exposure to environmental mycobacterial infections.

1.3.3 *Mycobacterium africanum*

This heterogeneous group of isolates responsible for human TB in Africa was first⁵⁴ isolated in a Senegalese patient in 1968. *M. africanum* has traditionally been defined on the basis of biochemical properties that place it at an intermediate position between *M. tuberculosis* and *M. bovis*. *M. africanum* has then been further subdivided into subtype I (West Africa) and subtype II (East Africa). *M. africanum* subtype II has however shown to correspond to a particular sublineage of *M. tuberculosis*^{4,35}. West African *M. africanum* can further be divided into two lineages where lineage 1 lacks RD711 and lineage 2 lack RD7, RD8, RD9, RD10, RD701 and RD702. These two lineages are responsible for almost half of the TB patients in West Africa^{55,56}. Although the host range of *M. africanum* is thought to be human, it is generally isolated at a much lower frequency than *M. tuberculosis*. This may represent a geographically restricted pathogen³⁵. Recent studies illustrate that *M. africanum* appears in various frequencies in West African countries. It has been reported that the proportion of *M. africanum* isolates have drastically decreased during the last decades in Cameroon⁵⁷ and Burkina Faso⁵⁸. DNA from both *M. africanum* and *M. tuberculosis* was recovered from Egyptian mummies⁵⁹, yet today *M. africanum* isolates are rarely found outside of West Africa except among first generation immigrants⁶⁰. This lack of spread of *M. africanum*, despite large migrations such as the slave trade to the New World that lasted several centuries, suggests that *M. africanum* has established a specific geographic niche in West Africa⁵⁶.

1.3.4 *Mycobacterium canettii*

M. canettii is a very rarely encountered strain that was earliest isolated in 1969 from a French farmer by Canetti. Isolates of *M. canettii* are characterized by a highly particular growth pattern, and colonies that appear smooth and glossy. All known cases of TB caused by *M. canettii* seem to have a connection to the Horn of Africa⁶¹. Although it shares identical 16S rRNA sequences with the other members of the MTC, *M. canettii* strains differ in many respects, including polymorphisms in certain house-keeping genes, insertion sequence (IS) 1081 copy number, colony morphology and the lipid content of the cell wall⁴⁶. The smooth and glossy colonies produced are highly exceptional for this species. This smooth phenotype is however unstable and can non-reversibly switch to a rough colony morphology³⁸. Deletion analyses have shown that RD, RvD (H37Rv related deletion) and TbD1 regions are conserved in the genome of *M. canettii*. In conjunction with the many described and observed differences it has been suggested that *M. canettii* diverged from the common ancestor of the MTC before RD, RvD, and TbD1 occurred in the lineages of tubercle bacilli. This hypothesis is

further supported by the finding that *M. canettii* carries 26 unique spacer sequences in the direct repeat region that are no longer present in any other member of the MTC⁴⁶.

1.3.5 *Mycobacterium microti*

M. microti is the vole bacillus discovered by Wells⁶² in 1937. It is mainly found in small rodents, but infection has also been documented in cats, pigs and llamas^{63,64}. The incidence of *M. microti* is probably underestimated as it grows poorly on traditional solid egg media, and modern automated liquid culture techniques do not seem to yield better results. Very little is known about the incidence and ecology of *M. microti* infection in farm and domestic animals. It was considered to be non-pathogenic for humans but recently there are, though few, reports on infections in humans. Human-to-human transmission of *M. microti* infection seems however rare⁶⁵. Based on biochemical properties, this bacterium is difficult to distinguish from *M. tuberculosis*, *M. africanum*, or *M. bovis* but *M. microti* strains have been shown to display characteristic IS6110 banding patterns and spoligotypes, distinct from types previously observed in other MTC strains. By spacer oligonucleotide typing (spoligotyping) several genotypes of *M. microti* have been recognized where the llama-type (presence of spacers 4-7, 23, 24, 26, 37, 38) and the vole-type (presence of spacers 37-38) have been well described; both types involved in human infections^{40,64}.

1.3.6 *Mycobacterium pinnipedii*

This organism was first isolated from cases of TB in Australian sea lions and seals⁴¹. *M. pinnipedii* transmission to humans has been reported in a zoo where animal keepers had been cleaning sea lions' enclosures⁶⁶. With biochemical testing it has been confirmed that the seal bacillus, isolated from pinnipeds, belongs to the MTC. Slight differences from typical *M. bovis* isolates were noted, in that the cord formation observed after Ziehl-Neelsen staining was loose and that the isolates were susceptible to pyrazinamide (PZA). The MPB70 antigen considered to be characteristic of *M. bovis* was not found among the seal isolates. The *gyrA* and *katG* gene sequences, utilized to confirm that isolates belong to the MTC, were identical in all seal isolates. When sequencing the *mtp40*, *pncA* and *oxyR* genes it was clearly demonstrated that the seal isolates were genetically more consistent with *M. tuberculosis* and *M. africanum* than with *M. bovis*. The isolates presented a distinct spoligotype pattern when compared to other members of the MTC⁶⁷.

1.3.7 *Mycobacterium caprae*

These mycobacteria first isolated from goats in Spain⁶⁸ possess clear traits that differentiate them from the classical species of the complex. Besides biochemical and epidemiological features, isolates of this unusual member of the MTC show a special combination of *pncA*, *oxyR*, *katG* and *gyrA* gene polymorphisms. Further differentiation amongst caprine mycobacterial isolates and other MTC members can be drawn from *gyrB* gene sequence polymorphism analyses. *M. caprae* isolates have special features by genetic fingerprinting. By means of spoligotyping they form a homogeneous cluster easily recognizable by the absence of spacers 1, 3-16, 30-33 and 39-43. The lack of spacers 39-43 has also been described in *M. bovis* and *M. microti*^{42,68}.

1.4 LABORATORY DIAGNOSIS OF TUBERCULOSIS

1.4.1 Identification of the *M. tuberculosis* complex

1.4.1.1 Microscopy

In many countries, direct microscopy for detection of acid-fast bacilli in sputum specimens remains the main tool for diagnosis of pulmonary TB. Sputum smear microscopy is rapid, inexpensive and technically simple. It can however not distinguish MTC from other mycobacteria. In addition, the sensitivity is low as 5,000 to 10,000 acid fast bacilli per mL of sputum must be present in order to be detected. The Ziehl-Neelsen staining technique where mycobacteria appear as red-stained rods on a blue background is predominantly used for light microscopy⁶⁹. Using a fluorescence microscope mycobacteria can also be stained with auramine⁷⁰ and then appear as yellow fluorescing rods on a dark background.

1.4.1.2 Culture

Culture is still considered the reference method for the detection of mycobacteria. Although it provides high sensitivity and specificity, the slow growth rate of *M. tuberculosis* and other mycobacteria complicates the use of cultivation as a diagnostic technique^{71,72}. In settings with limited resources mycobacterial culture is often not performed routinely as it requires quite advanced laboratories with specific biosafety measures.

1.4.1.3 Phenotypic identification of the *M. tuberculosis* complex

Accurate identification of MTC isolates at the species level is of interest, particularly in Africa where species other than *M. tuberculosis* have been characterized in human TB and *M. bovis* remains a huge problem for cattle. Traditional biochemical tests have been used with some success since the late 1950s to determine the phenotypic characters of some commonly encountered species of mycobacteria. Phenotypic methods of identification relying on colony morphology, oxygen preference, niacin accumulation, nitrate reductase activity, growth kinetics, resistance to thiophene-2-carboxylic acid hydrazide (TCH) and production of pyrazinamidase are hampered by the slow growth of MTC members and subjective interpretation of colony morphology and biochemical tests^{73,74}. High-performance liquid chromatography identification of mycobacteria was introduced into some laboratories during the late 1980s. The principle was based on the analysis of mycolic acids present in the cell wall⁷⁵. Another past method is gas-chromatography, either alone or in combination with mass spectrometry, to identify specific mycobacterial lipids such as tuberculostearic acid in clinical samples. As tuberculostearic acid is present not only in mycobacteria but also in other microorganisms constituting the normal microbial flora the method is unsuitable for the diagnosis of TB in sputum or other nonsterile specimen⁷¹.

1.4.1.4 Genotypic identification of *M. tuberculosis* complex

Molecular biological methods such as direct sequence determination of the 16S rRNA gene⁷⁶ and commercial tests have replaced conventional biochemical tests for the identification of mycobacteria. Assigning an isolate to the MTC can be made possible by using AccuProbe gene probes (Gen-Probe, San Diego, USA), which target the 16S rRNA. But these probes do not allow for identification at the species level^{77,78}. The reverse hybridization assay INNO-LiPA (Innogenetics NV, Ghent, Belgium) which amplifies the 16S-23S rRNA region identifies *Mycobacterium* species and differentiates the MTC⁷⁹. The newer available DNA strip assay GenoType MTBC (Hain Lifescience GmbH, Nehren, Germany) targeting the 23S rRNA can also be used for the differentiation of members of the MTC. This assay is based on *gyrB* DNA sequence polymorphisms and the RD1 deletion of *M. bovis* BCG⁸⁰. Recently it was shown that the sequencing of the Exact Tandem Repeat D (ETR-D; alias Mycobacterial Interspersed Repetitive Unit 4) allowed rapid, accurate, single-step identification of the MTC at the species level⁷³.

Comparative genomics have revealed a number of variable genomic regions in the MTC members. Behr *et al* identified several regions of difference (RD1-16) absent in *M. bovis* BCG compared to *M. tuberculosis* H37Rv⁸¹. Huard *et al* developed a PCR-based typing method with seven primer pairs run in separate but simultaneous reactions using the RD deletion loci to differentiate the species within the MTC⁸². As *M. caprae* and *M. pinnipedii* were recognized as members of the MTC this typing panel was further expanded to accommodate novel LSPs and single nucleotide polymorphisms (SNPs)⁸³.

1.4.2 Strain identification of the *M. tuberculosis* complex

The remarkably homogenous *M. tuberculosis* genome consists of 4.4 megabase pairs and has very densely packed coding regions. The genome is comprised of about 4000 protein-coding sequences and has a very high guanine and cytosine content⁸⁴. There are very few silent mutations in the genome and most recombination seems to occur through transposons. The simplest transposons, insertion sequences (IS), are often used to discriminate different strains.

A large number of different molecular biological methods have been developed to measure the genetic relationship between different MTC strains. Ideally, molecular genotyping tools should be inexpensive, highly discriminative, deliver rapid results, be straightforward to perform, and produce easily interpretable results that allow for accurate interlaboratory comparisons. In order to discriminate bacterial strains as much as possible, the best approach would possibly be whole genome sequencing of each strain. As this is at present too costly and time consuming only parts of the genome are being examined⁸⁵. Each molecular method provides specific genetic profiles referred to as fingerprints. When two or more strains have identical fingerprints they are referred to as the same cluster and may be epidemiologically linked. Molecular epidemiological typing methods constitute powerful tools for identifying outbreaks.

1.4.2.1 RFLP

The method IS6110 Restriction Fragment Length Polymorphism (RFLP) was in the early 1990s agreed upon as the standard typing method for TB epidemiology⁸⁶. IS6110 is present in different copy numbers and integrated at different chromosomal sites in MTC isolates. The fragments based on IS6110 are highly polymorphic but stable enough for epidemiological studies⁸⁷. Strains with fewer copies of IS6110 are more homogenous and the fingerprints of those are not as reliable concerning epidemiological links as of those containing multiple copies⁸⁸. In the RFLP method⁸⁶, genomic DNA is digested with the restriction endonuclease *PvuII*, the fragments are separated by gel electrophoresis and subsequently transferred to a nylon membrane by Southern blotting. The fragments are visualised by hybridizing the membrane with an IS6110 probe and using an enhanced chemiluminescence kit. The radiographs are then scanned and analysed using powerful softwares such as BioNumerics (Applied Maths, Kortrijk, Belgium). For the past decades this method has been the reference standard due to its high discriminatory power and its reproducibility. Nevertheless, the RFLP method which requires large amounts of DNA is labour intensive, time consuming and inconvenient when quick results are needed.

1.4.2.2 Spoligotyping

Spoligotyping is a very widespread method for TB epidemiology. It is a simple, rapid and reproducible method that simultaneously detects and differentiates the MTC without the need of purified DNA. The method is based on polymorphism in the direct repeat locus which consists of multiple direct variable repeats (DVR). Each DVR is composed of a 36 bp direct repeat and a non-repetitive short sequence also called spacer⁸⁹. In short, after using primers that amplify the direct repeat locus the denatured PCR product is hybridized to a membrane on which a set of 43 immobilised synthetic spacer oligonucleotides are covalently bound. Hybridization signals are detected using an enhanced chemiluminescence kit⁹⁰.

Spoligopatterns evolve through successive loss of spacers. Deletions of these spacers need not be independent events and identical patterns may develop in unrelated strains as a result of convergent evolution^{91,92}. The method has however been in use for a long time and many strains have been typed to describe the lineages of the MTC that dominate in various parts of the world. The discriminatory power of spoligotyping is low to medium depending on the geographical setting and a second method truly is needed to underline epidemiological clusters.

Spoligotypes can be assigned to the major phylogenetic lineages according to signatures provided in the international database SpolDB4.0⁹³, which defines 62 genetic lineages/sublineages. Spoligotypes can be entered into the SITVIT2 database⁹⁴ (Pasteur Institute of Guadeloupe), which is an updated version of the previously released SpolDB4.0 database. Up to date (October 2011) the SITVIT2 database contained genotyping information on about 86,000 *M. tuberculosis* clinical isolates from 160 countries of origin. In this database, SIT (Spoligotype International Type) designates spoligotypes shared by two or more patient isolates, as opposed to “orphan” which designates patterns reported for a single isolate.

1.4.2.3 MIRU-VNTR

This PCR-based method, providing a digital number as result, analyses multiple loci containing variable numbers of tandem repeats (VNTR) of different families of interspersed genetic elements collectively called mycobacterial interspersed repetitive units (MIRU). PCR products of the tandem repeat loci are sequenced to reveal the number of tandem repeats and the size of the DNA fragments located at both sides of the repeats^{95,96}. The discriminatory power of VNTR is low compared to IS6110 RFLP and should for epidemiological studies be used in combination with another typing method. Nonetheless, for evolutionary studies VNTR may be useful⁸⁵.

Supply *et al* proposed and evaluated the use of 12 MIRUs as strain differentiation markers^{97,98}. Later, a standardized format comprising 24 loci was proposed and evaluated^{99,100}. The MIRU technique has been shown to be reproducible and has a discriminatory power that resembles the IS6110 RFLP method⁸⁵. MIRU is also said to be a fast but sophisticated and costly method that requires sequencing technologies perhaps not available in resource-poor countries. It has been proposed (and in some countries implemented) that a standardized MIRU-VNTR genotyping method should be the new reference for epidemiological and phylogenetic screening of *M. tuberculosis* strains^{99,100}. In high TB-burden countries the use of genotyping is often complicated by dominance of geographically specific, genetically homogeneous strain lineages. A number of studies have revealed the problematic of the differentiation power when it comes to strains of the Beijing genotype¹⁰¹⁻¹⁰³. These findings suggest that the selection of MIRU-VNTR loci for optimal differentiation of *M. tuberculosis* requires further validation in different geographical settings. On the other hand a study in the high incidence country Brazil recently showed that 24-loci MIRU-VNTR typing combined with spoligotyping can reveal epidemiologically meaningful clonal diversity behind a dominant *M. tuberculosis* strain lineage. It was concluded that this combined typing may be an effective strategy for molecular epidemiological applications in Brazil and probably in the many other regions of the world where Latin-American-Mediterranean (LAM) strains are common¹⁰⁴.

1.4.2.4 LSP and SNP

The above mentioned genotyping tools make use of mobile (IS6110 RFLP) and repetitive (spoligotyping and MIRU-VNTR) DNA elements. Even though these tools have been invaluable for detecting ongoing TB transmission, the markers upon which they are based change relatively rapidly, making it difficult to define deep phylogenetic relationships. Because of this rapid change, identical fingerprinting patterns can emerge in unrelated strain lineages as a result of convergent evolution. To define phylogenetic associations clearly, genetic markers need to be unique and ideally irreversible. Such phylogenetically informative mutations have been identified in *M. tuberculosis* in the form of LSPs and SNPs^{4,55}. SNPs are likely to be a more exact tool for phylogenetic studies as they are less prone to converge (as can be the case with spoligotype or MIRU markers) or distort by selective pressure (as can take place in LSPs)¹⁰⁵. The MTC PCR typing panel developed by Huard *et al*⁸² was originally intended for use in the confirmation of MTC subspecies identity and as a control for cross-species contamination and other laboratory errors but SNP genotyping has provided insights into phylogenetic relationships among members of the MTC^{83,106}.

1.4.2.5 Whole genome sequencing

Whole genome sequencing provides the most comprehensive collection of a strain's genetic variation. Although currently this might be expensive to conduct on a large scale, many envision a shift towards this technology due to falling costs. To investigate whether genome sequencing yielded more useful markers than those currently used to study the epidemiology of TB, Schürch *et al*¹⁰⁷ applied it to three isolates of the Dutch Harlingen outbreak. Their findings suggested that SNPs can be used to identify transmission chains within RFLP clusters.

There are different approaches for assembling shotgun reads into longer contiguous sequences. De novo assembly is the only option if a closely related genome sequence does not yet exist. Here, the sequence reads are compared to each other then overlapped to build longer contiguous sequences. An alternative approach, reference based assembly, involves mapping each read to a reference genome sequence, then building a consensus sequence that is similar but not necessarily identical to the backbone reference. Obviously, the quality of a genome assembly, whether de novo or reference based, is the most important factor for the subsequent analyses of sequence variation between individual genomes¹⁰⁸.

For this thesis the Roche-454 technology (<http://www.454.com>) was used and is therefore mentioned in more detail. The Roche-454 technology is derived from the technological convergence of pyrosequencing and emulsion PCR. It relies on fixing nebulized and adapter-ligated DNA fragments to small DNA-capture beads in a water-in-oil emulsion. The DNA fixed to these beads is then amplified by PCR. Each DNA-bound bead is placed into a well on a PicoTiterPlate, a fiber optic chip. A mix of enzymes such as DNA polymerase, ATP sulfurylase, and luciferase are also packed into the well. The PicoTiterPlate is then placed into a GS FLX System for sequencing¹⁰⁹.

1.4.3 Lineages of the *M. tuberculosis* complex

Spoligotypes can be assigned to phylogenetic lineages which are often named after regions, countries, cities or places of high prevalence. The LAM lineage is dominating in South America, the Central Asian Strain (CAS) lineage is prevalent in the Middle-East, Central Asia and South-Asia while the East African Indian (EAI) lineage is more common in South-East Asia. Other lineages have been simply designated by the letters S, T, U and X where the T lineage is a common rather ill-defined lineage found all over the world⁹³. The Beijing lineage represents a large proportion of the strains in Far East-Asia but is now emerging globally. In Europe, many isolates belong to the Haarlem lineage but Haarlem strains have also been found to a large extent in Central America and the Caribbean suggesting a link to the post-Columbus European colonization^{93,110}.

The Beijing genotype, a highly virulent strain that originated out of China¹¹¹, has on many occasions been associated with drug resistance^{112,113}. Although endemic Beijing strains in China may be pan-susceptible an association between Beijing strains and multidrug resistance¹¹⁴⁻¹¹⁶ has lead to concern that they may possess the ability to

acquire drug resistance. Beijing strains have been observed in institutional^{117,118} and nosocomial¹¹⁹ outbreaks and in ongoing community transmission^{112,120,121}.

As previously mentioned *M. tuberculosis* can by the presence or absence of a *M. tuberculosis* specific deletion (TbD1) be divided into ancient (TbD1 positive) and modern (TbD1 negative) strains⁴⁶. This classification superimposes well with the principal genetic group (PGG) classification¹. This classification links strains to ancestral (PGG1) or evolutionary younger (PGG2/3) groups of the MTC due to SNPs in the *katG* and *gyrA* genes. Examples of lineages belonging to PGG1 are EAI, Beijing and CAS and to PGG2/3 Haarlem, LAM, X and T¹²².

1.4.4 Drug resistant tuberculosis

In order to avoid the development of resistance to the few efficient drugs that are currently available, combinatory regimens are used for treatment of TB infections. If resistance is not suspected or detected the standard four-drug regimen is used. The treatment is initiated with four drugs [isoniazid (INH), rifampicin (RIF), PZA and ethambutol (EB)] for two months and then continued with RIF and INH for the following four months¹²³. In high-income countries PZA is used whereas resource poor countries for the most part use streptomycin (SM). If the TB strain is resistant to one or more drugs, the treatment is if possible altered. As drug-resistant TB requires treatment regimens even more prolonged, patient compliance problems may arise.

There are two forms of drug resistance; primary drug resistance, which is when an already resistant organism is transmitted, and acquired resistance, which is when the organism becomes resistant within the same host due to inadequate treatment. TB strains are defined as MDR if they are resistant to the two most important drugs used for treatment; INH and RIF. When MDR-TB strains develop further resistance to any fluoroquinolon and at least one of the injectable second line drugs (kanamycin, capreomycin or amikacin) they are defined as XDR-TB¹²⁴

There are many methods used for DST, varying from the use of solid- or liquid-culture based methods to DNA-based techniques. The need of rapid and affordable DST techniques has led to the development of alternative methods. One of the tests endorsed by the WHO is Line probe assay. In this method PCR-products of amplified resistance-determining regions are hybridized on a strip and mutations are detected by lack of binding to wild-type probes as well as by binding to probes for the most commonly occurring mutations¹²⁵. Another test, also endorsed by the WHO, is a rapid fully-automated nucleic acid amplification test known as the Xpert MTB/RIF assay¹²⁶. This table-top device, with fully integrated and automated sample preparation, amplification and detection required for real-time PCR, significantly cuts the time it takes to provide an accurate diagnosis of TB and RIF resistance.

2 THE PRESENT INVESTIGATION

2.1 PURPOSE OF THE STUDY

The aim of this study was to phylogenetically and epidemiologically characterize MTC isolates obtained from patients with TB in Sweden and Guinea-Bissau using molecular techniques. This was performed with the view to understand species and strain diversity as well as transmission patterns.

2.1.1 Specific objectives

The specific objectives were:

- to use spoligotyping to assign strains to the major phylogenetic lineages defined in the international spoligotyping database SpolDB4.0 or SITVIT2 (addressed in papers I-V)
- to characterize the molecular epidemiology of strains by means of *IS6110* RFLP (addressed in papers I-V) and 24-loci MIRU-VNTR (addressed in paper II)
- to differentiate Beijing lineage strains into different arbitrary sublineages using molecular makers such as RD deletions, spoligotyping, *IS1547* polymorphisms, *mutT* gene polymorphism and *Rv3135* gene analysis (addressed in paper II)
- to study the genomic pattern of the cluster SMI-049 outbreak strain by whole genome sequencing (addressed in paper III)
- to infect macrophages with the above mentioned outbreak strain and subsequently determine the concentration of TNF from culture supernatants (addressed in paper III)

2.2 MATERIALS AND METHODS

Below is a summary of the general methods used in papers I through V.

2.2.1 Study area

2.2.1.1 Paper I

During the years 1994-2005 all drug-resistant (resistant to at least one of the drugs, INH, RIF, EB or SM) MTC isolates were sent to the reference TB-laboratory at the Swedish Institute for Communicable Disease Control (SMI). These isolates studied in paper I had been obtained from all Swedish TB-laboratories situated in Stockholm, Gothenburg, Malmö/Lund, Linköping and Umeå.

2.2.1.2 Paper II

As for paper I all drug-resistant isolates were studied but this time for the period of 1994-2008. All drug-resistant isolates of the Beijing genotype were subjected to further analysis.

2.2.1.3 Paper III

As for papers I and II all drug-resistant isolates were studied but this time for the period of 1994-2010. Three isolates were selected from the cluster SMI-049 outbreak and subjected to further analysis.

2.2.1.4 Paper IV

All isolates collected at SMI during the years 1994-2009 belonging to patients born in Sweden before 1945 were studied.

2.2.1.5 Paper V

At three separate time periods isolates were collected from Guinea-Bissau. During the years 1989-1994 samples were collected from patients referred to the reference TB hospital Raoul Follereau in Bissau. Secondly, during the years 1994-1998 samples were collected from the above mentioned reference TB hospital as well as from two different health stations (Bissorá and Catão). Thirdly, in 2007 and 2008 samples were again collected from the reference TB hospital. Samples collected during the first, second and late part of the third study were processed at the Laboratório Nacional de Saúde Pública in Bissau while samples from the early part of the third study were processed at the TB-laboratory at the Karolinska University Hospital, Stockholm, Sweden.

2.2.2 Sample processing

Samples were decontaminated of nonmycobacterial microorganisms by the sodium lauryl sulfate method¹²⁷ and an aliquot of the specimen was then inoculated onto both conventional Löwenstein-Jensen egg medium (LJ) as well as LJ supplemented with 0.6% pyruvate. The samples were incubated at 37°C and examined weekly for seven weeks. Growth of mycobacteria was confirmed by microscopic observation of acid-fast bacilli. In Sweden, mycobacteria were stained with auramine^{70,128} and in Guinea-Bissau the Ziehl-Neelsen method was used.

2.2.3 Chromosomal DNA isolation

In short, mycobacteria were harvested, heat killed at 80°C for 20 minutes and then subjected to repeated freeze thawing. Bacteria were resuspended in TE (Tris; EDTA) buffer and lyzed for two hours at 37°C. Incubations were made at 65°C with sodium dodecyl sulphate, proteinase K and finally cetyl trimethyl ammonium bromide. A mixture of chloroform-isoamyl alcohol was added and DNA was at last precipitated using isopropanol. The pellet was centrifuged, washed with 70% ethanol and redissolved in TE buffer.

2.2.4 Spoligotyping

Spoligotyping, previously described in the introduction, relied on the amplification of the direct repeat region so as to obtain hybridization patterns of the amplified DNA using multiple synthetic spacer oligonucleotides. In short, standard spoligotyping was performed using the DRa and biotinylated DRb primers to amplify the whole direct repeat region by PCR. The denatured PCR product was pipetted into the parallel channels in such a way that the channels of the miniblotted apparatus were perpendicular to the rows of immobilized oligonucleotides. Hybridization signals were detected using an enhanced chemiluminescence kit. The spoligotypes were compared to those previously defined in the international spoligotyping database SpolDB4.0 or SITVIT2 and assigned to the major phylogenetic lineages.

2.2.5 RFLP

IS6110 RFLP genotyping, previously described in the introduction, was performed in accordance with standard protocols. Genomic MTC DNA was digested with the restriction endonuclease *Pvu*II, the fragments were separated by gel electrophoresis and subsequently transferred by Southern blotting onto a nylon membrane. Hybridization was performed with a 245-bp PCR fragment of the IS6110 sequence as a probe, which was labelled with peroxidase and later visualized using an enhanced chemiluminescence kit.

IS6110 RFLP patterns were analyzed using the BioNumerics software (Applied Maths, Kortrijk, Belgium). The different fingerprint types were compared on the basis of the molecular sizes of the hybridizing fragments and the number of IS6110 copies. The distance between patterns was computed using the un-weighted pair-group method of arithmetic averaging and the Jaccard coefficient.

2.2.6 MIRU-VNTR

Standardized 24-loci MIRU-VNTR, previously described in the introduction, was performed using the MIRU-VNTR typing kit (Genoscreen, Lille, France). The PCR product were run with 1200 LIZ size standard on ABI3131xl sequencers. Sizing of the PCR-fragments and assignments of MIRU-VNTR alleles were done with the GeneMapper software version 4.0 (Applied Biosystems) according to the manufacturers' instructions.

2.2.7 RD polymorphism

Genomic deletions provide information on the diversity and frequency of polymorphism in mycobacterial populations¹²⁹. These deletions could therefore be connected to other genotyping techniques to study the evolution of the MTC. As previously mentioned comparative genomics have revealed differences in the presence or absence of RDs and TbD1 between *M. tuberculosis* H37Rv and *M. bovis* BCG. These RD patterns can help us understand the historical origins and genealogy of the present-day mycobacterial populations^{46,81,130}.

In paper II the genomic deletions RD105, RD142, RD150, and RD181 classifying the Beijing lineage were studied and in paper V RD702, which delineates the branch of *M. africanum* West African 2 lineage, was considered. The different PCR products were analyzed by agarose gel electrophoresis.

2.2.8 Rv3135 gene analysis

In paper II size variation in the *Rv3135* gene encoding a member of the Pro-Pro-Glu (PPE) family of *M. tuberculosis* proteins was studied by PCR. Typical Beijing strains are believed to have the 1.02 kb *Rv3135* gene whereas atypical Beijing strains have the 1.97 kb *Rv3135* gene^{131,132}.

2.2.9 *mutT* gene polymorphism

Another source of polymorphism, characteristic and unique within the Beijing lineage, is represented by single nucleotide substitutions in the genes *mutT4* and *mutT2*, coding for DNA repair enzymes^{133,134}. It has been proposed that mutations in DNA repair genes would confer a mutator phenotype allowing an increased mutation rate, which in turn might lead to selective advantages to harmful conditions, such as the exposure to antimycobacterial drugs or immune system effector mechanisms¹³⁵. On the other hand, it has been shown that Beijing genotype strains are not more prone to gaining drug resistance compared to non-Beijing strains¹³⁶.

In the most ancient evolutionary lineages, the genes *mutT4* and *mutT2* were in wild type configuration; the *mutT4* mutation was acquired subsequent to the RD181 deletion in a progenitor strain that in turn gave rise to a sublineage bearing the *mutT2* mutation¹³³. In paper II mutation analysis of *mutT2* and *mutT4* was performed by sequencing these genes.

2.2.10 Genotyping for specific resistance mutations

In paper II sequencing of the *katG* gene and *inhA* gene (INH-resistance) and *rpoB* gene (RIF-resistance) was performed.

In paper III the INH-resistance was investigated by sequencing the *inhA* gene. A -15C to T transition in the promoter region of the *inhA* gene has often shown to be associated with low level INH-resistance¹³⁷.

2.2.11 Sanger sequencing of the *inhA* promotor

In paper III the differences found in the generated whole genome sequences were verified by Sanger sequencing. A fragment of the *inhA* promotor region was sequenced to look at specific variations in the *polA*, *PPE55* and *cyp138* genes.

2.2.12 Whole genome sequencing

Massive parallel sequencing using the 454-technology, previously described in the introduction, was performed on three isolates from the cluster SMI-049 outbreak (paper III). Chromosomal DNA of the isolates was extracted as for the RFLP analysis except for two additional extractions with chloroform-isoamyl alcohol. DNA was sequenced using the 454-FLX technology following the manufacturer's instructions.

Assembly and analysis of 454-sequencing data was performed with the CLC Genomics Workbench v3.7. Assemblies were performed for each isolate individually and also as a combined assembly including all sequence data for a deeper coverage of the common features of the cluster SMI-049 isolates. LSPs between cluster SMI-049 isolates and the *M. tuberculosis* reference strain H37Rv were identified by detecting assembly positions with partially matching reads and the source of variation elucidated by a combination of de novo assembly of non-assembled reads, reference assembly of the SMI-049 data on the other three fully sequenced *M. tuberculosis* genomes H37Ra, CDC1551, and F11 and manual sequence identification of partially matching reads.

SNP analysis was performed for each sequenced isolate after the LSPs had been introduced into a modified H37Rv-SMI049 reference genome to avoid false calls related to assembly problems at locations of large rearrangements. The SNP results for the three sequenced genomes were compared and all SNP calls that were not found in all three or had a variation frequency <90% were manually verified in the assemblies.

2.2.13 Growth rate and TNF induction in macrophages

M. tuberculosis is usually transmitted via aerosols and establishes infection in the lung. One of the first cell types to encounter the bacilli is alveolar macrophages within which *M. tuberculosis* can survive and proliferate. Macrophages activated by the infection can become efficient effector cells that express microbicidal substances and cytokines, of which tumor necrosis factor (TNF) contributes to promote bacterial killing and granuloma formation^{138,139}. Although activated macrophages usually kill intracellular bacteria, mycobacteria may persist. Some macrophages kill the organisms, others undergo necrosis and release their bacterial contents, and still others harbor dormant bacteria for long periods.

In paper III peripheral blood mononuclear cells (PBMCs) were isolated from buffy coats of blood samples from healthy donors by density gradient centrifugation. Monocytes were separated from PBMCs using CD14-positive magnetic beads, subsequently cultured in the presence of Macrophage Colony-Stimulating Factor (M-CSF) and differentiated into macrophages. The macrophages were seeded in

24-well plates or chamber slides and after 24 hours infected with *M. tuberculosis* H37Rv or one of two cluster SMI-049 isolates. After the cells had been washed to remove extracellular bacteria they were cultured and harvested on day 0, 1, 3 and 6. At the specified time points infected macrophages were lysed and plated onto 7H11 Middlebrook agar and colony-forming units (cfu) were enumerated after 3-4 weeks. The macrophages in chamber slides were fixed and intracellular acid fast bacteria were visualized using Kinyoun stain.

Upon infection, macrophages secrete a variety of cytokines and other molecules¹⁴⁰. To determine if this differed between *M. tuberculosis* H37Rv and the cluster SMI-049 isolates culture supernatants were harvested at different time points and the concentration of TNF was determined using an ELISA.

2.2.14 Single nucleotide polymorphism in the *narGHJI* operon

A selection of Guinea-Bissau isolates (paper V) with different nitrate activity was investigated by the single nucleotide polymorphism at position -215 within the nitrate reductase (*narGHJI*) operon. The fact that *M. tuberculosis* rapidly reduces nitrate, leading to the accumulation of nitrite, has been utilized to differentiate *M. tuberculosis* from other members of the MTC. At nucleotide -215, *M. tuberculosis* carries a thymine, whereas *M. bovis* carries a cytosine residue¹⁴¹. The selected isolates were characterized by PCR-RFLP where the PCR products were digested with the restriction endonuclease *Sau3AI* that produces two bands in case of -215C sequence (*M. bovis*) and one band in case of -215T sequence (*M. tuberculosis*).

2.3 RESULTS AND DISCUSSION

2.3.1 Paper I

The objective of paper I was to characterize the epidemiology of drug-resistant TB in Sweden. In total, 400 isolates from 199 men and 201 women (53 born in Sweden and 347 born abroad) were studied. Of the 400 isolates investigated, 46 were MDR and one was classified as XDR. It was found that 120 patients were culture positive for an extended period of time, allowing 2-10 positive samples (with identical RFLP pattern) from each patient. **Table 1** illustrates that the foreign-born patients originated from several regions, with a clear dominance for Africa (n=213), in particular the Horn of Africa (n=186).

Table 1 Global strain lineages of isolates in relation to the origin of patients (number and region or country of origin of patients with drug-resistant TB)

Lineage	Sweden	Africa	Middle East		Europe	South America	Unknown	Total	Number of clusters	Patients in cluster
			South East	Central Asia						
T1	13	15	2	9	9		1	49	7	23
T2*	13	87		3	3	3		109	3	103
T3		14			2			16	3	12
Beijing	5	6	20	17				48	6	16
EAI	2	21	13					36	3	12
Haarlem	5	13	3	10	3	2		36	5	16
LAM	7	12		4	3	5	1	32	4	10
CAS		24		7				31	2	7
U	2	2	1	3	1			9	2	4
S		1			2	1		4		
X		2						2		
AFRI3		1						1		
Undefined	6	15	2	2		1	1	27		
Total	53	213	41	55	23	12	3	400	35	203

East-African Indian (EAI): EAI1Som; EAI1IND; EAI2Manilla; EAI3; EAI4; EAI5; EAI6BDG1; EAI8MDG. Haarlem (H): H1-H4. Central Asian (CAS): CAS undefined; CAS1KILI; CAS1Dehli. Latin-American-Mediterranean (LAM): LAM1, LAM3-7; LAM9; LAM10; LAM11-ZWE.

* Including all 96 patients in cluster SMI-049.

When the spoligotypes were compared to the ones in the international database SpolDB4.0 it was established that the majority of the isolates were of the T lineage and that the isolates of the foreign-born patients to a large extent reflected genotypes common in their country of origin. For instance, it was found that a considerable majority of Vietnamese and Philippine patients were infected with strains of the Beijing and EAI genotype. Beijing strains, which have disseminated globally in the recent years, have often been associated with drug resistance⁹². In this study, 48 isolates of the Beijing genotype were found of which ten were MDR and one was XDR. The fact that only five Beijing strains were found among the Swedish-born patients indicates that the

Beijing lineage has not been able to disseminate within the country. With more than half of the patients in this study originating from Africa, it was surprising that only one isolate of the AFRI lineage (*M. africanum*) was seen.

When IS6110 RFLP and spoligotyping patterns were combined, 203 (52%) of all 400 isolates fell into 35 different clusters. The size of the clusters varied comprising between 2 and 96 isolates. The majority of the clusters included no more than two individuals but one exceptionally large INH-resistant cluster (SMI-049) comprising 96 individuals was revealed. In total, 25 Swedish-born patients were present in 12 different clusters, of which 23 belonged to 11 clusters that also included foreign-born patients. Eight of these 25 patients were born to immigrants and there was merely one cluster (of two patients) that contained only Swedish-born patients. In four out of 11 clusters a connection could be established between the Swedish- and foreign-born patients. For the most part it was not possible to find an epidemiological link for clustered cases upon conventional contact tracing. In some cases, the epidemiological typing confirmed a link that had already been established by contact tracing. Nevertheless, some unexpected links were revealed as a result of the molecular typing. The remarkably large cluster (SMI-049) identified during the study period²⁸ represents one of the largest ever reported in a low incidence country. Starting in 1996, about two patients were identified per year until 1999 when suddenly 19 new patients were seen. Following that, a more thorough investigation was initiated by the Swedish Board of Health and Welfare and in the end it was confirmed by careful contact investigation that this outbreak was indeed due to active transmission within Sweden²⁵. However, in spite of awareness of the outbreak, and strengthened contact investigations, new cases continued to appear. The majority of patients belonging to cluster SMI-049 were born in Africa (mainly Somalia and Ethiopia), their median age was 26 years and apart from six individuals they all lived in the Stockholm area. Poor compliance was noted among as many as 40% of culture-positive cases²⁵. **Figure 4** illustrates the number of new cases belonging to this cluster identified per year. Up to date (October 2011) 117 patients infected with the SMI-049 strain have been identified.

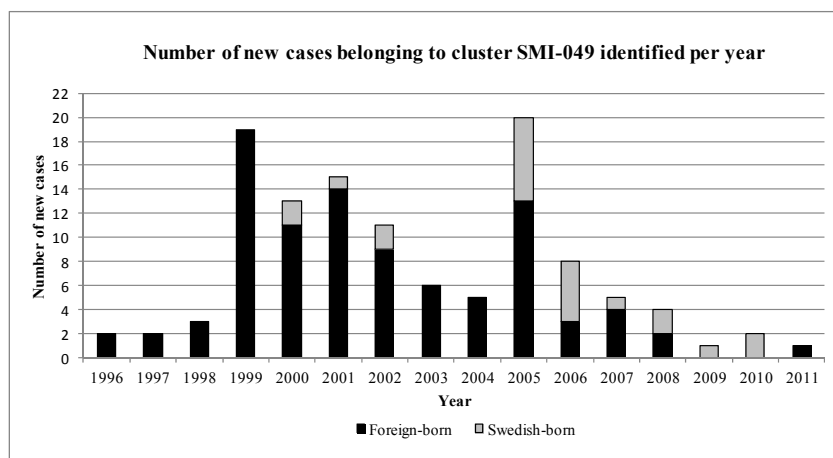


Figure 4 Number of new cases belonging to cluster SMI-049 identified per year. Up to date (October 2011) 117 patients have in total been identified.

2.3.2 Paper II

With paper II the intention was to investigate the presence and transmission of drug-resistant Beijing strains in the Swedish population. Among 536 drug-resistant strains from 535 patients we identified 70 (13.1%) strains of the Beijing genotype as defined by spoligotyping. In one case multiple strains were found in the same patient. This patient had in May 2003 a sputum sample from which three different isolates were cultured, one of which was of a Beijing genotype resistant to INH. In December 2003 the patient presented a sample containing yet another Beijing strain with a different RFLP pattern. This patient came from Azerbaijan and was reported to have been in prison.

Of the 69 patients with drug-resistant Beijing strains six (8.7%) were born in Sweden and 63 born abroad. Of the 466 patients with drug-resistant non-Beijing strains 63 (13.5%) were born in Sweden and 403 born abroad. As depicted in **Table 2**, 42 (67%) of the 63 foreign-born patients infected with drug-resistant Beijing strains were women and 27 were men. Using a one-sided test of hypothesis at 5% significance level, it was also found that patients with Beijing strains were more often women than were those infected with strains of other genotypes. The majority (58%) of the patients with drug-resistant Beijing strains were in the age group 25-44 and 20% in the age group 15-24, (**Table 2**) which is consistent with the age distribution for patients with drug-resistant TB in general in Sweden.

Table 2 Age and sex distribution of 535 patients with drug-resistant strains and relative proportion of patients with Beijing strains.

Age (years)	Male			Female		
	Total	Beijing	%	Total	Beijing	%
0-14	14	0	0	16	1	6.3
15-24	50	4	8.0	63	10	15.9
25-44	143	16	11.2	136	24	17.6
45-64	48	7	14.6	32	4	12.5
65+	12	0	0	21	3	14.3
Total*	267	27	10.1	268	42	15.7

* Patients with Beijing strains were more often women than were those infected with strains of other genotypes (p-value=0.0268, using a one-sided test of hypothesis at 5% significance level).

Of the 69 patients with drug-resistant Beijing strains that were foreign-born a large proportion came from Asia (75%); the majority of these patients were from Vietnam (n=19) and Thailand (n=7). When looking at all (n=536) drug-resistant TB it was found that most of the patients from Asia had strains of Beijing genotype and that only 11 of 280 patients from Africa had Beijing strains. Of the six Swedish-born patients that were infected with drug-resistant Beijing strains three had parents who were born outside of Sweden (Tibet, Hungary and Chile).

While 17 (24%) of the Beijing strains were MDR, only 59 (13%) of 466 drug-resistant non-Beijing strains were MDR during the same period, which is a

statistically significant difference [p-value=0.0134, 95%CI=(0.0137;0.2258)]. One strain was defined as XDR, in this case resistant to amikacin and ofloxacin. When the Beijing strains were further investigated for specific resistance mutations it was for instance found that 96% of the Beijing strains with phenotypic INH resistance showed resistance mutations in the *katG* gene and/or in the *inhA* promoter region.

The collection of Beijing strains isolated from patients originating from various parts of the world illustrates the genomic diversity of the Beijing family. Based on a number of genetic markers the 70 Beijing isolates could be allocated to 11 different sublineages, from more “modern” or “typical” to more “ancient” or “atypical” variants¹³².

Table 3 Polymorphism of *M. tuberculosis* strains of Beijing genotype.

Genomic sublineage	Number of isolates (n=70)	SIT ^c	Region of difference (RD) ^a				Rv3135 ^d	Mutations in <i>mutT</i> genes ^b		IS1547 ^e
			105	181	150	142		<i>mutT2</i>	<i>mutT4</i>	
1	1	265	+	+	+	+	0.15 kb	-	-	8
2	13	1	-	-	+	+	at	-	+	5
3	1	1	-	-	+	+	at	-	+	6
4	1	1	-	-	+	+	at	-	+	7
5	41	1	-	-	+	+	t	+	+	1
6	3	1	-	-	+	+	t	+	+	2
7	1	1	-	-	+	+	t	+	+	3
8	3	265	-	-	+	+	t	+	+	1
9	1	1	-	-	+	-	t	+	+	1
10	4	1	-	-	-	+	t	+	+	1
11	1	1	-	-	-	+	t	+	+	4

Isolates are grouped into 11 arbitrary sublineages on the basis of the polymorphisms relative to genomic deletion RD105, RD181, RD150 and RD142, Rv3135 IS1547 RFLP pattern and mutations in genes *mutT2* and *mutT4*.

^a presence (+) or absence (-) of the specific genomic region.

^b presence (+) or absence (-) of the specific mutation.

^c spoligotype international type.

^d sublineage 1: a 0.15 kb PCR product, sublineage 2-4: “atypical” isolates, with a 1.97 kb PCR product, sublineage 5-11: “typical” isolates, with a 1.02 kb PCR product.

^e pattern 1: three bands (1.7, 2.1, and 2.5 kb), pattern 2: two bands (2.1 and 2.5 kb), pattern 3-7: individual patterns, all containing the 2.5 kb band, except pattern 8 that had no 2.5 kb band.

Table 3 portrays the polymorphisms of the Beijing isolates. The Beijing isolates either corresponded to SIT1 (n=66) or SIT265 (n=4). The majority (n=69) of the isolates had the RD105 and RD181 deletions. PCR of the Rv3135 genome region showed that the majority (n=54) of the isolates had the 1.02 kb gene believed to correspond to “typical” Beijing strains^{131,132}. All except one (n=69) had the *mutT4* mutation and 53 isolates had the *mutT2* mutation indicating the majority of isolates were not of ancient evolutionary

lineages. Eight different *IS1547* RFLP patterns were seen. All isolates (except one) had a 2.5 kb band reported¹³² to be common among Beijing strains. Thirteen isolates showed this single *IS1547* band corresponding to the pattern of “atypical” strains¹³² while a three band (1.7, 2.1, and 2.5 kb) pattern was found for 49 strains, corresponding to the pattern of “typical” Beijing strains¹³².

As identified by *IS6110* RFLP, 28 (41%) of the 69 patients with Beijing strains were found in altogether 10 clusters (2-5 per cluster) while 240 (52%) of 466 patients with non-Beijing strains during the same time period were clustered. All strains that clustered by *IS6110* RFLP were identical also by RD deletions, spoligotyping, *IS1547*, *mutT* gene polymorphism and *Rv3135* gene analysis. Of the six Swedish-born patients two were in clusters, one with two patients from Vietnam, and one with a patient from Bhutan. Contact tracing established possible epidemiological linkage for the latter cluster where the two men had shared the same indoor recreational activity. In the other clusters no epidemiological link was found by conventional contact tracing.

By MIRU-VNTR, 31 (45%) of the 69 patients with Beijing strains were found in altogether 7 clusters (2–11 per cluster). The MIRU-VNTR typing, with fewer and larger clusters, was thus less discriminatory than *IS6110* RFLP. The two strains where a possible epidemiological linkage was established differed in one allele and thus did not cluster in MIRU-VNTR. All strains that clustered by MIRU-VNTR were identical also by RD deletions, *mutT* gene polymorphism and *Rv3135* gene analysis, but not by spoligotyping and *IS1547*. Four of the *IS6110* RFLP clusters contained isolates that differed by MIRU-VNTR. The combination of MIRU-VNTR with RFLP resulted in the disappearance of two clusters, and a reduction of the number of isolates in two clusters, compared to the clustering observed with *IS6110* RFLP clustering alone. As previously reported¹⁴², Beijing strains that exhibit the same fingerprinting pattern can harbor substantial genomic diversity demonstrating that clustering more likely represents older transmission events.

2.3.3 Paper III

In paper I it was reported on the extensive spread of the cluster SMI-049 strain in the community. Despite this great spread, the strain had stayed genetically unchanged since its discovery with regard to drug resistance (except for one case), *IS6110* RFLP- and spoligotyping patterns. The apparent genetic stability together with the extensive spread of this strain in a low incidence country like Sweden led us to further characterize if the strain possessed any particular genetic factors that could explain its epidemiological success and genetic stability. As a result, paper III presented a detailed analysis of the *M. tuberculosis* strain causing the cluster SMI-049 outbreak.

The genomes of the isolate of the index case, and two SMI-049 isolates with isolation dates differing from the index isolate by 9 years, were sequenced by massive parallel DNA sequencing using the 454-platform. In a chain of transmission one of the later isolates (BTB05-552) had been transmitted through at least three patients, more probable four patients, excluding the source case and the patient from whom it was isolated. The other late isolate (BTB05-559) had been transmitted from the source

case through two or three patients before infecting the patient from whom it was isolated²⁵.

Comparison of the generated genome sequences revealed that the three sequenced isolates did not differ by any LSPs. The only differences found were four SNPs in BTB05-559 and a four amino acid in-frame deletion in BTB05-552 compared to the index isolate S96-129. All differences were verified by manual Sanger sequencing. The SNPs in BTB05-559 consisted of a synonymous C→T transition at nucleotide position 579 in the *polA* gene (encoding DNA polymerase I) and three SNPs in the PPE55 gene that all result in amino acid changes (N1496Y, T1517A, I1520V). PPE55 belongs to a group of highly variable *M. tuberculosis* specific proteins with unknown function but that are very immunogenic. The in-frame deletion in BTB05-552 removes amino acids 404-407 (out of 441) of the putative cytochrome P450 protein Cyp138.

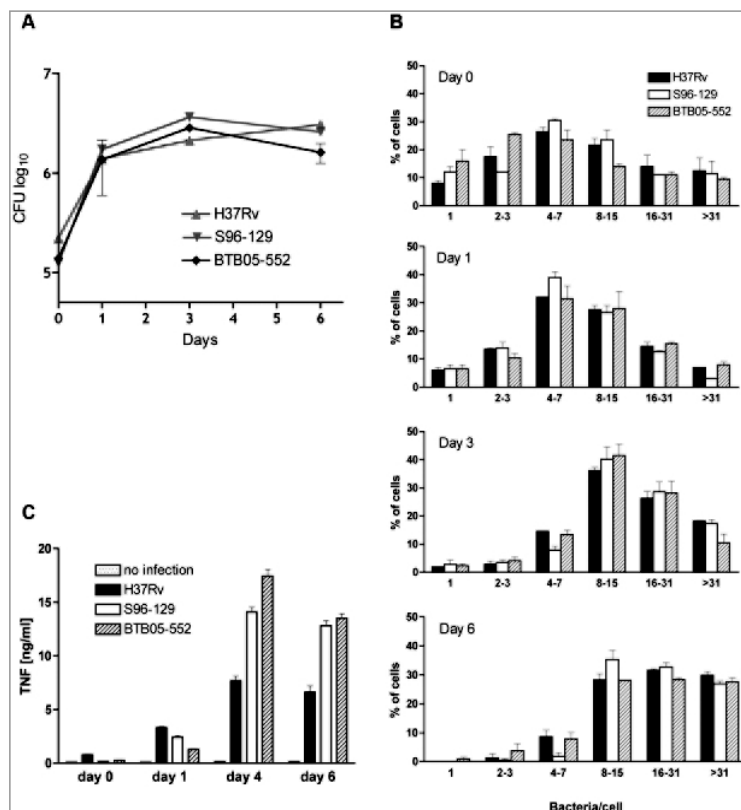


Figure 5 Virulence of SMI-049 isolates in human macrophages.

Monocyte-derived human macrophages were infected with *M. tuberculosis* reference strain H37Rv or cluster SMI-049 isolates S96-129 or BTB05-552, at multiplicity of infection 1:1. Intracellular bacterial growth was determined by enumeration of cfu in plated macrophage lysates (A) or by estimation of proportion of cells containing varying amounts of bacteria (B) over time. TNF induction in infected macrophages was quantified by ELISA using culture supernatants harvested at different time points post-infection (C).

Figure 5 shows that the cluster SMI-049 isolates S96-129 and BTB05-552 grew in macrophages as indicated by a successive increase in cfu in macrophage lysates (**Figure 5A**) and an increase in proportion of cells containing higher numbers of bacteria over time (**Figure 5B**). The growth rate of both SMI-049 cluster isolates was similar and comparable to that of *M. tuberculosis* H37Rv until day 3. Of note was that at day 6 fewer bacteria were recovered from cluster SMI-049-infected macrophages as compared with day 3, while macrophages infected with *M. tuberculosis* H37Rv showed further increment in cfu at that time point (**Figure 5A**). Microscopic analysis of chamber slide cultures revealed that there was increased death of macrophages infected with the SMI-049 isolates on day 6 as indicated by increased number of cells with fragmented cytoplasm and pycnotic nuclei; no increased cell death was observed in *M. tuberculosis* H37Rv-infected macrophages. Interestingly, from day 4 post infection SMI-049-infected macrophages produced much greater amounts of TNF than macrophages infected with *M. tuberculosis* H37Rv (**Figure 5C**).

Phylogenetically the SMI-049 isolates belonged to modern PGG2 and by extended SNP analysis¹⁰⁵ they were found to belong to SNP cluster group 5 (SCG-5), which is predominant in Uganda¹⁰⁵. The RFLP pattern of the cluster SMI-049 strain matches (92%) with a strain (BEA000007341) from Rwanda recorded in the international database of IS6110 RFLP patterns maintained at RIVM, Bilthoven, the Netherlands, also indicating that the strain may have its origin in sub-Saharan Africa. SNPs of the three sequenced genomes placed the SMI-049 strain in the Lineage 4 (red) Europe and Americas lineages^{143,144}, with 2/3 matches to the “Uganda” cluster within the PGG2 group. By spoligotyping all cluster SMI-049 isolates were of the T2 lineage, which is geographically linked to East Africa⁹³ and they were all of the type SIT52. The sequenced isolates also lacked RD724, a polymorphism that defines one major sub-lineage of *M. tuberculosis* commonly seen in Uganda¹⁴⁵ again supporting the concept that the cluster SMI-049 strain has its origin in Central/Eastern Africa.

2.3.4 Paper IV

In paper IV a population-based search for distinct strains of *M. tuberculosis* isolated from patients born in Sweden before 1945 was conducted. The majority of the isolates represented the *M. tuberculosis* population that fueled the TB epidemic in Sweden during the first half of the 20th century. Genetic relationships between strains that caused the epidemic and present day strains were studied by spoligotyping and RFLP. In total, 469 isolates from 280 (60%) men and 189 (40%) women born in Sweden between 1908 and 1945 (**Table 4**) were analysed.

Table 4 Number of patients born in Sweden between the years 1908-1945

Born	Number of patients (%)	Mean (median) age at diagnosis
<1914	47 (10%)	88 (87)
1915-1919	84 (18%)	85 (85)
1920-1924	126 (27%)	81 (81)
1925-1929	78 (17%)	77 (78)
1930-1934	53 (11%)	73 (74)
1935-1939	37 (8%)	67 (68)
1940-1945	44 (9%)	61 (62)

When compared with SITVIT2, the majority (416/469) were shared-types, i.e. had an identical pattern shared by two or more isolates worldwide (within this study, or matching a strain in the SITVIT2 database). A SIT number was attributed to each pattern according to the SITVIT2 database. Twenty-nine patterns corresponded to orphan strains that were unique among the 86,000 strains at that time recorded in the SITVIT2 database and were designated as “Unknown”.

Table 5 illustrates that the absolute majority of the isolates from the elderly population were evolutionary recent PGG2/3 strains (404/469 or 86.1%), and only a low proportion were of ancient PGG1 lineages (36/469 or 7.7%).

Table 5 Number of isolates corresponding to defined Principal Genetic Groups including different lineages found among a total of 469 *M. tuberculosis* complex strains isolated from patients born in Sweden before 1945.

Principal Genetic Group	Lineage	Number of isolates (%)
PGG1	BOVIS, Manu	20 (4.3%)
PGG1	Beijing, EAI, CAS	16 (3.4%)
PGG2/3	T, H, LAM, X, S	404 (86.1%)
Unknown	Unknown	29 (6.2%)

BOVIS: BOV_1; BOV_LIKE, BOV-1, Manu: MANU_ancestor; MANU1; MANU2,
EAI: EAI2-Manilla; EAI4-VNM; EAI5, CAS: CAS1-Dehli, T: T; T1; T1-RUS2; T2; T3; T4; T4-CEU1;
T5; T5-RUS1, H (Haarlem): H; H1; H3; LAM: LAM; LAM1; LAM6; LAM7-TUR; LAM9;
LAM10-CAM; LAM11-ZWE; LAM12-Madrid1, X: X1; X3.

The isolates demonstrated a highly homogenous population where the T- (38.8%), Haarlem- (37.5%) and LAM (7.7%) lineages dominated and only 16% belonged to other lineages. The most common spoligotypes (**Table 6**) were SIT50 of the H3 sublineage, SIT53 of the T1 sublineage and SIT47 of the H1 sublineage and SIT42 of the LAM9 sublineage. In addition to the T1 sublineage prototype, SIT53, two more T clade SITs, SIT52 and SIT153 were among the seven predominant SITs.

Table 6 The seven most predominant Spoligotype International Types among 469 *M. tuberculosis* complex isolates from patients born in Sweden before 1945.

Sublineage	Spoligotype International Type	Number of isolates (%)
H3	SIT50	58 (12.4%)
T1	SIT53	52 (11.1%)
H1	SIT47	35 (7.5%)
LAM9	SIT42	15 (3.2%)
T2	SIT52	12 (2.6%)
T2	SIT153	12 (2.6%)
H3	SIT49	12 (2.6%)

The 182 patients with T lineage isolates were born between 1910-1945 (median 1923), with a mean age of 78.4 years at diagnosis (range 53-97), the 176 patients with H lineage isolates were born between 1908-1945 (median 1924), with a mean age of 77.6 years at diagnosis (range 52-98) and the 36 patients with LAM lineage isolates were born between 1913-1943 (median 1923), with a mean age of 78.9 years at diagnosis (range 63-94). The hypothesis that the 404 “modern” isolates of the T, H, LAM, X and S lineages differed in patient characteristics compared to the 36 “ancient” isolates split into two groups [group one (n=16): Beijing, EAI and CAS and group two (n=20): *M. bovis*/*M. bovis* like and Manu] with regards to the age of the patients at diagnosis and date of birth was tested. Patients with modern isolates of the T, H, LAM, X and S lineages were shown to be significantly older at diagnosis and were born significantly earlier than patients with “ancient” isolates.

The low number of isolates clustered by RFLP in this elderly population supported the concept that most isolates represented reactivation cases, without active TB transmission. My belief is that these MTC isolates displaying a low rate of diversity have had a common epidemiological background and represent the remaining cases of the TB epidemic that tormented Sweden during the first half of the 20th century.

To summarize the probable evolutionary relationships between the genotypes obtained, two separate spoligotyping-based minimum spanning trees (MST) were drawn (**Figure 6**). MST is an undirected network in which all of the samples are linked together with the fewest possible linkages between nearest neighbors. Using this approach, one assumes that all intermediate stages are present within the sample analyzed by first including the individual that shows the greatest number of possible linkages to other individuals in the population studied. **Figure 6A** is based only on SITs, while **Figure 6B** shows combined data for both SITs and orphan patterns pooled together. The SIT-based MST (**Figure 6A**) shows a tree split into distinct families: the top section displays the ancient PGG1 lineages (EAI, Bovis, Manu and CAS1-Delhi lineages) whereas the bottom shows the evolutionary modern PGG2/3 strains belonging to the Haarlem, LAM, and T lineages. The tree shown after combining orphans with SITs in **Figure 5B** is overlapping with the tree shown in **Figure 6A** in the sense that the two ancestral PGG1 lineage orphans were grouped together with their own lineage

(*M. bovis*), while evolutionary modern PGG2/3 (Haarlem, LAM, T) lineage SITs grouped with their own orphans. Almost all of the evolutionary modern orphan strains appeared at terminal positions within their respective genotypic families on the tree and not as the central nodes. This observation suggests that most of the orphan strains that existed in Sweden a century ago probably underwent extinction because they evolved as terminal members within these lineages that were unfit to proliferate.

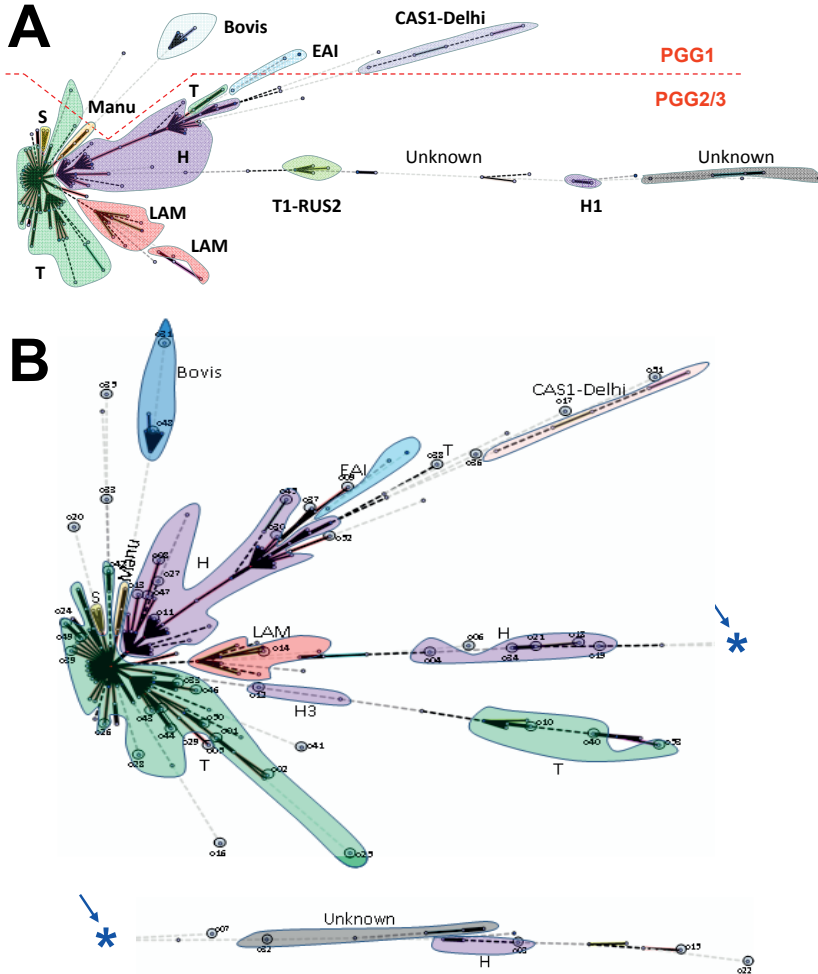


Figure 6 A minimum spanning tree (MST) illustrating evolutionary relationships between the Swedish spoligotypes obtained for *M. tuberculosis* complex strains presumably causing the TB epidemic a century ago. (A) SITs (B) all patterns, i.e. SITs and orphans pooled together. The tree connects each genotype based on degree of changes. The structure of the tree is represented by branches (continuous vs. dotted lines) and circles representing each individual pattern. The length of the branches represents the distance between patterns while the complexity of the lines (continuous, black dotted and gray dotted) denotes the number of spacer changes between two patterns: solid lines, one change; dotted lines, two or more changes.

2.3.5 Paper V

In paper V the molecular epidemiology of strains from Guinea-Bissau with regards to the phylogeography of *M. africanum* in West Africa was reinvestigated. Altogether, 414 isolates were investigated where 229 had been collected in 1989-1994, 138 in 1994-1998 and 47 in 2007 and 2008. Spoligotyping was performed on all isolates and the spoligotyping results from the SITVIT2 database comparison were summarized in Table 1 in paper V. Of 120 different patterns for the 414 strains studied 50 patterns corresponded to orphan strains that were unique among the at that time 75,000 strains recorded in the SITVIT2 database, as opposed to 70 patterns from 364 clinical isolates that corresponded to shared types.

Table 7 illustrates that the most predominant lineage in Guinea-Bissau was AFRI, which accounted for almost half of all TB cases. All the cases with AFRI, with the exception of a single orphan isolate classified as AFRI_3, were caused by the AFRI_1 sublineage split into two spoligopatterns; SIT181 (the prototype of the AFRI_1 sublineage) and SIT187. As compared to the SITVIT2, these two SITs were highly predominant in Guinea-Bissau and showed extremely high phylogeographical specificity for Western Africa.

Table 7 The major circulating lineages among 414 isolates from Guinea-Bissau.

Lineage	Number of isolates (%)
AFRI	195 (47.1%)
LAM	75 (18.1%)
T	53 (12.8%)
H	37 (8.9%)
EAI	25 (6.0%)
Unknown	12 (2.9%)
Beijing	7 (1.7%)
X	4 (1.0%)
MANU	4 (1.0%)
CAS	2 (0.5%)

AFRI: AFRI_1; AFRI_3, LAM: LAM1; LAM2; LAM3; LAM4; LAM5; LAM9; LAM10-CAM, T: T; T1; T2; T5, H (Haarlem): H1; H3, EAI: EAI-SOM; EAI5; EAI6-BGD, X: X1; X3, MANU_ancestor; MANU1; MANU2, CAS: CAS1-Dehli

Other observations made were the presence of the evolutionary ancient (PGG1) EAI5 sublineage (SIT527, n=9) with 100% of the cases being identified so far exclusively in Guinea-Bissau and the evolutionary recent (PGG2/3) LAM (SIT42 and SIT20) and Haarlem (SIT47 and SIT50) lineage. Regardless, when all the PGG1 isolates in Guinea-Bissau were added up, i.e. AFRI n=195, EAI n=25, Beijing n=7, Manu n=4, CAS n=2, and one unknown lineage pattern (SIT1200) that could be nonetheless classified as PGG1, the total number of ancestral strains in Guinea-Bissau was 234/414 or 56.5% of all strains. Only seven Beijing strains were identified, representing 1.7% of the study sample. The Beijing family of strains distributed worldwide¹⁴⁶, predominates in Asia and is except for parts of South Africa¹⁴⁷ quite rare

in Africa. During the third part of the study two isolates belonging to the Cameroon family had found their way into this setting. The recently expanding Cameroon family of *M. tuberculosis* has been reported to have out conquered *M. africanum* in Cameroon⁵⁷.

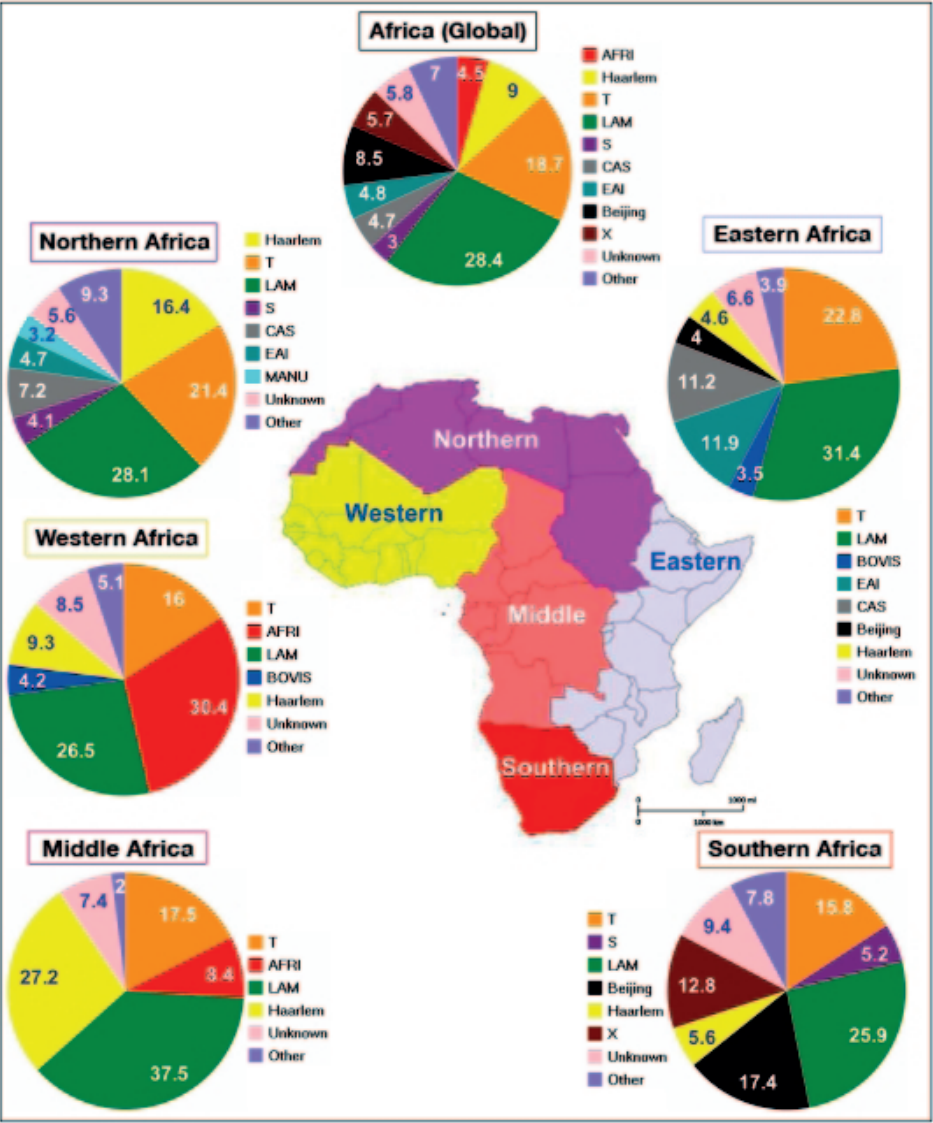


Figure 7 Geographical distribution of spoligotyping-based genotypic lineages of *M. tuberculosis* in various subregions of Africa (n=11956 clinical isolates) in the SITVIT2 database.

Results were only tabulated for all major lineages with a cut-off at 3% (strains belonging to all minor lineages below the cut-off 3% were pooled together and shown as “Other”). All spoligotyping signatures that were not yet associated to a well defined genotypic lineage were designated as “Unknown”.

To get an indication of the distribution of the Guinea-Bissau family in Africa all data obtained in Guinea-Bissau was compared to isolates in the SITVIT2 from various subregions of Africa (namely Eastern Africa, Middle Africa, Northern Africa, Southern Africa, and Western Africa (which included Guinea-Bissau). **Figure 7** highlights the major differences in the population structure of circulating TB bacilli in Africa. One can see that *M. africanum* is predominant in Western and Middle Africa, EAI and CAS in Northern and Eastern Africa, Beijing and X essentially limited to Southern Africa, while LAM, Haarlem and T are distributed throughout.

This study of MTC isolates from Guinea-Bissau, collected during two decades, shows that TB in Guinea-Bissau is essentially caused by predominant genotypes characteristic of the Western and Middle African region, with the highest prevalence of *M. africanum* (AFRI_1/West African 2) in the whole African continent. Forty-seven percent of all isolates were of what we call the Guinea-Bissau family, a member of the *M. africanum* AFRI_1/West African 2 lineage.

When a selection of Guinea-Bissau isolates (n=28, representing all five biovars¹⁴⁸) with different nitrate activity was investigated by the single nucleotide polymorphism at position -215 within the nitrate reductase (*narGHJI*) operon it was found that all but one harboured the -215C sequence (as does *M. bovis* and some ancestral *M. tuberculosis*¹⁴⁹). Thus, the variation in biochemical traits classically used to differentiate *M. tuberculosis* from *M. bovis* is not related to any genes (*narGHJI* and *pncA*) by us so far investigated.

To sum up, the Guinea-Bissau family of strains has an AFRI_1 spoligotype pattern lacking spacers 7 to 9 and 39, a distinct RFLP pattern with low numbers of IS6110 insertions and they lack the regions of difference RD7, RD8, RD9 and RD10¹⁵⁰, as well as RD701 and RD702³⁵. Altogether, this profile classifies the Guinea-Bissau family strains, irrespective of phenotypic biovar, as part of the *M. africanum* West African 2 lineage⁵⁵ or the AFRI_1 sublineage according to the spoligotyping nomenclature. This family shows extremely high phylogeographical specificity for Western Africa, with Guinea-Bissau being the epicenter. During the 15-year period since isolates were first collected, the proportion of isolates of the Guinea-Bissau family of strains significantly decreased over time. A potential reason for the decline of *M. africanum* in West Africa could be a lower transmission capacity. There is a huge uncertainty about the virulence of *M. africanum*¹⁵¹. An additional reason for the decline could be that certain modern lineages may possess advantages in their ability to disseminate within a community in relation to the more ancient lineages such as *M. africanum*. The appearance of Beijing and Cameroon family isolates during the late phase of the study is thereby of particular interest.

2.4 CONCLUDING REMARKS

With the aim of characterizing MTC isolates phylogenetically and epidemiologically, isolates from the two entirely different settings, Sweden and Guinea-Bissau, were collected. In view of what was found in paper I it was especially interesting to continue investigating the different genotypes in papers II-V. It was obvious that immigration from countries with a high incidence of TB had a strong impact on the TB epidemiology in Sweden. The Swedish TB setting in paper I was made up of more than half of the patients originating from Africa, primarily the Horn of Africa. In contrast, paper V investigated the TB setting in the West African country Guinea-Bissau.

Different genotypes of MTC predominate in different geographical regions of the world and strain-to-strain variations may have important consequences for instance when it comes to transmissibility. A large proportion of the global TB burden is caused by strain lineages that are clearly distinct from those usually used in TB product development⁵⁵. If future diagnostics, drugs and vaccine candidates are to be proven globally effective they may need to be evaluated against a range of MTC genotypes and host ethnicities. It is therefore of great importance to establish the whole spectrum of strains of the MTC worldwide.

Paper I explained that molecular epidemiological techniques revealed one of the largest outbreaks ever reported in a low incidence country. Early warning signals of cluster SMI-049 were not listened to and there were several false assumptions made that contributed to the successful spread of this strain. The finding of Swedish-born patients during the later phase of the outbreak shows that the outbreak has gradually taken hold of the Swedish-born population. It was illustrated that molecular epidemiological typing is a powerful tool to monitor and identify deficiencies in national TB control programs. Screening of asylum seekers and immigrants from high-incidence countries is at present not mandatory in Sweden. A weak TB control program where patients are not identified, do not complete their treatment and are lost to follow-up will promote the development and spread of drug-resistant TB. In my opinion, many of the cases of cluster SMI-049 could have been prevented had all patients taken their medications. DOTS is only on exceptional cases basis used in Sweden and if DOTS is not to be fully integrated into the Swedish national TB control program I feel it is absolutely essential to explain and thoroughly motivate the patient as to why it is important for them to conclude their treatment.

On the contrary to other reports, for instance on the Beijing lineage¹⁴², it was in paper III shown that the outbreak strain of cluster SMI-049 was exceptionally stable genetically, epidemiologically successful and that clonally disseminated *M. tuberculosis* strains can stay virtually unchanged over many years and multiple transmission cycles. This conclusion was drawn from many observations, one of them being that although the three isolates investigated were obtained with nine years in between and had been transmitted through a number of patients, the only differences found on the whole genome scale were four SNPs in one of the later isolated strains and a four amino-acids in-frame deletion in the other late strain compared to the index strain.

In contrast to the very successful spread of cluster SMI-049 it was discovered that drug-resistant Beijing lineage strains (paper II), which elsewhere in the world have caused large outbreaks, have not been able to spread within Sweden. In spite of the proximity to high prevalence countries such as Russia and the Baltic countries, Beijing lineage strains have not been observed to spread extensively within any of the Scandinavian countries.

The global distribution of Beijing lineage strains and their ability to predominate suggests that they have advantages to other strains that make them better adapted to infect and cause disease in humans. It is assumed that “modern” Beijing lineage strains have a higher degree of transmissibility than more “ancient” variants. In paper II it was shown that the collection of Beijing strains studied, originating from various parts of the world, displayed large genomic diversity. Based on a number of genetic markers it was possible to allocate these Beijing strains into 11 different arbitrary sublineages from more “modern” to more “ancient” variants. If allowed to speculate, it is possible that the Beijing sublineages so far introduced in Sweden may not be adapted to spread in Scandinavian or immigrant populations.

Today, Sweden is a high-income nation with a low prevalence of TB. A hundred years ago however, the estimated TB incidence was 890/100,000 population, higher than in most high incidence countries of sub-Saharan Africa. Little is known about the MTC population that dominated Sweden at that time and it appears as though this bacterial population has been successfully reduced from representing the major public health problem to its current level of near elimination. In paper IV isolates from patients born in Sweden before 1945 were studied and a highly homogenous bacterial population with a domination of the T, Haarlem and LAM lineages was found. It was shown that evolutionary recent PGG2/3 strains restricted to Sweden and its immediate neighbours appear to have caused the epidemic during the first half of the 20th century, while ancestral PGG1 strains are usually linked to immigrant populations in today’s Sweden.

Despite the high prevalence of TB in Africa, relatively little is known about the MTC genetic diversity in this continent. Studies in low-incidence countries like Sweden, which has a fairly high proportion of immigrants from Africa, could possibly contribute to increase this knowledge.

Paper V revisited Guinea-Bissau and studied the molecular epidemiology of strains with regards to the phylogeography of *M. africanum* in West Africa. Guinea-Bissau was shown to have the highest prevalence of *M. africanum* recorded in the African continent and the Guinea-Bissau family of strains demonstrated high phylogeographical specificity for Western Africa. During the 15-year period since isolates were first collected, the proportion of isolates of the Guinea-Bissau family of strains significantly decreased over time. One reason for the decline of *M. africanum* in West Africa could be that certain “modern” lineages possess advantages in their ability to disseminate within a community, in relation to more “ancient” lineages such as *M. africanum*. Yet the reason why *M. africanum* has not established itself outside of West Africa remains unknown. Others¹⁵² have potentially explained the limited geographic spread of *M. africanum*, despite massive migrations of West Africans to the Americas at the time of the slave trade, with the possibility that diseased slaves did not survive the crossing

of the Atlantic, or that *M. africanum* did initially establish itself in the New World, but was then outcompeted by *M. tuberculosis*. Or, *M. africanum* may not have been established in the indigenous American people and in their European colonizers if this organism has a host preference for ethnically West African persons. In a recent study from Ghana¹⁵³ a host polymorphism was found that is associated with protection against Euro-American *M. tuberculosis* but not against *M. africanum*, and may therefore have provided a selective advantage for *M. africanum* in West African populations. It has been hypothesized that lineages that are rare in a specific human population are not adapted to transmit and cause secondary cases in this specific human population¹⁵⁴⁻¹⁵⁶. As previously mentioned there are observations suggesting that particular lineages of the MTC might be adapted to specific human populations and maladapted to others⁴. Gagneux *et al* assigned strains to six main phylogenetic lineages and found that most areas were associated with only one or two lineages, whereas in Africa all six main lineages were represented⁴. Other studies¹⁵⁷ have found that no single *M. tuberculosis* lineage dominates in African-born patients and that Africa appears to be a melting pot for genetic diversity.

Molecular epidemiological tools, such as those described in this thesis, emerged with the objective of complementing classical epidemiology, tracking the movement of strains through space and time, and thereby enhancing the accuracy of the epidemiological picture¹⁵⁸. As in many other low-burden countries, Sweden has generally two distinct components of the TB situation; one is reactivation of latent infections in the native population and the other is influx in immigrant populations. This is in contrast to high-burden countries where the epidemic commonly is driven by recent transmission. It is important to remember that a molecular evolutionary marker that may answer well in a low-incidence setting with an effective TB control program, and where outbreaks of particular strains are well contained and do not persist within the community, might not be the most suitable choice in a high-incidence setting which often has epidemics of closely related TB strains that would require molecular markers exhibiting a much higher discriminatory power. The properties of molecular markers required to address both local and global levels of bacterial diversity are unlikely to be met by a single marker¹⁴⁴.

To conclude this thesis, it is my belief that molecular typing has contributed significantly to our understanding of the TB epidemiology. The papers included in this thesis revealed one of the largest outbreaks ever reported in a low incidence country and further characterization of this outbreak strain explained it to be remarkably stable genetically. It was also shown that the otherwise successfully disseminating Beijing lineage has not been able to spread within Sweden. One of the papers further reported on evolutionary recent PGG2/3 strains among elderly Swedes that are believed to have fueled the TB epidemic in Sweden a hundred years ago. Finally, by comparing the utterly different settings of Sweden and Guinea-Bissau some insight into the genetic diversity of the MTC in West versus East Africa was given.

3 ACKNOWLEDGEMENTS

I would like to express my genuine gratitude and sincere thanks to all those who have helped and encouraged me during this work, especially

Tuija Koivula, my supervisor, for her never-ending support, understanding, belief and friendship. Thanks for always having time!

Gunilla Källenius, my co-supervisor, for her kind and continuous interest, superior scientific knowledge and countless suggestions. Thanks for providing great inspiration.

Sven Hoffner, my co-supervisor, for his guidance, kind support and for creating opportunities.

Victoria Romanus, my former co-supervisor, for her encouragement, valuable discussions and fine collaborations over the years.

Other co-authors for their help; Ingela Berggren, Björn Petrini, Emmi Andersson, Judith Bruchfeld, Abdolreza Advani, Elsie Castro, Jolanta Mazurek, Andrzej Pawlowski, Boris Kan, Öjar Melefors, Jerker Jonsson, Nalin Rastogi, Jenny Svensson, Paulo Rabna, Raffaella Colombatti, Fabio Riccardi, David Couvin and Véronique Hill.

Solomon Ghebremichael, for his true friendship, back-up and generosity.

Alexandra Pennhag, for always helping out and keeping us afloat.

Past (and somewhat present) members of the amazing TB group, especially; Jim, Pontus, Anna, Melles, Juan Carlos, Emma, Maria, Andrzej, Jolanta, Lech and Lisbeth. I have truly appreciated all the help, support and fun times during fika.

As new constellations have been formed at SMI I have had the opportunity to meet and work with other people. I look forward to new collaborations and the times to come with the Unit of highly pathogenic microorganisms. In particular I would like to thank Tara, Talar and Andreas for their recent support and encouragement.

Over the years there have been many wonderful colleagues at SMI, especially at the former Department of Bacteriology. Many thanks to Britt-Marie and all others, ingen nämnd och absolut ingen glömd.

Friends and family for at all times showing honest interest (even though sometimes not having a clue what they were listening to).

Christopher and Lukas for distracting me with what is relevant in life!

4 REFERENCES

- 1 Sreevatsan, S. *et al.* Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci U S A* **94**, 9869-9874 (1997).
- 2 Kapur, V., Whittam, T. S. & Musser, J. M. Is Mycobacterium tuberculosis 15,000 years old? *J Infect Dis* **170**, 1348-1349 (1994).
- 3 Gutierrez, M. C. *et al.* Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. *PLoS Pathog* **1**, e5 (2005).
- 4 Gagneux, S. *et al.* Variable host-pathogen compatibility in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **103**, 2869-2873 (2006).
- 5 Zink, A., Haas, C. J., Reischl, U., Szeimies, U. & Nerlich, A. G. Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. *J Med Microbiol* **50**, 355-366 (2001).
- 6 Jones, G. *Captain of all these men of death - The history of tuberculosis in nineteenth and twentieth century Ireland.* (Editions Rodopi B.V., 2001).
- 7 Harries, A. D. & Dye, C. Tuberculosis. *Annals of Tropical Medicine & Parasitology* **100**, 415-431 (2006).
- 8 Grange, J. M. & Zumla, A. The global emergency of tuberculosis: what is the cause? *J R Soc Health* **122**, 78-81 (2002).
- 9 A.D., H. & C., D. Tuberculosis. *Annals of Tropical Medicine & Parasitology* **100**, 415-431 (2006).
- 10 Murray, J. F. A century of tuberculosis. *Am J Respir Crit Care Med* **169**, 1181-1186 (2004).
- 11 WHO. Global tuberculosis control. *WHO report* (2010).
<http://whqlibdoc.who.int/publications/2010/9789241564069_eng.pdf [cited 2011 Aug 17]>.
- 12 Population Reference Bureau - 2008 World Population Data Sheet,
<<http://www.prb.org/Publications/Datasheets/2008/2008wpds.aspx> (cited Sep 20, 2011)> (
- 13 Murray, J. F. Tuberculosis and HIV infection: a global perspective. *Respiration* **65**, 335-342 (1998).
- 14 Sotgiu, G. *et al.* Epidemiology and clinical management of XDR-TB: a systematic review by TBNET. *Eur Respir J* (2009).
- 15 Streicher, E. M. *et al.* Emergence and treatment of multidrug resistant (MDR) and extensively drug-resistant (XDR) tuberculosis in South Africa. *Infect Genet Evol.* doi:S1567-1348(11)00274-7 [pii]10.1016/j.meegid.2011.07.019 (2011).
- 16 Velayati, A. A. *et al.* Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in iran. *Chest* **136**, 420-425, doi:chest.08-2427 [pii]10.1378/chest.08-2427 (2009).
- 17 WHO. 2010/2011 Tuberculosis Global Facts. (2010).
<http://www.who.int/tb/publications/2010/factsheet_tb_2010_rev21feb11.pdf [cited 2011 Aug 17]>.
- 18 WHO. Consolidated action plan to prevent and combat multidrug- and extensively drug-resistant tuberculosis in the WHO European region 2011-2015., (WHO Regional Committee for Europe, Baku, Azerbaijan, 2011).
- 19 WHO. Treatment of Tuberculosis: guidelines for national programmes **3rd edition** (2003).
- 20 WHO. *Factsheet What is DOTS?* ,
<http://www.searo.who.int/en/Section10/Section2097/Section2106_10678.htm (Cited 26 Aug 2011)> (
- 21 Tamm, G., Dövertie, G., Johansson, J., Nylander, O., Östberg, G. Betänkande och förslag angående tuberkulossjukvårdens ordnande i riket. (1907).
- 22 Smittskyddsinstitutet. Statistik för tuberkulos 2010. (2011).
<<http://www.smi.se/statistik/tuberkulos/> [cited 2011 Aug 17]>.

- 23 Winqvist, N. *Dynamics of tuberculosis infection in Sweden* Doktorsexamen thesis, Lunds universitet(2011-05-31).
- 24 Romanus, V. Selective BCG vaccination in a country with low incidence of tuberculosis. *Euro Surveill* **11**, 14-17 (2006).
- 25 Kan, B. *et al.* Extensive transmission of an isoniazid-resistant strain of *Mycobacterium tuberculosis* in Sweden. *Int J Tuberc Lung Dis* **12**, 199-204 (2008).
- 26 Berggren-Palme, I., Larsson, L. O., Zedenius, I., Gillman, A. & Bennet, R. Outbreak of tuberculosis in a Stockholm nursery affects 35 children. *Euro Surveill* **10**, E051027 051025 (2005).
- 27 Romanus, V. Smittskyddsinstitutets årsrapport 2008: Statistik för tuberkulos. . (2008).
- 28 Ghebremichael, S. *et al.* [Resistant tuberculosis is spreading in Sweden. Molecular epidemiological strain identification by "fingerprinting" can make the infection tracing easier]. *Lakartidningen* **99**, 2618-2619, 2622-2613 (2002).
- 29 Socialstyrelsen. Tuberkulos - Vägledning för sjukvårdspersonal. (2009).
- 30 WHO. *Country profile Guinea-Bissau 2009*,
<<http://www.who.int/countries/gnb/en/> (Cited 2011 Aug 19)> (
- 31 CIA. *The 2008 World Factbook*. (Central Intelligence Agency, 2008).
- 32 WHO. *Global Health Observatory Data Repository Country Statistics Guinea-Bissau*, <<http://apps.who.int/ghodata/?vid=9800&theme=country> (cited Oct 14, 2011)
- 33 Mansson, F. *et al.* Trends of HIV-1 and HIV-2 prevalence among pregnant women in Guinea-Bissau, West Africa: possible effect of the civil war 1998 1999. *Sex Transm Infect* **83**, 463-467 (2007).
- 34 Imaeda, T. Deoxyribonucleic Acid Relatedness Among Selected Strains of *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium bovis* BCG, *Mycobacterium microti*, and *Mycobacterium africanum*. *Int. J. Syst. Bacteriol*, 147 - 150 (1985).
- 35 Mostowy, S. *et al.* Genomic analysis distinguishes *Mycobacterium africanum*. *J Clin Microbiol* **42**, 3594-3599 (2004).
- 36 Kallenius, G. *et al.* Evolution and clonal traits of *Mycobacterium tuberculosis* complex in Guinea-Bissau. *J Clin Microbiol* **37**, 3872-3878 (1999).
- 37 Pfyffer, G. E., Auckenthaler, R., van Embden, J. D. & van Soolingen, D. *Mycobacterium canettii*, the smooth variant of *M. tuberculosis*, isolated from a Swiss patient exposed in Africa. *Emerg Infect Dis* **4**, 631-634 (1998).
- 38 van Soolingen, D. *et al.* A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: characterization of an exceptional isolate from Africa. *Int J Syst Bacteriol* **47**, 1236-1245 (1997).
- 39 Rastogi, N., Legrand, E. & Sola, C. The mycobacteria: an introduction to nomenclature and pathogenesis. *Rev Sci Tech* **20**, 21-54 (2001).
- 40 van Soolingen, D. *et al.* Diagnosis of *Mycobacterium microti* infections among humans by using novel genetic markers. *J Clin Microbiol* **36**, 1840-1845 (1998).
- 41 Cousins, D. V. *et al.* Tuberculosis in wild seals and characterisation of the seal bacillus. *Aust Vet J* **70**, 92-97 (1993).
- 42 Aranaz, A., Cousins, D., Mateos, A. & Dominguez, L. Elevation of *Mycobacterium tuberculosis* subsp. *caprae* Aranaz *et al.* 1999 to species rank as *Mycobacterium caprae* comb. nov., sp. nov. *Int J Syst Evol Microbiol* **53**, 1785-1789 (2003).
- 43 Koch, R. The aetiology of tuberculosis. A translation by Berna Pinner and Max Pinner, with an introduction by Allen K Krause. . *Am Rev Tuberc* **25** (1932).
- 44 Wayne, L. G. Synchronized replication of *Mycobacterium tuberculosis*. *Infect Immun* **17**, 528-530 (1977).
- 45 Stead, W. W. *et al.* When did *Mycobacterium tuberculosis* infection first occur in the New World? An important question with public health implications. *Am J Respir Crit Care Med* **151**, 1267-1268 (1995).
- 46 Brosch, R. *et al.* A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A* **99**, 3684-3689 (2002).

- 47 Cosivi, O. *et al.* Zoonotic tuberculosis due to *Mycobacterium bovis* in developing countries. *Emerg Infect Dis* **4**, 59-70 (1998).
- 48 Calmette, A. La Vaccination Preventive Contre la Tuberculose. . *Masson et cie* (1927).
- 49 Andersen, P. & Doherty, T. M. The success and failure of BCG - implications for a novel tuberculosis vaccine. *Nat Rev Microbiol* **3**, 656-662 (2005).
- 50 Lanckriet, C., Levy-Bruhl, D., Bingono, E., Siopathis, R. M. & Guerin, N. Efficacy of BCG vaccination of the newborn: evaluation by a follow-up study of contacts in Bangui. *Int J Epidemiol* **24**, 1042-1049 (1995).
- 51 Mittal, S. K., Aggarwal, V., Rastogi, A. & Saini, N. Does B.C.G. vaccination prevent or postpone the occurrence of tuberculous meningitis? *Indian J Pediatr* **63**, 659-664 (1996).
- 52 Colditz, G. A. *et al.* The efficacy of bacillus Calmette-Guerin vaccination of newborns and infants in the prevention of tuberculosis: meta-analyses of the published literature. *Pediatrics* **96**, 29-35 (1995).
- 53 Fine, P. E. Variation in protection by BCG: implications of and for heterologous immunity. *Lancet* **346**, 1339-1345, doi:S0140-6736(95)92348-9 [pii] (1995).
- 54 Castets, M., Boisvert, H., Grumbach, F., Brunel, M. & Rist, N. [Tuberculosis bacilli of the African type: preliminary note]. *Rev Tuberc Pneumol (Paris)* **32**, 179-184 (1968).
- 55 Gagneux, S. & Small, P. M. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect Dis* **7**, 328-337 (2007).
- 56 de Jong BC, A. M., Awine T, Ogunbemi K, de Jong YP, Gagneux S, Deriemer K, Zozio T, Rastogi N, Borgdorff M, Hill PC, Adegbola RA. Use of spoligotyping and large-sequence polymorphisms to study the population structure of the *Mycobacterium tuberculosis* complex in a cohort study of consecutive smear positive tuberculosis cases in the Gambia. *J Clin Microbiol*. (2009).
- 57 Niobe-Eyangoh, S. N. *et al.* Genetic biodiversity of *Mycobacterium tuberculosis* complex strains from patients with pulmonary tuberculosis in Cameroon. *J Clin Microbiol* **41**, 2547-2553 (2003).
- 58 Godreuil, S. *et al.* First molecular epidemiology study of *Mycobacterium tuberculosis* in Burkina Faso. *J Clin Microbiol* **45**, 921-927 (2007).
- 59 Zink, A. R. *et al.* Characterization of *Mycobacterium tuberculosis* complex DNAs from Egyptian mummies by spoligotyping. *J Clin Microbiol* **41**, 359-367 (2003).
- 60 Desmond, E. *et al.* *Mycobacterium africanum* cases, California. *Emerg Infect Dis* **10**, 921-923 (2004).
- 61 Miltgen, J. *et al.* Two cases of pulmonary tuberculosis caused by *Mycobacterium tuberculosis* subsp *canetti*. *Emerg Infect Dis* **8**, 1350-1352 (2002).
- 62 Wells, A. Q., and D. M. Oxon. Tuberculosis in wild voles. . *Lancet i* (1937).
- 63 Huitema, H., and F. H. J. Jaartveld. . *Mycobacterium microti* infection in a cat and some pigs. . *Antonie Leeuwenhoek* 209-212 (1967).
- 64 Pattyn, S. R., F. A. Portaels, P. Kageruka, and P. Gigase. . *Mycobacterium microti* infection in a zoo-llama: lama Vicugna (molina). . *Acta Zool. Pathol. Antverp.* , 17-24 (1970).
- 65 Xavier Emmanuel, F. *et al.* Human and animal infections with *Mycobacterium microti*, Scotland. *Emerg Infect Dis* **13**, 1924-1927 (2007).
- 66 Kiers, A., Klarenbeek, A., Mendelts, B., Van Soolingen, D. & Koeter, G. Transmission of *Mycobacterium pinnipedii* to humans in a zoo with marine mammals. *Int J Tuberc Lung Dis* **12**, 1469-1473 (2008).
- 67 Cousins, D. V. *et al.* Tuberculosis in seals caused by a novel member of the *Mycobacterium tuberculosis* complex: *Mycobacterium pinnipedii* sp. nov. *Int J Syst Evol Microbiol* **53**, 1305-1314 (2003).
- 68 Aranaz, A. *et al.* *Mycobacterium tuberculosis* subsp. *caprae* subsp. nov.: a taxonomic study of a new member of the *Mycobacterium tuberculosis* complex isolated from goats in Spain. *Int J Syst Bacteriol* **49 Pt 3**, 1263-1273 (1999).

- 69 Kent, P. & Kubica, G. *Public Health Mycobacteriology. A guide for the level III laboratory*. . (Centers for Disease Control 1985).
- 70 Bloom, B. R. *Tuberculosis pathogenesis, protection and control*. (ASM Press, 1994).
- 71 Kallenius, G., Hoffner, S. E., Miorner, H. & Svenson, S. B. Novel approaches to the diagnosis of mycobacterial infections. *Eur Respir J* **7**, 1921-1924 (1994).
- 72 Walker, D. Economic analysis of tuberculosis diagnostic tests in disease control: how can it be modelled and what additional information is needed? *Int J Tuberc Lung Dis* **5**, 1099-1108 (2001).
- 73 Djelouadji, Z., Raoult, D., Daffe, M. & Drancourt, M. A Single-Step Sequencing Method for the Identification of Mycobacterium tuberculosis Complex Species. *PLoS Negl Trop Dis* **2**, e253 (2008).
- 74 Hall, L. & Roberts, G. Non-molecular identification of nontuberculous mycobacteria in the clinical microbiology laboratory: What's the real deal? . *Clinical Microbiology Newsletter* **28**, 73-80 (2006).
- 75 Butler, W. R. & Guthertz, L. S. Mycolic acid analysis by high-performance liquid chromatography for identification of Mycobacterium species. *Clin Microbiol Rev* **14**, 704-726, table of contents (2001).
- 76 Kirschner, P. *et al.* Genotypic identification of mycobacteria by nucleic acid sequence determination: report of a 2-year experience in a clinical laboratory. *J Clin Microbiol* **31**, 2882-2889 (1993).
- 77 Richter, E., Weizenegger, M., Fahr, A. M. & Rusch-Gerdes, S. Usefulness of the GenoType MTBC assay for differentiating species of the Mycobacterium tuberculosis complex in cultures obtained from clinical specimens. *J Clin Microbiol* **42**, 4303-4306 (2004).
- 78 Evans, K. D., Nakasone, A. S., Sutherland, P. A., de la Maza, L. M. & Peterson, E. M. Identification of Mycobacterium tuberculosis and Mycobacterium avium-M. intracellulare directly from primary BACTEC cultures by using acridinium-ester-labeled DNA probes. *J Clin Microbiol* **30**, 2427-2431 (1992).
- 79 Miller, N., Infante, S. & Cleary, T. Evaluation of the LIPA MYCOBACTERIA assay for identification of mycobacterial species from BACTEC 12B bottles. *J Clin Microbiol* **38**, 1915-1919 (2000).
- 80 Richter, E., Weizenegger, M., Rusch-Gerdes, S. & Niemann, S. Evaluation of genotype MTBC assay for differentiation of clinical Mycobacterium tuberculosis complex isolates. *J Clin Microbiol* **41**, 2672-2675 (2003).
- 81 Behr, M. A. *et al.* Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**, 1520-1523 (1999).
- 82 Huard, R. C., Lazzarini, L. C., Butler, W. R., van Soolingen, D. & Ho, J. L. PCR-based method to differentiate the subspecies of the Mycobacterium tuberculosis complex on the basis of genomic deletions. *J Clin Microbiol* **41**, 1637-1650 (2003).
- 83 Huard, R. C. *et al.* Novel genetic polymorphisms that further delineate the phylogeny of the Mycobacterium tuberculosis complex. *J Bacteriol* **188**, 4271-4287 (2006).
- 84 Cole, S. T. *et al.* Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393**, 537-544 (1998).
- 85 Mostrom, P., Gordon, M., Sola, C., Ridell, M. & Rastogi, N. Methods used in the molecular epidemiology of tuberculosis. *Clin Microbiol Infect* **8**, 694-704 (2002).
- 86 van Embden, J. D. *et al.* Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* **31**, 406-409 (1993).
- 87 van Soolingen, D., Hermans, P. W., de Haas, P. E., Soll, D. R. & van Embden, J. D. 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).
- 88 van Soolingen, D., de Haas, P. E., Hermans, P. W., Groenen, P. M. & van Embden, J. D. Comparison of various repetitive DNA elements as genetic

- markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. *J Clin Microbiol* **31**, 1987-1995 (1993).
- 89 Hermans, P. W. *et al.* Insertion element IS987 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).
- 90 Kamerbeek, J. *et al.* Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* **35**, 907-914 (1997).
- 91 Flores, L. *et al.* Large sequence polymorphisms classify *Mycobacterium tuberculosis* strains with ancestral spoligotyping patterns. *J Clin Microbiol* **45**, 3393-3395, doi:JCM.00828-07 [pii]10.1128/JCM.00828-07 (2007).
- 92 Driscoll, J. R. Spoligotyping for molecular epidemiology of the *Mycobacterium tuberculosis* complex. *Methods Mol Biol* **551**, 117-128, doi:10.1007/978-1-60327-999-4_10 (2009).
- 93 Brudey, K. *et al.* *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol* **6**, 23 (2006).
- 94 *SITVIT database*, <<http://www.pasteur-guadeloupe.fr:8081/SITVITDemo> [cited 2011 Aug 17]> (
- 95 Frothingham, R. & Meeker-O'Connell, W. A. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology* **144** (Pt 5), 1189-1196 (1998).
- 96 Kremer, K. *et al.* 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).
- 97 Supply, P. *et al.* Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Mol Microbiol* **36**, 762-771 (2000).
- 98 Supply, P. *et al.* 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).
- 99 Allix-Beguec, C., Fauville-Dufaux, M. & Supply, P. Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* **46**, 1398-1406 (2008).
- 100 Oelemann, M. C. *et al.* Assessment of an optimized mycobacterial interspersed repetitive- unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. *J Clin Microbiol* **45**, 691-697 (2007).
- 101 Iwamoto, T. *et al.* Hypervariable loci that enhance the discriminatory ability of newly proposed 15-loci and 24-loci variable-number tandem repeat typing method on *Mycobacterium tuberculosis* strains predominated by the Beijing family. *FEMS Microbiol Lett* **270**, 67-74 (2007).
- 102 Jiao, W. W. *et al.* Evaluation of new variable-number tandem-repeat systems for typing *Mycobacterium tuberculosis* with Beijing genotype isolates from Beijing, China. *J Clin Microbiol* **46**, 1045-1049 (2008).
- 103 Hanekom, M. *et al.* Discordance between mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing and IS6110 restriction fragment length polymorphism genotyping for analysis of *Mycobacterium tuberculosis* Beijing strains in a setting of high incidence of tuberculosis. *J Clin Microbiol* **46**, 3338-3345 (2008).
- 104 Cardoso Oelemann, M. *et al.* The forest behind the tree: phylogenetic exploration of a dominant *Mycobacterium tuberculosis* strain lineage from a high tuberculosis burden country. *PLoS One* **6**, e18256, doi:10.1371/journal.pone.0018256 (2011).
- 105 Filliol, I. *et al.* Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and

- recommendations for a minimal standard SNP set. *J Bacteriol* **188**, 759-772, doi:188/2/759 [pii]10.1128/JB.188.2.759-772.2006 (2006).
- 106 Gutacker, M. M. *et al.* Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. *Genetics* **162**, 1533-1543 (2002).
- 107 Schurch, A. C. *et al.* High-resolution typing by integration of genome sequencing data in a large tuberculosis cluster. *J Clin Microbiol* **48**, 3403-3406, doi:JCM.00370-10 [pii]10.1128/JCM.00370-10 (2010).
- 108 Ng, P. C. & Kirkness, E. F. Whole genome sequencing. *Methods Mol Biol* **628**, 215-226, doi:10.1007/978-1-60327-367-1_12 (2010).
- 109 Voelkerding, K. V., Dames, S. & Durtsche, J. D. Next generation sequencing for clinical diagnostics-principles and application to targeted resequencing for hypertrophic cardiomyopathy: a paper from the 2009 William Beaumont Hospital Symposium on Molecular Pathology. *J Mol Diagn* **12**, 539-551, doi:S1525-1578(10)60097-9 [pii]10.2353/jmoldx.2010.100043 (2010).
- 110 Duchene, V. *et al.* Phylogenetic reconstruction of *Mycobacterium tuberculosis* within four settings of the Caribbean region: tree comparative analyse and first appraisal on their phylogeography. *Infect Genet Evol* **4**, 5-14, doi:10.1016/j.meegid.2003.09.001S1567134803000972 [pii] (2004).
- 111 Vansoolingen, D. *et al.* Predominance of a Single Genotype of *Mycobacterium-Tuberculosis* in Countries of East-Asia. *Journal of Clinical Microbiology* **33**, 3234-3238 (1995).
- 112 Bifani, P. J., Mathema, B., Kurepina, N. E. & Kreiswirth, B. N. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. *Trends Microbiol* **10**, 45-52, doi:S0966842X01022776 [pii] (2002).
- 113 European Concerted Action on New Generation Genetic Markers and Techniques for the Epidemiology and Control of Tuberculosis - Beijing/W genotype *Mycobacterium tuberculosis* and drug resistance. *Emerg Infect Dis* **12**, 736-743 (2006).
- 114 Bifani, P. J. *et al.* Origin and interstate spread of a New York City multidrug-resistant *Mycobacterium tuberculosis* clone family. *JAMA* **275**, 452-457 (1996).
- 115 Kruuner, A. *et al.* Spread of drug-resistant pulmonary tuberculosis in Estonia. *J Clin Microbiol* **39**, 3339-3345 (2001).
- 116 Diaz, R. *et al.* Molecular epidemiology of tuberculosis in Cuba outside of Havana, July 1994-June 1995: utility of spoligotyping versus IS6110 restriction fragment length polymorphism. *Int J Tuberc Lung Dis* **2**, 743-750 (1998).
- 117 Johnson, R. *et al.* An outbreak of drug-resistant tuberculosis caused by a Beijing strain in the western Cape, South Africa. *Int J Tuberc Lung Dis* **10**, 1412-1414 (2006).
- 118 Pfyffer, G. E. *et al.* Multidrug-resistant tuberculosis in prison inmates, Azerbaijan. *Emerg Infect Dis* **7**, 855-861 (2001).
- 119 Narvskaya, O. *et al.* Nosocomial outbreak of multidrug-resistant tuberculosis caused by a strain of *Mycobacterium tuberculosis* W-Beijing family in St. Petersburg, Russia. *Eur J Clin Microbiol Infect Dis* **21**, 596-602, doi:10.1007/s10096-002-0775-4 (2002).
- 120 Caminero, J. A. *et al.* Epidemiological evidence of the spread of a *Mycobacterium tuberculosis* strain of the Beijing genotype on Gran Canaria Island. *Am J Respir Crit Care Med* **164**, 1165-1170 (2001).
- 121 Moss, A. R. *et al.* A city-wide outbreak of a multiple-drug-resistant strain of *Mycobacterium tuberculosis* in New York. *Int J Tuberc Lung Dis* **1**, 115-121 (1997).
- 122 Rastogi, N. & Sola, C. *Molecular evolution of the Mycobacterium tuberculosis complex. Amadeo Online Textbooks 2007:* <http://www.tuberculosis textbook.com/index.htm> [cited 2011 Sep 19].
- 123 WHO. *Treatment of tuberculosis guidelines - fourth edition*, <http://whqlibdoc.who.int/publications/2010/9789241547833_eng.pdf (cited Och 10, 2011)> (2010).
- 124 CDC. Revised Definition of Extensively Drug-Resistant Tuberculosis. *MMWR Recomm Rep* **55** (2006).

- 125 WHO. *Molecular Line Probe Assays for Rapid Screening of patients at Risk of Multidrug-resistant Tuberculosis - Policy Statement*,
<http://www.who.int/tb/laboratory/lpa_policy.pdf (cited Oct 10, 2011)> (2008).
- 126 WHO. *Xpert MTB/RIF - rapid TB test - WHO publishes policy and guidance for implementations*
<http://www.who.int/tb/features_archive/xpert_rapid_tb_test/en/index.html
(cited 2011 Aug 31)> (2011).
- 127 Narasiman, R., Mathur, G. P. & Pamra, S. P. Sodium lauryl sulphate method of culturing sputum for Mycobacterium tuberculosis. *Indian Journal of Tuberculosis* **19** 125-130 (1972).
- 128 Fernstrom, M. C., Dahlgren, L., Ranby, M., Forsgren, A. & Petrini, B. Increased sensitivity of Mycobacterium tuberculosis Cobas Amplicor PCR following brief incubation of tissue samples on Lowenstein-Jensen substrate. *Apmis* **111**, 1114-1116 (2003).
- 129 Goquet de la Salmonière, Y., Kim, CC., Tsolaki, AG., Pym, AS., Siegrist, MS., Small, PM. High-throughput method for detecting genomic-deletion polymorphisms *J Clin Microbiol* **42** (7), 2913-2918 (2004).
- 130 Rao, K. R. *et al.* Analysis of genomic downsizing on the basis of region-of-difference polymorphism profiling of Mycobacterium tuberculosis patient isolates reveals geographic partitioning. *J Clin Microbiol* **43**, 5978-5982, doi:10.1128/JCM.43.12.5978-5982.2005 (2005).
- 131 Musser, J., Amin, A., Ramaswamy, S.. Negligible genetic diversity of mycobacterium tuberculosis host immune system protein targets: evidence of limited selective pressure *Genetics* **155**(1), 7-16 (2000).
- 132 Mokrousov, I. *et al.* Phylogenetic reconstruction within Mycobacterium tuberculosis Beijing genotype in northwestern Russia. *Res Microbiol* **153**, 629-637 (2002).
- 133 Garzelli, C., Rindi, L., Lari, N. & Cuccu, B. Evolutionary pathway of the Beijing lineage of Mycobacterium tuberculosis based on genomic deletions and mutT genes polymorphisms. *Infection Genetics and Evolution* **9**, 48-53, doi:10.1016/j.meegid.2008.09.006 (2009).
- 134 Gicquel, B. *et al.* Mutations in putative mutator genes of Mycobacterium tuberculosis strains of the W-Beijing family. *Emerging Infectious Diseases* **9**, 838-845 (2003).
- 135 Mizrahi, V. & Andersen, S. J. DNA repair in Mycobacterium tuberculosis. What have we learnt from the genome sequence? *Molecular Microbiology* **29**, 1331-1339 (1998).
- 136 Werngren, J. & Hoffner, S. E. Drug-susceptible Mycobacterium tuberculosis Beijing genotype does not develop mutation-conferred resistance to rifampin at an elevated rate. *J Clin Microbiol* **41**, 1520-1524 (2003).
- 137 Banerjee, A. *et al.* inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. *Science* **263**, 227-230 (1994).
- 138 North, R. J. & Jung, Y. J. Immunity to tuberculosis. *Annu Rev Immunol* **22**, 599-623, doi:10.1146/annurev.immunol.22.012703.104635 (2004).
- 139 van Crevel, R., Ottenhoff, T. H. & van der Meer, J. W. Innate immunity to Mycobacterium tuberculosis. *Clin Microbiol Rev* **15**, 294-309 (2002).
- 140 Park, J. S., Tamayo, M. H., Gonzalez-Juarrero, M., Orme, I. M. & Ordway, D. J. Virulent clinical isolates of Mycobacterium tuberculosis grow rapidly and induce cellular necrosis but minimal apoptosis in murine macrophages. *J Leukoc Biol* **79**, 80-86, doi:10.1189/jlb.0505250 [pii]10.1189/jlb.0505250 (2006).
- 141 Stermann, M., Bohrsen, A., Diephaus, C., Maass, S. & Bange, F. C. Polymorphic nucleotide within the promoter of nitrate reductase (NarGHJI) is specific for Mycobacterium tuberculosis. *J Clin Microbiol* **41**, 3252-3259 (2003).
- 142 Niemann, S. *et al.* Genomic diversity among drug sensitive and multidrug resistant isolates of Mycobacterium tuberculosis with identical DNA fingerprints. *PLoS One* **4**, e7407, doi:10.1371/journal.pone.0007407 (2009).
- 143 Hershberg, R. *et al.* High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. *PLoS Biol* **6**, e311, doi:10.1371/journal.pbio.0060311 [pii]10.1371/journal.pbio.0060311 (2008).

- 144 Comas, I., Homolka, S., Niemann, S. & Gagneux, S. Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS One* **4**, e7815, doi:10.1371/journal.pone.0007815 (2009).
- 145 Asiimwe, B. B. *et al.* *Mycobacterium tuberculosis* Uganda genotype is the predominant cause of TB in Kampala, Uganda. *Int J Tuberc Lung Dis* **12**, 386-391 (2008).
- 146 Glynn, J. R., Whiteley, J., Bifani, P. J., Kremer, K. & van Soolingen, D. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. *Emerg Infect Dis* **8**, 843-849 (2002).
- 147 Cowley, D. *et al.* Recent and rapid emergence of W-Beijing strains of *Mycobacterium tuberculosis* in Cape Town, South Africa. *Clin Infect Dis* **47**, 1252-1259 (2008).
- 148 Hoffner, S. E. *et al.* Biochemical heterogeneity of *Mycobacterium tuberculosis* complex isolates in Guinea-Bissau. *J Clin Microbiol* **31**, 2215-2217 (1993).
- 149 Goh, K. S., Rastogi, N., Berchel, M., Huard, R. C. & Sola, C. Molecular evolutionary history of tubercle bacilli assessed by study of the polymorphic nucleotide within the nitrate reductase (narGHJI) operon promoter. *J Clin Microbiol* **43**, 4010-4014, doi:43/8/4010 [pii]10.1128/JCM.43.8.4010-4014.2005 (2005).
- 150 Koivula, T. *et al.* Genetic characterization of the Guinea-Bissau family of *Mycobacterium tuberculosis* complex strains. *Microbes Infect* **6**, 272-278 (2004).
- 151 de Jong, B. C. *et al.* Progression to active tuberculosis, but not transmission, varies by *Mycobacterium tuberculosis* lineage in The Gambia. *J Infect Dis* **198**, 1037-1043 (2008).
- 152 de Jong, B. C., Antonio, M. & Gagneux, S. *Mycobacterium africanum*--review of an important cause of human tuberculosis in West Africa. *PLoS Negl Trop Dis* **4**, e744, doi:10.1371/journal.pntd.0000744 (2010).
- 153 Intemann, C. D. *et al.* Autophagy gene variant IRGM -261T contributes to protection from tuberculosis caused by *Mycobacterium tuberculosis* but not by *M. africanum* strains. *PLoS Pathog* **5**, e1000577, doi:10.1371/journal.ppat.1000577 (2009).
- 154 Hanekom, M. *et al.* Evidence that the spread of *Mycobacterium tuberculosis* strains with the Beijing genotype is human population dependent. *J Clin Microbiol* **45**, 2263-2266, doi:JCM.02354-06 [pii]10.1128/JCM.02354-06 (2007).
- 155 Hirsh, A. E., Tsolaki, A. G., DeRiemer, K., Feldman, M. W. & Small, P. M. Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. *Proc Natl Acad Sci U S A* **101**, 4871-4876, doi:10.1073/pnas.03056271010305627101 [pii] (2004).
- 156 Caws, M. *et al.* The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. *PLoS Pathog* **4**, e1000034, doi:10.1371/journal.ppat.1000034 (2008).
- 157 Baker, L., Brown, T., Maiden, M. C. & Drobniowski, F. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. *Emerg Infect Dis* **10**, 1568-1577 (2004).
- 158 van der Spuy, G. D., Warren, R. M. & van Helden, P. D. The role of molecular epidemiology in low-income, high-burden countries. *Int J Tuberc Lung Dis* **13**, 419-420 (2009).