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

MOLECULAR EPIDEMIOLOGY OF TUBERCULOSIS

Ramona Petersson

Karolinska Institutet

SMITTSKYDDSINSTITUTET

Swedish Institute for Infectious Disease Control

Published by Karolinska Institutet.

ISBN 978-91-7409-456-5

Printed by
REPROPRINT AB
Stockholm 2009
www.reproprint.se
Gårdsvägen 4, 169 70 Solna

"It was the best of times, it was the worst of times, it was the age of wisdom…"

Charles Dickens (1812-1870)

ABSTRACT

Tuberculosis (TB) is a global epidemic, with one third of the world's population infected, at least 9 million new active cases per year and about 2 million deaths per year. Without adequate chemotherapy treatment TB may be a mortal disease.

In countries like Sweden the mortality was dramatically reduced about half a century ago when modern TB drugs where introduced. These gains are however seriously jeopardized by the now emerging multidrug resistant (MDR) and extensively drug-resistant (XDR) TB. In Sweden the majority of patients with TB are immigrant from countries with a high incidence of TB. The incidence among the Swedish-born population has continued to decrease while it has increased among the foreign-borns.

In the West African country Guinea-Bissau TB is a common disease where the incidence is believed to be further increased by the epidemic of the human immunodeficiency virus (HIV). 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. The national TB control programme has regrettably been having difficulties in its function.

Different genotypes of *Mycobacterium tuberculosis* complex (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. 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 MTC world wide. Despite the high prevalence of TB in Africa, relatively little is known about the MTC genetic diversity on this continent. This study phylogenetically and epidemiologically characterized MTC isolates obtained from TB patients in Sweden and Guinea-Bissau using the two molecular techniques Restriction Fragment Length Polymorphism (RFLP) and spacer oligonucleotide typing (spoligotyping). The work was performed with the view to understand species and strain diversity as well as transmission patterns.

In Sweden, 400 drug-resistant isolates were collected during the years 1994-2005. The great majority (87%) of the patients was foreign-born and originated from several regions, with a clear dominance for Africa, in particular the Horn of Africa. It was found that the T lineage was predominant and that the isolates of the foreign-born patients to a large extent reflected genotypes common in their country of origin. One large outbreak of isoniazid resistant TB was identified, at the time (2005) involving 96 patients, mainly from the Horn of Africa. This outbreak represents one of the largest outbreaks of TB ever reported and was an important warning signal to the Swedish authorities. This study showed that molecular epidemiological typing is a powerful tool to monitor and identify chains of transmission which indicate deficiencies in national TB control programs. It was our strong recommendation that Swedish health authorities should integrate full Directly Observed Treatment Short course (DOTS) in the Swedish national TB control program.

In Guinea-Bissau we reinvestigated the molecular epidemiology of strains with regards to the phylogeography of *M. africanum* in West Africa. Altogether, 418 isolates collected at three different points in time were investigated. The main (47%) lineage present was the AFRI (*M. africanum*) lineage which included the genotypically closely related but phenotypically heterogenic Guinea-Bissau family of strains. This family of strains appeared to be decreasing over time and may be out conquered by other more successful lineages such as the Beijing and Cameroon family of strains.

LIST OF PUBLICATIONS

This thesis is based on the following articles, which in the text will be referred to by their roman numerals.

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; 10(6):699-705.

II. **Petersson R**, Ghebremichael S, Källenius G, Rabna P, Colombatti R, Riccardi F, Koivula T. The Guinea-Bissau family of *Mycobacterium tuberculosis* complex revisited. Manuscript.

CONTENTS

LIST OF ABBREVIATIONS

AFRI	*Mycobacterium africanum* lineage
BCG	Bacillus Calmette-Guérin
CAS	Central Asian
DNA	Deoxyribonucleic Acid
DOTS	Directly Observed Treatment Short course
DR	Direct Repeat
DVR	Direct Variable Repeat
EAI	East-African-Indian
HIV	Human Immunodeficiency Virus
IS	Insertion Sequence
LAM	Latin-American-Mediterranean
LJ	Löwenstein-Jensen (egg medium)
LSP	Large Sequence Polymorphism
MDR	Multidrug Resistant
MIRU	Mycobacterial Interspersed Repetitive Unit
MTC	*Mycobacterium tuberculosis* Complex
PCR	Polymerase Chain Reaction
PZA	Pyrazinamide
RD	Region of Difference
RFLP	Restriction Fragment Length Polymorphism
rRNA	Ribosomal Ribonucleic Acid
ST	Shared Type
SNP	Single Nucleotide Polymorphism
Spoligotyping	Spacer Oligonucleotide Typing
TB	Tuberculosis
TCH	Thiophene-2-Carboxylic Acid Hydrazide
WHO	World Health Organization
VNTR	Variable Numbers of Tandem Repeats
XDR	Extensively Drug-Resistant

1 INTRODUCTION

1.1 GLOBAL HEALTH BURDEN OF TUBERCULOSIS

In our day, tuberculosis (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 cured[1]. Despite remarkable achievements TB is still one of the leading causes of death from an infectious disease worldwide. Current estimates indicate that one third of the world's population is infected; there are more than 9 million new active cases per year and over 2 million deaths per year[2]. Although it is a global epidemic, TB predominantly affects the populations of resource-poor countries, where 98% of all TB deaths occur[3]. The great majority (86%) of the world's total population lives in low resource countries which is also where 95% of all TB cases occur. These countries are furthermore now being ravaged by the pandemic of human immunodeficiency virus (HIV) infection - the most powerful factor ever known to favour the development of TB[4]. In our part of the world the situation is to some extent more hopeful. During the nineteen hundreds a drastic decline in incidence rate took place in Western Europe and USA as social factors improved, public health measures were taken, treatments were made available and living conditions were improved[3].

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. 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 poor nations[3, 5].

TB is not a novel disease. Early estimates suggested that the causative agent of TB, *Mycobacterium tuberculosis*, was 15,000 years old[6, 7], whereas recent evidence suggests that ancestral mycobacteria may have already affected early hominids in East Africa 3 million years ago[8]. It is believed that human TB originated in East Africa and was expanded and diversified during its spread out of Africa[9].

Epidemiological studies are important for understanding the transmission of TB and for tracing sources of outbreaks in different settings. Genotyping tools are recommended[1] 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 MYCOBACTERIUM TUBERCULOSIS COMPLEX

The MTC consists of a highly related[10] group of acid-alcohol-fast bacilli which are human and animal pathogens. Detection of the different species within the complex has mainly been based on the analysis of phenotypic characteristics such as acid-fast microscopy, 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 promising. 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)[11]. 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* accounts for TB in West Africa[12]; *M. canettii* appears be linked to the Horn of Africa[13, 14]; *M. microti* infects voles[15, 16]; *M. pinnipedii* causes disease in seals[17] and *M. caprae* in goats[18].

1.2.1 Mycobacterium tuberculosis

This organism, first described by Koch[19] 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 replication[20]. It was previously speculated that *M. tuberculosis* evolved from *M. bovis* by specific adaptation of an animal pathogen to human host[21]. Conversely, comparative genomics later uncovered several variable genomic regions in the members of MTC. Based on the presence or absence of an *M. tuberculosis* specific deletion (TbD1), *M. tuberculosi*s 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 already[22].

1.2.2 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 distinguishable 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 practiced[23]. The world wide used live attenuated vaccine strain *M. bovis* BCG (Bacillus Calmette-Guérin) has its origin in a virulent *M. bovis* strain that lost its virulence in 1921 after 230 *in vitro* passages[24]. BCG can however cause disease in humans[25, 26].

1.2.3 Mycobacterium africanum

This heterogeneous group of isolates responsible for human TB in Africa was first[27] 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*[9, 11]. 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[28, 29]. 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 pathogen or a laboratory misclassification[11]. 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 Cameroon[30] and Burkina Faso[31]. DNA from both *M. africanum* and *M. tuberculosis* was recovered from Egyptian mummies[32], yet today *M. africanum* isolates are rarely found outside of West Africa except among first generation immigrants[33]. 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 Africa[29].

1.2.4 Mycobacterium canettii

M. canettii is a very rare, smooth variant of *M. tuberculosis* that was earliest isolated in 1969 from a French farmer by Canetti. Isolates of *M. canettii* are characterized by a highly particular growing 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 Africa[34]. 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, IS*1081* copy number, colony morphology, and the lipid content of the cell wall[22]. 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 morphology[14]. Deletion analyses have shown that RD, RvD, 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 DR region that are no longer present in any other member of the MTC[22].

1.2.5 Mycobacterium microti

M. microti is the vole bacillus discovered by Wells[35] in 1937. It is mainly found in small rodents, but infection has also been documented in cats, pigs and llamas[36, 37]. 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 rare[38]. 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 IS*6110* 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[16, 37].

1.2.6 Mycobacterium pinnipedii

This organism was first[17] isolated from cases of TB in Australian sea lions and seals. Later on, similar organisms were recovered from the same animals in South America. Cousins *et al* confirmed with biochemical testing that the seal bacillus, isolated from pinnipeds, belonged to the MTC. In most cases, the seal isolates grew preferentially on media that contained sodium pyruvate. 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 MTC[39].

1.2.7 Mycobacterium caprae

These mycobacteria first isolated from goats in Spain[40] possess clear traits that differentiate them from the classical species of the complex. Besides biochemical (sensitivity to PZA) and epidemiological features, strains 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*[18, 40].

1.3 LABORATORY DIAGNOSIS OF TUBERCULOSIS

1.3.1 IDENTIFICATION OF THE MTC

1.3.1.1 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 microscopy[41]. Using a fluorescence microscope mycobacteria can also be stained with auramine[42] and then appear as yellow fluorescing rods on a dark background.

1.3.1.2 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[43, 44]. In settings with limited resources mycobacterial culture is often not performed routinely as it requires quite advanced laboratories with biosafety facilities.

1.3.1.3 Phenotypic identification of the MTC

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 and resistance to thiophene-2-carboxylic acid hydrazide (TCH) and production of pyrazinamidase are hampered by slow growth of MTC members and subjective interpretation of colony morphology and cross-resistance to drugs. They do not always allow unambiguous species identification in every case[45, 46]. 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 wall[47].

1.3.1.4 Genotypic identification of MTC

Molecular biological methods such as direct sequence determination of the 16S rRNA gene[48] 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[49, 50]. The reverse hybridization assay INNO-LiPA (Innogenetics NV, Ghent, Belgium) which amplifies the 16S-23S rRNA region identifies *Mycobacterium* species and differentiates the MTC[51]. The newer available DNA strip assay GenoType MTBC (Hain Lifescience GmbH, Nehren, Germany) targeting the 23S rRNA can conversely 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* BCG[52]. 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 level[45].

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* H37Rv[53]. 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 MTC[54]. 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)[55].

1.3.2 STRAIN IDENTIFICATION OF THE MTC

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 4,000 protein-coding sequences and has a very high guanine and cytosine content[56]. 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 examined[57]. 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.

1.3.2.1 RFLP

The method IS*6110* Restriction Fragment Length Polymorphism (RFLP) was in the early 1990s agreed upon as the standard typing method for TB epidemiology[58]. The insertion sequence IS*6110* is present in different copy numbers and integrated at different chromosomal sites in MTC isolates. The fragments based on IS*6110* are highly polymorphic but stable enough for epidemiological studies[59]. Strains with fewer copies of IS*6110* are more homogenous and the fingerprints of those are not as reliable concerning epidemiological links as of those containing multiple copies[60]. In the RFLP method[58], genomic DNA is digested with the restriction endonuclease *Pvu*II, 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 IS*6110* 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. Still, the RFLP method which requires large amounts of DNA is labour intensive, time consuming and inconvenient when quick results are needed.

1.3.2.2 Spoligotyping

Next to IS*6110* RFLP, spoligotyping is the most widespread method for TB epidemiology. Spoligotyping 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 (DR) locus which consists of multiple direct variable repeats (DVR). Each DVR is composed of a 36 bp-DR and a

non-repetitive short sequence also called spacer[61]. In short, after using primers that amplify the DR 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 kit[62]. Although widely used, the discriminatory power of spoligotyping is low to medium depending on the geographic setting and a second method is needed to underline epidemiological clusters. Spoligotypes can be assigned to the major phylogenetic lineages according to signatures provided in the international database SpolDB4.0[63]. This database defines 62 genetic lineages/sublineages of MTC isolates and contains to date more than 35,000 spoligopatterns for clinical isolates, with more than 1,900 shared types (STs; a common pattern shared by two or more isolates).

1.3.2.3 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[64, 65]. The discriminatory power of VNTR is low compared to IS*6110* RFLP and should for epidemiological studies be used in combination with another typing method. Nonetheless, for evolutionary studies VNTR may be useful[57].

Supply *et al* proposed and evaluated the use of 12 MIRUs as strain differentiation markers[66, 67]. Later, a standardized format comprised of 24 loci was proposed and evaluated[68, 69]. The MIRU technique has been shown to be reproducible and has a discriminatory power that resembles the IS*6110* RFLP method[57]. 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 that a standardized MIRU-VNTR genotyping method could be the new reference for epidemiological and phylogenetic screening of *M. tuberculosis* strains[68, 69]. In opposition, a number of studies have revealed the problematics of the differentiation power when it comes to strains of the Beijing genotype[70-72]. These findings suggest that the selection of MIRU-VNTR loci for optimal differentiation of *M. tuberculosis* requires further validation in different geographical settings.

1.3.2.4 LSP and SNP

The above mentioned genotyping tools make use of mobile (IS*6110* 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 unambiguously, 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[9, 28]. 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)[73]. The MTC PCR typing panel developed by Huard *et al*[54] 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[55, 74].

1.4 TUBERCULOSIS CONTROL

In 1993, the World Health Organization (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 results[75]. The programme comprises five components[76]:

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 counties by 2002 had accepted the DOTS strategy only 27% of infectious pulmonary TB cases were treated under DOTS programmes[75].

1.4.1 Tuberculosis in Sweden

In 1975, the 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 osteomylitis had been noticed. The now increasing proportion of foreign-born TB patients indicates the need to continue selective vaccination of children in families originating from countries with high TB incidence[77].

Sweden has one of the lowest TB incidences in the world. Figure 1 illustrates the number of new TB cases in Sweden during the past 20 years. From 1989 to 2003 there was a decrease in incidence that could be ascribed to a decrease among Swedish-born cases. During the past five years the number and proportion of foreign-born TB patients among all cases have increased. The particularly observed increase of TB in 2005, partly attributed to an outbreak at a day nursery, is a reminder of the serious consequences of delayed diagnosis[78, 79].

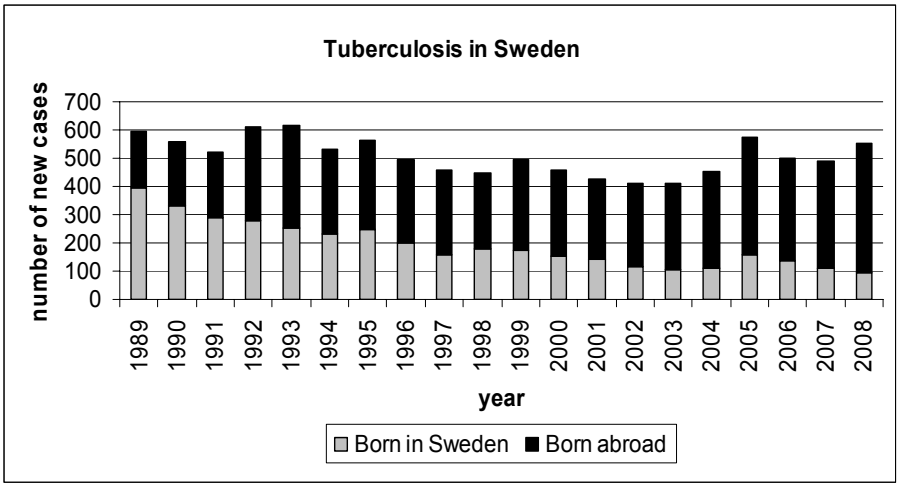

Figure 1 Number of new TB cases in Sweden among Swedish-born (grey) and foreign-born (black).

In 2008, 554 new cases (273 men and 281 women) of TB were reported generating an overall incidence of 6.0 cases per 100,000 population. The increase of 13% compared to 2007 was especially seen among foreign-born patients and highlights the need for physical examinations of individuals entering Sweden. The great majority (83%) was foreign-born patients with a median age of 30 years. Several (44%) originated from Africa, in particular Somalia. The TB incidence among foreign-born patients was 37.5 cases per 100,000 population and 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. The median age of Swedish-born TB patients is 65 years and the incidence 1.2 cases per 100,000 population. 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 neglected[80].

Of all new cases, 79% were culture verified and drug-resistance was seen among 13%. An increase of MDR TB has been noted with 14 new cases (3.2%) in 2008 and 15 new cases (4.1%) in 2007 compared to about four new cases in 1994-2006[80].

Although the DOTS programme has showed promising results in many countries it is only on exceptional cases basis employed in Sweden[81]. There are however recommendations that the treatment should be observed directly for the first two months.

1.4.2 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 48 years[82, 83]. 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. The WHO has estimated the TB incidence to be 219 cases per 100,000 population[2, 83]. 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 reported[84].

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.

2 THE PRESENT INVESTIGATION

2.1 PURPOSE OF THE STUDY

The aim of the 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.
- to characterize the molecular epidemiology of strains by means of IS*6110* RFLP.

2.2 MATERIALS AND METHODS

Below is a summary of the general methods used in papers I and II.

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, isoniazid, rifampicin, ethambutol or streptomycin) MTC isolates were sent to the reference TB-laboratory at the Swedish Institute for Infectious Disease Control. These isolates were obtained from all Swedish TB-laboratories situated in Stockholm, Gothenburg, Malmö/Lund, Linköping and Umeå.

2.2.1.2 Paper II

At three separate time points 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 Catío). Thirdly, in 2007 and 2008 samples were again collected from the reference TB hospital. Samples collected during the first, second and late parts of 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

Sputum samples were decontaminated of nonmycobacterial microorganisms by the sodium lauryl sulfate method[85] 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^{o}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[42, 86] 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^{o}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^{o}C. Incubations were made at 65^{o}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 DR 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 DR

region by PCR. The denatured PCR product was pipetted into the parallel channels in such a way that the channels of the miniblotter 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 and assigned to the major phylogenetic lineages.

2.2.5 RFLP

IS*6110* 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 IS*6110* sequence as a probe, which was non-radioactively labelled with peroxidase and later visualized using an enhanced chemiluminescence kit.

IS*6110* RFLP patterns were analyzed using the BioNumerics software version 5.10 (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 IS*6110* copies. The distance between patterns was computed using the un-weighted pair-group method of arithmetic averaging and the Jaccard index.

2.2.6 Single nucleotide polymorphism in the *narGHJI* operon

A selection of Guinea-Bissau isolates (paper II) 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 residue[87]. 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, we studied 400 isolates from 199 men and 201 women (53 born in Sweden and 347 born abroad). This represented 8.5% of all culture positive cases in Sweden during the study period. Of the 400 isolates investigates, 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 (but with identical RFLP pattern) samples from each patient. Four patients developed additional drug resistance, mainly due to poor compliance. 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). 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 East-African-Indian (EAI) genotype. Beijing strains, which have disseminated globally in the recent years, have been associated with drug-resistance[92]. In this study, 48 isolates of the Beijing genotype were found where ten of them 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 that 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.

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	South East Asia	Middle East Central Asia	Europe	South America	Unknown	Total	Cluster	Patients in cluster
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 IS*6110* 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 isoniazid-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 comprised 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 period[81] represents one of the largest ever reported. 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 Sweden[78]. 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 cases[78]. Figure 2 illustrates the number of new cases belonging to this cluster identified per year. Up to date (April 2009) 113 patients infected with the SMI-049 strain have been identified.

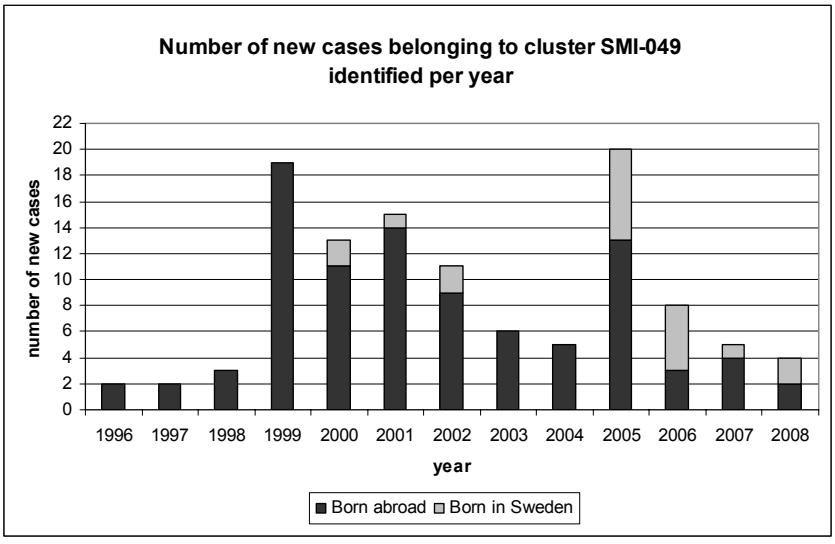

Figure 2 Number of new cases belonging to cluster SMI-049 identified per year.

2.3.2 Paper II

In paper II we reinvestigated the molecular epidemiology of strains from Guinea-Bissau with regards to the phylogeography of *M. africanum* in West Africa. Altogether, 418 isolates were investigated where 229 had been collected at the first, 142 at the second and 47 at the third time point. Spoligotyping was performed on all isolates and the spoligotypes were compared to the ones in the international database SpolDB4.0. Table 2 shows that 197 (47%) isolates belonged to the AFRI lineage and some other lineages also represented in Guinea-Bissau included EAI, Haarlem (H), Latin-American-Mediterranean (LAM), and the T family. Only seven Beijing strains were identified, representing 1.7% of the study sample. The Beijing family of strains distributed worldwide[88], predominates in Asia and is except for parts of South Africa[89] 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 Cameroon[30].

STs were assigned to the isolates present in SpolDB4.0. The major shared spoligotypes in the study were ST181 (AFRI1) and ST187 (AFRI1) and it was revealed that 84% of the AFRI isolates belonged to a spoligocluster.

When a selection of Guinea-Bissau isolates (n=29, representing all five biovars[90]) 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*).

The 197 AFRI spoligotype isolates all belonged to the previously described[12, 91] Guinea-Bissau family of strains. All, except for two isolates, had a spoligotype lacking spacers 7 to 9 and 39 and a distinct RFLP pattern with low numbers of IS*6110* insertions. The Guinea-Bissau family of strains lacks the regions of difference RD7, RD8, RD9 and RD10[91], as well as RD701 and RD702[11]. Altogether this profile classifies the Guinea-Bissau family of strains, irrespective of phenotypic biovar, as a part of the *M. africanum* West African 2 lineage. During the 15-year period since isolates were first collected, the proportion of isolates of the Guinea-Bissau family of strains appears to have diminished over time (table 2). A potential reason for the decline of *M. africanum* in West Africa could be a lower transmission capacity. There is huge uncertainty about the virulence of *M. africanum*[92]. 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.

Apart from the Guinea-Bissau family of strains other strains in this study showed a high degree of genetic diversity. Nine isolates belonged to the "ancient" EAI lineage whereas the rest were "modern" TB strains most likely imported from countries where such strains predominate. Guinea-Bissau, a former colony of Portugal, probably still today has more contacts with Portugal than neighbouring African countries.

Table 2 Global strain lineages of isolates collected in 1989-1994 (study A), 1994-1998 (study B) and 2007 and 2008 (study C).

Lineage	ST	Study A	Study B	Study C	Total
AFRI	188	2	1		3
AFRI	536	1	4	1	6
AFRI	537	1			1
(AFRI)		3	1		4
AFRI1	181	61	27	9	97
AFRI1	187	24	16	6	46
AFRI1	318	1	1		2
AFRI1	324		1		1
AFRI1	326	2			2
AFRI1	525	2			2
AFRI1	530	2			2
AFRI1	532	2			2
(AFRI1)		17	10	2	29
Beijing	1	1	4	2	7
EAI1 SOM	529	1	1		2
EAI1 SOM	538	1			1
EAI5	129		1		1
EAI5	236	1			1
EAI5	528	2	2		4
(EAI)		2			2
H1	47	8	8	2	18
H1	62		1		1
H1	531	1			1
H3	50	8	2	2	12
H3	75		1		1
H3	533	2			2
(H)		3			3
LAM 3 and S	4	1			1
LAM1	20	1	9	1	11
LAM2	604	1	1		2
LAM4	60	4	2	1	7
LAM4	828	1			1
LAM5	93		1		1
LAM9	42	21	13	2	36
LAM9	866	1			1
LAM10 CAM	61			2	2
(LAM)		6	5	1	12
MANU 2	54		1		1
(MANU)		1		1	2
T1	53	7	2	1	10
T1	196		2		2
T1	230	1			1
T1	244	2	3	2	7
T1	334	3			3
T1	521	1			1
T1	522	1		1	2
T1	535	2			2
T1	611	1			1
T1	766		1		1
T1	801			1	1
T1	804		1	2	3
T1	888	1			1
T2-T3	73	1	4	1	6
T5	44			1	1
(T)		4	4	2	10
U	458	2			2
U	523	1			1
U	527	7	2	1	10
U	534	3			3
U	1200		1		1
U	1204		1		1
(U)		1	1		2
X1	119	1	1		2
X3	92	1	1		2
Undefined	Undefined	4	5	3	12
Total		**229**	**142**	**47**	**418**

First column: lineage/sub lineage name [*M. africanum* (AFRI), Beijing, East-African-Indian (EAI), Haarlem (H), Latin-American-Mediterranean (LAM), Manu (an Indian family), T ("modern ill-defined family), U and Undefined]. Second column: shared-type (ST) number. Third to fifth column: number of isolates in each study belonging to the different lineages and shared types

2.4 CONCLUDING REMARKS

With the aim of characterizing MTC isolates phylogenetically and epidemiologically we collected isolates from the two entirely different settings, Sweden and Guinea-Bissau. In view of what was found in paper I it was especially interesting to continue investigating the different genotypes in Africa in paper II. 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 II 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 development[28]. 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 MTC world wide.

Paper I explains that molecular epidemiological techniques revealed one of the largest outbreaks ever reported. 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 epidemic has gradually taken hold of the Swedish-born population. Whether or not this strain has increased virulence and/or sociological factors are causing the extensive spread needs further investigation. It was illustrated that molecular epidemiological typing is a powerful tool to monitor and identify chains of transmission which indicate deficiencies in national TB control programs. The findings from our study should encourage Swedish health authorities to integrate full DOTS in the Swedish national TB control program. Screening of asylum seekers and immigrants from high-incidence countries is not mandatory in Sweden. A weak TB control program where patients are not identified, to not complete their treatment and are lost to follow-up will promote the development and spread of drug-resistant TB.

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

Paper II revisited Guinea-Bissau and studied the molecular epidemiology of strains with regards to the phylogeography of *M. africanum* in West Africa. During the 15-year period since isolates were first collected, the proportion of isolates of the Guinea-Bissau family of strains appeared to have diminished over time.

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 population. Observations suggest that particular lineages of the MTC might be adapted to specific human populations and maladapted to others[9]. 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 represented[9]. Other studies[93] 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.

Comparing the utterly different settings of Sweden and Guinea-Bissau provided us with some insight into the genetic diversity of the MTC in West versus East Africa. The low-resource country Guinea-Bissau has implemented DOTS and is struggling to reach the global targets set by the WHO. In Sweden, this therapy course is only on exceptional cases basis employed and in view of our findings it is our strong recommendation that Swedish health authorities should integrate full DOTS in the Swedish national TB control program.

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 no matter what the world looks like.

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

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

Sven Hoffner, for his direction, guidance, kind support and for creating opportunities.

Solomon Ghebremichael, for his true friendship, back-up, generosity and being an excellent travel companion.

The co-authors for their help; Ingela Berggren, Björn Petrini, Paulo Rabna, Raffaella Colombatti and Fabio Riccardi.

Past and present members of the amazing TB group; particularly Alexandra, Emma, Jim, Pontus, Melles, Maria, Anna, Andrzej, Jolanta, Lech and Lisbeth, for the recent actual support.

All superb colleages at SMI, especially at the Department of Bacteriology. Ingen nämnd, ingen glömd.

Britt-Inger Marklund, who strongly influenced me during my time at the University of Kalmar and bares the sole responsibility of me being at SMI today.

Diana Axelsson Olsson, who was there from the very beginning, for the fun times in the lab and for the everlasting practice of presentations. Thanks for knowing my environment and for being able to discuss it with me.

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

Christopher, my soon to be husband, for his endless love and patience while going through this. Our new life begins soon…

4 REFERENCES

1. Murray, J. F. A century of tuberculosis. Am J Respir Crit Care Med 169, 1181-6 (2004).
2. WHO. Report - Global tuberculosis control - surveillance, planning, financing. (2008).
3. Grange, J. M. & Zumla, A. The global emergency of tuberculosis: what is the cause? J R Soc Health 122, 78-81 (2002).
4. Murray, J. F. Tuberculosis and HIV infection: a global perspective. Respiration 65, 335-42 (1998).
5. Sotgiu, G. et al. Epidemiology and clinical management of XDR-TB: a systematic review by TBNET. Eur Respir J (2009).
6. 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, 9869-74 (1997).
7. Kapur, V., Whittam, T. S. & Musser, J. M. Is Mycobacterium tuberculosis 15,000 years old? J Infect Dis 170, 1348-9 (1994).
8. Gutierrez, M. C. et al. Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. PLoS Pathog 1, e5 (2005).
9. Gagneux, S. et al. Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 103, 2869-73 (2006).
10. 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).
11. Mostowy, S. et al. Genomic analysis distinguishes Mycobacterium africanum. J Clin Microbiol 42, 3594-9 (2004).
12. Kallenius, G. et al. Evolution and clonal traits of Mycobacterium tuberculosis complex in Guinea-Bissau. J Clin Microbiol 37, 3872-8 (1999).
13. 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, 631-4 (1998).
14. 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, 1236-45 (1997).
15. Rastogi, N., Legrand, E. & Sola, C. The mycobacteria: an introduction to nomenclature and pathogenesis. Rev Sci Tech 20, 21-54 (2001).
16. van Soolingen, D. et al. Diagnosis of Mycobacterium microti infections among humans by using novel genetic markers. J Clin Microbiol 36, 1840-5 (1998).
17. Cousins, D. V. et al. Tuberculosis in wild seals and characterisation of the seal bacillus. Aust Vet J 70, 92-7 (1993).
18. 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, 1785-9 (2003).
19. 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).
20. Wayne, L. G. Synchronized replication of Mycobacterium tuberculosis. Infect Immun 17, 528-30 (1977).
21. 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, 1267-8 (1995).
22. Brosch, R. et al. A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci U S A 99, 3684-9 (2002).
23. Cosivi, O. et al. Zoonotic tuberculosis due to Mycobacterium bovis in developing countries. Emerg Infect Dis 4, 59-70 (1998).

24. Calmette, A. La Vaccination Preventive Contre la Tuberculose. . Masson et cie (1927).
25. Quinn, T. C. Interactions of the human immunodeficiency virus and tuberculosis and the implications for BCG vaccination. Rev Infect Dis 11 Suppl 2, S379-84 (1989).
26. Andersen, P. & Doherty, T. M. The success and failure of BCG - implications for a novel tuberculosis vaccine. Nat Rev Microbiol 3, 656-62 (2005).
27. Castets, M., Boisvert, H., Grumbach, F., Brunel, M. & Rist, N. [Tuberculosis bacilli of the African type: preliminary note]. Rev Tuberc Pneumol (Paris) 32, 179-84 (1968).
28. Gagneux, S. & Small, P. M. Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. Lancet Infect Dis 7, 328-37 (2007).
29. de Jong BC, et al. 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.47, 994-1001 (2009).
30. Niobe-Eyangoh, S. N. et al. Genetic biodiversity of Mycobacterium tuberculosis complex strains from patients with pulmonary tuberculosis in Cameroon. J Clin Microbiol 41, 2547-53 (2003).
31. Godreuil, S. et al. First molecular epidemiology study of Mycobacterium tuberculosis in Burkina Faso. J Clin Microbiol 45, 921-7 (2007).
32. Zink, A. R. et al. Characterization of Mycobacterium tuberculosis complex DNAs from Egyptian mummies by spoligotyping. J Clin Microbiol 41, 359-67 (2003).
33. Desmond, E. et al. Mycobacterium africanum cases, California. Emerg Infect Dis 10, 921-3 (2004).
34. Miltgen, J. et al. Two cases of pulmonary tuberculosis caused by Mycobacterium tuberculosis subsp canetti. Emerg Infect Dis 8, 1350-2 (2002).
35. Wells, A. Q., and D. M. Oxon. Tuberculosis in wild voles. . Lancet i (1937).
36. Huitema, H., and F. H. J. Jaartsveld. . Mycobacterium microti infection in a cat and some pigs. . Antonie Leeuwenhoek 209-212 (1967).
37. 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).
38. Xavier Emmanuel, F. et al. Human and animal infections with Mycobacterium microti, Scotland. Emerg Infect Dis 13, 1924-7 (2007).
39. 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, 1305-14 (2003).
40. 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, 1263-73 (1999).
41. Kent, P. & Kubica, G. Public Health Mycobacteriology. A guide for the level III laboratory. (Centers for Disease Control Atlanta, 1985).
42. Bloom, B. R. Tuberculosis pathogenisis, protection and control (ASM Press, Washington D.C. , 1994).
43. Kallenius, G., Hoffner, S. E., Miorner, H. & Svenson, S. B. Novel approaches to the diagnosis of mycobacterial infections. Eur Respir J 7, 1921-4 (1994).
44. 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, 1099-108 (2001).
45. 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).
46. 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).

47. Butler, W. R. & Guthertz, L. S. Mycolic acid analysis by high-performance liquid chromatography for identification of Mycobacterium species. Clin Microbiol Rev 14, 704-26, table of contents (2001).
48. 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, 2882-9 (1993).
49. 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, 4303-6 (2004).
50. 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, 2427-31 (1992).
51. 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, 1915-9 (2000).
52. 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, 2672-5 (2003).
53. Behr, M. A. et al. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science 284, 1520-3 (1999).
54. 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, 1637-50 (2003).
55. Huard, R. C. et al. Novel genetic polymorphisms that further delineate the phylogeny of the Mycobacterium tuberculosis complex. J Bacteriol 188, 4271-87 (2006).
56. Cole, S. T. et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393, 537-44 (1998).
57. 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).
58. van Embden, J. D. et al. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 31, 406-9 (1993).
59. 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, 2578-86 (1991).
60. 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, 1987-95 (1993).
61. 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, 2695-705 (1991).
62. Kamerbeek, J. et al. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol 35, 907-14 (1997).
63. 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).
64. 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), 1189-96 (1998).

65. 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, 2607-18 (1999).
66. Supply, P. et al. Variable human minisatellite-like regions in the Mycobacterium tuberculosis genome. Mol Microbiol 36, 762-71 (2000).
67. 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, 3563-71 (2001).
68. 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, 1398-406 (2008).
69. 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, 691-7 (2007).
70. 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).
71. 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, 1045-9 (2008).
72. 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, 3338-45 (2008).
73. 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, 759-72 (2006).
74. 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, 1533-43 (2002).
75. WHO. Treatment of Tuberculosis: guidelines for national programmes 3rd edition (2003).
76. WHO. Factsheet What is DOTS? .
77. Romanus, V. Selective BCG vaccination in a country with low incidence of tuberculosis. Euro Surveill 11, 14-7 (2006).
78. 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).
79. 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 5 (2005).
80. Romanus, V. Smittskyddsinstitutets årsrapport 2008: Statistik för tuberkulos. . (2008).
81. Ghebremichael, S. et al. [Resistant tuberculosis is spreading in Sweden. Molecular epidemiological strain identification by "fingerprinting" can make the infection tracing easier]. Lakartidningen 99, 2618-9, 2622-3 (2002).
82. CIA. The 2008 World Factbook (Central Intelligence Agency, USA, 2008).
83. WHO. TB country profile Guinea-Bissau. (2007).
84. 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, 463-7 (2007).

85. 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).
86. 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, 1114-6 (2003).
87. Stermann, M., Bohrssen, 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, 3252-9 (2003).
88. 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, 843-9 (2002).
89. Cowley, D. et al. Recent and rapid emergence of W-Beijing strains of Mycobacterium tuberculosis in Cape Town, South Africa. Clin Infect Dis 47, 1252-9 (2008).
90. Hoffner, S. E. et al. Biochemical heterogeneity of Mycobacterium tuberculosis complex isolates in Guinea-Bissau. J Clin Microbiol 31, 2215-7 (1993).
91. Koivula, T. et al. Genetic characterization of the Guinea-Bissau family of Mycobacterium tuberculosis complex strains. Microbes Infect 6, 272-8 (2004).
92. 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, 1037-43 (2008).
93. Baker, L., Brown, T., Maiden, M. C. & Drobniewski, F. Silent nucleotide polymorphisms and a phylogeny for Mycobacterium tuberculosis. Emerg Infect Dis 10, 1568-77 (2004).