Analysis of Mycobacterium tuberculosis H37Rv antigens, and their potential use in the development of a diagnostic kit for detection of active TB

Analyse av Mycobacterium tuberculosis H37Rv antigener, og deres potensielle bruk i utviklingen av et diagnose-kit for deteksjon av aktiv TB

Nermin Zecic

Norwegian University of Life Sciences
Department of Chemistry, Biotechnology and Food Science
Master Thesis 60 credits 2012

Acknowledgements

I would like to thank everyone at the Norwegian Institute of Public Health for allowing me to write my thesis, as well as helping me along the way. The department of SMBI has been especially helpful; this would not have been possible without your support.

My deepest gratitude goes to Ulf R. Dahle and Øistein Ihle, for their guidance and help throughout this research. My time was short, but it was none the less a rich learning experience, both practical and theoretical.

I would also like to thank Prof. Tor Lea at the University of Life Science, UMB.

Finally, I would like to express my gratitude to all my friends and family for their support and encouragement.

Oslo, 2012

……..

Nermin Zecic

Abstract

TB is a serious disease caused by *Mycobacterium tuberculosis*. The bacterium causes inflammation that may ultimately lead to tissue death around the area of infection. The tubercle bacilli usually manifests in the lungs, but can also be found in extra pulmonary parts of the body. TB has become a worldwide disease that has been plaguing humanity for millennia. The bacterium was isolated and characterized for the first time in 1882 by Dr. Robert Koch. Since then, there have been numerous developments in both available treatments and diagnostic methods. Before the discovery of chemotherapy, about 50 % of TB patients died within 5 years following infection. The discovery of antibiotics such as streptomycin gave hope that the disease could be completely eradicated. This however did not last long, as reports of drug resistant TB, multi-drug resistant TB (MDR-TB), and later XDR-TB (greater range of antibiotic resistance) began to surface.

In Norway, the disease was very common up until 1960s, but declined dramatically when living conditions improved and vaccination, chemotherapy and antibiotics became available. In 2006 nearly 80 % of TB incidences in Norway were associated with immigrants. Many other countries are also challenged by TB import brought on by foreign-borne individuals. Despite the control programs present and the available treatment, it is estimated that TB will claim some 50-70 million lives in the timeframe 1998-2020

About 80 % of the global TB incidences happening in the present are mostly associated to 22 countries. These areas are referred to as “high-burden countries” by the World Health Organization (WHO). TB is mainly transmitted through droplets expelled when a sick individual coughs or sneezes in close contact with others. Most of these countries are either underdeveloped (low livings standards and poor hygiene) or contain overpopulated areas, creating the ideal environment for transmission. This has been especially the case in India and China. However for the countries in Africa, the high rate of TB has mainly been influenced by the HIV pandemic that started in the early 1990s.

A conclusive diagnosis of TB can only be made by culturing *M. tuberculosis* from a specimen taken from the patient. This is generally time-consuming and uncertain, due to the nature of

the slow-growing organisms. A diagnosis made other than by culture (microscopy, clinical, serological etc) may only be classified as “probable” or “likely”. Most of these methods require advanced technology, which may not be available in resource-limited settings.

Earlier studies have previously demonstrated elevated levels of H37Rv proteins in sera from patients infected by this particular strain. However, for this study, we chose to examine if some of these proteins could also be used to detect active TB infections caused by other strains of *M. tuberculosis*.

We selected 42 unknown strains of *M. tuberculosis* from *The National Reference Laboratory for Mycobacteria* at Norwegian Institute of Public Health. All of the strains used throughout this study were characterized with spoligotyping and RFLP. Upon having retrieved our results, we assigned the strains to various lineages of *M. tuberculosis* and selected 8 strains with focus on those that originated from the high-burden countries. Our strains were grown on Löwenstein-Jensen over a period of 4-6 weeks. The colonies were heat-killed, and their DNA extracted and purified. The selected strains were tested for their ability to produce 5 different TB proteins by the use of PCR amplification and DNA sequencing. The primers used to amplify the desired DNA sequences were derived from 5 different strains of H37Rv; Rv3881c, Rv0934, Rv1886c, Rv1759c and Rv3874c.

Based on the sequencing results, we confirmed that 4 out of the 5 tested proteins had very high sequence similarities with our 8 TB-strains. They should all therefore be able to produce these peptides. Thus the diagnostic kit under development could potentially be used to identify infections caused by either of these diverse strains of *M. tuberculosis*. If a diagnostic kit could be developed based on these peptides, the kit could be equally sensitive in different epidemiological settings throughout the world.

Sammendrag

TB er en alvorlig sykdom forårsaket av *Mycobacterium tuberculosis*. Sykdommen forårsaker inflammasjon og kan føre til at vevet dør rundt infeksjonsområdet. Tuberkelbasillen manifesterer seg vanligvis i lungene, men kan også finnes i andre deler av kroppen. I dag er TB en verdensomfattende sykdom, og har plaget menneskeheten i årtusener. Bakterien ble isolert og karakterisert av Dr. Robert Koch for første gang i 1882. Siden da har det blitt utviklet en rekke behandling- og diagnosemetoder. Før oppdagelsen av kjemoterapi, døde rundt 50 % av TB pasienter i løpet av de 5 første årene som følge av infeksjonen. Oppdagelsen av antibiotika slik som streptomycin ga håp om at sykdommen kunne bli utryddet en gang for alle. Dette varte dessverre ikke lenge, da rapporter om resistens, multi-resistens (MDR-TB), og senere XDR-TB (utvidet multiresistens) begynte å komme.

I Norge var TB en folkesykdom helt frem til 1960-tallet, men ble dramatisk redusert når levestandarden ble forbedret og vaksinasjon, kjemoterapi og antibiotika ble tilgjengelig. I 2006 var 80 % av alle TB-tilfellene i Norge var knyttet til innvandrere. Mange andre land sliter også grunnet import av TB fra personer født i utlandet. Selv med dagens kontrollprogrammer inkluderer tilgjengelig behandling, estimeres det at TB vil føre mellom 50-70 millioner mennesker i døden i løpet av tidsperioden 1998-2020.

TB blir spredt hovedsakelig via dråpesmitte som følge av at en infisert person hoster eller nyser i nærheten av andre. Omlag 80 % av de globale TB-tilfeller i dag kan assosieres med 22 land. Disse områdene blir referert til som høy-byrde land av Verdens Helseorganisasjon. De fleste av disse landene er enten underutviklet (lav levestandard og hygiene) eller inneholder overbefolkede områder, noe som skaper det ideelle miljø for smitte. Dette har vert spesielt tilfelle i India og Kina. I Afrika derimot, har TB-smitten stort sett vert påvirket av HIV pandemien som startet på 1990-tallet.

En sikker TB-diagnose krever at M. tuberculosis påvises. Dette kan gjøres I form av dyrkning, mikroskopi eller indirekte ved PCR-påvisning av M. tuberculosis i prøver tatt fra en pasient. Dette er tidkrevened og upålitelig da *M. tuberculosis* vokser svært sakte og er vanskelig å dyrke eller påvise. Diagnose stillet ved kliniske funn eller serologiske undersøkelser er kun indikativ og krever ofte avansert teknologi som ikke alltid er tilgjengelig i fattige deler av verden.

Tidligere undersøkelser har påvist høye nivåer av H37Rv proteiner i serum fra pasienter smittet av denne stammen. Disse proteinene burde derfor gi en måte å oppdage pasienter med aktiv TB forårsaket av denne stammen. I denne studien derimot, valgte vi å undersøke om disse proteinene også kunne brukes til å detektere infeksjoner fra andre TB-stammer.

Vi valgte 42 ukjente stammer av *M. tuberculosis* fra Det Nasjonale Referanselaboratoriet for Mycobacterier ved Folkehelseinstituttet. Alle stammene ble karakterisert ved bruk av spoligotyping og RFLP. På bakgrunn av disse resultatene, identifiserte vi isolatenes familietilhørighet og selekterte 8 stammer med fokus på de som dominerer i høy-byrde land. De utvalgte stammene ble dyrket på Löwenstein-Jensen agar over en periode på 4-6 uker. Koloniene ble varme-drept og deres DNA ekstrahert og renset. Det ble undersøkt om de selekterte stammene var i stand til å produsere 5 ulike H37Rv proteiner ved bruk av PCR og DNA sekvensering. Primere som ble brukt til å amplifisere de ønskede DNA-sekvensene stammet fra 5 ulike stammer av H37Rv; Rv3881c, Rv0934, Rv1886c, Rv1759c and Rv3874c.

På bakgrunn av sekvenseringsresultatene, bekreftet vi at 4 av de 5 testede proteinene hadde veldig høy sekvenslikhet med våre stammer. De testede stammene bør derfor i teorien være i stand til å produsere disse proteinene. Diagnose Kit'et som er under utvikling, har derfor potensiale til å identifisere infeksjoner fra alle *M. tuberculosis* familier som vi testet. Hvis et kit'et utvikles basert på disse peptidene, kan den være like senestivt i ulike epidemilogiske lokalisasjoner verden over.

Table of Contents

Section I: Introduction……..6

1.1.1 The current state of TB……6

1.2 Mycobacterium tuberculosis complex (MTC)……12

1.3 The Origin of *Mycobacterium tuberculosis*……...14

1.4 Diagnosis……15

1.4.1 Microbiological diagnosis……...15

1.4.2 Immunological diagnosis……17

1.4.3 Clinical diagnosis……18

1.5 Pathogenesis……...20

1.5.1 Disease progression……20

1.5.2 Treatment……...21

1.6 Aim of the study……23

Section II: Materials and Methods……...24-38

Section III: Results……39-46

Section IV: Discussion……47

4.1 Spoligotyping……48

4.2 Restriction Fragment Length Polymorphism……49

4.3 Molecular methods……...49

4.4 DNA sequencing……...49

Section V: Conclusion……50

Section VI: References……..51-55

Section I: Introduction

1.1.1 The Current state of Tuberculosis (TB)

During the 20th century, the developed world witnessed a steady decline in TB incidents (**Figure 1.1.4**), which lead to dramatic cutbacks in funds for both research and control (***Kaufmann SH & Parida SK, 2007***). In the last decade however, the incidence rates of TB have fallen in wealthy countries but risen in developing countries. This development has been fueled by the HIV pandemic that started in the early 1990s, in sub-Saharan Africa (***Frieden TR et al 1995***). Although great efforts have been made to eradicate the disease, the numbers of incidents today are higher than ever (***Kaufman et al, 2005***). It is estimated that around 9 million individuals are infected each year, 2 million of these are fatal, and as much as one third of the entire human population is infected. However most of the infected individuals do not develop active TB, rather they remain healthy carriers of latent tuberculosis (LTB) (***Lawn SD et al, 2011***). Individuals with active TB are able to infect others given a longer period of exposure, either through coughing or sneezing. Carriers of latent TB however, do not display any symptoms, nor are they able to transmit the disease onto others.

The World Health Organization has made a list over the 22 areas of the world with the highest burden of TB (**Table 1.1.1)**. Together they account for over 80 % of the total estimated TB cases around the world (**Figure 1.1.1**).

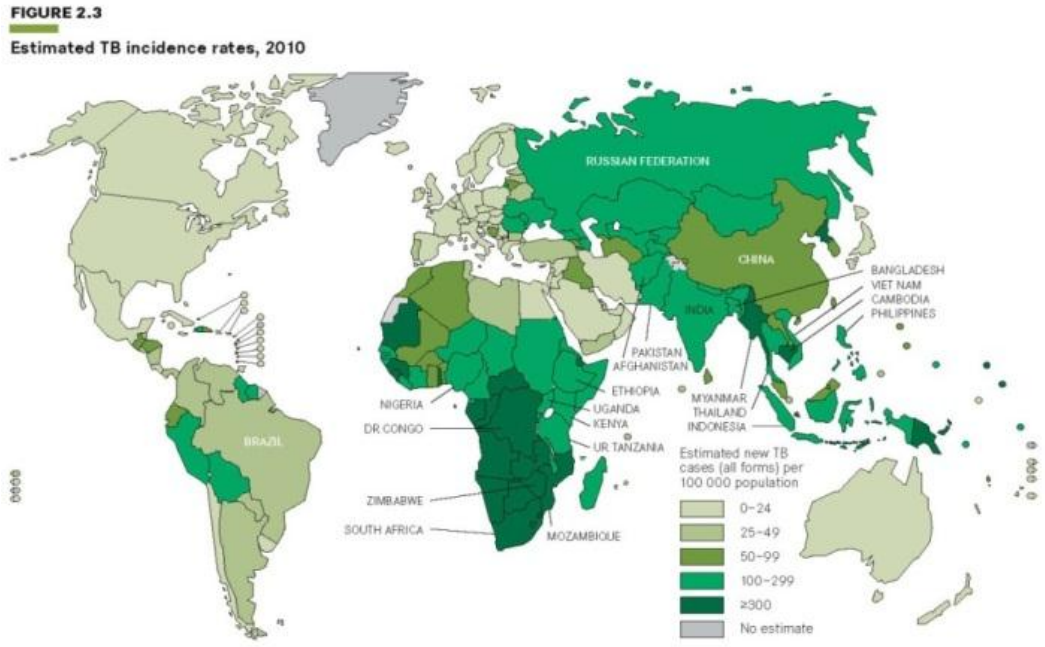


Figure 1.1.1 Shows the 22 high-burden countries, and the estimated number of TB incidents as of 2010. The numbers in Africa and Asia are severe compared to the other regions. Source: World Health Organization.

Neglect of well-known public health principles fuelled the global TB epidemic is the reason why WHO declared TB a global health emergency in 1993. Up until the early-1990s, very little was known about the biology of the *M. tuberculosis* bacillus. The sudden spread that followed caught health-care workers, scientists and governments' off-guard. There are numerous examples where financial cutbacks in the national health systems have led to re-emergence of different diseases. The greatest impact however has been observed in developing countries that are unable to sustain public health programs such as those that are established in industrialized parts of the world.

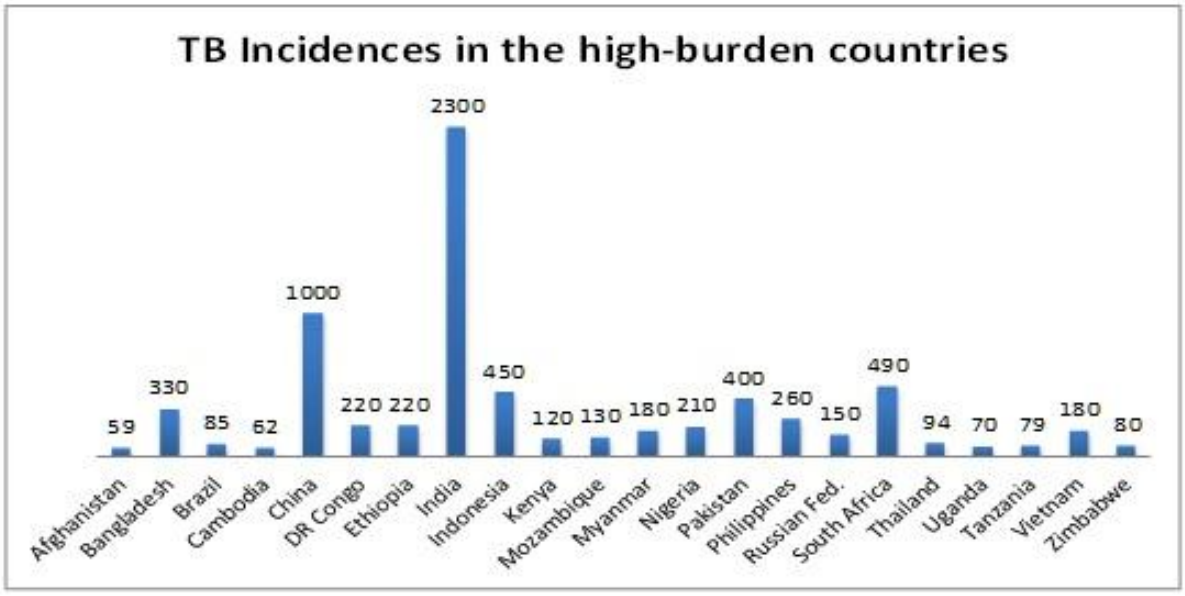


Figure 1.1.2 Shows the global TB incidences for the 22 high burden countries, the numbers displayed are in thousands. The incidence numbers in India and China is exceptionally high. Source: WHO

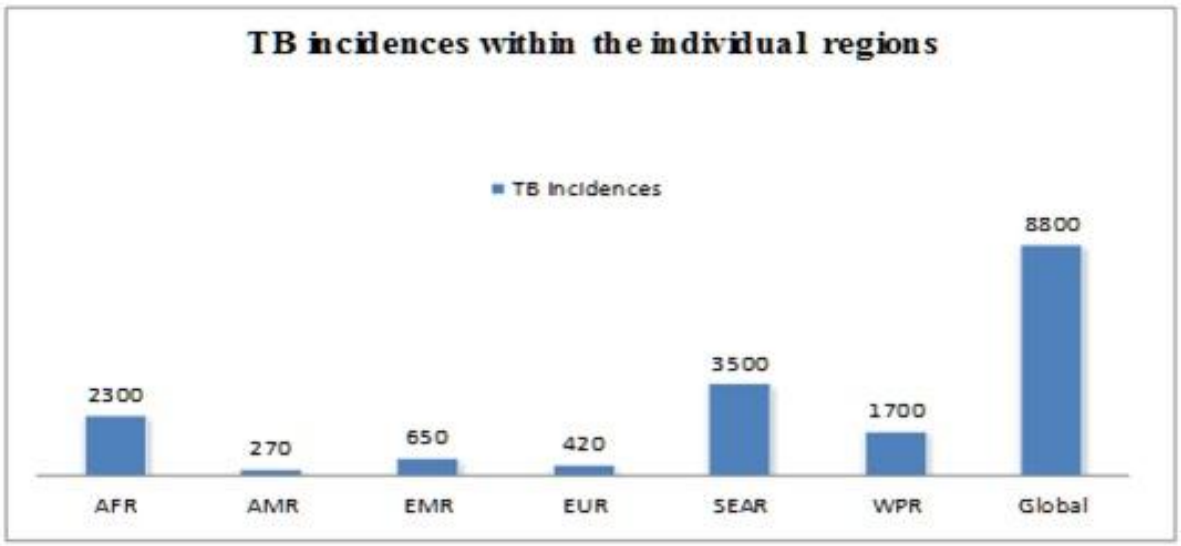


Figure 1.1.3 Shows the distribution of TB incidences between the different regions in comparison to the total global numbers. The numbers displayed here are in thousands. AFR=Africa; AMR=Americas; EMR=Eastern Mediterranean; EUR=Europe; SEAR=South-East Asia; WPR=Western Pacific. Source: WHO.

Ineffective tools for the diagnosis of TB, combined with poverty and a high mortality rate, may be one of the greatest challenges to an already suffering population. A large number of young adults die each day, effectively crippling the workforce and further slowing down development within the country, as well as the fight against TB. The Millennium Development Goal of 2015, signed by 193 nations, has listed the eradication of several worldwide diseases as one of its top priorities, which includes TB in addition to HIV and malaria, under **Target 6C**. The aim of the MDG is to reduce global TB incidence by 50 % within the year of 2015, and hopefully eliminate the disease by the end of 2050. However the 2010 mid-term report revealed that the goals are far from being met, especially in Africa (***Richard E et al, 2008***). There is also evidence that the current diagnostic and treatment methods will not be sufficient in achieving this goal (***Young DB et al, 2008***)

Table 1.1.1 Displays the amount of TB and HIV infections in the 22 high-burden countries, the numbers are in thousands, and represent the situation as of 2010, as listed by the World Health Organization.

		Mortality			Relevance			Incidence			HIV-positive TB cases		
Country	Population	Best	Low	High	Best	Low	High	Best	Low	High	Best	Low	High
Afghanistan	31412	12	8.6	16	110	51	180	59	49	71	-	-	-
Bangladesh	148692	64	47	85	610	280	1000	330	270	400	0.7	0.3	1.1
Brazil	194946	5	3.1	8.3	92	34	160	85	70	100	18	15	22
Cambodia	14138	8.6	6.2	12	93	42	150	62	53	72	4	3.4	4.7
China	1341335	54	52	56	1500	1300	1700	1000	910	1200	18	10	28
DR Congo	65966	36	27	45	350	160	560	220	190	250	18	13	24
Ethiopia	82950	29	23	35	330	140	520	220	200	230	-	-	-
India	1224614	320	210	470	3100	2000	4600	2300	2000	2500	110	75	160
Indonesia	239871	64	42	91	690	300	1200	450	370	540	18	9.9	29
Kenya	405413	6.9	4.9	9.4	110	49	180	120	120	130	50	45	55
Mozambique	23391	11	7.0	17	110	54	200	130	87	170	77	53	110
Myanmar	47963	20	12	31	250	180	310	180	160	210	37	21	57
Nigeria	158423	33	11	68	320	110	690	210	99	360	51	25	87
Pakistan	173593	58	39	84	630	270	1100	400	330	480	1.2	0.7	1.9
Philippines	93261	31	21	43	470	410	530	260	210	310	1	0.5	1.8
Russian Fed.	142958	23	16	41	190	70	330	150	130	180	8.1	6.8	9.4
South Africa	50133	25	16	38	400	180	630	490	400	590	300	240	350
Thailand	69122	11	7	16	130	55	210	94	78	110	15	13	18
Uganda	33452	5.1	3.3	7.3	64	32	100	70	56	85	38	30	46
Tanzania	44841	5.8	4.7	6.9	82	39	130	79	75	85	30	28	32
Vietnam	87848	29	19	43	290	130	510	180	130	220	7.6	4.6	11
Zimbabwe	12571	3.4	2.1	5.1	51	23	80	80	61	100	60	47	76
High Burden	4321967	860	730	1000	10000	8500	12000	7200	6800	7500	860	780	950
AFR	836970	250	220	280	2800	2300	3300	2300	2100	2500	900	820	980
AMR	933447	20	17	23	330	260	410	270	250	280	35	31	38
EMR	593747	95	74	120	1000	670	1500	650	580	730	12	9.8	15
EUR	896480	61	48	75	560	430	720	420	390	450	20	19	22
SEAR	1807594	500	370	640	5000	3700	6500	3500	3200	3700	190	140	230
WPR	1798335	130	120	150	2500	2200	2800	1700	1500	1800	35	26	45
Global	6869573	1100	920	1200	12000	11000	14000	8800	8500	9200	1100	1000	1200

Andorra, Antigua and Barbuda, Australia, Austria, Barbados, Belgium, Bermuda, British Virgin Islands, Canada, Cayman Islands, Chile, Costa Rica, Cuba, Cyprus, Czech Republic, Denmark, Dominica, Finland, France, Germany, Greece, Grenada, Iceland, Ireland, Israel (including the Occupied Palestinian Territory, and including East Jerusalem), Italy, Jamaica, Jordan, Lebanon, Libyan Arab Jamahiriya, Liechtenstein, Luxembourg, Malta, Monaco, Montserrat, Netherlands, Netherlands Antilles, New Zealand, Norway, Oman, Puerto Rico, Saint Kitts and Nevis, Saint Lucia, San Marino, Slovenia, Sweden, Switzerland, Trinidad and Tobago, Turks and Caicaos, United Arab Emirates, United Kingdom, United States of America, United States Virgin Islands and Vatican City.

Figure 1.1.4 Shows a list of the low-incidence countries, most of these are developed with high living standards and control programs. Source: WHO

Considering the current situation from a global perspective (**Figure 1.1.1**), it is obvious that low-incidence countries are challenged by import of TB from high-incidence regions.

It is well known that a lot of TB incidences in Australia, Canada, Europe, and the United States occur in foreign-born individuals, and that the risk of TB is great in the first years after immigrants from high-burden countries arrive in a host country. However, TB can persist within immigrant communities for extended periods, and the prevalence of the disease increases with the age of said individuals. It is therefore less clear how importation of TB influence populations over extended periods, such as years or decades, because detailed characterization of complete microbial collections are difficult to obtain. Also due to chronic nature of TB and the persistence of substantial immigrant populations with latent TB in low-incidence countries, it may take many years before the full impact of imported TB can be measured.

Programs to keep TB from rising in many low-incidence countries are perceived to be challenged by international travel and immigration from high-burden countries. Therefore, TB has been the subject of much concern in the recent years. Although immigrants from high-burden countries are often stigmatized, and considered a threat to local control programs and public health, it has been demonstrated that local control programs can control the disease despite high rates of TB import.

A study conducted in Norway stretching over a period of 12 years (1994-2005), demonstrated that TB import from high-burden countries (**Table 1.1.1**) played only a minor role in the total domestic rate of transmission, and that the current control strategies in Norway were adequate to keep the transmission at minimum ***(Dahle, U et al 2007).*** It is therefore clear that immigrants from high-burden regions do bring with them more strains of *M. tuberculosis*, but they do not significantly contribute to the spread of TB within the resident population.

It has been hypothesized that a co-evolution between human hosts and bacilli must have been one of the main reasons why TB has re-emerged steadily and regionally throughout the centuries ***(Sola, C et al 2001)***. Based on earlier studies, we are able to link a host's geographical origin to a strain of TB, indicating a stable association between host, TB bacilli and environment (***Hirsh, Ae et al 2004***).

Due to the great variety of TB strains found around the world, combined with difficult diagnosis, the incidence of TB today is at an all-time high. It should also be noted that a growing number of TB infections can be linked to a growing number of HIV cases in some parts of the world ***(Alison, E et al 1990)***. These individuals are especially exposed due to an already weakened immune system, therefore an HIV infection in many cases may well be an indicator that the patients is also suffering from TB ***(Elizabeth, L et al 2006)***.

The challenge to fight TB appears overwhelming in some parts of the world. The TB incidence rate is higher today than any other time in history, and we still lack reliable means to diagnose the disease, individuals cannot protect themselves from transmission and the disease induces no immunity in the host. Surveillance programs have been questioned, as have the policies for Bacillus Calmette-Guerin (BCG) immunization and screening of immigrants. The latter issue coincides with the fact that previous exposure to MTC does not provide protection against subsequent infections. The human body is unable to develop immunity, and a person could contract TB shortly after being cured (***Small, PM et all 1993***). Vaccine development and serological diagnostics are seriously challenged. Tackling TB still depends on effective surveillance, prevention and control as well as established systems that are already in place in low-incidence countries. Ideally, all of these should be established throughout the world.

1.2 Mycobacterium tuberculosis complex (MTC)

The cause of TB was unknown until the end of 18th century. It wasn't until 1882 that the *M. tuberculosis* bacteria was isolated and characterised by Dr. Robert Koch, for which he was awarded the Nobel Prize in 1905. Every year on the 24.03 has since been marked as the world tuberculosis day.

TB is caused by bacterial species within the *Mycobacterium tuberculosis* complex (MTC), commonly referred to as tubercle bacilli. The members include M*ycobacterium tuberculosis, Mycobacterium africanum, Mycobacterium bovis, Mycobacterium bovis BCG, Mycobacterium microt, Mycobacterium canetti, Mycobacterium pinnipedii and Mycobacterium mungi.* The most common member is *M. tuberculosis* which is a non-motile, gram-positive, obligate aerobe, large pleomorphic rod. It is a facultative intracellular parasite (macrophages), with a growth rate of about 18-20 hours.

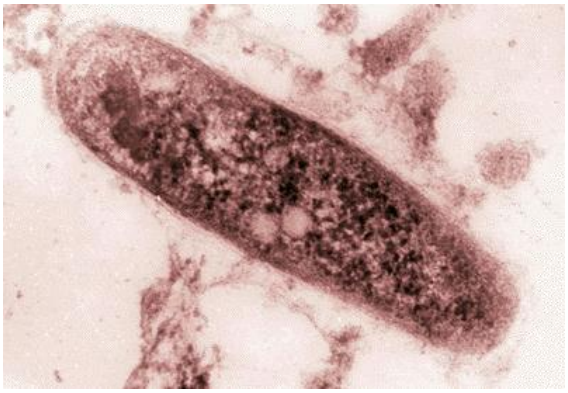


Figure 1.2.2 Shows an electron micrograph of *M. tuberculosis*. Source: *wadsworth.org*

M. tuberculosis thrives in parts of the body with high oxygen levels, hence being primarily a pulmonary disease. Due to the nature of their wax-like cell membranes, they often clump together **(Figure 1.2.3),** making them impermeable to traditional gram staining, unless the dyes are combined with phenol. Once the bacterium has been stained however, it resists de-colorization with acidified organic solvents. The bacterium is therefore known as *acid fast bacilli*, and the method is referred to as acid fast staining.

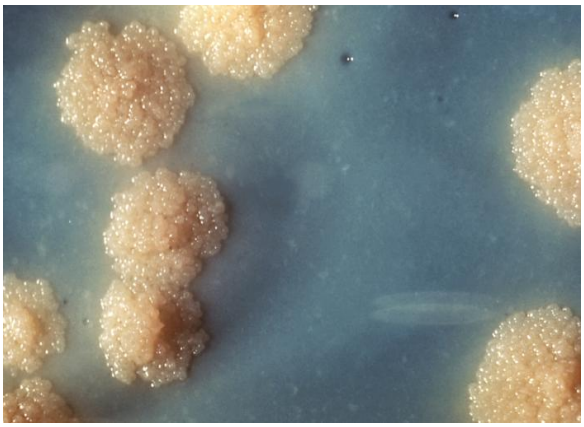


Figure 1.2.1 Shows an image containing several colonies of *M. tuberculosis* growing on Lownstein Jensen medium. The bacteria are tightly lumped together due to the nature of their membrane. Source: *redOrbit.com,* 25 April, 2011.

The membrane of *M. tuberculosis* differs greatly from other microbes, due to the production of three distinct types of mycolic acids; alpha- (70 %), methoxy- (10-15 %), and keto mycolic acid (10-15 %) (**Figure 1.2.1**). The bacterium is therefore adapted to survival in harsh environments, and is resistant to hydrophobic antibiotics, as well as chemical damage and dehydration.

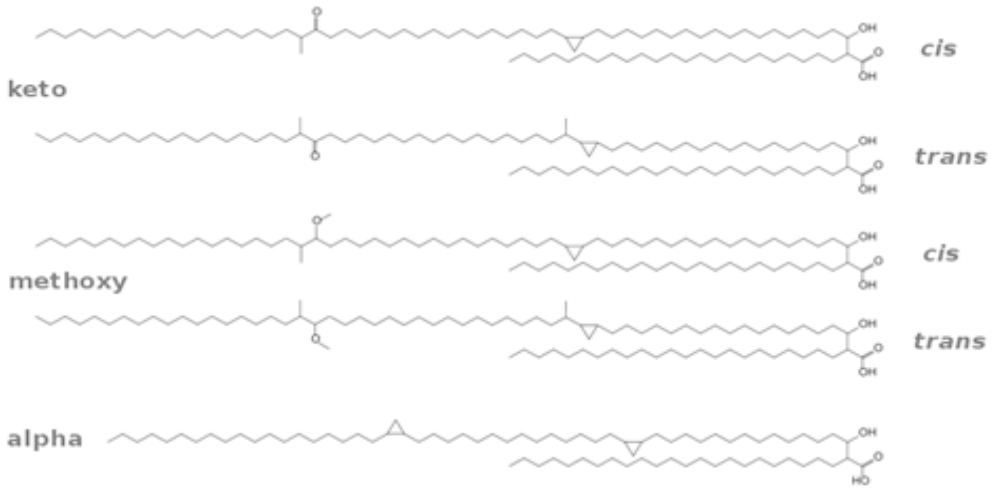


Figure 1.2.3 Shows the different types of mycolic acids produced by *M. tuberculosis*, the image was made using a ChemDraw. (Illustration: stud. chem., UMB, Simen Gjelseth Antonsen)

1.3 The origin of Mycobacterium tuberculosis

Although there are a number of different strains of TB, they are usually adapted to different mammalian hosts. For some time it was believed that TB transitioned from animal to human host when early nomads settled and started using livestock, approximately 8000 BC. This belief may have been linked to the discovery of TB in the remains of a bison, which lived 17 000 years ago ***(Rothschild, B 2001)***. Studies of Egyptian mummies have also shown traces of TB in both the spine and the head of the deceased individuals. This confirms that modern strains of TB have been around for at least 4 000 years ***(Zink, A et al 2006).*** It has been suggested that the genus Mycobacterium has been around for 150 million years, there is also evidence that TB has been infecting early hominids some 3 million years ago.

The belief that *M. tuberculosis* was a specialized form of *M. bovis* was however challenged in more recent times. Newer studies have proposed that *M. bovis* and *M. tuberculosis* both evolved from a common ancestor, rather than host transition ***(Brosch, R et al 2002)***.

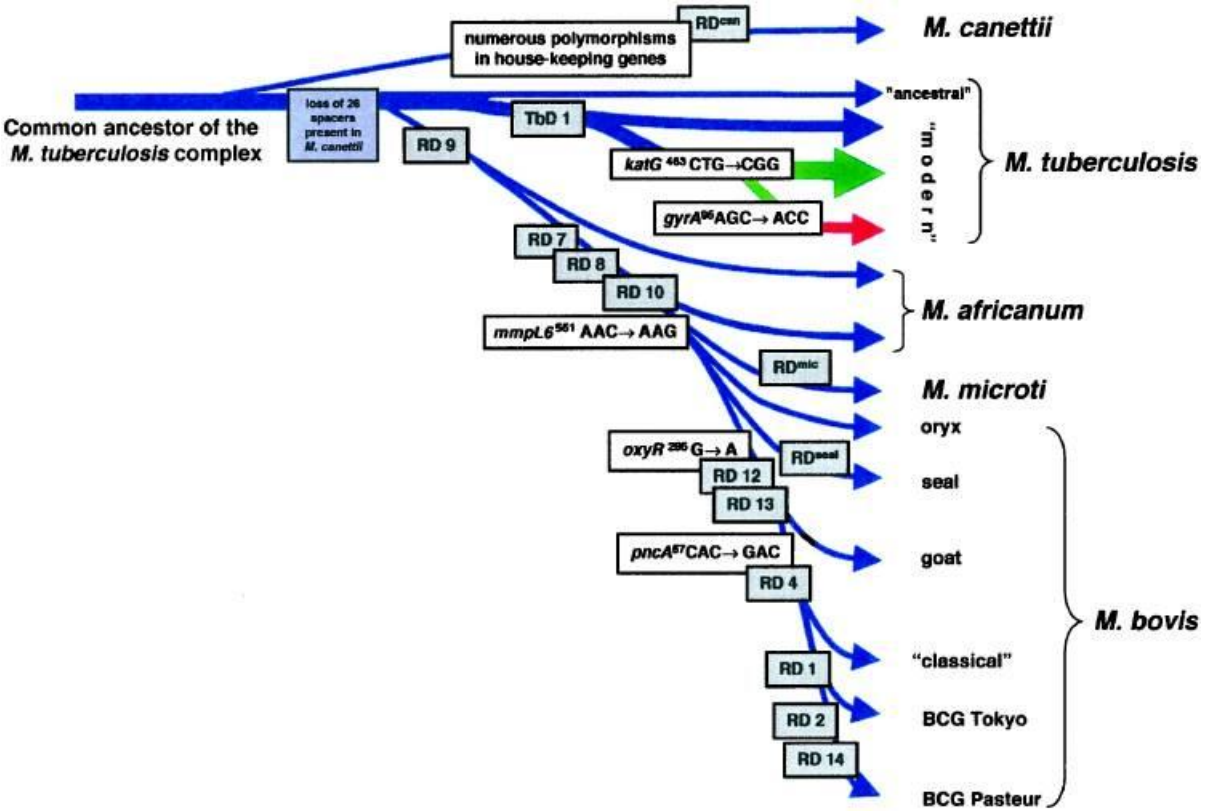


Figure 1.3.1 Shows the proposed evolution of *M. tuberculosis* complex. The image was retrieved from Brosch, R et al 2002; A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc National Acad Sci 2002;3684-9

1.4 Diagnosis

Diagnosis of active TB in patients has proven to be a rather difficult task, and is further complicated by the sheer amount of carriers of latent TB. Another major problem lies in the current methods available (described below). Many of the diagnostic techniques used today are severely outdated, unspecific, and fail to detect active TB in over half the patients.

1.4.1 Microbiological diagnosis

Microbiological tests are still the core of diagnosis, and only by successfully culturing *M. tuberculosis* on Lowenstein–Jensen medium, is one able to determine that the patient has an active TB infection. In addition to cultivation, the bacterium can be stained and observed via a microscope, however due its wax-like membrane, a special form of staining is required, referred to as acid-fast staining. Microscopic detection of mycobacteria does not distinguish *M. tuberculosis* from non-tuberculous mycobacteria. Worldwide, microcopy still represents the recommended procedure to diagnose TB. However, a specimen must contain at least 10 CFU (colony forming units) per millilitre sample to yield a positive smear ***(Allen BW, Mitchinson DA SO Med Lab Sci. 1992;49(2):94).*** Microscopy of specimens stained with a fluorchrome dye, such as auramine O provides a more sensitive, but also more expensive alternative.

Microscopy is rapid, cheap and easily replicable, but may not distinguish *M. tuberculosis* from non-TB mycobacteria (MOTT). The samples are usually taken from the sputum, but can also be from pus and cerebrospinal fluid. It should be noted however, that even though the wax-like membrane is unique, it is not limited to *M. tuberculosis*. Therefore it is also recommended to use PCR or gene probe test in order to distinguish other mycobacteria from *M. tuberculosis.* In order for a patient to test negative for TB, it is recommended that at least two separate cultivation testes show negative results. Although the method of choice, it is not to say it is without risk, if TB is handled in an unfit laboratory, which is not unheard of, there is always an ever-present danger of either inhaling or spreading it into the environment.

Figure 1.4.1 Shows an image of a clinical sputum sample. The patient is infected by acid fast bacteria, indicative of *M. tuberculosis* infection. Source: Wikimedia Commons

One of the important elements in clinical management of TB and infection control is knowledge of the drug susceptibility of the *M. tuberculosis* isolate. This requires cultivation of the bacteria. Artificial media used to cultivate *M. tuberculosis* include potato and egg base media such as Middlebrook 7H10 or 7H11, or albumin in agar base such as Lowenstein-Jensen (LJ) medium. Although detection of *M. tuberculosis* by culture may be more sensitive than microscopy, three to four weeks are required to recover the organism, and the method cannot be used to predict the risk of ineffectiveness a patient represents.

Once the organism is isolated however, it may also be identified as *M. tuberculosis* based upon morphologic and biochemical characteristics. Nucleic acid-based detection methods are also becoming increasingly used as supplements to these conventional tests.

Drug susceptibility testing is of great importance with emergence of resistant *M. tuberculosis* strains. Drug-resistant TB refers to *M. tuberculosis* that is resistant to one of the first-line anti-tuberculosis drugs; isoniazid, rifampin, pyrazinamide and ethambutol. Multidrug-resistant tuberculosis (MDR-TB) refers to *M. tuberculosis* that is resistant to at least one of the injectable second-line drugs (capreomycin, kanamycin and amikacin), as well as fluorquinolone.

1.4.2. Immunological diagnosis

Interferon-γ release assays (IGRA), are commercially available blood tests (QuantiFERON-TB Gold In Tube & T-SPOT.TB), which can be used as supplements in diagnosis of TB infections. These methods are based on the fact that certain *M. tuberculosis* antigens (ESAT-6, CFP-10 & TB7.7) are able to stimulate production of interferon- γ within the host, which are not be present in non-tuberculous mycobacteria. The methods are advanced, but relatively rapid as results can be retrieved within 24 hours, and will eliminate false-positives (BCG vaccine). They do however also have some limitations and drawbacks; In order for the blood samples to remain viable, they must be analyzed within a relatively short timeframe (8-30h), which often may be impossible in some parts of the world. There is also significant lack of data regarding usage on children less than 5 years of age, immune comprised patients (like those with HIV), and those who have had previous exposure to *M. tuberculosis*. A study conducted on these two methods showed that they often may give different results in a clinical everyday use, and that QuantiFERON-TB was more accurate when tested on those with confirmed active TB (***Giovanni Ferrara et al, 2006***). Both QuantiFERON-TB Gold In Tube & T-SPOT.TB rely essentially on blood samples, which are incubated with synthetic TB antigens, but function a bit differently. QuantiFERON-TB Gold In-Tube is used to measures the total amount of interferon-γ production, which is carried out by T-lymphocytes in response to TB antigens.

T-SPOT.TB on the other hand is used to measure the amount of activated T-lymphocytes, rather than interferon- γ production. These methods are relatively advanced, expensive and are unable to separate those with active from latent TB.

Other methods include a simple skin test (Mantoux skin test), in which the patient is injected with tuberculin under the forearm. If a reaction occurs around the site of injection within 72 hours, it may indicate that the person has a TB infection. However this method is also not able to distinguish between active and latent TB, and can even be triggered by other TB-related microbes.

1.4.3. Clinical diagnosis

Basic clinical diagnosis usually starts based on described symptoms, which include but are not limited to; coughing, fever, loss of appetite and night-sweat. However these symptoms are very nonspecific and vague *i.e.* they could even match the common cold, making it very difficult to diagnose as TB. Considering the symptoms, many do not seek medical attention straight away. However, during this period they are able to transmit the disease onto others during this period unknowingly. This creates a "diagnostic delay" situation, in which a lot of time is lost due to a patients not seeking treatment. It is imperative that the health-care system diagnose TB as fast as possible once the patient presents his or hers symptoms, so that the "diagnostic delay" is not prolonged further.

Supplemental examination such as radiography is usually required. In active pulmonary TB, infiltrates of cavities are often seen in the upper lings with or without lymphadenopathy or pleural effusions. However, lesions may appear anywhere in the lungs. In disseminated TB, a pattern of many tiny nodules throughout the lung fields is common (miliary TB). In immunosuppressed individuals, any abnormality may indicate TB. The chest X-ray may however also appear normal.

Thus abnormalities on chest radiographs may be suggestive of, but are not conclusive of TB. However, chest radiographs may be used to rule out the possibility of pulmonary TB in a person who has a positive reaction to tuberculin skin test and no symptoms of the disease.

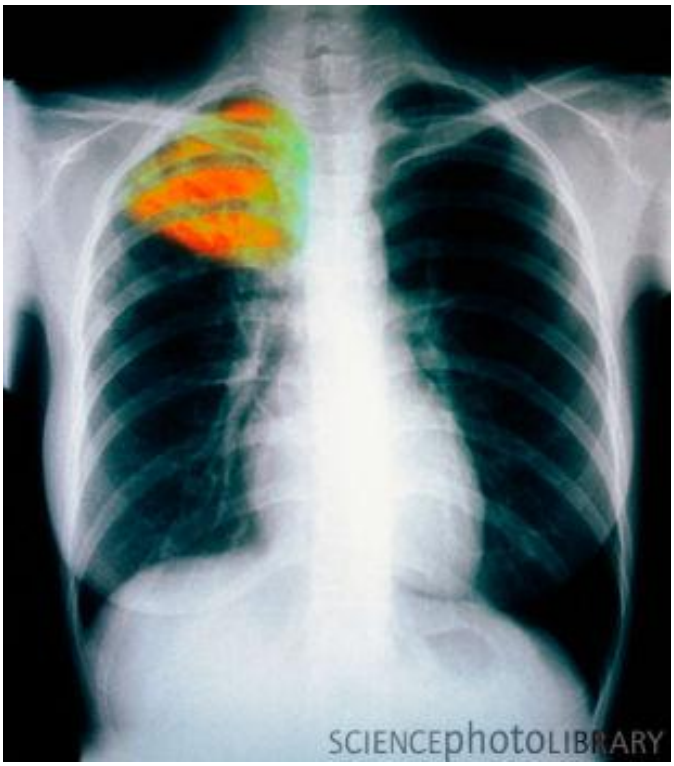


Figure 1.4.2 Shows an image of a chest X-ray. The patient is a 31 year-old female infected by *M. tuberculosis*. The red and green area in the lung is where the bacteria are developing. The black parts of the lungs are healthy. Source: Scott CAMAZINE/SCIENCE PHOTO LIBRARY.

1.5 Pathogensis

A key feature in the pathogenicity of *M. tuberculosis* is the ability to change its metabolism. By injecting alveolar macrophages with virulent strains of *M. tuberculosis*, it was demonstrated that the bacterium halted all biomass accumulation, and instead shifted focus on accumulating mycolic and fatty acids on its cell wall ***(Schnappinger, D et al 2003***). This helps the bacterium survive the harsh environment found within the phagsome, which is hypoxic, nitrosative, oxidative and nutrient poor.

1.5.1 Disease progression

TB spreads between people mainly through the air. When a person with infectious pulmonary TB coughs or sneezes, tiny droplets containing *M. tuberculosis* may be expelled into the air. Transmission may occur when another person inhales droplet nuclei, which contains tubercle bacilli. These enter the lungs, and deposit in alveoli. Here they may multiply and a small number may spread to other parts of the body via the blood-stream. Generally the immune system may keep these *M. tuberculosis* under control. The patient is considered to have a latent TB infection (LTB) at this stage. Following exposure to TB, the bacterium is either eliminated by an innate immune response (via dendritic cells), or through an adaptive response (via T-cells), usually within 4-6 weeks. If macrophages are unable to dispose of the infectious agents however, they can cluster around it, forming a prison-like structure referred to as tuberculoma. These structures can be visualized by X-ray in a patient, and as long the tuberculomas hold, they keep the bacterium isolated and prevent any interaction with its surroundings ***(Mukhopadhyay, S et al 2011)***.

Individuals who are exposed to TB may or may not develop a TB infection. Active TB develops when the immune system can no longer keep the bacilli under control, and the bacilli begin to multiply. The highest risk of developing active TB is in the first 2 years following infection. About 10 % of those with latent TB will develop an active infection during their lifetime, but the exact mechanism as to why this happens is still unknown. The remaining 90 % will stay latently infected, but otherwise free of the disease.

The MTC is able to grow in many different parts of the body, especially in areas with high oxygen levels, making the human lungs ideal.

Depending on the patient's origin of infection, the disease can develop either as pulmonary, meningitis, lymph nodes or any other extra-pulmonary areas. When TB does become active, it most commonly involves infection in the lungs (pulmonary TB). Symptoms include chest pain and a productive, prolonged cough. This occurs commonly in immunosuppressed individuals and young children.

Extrapulmonary infections commonly include sites such as the pleura (tuberculous pleusi), the central nervous system (meningitis), the lymphatic system, the genitourinary system and the bones, joints or the spine (Pott's disease).

In the interest of emphasizing the complicated clinical presentation of TB and the resulting diagnostic challenges, it should be pointed out that any organ or tissue in the body may be the main site infection. Recently a case of TB infection inside the eye was notified in Norway. This case represented a great diagnostic challenge as no tissue biopsy could be provided for microbiological sampling.

1.5.2 Treatment

It wasn't until 1943 that scientists revolutionized medicine by demonstrating that streptomycin could cure TB ***(Schatz A, Waksman SA et al 1945).*** During the same time period, another group of scientists (lead by Jørgen Lehmann and Karl Gustav) were able to synthesize para-aminosalcyclicacid, which was also very effective against TB. Over the next few years (1950-60) several other antibiotics began to emerge. At this point, there was great optimism, and even belief that TB could be completely eradicated. The joy however was somewhat short lived, as it only took a few years before the bacterium began to build resistance. It was possible to slow down the gradual resistance buildup by combining streptomycin and para-aminosalcyclicacid, however there was no way to stop it. During the 1990s, reports of multi-resistant TB began to emerge **(MDR-TB)**, and in 2005 strains with even a wider range of resistances were clinically isolated **(XDR-TB)**.

Treatment is difficult and requires a long course of multiple antibiotics. Social contacts are also screened and treated when necessary. There has been very little development in available treatment over the past 40 years. The treatment can be divided in 2 main categories, referred to as first and second line.

First line drug-treatment lasts about 6 months, and is divided into two stages. During the first stage, the patients are given a daily combination of antibiotics for two months; isoniazid, rifampicin, ethambutol and pyrazinamide. The second stage lasts the remaining four months, and the patient is given isoniazid and rifampicin three times a week. Streptomycin used to be included in the first line category, but is no longer considered viable due to a gradual resistance buildup by the microbes. After the first line treatment is done, and if the microbe has been successfully identified as a multi-resistant strain, second line treatment must commence.

The second line drugs are classified as such for one of three reasons; they are less effective, have severe side effects or are unavailable in many parts of the world. The drugs are administered in a specific order i.e. based on known sensitivities, starting with the most effective and least toxic, to least effective and most toxic; amikacin/kanamycin, pyrazinamide, ethambutol, moxifloxacin, rifabutin, cycloserine, prothionamide/ethionamide, aminosalicyclic acid, clarithromycin, linezolid, interferon-γ, thiordazine and ampicillin. There is also a Third line of drugs, but insufficient data is lacking to prove their viability.

1.6 Aim of the study

As presented above, TB is a serious disease that represents a great challenge to individual and public health throughout the world. This current study, in part, is connected a larger study attempting to develop a method for rapid diagnosis of TB. It is estimated that TB is one of the leading causes of death among known curable disease (***WHO, World Health report 2004***). Even though there are numerous measures available, including vaccination, antibiotics and chemotherapy, the rate of TB is a growing worldwide problem.

Cloning and expression studies have demonstrated that several proteins from the *H37RV* genome have potentials as diagnostic markers for active TB. Therefore, we consider them candidates for use in a future rapid serological test to diagnose TB. However, the genetic variations of these diagnostic candidates among the different strains of TB are not clear. By examining the amount of genetic variation of these sequences in different strains, we will get an indication of the diversity of this gene in an *M. tuberculosis* population. We may thus predict their ability/suitability to detect infection from other strains of *M. tuberculosis*, than the original *H37RV.*

The overall aim of the study was to confirm if given peptides that were identified as putative diagnostic markers for active TB, were present in strains of *M. tuberculosis* from various parts of the world.

The specific aims of the study were;

1. To identify candidate-isolates of *M. tuberculosis* in the national strain collection at the Norwegian Institute of Public Health.
2. To verify the strain characteristics of these strains, and to assign each isolate to *M. tuberculosis* lineages.
3. To examine the genetic variation of the sequences coding for peptides that may have a potential as diagnostic markers in these different strains of *M. tuberculosis*.

Section II: Materials & Methods

2.1 Spoligotyping

Spoligotyping is a method that utilizes PCR to both detect and type strains of the MTC (Mycobacterium tuberculosis complex) simultaneously. The basic principle revolves around polymorphism in the direct repeat (DR) locus of the mycobacterial chromosome (***Kamerbeek et al 1997***). The DR region consists of two or more repeats of a specific sequence, the exact function of the DR region in MTC is however still unknown (***van Emdebn et al, 2000***). The DR is highly conserved, indicating an evolutionary importance, and consists of 36 bp with unique spacer sequences, which vary from 35 to 41 bp in length. A combination of DR and spacer sequence is referred to as a *direct variable repeat* (DVR) (***Groenen et al, 1993***). It is believed that spoilgotypes are evolved by mutations of DVRs by means of recombination, replication, or point mutation (***Warren et al, 2002***).Clinical isolates can therefore be differentiated based on the presence or absence of one or more spacers.

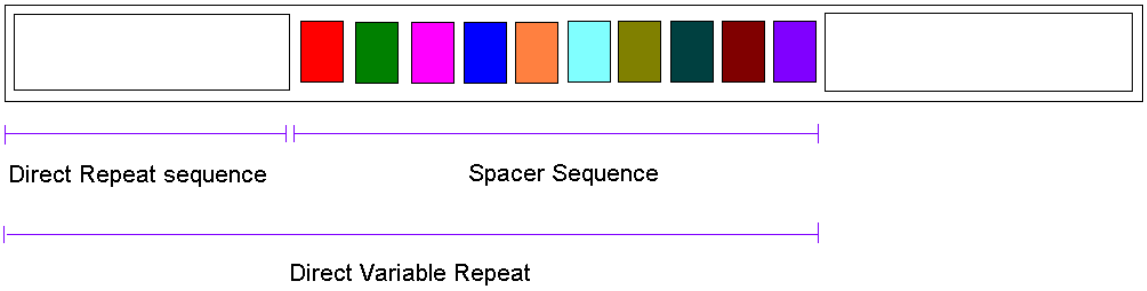


Figure 2.1.1 Shows an image of the DVR (Direct Variable Repeat), containing both DR and spacer sequences. This pattern is used to differentiate strains of the MTC. (Illustration: Nermin Zecic, Adobe Photoshop CS5)

All spoligotypes can be represented by a 43-dimensional binary code, (see section below for details), and converted to an octal code (***Dahle, UR et al, 2001***). These codes can then be used to browse and search online databases like SpolDB3 and SpolDB4.

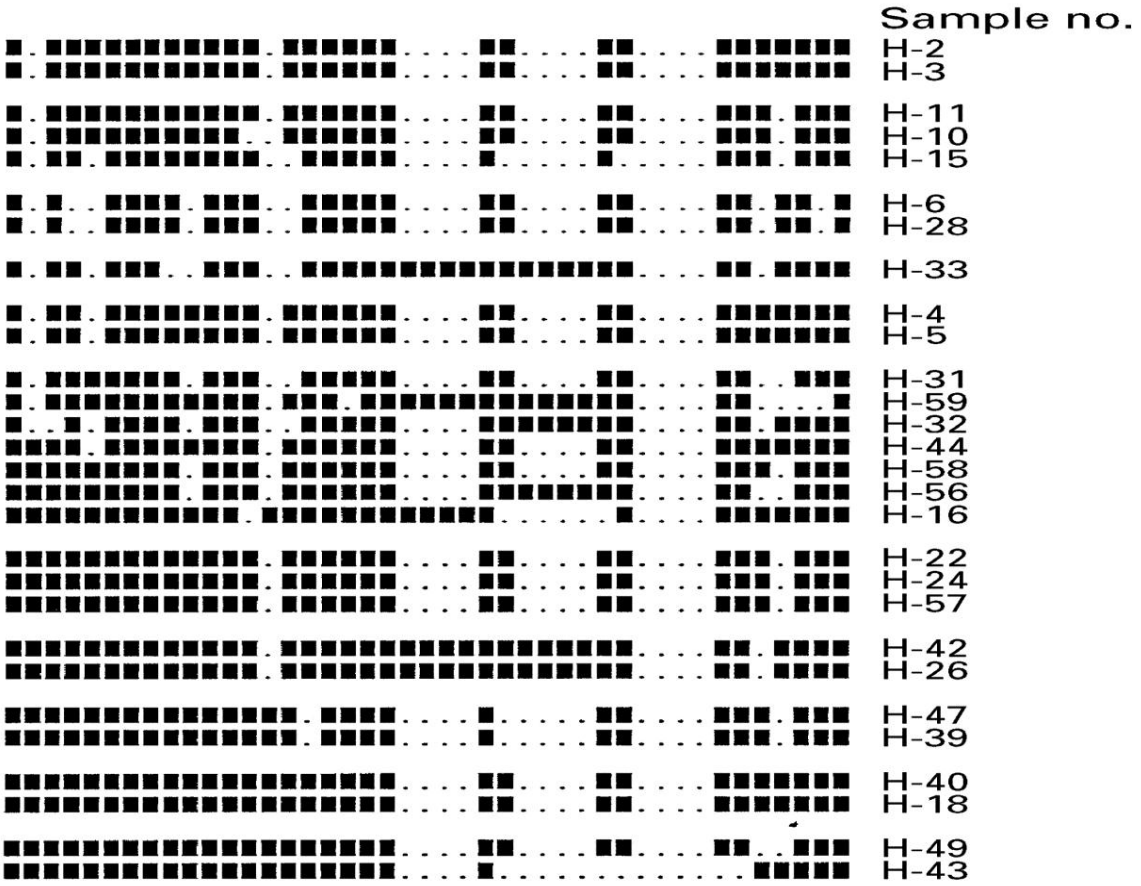


Figure 2.1.2 Shows an example of a spoligo-pattern. The barcode pattern is generated by DNA hybridisation to the BioDyne C membrane, and is visualised by chemiluminescence. ***(R S Heydermana et al, 1997)***

The membrane pattern (**Figure 2.1.2**) is read from left to right (the whole row), and dark patches are referred to as either "***G or 1***", while the clear areas are referred to as "***C or 0***". Following this principle, one is able to convert a series of patches and clear areas into a character/binary code, consisting of only ***C*** and ***G*** or ***1*** and ***0*** (**Table 2.1.1**).

Table 2.1.1 Shows how to read and interpret the spoligo pattern, the numbers/letters are read three at the time, where each variation of triplets refers to a unique number.

Triplet:	001/CCG	010/CGC	011/CGG	100/GCC	101/GCG	110/GGC	111/GGG
Numbers:	1	2	3	4	5	6	7

These numbers are then combined, and used to search in an online database; Spotclust. An example of a search result is shown below.

- Membrane pattern: 
- Letter code: GGG, CCC, CGG, GGG, GGG, GGG, GGG, GCC, CCC, CCC, CCG, GGG, GGG, G
- Combined triplet code: 703777740001771
- Result: CAS, 0.999999999786093 %

2.1.1 Spoligotyping method

The samples were chosen from the most exposed parts of the world, *i.e*. where TB is an ever present and ever growing problem. In this study the DNA sequences from the DR (Direct repeat region) was derived from *Mycobacterium tuberculosis* H37Rv and *M. bovis* BCG P3, and all the spacer-oligonucleotides were synthesized with a 5' terminal amino group. This allows the spacer-oligos to be covalently linked to a negatively charged Biodyne C membrane, and the membrane pattern visualised using chemiluminescence.

2.1.2 Membrane containing spacer oligonucleotides

The spacer oligo-nucleotides were diluted in 150 µl 500 mM NaOCO3 pH 8.4. The membrane was than activated by incubation in 10 ml of 16 % (w/v) 1-ethyl-3(3-dimethylaminorpopyl) carbodiimide (EDAC) in deminaralized water, which had to be freshly made.

2.1.3 In vitro Amplification of spacer DNA by the use of PCR

A reaction mixture had to be prepared prior to running PCR, the amount of each ingredient (see below) had to be calculated based on the amount of samples chosen, i.e. X amount of samples meant multiplying each of the values below by X. DNA from the strains *M. tuberculosis* H37Rv and *M. bovis* BCG were used as positive control units, while water was used as a negative control.

- 5 µl 10x PCR-buffer.
- 4 µl primer DRa.
- 4 µl primer DRb.
- 4 µl dNTP mix.
- 0,25 µl Taq polymerase.
- 30,75 µl H2O.

Each sample contained a total of 50 µl, which consisted of 48 µl of reaction mixture and 2 µl of DNA.

The PCR procedure can be performed in a number of ways, usually by the use of 10 ng purified chromosomal DNA, however for this particular experiment, frozen DNA from lysed mycobacterium was chosen. Amplification of the spacers was accomplished by the use of two specific primers, which correspond to the DR (direct repeat) sequences of the MTC (Mycobacterium tuberculosis complex), DR**a** (5'-GGT-TTT-GGG-TCT-GAC-GAC 3') and DR**b** (5'-CCG-AGA-GGG-GAC-GGA-AAC-3'). The DRa primer is biotinylated at 5' end, which enables the detection of the biotin-labelled PCR end-products.

Table 2.1.1 Shows the PCR program used during the amplification of the DR locus. The program was cycled 30 times.

Process	Temperature:	Time
Denaturing	96	3 min
Denaturing	96	1 min
Annealing	55	1 min
Extension	72	30 sec
Extension	72	5 min
Storage	4	∞

2.1.4 Hybridization with PCR-product and detection

Buffer solutions containing 2xSSPE (saline-sodium-phosphate-EDTA) and SDS (Sodium dodecyl sulphate) were prepared in the following concentrations:

- 250 ml 2xSSPE/0.1 % SDS, 60 °C.
- 500 ml 2xSSPE/0.5 % SDS, 60 °C.
- 500 ml 2xSSPE/0.5 % SDS, 42 °C.
- 500 ml 2xSSPE, room temperature.

The PCR products (see **2.1.3**) were diluted in eppendorf tubes, which contained a premade solution of 150 µl of 2xSSPE/0.1 %. The samples were then placed in a washtub at 99 °C for 10 min to heat-denature the DNA. The membrane (see Section III 3.1) was then placed in a rolling bottle, and incubated at 60 °C in what remained of the 2xSSPE/0.1 % SDS buffer. Following the heat treatment, the membrane had to be placed on a support cushion inside a mini-blotter system (see **Figure 2.1.3**), and aligned parallel to the line pattern. 160 µl of the diluted PCR products were then applied to the mini-blotter system, and placed in an incubator oven at 60 °C for one hour.

Figure 2.1.3 Shows the mini-blotter system used during spoligotyping. The instrument has 45 wells, and can handle up to 42 samples in addition to 3 controls units. When the BioDyne C membrane is placed within, the lid forces it in place and creates the narrow wells on its surface. Source: Webscientific.

It's important to keep the mini-blotter steady, and avoid shaking at all costs, as this may lead to contamination between the different wells. After the heat treatment was done, the membrane had to be removed from the mini-blotter, and washed twice in a rolling bottle using 250 ml of 2xSSPE/0.5 % SDS buffer, for 10 min at 60 °C. At this point, it was crucial to let the membrane cool off properly, in order prevent the inactivation of streptavidin-peroxidase, and was therefore stored overnight at 4 °C. The streptavidin-peroxidase (2.5 µl) was diluted in 10 ml of 2xSSPE/0.5 % buffer, and placed in a rolling bottle along with the membrane for 10 min at 42 °C. The membrane was then washed twice in 250 ml of 2xSSPE/0.5 % buffer for 10 min at 42 °C, and then again twice in 250 ml of SSPE for 5 min at room temperature.

Finally the membrane was incubated in 20 ml of ECL detection liquid for 1 minute, in order to visualise the spoligo-pattern by chemiluminescent detection.

2.2 Restriction fragment length polymorphism

2.2.1 Lysozyme treatment

Before starting it was important to properly thaw the frozen cultures, which was done in room temperature. The samples were then placed in a centrifuge and run at 13 000 rpm for 5 minutes. This process allows the bacterium to be separated from the solution by forming a pellet at the bottom. The remaining solution was then removed and discarded, which was done using a glass pipette. In order to lyse the cells, a solution containing 10 mg lysozyme/1 ml distilled H_2O was made, each pellet was then dissolved in 50 µl lysozyme solution, and heat-treated for 2 hours on a water bath at 37 °C.

2.2.2 DNA purification

To each sample, 5 µl *Proteinase K* was added (to degrade proteins), followed by the addition of 70 µl of 10 % SDS (sodium dodecyl sulfate) to lyse the cells. The samples were mixed gently using a vortex, and incubated on a water bath at 65 °C, for 10 min. 100 µl of 5M NaCl was added along with 100 µl CTAB/NaCl, and incubated again at 65 °C for 10 min. The samples were then added 750 µl of chloroformisoamylalcohol, and centrifuged at 13 000 rpm for 5 min.

The centrifugation process allows the DNA to be separated from the other substances by creating different layers in the eppendorf tubes. The chloroform fell to the bottom, the proteins in the middle (seen as a white foamy substance), and the DNA on the top.

The DNA was then transferred to new eppendorf tubes, added 450 µl isopropanol, and mixed well using a vortex before being frozen at -20 °C for the next step.

2.2.3 Concentration of DNA

Samples were taken out of the freezer, thawed at room temperature, and then centrifuged for 10 min at 13 000 rpm. The supernatant was then carefully removed, and 900 µl of cold 70 % ethanol was added. The previous step was then repeated again, i.e. centrifugation and the removal of the supernatant. The samples were then placed in an incubator oven at 37 °C for 30 min, in order to allow the ethanol to vaporize. 20 µl of TE-buffer was then added to make the DNA soluble again, and the samples placed at -20 °C over night.

2.2.4 Restriction enzyme cutting

The Pvu II enzyme is derived from *Eschericia coli* strains that carry the PvuII gene from *Proteus vulgaris*. The enzyme cuts DNA at the following sequences; CAG|CTG
GTC|GAC

From each sample, 8 µl of DNA was transferred to new eppendorf tubes, and kept on ice. The following reagents were then added to each sample:

- 2 µl of M-buffer
- 9 µl of distilled H2O
- 1 µl of Pvu II

As described in the previous steps, the samples had to be homogenized by gentle mixing, and then centrifuged at 13 000 rpm for a few seconds, before being placed in a water bath at 37 °C over night.

2.2.5 Separation of DNA fragments by the use of gel electrophoresis

In order to separate the DNA fragments, and to verify the presence and quality of the purified DNA, a gel had to be made containing 0.8 % agarose. 1.2 g of agarose was weighed in, and mixed with 150 ml 1x TBE-buffer. The solution was then placed in a microwave for 4 min, at 350 W, before applying 15 µl of SYBR-green. The gel sets as it cools, and it was therefore important to pour it onto a container while it was warm. When loading the samples, the first and last well was added a DNA-ladder, and the samples were loaded between said ladders.

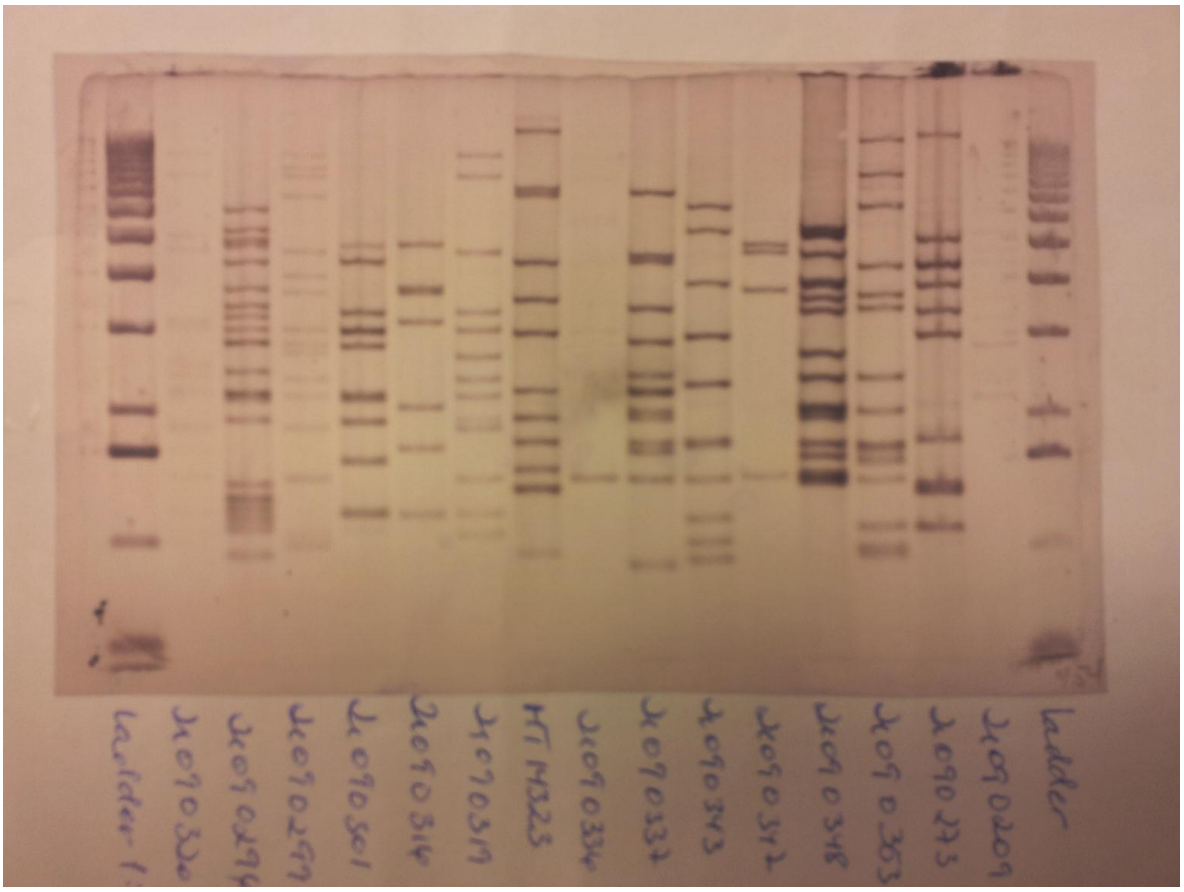


Figure 2.2.1 Shows an image of results obtained by RFLP. The first and last samples contain DNA ladders, while the 15 samples in between are isolates of *M. tuberculosis*. (Photo: Nermin Zecic)

2.3 DNA purification of selected samples

Following spoligotyping results (see **Section III, Table 3.1-3.3**), additional DNA was required in order to proceed with the next step. The 8 chosen samples (**Table 3.4**) were cultured, heat-killed, and their DNA purified using the method previously described in RFLP. The purification procedure was followed until and including the adding of *Proteinase K*. From this point on another method was applied using a DNA extraction and purification kit, Genomic DNA Clean & Concentrator™. Before proceeding with this kit, 24 ml of 100 % ethanol had to be added to the provided wash buffer. A solution containing 5 ml of ChIP DNA Binding-Buffer was added to each DNA sample, and mixed thoroughly. The DNA mixture was then transferred to *Zymo-Spin™ IC-XL Column in collection tubes*, centrifuged for 30 seconds at 13 000 rpm, and the flow-through discarded. To each column 200 µl of wash buffer was added, and then centrifuged for one minute, this step was repeated twice. Finally 10 µl of distilled H_2O was added to the column matrix and placed in a new microcentrifuge tube. The samples were than centrifuged for 30 seconds, and the eluted DNA collected for subsequent use in further downstream PCR.

2.4 Amplification of genes coding for H37Rv proteins

2.4.1 Amplifying genes coding for the diagnostic candidates from genomic DNA of the different strains.

Prior to running gel electrophoresis, each of the target sequences were first amplified using PCR.

Table 2.4.1 Shows the concentration and ingredients used in the PCR reaction mixture, the concentrations shown are per sample.

Ingredient:	Amount
DNA template	3 µl
of distilled H_2O	73 µl
Primer (X)A + Primer (X)B	0,5 µl +0,5 µl
dNTP	2 µl
5x Herculase Buffer	20 µl
Herculase enzyme	2 µl
Total	100 µl

In case of observing low yields of DNA product, the reactions were adjusted with respect to:

1. Increasing the DNA concentration, which were be correlated with a decreased amount of added H_2O.
2. Wait for the cooling block on the PCR machine to reach 4 ºC before adding the final ingredient; Herculase enzyme.

The samples were than run on a PCR program with the following settings.

Table 2.4.2 Shows the PCR program used during PCR.

Temperature	Time	Cycles
93 ºC	3 min	
96 ºC	1 min	x30
55 ºC	1 min	
72 ºC	30 sec	
72 ºC	5 min	
4 ºC	∞	

After the PCR was done, 10 µl of each sample was applied to a gel (1%) to verify that a PCR product was present. The samples were prepared as listed below.

1. 10 ml of Triss-acetat EDTA was diluted in 500 ml of distilled H_2O.
2. From the diluted solution, 100 ml was transferred to a new container, and added 1 g of agarose (powder form). The container was then put in a microwave at 550W for 1 min and 30 seconds (or until the solution becomes completely clear). The container had then to be cooled on ice or running water, before adding 2.5 µl Ethidium Bromide into the mix.
3. The agarose solution was left in room temperature for a few minutes, before being poured onto the plastic plates, which ultimately formed the gel. The sample wells were made using the provided plastic comb.
4. The gel was transferred into the electrophoresis container, and to each chamber side 2.5 µl Ethidium Bromide was added. The remaining solution made in the first step (400 ml of Triss-acetat EDTA + H_2O) was used to cover the gel.
5. To each PCR product, 2 µl of DNA loading buffer was added, and the samples applied to their respective wells.
6. Electrophoresis was run for 20 minutes at 1.8V.
7. The gel was visualised using UV-light in a dark-room.

2.5 Purification of the PCR product

Based on the individual results obtained in (**Section II, Part 2.3**), the bands which contained DNA of expected sizes were excised from the agarose gel, and dissolved in *buffer QG* (Qiagen) for approximately 15 minutes at room temperature. After the agar had been completely dissolved, the DNA containing QG buffer was allowed to pass through two *Qiagn Gel Purification columns*. These columns were then washed with 800 µl *Qiagen QC washing buffer* and centrifuged. The purified DNA was recovered in 20 µl *Qiagen elution buffer EB*. The two eluted products were combined, and the residual salts and ethanol were removed by using *S200 Microspin Column* (VWR). The DNA was stored in -20 ºC until the next step.

2.6 DNA sequencing of genes coding for potential diagnostic markers

The method used in this study is well suited for sequence-analysis of DNA-fragments amplified from bacterial DNA. By sequencing a certain area of the 16s rRNA gene, one is able to identify the bacterium by comparison in online databases, such as Basic Local Alignment Search Tool (BLAST). The method works with both single-, as well as double-stranded DNA (ds DNA has to be denatured prior). For best results, both strands should be sequenced.

When working with ss DNA, a primer has to be added to the samples, which hybridizes complementary to the DNA template. The primer provides a 3'–end which is essential for the DNA- polymerase. Without the 3' –end the polymerase is unable to add deoxyribonucleotides (dNTP) during the elongation process. In addition to the primer, the samples are also added four color-coded 2'-3'-dideoxyribonucleotidephosphate (ddNTP[A,T,C,G]), which label the DNA and terminate the elongation process as they lack a 3' hydroxyl end. The fragments are then separated by capillary gel-electrophoresis, and detected by fluorescence. The sequence is read as the fragments pass a CCD-chip, which functions as a detector, and is able to distinguish the emitted base-specific wavelengths. These light-signals are then turned into electric signals, which are read by a computer. The detection can also be performed by the use of autoradiography, which relies on incorporated radioactive labeled substance, such as ^{32}P-dATP.

2.6.1 Cyclic sequencing reaction

The sequencing reactions were performed on **ABI 3730 DNA Analyser** using microtiter plates, which contain 96 wells. Each well had to contain the following ingredients:

Table 2.6.1 Shows the reagents used to make the reaction mixtures, each well contained 20 µl in the microtiter plate.

Reagent	Amount in µl
Sterile dH2O	6.5 µl
5x Sequencing Buffer	4.0 µl
Big-Dye Terminator mix	1.5 µl
1.6 µM primer	3.0 µl
DNA template	5.0 µl
Total reaction volume	20.0 µl

After adding all of the listed reagents, a rubber lid was placed on the microtiter plate, and placed in micro-shaker running at 400 rpm for 2 minutes. The plate was then loaded into a PCR machine with the program listed below:

Table 2.6.2 Shows the different temperatures and durations, which were preprogrammed into the PCR machine.

Temp	Time	Cycles
94 ºC	5.0 min	
96 ºC	10 sec	40 cycles
50 ºC	5.0 sec	
60 ºC	4.0 min	
4.0 ºC	∞	

2.6.3 Purification of the sequencing reaction product

To each well in the microtiter plate, 10 µl of distilled H_2O was added before being left in room temperature until a mew mixture was prepared. The new solution was prepared by combining 280 µl of 3M NaAc pH 5.2 with 7 ml of EtOH. 100 µl of the newly made solution was added to each sample-containing well. Finally the plate was sealed with a plastic film, before being left in room temperature for 45 min. The microtiter plate was now ready for centrifugation, which was run on **Program 1**, as described below.

Table 2.6.3 Shows the different programs which were used in the PCR-product purification.

Program nr	Temperature	G	Duration
1	4 ºC	4000	30 min
2	4 ºC	1000	30 sec
3	4 ºC	5	5 min
4	4 ºC	30	30 sec

Following centrifugation on **Program 1**, the plastic film was removed from the plate so that the ethanol could be shaken out on a piece of absorbent paper, and then centrifuged on **Program 2**. Note that if residual ethanol is still present, a rerun of **Program 2** is necessary. In the next step, 150 µl of cold 70 % EtOH was applied to each of the wells with PCR product, before being centrifuged on **Program 3**. The previous step above was then repeated *i.e.* ethanol was shaken out followed by a rerun of **Program 2**. The final step was performed by placing the plate upside-down on a Whatman 3MM Chromatography paper, and was than centrifuged on **Program 4**. The plate was stored in -20 ºC until capillary separation could be performed.

2.6.4 Preparing the plate for DNA sequencing

To each well, 10 µl of Hi-DI™ Formamide was added, and the plate sealed with a SEPTA matt. The plate was then left in room temperature for 10 minutes, before being mixed gently on a Whirl-Mixer. The final step involved centrifugation at 4000G for 30 seconds, after which the plate was placed inside the ABI 3730 DNA Analyzer machine.

The raw results from the sequence reactions were obtained after 2 hours, and the DNA sequences were read and edited using BioEdit.

Section III: Results

A total of 3500 *M. tuberculosis* strains, present in the strain collection at the Norwegian Institute of Public Health were evaluated in order to identify representative candidates of different *M. tuberculosis* lineages. The 42 isolates were further characterized and assigned to *M. tuberculosis* lineages (**Figure. 3.1., Tabs 3.1, 3.2**)

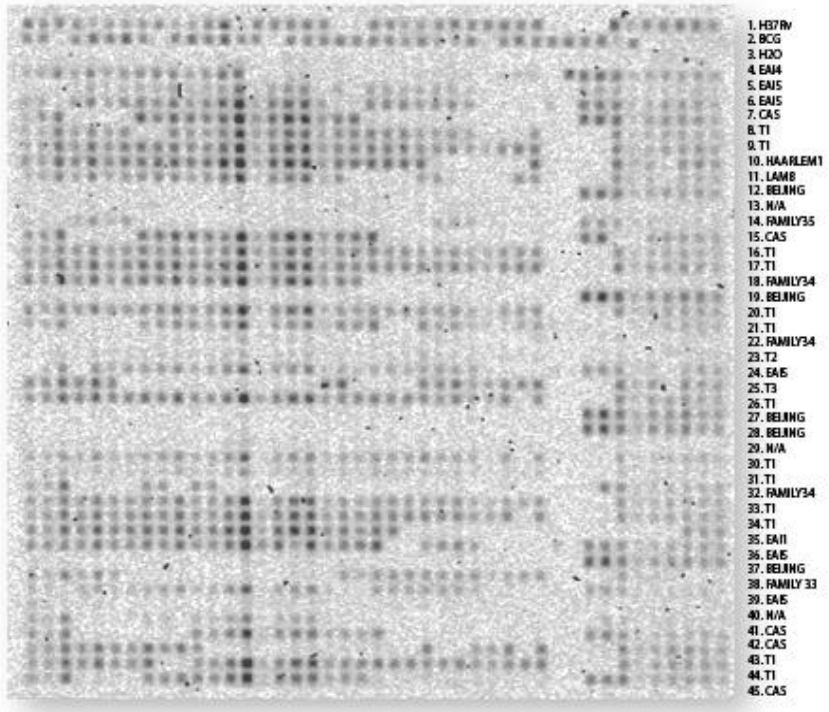


Figure 3.1 Shows the obtained spoligo-patterns on the *BioDyne C* membrane. The translated results are displayed in **Table 3.3**. The first 2 samples contain positive control units (*M. tuberculosis* H37Rv and *M. bovis* BCG P3), while the third is used as negative control (H_2O). Each pattern represents a different strain of TB, and is read one line at the time from left to right. (Photo: Nermin Zecic)

Table 3.1 Shows the spoligo-pattern in both binary and character codes, as seen on the BioDyne C membrane. The dark patches represent "**g or 1"**, while the clear areas represent "**c or 0"**. The binary/character codes had to further be translated into octal codes, prior to being used to brows SpolDB4.

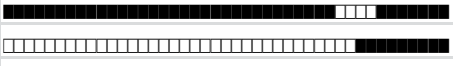
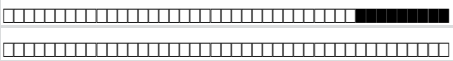
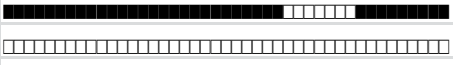
21100387		ggg	ggg	ggg	ggg	ggg	ggg	ggc	ccc	ccc	ccc	ccc	cgg	ggg	ggg	g
21100302		ggg	ggg	ggg	ggg	ggg	ggg	ggc	ggg	ggg	gcc	ccc	ccg	ggg	ggg	g
21000446		ggg	ggg	ggg	ggg	ggg	ggg	ggc	ggg	ggg	gcc	ccc	ccg	ggg	ggg	g
21090592		ggg	ccc	cgg	ggg	ggg	ggg	ggg	ccc	ccc	ccc	ccc	cgg	ggg	ggg	g
21100059		ggg	ggg	gcc	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggc	ccc	ggg	ggg	g
21090029		ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggc	ccc	ggg	ggg	g
21070505		ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	gcc	ccc	cgc	ccc	ggg	ggg	g
21070181		ggg	ggg	ggg	ggg	ggg	ggg	ggc	ccc	ggc	ccc	ggc	ccc	ggg	ggg	g
21070239		ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	cgg	ggg	ggg	g
21100210		ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	c
21070545		ccc	ggg	gcc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	cgg	ggg	ggg	g
21070456		ggg	ccc	cgg	ggg	ggg	ggg	ggg	gcc	ccc	ccc	ccc	cgg	ggg	ggg	g
21090431		ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggc	ccc	ggg	ggg	g
21090466		ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggc	ccc	ggg	ggg	g
21090219		ggg	ggg	ggg	ggg	ggg	ggg	ggg	ccc	ccc	ccc	ccc	ccc	ccc	ccc	c
21090605		ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	cgg	ggg	ggg	g
21100428		ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggc	ccc	ggg	ggg	g
10201182		ggg	ccc	cgg	ggg	ggg	ggg	ggg	gcc	ggg	ggg	ggc	ccc	ccc	ggg	g
21060115		ggg	ggg	ggg	ggg	ggg	ggg	ggc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	c
21060307		ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	gcc	ccc	ccc	ccc	ccc	c
21060150		ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	gcc	ccc	cgg	gcc	ggg	g
21060189		ggg	ggg	ccc	ccc	ccc	ccc	ggc	ccc	ggg	ggg	ggc	ccc	ggg	cgg	g
10107078		ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggc	ccc	ggg	ggg	g
21100266		ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	cgg	ggg	ggg	g
10109935		ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	cgg	ggg	ggg	g
F11/99		ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	c
10201215		ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggc	ccc	ggg	ggg	g
10201330		ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggc	ccc	ggg	ggg	g
10007774		ggg	ccc	cgg	ggg	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccg	ggg	ggg	g
21090295		ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggc	ccc	ggg	ggg	g
21070417		ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggc	ccc	ggg	ggg	g
21070547		ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggc	ccc	ccc	ccc	ccc	ccc	cgg	g
21080127		ggg	ggg	ggg	ggg	ggg	ggg	ggg	gcc	ggg	gcc	ccc	cgg	ggg	ggg	g
21080002		ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	cgg	ggg	ggg	g
21070272		ggg	ggg	ccc	ccc	cgg	ggg	ggg	ggg	ggg	ggg	ccc	cgg	ggg	ggg	g
21080645		ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ccc	ccc	cgg	ggg	ggg	g
21090468		ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	ccc	c
10112920		ggg	ccc	ccc	cgg	ggg	ggg	ggg	gcc	ccc	ccc	ccc	cgg	ggg	ggg	g
21090276		ggg	ccc	ccc	cgg	ggg	ggg	ggg	gcc	ccc	ccc	ccc	cgg	ggg	ggg	g
10108375		ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggc	ccc	ggg	ggg	g
21090415		ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggc	ccc	ggg	ggg	g
21090481		ggg	ccc	cgg	ggg	ggg	ggg	ggc	ccc	ccc	ccc	ccc	cgg	ggg	ggg	g

Table 3.2 Displays the results obtained from **Table 3.1**, the octal numbers were used to browse the online database known as SpolDB4 for strain identification. Refer to Section II: 2.1 Spoligotyping, for conversion details.

Nr	Octal code	Family	Sample	Match in %
1	777777600003771’	EAI4	21100387	0.993487267513416
2	777777677401771’	EAI5	21100302	0.95255037315678
3	777777677401771’	EAI5	21000446	0.95255037315678
4	703777700003771’	CAS	21090592	0.999999999786104
5	774777777760771’	T1	21100059	0.999991107775175
6	777777777760771’	T1	21090029	0.999991146319657
7	777777774020771’	Haarlem1	21070505	0.999963762792972
8	777777606060771’	LAM8	21070181	0.93651817282834
9	000000000003771’	Beijing1	21070239	0.999999999982487
10	000000000000000’	N/A	21100210	N/A
11	074000000003771’	Family35	21070545	0.999980028501071
12	703777740003771’	CAS	21070456	0.999999999514475
13	777777777760771’	T1	21090431	0.999991146319657
14	777777777760771’	T1	21090466	0.999991146319657
15	777777700000000’	Family34	21090219	0.998824987666382
16	000000000003771’	Beijing1	21090605	0.999999999982487
17	777777777760771’	T1	21100428	0.999991146319657
18	703777747760071’	T1	10201182	0.999990261890011
19	777777600000000’	Family34	21060115	0.972016516493734
20	777777777400000’	T2	21060307	0.938490717863386
21	777777777403471’	EAI5	21060150	0.959979708030938
22	770000607760731’	T3	21060189	0.999986998352701
23	777777777760771’	T1	10107078	0.999991146319657
24	000000000003771’	Beijing1	21100266	0.999999999982487
25	000000000003771’	Beijing1	10109935	0.999999999982487
26	000000000000000’	N/A	F11/99	N/A
27	777777777760771’	T1	10201215	0.999991146319657
28	777777777760771’	T1	10201330	0.999991146319657
29	703700000001771’	Family35	10007774	0.999769686964576
30	777777777760771’	T1	21090295	0.999991146319657
31	777777777760771’	T1	21070417	0.999991146319657
32	777777760000031’	EAI1	2107547	0.981920224737622
33	777777747403771’	EAI5	21080127	0.952969669377335
34	000000000003771’	Beijing1	21080002	0.999999999982487
35	770037777703771’	Family33	21070272	0.999954150905156
36	777777777003771’	EAI5	21080645	0.944264550157429
37	000000000000000’	N/A	21090468	N/A
38	700377740003771’	CAS	10112920	0.999999999821779
39	700377740003771’	CAS	21090276	0.999999999821779
40	777777777760771’	T1	10108375	0.999991146319657
41	777777777760771’	T1	21090415	0.999991146319657
42	703777600003771	CAS	21090481	0.999999999891887

Eight strains, representative of the *M. tuberculosis* epidemics In various parts of the world were selected for sequence analyses (tab 3.3). These included to identical strains assigned to the Beijing 1 lineage, abundant in the former Sovjet Union. These two strains served as an internal control to verify the reproducibility of the analyses performed. The selection of these strains was based on the homology the strains exhibited to the Spol DB4.

Table 3.3 Displays the 8 strains, which were found to originate from various high-burden countries. Strain **nr.7** was replaced at a later stage due laboratory challenges and the scope of the thesis.

No	Sample	Strain	Match in %
1	10109935	Beijing1	0.999999999982487
2	21070181	Beijing1	0.936518172828345
3	21060189	T3	0.999986998352701
4	21100059	T1	0.999991107775175
5	21090481	CAS	0.999999999891887
6	21080002	Beijing1	0.999999999982487
7	21100373	EAI1	N/A
8	21070505	LAM8	0.999963762792972

Primers were designed in order to amplify the sequences that code for the selected proteins, presented in Table 3.4. The sequences were amplified in all 8 strains and verified by electrophoresis (Figs 3.2, 3.3., 3.4.). We were able to confirm the presence of 4 of these proteins in all strains (Tab 3.5). The Primer-set 4, could not produce a product probably due to the large size of the sequence. After intense troubleshooting the primer-set was abandoned. It was not considered within the scope of this thesis to scrutinize the PCR methodology further.

Table 3.4 Shows the 5 proteins which served as basis for our research. All are derived from different *M. tuberculosis* H37Rv strains. Their function was determined by using BLAST. The genes coding for these proteins served as reference-sequences for the synthesis of the primers, which were used during PCR in the following steps.

Nr	Strain	Protein function
1	Rv3881	Conserved hypothetical alanine and glycine rich protein
2	Rv1884	Probable resuscitation-promoting factor RPFC
3	Rv0934	Active transport of inorganic phosphate across the membrane.
4	Rv1759	Unknown, has fibronectin-binding activity. Bacterial attachment to host cell
5	Rv3874	Unknown.

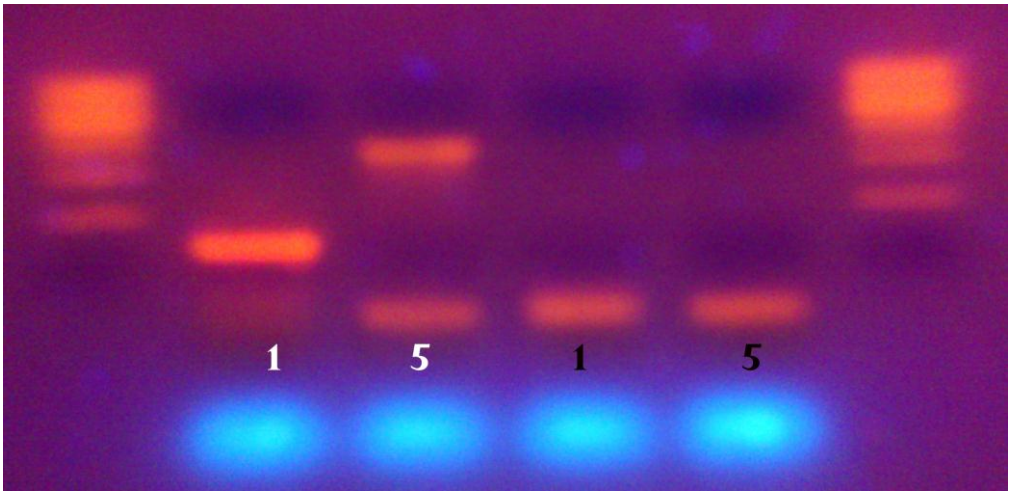


Figure 3.2 Shows the amplified PCR product using Primer 1 and 5 for sample 1 and 2, which were run separately at an early stage to confirm primer compatibility. The missing results from sample 2 were obtained at a later stage, and are displayed in **Figure 3.4**.

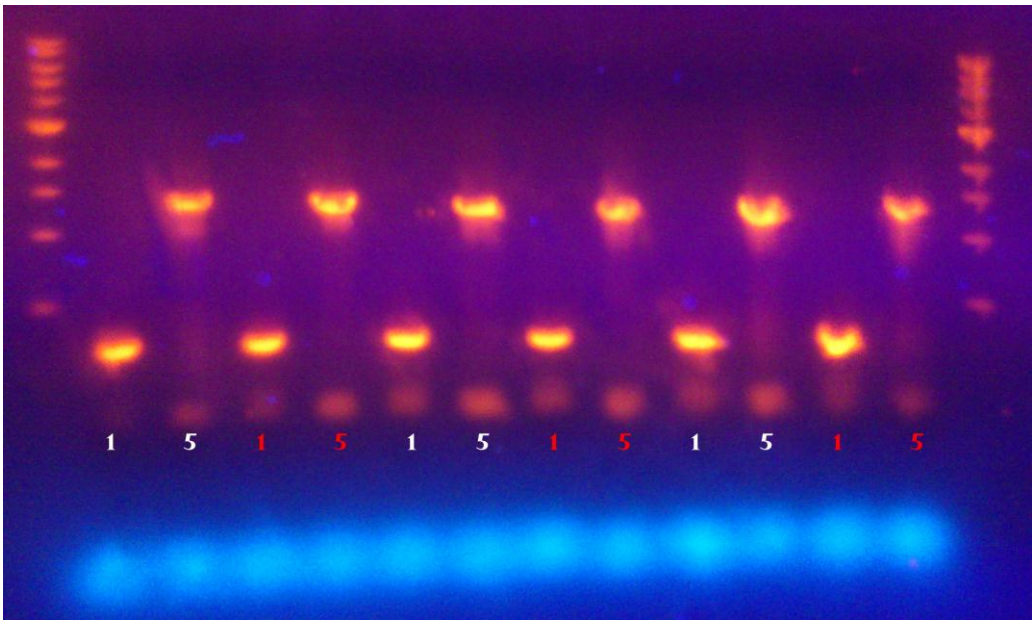


Figure 3.3 Shows the amplified PCR product for Primer 1 and 5 for the remaining samples 2-8. The gel was visualized after being run on gel electrophoresis for 20 min.

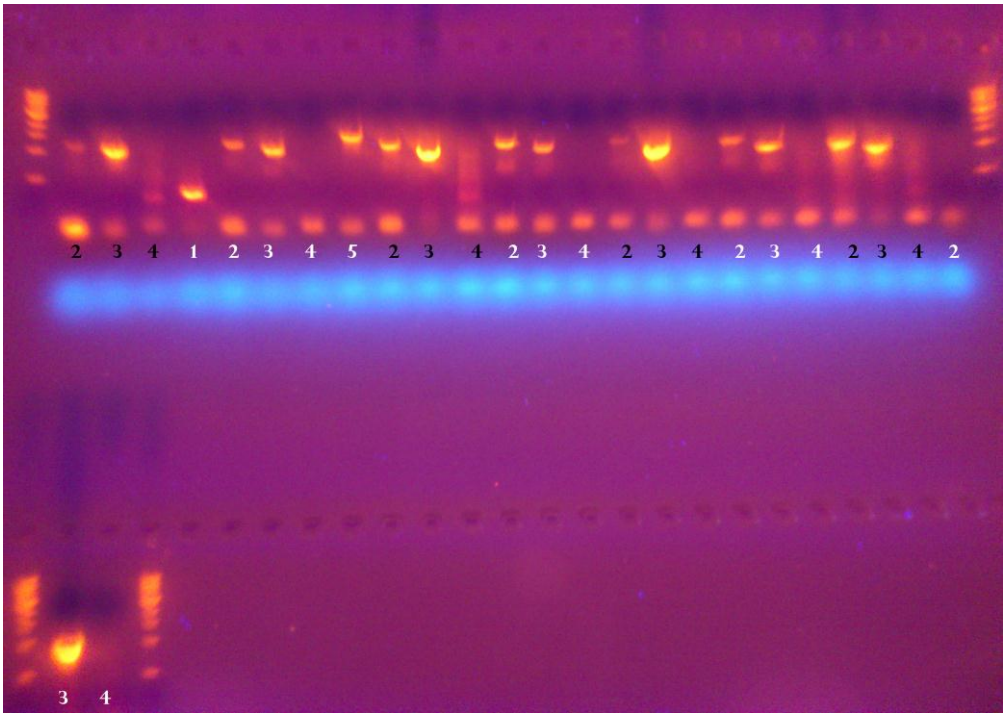


Figure 3.4 Shows all 8 samples and their respective primers, the sample-numbers are color-coded to reflect where one sample ends and another begins. Primer 4 was the only one to yield no results throughout the testing, along with sample 8 primer 2. Sample order is as shown in **Table 3.5.**

Table 3.5 Shows the combined results for all the TB samples and their respective primers. The areas marked by **X** are those that yielded no results. It was at this point that we decided to abandon primer 4.

Sample	10109935	21070181	21060189	21100059	21090481	21080002	21100373	21070505
Primer 1	√	√	√	√	√	√	√	√
Primer 2	√	√	√	√	√	√	√	X
Primer 3	√	√	√	√	√	√	√	√
Primer 4	X	X	X	X	X	X	X	X
Primer 5	√	√	√	√	√	√	√	√

The PCR products were sequences to verify their authenticity and to compare their homology to the *M. tuberculosis* H37Rv genome. The 8 isolates all carried sequences that were highly homologous to H37Rv (tabs 3.6 and 3.7). The homology to the Rv3881 varied from 93-99 %, Rv1886 varied between 82-99 %, Rv0934 varied between 81-94 % and Rv3874 varied from 96-97 %. Rv1759 is not included.

Table 3.6 Shows the sequences amplified by PCR. The product from each strain was compared to complete H37Rv genome. The comparison was done using an online database; Basic Local Alignment Search Tool.

Sample	10109935	21070181	21060189	21100059	21090481	21080002	21100373	21070505
Strain	Beijing1	Beijing1	T3	T1	CAS	Beijing	EAI1	LAM
Rv3881	Identities = 254/272 (93%), Gaps = 6/272 (2%	Identities = 273/277 (99%), Gaps = 3/277 (1%	Identities = 273/280 (98%), Gaps = 4/280 (1%)	Identities = 273/276 (99%), Gaps = 1/276 (0%)	Identities = 275/280 (98%), Gaps = 3/280 (1%)	Identities = 271/277 (98%), Gaps = 3/277 (1%)	Identities = 271/274 (99%), Gaps = 0/274 (0%)	Identities = 263/267 (99%), Gaps = 2/267 (1%)
Rv1886	Identities = 491/514 (96%), Gaps = 5/514 (1%)	Identities = 507/516 (98%), Gaps = 1/516 (0%)	Identities = 855/882 (97%), Gaps = 16/882 (2%)	Identities = 742/779 (95%), Gaps = 13/779 (2%)	Identities = 224/260 (86%), Gaps = 11/260 (4%)	Identities = 240/243 (99%), Gaps = 1/243 (0%)	Identities = 1031/1082 (95%), Gaps = 21/1082 (2%	Identities = 145/177 (82%), Gaps = 6/177 (3%)
Rv0934	Identities = 640/745 (86%), Gaps = 34/745 (5%)	Identities = 321/342 (94%), Gaps = 5/342 (1%)	Identities = 494/531 (93%), Gaps = 5/531 (1%)	Identities = 311/338 (92%), Gaps = 4/338 (1%)	Identities = 713/795 (90%), Gaps = 18/795 (2%)	Identities = 379/420 (90%), Gaps = 8/420 (2%)	Identities = 601/741 (81%), Gaps = 53/741 (7%)	Identities = 284/346 (82%), Gaps = 15/346 (4%)
Rv1759	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Rv3874	Identities = 1177/1213 (97%), Gaps = 26/1213 (2%)	Identities = 1166/1212 (96%), Gaps = 28/1212 (2%)	Identities = 1186/1237 (96%), Gaps = 34/1237 (3%)	Identities = 1153/1200 (96%), Gaps = 27/1200 (2%	Identities = 1169/1211 (97%), Gaps = 26/1211 (2%)	Identities = 1127/1168 (96%), Gaps = 27/1168 (2%)	Identities = 1161/1196 (97%), Gaps = 23/1196 (2%)	Identities = 1093/1131 (97%), Gaps = 25/1131 (2%)

Section IV: Discussion

The 5 TB-proteins **(Table 3.4**) were identified in *M. tuberculosis* H37Rv. They should therefore (in theory) be applicable in the detection of active TB infections by this particular strain. However, earlier studies have shown that these proteins can also be used to detect other strains, but exactly how many is still unknown. The protein sequences were derived from H37Rv3881c, Rv0934, Rv1886c, Rv1759c and Rv3874c. If the different sequences and their proteins are the same in all the strains (that we have analyzed during our own research), it could potentially be applicable in the development of a rapid TB test. If the genes should vary from strain to strain however, it may be limited to the detection of H37Rv.

The TB-strains (**Table 3.3**) were obtained from *The National Reference Laboratory for Mycobacteria* at the NIPH. The laboratory has since 1994 genotyped all known strains of *M. tuberculosis* isolated in Norway. The data has been stored in a local database, which consists of more than 3500 strains and 2000 spoligo-patterns. After analysis of this collection, a total of 42 different TB strains were initially chosen, since they were found to represent the heterogeneity of this strain collection (**Table 3.1**). The characteristics of these strains were confirmed by the use of two internationally standardized methods, described in *Section II*; Spoligotyping and IS*6110*-restriction fragment length polymorphism (RFLP). These methods are often used in combination, when characterizing *M. tuberculosis* strains. Spoligotyping is rapid, and can be applied directly on clinical samples by PCR. This method is well suited for assigning TB into defined phylogenetic groups (***Brudey, K et al 2006***) and for strain characterization of isolates carrying a low number (<5) of IS*6110* copies. RFLP on the other hand, is more reliable when used on isolates with multiple copies of IS*6110*, as is the case with most strains assigned to the Beijing lineage, indicating that it also has higher discriminatory capabilities than spoligotyping ***(Goyal, M et al 1997)***. For this discriminating ability RFLP has been the choice for contact tracing and deciphering outbreaks (***Dahle, U et al***).

The samples were chosen to represent those abundant in both high burden, as well as low burden countries, with focus on the former. The *Beijing* family is largely abundant in Asia and Russia, CAS in East Africa, EAI stretching from India to China as well as throughout parts of Europe. The T-family can be found for the most part in the USA, and includes more than 600 unclassified lineages. LAM is common in South- and Western Africa, most of Europe, as well as parts of North and South America (***Brudey, K et al 2006***).

From the 42 strains that were characterized in detail and assigned to given families, we selected 8. This sub-selection was based on the geographic origin found by analyzing the clinical data, and genetic diversity found by use of the molecular methods. The *M. tuberculosis* strains were studied for the presence of 5 distinct peptides. These peptides were selected based on previous studies (not considered part of this thesis). They were found in high concentrations in the sera of patients with active TB, and in lower concentrations in healthy people or people diagnosed with latent TB infection.

The primers used to amplify the sequences that corresponded to these peptides were derived from fully sequenced strains of *M. tuberculosis* H37Rv. The primers were used in PCR-amplification of the desired DNA sequences. This was done for all isolates and with all 5 sets of primers. The quality of the amplified DNA-sequences was determined by the use of gel electrophoresis and confirmed by DNA sequencing.

4.1 Spoligotyping

For spoligotyping we used genomic DNA. The hybridization pattern obtained from spoligotyping (**Figure 3.1**) was visualized by streptavidin-peroxidase autoradiography, and the most representative candidates were selected based on these results (**Table 3.2-3.3**). The spoligo-patterns are evolved from successive loss of DNA spacers, which separate the short DR sequences ***(Flores, L et al 2007)***. One of the selected samples (**nr.7**) was however replaced by a known strain (EAI1) at a later stage due to sub-optimal DNA quality. The membrane also featured two positive control strains (H37Rv & *M. bovis BCG),* as well as a negative control (H_2O). The 42 distinct patterns displayed on the *BioDyne C* membrane each represented a different strain of TB. These membrane patterns were initially converted to binary numbers (**Table 3.1**), and subsequently to octal numbers (**Table 3.3**), refer to *Section II: Materials & Methods* for conversion details. The obtained octal numbers were then used to identify the strains through an international database (SpolDB4). SpolDB4 is an updated version of SpolDB3, and contains clinical isolates from a total of 141 countries, which makes it ideal for MTC strain identification.

After the strains had been identified by the use of the methods described above, the 8 chosen samples had to be re-cultured on Lowenstein Jensen agar over a 4-6 week period. The colonies were then heat-killed over the course of 20 minutes at 95 ºC. The DNA was extracted and purified. This was done in order to prevent the infection hazard, obtain more DNA, and to optimize the quality of our results.

4.2 Restriction Fragment Length Polymorphism

Due to the amount of time required to grow new TB cultures, we only had once chance to produce results by the use of RFLP. This however failed most likely due to wrong temperatures in the heat-inactivation step. Thus the archived results could not be confirmed in the current study, but the methodology was.

4.3 Molecular methods

A series of preliminary tests were conducted to check the compatibility of the selected TB-strains with *primers 1* and *5*. After the gel had been visualized (**Figure 3.2-3.3**), it was clear that these early results were promising. Sample 2 however, failed to show initial results. The 3 remaining primers (in addition to *primer 1* and *5* for sample 2) were verified in the same way as described above. All of the primers repeatedly demonstrated results following the PCR amplification (**Figure 3.4**). The only exception was *primer 4 (*H37Rv1759c*)*, which failed to yield any results following multiple runs. We suspected that this was due to the size of this particular sequence, and adjusted the PCR-program to run for an extended period of time. This however still yielded no visible product during gel electrophoresis (results not shown). It was therefore decided to abandon it for the remainder of this part of the study. **Figures 3.2-3.4** demonstrate that results were obtained for all samples, excluding sample 8, primer 2. After the amplification of all the target sequences was completed, the PCR products were excised from the gel and purified for sequencing.

4.4 Sequencing

The raw data containing the DNA sequences was obtained from the ABI 3730 DNA Analyzer after about 2 hours. In order to get the most accurate representation, the sequences were opened and examined using BioEdit. Obvious reading-errors were corrected.

Each of the retrieved nucleotide sequences (32 in total) was compared to complete H37Rv genome. This was done using the online database; Basic Local Alignment Search Tool (BLAST). The nucleotide coverage (in %) between our results and the complete H37Rv genome was retrieved, and is displayed in **Table 3.6-3.7**. The sequence results were promising with mostly above 90% match.

Our results showed that the amplified target-sequences, confirmed that the PCR products were close to identical to those expected to produce the selected peptides.

Section V: Conclusion

It was obvious that the collection of *M. tuberculosis* at the reference laboratory at the Norwegian Institute of Public Health represented a wide diversity of strains that originated from various parts of the world.

This study confirmed that spoligotyping and RFLP are well suited for characterizing different strains of *M. tuberculosis*, and that they both have strengths and weaknesses.

The selected peptides were promising for a future diagnostic test for *M. tuberculosis* infection. This study concluded that all of the TB families that were tested should (in theory) be capable of producing these peptides. If a diagnostic kit could be developed based on these peptides, the kit could be equally sensitive in different epidemiological settings throughout the world.

Section VI: References

1. WHO. The World Health Report 2004: Changing History. Geneva: World Health Organization; 2004.
2. Kristin Kremer, Annelies Bunschoten, Leo Schouls, Dick van Sooligen and Jan van Embden. National Institute of Public Health and the Environment, Blithoven, The Netherlands. Version Number 2002 "Spoligotyping" A PCR-based method to simultaneously detect, and type *Mycobacterium tubercolosis* complex bacteria.
3. Saleh A. Bedeir. Tuberculosis in Ancient Egypt.
4. J D van Embden, M D Cave, J T Crawford, J W Dale, K D Eisenach, B Gicquel, P Hermans, C Martin, R McAdam and T M Shinnick. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology.
5. P W Hermans, D van Sooligen, J W Dale, A R Schuitema, RA McAdam, D Catty and J D van Embden. Insertion element IS986 from Mycobacterium tuberculosis: A useful tool for diagnosis and epidemiology of tuberculosis.
6. Inna Vitol, Jeffrey Driscoll, Barry Kreiswirth, Natalia Kurepina, Kristin P. Bennett, "Identifying Mycobacterium tuberculosis Complex Strain Families using Spoligotypes", Infection, Genetics and Evolution, Volume 6, Issue 6, November 2006, Pages 491-504.
7. J. Kamerbeek, L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology
   J Clin Microbiol, 35:5907-5914, 1997.
8. J.D.A. van Embden, T. van Gorkom, K. Kremer, R. Jansen, B.A. van Der Zeijst, and L.M. Schouls. Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria
   J Bacteriol, 9:2393-2401, 2000.
9. P.M. Groenen, A.E. Bunschoten, D. van Soolingen, and J. D. van Embden
   Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method
   Mol Microbiol, 10:1057-1065, 1993.
10. R.M. Warren, E.M. Streicher, S.L. Sampson, G.D. van der Spuy, M. Richardson, D. Nguyen et al. Microevolution of the Direct Repeat Region of *Mycobacterium*

tuberculosis: Implications for Interpretation of Spoligotyping Data
J Clin Microbiol, 40:4457-4465, 2002.

11. Karine Brudey[1], Jeffrey R Driscoll[2], Leen Rigouts[3], Wolfgang M Prodinger[4], Andrea Gori[5], Sahal A Al-Hajoj[6], Caroline Allix[7], Liselotte Aristimuño[8], Jyoti Arora[9], Viesturs Baumanis[10],Lothar Binder[11], Patricia Cafrune[12], Angel Cataldi[13], Soonfatt Cheong[14], Roland Diel[15],Christopher Ellermeier[16], Jason T Evans[17], Maryse Fauville-Dufaux[7], Séverine Ferdinand[1], Dario Garcia de Viedma[18], Carlo Garzelli[19], Lidia Gazzola[5], Harrison M Gomes[20], M Cristina Guttierez[21], Peter M Hawkey[17], Paul D van Helden[22], Gurujaj V Kadival[23], Barry N Kreiswirth[24], Kristin Kremer[25], Milan Kubin[26], Savita P Kulkarni[23], Benjamin Liens[1], Troels Lillebaek[27], Ho Minh Ly[28], Carlos Martin[29], Christian Martin[30], Igor Mokrousov[31], Olga Narvskaïa[31],Yun Fong Ngeow[14], Ludmilla Naumann[32], Stefan Niemann[33], Ida Parwati[34], Zeaur Rahim[35],Voahangy Rasolofo-Razanamparany[36], Tiana Rasolonavalona[36], M Lucia Rossetti[12], Sabine Rüsch-Gerdes[33], Anna Sajduda[37], Sofia Samper[38], Igor G Shemyakin[39], Urvashi B Singh[9], Akos Somoskovi[40], Robin A Skuce[41], Dick van Soolingen[25], Elisabeth M Streicher[22], Philip N Suffys[20], Enrico Tortoli[42], Tatjana Tracevska[10], Véronique Vincent[21], Tommie C Victor[22], Robin M Warren[22], Sook Fan Yap[14], Khadiza Zaman[35], Françoise Portaels[3], Nalin Rastogi[1]. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology.
12. Stephen D Lawn MRCP & Prof Almuddin I Zumla. Tuberculosis: The Desmond Tutu HIV Centre, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, South Africa..
13. WHO. Global tuberculosis control. Geneva: World Health Organization, 2010.
14. Mukhopadhyay S, Farver CF, Vaszar LT, Dempsey OJ, Popper HH, Mani H, Capelozzi VL, Fukuoka J, Kerr KM, Zeren EH, Iyer VK, Tanaka T, Narde I, Nomikos A, Gumurdulu D, Arava S, Zander DS, Tazelaar HD. Source: Department of Pathology, State University of New York Upstate Medical University, Syracuse, New York, USA. Causes of pulmonary granulomas: a retrospective study of 500 cases from seven countries
15. Hirsh AE, Tsolaki AG, DeRiemer K, Feldman MW, Small PM. Stable association between strains of Mycobacterium tuberculosis and their human host populations.

16. Sola C, Filliol I, Legrand E, Mokrousov I, Rastogi N. Mycobacterium tuberculosis phylogeny reconstruction based on combined numerical analysis with IS1081, IS6110, VNTR, and DR-based spoligotyping suggests the existence of two new phylogeographical clades.
17. Giovanni Ferrara MD a, Monica Losi PhD a, Roberto D'Amico PhD b, Pietro Roversi MD a, Roberto Piro MD a, Marisa Meacci BSc c, Barbara Meccugni BSc c, Ilaria Marchetti Dori BSc c, Alessandro Andreani MD a, Barbara Maria Bergamini MD d, Cristina Mussini MD e, Fabio Rumpianesi BSc c, Prof Leonardo M Fabbri MD a, Dr Luca Richeldi MD. Use in routine clinical practice of two commercial blood tests for diagnosis of infection with Mycobacterium tuberculosis: a prospective study.
18. Kaufmann SH, Schaible UE. 100th anniversary of Robert Koch's Nobel Prize for the discovery of the tubercle bacillus.
19. Dahle U, Eldholm V, Winje BA, Mannsåker T, Heldal E. Impact of immigration on the molecular epidemiology of Mycobacterium tuberculosis in a low-incidence country.
20. Alison M Elliott, Nkandu Luo, George Tembo, Benita Halwiindi, Ger Steenbergen, Lieve Machiels, Joseph Pobee, Paul Nunn, Richard J Hayes, Keith P W J McAdam Impact of HIV on tuberculosis in Zambia: a cross sectional study
21. Elizabeth L Corbett, Barbara Marston, Gavin J Churchyard, Kevin M De Coc. Tuberculosis in sub-Saharan Africa: opportunities, challenges, and change in the era of antiretroviral treatment
22. Richard E. Chaisson, M.D., and Neil A. Martinson, M.B., B.Ch., M.P.H. Tuberculosis in Africa — Combating an HIV-Driven Crisis
23. Stop TB Partnership (2006). The global plan to stop TB 2006–2015. Geneva: WHO
24. Young DB, Perkins MD, Duncan K, Barry CE (2008) Confronting the scientific obstacles to global control of tuberculosis. J Clin Invest 118: 1255–1265
25. Kaufmann SH, Parida SK (2007) Changing funding patterns in tuberculosis. Nat Med. 13: 299–303.
26. Frieden TR, Fujiwara PI, Washko RM, Hamburg MA (1995) Tuberculosis in New York City–turning the tide. N Engl J Med 333: 229–233
27. Small PM, Shafer RW, Hopewell PC, Singh SP, Murphy MJ, et al. (1993) Exogenous reinfection with multidrug-resistant Mycobacterium tuberculosis in patients with advanced HIV infection. N Engl J Med 328: 1137–1144

28. R S Heydermana, M Goyalc, P Roberts, S Ushewokunzea, S Zizhoua, B G Marshallc, Robert Makombed, J D A Van Embdene, P R Masonb, R J Shawc. Pulmonary tuberculosis in Harare, Zimbabwe: analysis by spoligotyping
29. H.S. Lee, J.Y. Oh, J.H. Lee, C.G. Yoo, C-T. Lee, Y.W. Kim, S.K. Han, Y-S. Shim and J-J. Yim. Response of pulmonary tuberculomas to anti-tuberculous treatment
30. Department of Biology, Davidson College, Davidson, NC 28036
31. National TB Collectors Association / CDC Advisory Group of Tuberculosis Genotyping. Guide to Application of Genotyping to Tuberculosis Prevention and Control. Atlanta, GA: US Department of Health and Human Services, CDC; June 2004.
32. G Marchetti, A Gori, L Catozzi, M C Rossi, M Moroni, F Franzetti. Comparison of spoligotyping vs RFLP DNA fingerprinting analysis in M. tuberculosis epidemiological typing
33. Schatz A, Waksman SA. Strain specificity and production of antibiotic substances: IV. Variations among Actionomycetes, with special reference to Actinomyces griseus. Proc Natl Acad Sci USA 1945; 31: 129–37.
34. Feldman WH, Karlson AG, Hinshaw HC. Streptomycin-resistant tubercle bacilli; effect of resistance on therapeutic results. Am J Pathol 1947;23: 874.
35. M Goyal; N A Saunders; J D van Embden; D B Young; R J Shaw Journal of Clinical Microbiology -- 35(3) 1997/03/01.Differentiation of Mycobacterium tuberculosis isolates by spoligotyping and IS6110 restriction fragment length polymorphism. Medline/Pubmed
36. Flores L, Van T, Narayanan S, DeRiemer K, Kato-Maeda M, Gagneux S.Large sequence polymorphisms classify Mycobacterium tuberculosis strains with ancestral spoligotyping patterns.
37. Sebastien Gagneux, Kathryn DeRiemer, Tran Van, Midori Kato-Maedab, Bouke C. de Jongb, Sujatha Narayanan, Mark Nicol, Stefan Niemann, Kristin Kremer, M. Cristina Gutierrez, Markus Hilty, Philip C. Hopewelle, and Peter M. Small. Variable host–pathogen compatibility in Mycobacterium tuberculosis.
38. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, Garnier T, Gutierrez C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, van Soolingen D, Cole ST.Source. Unité de Génétique Moléculaire Bactérienne, Laboratoire de Génomique des Microorganismes Pathogènes, and Centre National de Référence des Mycobactéries, Institut Pasteur, 25-28 Rue du Docteur Roux, 75724 Paris Cedex 15, France. A new evolutionary scenario for the Mycobacterium tuberculosis complex.

39. Basic Local Alignment Search Tool, BLAST. Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "A greedy algorithm for aligning DNA sequences", J Comput Biol 2000; 7(1-2):203-14.
40. Cole. ST., Eiseach D. K., McMurry D. N., Jacobs W. R,. Tuberculosis and the Tubercle Bacillus. 2004: ASM Press. 584 pp.
41. Transcriptional Adaptation of Mycobacterium tuberculosiswithin Macrophages
42. Dirk Schnappinger , Sabine Ehrt, Martin I. Voskuil 1 ,Yang Liu, Joseph A. Mangan , Irene M. Monahan, Gregory Dolganov, Brad Efron, Philip D. Butcher, Carl Nathan, and Gary K. Schoolnik 1. Insights into the Phagosomal Environment
43. Rothschild B. M., Martin L. D., Lev G., Bercovier H., Bar-Gal G. K., Greenblatt. C., Donoghue H., Spigelman M. and Brittain D., *Mycobacterium tuberculosis* complex DNA from an extinct bison dated 17, 000 years before the present. Clin Infect Dis, 2001. 33(3): p. 305-11
44. Zink A. R, Sola. C, Reischl. U, Grabner. W, Rastogi, N, Wolf. H, and Nerlich. A. G. Characterization of *Mycobacterium tuberculosis* complex DNA from Egyptian mummies by Spoligotyping.
45. Tiago Dos Vultos; Olga Mestre; Jean Rauzier; Marcin Golec; Nalin Rastogi; Voahangy Rasolofo; Tone Tonjum; Christophe Sola; Ivan Matic; Brigitte Gicquel. Evolution and diversity of clonal bacteria: the paradigm of Mycobacterium tuberculosis.