Genetic diversity and resistance to anti-tuberculosis drugs of Mycobacterium tuberculosis in China

Kanglin Wan

UNIVERSITE PARIS-SUD XI

THESE

Pour obtenir le grade de

DOCTEUR EN SCIENCES DE L'UNIVERSITE PARIS-SUD

Présentée par

Kanglin WAN

Study on Genetic diversity and resistance to anti-TB drugs of *Mycobacterium tuberculosis* in China

Soutenue le 8 Octobre 2007 devant la commission d'examen :

Mr Dirk van Soolingen	Rapporteur
Mr. Christophe Sola	Rapporteur
Mr. Longfei Wu	Examinateur
Mr. Gérard Leblon	Président du Jury
Mr. Gilles Vergnaud	Directeur de Thèse
Mr. Xu Jianguo	Directeur de Thèse

At this time, feeling deeply grateful, I would like to thank every one, teachers, friends, colleagues and families, for your kind help and understanding in the course of my PhD thesis studies.

I would like to thank my supervisors Dr. Gilles Vergnaud and Prof. Jianguo Xu for giving me this very precious opportunity to study and improve my scientific knowledge.

Especially from the bottom of my heart I would like to thank Dr. Christine Pourcel for doing a lot of work on the project design, organization and implementation, especially for your guidance during the thesis writing, for helping me correct the manuscript, and for many other things. Thank you very much for your kindest help, patience and understanding.

I would like to thank Prof. Christophe Sola for his kindest help to correct and improve my manuscript. I would like to thank Prof. Dick van Soolingen for his kindest guidance and help for the project and for suggesting many valuable improvements of the thesis manuscript. I would like to thank every Professors of the committee for accepting to be members of my thesis jury.

I would like to thank Philippe Le Flèche, Yolande Hauck and other staff members of GPMS, Institut de Génétique et Microbiologie, Université Paris-Sud for their help.

I would like to thank my colleagues Liu Jinghua, Liu Zhiguang, Zhang Yuanyuan, Lv Bing, Dong Haiyan, Jiang Yi, Zhao Xiuqin, and other staff members and students of National Laboratory of Tuberculosis, State Key

Laboratory for Infectious Disease Prevention and Control / National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention for their help.

I would like to thank staff members of the Tuberculosis laboratories who provided samples and information for my thesis project in Beijing, Henan, Jiang Su, Sichuan, Hunan, Anhui, Zhejiang, Fujian, Ji Lin, Shannxi, Gansu, Xinjiang, Xizang, and Guangxi.

I would like to thank my wife and daughter for their help, encouragement, patience and understanding.

This work could not have been done without the support of the European Community through the SSA grant "Tuberculosis China" N° 012166.

Thank every one very much again.

Table of Contents

List of Abbreviations

ADR	Secondary Resistance or acquired drug resistance
AFB	Acid-Fast Bacilli
AFLP	amplified fragment length polymorphism
BAC	bacterial artificial chromosome
BCG	Bacillus Calmette-Guerin
bp	base pairs
CRISPR	clustered regularly interspaced short palindromic repeats
DNA	deoxyribonucleic acid
dNTP	Deoxynucleoside tRFPhosphate
DR	direct repeat
EAI	the East African Indian
EDTA	Ethylene diamine tetraacetic acid
EMB	Ethambutol
ETR	exact tandem repeat
g	Gram
HGDI	Hunter-Gaston diversity index
IDR	Initial drug resistance
INH	Isoniazid
IS6110	insertion sequence *6110*
IS	insertion sequence
kb	kilobase
kg	Kilogram
LAM	the Latin-American and Mediterranean
LSP	Large-sequence polymorphisms
M	Mole
M. africanum	*Mycobacterium africanum*
M. bovis	*Mycobacterium bovis*
M. canettii	*Mycobacterium canettii*
M. caprae	*Mycobacterium caprae*
MDR	multi-drug resistance
MDR-TB	multi-drug resistance TB
mg	Milligram
MIC	minimal inhibitory concentrations
Min	Minute
MIRU	mycobacterial interspersed repetitive units
MLVA	multiple loci VNTR analysis
ml	Millilitre
M. microti	*Mycobacterium microti*
mM	Millimole
M. pinipedii	*Mycobacterium pinipedii*
MPTR	major polymorphic tandem repeats
M. tuberculosis	*Mycobacterium tuberculosis*
NAP	P-nitro-alpha-acetylamino-beta-hydroxypropiopheneone
nsSNP	nonsynonymous single-nucleotide polymorphisms
NTM	non-tuberculous mycobacteria
PCR	polymerase chain reaction
PDR	primary drug resistance
PFGE	pulsed field gel electrophoresis
PPD	purified protein derivative
SNP	single nucleotide polymorphism
RAPD	the DNA polymorphism of randomly-amplified-PCR
RD	Region of deletions

RDH	Reverse-line Blot Hybridation
RFP	Rifampicin
RFLP	restriction fragment length polymorphism
rpm	revolutions per minute
rpoB	the RNA polymerase B subunit
PNB	4-Nitrobenzoic acid
SDS	Sodium dodecyl sulphate
SM	Streptomycin
SNP	single nucleotide polymorphism
SPIDR	the spacer interspersed DR
Spoligotyping	spacer oligonucleotide typing
sSNP	synonymous single-nucleotide polymorphisms
TB	Tuberculosis
TbD	M. tuberculosis-specific deletion
TCH	2-Thiophencarboxtylic acid hydrazide
Tris	Tri hydroxymethyl aminomethance
µg	Microgram
µl	Microlitre
µM	Micromole
VNTR	variable number of tandem repeats
WHO	the World Health Organization
XDR	Extensively drug-resistant

1. INTRODUCTION

1.1. Introduction on tuberculosis

Tuberculosis (TB) has occurred in humans for at least several thousands years with archeological findings from a number of Neolithic sites in Europe and sites from ancient Egypt to the Greek and Roman empires showing evidence of a disease consistent with modern TB.

TB caused by Mycobacterium tuberculosis（M. tuberculosis） which was first identified by Robert Koch in 1882 (Koch, 1882), is transmitted primarily by inhalation of aerosolized droplets containing the organisms. Before molecular technic could be applied to the characterisation of bacilli responsible for tuberculosis, biochemical technics allowed their separation into several subspecies which formed the *Mycobacterium tuberculosis* complex (MTBC): *M. tuberculosis* and *Mycobacterium africanum* (*M. africanum*) were strictly found in humans whereas *Mycobacterium bovis* (*M. bovis*) *M. bovis* was principally found in cattle but could also infect humans. *Mycobacterium microt (M. microti), Mycobacterium pinipedii* (*M. pinipedii*) and *Mycobacterium caprae* (*M. caprae*) are nonhuman pathogens. In addition rares isolates, the *Mycobacterium canettii* (*M. canettii*)., forming smooth colonies and showing characteristic phenotypic properties can be found in a very limited geographical area.

Human beings are infected through the respiratory tract and the tubercle bacilli spread by lymphatic system and blood stream to many different organs.

TB is spread through by airborne contamination from one person to another. The bacteria are transmitted when a person with active TB disease of the lungs or throat coughs or sneezes. People nearby may breathe in these bacteria and become infected. The course of the TB disease is divided into primary pulmonary TB, latent TB infection and active TB disease including tuberculo-meningitis, skin TB, scrofulous TB, pleurisy TB, cordis TB, urinary systemic TB, digestive systemic TB, skeletal TB, and so on. People with active TB disease can be treated and cured if they seek medical help. Even better, people with latent TB infection can take medicine so that they will not develop active TB disease.

Primary pulmonary TB (the first infection with the TB bacteria) which often occurs in older infants and children usually produces no signs or symptoms, and a chest X-ray shows no signs of infection. Rarely, there may be an enlargement of the lymph nodes and possibly some coughing. In most cases, only a tuberculin skin test is positive, indicating that the child has been infected. This

primary infection usually resolves on its own as a child develops immunity over a 6- to 10-week period. But in some cases, it can progress and spread all over the lungs (called progressive TB) or to other organs. This causes signs and symptoms such as fever, weight loss, fatigue, loss of appetite, and cough.

Another type of infection is called reactivation TB. Here, the primary infection has resolved, but the bacteria are dormant, or latent. When conditions become favorable (for instance, a lowered immunity), the bacteria become active. Reactivation TB may occur in older children and adults. The most prominent symptom is a persistent fever, with sweating during the night. Fatigue and weight loss may follow. If the disease progresses and cavities form in the lungs, the person may experience coughing and the production of saliva, mucus, or phlegm that may contain blood.

Latent TB infection occurs when TB bacteria manage to escape the immune system. The body infected with *M. tuberculosis* has no symptoms and is not contagious, but the "germ" is not completely eliminated from the body. For this reason it is called a "latent infection." People who have latent TB infection do not feel sick, do not have any symptom, and cannot spread TB to others. Many people who have latent TB infection will never develop active TB disease. In these people, the TB bacteria may remain inactive for a lifetime without causing disease. The problem with TB is when the latent infection turns "active."

Generally, the primary pulmonary TB infections are asymptomatic and a positive result of tuberculin skin test is the only indication of it. Approximately 10% of infected individuals develop active disease. Immunosuppression caused by systemic diseases or medication increases the risk of developing TB. In addition to the pulmonary tract, TB may affect many organs and systems, including lymph nodes, larynges, middle ear, genitourinary tract, musculoskeletal system, central nervous system, gastrointestinal tract, pericardium, and skin. The diagnosis is based upon the detection of mycobacteria in fluids and tissues. If there is no available material for analysis, a presumptive diagnosis is made and therapeutic test initiated. The therapy is based upon anti-tuberculous drugs and corticosteroids.

Some people (approximately 10%) with latent TB infection will go on to activate the TB disease. Active TB is contagious. Some people develop active TB disease soon after becoming infected, before their immune system can fight the TB bacteria. Other people may get sick later, when their immune system becomes weak for another reason. People with active TB disease are most likely to spread it to people they spend time with every day. This includes family members,

friends, and coworkers. Symptoms of active TB disease include: weakness or fatigue, weight loss, no appetite, chills, fever, and sweating at night.

TB is infectious when it is airborne and can be inhaled by others. In general, children are not considered infectious, and usually get the infection from infected adults. The incubation period (the time it takes for a person to become infected after being exposed) varies from weeks to several years, depending on the individual and whether the infection is primary, progressive, or reactivation TB.

The prevention of TB depends on: avoiding contact with those who have an active disease; using antibioprophylaxy as a preventive measure in high-risk cases (preventive administration of antibiotics); maintaining good living standards. Isolation (quarantine) of patients harboring multi-drug-resistant (MDR) and extensively-drug-resistant (XDR) strains is a required measure to lower MDR transmission. New cases and potentially contagious patients are identified through proper use and interpretation of the tuberculin skin test. A vaccine called Bacille Calmette-Guérin (BCG) is considered controversial because it is not very effective in countries with a low incidence of TB. Before considering BCG as “controversial”, it would be good to say some word about this vaccine in a dedicated paragraph. In particular, say that this vaccine is efficient in some settings, and has been used for decades to control and almost eradicate TB in developed countries; also say that its efficiency is related to the level of non-tuberculous mycobacteria (NTM) in the environment – the vaccine is less efficient in India than in Europe for these reasons, and finally say that the vaccine probably evolved and is perhaps less efficient today than before because the vaccine stocks is not the same as initially (evolution) and may be because BCG may have induced a selection pressure for more virulent strains, like *M. tuberculosis* Beijing family strains. (Brosch et al., 2007) For this reason, BCG is not usually given in the United States. However, it may be considered for children immigrating to countries where TB is prevalent.

1.2. The epidemiology of tuberculosis

1.2.1. Nomenclature and definitions of TB epidemiology

Annual Infection Rate

In the present stage of development where TB mortality rate has lost its statistical significance because of effective chemotherapy, it is generally recognized that the most reliable measure of the

extent of the TB problem in a population is the "annual TB infection rate". In countries where infection with the bovine type of tubercle bacilli no longer exists, as is the case in the majority of low prevalence countries, the annual infection rate expresses that proportion of the population under study which will be primarily infected, or re-infected with tubercle bacilli from a human source in the course of one year. The annual TB infection rate is also the best measure for following the trend of the TB problem in a given population and for evaluating the total effects of organized efforts to control TB.

Infection Rate or incidence of infection

When we conduct an investigation on TB in a population to understand the situation of *M. tuberculosis* infection, we usually use the purified protein derivative (PPD) test (tuberculin skin test) with the injection of a TB-like substance (tuberculin) sub-cutaneously, using a very small needle. About two days later, the area is examined for a red, rash-like reaction with a size of more than 10 mm (millimeters). Then we calculate the number of PPD positive. Infection rate or incidence of infection is the number of PPD positive per 100 people (Seibert, 1934).

Smear positive rate

Sputum smear microscope examination by Ziehl-Neelsen staining is usually used in patients suspected to have TB. In the diagnosis of TB in suspected patients, smear positive rate is the rate of smear positive for Acid-Fast Bacilli (AFB) among patients under examinations (positive number/examination number × 100%). But in an epidemiological investigation of TB in the whole population, the incidence of smear positive patients is the number of the smear positive TB cases per 100,000 people.

Culture positive rate

Culture positive is usually the *M. tuberculosis* positive result of sputum isolation examination in the patients suspected of TB, by culturing on Löwenstein-Jensen medium. Culture positive rates and the incidence of culture positive cases are calculated as for smear positive rates.

Incidence and Prevalence

Incidence is the rate of new TB cases in 100,000 people, and prevalence is the rate of all TB cases (including new and old cases) per 100,000 people.

Mortality

Mortality refers to the rate of death due to TB (or other reason) per 100,000 people.

Drug Resistance Rate

Drug resistance rate is the rate of drug resistance strains as assessed by drug susceptibility tests (e.g. the number of drug resistance strains/the number of examined strains × 100%) including single drug resistance rate and multiple drug resistance rate.

Primary Resistance or Initial Resistance

This reflects the fact that the bacterium has a characteristic resistance against at least one drug before any treatment. According to the guidelines of drug resistance surveillance published by WHO, primary drug resistance (PDR) concerns *M. tuberculosis* isolated from patient which never used any anti-*M. tuberculosis* drug or used anti-*M. tuberculosis* drug for less than one month. Initial drug resistance (IDR) is identical to PDR. Because some time it is difficult to know whether a patient used anti-*M. tuberculosis* drugs or not when we ask the history of the patient, we consider that the patient has never used anti-*M. tuberculosis* drug, and this kind of drug resistance is called IDR.

Secondary Resistance or acquired drug resistance (ADR)

After treatment has started, drug-resistant variants may emerge although the bacterium initially isolated from the patient was sensitive to drugs.

Multi Drug Resistance TB (MDR-TB)

MDR-TB usually means that *M. tuberculosis* has resistance against≥2 anti -*M. tuberculosis* drugs. More precisely, and according to the guideline of drug resistance surveillance published by WHO, the term MDR-TB reflects resistance against ≥2 anti-*M. tuberculosis* drugs at least including Isoniazid (INH) and Rifampicin (RFP).

Extensive Drug Resistance

(XDR ; http://www.who.int/tb/xdr/taskforcereport_oct06.pdf)

XDR-TB means that the clinical isolate is MDR and is also resistant to fluoroquinolones and to either one injectable drug aminoglycosides (amikacin, kanamycin) or capreomycin or both.

1.2.2. The General prevalence situation of TB in the world

TB is a severe chronic infectious disease principally affecting the respiratory system, and is one of the infectious diseases which still threats people's health heavily. In the past 15 years, the epidemiology situation of TB has worsened sharply and the number of TB patients has been increasing in either advanced countries or developing countries due to multiple causes including increase of drug resistance, especially MDR, prevalence of HIV/AIDS, migration, ignorance of TB control and prevention. TB is spread by infectious droplets through coughing, sneezing, or spitting and thrives in conditions of poverty and overcrowding. A person with active TB can infect an average of 15 people a year.

TB was declared as a public health world-wide emergency in 1993. According to WHO, nearly 2,000 millions people have been infected by *M. tuberculosis*, with 20 million TB patients all over the world (Raviglione, 2003; WHO, 2004b). There is a new infection occurring every minute, and each year 8-10 millions new TB cases occur and 2 millions people die of TB. In 1995, TB death toll, which broke the record after 1850, reached 3 millions. More than 2 third of patients with TB are aged 16 to 50 years old. So TB has become one of the most important infectious diseases at present. In 1993 the WHO declared TB a Global Health Emergency. Therefore efforts toward TB control and prevention must be enhanced and improved. The objective of the Millennium Development Goals (MDG) of the United Nations Organization is to reach a 70% detection of active TB cases and 85% cure rate, however these goals are far from being reached.

1.2.3. The general prevalence situation of TB in China

The following paragraphs were modified basically from Report on fourth national epidemiological sampling survey of TB (Tuberculosis, 2002).

In China, the situation of TB prevalence is very serious too. China is the second of the 22 countries with 79% of the world's burden of TB, registering 17% of global cases. (Raviglione, 2003; WHO, 2004b) (Table 1).

Table 1 The estimated incidence rate and new cases/year of 22 high-burden countries by WHO

Country	Incidence (1/ 100,000)	% of total cases in the world
India	1.86	23
China	1.37	17
Indonesia	0.60	8
Nigeria	0.35	4
Bangladesh	0.33	4
…	…	…
Total of 22	0.69	79

There have been three nationwide epidemiological investigations of TB by random sampling in China (except Taiwan, Hongkong and Macao), performed in 1979, 1990, and 2000, respectively. Data were stratified by age, sex, region (west, central, or east), and Project (which is the special project for TB control from The Global Fund) and non-Project area. Proportional sampling was used for the country as a whole and prevalence characteristic of TB in China was determined.

According to the 2000 Nationwide Random Survey for the Epidemiology of TB in China (except Taiwan, Hong Kong and Macao) to identify previously undiagnosed patients by the multistratified random sampling method, the infection rate is 44.5% (Figure 1), and 550 million people were infected with *M. tuberculosis*. The prevalence of active pulmonary TB was 367/100,000, the prevalence of smear positive pulmonary TB was 122/100,000 and the prevalence of bacteriological positive pulmonary TB was 160/100,000. From these results it was calculated that there are probably 4.51 million active patients in the whole country, including 1.50 million smear-positive cases, which will become the infectious sources. So the task that we are facing to control TB is still very hard in China.

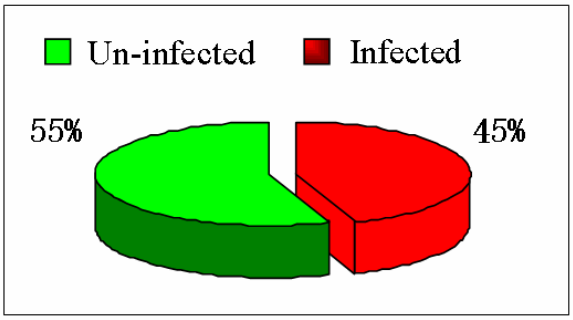


Figure 1 The situation of TB infection in China

The epidemiological investigations of TB performed in 1979, 1990 and 2000 showed that the incidence had decreased. With regard to the annual average decreasing rate in total population from 1979 to 2000, the morbidity rate of active TB decreased 2.6%, and the smear-positive rate of active TB decreased 2.1% (Table 2, Figure 2).

Table 2 Adjusted incidence rate, decreasing rate and annual decreasing rate of TB from 1979 to 2000 in China

Kind of cases	Incidence rate (1/100 000)			Decreasing rate (%) *			Annual average decreasing rate (%) #		
	1979	1990	2000	1979 -1990	1990 -2000	1979 -2000	1979 -1990	1990 -2000	1979 -2000
Active TB	796	523	367	34.3	29.8	53.9	3.1	3.0	2.6
Smear positive	187	134	122	28.3	9.0	34.8	2.6	0.9	1.7
Culture positive		177	124		29.9			3.0	

*** Decreasing rate of incidence rate: Active TB, 1979-1990: (796-523)/796; 1990-2000: (523-367)/523; 1979-2000: (796-367)/796; Smear positive, 1979-1990: (187-134)/187; 1990-2000: (134-122)/134; 1979-2000: (187-122)/187; Culture positive, 1990-2000: (177-124)/177. # Annual average decreasing rate of morbidity: Active TB, 1979-1990:(796-523)/796/11; 1990-2000: (523-367)/523/10; 1979-2000: (796-367)/796/21; Smear positive, 1979-1990: (187-134)/187/11; 1990-2000: (134-122)/134/10; 1979-2000: (187-122)/187/21; Culture positive, 1990-2000: (177-124)/177/10.**

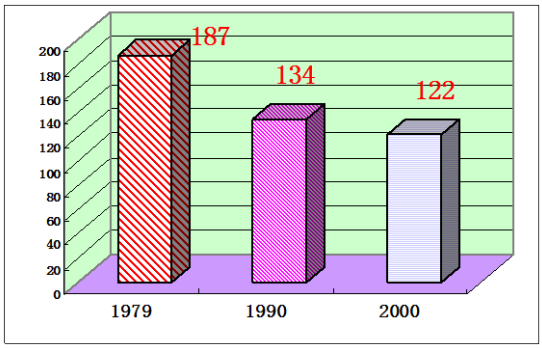


Figure 2 The morbidity rate (1/100,000) of the smear-positive in three investigations

Though the mortality has decreased, it is still very high (9.8 per million), with 250,000 patients dying of TB each year, because the population is increasing rapidly. The number of deaths caused by TB is twice the total number of deaths caused by the other infectious diseases (including verminosis).

In China there are 31 provinces (municipality, autonomous region) except Hong Kong Macao and Taiwan.

The Eastern region include the following provinces: Fujian, Guangdong, Hainan, Hebei, Jiangsu, Liaoning, Shandong, Zejiang, and the three municipalities of Beijing, Shanghai, and Tianjin. The Central region included Anhui, Jilin, Heilongjiang, Henan, Hubei, Hunan, Jiangxi and Shanxi. The Western region consisted of Sichuan, Chongqing, Gansu, Qinghai, Xinjiang, Ningxia, Xizang, Guizhou, Shaanxi, Yunnan, Neimeng and Guangxi. The geographical location of the provinces is in Figure 3.

Figure 3 The map of the People's Republic of China

The TB patients were classified according to sex (female and male), age group (0-, 5-, 10-, 15-, 20-, 25-, 30-, 35-, 40-, 45-, 50-, 55-, 60-, 65-, 70-, 75-, and 80- years), geographical region (Eastern, Central and Western, or rural area, town and city), and drug susceptibility pattern. Upon comparison of the results of the three investigations on TB in China (except Taiwan, Hongkong and Macao), the morbidity rate of smear positive patients was very low and was no different in the 4 groups under 16 years of age, but in most of the others groups the morbidity increased with age in each investigation (Table 3, Figure 4).

Table 3 The distribution of TB cases in age groups

Age group	Active TB (%)	Smear-positive (%)	Total (%)
16 ~ 50	67.9	76.3	62
others	32.1	23.7	38
Total (%)	100	100	100

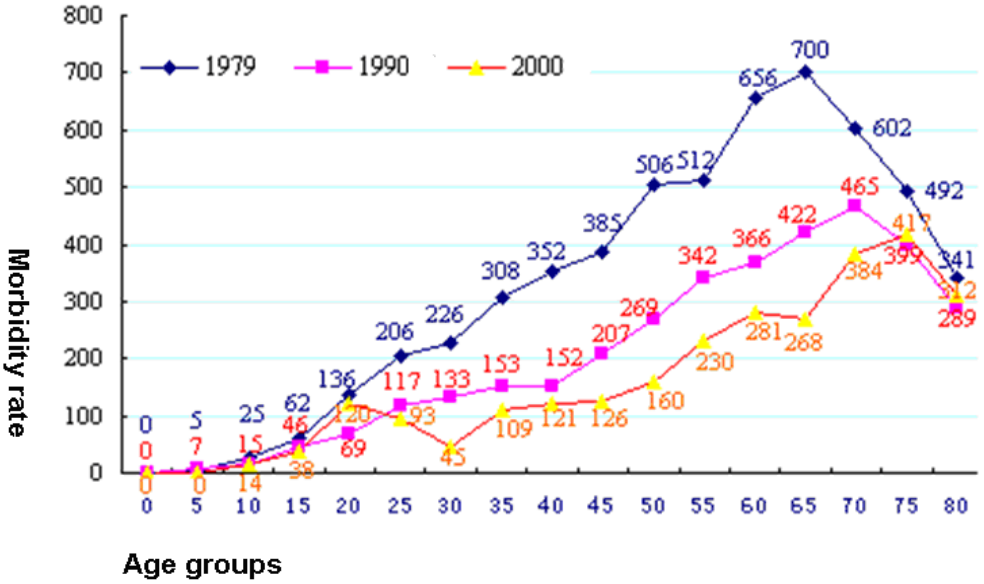


Figure 4 The morbidity rate (1/100,000) of the smear-positive of patients according to age groups in the three investigations

From 1979 to 2000 the overall morbidity rates decreased. In the 2000 investigation in China (except Taiwan, Hong Kong and Macao), the morbidity rate of smear positive patients was very low in the ＜16 age groups and was not different in both male and female groups. However in most of the other age groups the morbidity rate of smear positive patients was higher in males than in females and increased with age in both groups (Figure 5).

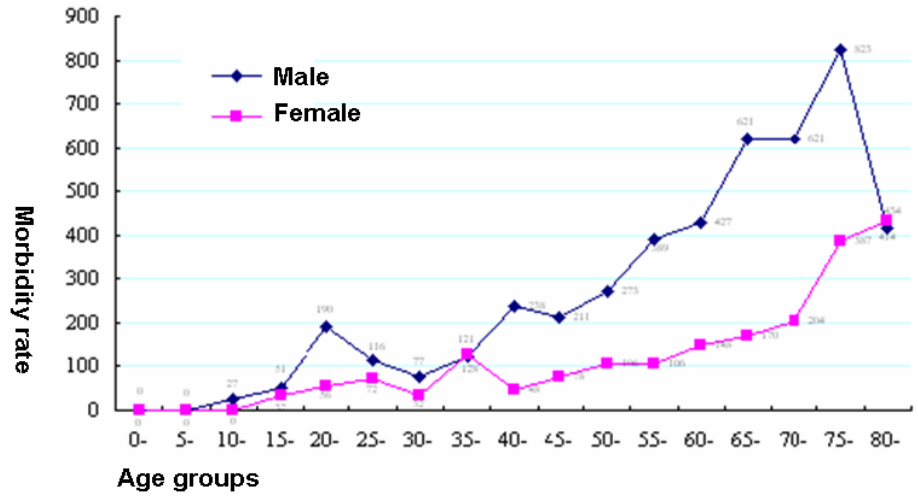


Figure 5 The morbidity rate (1/100,000) of the smear-positive male and female TB patients in the 2000 investigation

The epidemic situation differs greatly in different geographical regions. The infection rate in rural area is much higher than in cities. The infection rate of population in rural area, town and city was 59.4%, 55.1% and 35.9% respectively. The morbidity rate in rural area was twice the morbidity rate in cities (Figure 6).

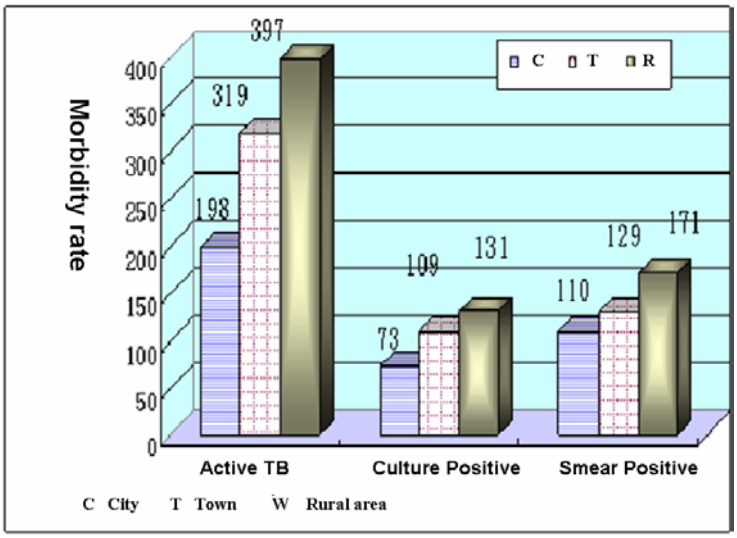


Figure 6 The morbidity rate of different kinds of TB in rural area and in cities

The morbidity rate was 1.7 times higher in the western part of China than in the Eastern part (Figure 7).

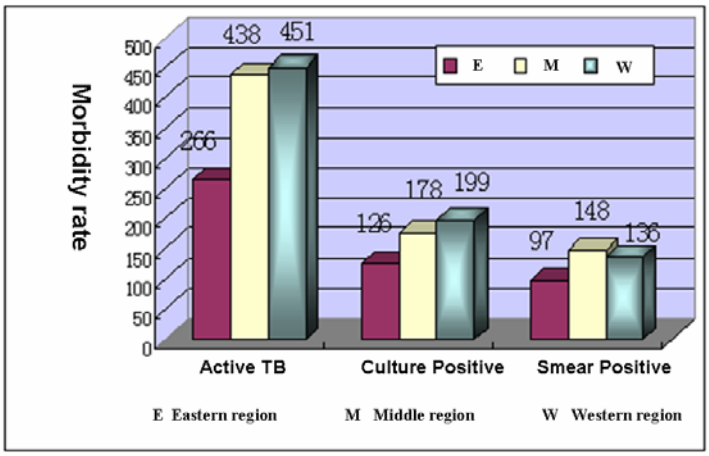


Figure 7 The morbidity rate of different kinds of TB in different geographical regions

In recent years, epidemiological situation of TB worldwide shows a tendency toward a raise. One of the main causes is that infection by drug-resistant *M. tuberculosis* (DR-TB) strains, especially the multidrug-resistant *M. tuberculosis* (MDR-TB), is prevailing. The situation of DR-TB infection in China is much worse than the world average. According to the statistics, the total rate of drug-resistance was 27.8%, and MDR rate was as high as 10.7% (Table 4). WHO estimates that a third of the world's cases of MDR TB are in China, even though the country has only 17% of the global burden of TB.

Table 4 The situation of drug-resistance in China in 2000

Kind of resistance	Total resistance (%)	MDR (%)
Primary resistance rate	18.6	7.6
Secondary resistance rate	46.5	17.1
Total	27.8	10.7
World average	10.4	1.7

1.3. Drug resistance and drug resistance detection

The following contents are modified basically from Anandi Martin and Françoise Portaels，Chapter 19: Drug Resistance and Drug Resistance Detection of Juan Carlos Palomino, Sylvia Cardoso Leão and Viviana Ritacco, TB 2007, www.TB.Textbook.com

Drug resistance in TB is a matter of great concern for TB control programs since there is no cure for MDR-TB strains of *M. tuberculosis* which could spread around the world. MDR-TB, defined as resistance to at least rifampicin (RFP) and isoniazid (INH), is a compounding factor for the control of the disease. XDR-TB now constitutes an emerging threat for the control of the disease and the further spread of drug resistance. For this reason, rapid detection of drug resistance to both first- and second line anti-TB drugs has become a key component of TB control programs (XDR, 2006).

1.3.1. Conventional phenotypic methods

Phenotypic methods for drug resistance tested in TB are based on cultivation of *M. tuberculosis* in the presence of antibiotics, which have been most commonly performed on egg-based or agar-based solid media, including the proportion method (Kent and Kubica, 1985; Heifets, 2000), the resistance ratio method (Kent and Kubica, 1985; Heifets, 2000), the absolute concentration method (Heifets, 2000), the BACTEC radiometric method (Becton Dickinson, Sparks, MD) (Roberts et al., 1983; Pfyffer et al., 1999; Heifets, 2000), and more recently the Mycobacteria Growth Indicator Tube (MGIT) (Becton Dickinson, Sparks, MD) (Pfyffer et al., 1997; Idigoras et al., 2000; Goloubeva et al., 2001; Johansen et al., 2004; Rusch-Gerdes et al., 2006).

In general, phenotypic methods assess inhibition of *M. tuberculosis* growth in the presence of antibiotics to distinguish between susceptible and resistant strains. This is possible since *M. tuberculosis* isolates from patients never treated before are very uniform in their level of susceptibility, as shown by the narrow ranges of minimal inhibitory concentrations (MIC) of the main anti-TB drugs (Heifets, 1996).

1.3.2. Genetic methods

With the development of molecular biological techniques, genetic changes were found in drug resistance *M. tuberculosis* strains. The mutation of some genes were described, for example, RFP-resistant strains show mutations in the RNA polymerase B subunit (*rpoB*) gene (Musser, 1995; Telenti et al., 1997); whereas INH-resistant strains have mutations principally in *katG*, pre-*inhA*, *inhA*, *ndh* and *oxyR-ahpC* genes (Lee et al., 2001).

Genetic methods for drug resistance tested in TB are based on the detection of mutations in the corresponding genes for drug resistance and on search for the genetic determinant of resistance. The detection at the genotypic level usually involves two basic steps: nucleic acid amplification such as polymerase chain reaction (PCR), to amplify the sections of *M. tuberculosis* genome known to be altered in resistant strains; and a second step assessing the amplified products for specific mutations correlating with drug resistance (Musser, 1995; Garcia de Viedma, 2003; Palomino, 2005).

Many genotypic methods for drug resistance detection on *M. tuberculosis* have been suggested, such as DNA sequencing, Solid-phase hybridization techniques, Real-time PCR techniques,

Reverse-line Blot Hybridation (RDH), Microarrays, etc. Some commercial kits are also available (Inno-Lipa, Hain diagnostics products)(see next paragraph).

DNA sequencing

At present, sequencing PCR-amplified products has become the most widely used genotypic method for analysing drug resistance in *M. tuberculosis*. Because of its accuracy and reliability it has become the reference standard for mutation detection. It is performed routinely for the detection of drug resistance mutations for several drugs (Hazbon, 2004). However, it is not an absolute method as not all molecular mechanisms of drug resistance for *M. tuberculosis* are known and furthermore it is still difficult and relatively expensive to implement on a systematic basis.

Solid-phase hybridization techniques

Solid-phase hybridization techniques are based on reverse hybridization of PCR-amplified DNA from cultured strains or clinical samples to probes specific for the drug resistant genes of *M. tuberculosis*, immobilized on a nitrocellulose stRFP (De Beenhouwer et al., 1995). There are currently two commercially available solid-phase hybridization techniques for the rapid detection of drug resistance in TB: the Line Probe Assay (INNO-LiPA Rif TB Assay, Innogenetics, Ghent, Belgium) for the detection of resistance to RIF and the GenoType *M. tuberculosis*DR assay (Hain Lifesciences, Nehren, Germany) for the simultaneous detection of resistance to INH and RFP. Four studies that applied LiPA directly to clinical specimens had 100 % specificity, and the sensitivity ranged from 80 % to 100 % (Morgan et al., 2005). In a recent study that evaluated the GenoType *M. tuberculosis*DR assay in 143 *M. tuberculosis* isolates, 99 % of the MDR strains were found to have mutations in the *rpo*B gene and 88.4 % of strains with mutations in the codon 315 of the *kat*G gene were also correctly identified (Hillemann et al., 2005).

Another solid-phase reverse hybridization test (rifoligotyping assay) was developed at the National Institute of Public Health and the Environment (http://www.rivm.nl/en/) in the Netherlands for rapid detection of RFP resistance-associated mutations in the *rpo*B gene of *M. tuberculosis*. It was initially evaluated at the Cetrángolo Hospital in Argentina (Morcillo et al., 2002)]. A total of 135 *M. tuberculosis* isolates were tested and the results compared with the proportion method and the MGIT960 system. The rifoligotyping assay correctly identified 90 of the 97 RIF-resistant isolates (sensitivity 92.8 %) while all the RFP-susceptible isolates were also correctly identified.

Real-time PCR techniques

Real-time PCR techniques have also been introduced recently for the rapid detection of MDR clinical isolates. Different probes have been used for detection, such as the TaqMan probe, Fluorescence Resonance Energy Transfer (FRET) probes, molecular beacons and bioprobes (Shamputa et al., 2004). Real-time PCR techniques have been applied to *M. tuberculosis* strains and directly to clinical samples (Ruiz et al., 2004; Sajduda et al., 2004; Espasa et al., 2005). Results are generally obtained in an average of 1.5-2.0 hours after DNA extraction.

Microarrays

Microarrays, also known as biochips or DNA chips, are based on the hybridization of DNA obtained from clinical samples to oligonucleotides immobilized on a solid support, such as miniaturized glass slides, which have been proposed as genotypic methods for detecting drug resistance in *M. tuberculosis*. They have been tested to detect resistance to INH and RFP (Gryadunov et al., 2005). The recently described CombiChip Mycobacteria Drug-Resistance detection DNA chip is an oligonucleotide microchip coupled with PCR for the detection of resistance to INH and RFP (Kim et al., 2006). It allowed identification of 84.1 % of INH-resistant isolates, based on the *katG* codon 315 and *inhA*15 mutations, and 100 % of RFP-resistant isolates based on seven codons: *rpoB*511, *rpoB*513, *rpoB*516, *rpoB*522, *rpoB*526, *rpoB*531, and *rpoB*533. The overall specificity was 100 % and 95.3 % for detecting INH and RFP resistance respectively.

1.4. Main methods of *M. Tuberculosis* Genotyping

Because of the characteristics of long latent *M. tuberculosis* infection, the infective mechanisms of TB could not be resolved by means of traditional epidemiological methods. Besides, the “stone in the pound” concept reaches its limit in a more globalize world where casual contacts may represent a driving force of TB transmission (Veen, 1992).

Modern molecular epidemiological methods have been developed to survey the outbreaks of infection at the pathogen molecular level, in order to track the spread of some *M. tuberculosis* strains and discover the mechanism of TB transmission (Alland et al., 1994; Small et al., 1994).

Genotyping of the *M. tuberculosis* strains now plays a key role in tracing the source of infection. The strains which possess the same genotype are called cluster strains and they are supposed to

come from a single progenitor, whereas the strains with unique genotypes are called idiosyncratic strains.

With the development of molecular biological technology, the genotyping methods have become the major tools in the epidemiological study of TB.

Since 1980, identification methods based on sequence polymorphism have gradually increased, for example, restriction fragment length polymorphism (RFLP), the analysis of a variety of "DNA fingerprint pattern" (as produced by amplified fragment length polymorphism (AFLP), pulsed field gel electrophoresis (PFGE), the hybridization of mobile elements on Southern blots, the DNA polymorphism of randomly-amplified-PCR (RAPD), DNA sequencing and gene chip, etc..

Both spacer oligotyping (Spoligotyping), and multiple loci variable number tandem repeat analysis (MLVA), developed more recently, are based on PCR. These methods are easy and fast.

The following paragraphs were modified basically from Barun Mathema et al "Molecular Epidemiology of TB: Current insights" (Mathema et al., 2006), Gilles Vergnaud, Christine Pourcel. Multiple Locus VNTR (Variable Number of Tandem Repeat) Analysis (Vergnaud and Pourcel, 2006), Karine Brudey, *et al*, *Mycobacterium* TB complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology (Brudey et al., 2006), and Anthony G. Tsolaki, *et al*, Genomic Deletions Classify the Beijing/W Strains as a Distinct Genetic Lineage of *Mycobacterium* TB (Tsolaki et al., 2005).

1.4.1. IS*6110*-RLFP

IS*6110*-RFLP is probably the best-known Southern blot hybridization approach. It takes advantage of the restriction fragment length polymorphism associated with the mobile genetic element called insertion sequence *6110* (IS*6110*). From most strain, a DNA fingerprint-like pattern can be produced, which most often shows a high discriminatory level. Patterns can be compared, so IS*6110*-RFLP typing has, for about a decade, been recommended as the gold standard DNA typing method (van Embden et al., 1993).

Insertion sequences (IS) are small mobile genetic elements, usually less than 2.5 kb in size, that are widely distributed in most bacterial genomes. IS elements in bacterial species are present in varying numbers of copies: IS1 in *Escherichia coli* strains is present in 2 to 17 copies, whereas the *Shigella* species contain from 2 to 40 copies (Chandler, 2002).

IS*6110* was first described as a 1,355-bp member of the IS3 family that, when intact, is so-far unique to the *M. tuberculosis* complex (Thierry et al., 1990). IS6110 has an imperfect 28-bp inverted repeat at its end and generates a 3-to 4-bp target duplication upon insertion. The number and distribution of IS*6110* elements within the genome of *M. tuberculosis* are different between strains. IS*6110* elements are more or less randomly distributed throughout the genome, with copy numbers ranging from rare strains lacking any IS*6110* elements to strains with 26 copies (Kurepina et al., 1998; McHugh and Gillespie, 1998).

IS*6110*-RFLP is a method based on the restriction fragment length polymorphism (RLFP) of IS*6110*. A standardized method for performing IS*6110*-based Southern blot hybridization analysis was proposed in 1993 (van Embden et al., 1993) based on the use of a common restriction endonuclease (PvuII, which cleaves IS*6110* at a single asymmetric site and yields reasonable-size *M. tuberculosis* chromosomal fragments), a hybridization probe (specific to the right side of IS6110, whereby each hybridizing band corresponds to a PvuII-PvuII chromosomal fragment with a single IS*6110* insertion), and standardized molecular weight markers (Hermans et al., 1990).

The applications of a software which assists in the analysis of IS*6110*-RFLP patterns has allowed for intra- and inter- laboratory comparisons of clinical isolates and the establishment of large national and international strain (and genotype) archives (Heersma et al., 1998; Kremer et al., 1999; Suffys et al., 2000). However, IS*6110*-RFLP genotyping method has some obvious limitations, for example, the limited resolution in analyzing clinical strains with six or fewer copies of IS*6110* (Bauer et al., 1999; Yang et al., 2000), the inability to distinguish among *M. tuberculosis* complex members and its labor intensiveness (Dale et al., 2001).

IS*6110* patterns have been used primarily to answer epidemiological questions, and their relationship in an evolutionary context had not been fully addressed (Cave et al., 1994). In addition, the two groupings of *M. tuberculosis* strains, i.e., low copy (≥five copies of IS*6110*) or high copy (> five copies IS*6110*), did not reveal clear evolutionary patterns. It was also unclear whether low copy number isolates represented a single evolutionary lineage, or whether they had evolved independently and demonstrated similar patterns due to transposition of IS*6110* into preferential integration sites (Fomukong et al., 1997).

1.4.2. Spacer oligonucleotide typing (Spoligotyping)

Studies show that *M. tuberculosis* complex strains contain a distinct chromosomal region consisting of multiple 36-bp direct repeats (DRs) interspersed by unique spacer DNA sequences (35 to 41 bp) also called DVR. Two forms of genetic rearrangements have been observed: one type consists of variation in one or a few discrete, contiguous repeats plus spacer sequences (DVRs), which is probably driven by homologous recombination between adjacent or distant chromosomal DRs; the other is driven by transposition of insertion sequence *6110* (IS*6110*), which is almost invariably present in the DR locus of *M. tuberculosis* complex strains (van Embden et al., 2000). The DR locus is a member of a larger family of repeats which have been designated as CRISPR (Jansen et al., 2002).

Spacer oligonucleotide typing (Spoligotyping) is a PCR-based technique for the detection of these interspersed spacer sequences in the genomic DR region of *M. tuberculosis* complex strains (Groenen et al., 1993; van Embden et al., 2000). Membranes spotted with synthetic oligonucleotides are hybridized with labeled PCR-amplified DR locus of the tested strain, resulting in a pattern that can be detected by chemiluminescence (Kamerbeek et al., 1997).

Spoligotyping method has many strongpoints as compared to the previously described techniques. It can be performed with considerably less DNA and in a fraction of the time; it also allows genotyping of thermolysates or impure DNA, nonviable specimens, paraffin-embedded material, and material from slides of Ziehl-Neelsen stainings (van der Zanden et al., 1998; Driscoll et al., 1999). The results are highly reproducible, and the binary (present/absent) data generated can be easily interpreted and computerized and are amenable to inter-laboratory comparisons. However, although Spoligotyping can be a powerful method to study the molecular epidemiology of *M. tuberculosis*, its discriminatory power in general is inferior to that afforded by IS*6110*-RFLP (Kremer et al., 1999).

1.4.3. Multiple Loci VNTR Analysis (MLVA)

Variable-number tandem repeats (VNTRs) consist in polymorphic head-to-tail direct uninterrupted repeats. They are microsatellites (1-to 8-bp repeats) and minisatellites (9-to 100-bp repeats) which are located in intergenic regions, in regulatory regions, or within open reading frames and are more or less abundant throughout bacterial genomes.

Multiple locus VNTRs analysis (MLVA) is a PCR-based molecular typing method based on VNTRs (Figure 8). It is very promising for the genotyping of *M. tuberculosis* isolates (Frothingham and Meeker-O'Connell, 1998; Le Flèche et al., 2001; Mazars et al., 2001; Smittipat et al., 2005; Supply et al., 2006).

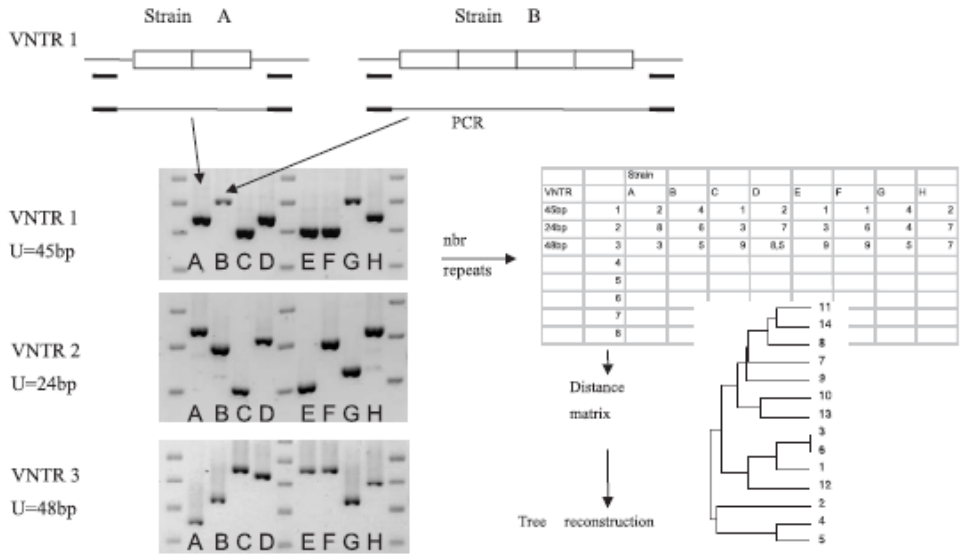


Figure 8 Schematic representation of a MLVA scheme. Primers are chosen on both sides of VNTR loci and PCR products are electrophoresed (here on agarose gel) together with size markers. The amplicon size is converted into a repeat number. Multiple markers are analyzed in the same way, a distance matrix is generated and a phylogenetic tree is produced

A systematic analysis of minisatellite VNTR loci in *M. tuberculosis* complex strains has been performed, allowing the description of more than 30 markers. The first analysis of genetic diversity using MLVA utilized the five so-called “major polymorphic tandem repeats” (MPTR) (A to E) (Hermans et al., 1992) and the six “exact tandem repeats” (ETR) (A to F) with repeat units ranging in size from 53 to 79 bp (Frothingham and Meeker-O'Connell, 1998). Since then, additional VNTR loci have been reported (Goyal et al., 1994; Magdalena et al., 1998; Namwat et al., 1998; Smittipat and Palittapongarnpim, 2000; Supply et al., 2000; Le Flèche et al., 2001; Skuce et al., 2002; Kovalev et al., 2005). MLVA has been conducted in a number of molecular epidemiologic studies, as well as to elucidate the phylogenetic relationships of clinical isolates (Sola et al., 2003; Fabre et al.; Kremer et al., 2005b; Vergnaud and Pourcel, 2006). VNTR analysis has also been used to evaluate *M. bovis* polymorphism (Roring et al., 2004). A comparative study of genotyping methods

which evaluated novel PCR-based typing techniques found MLVA analysis to have the greatest discriminatory power among amplification-based approaches (Kremer et al., 2005a; Vergnaud and Pourcel, 2006).

The discriminatory power of MLVA and the significance of the clustering achieved is related to the number and selection of loci which are used. A frequently used assay based upon a family of sequence-related tandem repeats called MIRUs comprises 12 loci. The genesis of some of these repeats in *M. tuberculosis*, which have highly similar sequences, may have started with an initial 53-bp single copy repeat, which then spread to different loci throughout the genome by recombination.This assay is less discriminating relative than IS*6110*-RFLP genotyping for isolates with high-copy-number IS*6110* insertions but more discriminating than IS*6110*-RFLP genotyping for isolates with low-copy-number IS*6110.* When additional loci are used, or MLVA is combined with Spoligotyping, the discriminatory power exceeds that of IS*6110*-RFLP. Variability at specific VNTR loci often depends on the sample collection (e.g., nationwide, population based, or convenience sampling), geographic origin, and inherent genetic diversity of the strains. For example VNTR2059 alias MIRU20 has been found to be polymorphic in some studies but not in others (Smittipat and Palittapongarnpim, 2000; Cowan et al., 2002).

MLVA is still quite new. So far, no official standard has been accepted for any of the bacteria species analyzed by this method (Vergnaud and Pourcel, 2006). One *M. tuberculosis* genotyping scheme was suggested to become “universal” in 2006 (Supply et al., 2006). Whether this scheme can indeed rapidly become a “golden standard” and replace IS*6110* is currently under investigation. One reason for this is that the MLVA assay is still in the development phase, in terms of the number of markers and strains tested, but it is very likely that, in the coming years, a standard will emerge. Another reason is that the resolution of a MLVA assay can be increased by adding markers (Figure 9), but requirements in terms of resolution depend upon the epidemiological question being asked, so that it is likely that different subsets of markers will eventually be used.

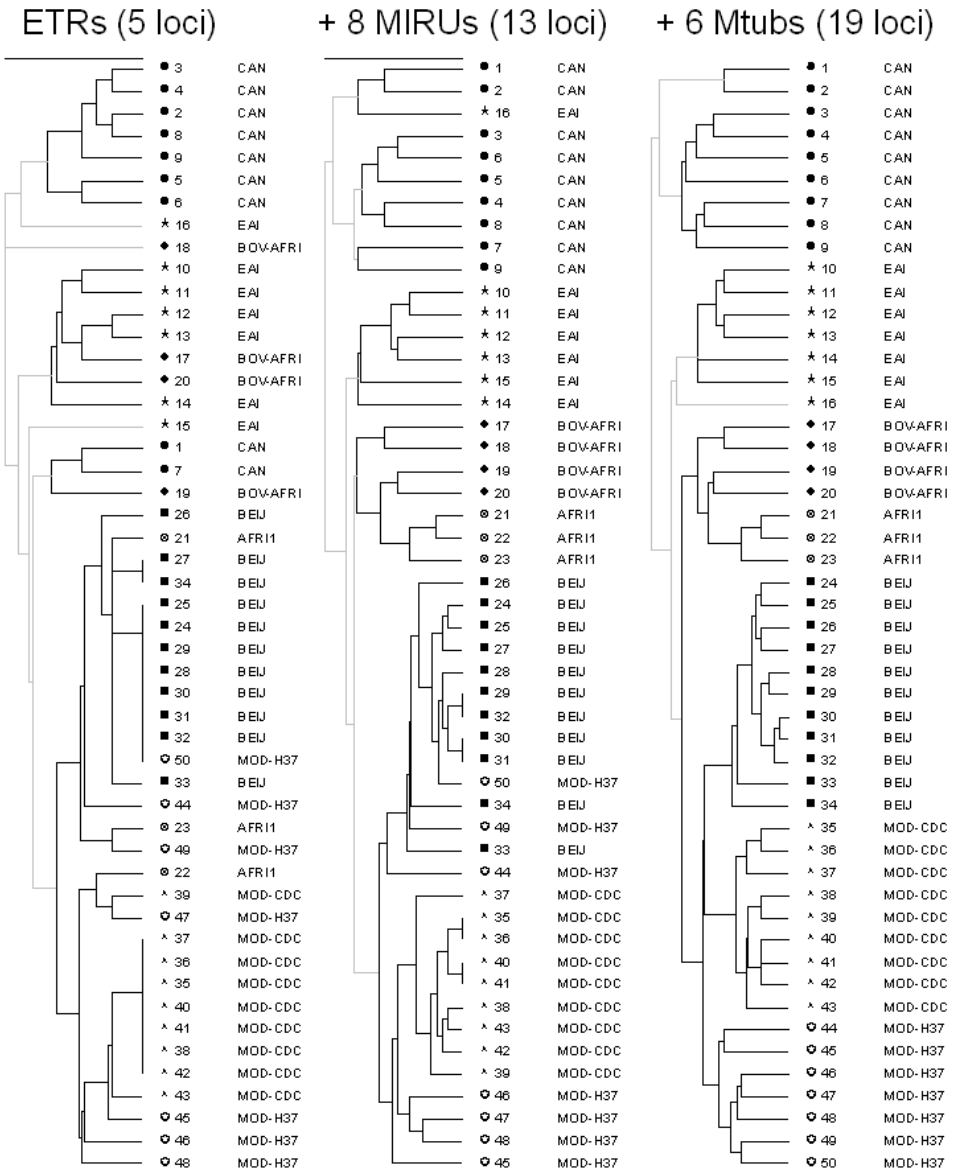


Figure 9 Comparison of the discrimination power of a MLVA analysis with 5, 13 or 19 VNTR

Acollection of 50 strains from the *M. tuberculosis* complex (MTBC) was typed using 19 markers; and clustering trees were produced using the data from either five markers (ETRs, left panel), 13 markers (5 ETRs + 8 MIRUs, middle panel) or 19 markers (5 ETRs + 8 MIRUs + 6 Mtubs, right panel). The strains were independently assigned to a MTBC group by classic biochemical assays and microdeletion typing (codes: CAN "M. canettii" strains, EAI East Africa/India, BOV-AFRI *M. bovis* and some *M. africanum* strains, AFRI1 the rest of *M. africanum* type 1 strains, BEIJ Beijing strains, MOD-CDC the group of modern *M. tuberculosis* strains, including the reference CDC1551 strain, MOD-H37 the group of modern *M. tuberculosis* strains, including the reference H37Rv strain). When all 19 markers are used in the analysis, 50 genotypes are identified (numbered from 1 to 50). The clustering fits with the independent and currently accepted

classification and phylogeny. When 13 markers are used, the discrimination is slightly reduced (46 different genotypes identified). The clustering achieved is still reasonable, with a few inconsistencies: genotype 16 (EAI strain) is grouped with "*M. canettii*" strains and three genotypes from modern *M. tuberculosis* strains are incorrectly assigned (genotypes 44, 49, 40) to the Beijing group of strains. When only five markers are used (left panel), 36 different genotypes are identified, which is still relatively high, but the clustering achieved is of little if any phylogenetic value.

There is evidence that this may depend on the genetic family to which the isolates being examined belong (Arnold et al., 2006).

1.4.4. single-nucleotide polymorphisms

Comparative genomic analysis of *M. tuberculosis* has revealed remarkable DNA conservation between chromosomes and noted genetic polymorphisms at the nucleotide level (Musser, 1995). The single-nucleotide polymorphisms (SNP) provide researchers with markers to differentiate clinical isolates as well as to study the phylogenetic relatedness of *M. tuberculosis* clinical strains. SNP (both nonsynonymous single-nucleotide polymorphisms [nsSNP] and synonymous SNP [sSNP]) provides useful genetic information that can be applied to differentiate *M. tuberculosis* strains (Sreevatsan et al., 1997).

In general, nsSNP polymorphisms produce an amino acid change that might be subject to internal or external selection pressure. As such, nsSNP changes in drug resistance-determining genetic loci can result in phenotypic drug resistance. Accordingly, *M. tuberculosis* resistance to anti-TB agents nearly always correlates with genetic alterations (nsNSP point mutations, small duplications, or deletions) in resistance-conferring chromosomal regions (Ramaswamy and Musser, 1998; Rengarajan et al., 2004; Maus et al., 2005; Zhang et al., 2005). nsSNP in genes that confer drug resistance can aid in understanding the nature and spread of resistance between and within populations.

In contrast, synonymous changes, which are considered functionally neutral, do not alter the amino acid profile. These neutral alterations, when in structural or housekeeping genes, can provide the basis to study genetic drift and evolutionary relationships among mycobacterial strains. Some investigations have used sSNP analysis to infer the phylogenetic structure of *M. tuberculosis* populations and have largely reported consistent findings (Gutacker et al., 2002; Baker et al., 2004;

Filliol et al., 2006; Gutacker et al., 2006a). In addition SNP analysis is amenable to targeting multiple polymorphisms that are informative in one platform, such as phylogenetic grouping, drug resistance, virulence, and other epidemiologically instructive markers.

1.4.5. Genomic deletion analysis

Since the complete annotated genome of *M. tuberculosis* laboratory strain H37Rv was published in 1998, and then that of the CDC 1551 clinical strain in 2002, the TB research community entered the genomic and comparative genomic era (Cole et al., 1998; Fleischmann et al., 2002; Garnier et al., 2003; Marmiesse et al., 2004; Brosch et al., 2007).

Large-sequence polymorphisms (LSP) in addition to SNPs have been revealed with comparative genomic analysis (Fleischmann et al., 2002). A study by Brosch et al. reported that LSPs occur as a result of genomic deletions and rearrangements rather than through recombination following horizontal transfer (Brosch et al., 2001).

Tsolaki et al. reported that up to 4.2% of the entire genome can be deleted in clinical isolates compared to the genome of laboratory strain H37Rv (Tsolaki et al., 2004). In the absence of horizontal gene transfer and because they are irreversible and often unique events, deletions have been proposed for genotyping as well as for constructing phylogenies (Brosch et al., 2002; Goguet de la Salmoniere et al., 2004; Tsolaki et al., 2004).

If data from the evolutionarily informative markers described above (IS*6110*-RLPF, spoligotype, deletion analysis and VNTR data) are combined for a large collection of diverse strains, an evolutionary lineage can be created.

1.5. Application of Genetic Markers for the phylogeny of *M. tuberculosis*

The ancestor of the MTBC is unknown, as, by contrast with other pathogenic mycobacteria, no environmental source has been found for this human pathogen. The first genetic investigations showed that the MTBC is highly clonal and it was suggested that the clonality resulted from the existence of a bottleneck some 40.000 years ago. Another possibility is that all the members of the MTBC derived from a bacteria which might be related to members of the smooth group of *M.*

tuberculosis but had acquired a highly significant selective advantage, such as the capacity to be transmitted from humans to humans. Indeed recent investigations demonstrate that the genetic heterogeneity (not taking into account deletions) within *M. canettii* is much larger than within the whole MTBC (Fabre et al., 2004; Gutierrez et al., 2005). There is no evidence of inter-human transmission of *M. canettii* which probably explain why this bacteria has not spread out of Africa and which suggests the existence of an environmental or animal reservoir. So far probably less than 100 *M. canettii* isolates have been recovered from patients almost exclusively in the Horn of Africa.

Different genetic markers indicate that the species *M. tuberculosis* consists of very distinct strain families, each with a single ancestor. An analysis of genetic marker deviation from the common ancestor of each of these families will improve understanding of the individual marker evolution rates and of the degree of convergence within genetic families.

1.5.1. The major genetic groups based on mutations at *kat*G and *gyr*A

In 1997 Sreevatsan *et al*. described nucleotide polymorphisms in two genes, *kat*G and *gyr*A, and showed that members of the *M. tuberculosis* complex could be separated in 3 major groups according to a few mutations in these genes (Sreevatsan et al., 1997). These genetic markers are highly valuable to anchor strains in phylogenetic analyses (Figure 10).

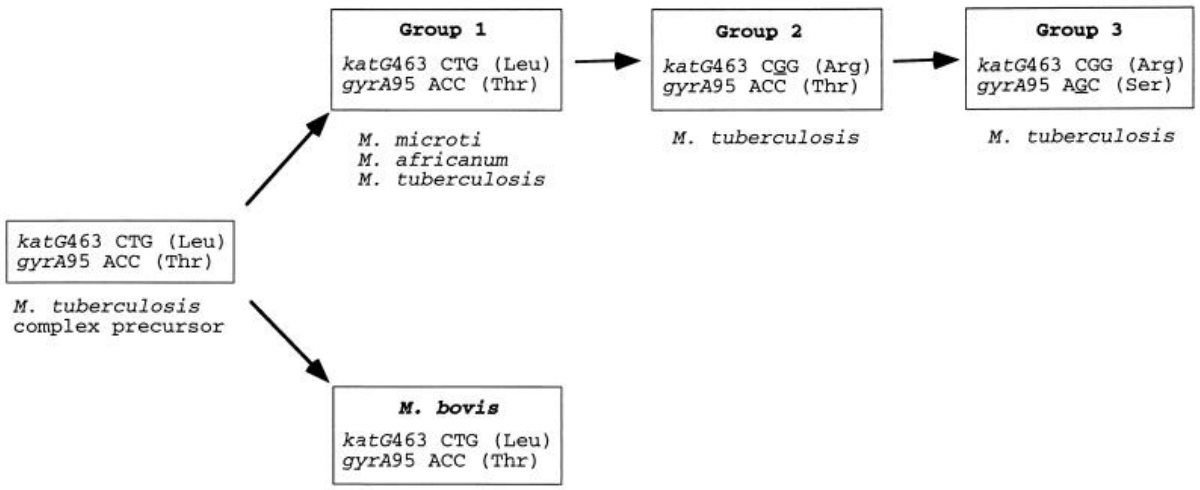


Figure 10: Broad evolutionary scenario for *M. tuberculosis* complex organisms. The precursor of *M. tuberculosis* complex organisms was characterized by KatG codon 463 (Leu) and GyrA codon 95 (Thr).

1.5.2. Deletion analyses

LSPs have been described in numerous studies of organisms from the *M. tuberculosis* complex, giving important insights into their evolution and biology (Brosch et al., 2002; Mostowy et al., 2002; Hirsh et al., 2004; Marmiesse et al., 2004; Tsolaki et al., 2004).

Successive loss of DNA was identified in an evolutionary lineage represented by *M. africanum*, *M. microti* and *M. bovis* strains that diverged from the progenitor of today's *M. tuberculosis* strains before the deletion of TbD1 occurred (Figure 11).

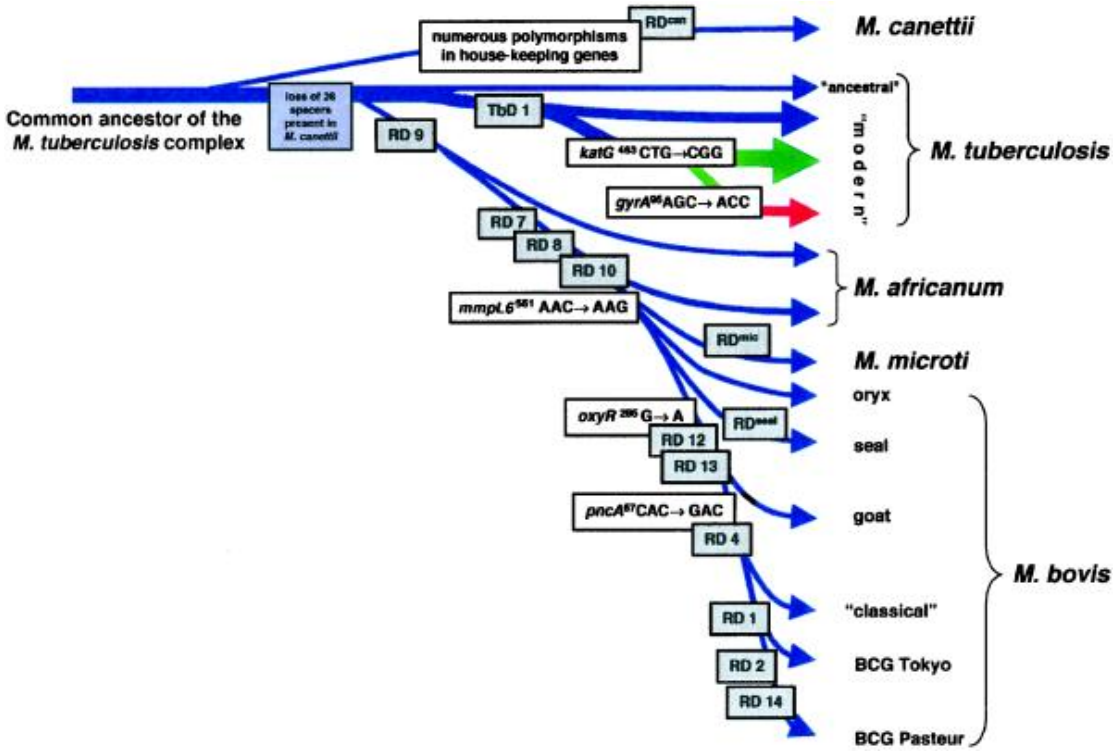


Figure 11 Scheme of the proposed evolutionary pathway of the tubercle bacilli illustrating successive loss of DNA in certain lineages (gray boxes). The scheme is based on the presence or absence of deleted regions and on sequence polymorphisms in five selected genes. Note that the distances between certain branches may not correspond to actual phylogenetic differences calculated by other methods. Blue arrows indicate that strains are characterized by *katG*$^{463.}$ CTG (Leu), *gyrA*^{95} ACC (Thr), defining "group 1" organisms. Green arrows indicate that strains belong to "group 2" characterized by *katG*^{463} CGG (Arg), *gyrA*^{95} ACC (Thr). The red arrow correspond to "group 3", characterized by *katG*^{463} CGG (Arg), *gyrA*^{95} AGC (Ser), as defined by Sreevatsan *et al.*

Until recently, it was assumed that cattle originally transmitted the disease to man, as the host range of *M. bovis* was much broader than that of *M. tuberculosis*. The findings of Brosch *et al.* argue against this hypothesis as both *M. canettii* and a subset of *M. tuberculosis* strains showed no loss of the specific regions identified in the study, and therefore seem to be direct descendants of tubercle bacilli that existed before the *M. africanum / M. bovis* lineage separated from the *M. tuberculosis* lineage (Brosch et al., 2002).

Restriction digested bacterial artificial chromosome (BAC) arrays of *M. tuberculosis* H37Rv were used in hybridization experiments with radiolabelled *M. bovis* BCG genomic DNA to reveal the presence of 10 deletions (RD1–RD10) relative to *M. tuberculosis*. RD1, a 9.5-kb DNA segment found to be deleted from all BCG substrains, was conserved in all virulent laboratory and clinical isolates of *M. bovis* and *M. tuberculosis* tested. The 9 open-reading frames predicted within RD1 may encode a novel secretion system. The reintroduction of RD1 into BCG repressed the expression of at least 10 proteins, including the TB-specific immunodominant antigens ESAT-6 and CFP-10, and resulted in a protein expression profile almost identical to that of virulent *M. bovis* and *M. tuberculosis*, as determined by two dimensional gel electrophoresis. These findings may be applicable to the rational design of a new attenuated TB vaccine and the development of new diagnostic tests to distinguish BCG vaccination from TB infection (Mahairas et al., 1996; Gordon et al., 1999; Lewis et al., 2003; Guinn et al., 2004).

Molecular epidemiological studies have identified a genetically related group of *M. tuberculosis* strains called Beijing/W that are widespread in many regions of the world (van Soolingen et al., 1995; Bifani et al., 2002; Glynn et al., 2002) (See paragraph 1.5.4 for detailed descRFPtion of this family). To enhance the understanding of their origin and evolution, Tsolaki and colleagues sought phylogenetically informative LSPs within the Beijing/W family. Comparative whole-genome hybridization of Beijing/W strains revealed 21 LSPs, 7 of which were previously unreported. They showed that some of these LSPs are unique event polymorphisms that can be used to define and subdivide the Beijing/W family. One LSP (RD105) was seen in all Beijing/W strains and thus serves as a useful additional marker for the identification of this family of strains. Additional LSPs (RD142, RD150, and RD181) further divided this family into four monophyletic subgroups, demonstrating a deeper population structure than could be previously appreciated. All Beijing/W strains were also observed to have an intact pks15/1 gene that is involved in the biosynthesis of a phenolic glycolipid, a putative virulence factor. A simple PCR

assay using these Beijing/W strain-defining deletions will facilitate molecular epidemiological studies and may assist in the identification of the molecular basis of phenotypes associated with this important lineage of *M. tuberculosis* (Figure 12).

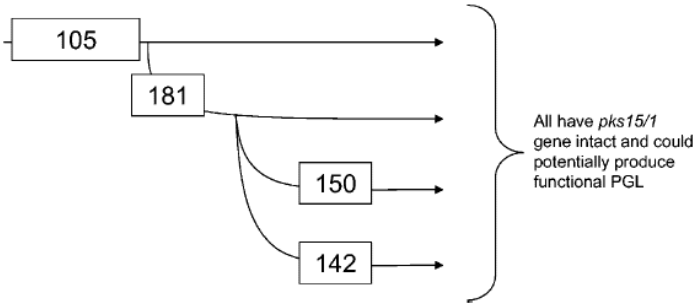


Figure 12 Beijing/W family of *M. tuberculo*sis is monophyletic. The phylogeny shows that RD105 defines the Beijing/W family and is further subdivided into four groups by RD181, RD150, and RD142. All of the Beijing/W strains contain an intact pks15/1 gene and may produce functional PGL.

1.5.3. Definition of spoligotype signatures for the identification of clades

Spoligotyping data can be used to elucidate evolution inside the *M. tuberculosis* complex, based on the assumption that the spacers in the DR region analysed can be lost (together or in groups) by homologous recombination between repeated elements or following movement of the IS*6110* element, and cannot be regained as little or no recombination appears to take place between strains (van Embden et al., 2000) and there appears to be no rearrangement within the DR (i.e., the DVRs are always in the same order). Genetic convergence has been demonstrated, and, although thought to be rare, this possibility should be borne in mind when using spoligotyping for evolution studies (Brudey et al., 2006).

The first evidence that members of the MTBC could be distributed into distinct families (also refered to as clades or families) with a precise geographical distribution (phylogeography) came from the comparison and clustering of spoligotypes. Because spoligotyping is easy, robust and cheap, it is widely used in clinical laboratory, for molecular epidemiology, evolution and population genetics. Thousands of spoligotypes have been produced worldwide and they have

been gathered and stored into databases thanks to the coordinating work of C. Sola and N. Rastogi, Pasteur Institute, Guadeloupe, France.

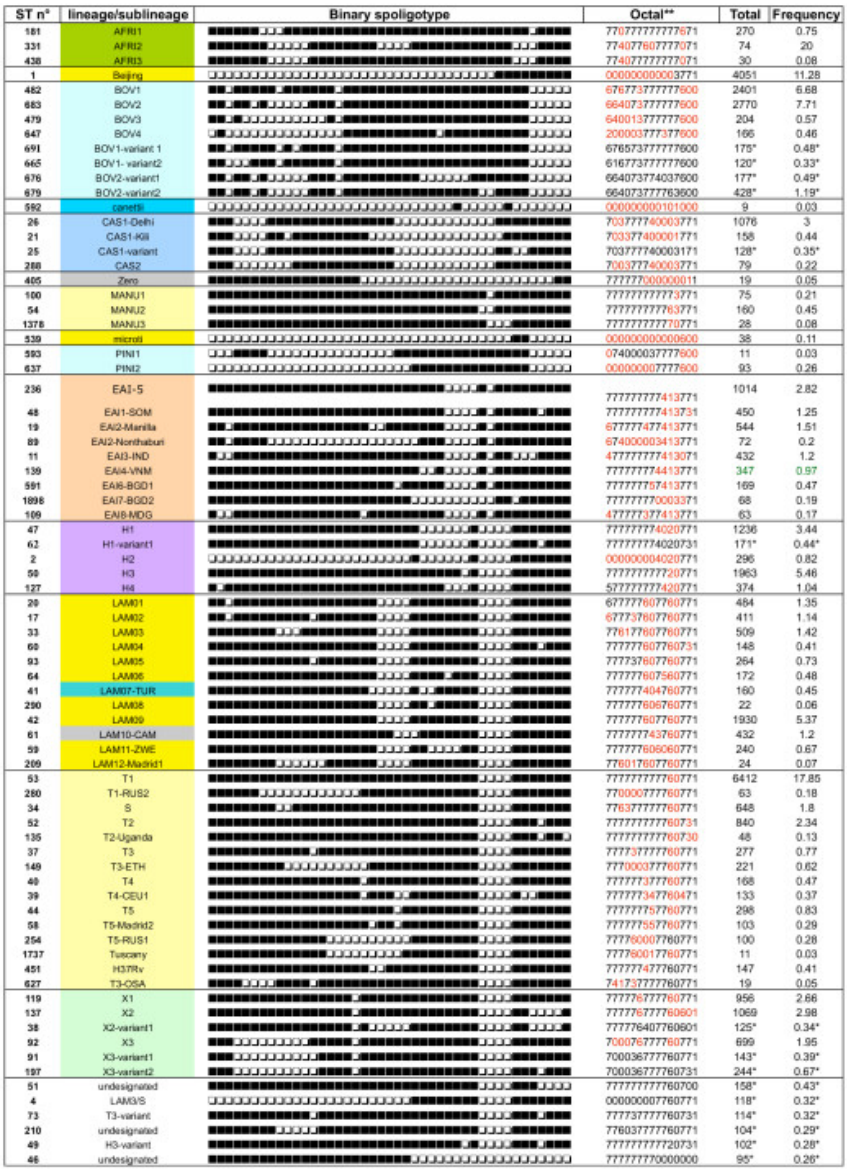
Four successive international spoligotype databases have revealed the global and local geographical structures of MTC bacilli populations. The fourth database, SpolDB4 (Brudey et al., 2006), describes 1939 shared-types (STs) representative of a total of 39,295 strains from 122 countries, which are tentatively classified into 62 clades/lineages using a mixed expert-based and bioinformatical approach (Figure 13). The results broaden the knowledge of the global phylogeography of the MTBC. The families have been named usually from their geographical origin. There are less than 10 clades that are easily recognized by a specific signature. Among them two families are frequently found in Asia, the EAI, which stands for East African Indian, and the CAS which stands for Central Asian.

All these investigations lead to the current view that the presence or absence of the *M. tuberculosis*-specific deletion TbD1, divides strains into so-called ‘ancestral’ and ‘modern’. In the *M. tuberculosis* complex, the East African Indian (EAI) lineage similarly to the *M. Africanum* and *M. bovis* subspecies possess the TbD1 region whereas the other large clades defined by spoligotyping e.g., the Central Asian (CAS), the Latin-American and Mediterranean (LAM), the Beijing and Haarlem clusters are all deleted.

Studies by Gutacker *et al.* (Gutacker et al., 2006b) using single-nucleotide polymorphisms confirm the existence of genetic clusters which are geographically restricted and correspond to the spoligotype clades.

By means of today's observed pattern of phylogeographical diversity of MTBC as the result of both a deep ecological differentiation and of a more recent demographic and epidemic history, there are some speculations about TB lineages. The publication of the recent studies done on *M. canettii* (Fabre et al., 2004; Gutierrez et al., 2005), suggests that TB is as old as humanity and probably appeared in the Horn of Africa. The EAI ancestral strains may have spread back from Asia to Africa through India concomitantly to human migrations (Cruciani et al., 2002; Gutierrez et al., 2005). Later-on evolution gave rise to the CAS lineage, and possibly to all "modern" TB lineages.

SpolDB4 is a very useful tool to better define the identity of a given MTBC clinical isolate, and to better analyze the links between its current spreading and previous evolutionary history.



ST n°	lineage/sublineage	Binary spoligotype	Octal**	Total	Frequency
181	AFRI1		770777777777671	270	0.75
331	AFRI2		774077607777071	74	20
438	AFRI3		774077777777071	30	0.08
1	Beijing		000000000003771	4051	11.28
482	BOV1		676773777777600	2401	6.68
683	BOV2		664073777777600	2770	7.71
479	BOV3		640013777777600	204	0.57
647	BOV4		200003777377600	166	0.46
691	BOV1-variant 1		676573777777600	175*	0.48*
665	BOV1- variant2		616773777777600	120*	0.33*
676	BOV2-variant1		664073774037600	177*	0.49*
679	BOV2-variant2		664073777763600	428*	1.19*
592	canetti		000000000101000	9	0.03
26	CAS1-Delhi		703777740003771	1076	3
21	CAS1-Kili		703377400001771	158	0.44
25	CAS1-variant		703777740003171	128*	0.35*
288	CAS2		700377740003771	79	0.22
405	Zero		777777000000011	19	0.05
100	MANU1		777777777773771	75	0.21
54	MANU2		777777777763771	160	0.45
1378	MANU3		777777777770771	28	0.08
539	microti		000000000000600	38	0.11
593	PINI1		074000037777600	11	0.03
637	PINI2		000000007777600	93	0.26
236	EAI-5		777777777413771	1014	2.82
48	EAI1-SOM		777777777413731	450	1.25
19	EAI2-Manilla		677777747413771	544	1.51
89	EAI2-Nonthaburi		674000003413771	72	0.2
11	EAI3-IND		477777777413071	432	1.2
139	EAI4-VNM		777777774413771	347	0.97
591	EAI6-BGD1		777777757413771	169	0.47
1898	EAI7-BGD2		777777773003371	68	0.19
109	EAI8-MDG		477777377413771	63	0.17
47	H1		777777774020771	1236	3.44
62	H1-variant1		777777774020731	171*	0.44*
2	H2		000000004020771	296	0.82
59	H3		777777777720771	1963	5.46
127	H4		577777777420771	374	1.04
20	LAM01		677777607760771	484	1.35
17	LAM02		677737607760771	411	1.14
33	LAM03		776177607760771	509	1.42
60	LAM04		777777607760731	148	0.41
93	LAM05		777737607760771	264	0.73
64	LAM06		777777607560771	172	0.48
41	LAM07-TUR		777777740760771	160	0.45
290	LAM08		777777606760771	22	0.06
42	LAM09		777777607760771	1930	5.37
61	LAM10-CAM		777777743760771	432	1.2
59	LAM11-ZWE		777777606060771	240	0.67
209	LAM12-Madrid1		776017607760771	24	0.07
53	T1		777777777760771	6412	17.85
280	T1-RUS2		770000777760771	63	0.18
34	S		776377777760771	648	1.8
52	T2		777777777760731	840	2.34
135	T2-Uganda		777777777760730	48	0.13
37	T3		777737777760771	277	0.77
149	T3-ETH		777000377760771	221	0.62
40	T4		777777737760771	168	0.47
39	T4-CEU1		777777347760471	133	0.37
44	T5		777777757760771	298	0.83
58	T5-Madrid2		777777557760771	103	0.29
254	T5-RUS1		777760007760771	100	0.28
1737	Tuscany		777760017760771	11	0.03
451	H37Rv		777777477760771	147	0.41
627	T3-OSA		741737777760771	19	0.05
119	X1		777776777760771	956	2.66
137	X2		777776777760601	1069	2.98
38	X2-variant1		777776407760601	125*	0.34*
92	X3		700076777760771	699	1.95
91	X3-variant1		700036777760771	143*	0.39*
197	X3-variant2		700036777760731	244*	0.67*
51	undesignated		777777777760700	158*	0.43*
4	LAM3/S		000000007760771	118*	0.32*
73	T3-variant		777737777760731	114*	0.32*
210	undesignated		776037777760771	104*	0.29*
49	H3-variant		777777777720731	102*	0.28*
46	undesignated		777777770000000	95*	0.26*

Figure 13 Bioinformatical (62 lineages/sub lineages prototype patterns) and statistical (50 most frequent) classification analysis of SpolDB4. (From (Brudey et al., 2006)).

1.5.4. The W/Beijing family

In 1995 van Soolingen *et al* reported that the analysis of the population structure of *M. tuberculosis* strains from the People's Republic of China showed that the vast majority belong to a genetically closely related group called the "Beijing family" (van Soolingen et al., 1995). These strains shared the majority of their IS*6110* DNA-containing restriction fragments, and DNA polymorphism associated with other repetitive DNA elements, like the polymorphic GC-rich sequence and the direct repeat, was very limited. Strains of this family were also found to dominate in neighboring countries such as Mongolia, South Korea, and Thailand, whereas a low prevalence of such strains was observed in countries on other continents. These data indicated that strains of the Beijing family very succesfuly expanded from a single ancestor which had a selective advantage. Now many studies showed that the Beijing *M. tuberculosis* genotype is globally widespread, with the highest prevalence found in Asia and the territory of the former Soviet Union.

It has been speculated that long-term *M. bovis* BCG vaccination may be one of the selective forces implicated in the successful spread of the Beijing genotype. However, Qian *et al* reported a study using stored biopsy specimens, suggesting that the prevalence of the Beijing family was already 90% in the 1950s and has not changed over time (Qian et al., 1999). It appears to have spread and become established as the predominant *M. tuberculosis* genotype in much of East and South East Asia at some time in the past, but there is now little evidence of increase. Demographic factors may also be responsible for this dominance. Besides, the relative homogeneity of Asian populations, at the anthropological level are well known, and could also explain the homogeneity of TB strains in China, based on a co-evolution hypothesis of the host and the pathogen agent (Oota et al., 2002).

Currently, Beijing/W strains are principally identified by the number of spacers in the direct repeat (DR) region of the *M. tuberculosis* genome, namely, spoligotype S00034, or ST1, which is characterized by the deletion of spacers 1 to 34 (van Soolingen et al., 1995). In a global survey of spoligotypes, type S00034 was found to be the most common (Filliol et al., 2003).

Members of the W-Beijing strain family have also been identified and classified by using the unique IS*6110* insertion site (A1) in the intergenic region in the origin of chromosome replication (oriC) as a marker (Kurepina et al., 1998; Kurepina et al., 2005). A specific assay to detect the W variant of the Beijing family had been developed by Plikaytis *et al.* using a multiplex PCR assay which targets a direct repeat of IS*6110* with a 556-bp intervening sequence (NTF-1) (Plikaytis et al.,

1994). Later, Beijing strains were subdivided into ancient and modern sublineages based on the analysis of the NTF-1 locus (Mokrousov et al., 2005). In a study by Mokrousov *et al.*, when strains were subjected to an IS6*110*-based inverse PCR assay, Beijing genotype strains were found to share identical “two-band” (ancient sublineage) or “three-band” (modern sublineage) profiles that were easily recognizable and distinct from the profiles of the non-Beijing strains (Mokrousov et al., 2006).

One LSP (RD105) was seen in all Beijing/W strains and thus serves as a useful marker for the identification of this family of strains. Additional LSPs (RD142, RD150, and RD181) further divide this family into monophyletic subgroups (Tsolaki et al., 2005).

Different studies suggest that the strains of the Beijing family might show a higher virulence. The bacterial genetic factors underlying this virulence are unknown, but the genetic relatedness of these Beijing/W strains suggests clonal expansion (Lopez et al., 2003; Rad et al., 2003; Smith et al., 2003). Recently, it was shown that the Beijing lineage may alter the innate immunological response of the host and overproduces triglycerides, through constitutive upregulation of the DosR regulon (Reed et al., 2004; Reed et al., 2007).

It was suggested that the Beijing genotype strains are associated with a higher frequency of drug resistance, including multidrug resistance, and are more transmissible, by an investigation on the relationships between *M. tuberculosis* genotypes and drug-resistant phenotypes and analysis of the transmission of drug-resistant TB (Sun et al., 2007). The role of mutator genes in the high frequency of drug-resistant variants in this lineage has been suggested by Rad *et al.* (Rad et al., 2003).

Nikolayevskyy *et al*. analysed the basis for drug resistance in Southern Ukraine. They used a non-commercial reverse hybridisation assay and DNA sequencing to detect mutations associated with rifampicin and isoniazid resistance (Nikolayevskyy et al., 2007). Genotyping was performed using an MLVA assay and spoligotyping. Mutations conferring rifampicin and isoniazid resistance were detected in 32.9% and 44.0%, respectively, of 225 *M. tuberculosis* isolates from individual consecutive patients. The prevalence of mutations was higher in Beijing strains sharing the VNTR signature 223325173533424 than in other Beijing strains (71.4% vs. 45.7%; RR 1.74; 95% CI 1.17-2.57), suggesting that this group may be responsible for rapid transmission of MDR TB in the southern Ukraine.

1.6. The transmission of drug-resistant strains

The emergence of MDR-TB is one of the main causes by which the epidemiology situation of TB has worsened sharply. Over the past decade, much has been learned of the drug targets and mechanisms of resistance to first-line and several second-line antiTB agents (Ramaswamy and Musser, 1998; Rengarajan et al., 2004; Maus et al., 2005). TB can usually be treated with a course of four standard, or first-line, anti-TB drugs. If these drugs are misused or mismanaged, MDR-TB can develop. MDR-TB takes longer to treat with second-line drugs, which are also more expensive and have more side-effects. XDR-TB can develop when these second-line drugs are also misused or mismanaged and therefore also become ineffective. Because XDR-TB is resistant to first- and second-line drugs, treatment options are seriously limited. It is therefore vital that TB control is managed properly (WHO, 2007).

Genotyping of multiresistant strains is performed to identify families that become resistant with a high frequency and to perform longitudinal studies in treated patients.

Genetic markers which are used in molecular epidemiological study can contain some “unique” characteristic that can distinguish target isolates, including unique fragment sequences, duplications, deletions, neutral SNP or polymorphisms associated with IS*6110*, or a drug resistance phenotype. With development of modern molecular techniques, rapid identification of TB transmission is greatly facilitated when strain-specific properties are targeted, as in the case of MDR-TB outbreaks (Palomino, 2006).

1.7. Molecular Epidemiology study of TB in China

TB represents a severe public health threat in the People’s Republic of China. China is one of the 22 countries representing 80% of the world’s burden of TB, registering 17% of global cases. The incidence rate is 113 per 100,000 population, and multidrug resistance (MDR, defined as resistance to at least isoniazid [INH] and rifampicin [RFP]) occurs in 5.3% of new TB cases. Treatment success under the DOTS strategy is high (96%), but the detection rate is very low (27%) compared to the objectives of the MDG (70% detection) (Millenium development goals: 70% active TB case detection objective as defined by the United nations).

Since the first study by van Soolingen *et al*. in 1995 showing the prevalence of the Beijing family of TB strains in China (van Soolingen et al., 1995), a number of studies have been conducted to analyse the genetic diversity of *M. tuberculosis* in this country, and to develop molecular epidemiological investigations. Below, we describe some examples of these research works.

To evaluate the use of VNTR typing in discriminating drug-resistant *M. tuberculosis* strains of the Beijing genotype family, Kam *et al*. analysed 102 multidrug-resistant (MDR) clinical isolates and 253 randomly picked non-MDR isolates collected from 2000 to 2003 in Hong Kong, using 12-locus MIRU typing, spoligotyping, and IS*6110* RFLP typing (Kam et al., 2005). Spoligotyping showed that 243 (68.5%) of 355 isolates belonged to the Beijing family genotype. This MLVA12 typing showed lower discrimination in differentiating between the Beijing family strains (Hunter-Gaston diversity index [HGDI] of 0.8827) compared with the IS*6110*-RFLP method (HGDI = 0.9979). For non-Beijing strains, MIRU typing provided discrimination (HGDI = 0.9929) comparable to that of the RFLP method (HGDI = 0.9961). There was no remarkable difference in discrimination power between the two methods in differentiating both within and between MDR and non-MDR strains of *M. tuberculosis*. Dendrograms constructed with the MIRU typing data showed a clear segregation between the Beijing and non-Beijing genotype. Addition of RFLP to MIRU typing offered a higher discrimination ability (92.6%) than did addition of MIRU typing to RFLP (40.0%). This supported the potential use of this method to analyze the global genetic diversity of MDR *M. tuberculosis* strains that may be at different levels of evolutionary divergence.

Li *et al*. sought to determine whether patients who had therapy failure with increasingly drug-resistant strains of *M. tuberculosis* had primary or acquired drug resistance, by genotyping the initial and subsequent drug-resistant clinical isolates of *M. tuberculosis* collected from patients by the Shanghai Centers for Disease Control and Prevention over the course of a 5-year period (Li et al., 2007). The vast majority of patients (27/32) had primary drug resistance, indicating transmission of a drug-resistant strain of *M. tuberculosis*. Only 16% (5/32) had acquired drug resistance because of a poor treatment regimen or no adherence to an adequate regimen.

Chin and his colleagues combined mycobacterial interspersed repetitive unit (MIRU) loci 26, 31, and ETR-A for differentiation of Beijing lineage isolates in 1,557 clinical isolates (Chin et al., 2007). This clearly showed that a method for rapid identification of Beijing lineage *M. tuberculosis* strain was still needed in regions of TB endemicity, especially if genotyping methods were not readily accessible.

Ma *et al*. identified *rpo*B gene mutations and performed the molecular characterization of rifampicin-resistant *M. tuberculosis* isolates from Shandong Province, China (Ma et al., 2006). Sixty rifampicin (RIF)-resistant and 75 RIF-susceptible *M. tuberculosis* isolates were analyzed. MIRU genotype 223325173533 (a very common Beijing-characteristic pattern) was overrepresented among RIF-resistant isolates.

Genotypes were determined by spoligotyping and IS*6110*-RFLP to determine the contribution of the *M. tuberculosis* Beijing genotype to the population structure of *M. tuberculosis* complex in China, using 441 mycobacterial strains isolated from a nationwide random survey for the epidemiology of TB conducted in 2000, for which the drug susceptibility patterns and epidemiological data were known. A total of 408 samples were identified as *M. tuberculosis* by spoligotyping. Of the *M. tuberculosis* strains, 64.9% (265/408) were of the Beijing genotype. Using the 2 tests, no statistically significant differences were observed in the proportion of Beijing genotype TB in patients of different sex, age or living in different areas of the country, but a significant difference was observed with multidrug-resistant (MDR) TB: the Beijing genotype represents 77.8% (42/ 54) of MDR isolates. However in the multivariate logistic regression model, the Beijing genotype was significantly associated with the region of origin and not with MDR. It indicated that the *M. tuberculosis* Beijing genotype had prevailed in China for at least five decades before this report and was associated with the East Asian continental region (Li et al., 2005).

Recently Jiao *et al.* (Jiao et al., 2007) reported the molecular characteristics of RIF and INH resistant *M. tuberculosis* strains from Beijing, China. They found that 91.7% of total strains were of the Beijing genotype of which 23.4% were ancestral. There was no significant difference between ancestral and modern group in prevalence of drug resistance-associated gene mutations.

As shown in the examples above, although many studies of genotyping or molecular epidemiology of *M. tuberculosis* have provided greatly significant data for the TB control and showed that molecular techniques could be applied in China, there are many limitations in these studies, such as the lack of systematic and comprehensive investigation on molecular epidemiology in China and the distribution of Chinese *M. tuberculosis* genotypes all over the world. There remains a lack of efficient and applicable molecular techniques including new genetic markers which would be available for Chinese *M. tuberculosis* genotyping and for rapid diagnosis for TB investigation and surveillance

A preliminary study of the genetic diversity of Chinese *M. tuberculosis* strains based on MLVA was performed by our research group using eleven tandem repeat loci analyzed by PCR and agarose gel electrophoresis. Sixty five *M. tuberculosis* isolates from patients with TB in Anhui, Hunan and Jiangsu provinces were analysed. Finally the results of seven VNTR loci (ETR-A , ETR-B , ETR-C , ETR-D , ETR-E , ETR-F , MPTR-A) were chosen. The strains were classified into 10 genotypes according to their VNTR characteristics. Sixty-nine percent of them belong to two main types. The result indicates that there is obvious polymorphism in the *M. tuberculosis* isolates and implies that there perhaps exists some epidemic strain clusters in *M. tuberculosis* in China (Publication 2) (Liu et al., 2004).

1.8. Aim of the thesis project: organisation of a nationwide study of the *M. tuberculosis* drug resistance and genetic diversity in the People's Republic of China

In China the vaccination of babies is systematic since the years 1950s and represents an efficient protection for young children. However it does not totally prevent infection and at the age of 16 a large fraction of the population is infected. In addition it is known that a high percentage of strains resist to one or several antibiotics.

As explained in the Introduction it is believed that the W/Beijing lineage, presently one of the most dreadful lineages spreading worldwide, emerged in China. In addition, although it appears to be highly prevalent in certain parts of China, almost nothing is known on the characteristics of other lineages that may exist.

To get a clear idea of the diversity over the whole country, to detect emerging clones and to investigate the basis for antibiotic resistance, a statistically relevant sampling of *M. tuberculosis* population, that would includes thousands of isolates, should be genotyped and their antibiotic-resistance profile should be determined. Information on the patients should be gathered to better understand the possible relationship between genotype and the severity of the disease.

Our long-term plan in order to achieve such a study is to collect and analyse a minimum of 200 isolates in each of the 31 provinces, half from males and half from females, and if possible distributed into three ages groups: 0 to 16, 17 to 50 and more than 50 years old. For this we have

sought the collaboration of physicians in the different provinces (municipality, autonomous region) to isolate strains from suspected patients with TB attending Province's level TB hospital or Institutes for TB control. Background information on patients were collected and stored in databases and the bacterial isolates were characterized and stored for further analyses.

To perform this large project, a collaboration between China CDC in Beijing and the IGM in Université Paris Sud Orsay was started in 2003. This is in this context that the present thesis work was conducted, helped by the EEC contribution for exchanges between the two laboratories.

At present, data have been obtained for isolates from 12 provinces. In this manuscRFPt, two main studies will be described. The first one deals with the analysis of antibiotic resistance and the second with genetic diversity.

1.8.1 Analysis of antibiotic resistance

A total of 2018 clinical *M. tuberculosis* strainsfrom 10 provinces were tested for their susceptibility to 4 anti-TB drugs (isoniazid (INH), rifampicin (RFP), streptomycin (SM), and ethambutol (EMB)) by means of the absolute concentration method and the proportional drug susceptibility testing method. Then the mutation characteristics of the *rpo*B gene confering resistance to rifampicin and *kat*G, *pre-inh*A *inh*A, *ndh* and *oxy*R-*ahp*C genes confering resistance to isoniazid, were investigated by DNA sequence analysis and oligonucleotide probe reverse dot blot hybridization based on PCR.

1.8.2 Genetic diversity study

In order to select the techniques to be used in a large scale study of several thousand isolates a pilot study was first conducted on 127 isolates from 5 Chinese Eastern Provinces, Hunan, Jiangsu, Anhui, Fujian and Beijing. Genetic polymorphisms of the strains were determined using several molecular typing techniques: Spoligotyping which analyses the DR locus, LSP for the presence of genomic deletions and MLVA. The different techniques are PCR-based, they can be performed using simple DNA purification and equipments present in every molecular biology laboratories.

The result of the survey performed in 12 provinces and involving almost two thousands isolates is reported. For each province we sought to measure the percentage of Beijing-family strains, describe new families, and assess the relationship between genotype and BCG vaccination and drug-resistance.

2. METHODS

2.1. *M. tuberculosis* strains

2.1.1. Standard strain H37Rv

For the research works in National Institute for Communicable Disease Control and Prevention (ICDC), Chinese Center for Disease Control and Prevention (China CDC), the standard strain H37Rv is cultured, inoculated, conserved and provided by the National Laboratory of TB, State Key Laboratory for Infectious Diseases Prevention and Control (SKLID)/ ICDC, China CDC in Beijing, P. R. China.

2.1.2. Collection and Identification of Clinical strains in China

The study obtained approval from the Ethics Committee of National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Contro**l** and Prevention (**ANNEX 1**). The patients with TB included in the present research protocol were given a Subject information sheet (**ANNEX 2**) and were asked for a written informed consent (**ANNEX 3**).

Bacterium Isolation

The sputum samples were collected from the patients suspected of TB who went to the institutes for TB control and cure or TB hospitals in each province. The bacteria were isolated and inoculated on Löwenstein-Jensen medium (Rieder et al., 1998). Culture was performed for all samples and the bacteria were kept in the National Laboratory of TB, SKLID/ ICDC, China CDC in Beijing, P. R. China. The epidemiological and clinical information of patients were collected by means of a special epidemiological questionnaire table (**ANNEX 4**) for future epidemiological studies.

Biochemical identification

P-nitro-alpha-acetylamino-beta-hydroxypropiopheneone (NAP) and 2-thiophene-2- carboxylic acid hydrazide (TCH) tests were used for biochemical identification of the bacterial strains. The NAP test was used to differentiate *M. tuberculosis* complex isolates from non-tuberculous mycobacteria (NTM), and the TCH test to recognize *M. bovis*. (Table 5).

Table 5 Identification result of *M. tuberculosis*、*M. bovis*、non-*M. tuberculosis*

Species of Mycobacterium	PNB	TCH	L-J
M. tuberculosis	-	+	+
M. bovis	-	-	+
non-*M. tuberculosis*	+	+	+

+：positive −: negative

2.2. Preliminary drug-resistance research

2.2.1. The Drug sensitivity test

Clinical isolates of *M. tuberculosis* from patients with TB were collected in Xizang, Hunan, Henan, Sichuan, Fujian, Anhui, and Shanxi provinces of China. The absolute concentration method and the proportion method were used to test the susceptibility of the isolates to 4 anti-TB drugs (INH, SRM, RFP, EMB) (Rieder et al., 1998).

2.2.2. Primary study on the mutation characteristics of Rifampicin (RFP)- and Isoniazid (INH)-resistance relationship genes of *M. tuberculosis*

To investigate potential *rpo*B gene mutations in *M. tuberculosis*, and to identify the characteristics and the distribution of the mutations, clinical isolates of *M. tuberculosis* from the patients with TB in Xizang, Hunan, Henan, Sichuan, Fujian, Anhui, and Shanxi provinces of China were investigated.

According to the sequence of RFP- and INH-resistance related genes of *M. tuberculosis* in GeneBank (*rpo*B, *kat*G, *inh*A, ...) we have decided to sequence the DNA fragment in the resistance genes, including the “hot spot mutation area ”, when known. We also used the DNA Star software and designed the primers for PCR amplification using Oligo6.0 software.

Primary study on the mutation characteristics of the RFP-resistance gene *rpo*B.

To investigate *rpo*B gene mutations in *M. tuberculosis*, and identify the characteristics and the distribution of mutations, *M. tuberculosis* clinical isolates including RFP-resistance strains (RFP-R) and RFP-sensitive (RFP-S) strains were selected.

A 286bp DNA fragment of the *rpo*B gene including the 81bp hot-spot region (rifampicin resistance determination region, RRDR) was analyzed by PCR-SSCP (single strand conformation polymorphism). 110 strains of *M. tuberculosis*, including 73 RFP-R strains, 11 RFP-S strains resistant to other drugs and 26 drug susceptible strains, were studied. The primers were P1: 5'-CAG GAC GTG GAG GCG ATC ACA-3' and P2: 5'-CCG ACA GCG AGC CGA TCA GAC-3'.

Then 300 *M. tuberculosis* clinical isolates including 119 RFP-S strains and 181 RFP-R strains were selected to investigate the mutations in the *rpo*B gene. For this the *rpo*B gene fragment (629bp) including almost every mutation sites previously published was amplified by PCR and the mutations in the *rpo*B gene fragment were detected by DNA sequencing. Based on the sequencing results, the mutations were analyzed, including the rate of mutation of each site, and the kind of *rpo*B gene mutation in *M. tuberculosis*. Then we selected some mutation sites and 15 oligonucleotide probes were designed (Table 6). A reverse dot blot hybridization protocol (RDH) was established based on PCR to detect the mutations of the *rpo*B gene. The primers are YY1: 5'-GAG CCC CCG ACC AAA GA-3' and YY2: 5'-ATG TTG GGC CCC TCA GG-3'.

Table 6 DNA sequence of 15 oligonucleotide probes

Num	Kind of Mutation	DNA Sequence	Tm(℃)	Concentration (pmol/150ul)
W1	511 Leu	5’agccagctgagccaattcat3’	52.3	12.5
W2	516 Asp	5’ttcatggaccagaacaacccg3’	59.8	12.5
W3	522 Ser	5’gctgtcggggttgacc3’	54.0	100
W4	526 His	5’ttgacccacaagcgccga3’	58.0	50
W5	531 Ser	5’ctgtcggcgctggggc3’	58.0	100
M6	511 ctg-ccg	5’agccagccgagccaattcat3’	59.4	50
M7	516 gac-tac	5’ttcatgggccagaacaacc3’	58.0	50
M8	516 gac-ggc	5’ttcatgtaccagaacaacccg3’	52.9	50
M9	526 cac-gac	5’ttgaccgacaagcgccga3’	58.0	100
M10	526 cac-cgc	5’ttgacccgcaagcgccg3’	58.0	200
M11	526 cac-tac	5’ttgacctacaagcgccg3’	58.0	200
M12	531 tcg-ttg	5’ctgttggcgctggggc3’	56.0	50

Primary study on the mutation characteristics of INH-resistance.

M. tuberculosis clinical isolates including INH-R strains and INH-S strains were selected to investigate the mutations in the *kat*G gene, the pre-*inh*A gene, the *inh*A gene, the *ndh* gene and the *oxy*R-*ahp*C gene respectively. Primers were designed for the polymerase chain reaction (PCR) amplification of these genes (Table 7). Then the PCR and the DNA sequence analysis were carried out on the *M. tuberculosis* standard strain H37Rv and on 291 *M. tuberculosis* clinical isolates from patients for which background information was available. Among these strains 64 were sensitive to 4 anti-TB drugs (INH, SRM, RFP, EMB), 48 were sensitive only to INH but resistant to the other drugs (SRM or/and RFP or/and EMB), 30 were INH single drug-resistant, and 149 were multi drug-resistant including INH.

Table 7 The primer sequences for *kat*G, pre-*inh*A, *inh*A, *oxy*R-*ahp*C and *ndh* gene amplification

Gene	Primer sequence	Beginning site	Primer annealing (℃)	DNA fragment (bp)
*kat*G	L: gatcgtcggcggtcacactt R: cgttgacctcccacccgact	795 1525	60	731
pre-*inh*A	L: cgtcaatacacccgcagcc R: cccggtgaggttggcgttg	29 564	63	536
*inh*A1	L: cgatatgacccgcgcgctgga R: gcgccacgaacctgttgacc	3 730	63	728
*Oxy*R-*ahp*C	L: cgtctggtcgcgtaggcagt R: cgtctggtcgcgtaggcagt	263 962	62	700
ndh	L: gtaagcagcgcaatgtccag R: caggtcgggcagcacttgg	230 927	62	698

2.3. A survey of the genetic diversity in 12 Provinces

2.3.1. DNA Sample Preparation, Extraction from Bacterial Cells

The *M. tuberculosis* strains were cultured on Löwenstein-Jensen medium at 37℃ constant temperature incubator for 3 to 4 weeks. The bacterial colonies were collected. DNA extracts were

prepared by suspending approximately 10mg wet bacterial cells in 100µl of sterile distilled water and subsequently heating at 80°C to 100℃ for 30 minutes to kill and lyse the cells (van der Zanden et al., 2002). Cell debris were removed by centrifugation at 13,000g for 2 minutes. The lysates were stored at -20℃ until further use.

2.3.2. Spacer oligonucleotide typing

Spoligotyping involves the PCR amplification of the whole CRISPR region using the DR as a target (Figure 14), followed by hybridization of the amplified DNA to a set of spacer oligonucleotides, covalently linked to a membrane. Because clinical isolates vary in the nature of spacer sequences, whereas the order of spacers is preserved (van der Zanden et al., 2002), the spoligotype patterns obtained are strain specific.

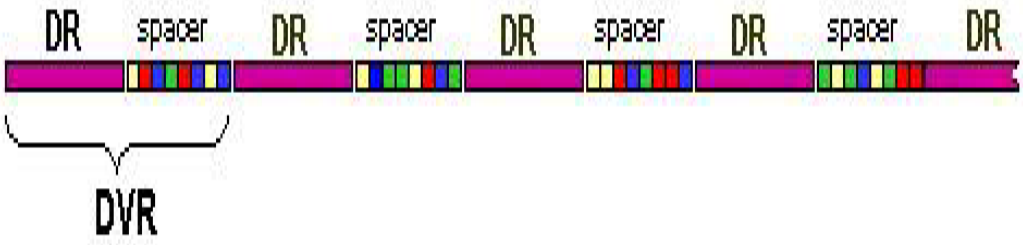


Figure 14 Scheme of the structure of a CRISPR locus (DR and Spacer location in the Bacterial DNA)

In this study, we used the classical spoligotyping method in which 43-spacer oligonucleotides are derived from the CRISPR region of *M. tuberculosis* reference strain H37Rv and BCG P3 (Kamerbeek et al., 1997) (Figure 15). Spoligotyping was performed according to the previously described method (Kamerbeek et al., 1997; van der Zanden et al., 2002).

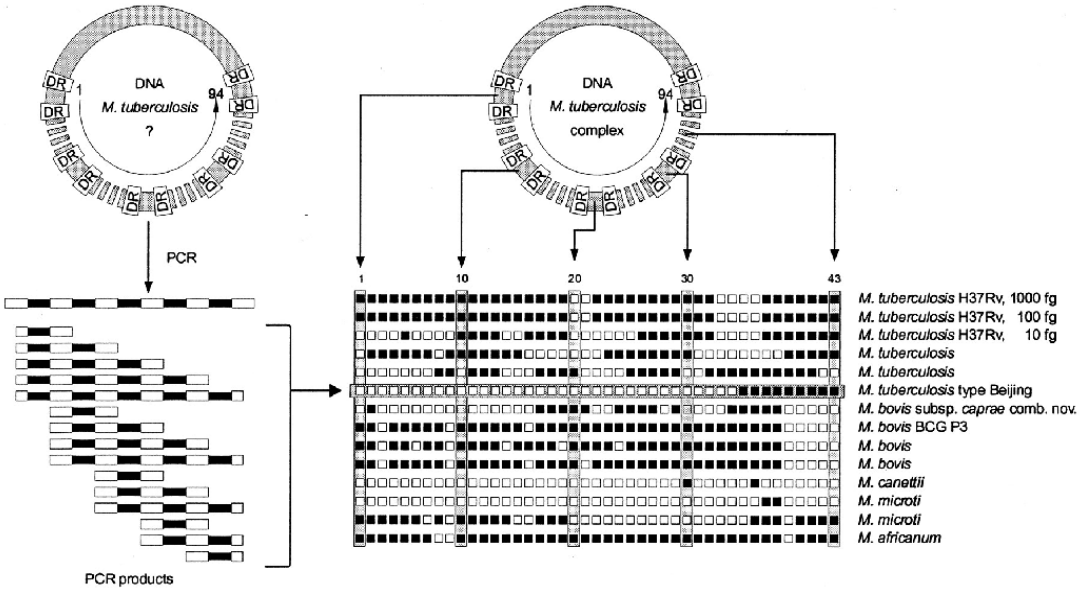


Figure 15 The DNA sequence of the CRISPR region, a synonym of the DR region, has been determined for the reference species *M. tuberculosis* H37Rv and M. bovis BCG P3. 43 oligonucleotides corresponding to the variable spacers have been synthesized, and introduced vertically onto the membrane used for traditional spoligotyping. The conserved DRs serve as targets for amplification of the entire locus from a species or samples containing a species of an unknown genotype and belonging to the *M. tuberculosis* complex. After hybridization and detection, the spoligotype pattern can be read from the horizontal axis of the various clinical isolates belonging to the *M. tuberculosis* complex, such as *M. tuberculosis* H37Rv and *M. bovis* BCG, *M. bovis, M. canettii, M. microti, M. Africanum*, and *M. bovis* subsp. *caprae* comb. Nov.

Ten microliters (μl) of the lysates obtained from the cultured *M. tuberculosis* strains was added to a PCR mixture which containinging 10μl dNTP(1mM), 5μl 10×Buffer, 3μl $MgCl_2$ (1.5mM), 3μl DRa (5μM), 3μl DRb (5μM), and 10μl DNA polymerase (3.0U), for 50μl total volume. The primers are DRa: Biotin-5'-GGT TTT GGG TCT GAC GAC-3' and DRb: 5'-CCG AGA GGG GAC GGA AAC-3'.

PCR amplification process: 3 minutes at 96°C for DNA denaturation; 35 cycles of: 1 minute at 96°C for DNA denaturation, 1 minute at 55°C for primer annealing and 30 second at 72°C for primer extension; followed by a last extension cycle of 6 minutes at 72°C. PCR products were kept at -20°C until further analysis.

Hybridization: PCR products were analyzed by hybridization using the reverse line blotting technique (Kaufhold et al., 1994) with 43 oligo probes (**ANNEX 5**). The hybridization conditions were 50℃, 45min (Kamerbeek et al., 1997).

Concerning washing condition, we used the operating procedures described in previous reports (van Soolingen et al., 1991; van Soolingen et al., 1993; Kaufhold et al., 1994). After hybridization, the membrane was washed three times in 250 ml of 23 SSPE–0.5% SDS for 10 min each time at 60℃ and then incubated in 1:4,000-diluted streptavidin-peroxidase conjugate (Boehringer) for 45 to 60 min at 42℃. The membrane was washed twice, for 10 min each time, in 250 ml of 2× SSPE–0.5% SDS at 42℃ and rinsed with 250 ml of 2× SSPE for 5 min at room temperature. For repeated use of membranes, the membranes were stRFPped by being washed two times for 30 min each time in 1% SDS at 80℃ and then incubated for 15 min in 20 nM EDTA (pH 8) at room temperature. Membranes were sealed in plastic and stored at 48°C until further use.

After hybridization and washing of the membrane, hybridized DNA was detected with peroxidase-labeled streptavidin as described previously (van der Zanden et al., 1998).

2.3.3. Genomic deletion analysis in strains of the Beijing family

PCR amplification

Two microliters (μl) of lysates obtained from the cultured *M. tuberculosis* strains were added to 13μl of PCR mixture which included 1.5μl dNTP(1mM), 1.5μl 10×Buffer (Including $MgCl_2$), 3μl Betain, 1.0μl Primers (Left and Right mixture), 0.05μl DNA polymerase (5U/μl), and 3μl distilled water, for a 15μl total volume. The primers were RD181Ext. / RD181Int. (Table 8)

Table 8 LSP Loci and primer sequences

Locus	Primer	Sequences of primer
RD181	Ext_L	CGCAACGGCCGCGGTGAACTCT
	Ext_R	CGGGCGGCTGCGGGAACCTT
RD181	Int_L	TAACAGCAGTGGGACCAAGC
	Int_R	GACTGCCGGTCTTAGTCTGC

PCR amplification process: 3 minutes at 94℃ for DNA denaturation; 35 cycles: 30 seconds at 94℃ for DNA denaturation, 1 minute at 62℃ for primer annealing and 30 second at 70℃ for primer extension; following by a last cycle of 10 minutes at 72℃ for primer extension. PCR products were analyzed by electrophoresis on 2% agarose gel.

2.3.4. Multiple Loci VNTR Analysis (MLVA)

VNTR loci

PCR amplification of 21 VNTRs (MLVA21) and electrophoresis of products on agarose gels were carried out as described in a previous report (Fabre et al., 2004) (Table 9). MLVA21 includes: ETR-A, ETR-B, ETR-C, ETR-D (MIRU04), ETR-E (MIRU31), MIRU02, MIRU10, MIRU16, MIRU23, MIRU26, MIRU27, MIRU39, MIRU40, Mtub01, Mtub02, Mtub12, Mtub21, Mtub29, Mtub30, Mtub39, Qub11a.

A MLVA scheme using 15VNTRs (MLVA 15) was performed using a subset of markers: ETR-A, ETR-B, ETR-C, ETR-D (MIRU04), ETR-E (MIRU31), MIRU10, MIRU16, MIRU23, MIRU26, MIRU27, MIRU39, MIRU40, Mtub21, Mtub30, Mtub39.

PCR amplification

Two microliters (μl) of the lysates obtained from the cultured *M. tuberculosis* strains were added to 13μl of PCR mixture which included 1.0μl dNTP (1mM), 1.5μl 10×Buffer, 3.0μl Betain (5M), 0.9μl $MgCl_2$ (1.5mM), 0.9μl primer L (5μM), 0.9μl primer R (5μM), 0.3μl DNA polymerase (3U), and 4.5μl distilled water, with a final 15μl total volume.

The PCR amplification process: 5 minutes at 94℃ for DNA denaturation; the following cycles were repeated 35 times: 30 seconds at 94℃ for DNA denaturation, 30 seconds at 62℃ for primer annealing and 45 second at 70℃ for primer extension; the last cycle included a 10 minutes step at 72℃ for final products extension. PCR products were analyzed by 2% agarose gel electrophoresis at once, or were kept at 4℃ until further analysis.

Table 9 Information of VNTR Loci in the MLVA21 scheme

Number	VNTR locus	H37Rv length	PCR product	Primers: L & R
1	H37Rv_2165_75bp_(**ETR-A**)	257	397	L)ATTTCGATCGGGATGTTGAT R)TCGGTCCCATCACCTTCTTA
2	H37Rv_2461_57bp_(**ETR-B**)	212	291	L)GCGAACACCAGGACAGCATCATG R)GGCATGCCGGTGATCGAGTGG
3	H37Rv_0577_58bp_(**ETR-C**)	211	346	L)GACTTCAATGCGTTGTTGGA R)GTCTTGACCTCCACGAGTGC
4	H37Rv_0580_77bp_(**MIRU04_ETR-D**)	252	353	L)GCGCGAGAGCCCGAACTGC R)GCGCAGCAGAAACGTCAGC

5	H37Rv_3192_53bp_(**MIRU31_ETRE**)	165	650	L)ACTGATTGGCTTCATACGGCTTTA R)GTGCCGACGTGGTCTTGAT
6	H37Rv_0154_53bp_ (**Miru02**)	128	507	L)TGGACTTGCAGCAATGGACCAACT R)TACTCGGACGCCGGCTCAAAAT
7	H37Rv_0959_53bp_(**MIRU10**)	157	643	L)GTTCTTGACCAACTGCAGTCGTCC R)GCCACCTTGGTGATCAGCTACCT
8	H37Rv_1644_53bp_(**MIRU16**)	119	670	L)TCGGTGATCGGGTCCAGTCCAAGTA R)CCCGTCGTGCAGCCCTGGTAC
9	H37Rv_2531_53bp_(**MIRU23**)	333	872	L)CAGCGAAACGAACTGTGCTATCAC R)CGTGTCCGAGCAGAAAAGGGTAT
10	H37Rv-2996_51bp_**(MIRU26)**	163	613	L)CCCGCCTTCGAAACGTCGCT R)TGGACATAGGCGACCAGGCGAATA
11	H37Rv_3006_53bp_(**MIRU27**)	183	656	L)TCGAAAGCCTCTGCGTGCCAGTAA R)GCGATGTGAGCGTGCCACTCAA
12	H37Rv_4348_53bp_(**Miru39**)	121	645	L)CGCATCGACAAACTGGAGCCAAAC R)CGGAAACGTCTACGCCCCACACAT
13	H37Rv_0802_54bp_(**MIRU40_Mtub08**)	54	407	L)GGGTTGCTGGATGACAACGTGT R)GGGTGATCTCGGCGAAATCAGATA
14	H37Rv_0024_18bp_(**Mtub01**)	178	328	L)GAGAAACAGGAGGGCGTTG R)TATTACGACGACCGCTATGC
15	H37Rv-0079_9bp_**(Mtub02)**	60	230	L)CGTGCACAGTTGGGTGTTTA R)TTCGTTCAGGAACTCCAAGG
16	H37Rv_1121_15bp_(**Mtub12**)	66	215	L)CTCCCACACCCAGGACAC R)CGGCCTACCCAACATTCC
17	H37Rv_1955_57bp_(**Mtub21**)	91	206	L)AGATCCCAGTTGTCGTCGTC R)CAACATCGCCTGGTTCTGTA
18	H37Rv_2347_57bp_(**Mtub29**)	219	350	L)AACCCATGTCAGCCAGGTTA R)ATGATGGCACACCGAAGAAC
19	H37Rv_2401_58bp_(**Mtub30**)	111	319	L)AGTCACCTTTCCTACCACTCGTAAC R)ATTAGTAGGGCACTAGCACCTCAAG
20	H37Rv_3690_58bp_(**Mtub39**)	151	341	L)AATCACGGTAACTTGGGTTGTTT R)GATGCATGTTCGACCCGTAG
21	H37Rv_2163_69bp_(**Qub11a**)	209	350	L)CCCATCCCGCTTAGCACATTCGTA R)TTCAGGGGGGATCCGGGA

Data management and analyses

Gel images were analyzed using the Bionumerics software package (version 4.5; Applied-Maths, Sint-Martens-Latem, Belgium) as previously described (Le Flèche et al., 2001; Fabre et al., 2004). The number of repeats in each allele was deduced from the amplicon size (Annexes 6). The resulting data were analyzed with Bionumerics as a character data set. Clustering analysis was done using the categorical parameter and the unweighted pair group method with

arithmetic averages coefficient. The minimum spanning tree (Feil et al., 2004) was constructed with the following options: (i) in case of equivalent solutions in terms of calculated distances, the selected tree was the one containing the highest number of links between genotypes differing at only one locus ("Highest number of single locus variants" option); (ii) the creation of hypothetical types (missing links) reducing the total length of the tree was allowed. Hunter-Gaston Index (HGDI) is calculated by $HGDI=1-\left[\frac{1}{N(N-1)}\sum_{j=1}^{s} nj\ (nj-1)\right]$ (Hunter and Gaston, 1988).

2.4. Main bioinformatic analysis softwares used in this study.

NCBI：http://www.ncbi.nlm.nih.gov/, Gel Compar 4.0 (Applied Maths, Kortrijk, Belgium), BioNumerics (Version 4.5) and SPSS (10.0).

3. RESULTS

3.1. Randomized drug-resistance research: identification of mutations in drug resistant genes

3.1.1. The Drug sensitivity test

Drug sensitivity test were carried out on a total of 2018 clinical M. tuberculosis strains isolated from patients with TB collected from 10 provinces in China.

The results showed that 1116 strains (55.30%) were sensitive to all 4 drugs (including INH, RFP, SM, and EMB), 902 (44.70%) were drug-resistant in which 313 (15.51%) were resistant to single drug and 589 (29.19%) to multiple drugs (Table 10, Table 11).

Table 10 Pattern of Drug-resistance in ten provinces

Province	Strains tested	Single drug-resistance	Rate (%)	Multi drug-resistance	Rate (%)	Total Rate (%)
Henan	75	7	9.33	32	42.66	52.00
Hunan	170	30	17.65	53	31.18	48.82
Xizang	216	24	11.11	118	54.63	65.74
Fujian	470	35	7.45	156	33.19	40.64
Gansu	219	62	28.31	47	21.46	49.77
Guangxi	206	27	13.11	32	15.53	28.64
Xinjiang	183	32	17.49	29	15.85	33.33
Zhejiang	190	42	22.11	36	18.95	41.05
Anhui	179	43	24.02	62	34.64	58.99
Shannxi	110	10	9.10	26	23.64	32.73
Total	2018	313	15.51	589	29.19	44.70

Table 11 Contribution of drug-resistance against four first line drugs in ten provinces, China

Kind of resistance	Henan	Hunan	Xizang	Fujian	Gansu	Guangxi	Xinjiang	Zhejiang	Anhui	Shannxi	Total	Rate (%)
sensitive to all drugs	36	87	75	279	110	147	122	112	74	74	1116	55.30
Drug-resistance	39	83	141	191	109	59	61	78	105	36	902	44.70
Single drug-resistance	7	30	25	35	62	27	32	42	43	10	313	15.51
INH	4	16	13	22	7	16	16	3	6	3	106	5.25
RFP	1	6	10	2	11	6	2	6	14	3	61	3.02
SM	2	6	2	11	43	5	10	16	22	4	121	6.00
EMB	0	2	0	0	1	0	4	17	1	0	25	1.24
Multi drug-resistance	32	53	116	156	47	32	29	36	62	26	589	29.19
INH + RFP	8	16	15	13	2	10	4	3	5	9	85	4.21
INH + EMB	0	2	3	2	0	1	2	3	0	0	13	0.64
INH + SM	5	1	11	16	9	4	9	6	2	7	70	3.47
SM + EMB	0	0	0	0	1	0	1	0	0	0	2	0.10
RFP + EMB	1	0	0	2	2	0	1	9	0	0	15	0.74
SM + RFP	0	0	3	2	22	2	2	0	12	1	44	2.18
INH + RFP + EMB	2	3	1	9	1	3	0	3	3	0	25	1.24
INH + SM + RFP	9	18	70	32	5	4	4	1	10	9	162	8.03
INH + SM + EMB	0	0	1	4	1	0	3	0	4	0	13	0.64
RFP + SM + EMB	0	1	0	1	2	1	0	1	0	0	6	0.30
INH+SM+RFP+EMB	7	12	12	75	2	7	3	10	26	0	154	7.63
Total	75	170	216	470	219	206	183	190	179	110	2018	

3.1.2. Preliminary study on the mutation characteristics of Rifampicin-resistance relationship gene *rpo*B

A 286bp DNA fragment of the *rpo*B gene including the 81bp hot-spot region (rifampin resistance determination region, RRDR Cluster I in Figure 16) was analyzed by PCR-SSCP (Liu et al., 2006).

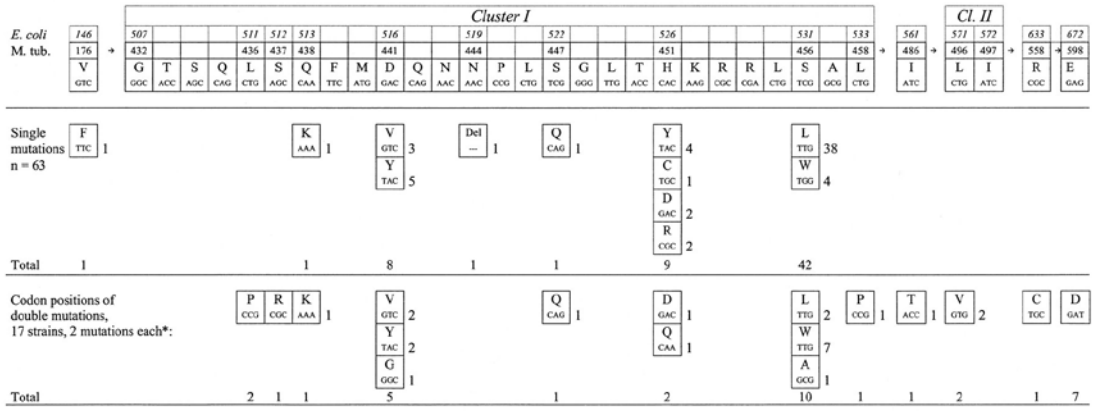


Figure 16 From (Heep et al., 2001)

In a first study, 47 rifampicin-resistant strains from Fujian Province were shown to have mutations by the PCR-SSCP method (Publication 1). In 76.6% rifampicin-resistant strains the *rpo*B gene had a single point mutation as shown by direct sequencing technique, which mainly located at 531-Ser (61.1%) (the amino acid number used thereafter is that of *rpo*B gene in *E. coli*) and 526-His (25.0%). Combinative mutation rate was 23.4%. In addition, 2 strains susceptible to rifampicin and resistant to other drugs and 1 strain susceptible to all tested drugs had mutations detected by PCR-SSCP method. Sequencing results showed that the mutations located at 511-Leu, 526-His and 535-Pro.

A second study was performed on 300 *M. tuberculosis* clinical isolates including 119 RFP-sensitive strains and 181 RFP-resistant strains, selected to investigate the mutations in the *rpo*B gene of *M. tuberculosis*. Mutations were found in 166 of 181 RFP-resistant strains in the fragment amplified by DNA sequencing (mutation rate is 91.71%) (results for 152 strains are shown in Table 13 and Table 12). The 531 position, 526 position and 516 position were the most common

sites to mutate; mutation rate were respectively 53.61%, 27.71%, 9.64%. The percentage of mutations at codon 511 was 6.02%. The sum of mutation frequencies of such four mutation sites was 96.99% in all mutation sites. Combinative mutations rate was 9.04% (15/166). An insertion was found in Cluster I in 1 strain, and deletions of single or multiple nucleotides were found in 5 strains. No mutation was found in 67 completely sensitive strains and a mutation of codon 511 and codon 526 was found respectively in 2 strains out of 52 strains sensitive only to RFP

Table 12 Multi loci mutations in 16 RFP-resistance strains

Num.	Mutation	Strains	Ratio (%)
1	509* AGC (Ser) –AG_, 510-511 (deletion), 512AGC (Ser) –_ _ C	1	0.61
2	510CAG (Gln) –CAC (His), 526CAC (His) –TAC (Tyr)	1	0.61
3	511CTG (Leu) –CCG (Pro), 505TTC (Phe) –CTC (Leu)	1	0.61
4	511CTG (Leu) –CCG (Pro), 515ATG (Met) –GTG (Val)	2	1.23
5	511CTG (Leu) –CCG (Pro), 512AGC (Ser) –ACC (Thr) 515ATG (Met) –ATC (Ile)	1	0.61
6	511 CTG (Leu) –CCG (Pro), 516GAC (Asp) –TAC (Tyr)	2	1.23
7	511 CTG (Leu) –CCG (Pro), 516GAC (Asp) –GGC (Gly)	2	1.23
8	511CTG (Leu) –CCG (Pro), 526CAC (His) –AAC (Asn)	1	0.61
9	516GAC (Asp) –GGC (Gly), 533CTG (Leu) –CCG (Pro)	1	0.61
10	516GAC (Asp) –GCC (Ala), 533CTG (Leu) –CCG (Pro)	1	0.61
11	516GAC (Asp) –TAC (Tyr), 526CAC (His) –CAA (Gln)	1	0.61
12	516 GAC (Asn) –TTC (Phe), 526CAC (His) –AAC (Asn)	1	0.61
13	516 GAC (Asn) – delation, 541GAG (Glu) –GGG (Gly)	1	0.61

.

Table 13 Single locus mutations in 136 RFP-resistant strains

DNA site	Changed Kind (a to b)	Amino acid site	Code changed (aaa to bbb)	Amino acid changed a to b	Strains	Ratio (%)
11223	c-t	531	tcg-ttg	Ser-Leu	88	53.99
11229	t-c	533	ctg-ccg	Leu-Pro	7	4.29
11166	t-c	511	ctg-ccg	Leu-Pro	2	1.23
11168	c-a	513	caa-aaa	Gln-Lys	1	0.61
11169	a-c	513	caa-cca	Gln-Pro	1	0.61
11171-6	ttc	514-515	ttc insertion	Phe insertion	1	0.61
11177	g-t	516	gac-tac	Asp-Tyr	3	1.84
11178	a-t	516	gac-gtc	Asp-Val	2	1.23
11178	a-g	516	gac-ggc	Asp-Gly	2	1.23
11195-6	tc-at	522	tcg-atg	His -Met	1	0.61
11207	c-a	526	cac-aac	His-Asn	4	2.46
11207	c-t	526	cac-tac	His-Tyr	11	6.75
11208	a-g	526	cac-cgc	His-Arg	6	3.68
11208	a-t	526	cac-ctc	His-Leu	2	1.23
11207-8	ca-tg	526	cac-tgc	His-Cys	3	1.84
11207-8	ca-ag	526	cac-agc	His-Ser	1	0.61
11207-8	ca-gg	526	cac-ggc	His –Gly	1	0.61

The mutation of *rpo*B gene in 300 strains which mutation had been analyzed by DNA sequencing was detected by means of oligonucleotide probe reverse dot blot hybridization (RDH) based on PCR as described in (Jiao et al., 2007) (Figure 17).

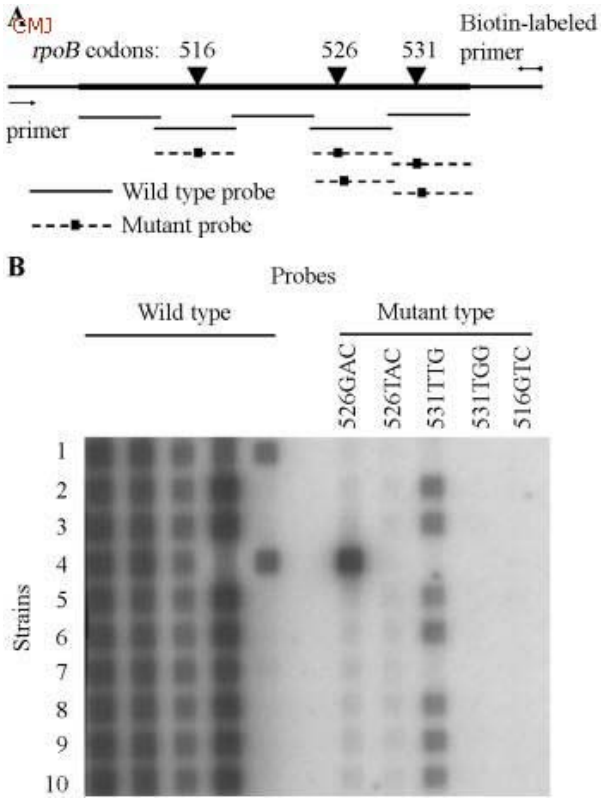


Figure 17 From (Jiao et al., 2007)

The results showed that no mutation was found in 67 completely sensitive strains whereas a mutation was found in 2 strains out of 52 strains sensitive only to RFP. Within 181 RFP-resistant strains, 165 were detected to bear mutations, whereas no mutation was found in 16 strains. Compared with DNA sequencing, the sensitivity and specificity of RDH were 97.62% and 97.73% respectively, the concordance with DNA sequencing of *M. tuberculosis* was 97.67. There was no significant difference between RDH and DNA sequencing by Fisher analysis (P>0.05). (Figure 18, Table 14, Table 15, Table 16, Table 17).

Table 14 Comparison between drug susceptibility test and DNA sequence method for RFP-resistance analysis on drug resistance strains

	Drug resistance	DNA	Sequence	Total
		Mutation	Non-mutation	
Proportional drug susceptibility test	Single resistance (RFP only)	24	4	28
	Multi drug resistance (including RFP)	143	10	153
	Total	167	14	181

P=0.2375 (fisher analysis)，P＞0.05

Table 15 Comparison between drug susceptibility test and DNA sequence method for RFP-resistance analysis on drug sensitivity strains

	Sensitivity to drug	DNA	Sequence	Total
		Mutation	Non-mutation	
Proportional drug susceptibility test	Sensitive to all 4 drugs	0	67	67
	Sensitive to RFP only	2	50	52
	Total	2	117	119

P=0.1889 (fisher analysis)，P＞0.05

Table 16 Comparison between drug susceptibility test and oligonucleotide probe reverse dot blot hybridization for RFP-resistance analysis on drug resistance strains

	Drug resistance	DNA	Sequence	Total
		Mutation	Non-mutation	
Proportional drug susceptibility test	Single resistance (RFP only)	24	4	28
	Multi drug resistance (including RFP)	144	9	153
	Total	168	13	181

P=0.1210 (fisher analysis)，P＞0.05

Table 17 Comparison between drug susceptibility test and reverse dot blot hybridization for RFP-resistance analysis on drug sensitivity strains

	Drug resistance	Oligonucleotide	probe blot	Total
		Mutation	Non-mutation	
Proportional drug susceptibility test	Sensitive to all 4 drugs	0	67	67
	Sensitive to RFP only	2	50	52
	Total	2	117	119

P=0.1889 (fisher analysis)，P＞0.05

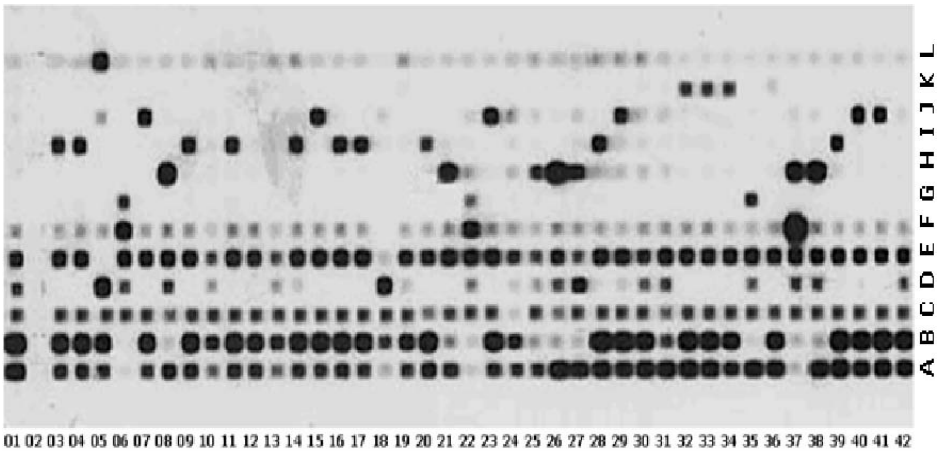


Figure 18 oligonucleotide probe reverse dot blot hybridization ; A to E: wild type probes, G-L: mutation type probes

When the wild type probes produce a positive reaction it shows that the corresponding nucleotide is present. When a wild type nucleotide is absent, the corresponding mutation is observed in one of the test lines. 01: H37Rv, 02: Negative control; line 03, 04, 09, 11, 14, 16, 17, 20, 28 and 39 were the mutation occurred on 526 location from CAC (His) to GAC (Asp); line 5 was the mutation occurred on 531 location from TCG (Ser) to TTG (Leu); line 06, 22 and 37 were the combinative mutations occurred on 511 and 516 locations from (511) CTG (Leu) to CCG (Pro) and (516) GAC (Asp) to GGC (Gly), (511) CTG（Leu）to CCG (Pro) and (516) GAC（Asp）to TAC (Tyr) respectively; line 07, 15, 23, 29, 40 and 41 were the mutations occurred on 526 location from CAC (His) to CGC (Arg); line 32, 33 and 34 were the mutation occurred on 526 location from CAC (His) to TAC (Tyr); line 08, 21, 25, 26, 27 and 38 were the mutation occurred on 516 location from GAC (Asp) to TAC (Tyr); line 10, 13, 30, 31 and 42 were non-mutation occurred; line 12, 18, 19, 24 and 36 were the mutation occurred on 526 location, but because of the mutation probes were negatitive the mutation were unidentified; line 35 was the mutation occurred on 516 location from GAC (Asp) to GGC (Gly).

3.1.3. Preliminary study on the mutation characteristics of INH-resistance relationship genes

We have designed a pairs of primer for *kat*G, pre-*inh*A *inh*A, *ndh* and *oxy*R-*ahp*C respectively. Then PCR and the DNA sequence analysis were carried out to detect the gene mutations in 291 *M. tuberculosis* clinical isolates including 64 strains absolutely sensitive to INH, 48 relatively sensitive to INH, 30 showing single drug-resistance to INH only, and 149 showing multi drug-resistance.

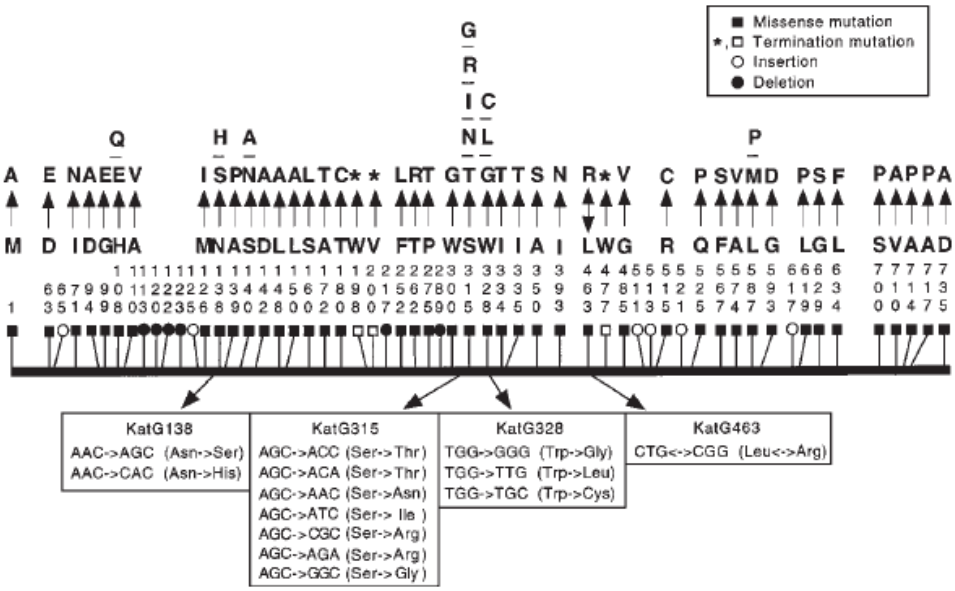


Fig. 2 Polymorphism in the KatG protein identified in INH^R *M. tuberculosis*. The data were compiled from mutations reported previously.[44,66–68,70–80] The variant amino acids are numbered vertically. The single-letter amino acid abbreviations are used. Shown below the schematic are nucleotide and amino acid changes occurring at codons with two or more variant codons. The KatG463 Leu<–>Arg substitution is a commonly-occurring natural polymorphism that is not associated with INH susceptibility levels judged to be clinically significant. A: alanine; C: cysteine; D: aspartic acid; E: glutamic acid; F: phenylalanine; G: glycine; H: histidine; I: isoleucine; L: leucine; M: methionine; N: asparagine; P: proline; Q: glutamine; R: arginine; S: serine; T: threonine; W: tryptophan; V: valine.

Figure 19 Polymorphism in the KatG protein (from Ramaswamy and Musser (Ramaswamy and Musser, 1998))

The results showed:

(1) A mutation in the “hot spot mutation area” of the *kat*G gene was found in 163 out of 179 INH-resistant strains and in 3 out of 112 INH-sensitive strains. Compared with the proportional drug susceptibility testing, the sensitivity and specificity of DNA sequencing were 91.06% (163/179) and 97.32% (109/112) respectively, the positive forecast value and the negative forecast value were 98.19% (163/166) and 87.20% (109/125) respectively. The concordance rate between the proportional drug susceptibility testing and DNA sequencing was 93.47% (272/291).

(2) In 179 INH-resistant strains, 129 strains (72.07%) have a mutation at *katG* codon 463, and in 112 drug-sensitive, 74 strains (66.07%) have the same. There was no significant difference by Chi-square statistics test (χ^2=1.1739, P > 0.05). Therefore, it is confirmed that the gene mutation at codon 463 has no correlation with INH-resistance (Figure 19).

(3) In the 179 INH-resistant strains there were 138 (77.09%) which had a gene mutation in *kat*G, of which 7 (Ile317→Val, Ile317→Met, Pro375→Pro, Asn329→Val, Thr344→Pro, Phe368→Leu, His400→Pro) had not been reported until now; in 131 strains (73.18%) mutation occured at codon 315, consisting of 4 kinds of mutation types, but the main mutation locus was AGC (Ser) to ACC (Thr), which occurred in 123 strains (93.89%).

(4) in 1 strain (0.56%) an *inh*A mutation was observed in INH-resistant strains while none was found in the sensitive strains.

(5) 18 (10.06%) INH-resistant strains showed a mutation in the *oxy*R-*ahp*C gene, of which 7 (3.39%) had a -9C to T mutation, 3 strains had a -5G to A mutation, 3 strains had a -14C to T change; while no mutation was found in the sensitive strains.

(6) 3 (1.68%) INH-resistant strains were mutated in *ndh* (Glu191Gly), while no change was observed in the sensitive strains.

(7) Concerning the pre-*inh*A gene, in 179 INH-resistant strains, there were 6 strains showing a change of 15C to T and 6 G to A. Other mutation, such as 1 T to C, was also found in some stains while no change was found in the sensitive strains.

(8) There was no mutation detected in *oxyR* and *ahpC* genes in any of the strains (Table 18 to Table 20)

Table 18 The contribution of mutations on INH-resistance relationship genes in 291 Chinese clinical *M. tuberculosis* isolates (R resistance S sensitivity)

Gene	Strains tested	R/S to INH	Mutation strains	Mutation rate (%)	Amino acid site	Nucleotide change	Amino acid changed
KatG	179	R	123	68.72	944	G→C	Ser315→Thr
			6	3.35	944	G→A	Ser315→Asn
			1	0.56	945	C→A	Ser315→Arg
			1	0.56	943	A→G	Ser315→Gly
			1	0.56	949	A→G	Ile317→Val
			1	0.56	951	C→G	Ile317→Met
			1	0.56	1125	G→T	Pro375→Pro
			1	0.56	986	A→T	Asn329→Val
			1	0.56	1030	A→C	Thr344→Pro
			1	0.56	1102	T→C	Phe368→Leu

			1	0.56	1199	A→C	His400→Pro
	112	S	1	0.89	943	A→G	Ser315→Gly
			1	0.89	1010	A→G	Tyr337→Cys
			1	0.89	1280	T→C	Leu427→Pro
Pre-inhA	179	R	6	3.35	-15	C→T	
			6	3.35	-19	G→A	
			1	0.56	-8	T→C	
			1	0.56	13	G→C	
inhA	179	R	1	0.56	280	T→G	Ser94→Ala
oxyR-ahpC	179	R	7	3.91	-9	C→T	
			3	1.68	-14	C→T	
			3	1.68	-5	G→A	
			2	1.12	-47	G→A	
			1	0.56	-8	G→A	
			1	0.56	-29	C→T	
			1	0.56	-60	C→G	
ndh	179	R	3	1.68	572	A→G	Glu191→Gly

Table 19 The mutations occurred on *katG*, pre-*inhA*, *inhA*, *ndh*, *oxy*R-*ahp*C in 179 Chinese *M. tuberculosis* INH-resistance strains

katG	pre-inhA	inhA	oxyR-ahpC	ndh	Strains	Mutation rate (%)
+	–	–	–	–	127	70.39
–	+	–	–	–	10	5.59
–	–	+	–	–	1	0.56
–	–	–	+	–	11	6.15
–	–	–	–	+	1	0.56
+	+	–	–	–	8	4.47
+	–	–	+	–	2	1.12
–	+	–	+	–	1	0.56
–	–	–	+	+	2	1.12
+	+	–	+	–	1	0.56
				Total	163	91.06

+: Mutation occurred; –: No mutation occurred

Table 20 Comparison between drug susceptibility test and DNA sequence method for INH-resistance analysis on drug resistance strains

	Drug resistance	DNA Sequence		Total
		Mutation	Non-mutation	
Proportional drug susceptibility test	Single resistance (INH only)	25	5	30
	Multi drug resistance (including INH)	138	11	149
	Total	163	16	179

P=0.1514 (fisher analysis)，P＞0.05

Table 21 Comparison between drug susceptibility test and DNA sequence method for INH-resistance analysis on drug sensitivity strains

	Sensitivity to drug	DNA Sequence		Total
		Mutation	Non-mutation	
Proportional drug susceptibility test	Sensitive to all 4 drugs	1	63	64
	Sensitive to INH only	2	46	48
	Total	3	109	112

P=0.5754 (fisher analysis)，P＞0.05

3.2. The genetic diversity of *M. tuberculosis* strains in China

3.2.1. Preliminary study (Publication N° 3)

A pilot study was first conducted on isolates from 5 Chinese Eastern Provinces, Hunan, Jiangsu, Anhui, Fujian and Beijing in order to select the techniques to be used in a large scale study of several thousand isolates.

One hundred and twenty seven isolates were genotyped by MLVA using 21 markers as described in previous work by our laboratory in orsay (Le Flèche et al., 2002; Fabre et al., 2004), and by spoligotyping.

The 21 VNTRs loci include ETR-A, ETR-B, ETR-C, ETR-D (MIRU04), ETR-E (MIRU31), MIRU02, MIRU10, MIRU16, MIRU23, MIRU26, MIRU27, MIRU39, MIRU40, Mtub01, Mtub02, Mtub12, Mtub21, Mtub29, Mtub30, Mtub39, Qub11a.

A total of 57 strains distributed into 37 genotypes were found to belong to the Beijing family and represent 55 to 95% of the isolates in the different provinces.

The Chinese isolates were compared to a collection of 125 isolates from other countries worldwide in which the major families are present. The MLVA group code described in the work by Le Flèche *and col.* was used and new groups were found (Le Flèche et al., 2002).

In addition to the Beijing family, MLVA21 distinguished two groups of strains, strongly associated with China and representing 17% of the total Chinese strains (GroupVIII/China2 and

GroupVIb/China3), and which may represent new phylogeographically specific genotypes (Figure 20). Strains of the ancestral EAI family were not found and only one isolate belonged to the CAS family. These preliminary results confirmed that the Beijing family was prevalent in China and revealed the existence of a limited number of other families apparently restricted to China (Publication 3).

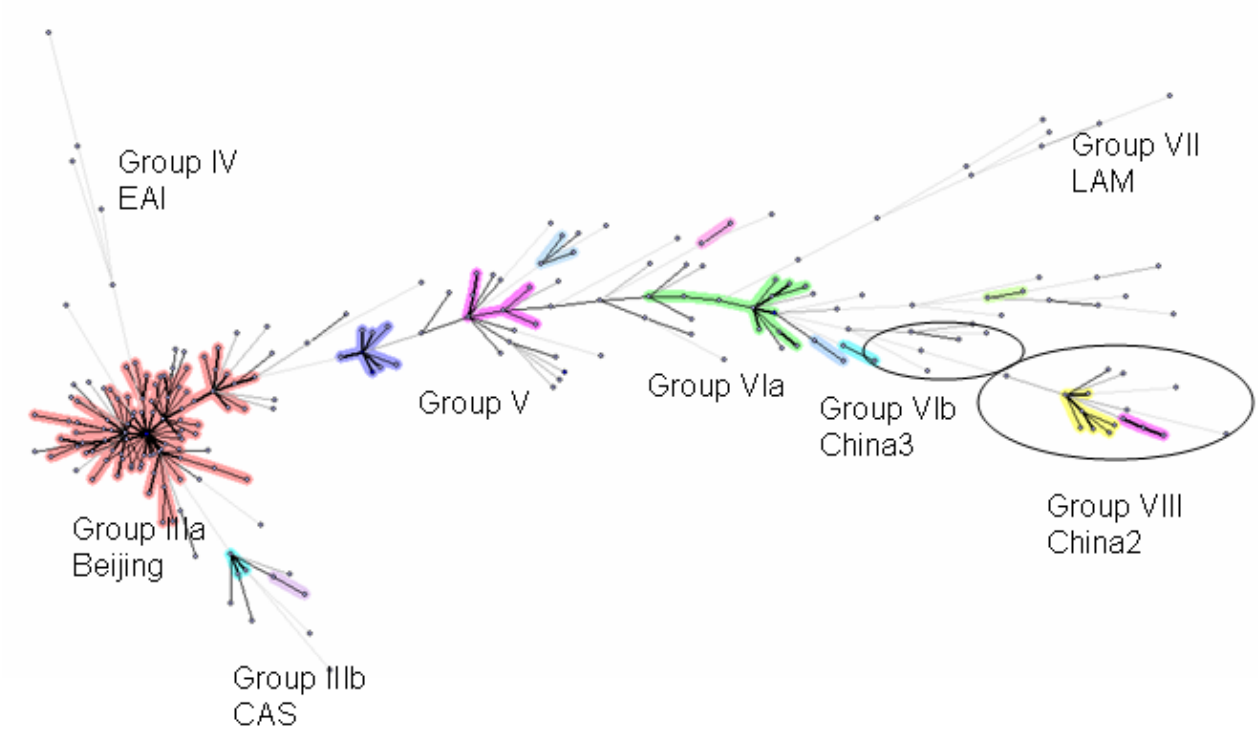


Figure 20 Minimum spanning tree showing the clustering by MLVA21 of 252 *M. tuberculosis* isolates comprising 127 Chinese isolates and 122 isolates from different countries worldwide.

The distribution into groups is based on the work by (Le Flèche et al., 2001). The correspondence with Clades defined by spoligotyping is indicated. CAS is Central Asian, LAM is Latin America.

3.2.2. Extended analysis of the DR/CRISPR locus in Beijing-family strains

Spoligotyping does not provide a complete description of the DR locus as some spacers are not included in the assay (spacers 55, 56 and 57 according to (van Embden et al., 1993)) and the presence of IS cannot be detected.

Therefore we used two PCR reactions to test for additional polymorphism and some samples were sequenced for further information (Figure 21).

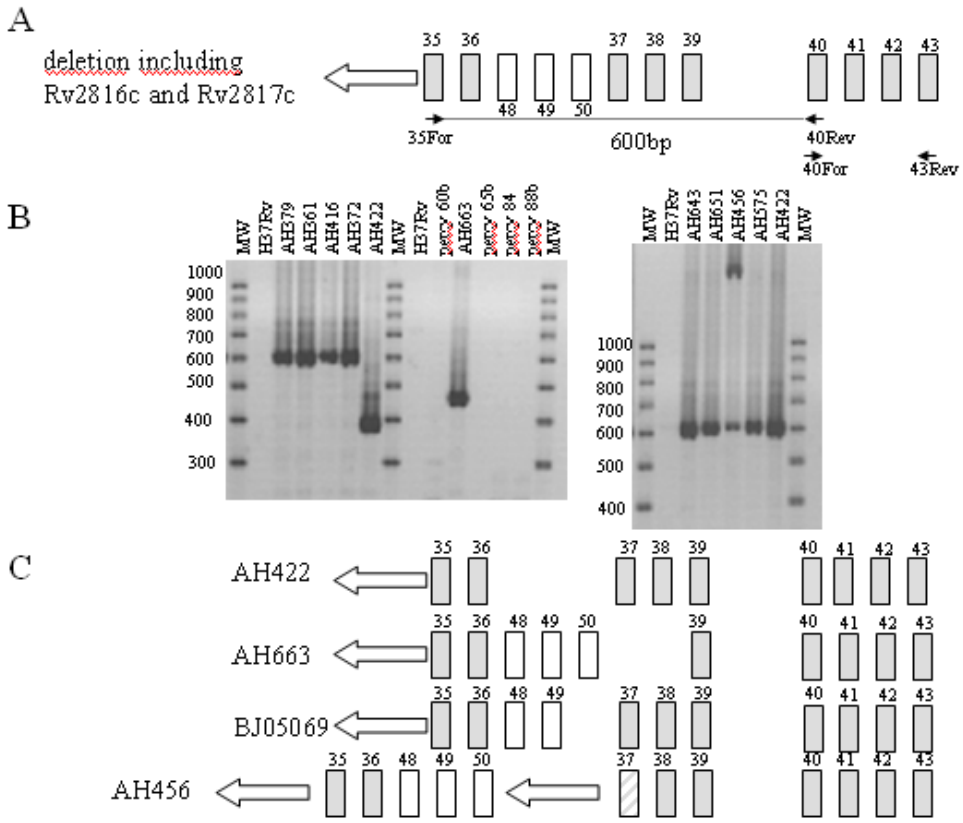


Figure 21 PCR-based analysis of the DR locus

A: Schematic representation of the DR locus where the spacers are shown by boxes bound by an IS*6110* element (large horizontal arrow). The numbers are from (van Embden et al., 2000) and represent the actual repeats order in the DR locus. PCR primers are shown as small arrows. B: PCR analysis using primers 35For and 40Rev of 11 Chinese strains, the H37Rv control strain and 4 EAI *M. tuberculosis* strains. C: representation of the DR locus of four non-typical Beijing strains showing the absence of some motifs or the insertion of an IS element.

In the majority of strains a classical arrangement of DR and spacer was observed whereas in 4 strains some of the spacers not included in the spoligotyping assay were deleted. Insertion of an IS element was observed in strain AH456.

To assay the presence of an IS*6110* element within the DR locus, a method designated as Left-Right spoligotyping or LR-spoligo has been developped and could be used to further investigate the structure of the DR locus (Filliol et al., 2000; Legrand et al., 2001).

3.2.3. VNTR locus selection for MLVA

In order to facilitate genotyping of large number of isolates we sought to select of smaller panel of VNTRs which would correctly cluster the bacteria into the main clades/families while providing additional informativity as compared to spoligotyping. Our first concern was to eliminate markers that were difficult to type such as Mtub02 (9bp repeat) or unstable such as Qub11a or were not very informative such as Mtub12. Clustering was performed with different combination of VNTRs and we retained a set that is amenable to manual reading and produce a satisfying clustering (Figure 22 and Figure 23). MLVA15 correspond to ETR-A, ETR-B, ETR-C, ETR-D (MIRU04), ETR-E (MIRU31), MIRU10, MIRU16, MIRU23, MIRU26, MIRU27, MIRU39, MIRU40, Mtub21, Mtub30, Mtub39.

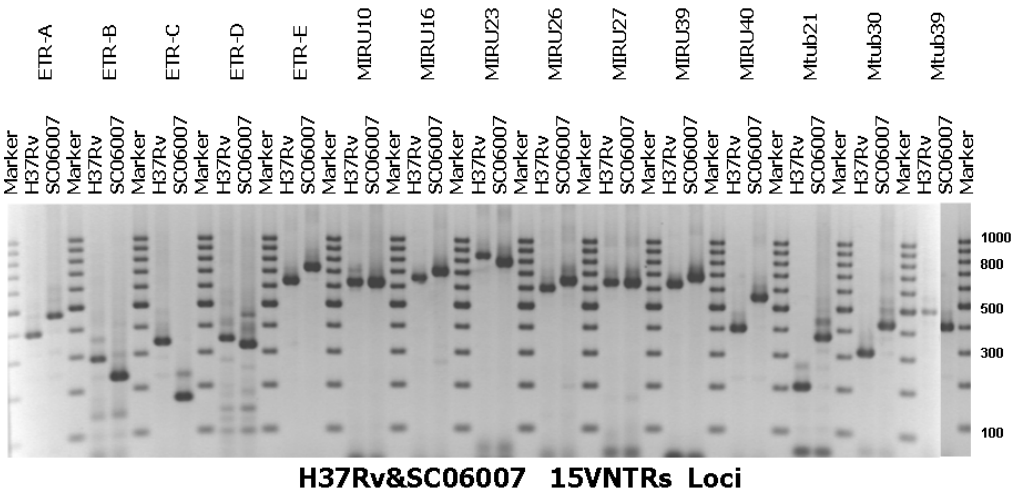


Figure 22 Length polymorphisms of 15 different VNTR loci in two strains H37Rv and SC0600

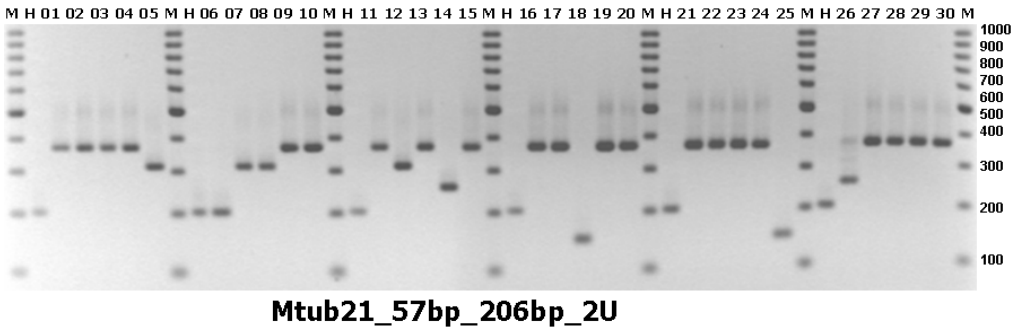


Figure 23 Length polymorphisms at the Mtub21 Locus in H37Rv and 30 studied strains

The following figure (Figure 24) shows a dendrogram in which 148 strains were clustered by UPGMA (all the Chinese strains and 21 others including EAI and *M. canettii* strains). The clustering was performed using either MLVA21 or MLVA15. The result shows that MLVA15 allows the correct clustering of strains into families although there is a difference in the position of the MOD family. The *M. canettii strains* cluster into a single outgroup as compared to all the other *M. tuberculosis* strains (Figure 24).

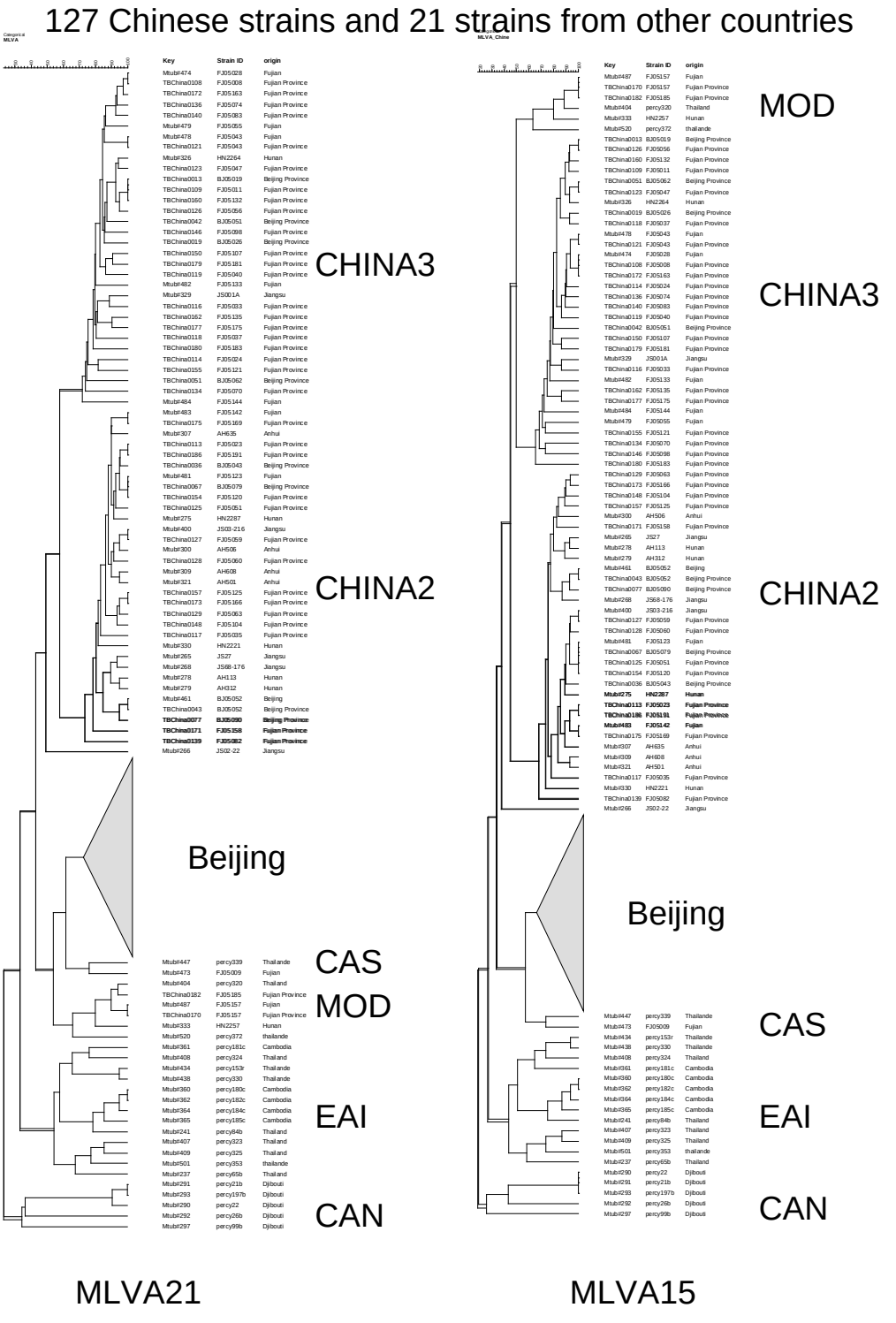


Figure 24 Dendrogram deduced from the clustering analysis of the 148 strains with MLVA21 or MLVA15

3.2.4. A survey of diversity in 12 Provinces

From 2004 to 2006, 2547 *M. tuberculosis* isolates were obtained from sputum samples in 13 provinces (municipality, autonomous region) distributed over the all country, including Beijing, Henan, Sichuan, Hunan, Anhui, Zhejiang, Fujian, Jilin, Shannxi, Gansu, Xinjiang, Xizang, and Guangxi (Figure 25). Then the bacteria were identified with biochemical tests, and the drug susceptibility was assessed by means of proportional drug susceptibility test.

Region	Strains
Beijing	108
Anhui	204
Fujian	485
Guangxi	209
Henan	98
Sichuan	103
Xizang	219
Xinjiang	205
Jilin	349
Shannxi	144
Gansu	221
Hunan	102
Zhejiang	100
Total	2,547

Figure 25 Bacterial samples from 13 provinces in China

Among all these isolates, 1937 *M. tuberculosis* clinical strains from 12 provinces have been studied for their genetic diversity. Molecular typing identification and genetic polymorphisms of the strains were determined by looking at different polymorphic markers, DR loci (Spoligotyping), RD (Genetic deletion, Large sequence polymorphisms (LSPs), or/and tandem repeats (MLVA) (Table 22).

Table 22 Bacterial Samples analysed by Spoligotyping, RD and /or MLVA

Province	1	2	3	4	5	6	7	8	9	10	11	12	Total
Spoligo	121	215	333	175	80	120	97	204	99	107	178	208	1937
MLVA	9	204	319	72	0	93	91	176	163	102	39	202	1470
RD	110	189	298	144	66	96	68	118	64	66	98	115	1432

1~12: Beijing, Xizang, Jilin, Guansu, Henan, Shannxi, Hunan, Xinjiang, Zhejiang, Sichuan, Fujian, and Guangxi.

Spoligotyping allowed the clear separation of strains into two main groups: those belonging to the Beijing family and the non-Beijing family group.

With MLVA15, 1470 Chinese *M. tuberculosis* clinical isolates from 11 provinces (municipality and autonomous regions) were genotyped, using H37Rv as reference strain. MLVA15 correctly aggregated the Beijing family into a large homogenous group although some of them showed more polymorphism (see following paragraphs).

There was an obvious genetic diversity in the *M. tuberculosis* strains isolated from different provinces of China, although the main epidemic strain cluster in the different provinces corresponds to the Beijing Family.

3.2.5. The Beijing family strains

By spoligotyping most of the isolates showed the presence of all the spacers from 36 to 43 but in some strains deletion of one or several spacers was observed (Figure 26).

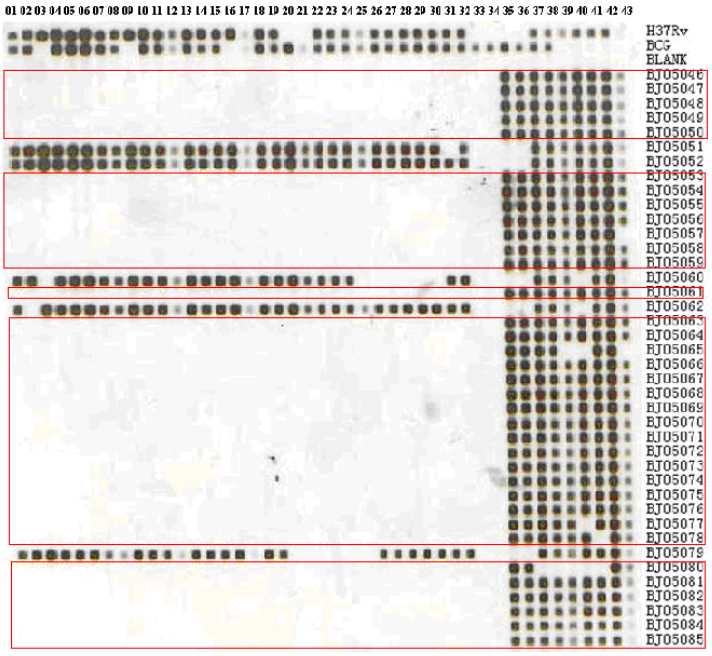


Figure 26 Genotype identification of BeiJing M. tuberculosis isolates with Spoligotyping

1 to 43 spacer oligonucleotide probes; H37Rv: and; BCG as the positive control; BLANK as the negative control; BJ05046 to BJ05085 are clinical isolates from Beijing.

At the moment the DR region of part of the Beijing-family isolates has been investigated by PCR and sequencing as shown in paragraphe 3.2.2. The presence of IS was demonstrated in some samples. The complete collection of isolates will be investigated by this method in the future.

Geographical distribution of *M. tuberculosis* Beijing family in 12 provinces

The mean percentage of Beijing family strains was 75.68% (1,466/1,928), with large differences in the 12 provinces (55.3% ~ 93.4%) (Table 23), whereas in other Asian countries they represent 50% or less of the *M. tuberculosis* strains. Geographical contribution of *M. tuberculosis* Beijing family in China could be divided roughly into four types. The results show that the highest number of Beijing family strains are observed in the Northern China including Xizang (Tibet), the second was in center of China, e. g. between Yellow River and Yangtze River, the lowest was in southern China, but Xinjiang might be a special area. (Figure 27).

Table 23 *M. tuberculosis* Beijing family strains distribution in twelve provinces, analyzed by Spoligotyping

Num	Province	Strains tested	Beijing strain	Rate (%)	Non-Beijing	Rate (%)
1	Beijing	121	113	93.40	8	6.60
2	Xizang	215	194	90.20	21	9.80
3	Jilin	333	298	89.50	.35	10.50
4	Gansu	175	149	85.10	26	14.90
5	Henan	80	67	83.80	13	16.20
6	Shannxi	120	97	80.80	23	19.20
7	Hunan	97	68	70.10	29	29.90
8	Xinjiang	204	135	66.20	69	33.80
9	Zhejiang	99	64	64.60	35	35.40
10	Sichuan	107	66	61.70	41	38.30
11	Fujian	178	100	56.20	78	43.80
12	Guangxi	208	115	55.30	93	44.70
13	Total	1,937	1,466	75.68	471	24.32

Figure 27 contribution of *M. tuberculosis* Beijing family in China

The role of BCG vaccination in the dissemination of the Beijing family

It has been hypothesized that BCG vaccination was responsible for the emergence of the Beijing family of *M. tuberculosis*. In Xizang (Tibet) municipality it was possible to check this hypothesis because some of the population is not vaccinated.

Among Beijing family strains, there were 87 strains (46.28%) from BCG vaccinated patients, while 101 strains (53.72%) were from non-vaccinated patients (Table 24). There was no significant difference in statistics (χ^2=0.012，P>0.05).

Table 24 Analysis on connecting Beijing family with BCG used in the clinical strains isolated from Tibet

BCG used	Beijing Family	Non-Beijing Family	Total
Yes	87	9	96
No	101	11	112
Total	188	20	208

The results show that there was no significant difference in the two groups whether BCG was used or not.

This strongly suggests that there might be no relationship between Beijing family of *M. tuberculosis* and BCG vaccination.

The role of Drug-resistance in the dissemination of the Beijing family of *M. tuberculosis.*

It has been hypothesized that strains of the Beijing family might become drug-resistant with a high efficiency thus implying that drug-resistance was responsible for the emergence of this family. In the present study it was possible to check this hypothesis with the results of drug susceptibility testing on 1286 *M. tuberculosis* clinical isolates from 10 provinces of China.

In Beijing family strains there were 499 (71.18%) sensitive strains, and 439 (75.04%) drug resistant strains. There was no significant difference in statistics (χ^2= 2.41， P > 0.05) between the drug resistant and drug sensitive strains. By comparison with the relative percentages of resistant and susceptible strains in the non Beijing families it appears that drug resistance is not more frequent in Beijing-family strains (Table 25, Table 26, Table 27).

Table 25 The relationship between Beijing genotype and drug-resistance in 10 provinces of China

Drug susceptibility test	Beijing family	Non-Beijing family	Total
drug susceptibility	499	202	701
drug-resistance	439	146	585
Total	938	348	1286

Chi-Square Tests, χ^2=2.41, $0.1 < P < 0.25$

Table 26 Comparison on the relationship between Beijing genotype and single drug-resistance to 4 drugs respectively

Kind of resistance	Beijing family	Non-Beijing	χ^2	OR	P	OR 95% CI
INH-resistance	313	110	0.36	1.08	0.5-0.75	(0.83,1.42)

INH-susceptibility	625	238				
RFP-resistance	284	93	1.55	1.19	0.1-0.25	(0.90,1.58)
RFP-susceptibility	654	255				
EMB-resistance	118	50	0.71	0.86	0.25-0.5	(0.59,1.24)
EMB-susceptibility	820	298				
SM-resistance	269	84	2.63	1.26	0.1-0.25	(0.94,1.69)
SM-susceptibility	669	264				

Table 27 Comparison on the relationship between Beijing genotype and drug-resistance in 12 provinces of China

Province	drug susceptibility test	Beijing family	Non-Beijing family	Total
Total of 12	drug susceptibility	499	202	701
province	Single drug-resistance	119	34	153
	Multi-drug Resistance	320	112	432
Beijing	drug susceptibility	76	5	81
	Single drug-resistance	10	1	11
	Multi-drug Resistance	12	0	12
Xizang	drug susceptibility	68	5	73
	Single drug-resistance	22	2	24
	Multi-drug Resistance	105	14	119
Ji Lin	drug susceptibility	5	0	5
	Single drug-resistance	3	0	3
	Multi-drug Resistance	18	4	22
Gansu	drug susceptibility	61	11	72
	Single drug-resistance	39	7	46
	Multi-drug Resistance	34	6	40
Henan	drug susceptibility	36	7	43
	Single drug-resistance	7	2	9
	Multi-drug Resistance	18	2	20
Hunan	drug susceptibility	37	13	50
	Single drug-resistance	4	3	7
	Multi-drug Resistance	15	9	24
Xin jiang	drug susceptibility	79	42	121

Province	drug susceptibility test	Beijing family	Non-Beijing family	Total
	Single drug-resistance	12	10	22
	Multi-drug Resistance	29	11	40
Sichuan	drug susceptibility	6	6	12
	Single drug-resistance	2	1	3
	Multi-drug Resistance	27	18	45
Fujian	drug susceptibility	54	44	98
	Single drug-resistance	1	0	1
	Multi-drug Resistance	45	33	78
Guangxi	drug susceptibility	76	69	145
	Single drug-resistance	19	8	27
	Multi-drug Resistance	18	15	33
Total		938	348	1286

Note: Total of 10 province （χ^2= 3.14 (3.19) ， 0.1＜*P*＜ 0.25 (P=0.2028) ）

Beijing, P=0.5835 (fisher analysis)

Xizang, χ^2= 1.35, *P*=0.5091 (Chi-Square Tests)

Ji Lin, *P*=0.7190 (fisher analysis)

Gansu, χ^2=0.0016, *P*=0.9992 (Chi-Square Tests)

Henan, *P*=0.6378 (fisher analysis)

Hunan, *P*=0.4568 (fisher analysis)

Xinjiang, χ^2=2.0396, *P*=0.3607 (Chi-Square Tests)

Sichuan, *P*=0.8884 (fisher analysis)

Fujian, *P*=0.8652 (fisher analysis)

Guangxi, χ^2=2.9721, *P*=0.2263 (Chi-Square Tests)

The genetic diversity of the Beijing strains

MLVA analysis

1107 Beijing-type isolates from 9 Provinces (Beijing-type isolates from Beijing, Fujian and Henan provinces have not been completely typed by MLVA) were analysed by UPGMA using the categorical coefficient, and colors were arbitrarily used to define groups differing by a maximum of two VNTRs. The minimum spaning tree shown on Figure 28 was performed for the purpose of showing the homogeneity of the population. The large majority of isolates differ by one to three

VNTRs whereas some of them show more diversity. Samples showing more than 80% similarity are colored in yellow, and the samples showing a higher diversity (less than 70% similarity) are colored in purple.

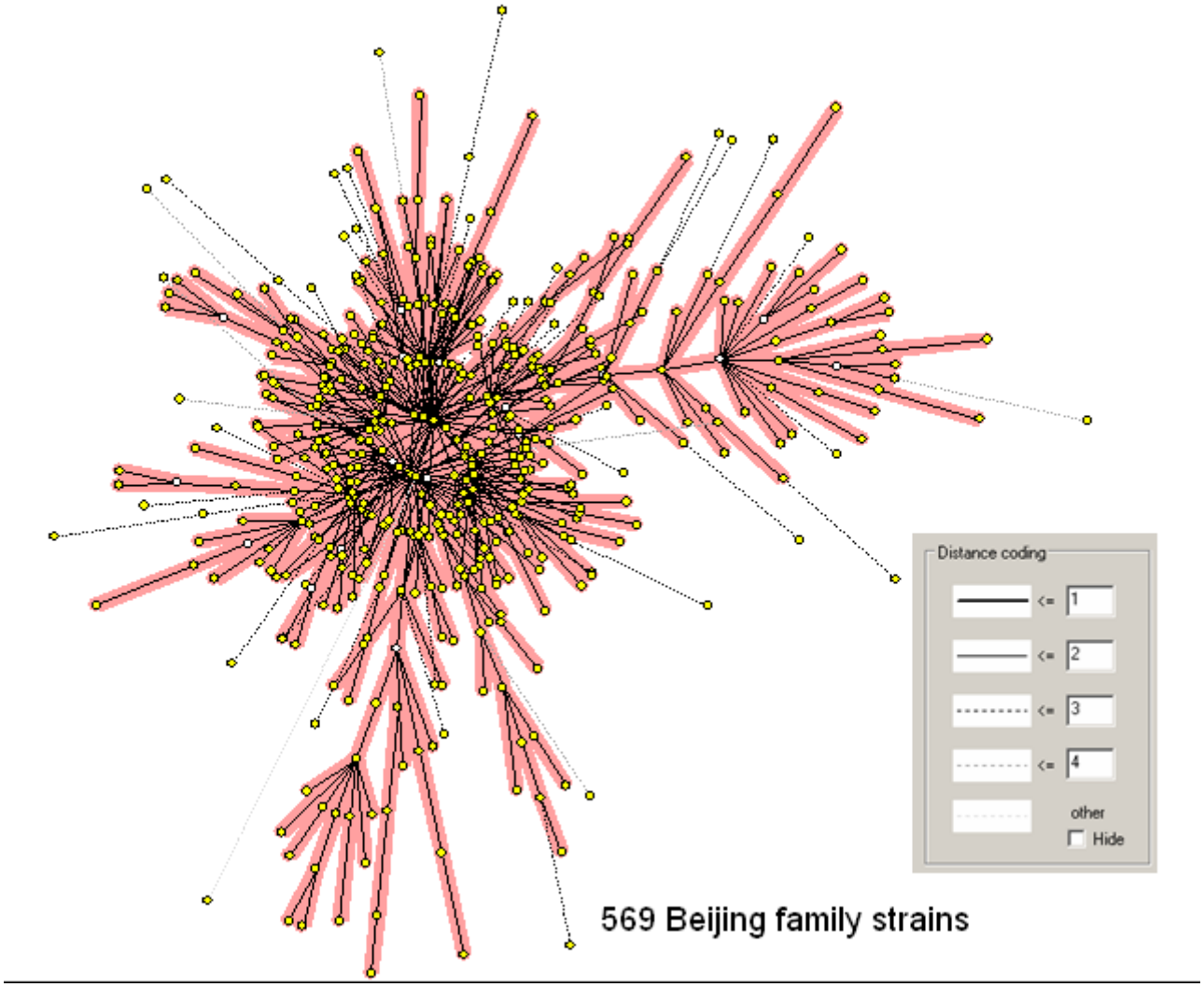


Figure 28 the minimum spanning tree deduced from the clustering analysis of 1107 Beijing-family strains by MLVA15

Genomic deletion analysis

We sought to identify the more ancestral strains of the Beijing family in order to investigate the geographical origin of this family. For this we tested for the presence/absence of the RD181 region by PCR using primers localized outside (external) or inside (internal) this region as shown on Figure 29.

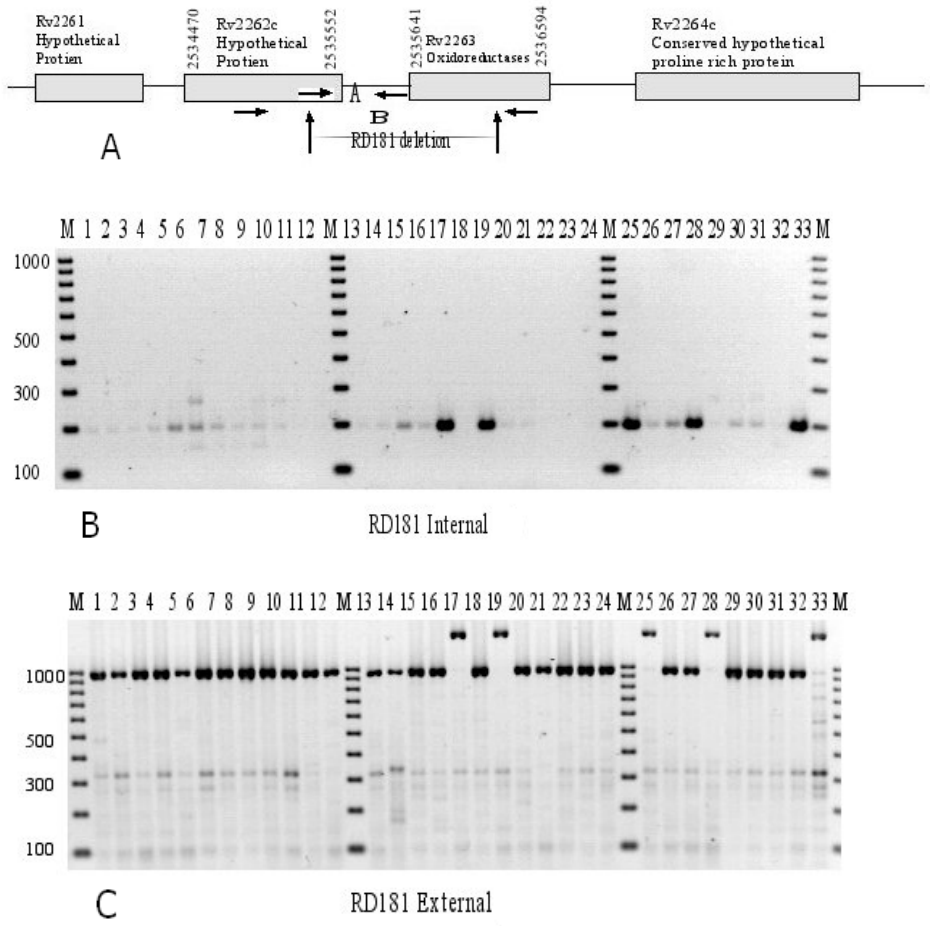


Figure 29 Genomic deletions for large sequence polymorphisms (LSPs) analyzed by RD181 in the Beijing family of *M. tuberculosis* clinical strains

A: The pattern of RD and Beijing-specific deletion RD181 location in the bacterial DNA; B: The results of some *M. tuberculosis* strains analyzed by PCR with RD181 internal primer; C: The results of the same strains analyzed by PCR with RD181 external primer. M: 100bp DNA Marker; 1 to 33: FJ05066, FJ05067, FJ05072, FJ05076, FJ05079, FJ05085, FJ05093, FJ05094, FJ05096, FJ05099, FJ05109, FJ05116, FJ05119, FJ05122, FJ05131, FJ05134 FJ05143, FJ05145, FJ05148, FJ05149, FJ05151, FJ05154, FJ05155, FJ05168, FJ05174, FJ05177, FJ05184, FJ05186, FJ05187, FJ05193, FJ05195, FJ05196, H37Rv.

Figure 29 shows that FJ05143, FJ05148, FJ05174, FJ05186 and H37Rv are positive with both RD181 internal and RD181 external primers, showing that the region is not deleted. These strains are considered to be ancestral. The other strains are positive with RD181 external primer and negative with RD181 internal primer, showing that the region is deleted.

A total of 1432 *M. tuberculosis* Beijing family strains identified by Spoligotyping from 12 provinces were used in the present study. The percentage of ancestral strain varies from 3.3% to

16.5% as shown in Table 28. GuangXi province in which the higher percentage of ancestral strains is found is also the province with the lower percentage of Beijing strains (55%).

Table 28 Percentage of Beijing family and RD181 non-deletion

No	Region	No. of Beijing family stains (%)	RD181 deletion (%)	RD181 non-deletion (%)
1	Gansu	149 (85.1)	122 (89.7)	14 (10.3)
2	Hunan	68 (70.1)	58 (96.7)	2 (3.3)
3	Jilin	298 (89.5)	252 (85.1)	44 (14.9)
4	Shannxi	97 (80.8)	75 (89.3)	9 (10.7)
5	Henan	67 (83.8)	58 (87.9)	8 (12.1)
6	Tibet	194 (90.2)	179 (95.7)	8 (4.3)
7	Xinjiang	135 (66.2)	104 (89.7)	12 (10.3)
8	Guangxi	115 (55.3)	96 (83.5)	19 (16.5)
9	Zhejiang	64 (64.6)	63 (98.3)	1 (1.7)
10	Beijing	113 (93.4)	103 (93.6)	7 (6.4)
11	Fujian	100 (56.2)	82 (91.1)	8 (8.9)
12	Sichuan	66 (61.7)	60 (93.7)	4 (6.3)
13	Total	1,466 (75.68)	1,073 (88.8)	136 (11.2)

3.2.6. The non-Beijing strains

Spoligotyping analysis

For a total of 342 non-Beijing isolates the spoligotype was of good quality and the data was analysed by clustering into BioNumerics using the categorical coefficient and ward (Figure 30). Clusters of isolates with similar spoligotypes are observed. In many cases the spoligotype has been already observed and major lineages can be recognized (**ANNEX 6**). The majority belong to the T and H families. A small group shows the spoligotype of the CAS lineage.

There are some inconsistent patterns which probably correspond to the presence of two strains in the sample, one belonging to the Beijing family and a non-Beijing family strain.

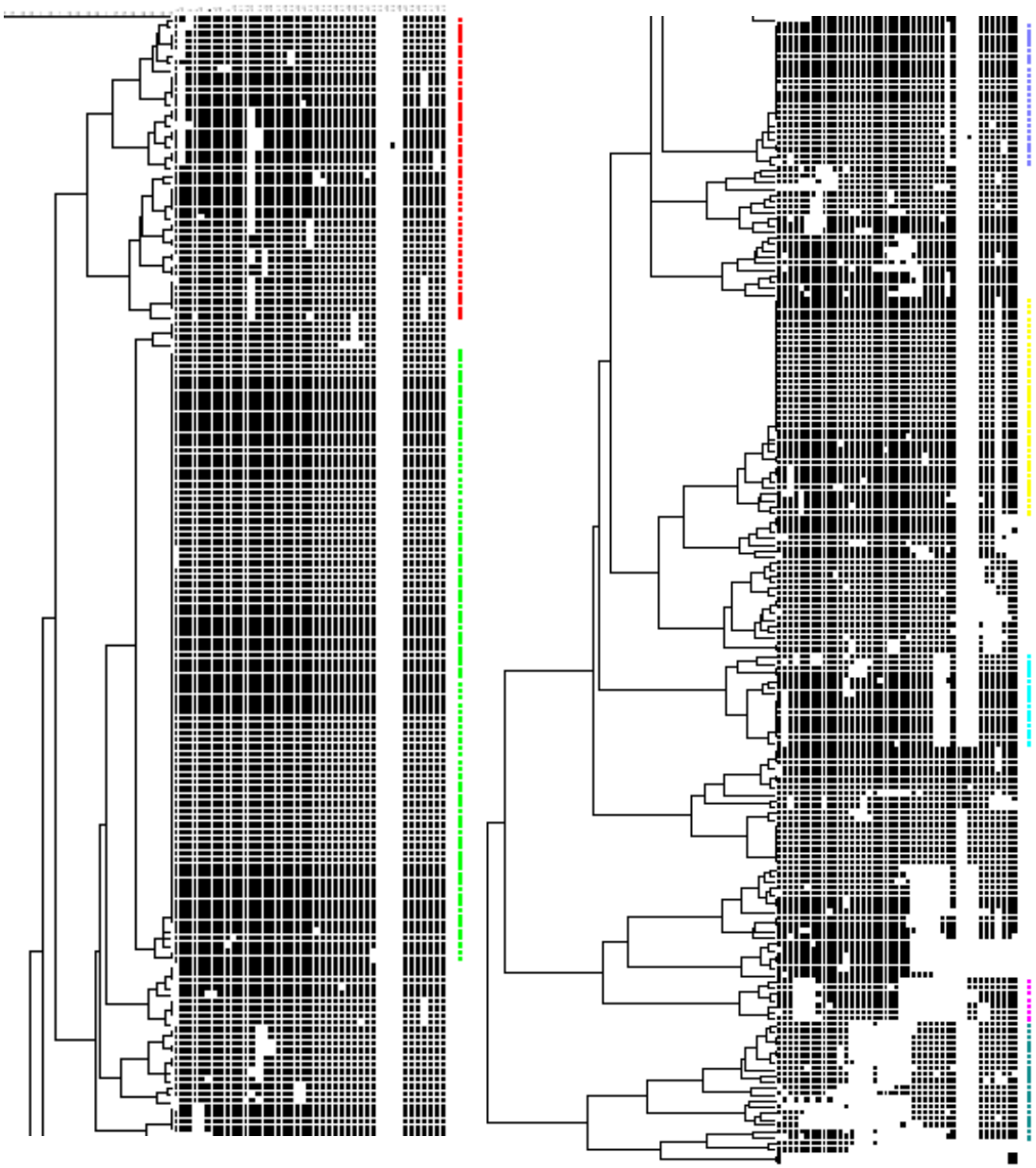


Figure 30 Clustering of non-beijing strains using the spoligotype

Large groups are shown with colors (red = T3 ; yellow = T2, green = Ural, blue=H4, pink = CAS, turquoise=undefined)

Repartition of isolates into three main groups by MLVA15

The same collection of samples was typed by MLVA15 and a clustering was performed using the categorical coefficient and UPGMA (Figure 31).

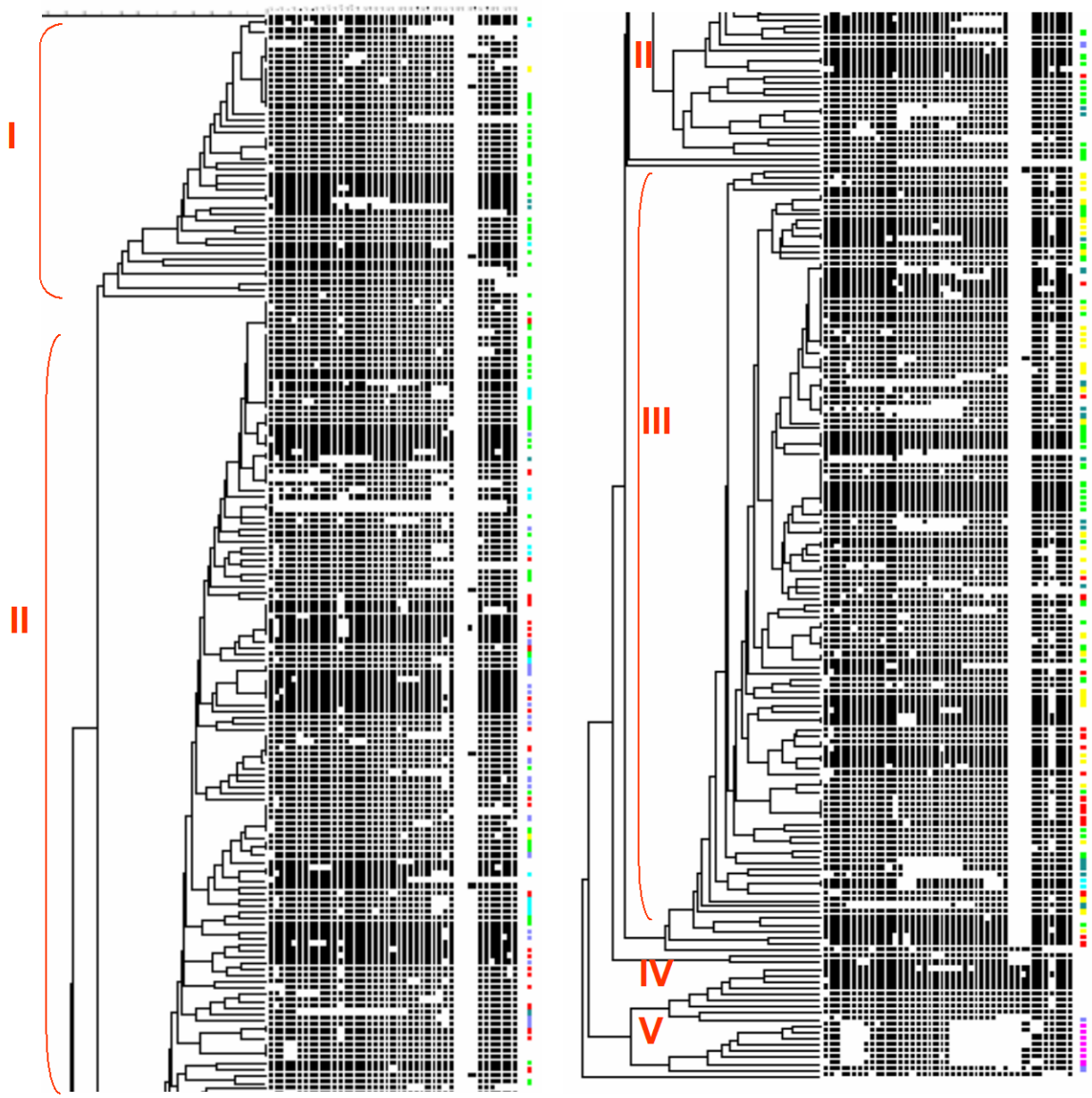


Figure 31 Clustering of MLVA15 data. The code color is that defined by spoligotypes

Five groups can be observed, one of them (group V) corresponding to the CAS lineage. GroupIV could correspond to an ancestor for the Beijing family (see discussion). The other spoligotype clusters are now scattered although there remain some aggregation.

New colors were assigned to the MLVA15 groups and a minimum spanning tree was performed as shown in Figure 32, in red green and blue, encompassing the majority of isolates in the total collection of strains. The previously observed China 2 and China 3 groups are shown in green and red respectively. By comparison with isolates from different part of the world, China 2 and China 3 seem to be majoritarily found in China and represent about 8.16% and 9.21% of the total strains respectively. A third group colored in blue (cluster I in Figure 30) belongs to the Modern family and is found in other countries worldwide.

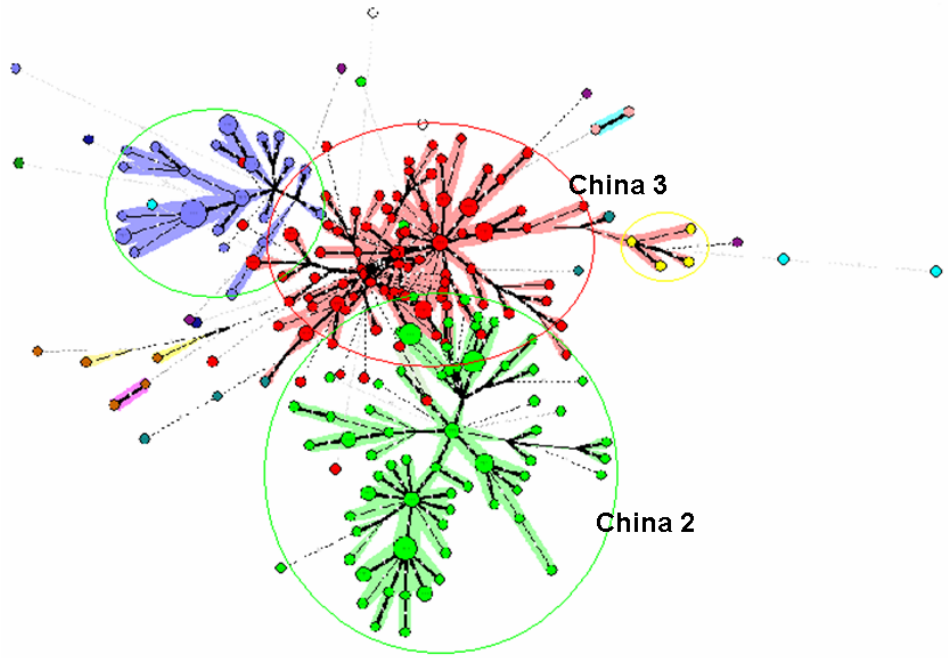


Figure 32 Minimum Spanning Tree for 344 *M. tuberculosis* strains of non-Beijing family identified by Spoligotyping

A combined analysis of the two fingerprints (spoligotype and MLVA5) was performed using the categorical coefficient and UPGMA. A new dendrogram was generated which did not improve the clustering.

3.2.7. Repartition of the different families in each province

Clustering by MLVA15 of all isolates was performed for each province separately in order to better evaluate the importance of each major group and to show the existence of new groups (Table 30, Figure 33~Figure 40).

Table 29 Distribution of *M. tuberculosis* China 2 and China 3 in nine provinces, analyzed by MLVA

No	Province	Strains tested	China 2 strains	Rate (%)	China 3 strains	Rate (%)
1	Xizang	204	6	2.94	7	3.43
2	Xinjiang	176	11	6.25	17	9.66
3	Ji Lin	319	14	4.39	14	4.39
4	Gansu	72	2	2.78	7	9.72
5	Shannxi	93	4	4.30	8	8.60
6	Guangxi	202	30	14.85	34	16.83
7	Sichuan	102	19	18.63	14	13.73
8	Hunan	91	10	10.99	10	10.99
9	Zhejiang	163	20	12.27	20	12.27
10	Total	1422	116	8.16	131	9.21

The repartition of families in 9 provinces is shown by a minimum spanning tree analysis in Figure 33 to Figure 41.

In the Xinjiang autonomous region, in addition to the three major groups described above, a group of strains which spoligotype is characteristic of the CAS clade is observed.

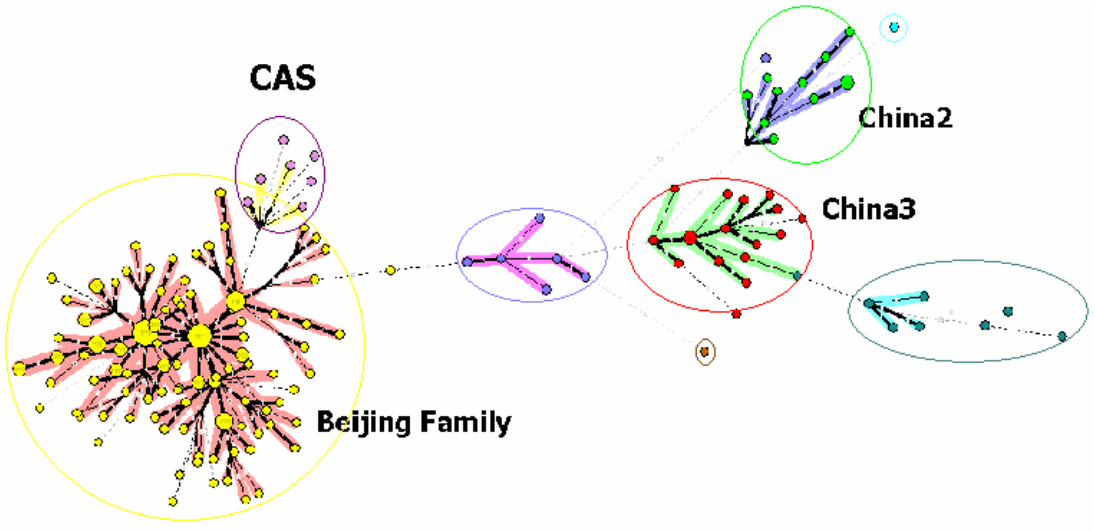


Figure 33 Minimum Spanning Tree for 202 *M. tuberculosis* strains isolated from Xinjiang Autonomous Region

In Xizang, Gansu, Jilin, Shannxi, Sichuan, Hunan and Zhejiang, the majority of non-Beijing strains is displayed in three groups.

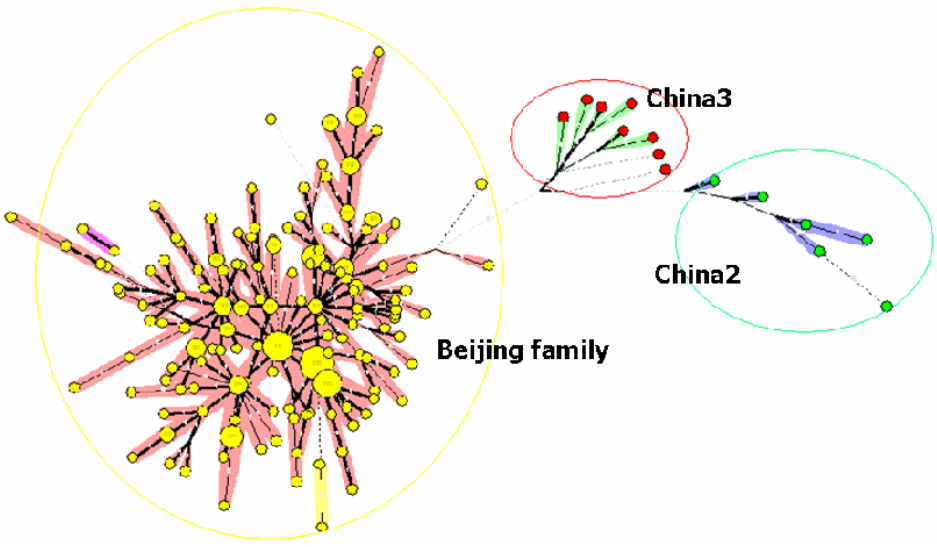


Figure 34 Minimum Spanning Tree for 204 *M. tuberculosis* strains isolated from Xizang (Tibet) Autonomous Region

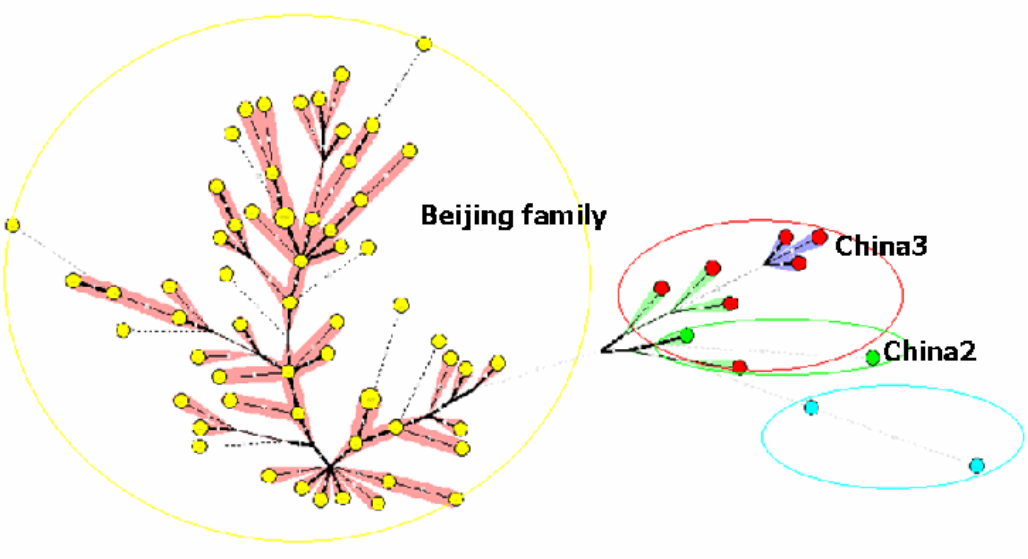


Figure 35 Minimum Spanning Tree for 72 *M. tuberculosis* strains isolated from Gansu province

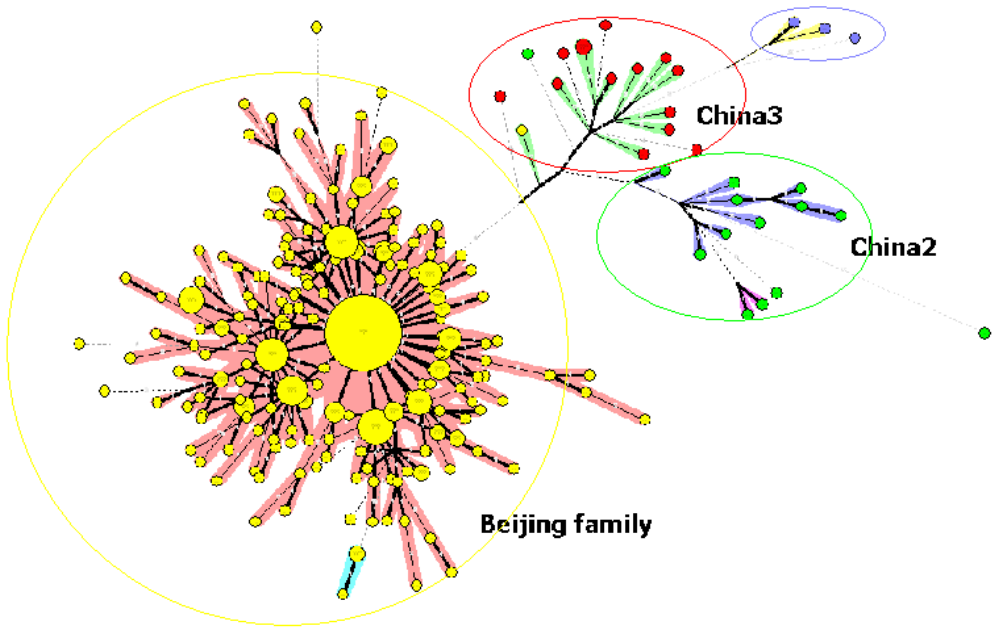


Figure 36 Minimum Spanning Tree for 319 *M. tuberculosis* strains isolated from Ji Lin province

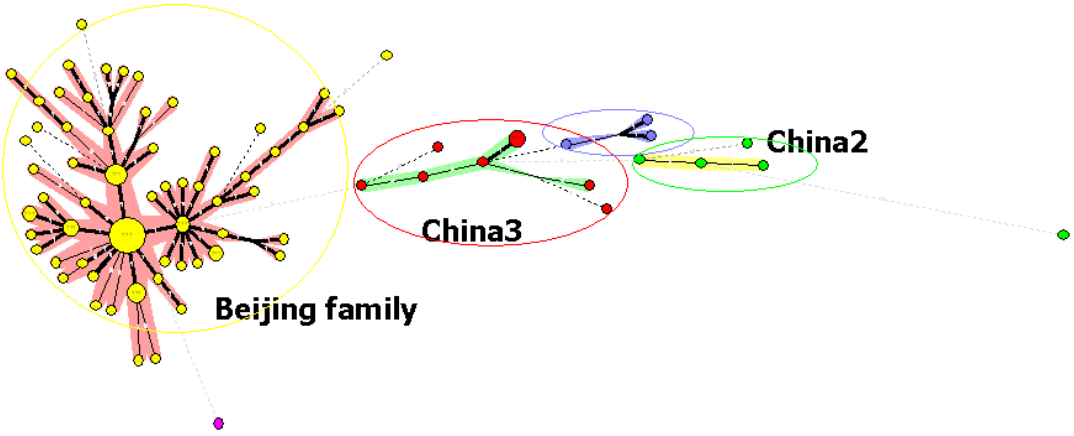


Figure 37 Minimum Spanning Tree for 93 *M. tuberculosis* strains isolated from Shannxi province

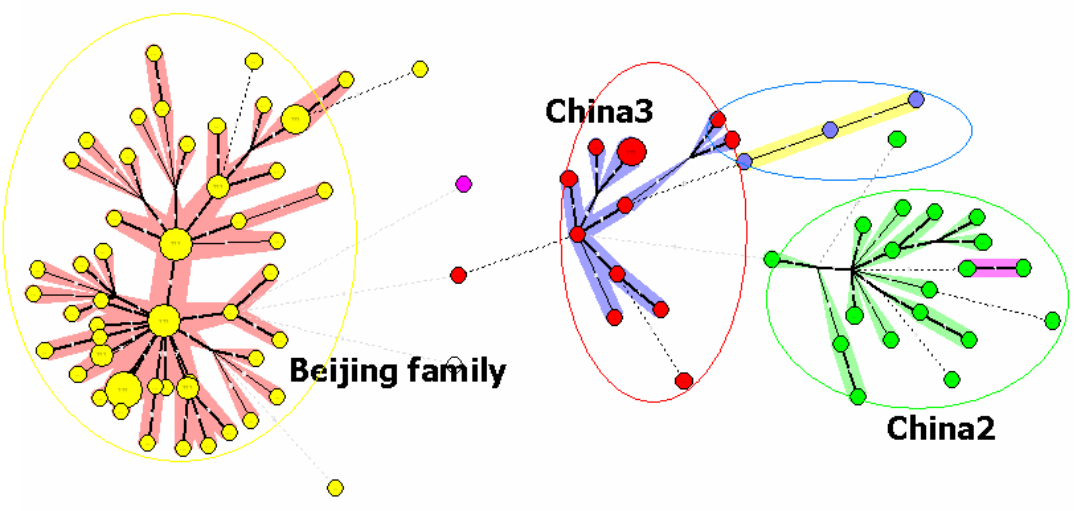


Figure 38 Minimum Spanning Tree for 102 *M. tuberculosis* strains isolated from Sichuan province

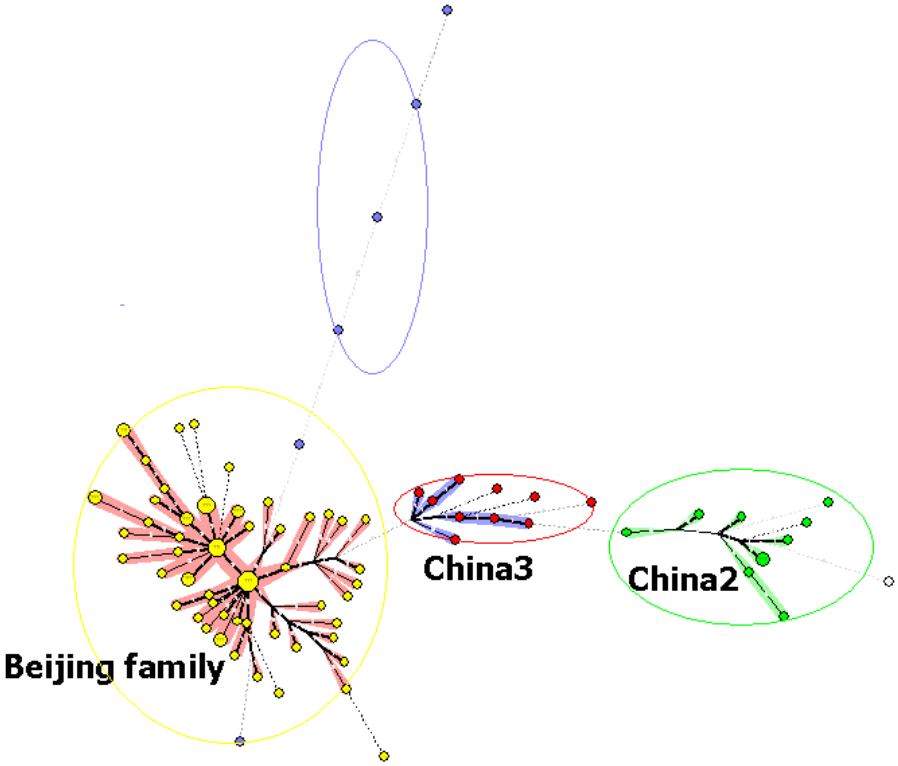


Figure 39 Minimum Spanning Tree for 91 *M. tuberculosis* strains isolated from Hunan province

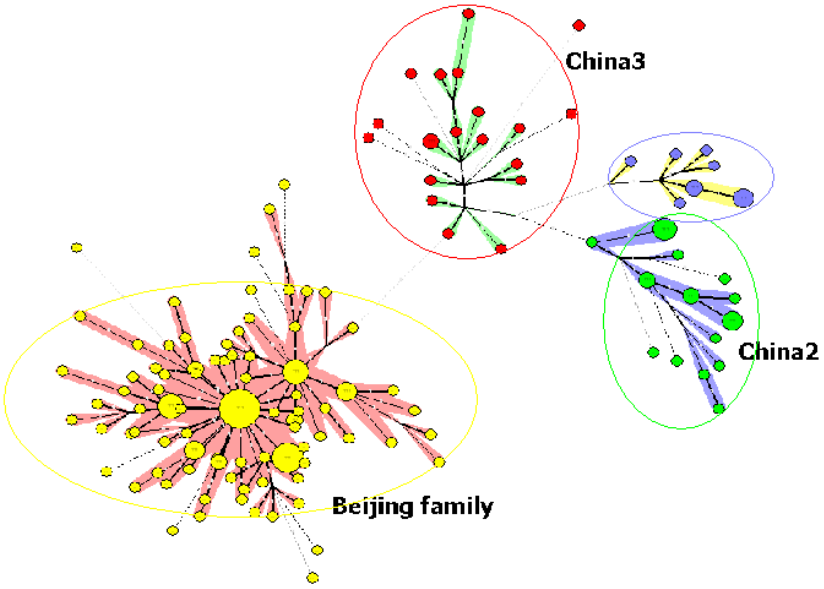


Figure 40 Minimum Spanning Tree for 163 *M. tuberculosis* strains isolated from Zhejiang province

The Guang Xi province possess the lowest percentage of Beijing-family strains and the larger diversity in non-Beijing strains with the existence of a group of isolates phylogenetically related to Beijing and CAS families and which could represent an ancestor to these two families.

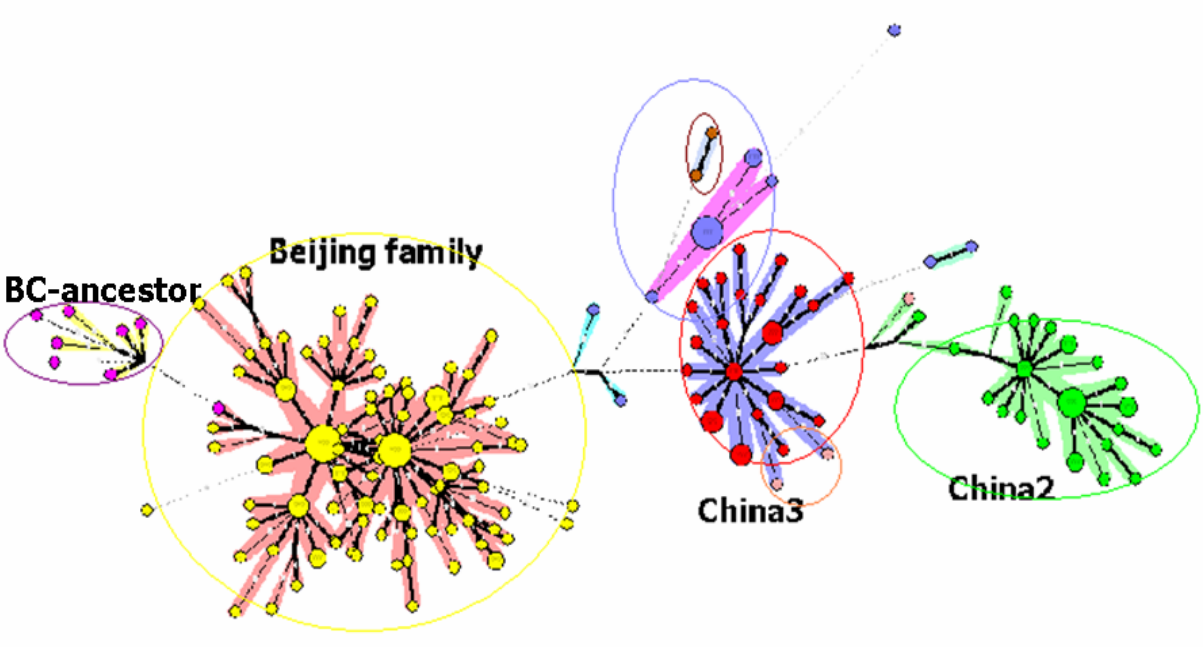


Figure 41 Minimum Spanning Tree for 202 *M. tuberculosis* strains isolated from Guangxi Autonomous Region

The MLVA15 clustering places the BC-ancestor group in an ancestral position as compared to the Beijing and CAS groups (Figure 42), and the spoligotype is consistent with this hypoyhesis (Figure 43).

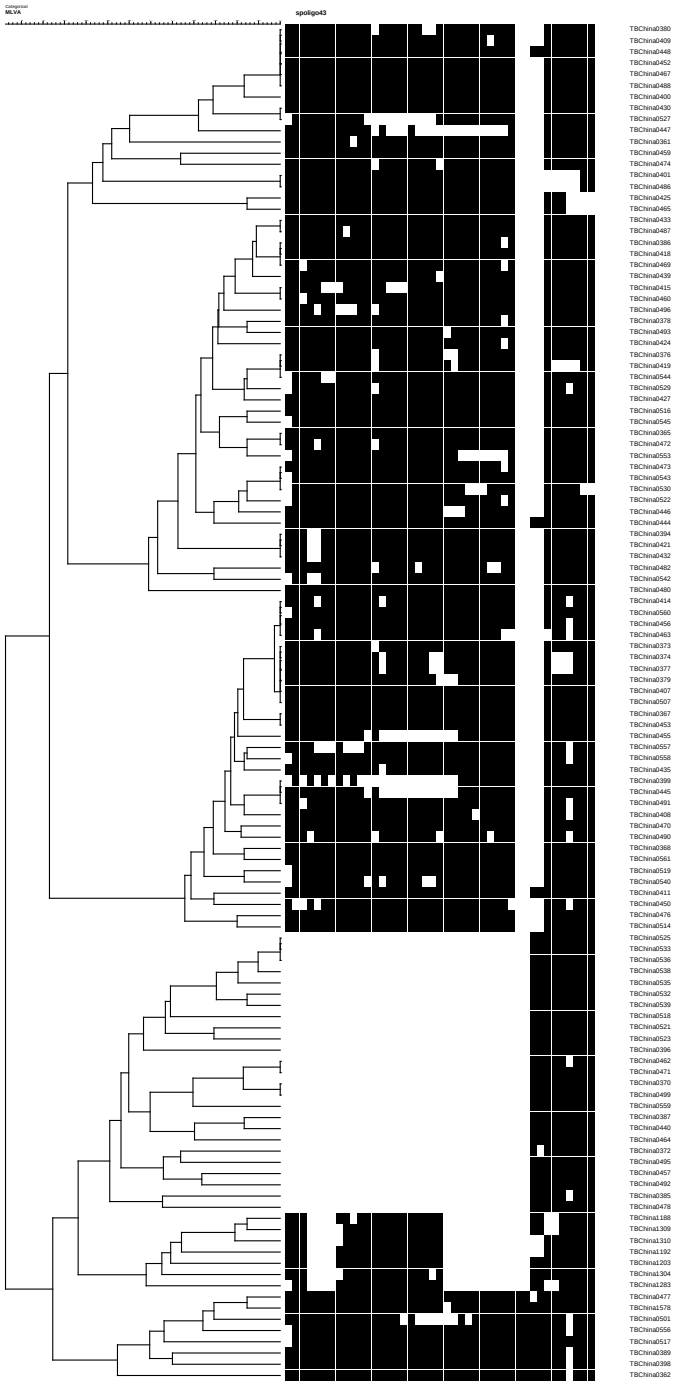


Figure 42 MLVA15 clustering of a selection of isolates from GangXi and Zhejiang provinces showing the ancestral position of BC-ancestor in pink as compared to the CAS group in purple and Beijing in yellow.

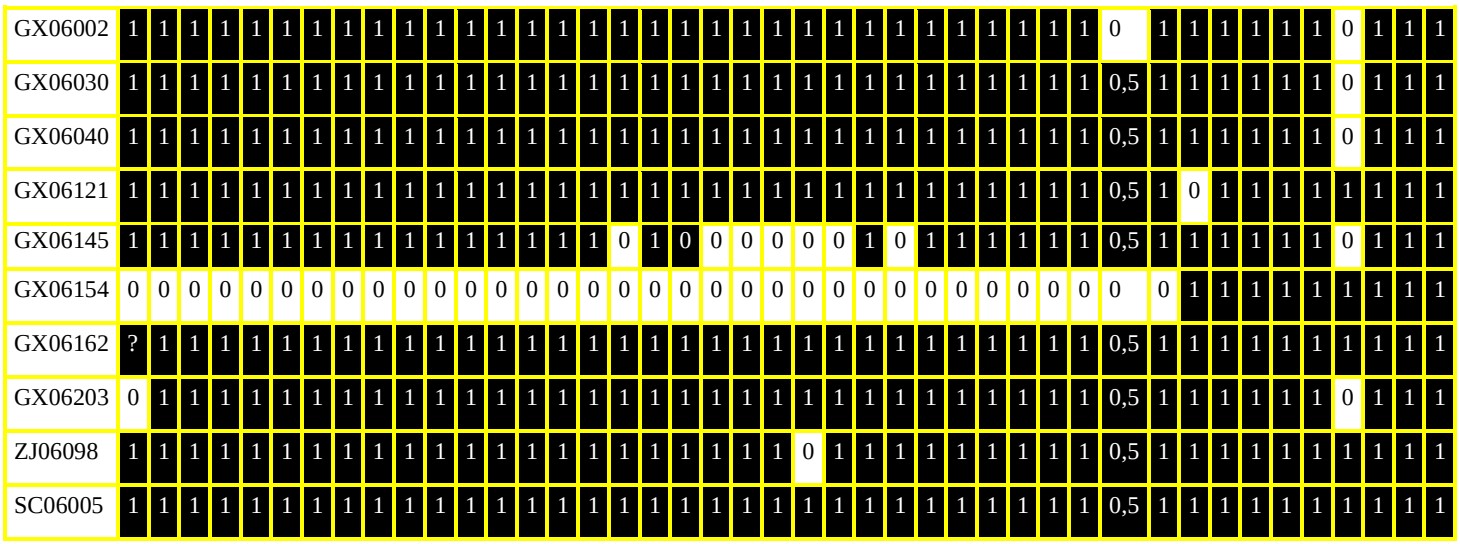


Figure 43 Spoligotype of a selection of isolates from GangXi, ZheJiang and SiChuan provinces including one Beijing-family sample and 9 BC-ancestor isolates.

An analysis of different regions of deletions was performed showing that the TbD1 region is absent and that RD105 is deleted. A preliminary study of the DR locus by PCR shows the presence of a group of spacers not detected by spoligotyping and which are present in Beijing family strains. This strongly suggests that this group of strain might represent the ancestor of the Beijing family.

4. DISCUSSION AND CONCLUSION

TB is the greatest killer of human-being in history. In China, 200 BC, Zhang Zhong-jing who was a famous doctor of the Han Dynasty described “phtisis” systematically in his Chinese traditional medicine book called ”Textbook Synopsis of PrescRFPtions of the Golden Chamber“. He first thought there were a relationship between “phtisis” and “scrofula”.

In 1882, Robert Koch made the landmark discovery that TB is caused by an infectious agent, *M. tuberculosis* introducing the possibility that antimicrobial agents could be developed to combat this age-old scourge (Koch, 1882). Morbidity due to TB steadily declined during the 20th century in the developed world, aided by improvements in both living conditions and public health practices, as well as widespread vaccination with Bacille-Calmette-Guerin (BCG), and the use of antibiotics (Smith, 2003). But today, despite the availability of effective antituberculosis chemotherapy for over 50 years, TB remains a major global health problem. As the rates of TB infection have fallen dramatically in advanced countries in the past century, developing countries now bear over 90% of all cases globally. In fact, there are more cases of TB today than ever recorded. While much is known about the epidemiology of TB, key questions have eluded classical epidemiologists for decades. These include the current rates of active transmission by differentiating disease due to recent or previous infection; the determination of whether recurrent tuberculosis is attributable to exogenous reinfection; whether all *M. tuberculosis* strains exert similar epidemiologic characteristics in populations; and an understanding of transmission dynamics on a population- or group-specific level, as well as in identifying extensive transmission or outbreaks from what appear to be sporadic, epidemiologically unrelated cases. Molecular epidemiologic methods have facilitated studies that address some of these very questions.

4.1. drug-resistance research

In recent years, the epidemiological situation of tuberculosis is presenting a tendency toward a raise. One of the main causes is that the disease caused by drug-resistant *M. tuberculosis* strains, especially the multidrug-resistant strains, is prevailing. The main problem in TB control is the drug resistance since there is no cure for some MDR-TB and XDR-TB strains. The fear is that such strains could spread around the world, stressing the need for additional control measures, such as new diagnostic methods, better drugs for treatment, and a more effective vaccine. The situation of drug resistance in China is much worse. According to the statistics, the total rate of drug-resistance

was 27.8%, and the multi drug-resistance rate reached as high as 10.7%. Because of this the tuberculosis prevalence rate and the mortality rate stay at a high level. So at present there is an emergency to develop rapid and precise methods for drug resistance diagnosis, in order to achieve the national plan for tuberculosis prevention and control and to develop an efficient chemotherapy to the disease.

The methods for the detection of drug resistance include both phenotypic and genotypic approaches. Phenotypic methods are in general simpler to perform and might be closer to implementation on a routine basis in clinical mycobacteriology laboratories. However these methods necessitate the cultivation of *M. tuberculosis* in the presence of antibiotics, demonstrating obvious limitations, and higher biohazard risk. Genotypic methods have the advantage of a shorter turnaround time, no need for growth of the organism, the possibility of direct application on clinical samples, lower biohazard risks, and the feasibility of automation.

4.1.1. Conventional phenotypic resistance test

According to the global report (WHO, 2004a) produced by the WHO/IUATLD Global Project on Anti-tuberculosis Drug Resistance Surveillance, in new TB cases with data available from 75 settings (55,779 patients) the median prevalence of resistance to at least one drug (any resistance) was 10.2 %, including: SM, 6.3 %; INH, 5.9 %; RFP, 1.4 %; and EMB, 0.8 %. Prevalence of MDR-TB was also measured; it was 1.1 % in Kazakhstan and Israel, and the highest prevalence were 13.7 % in Tomsk Oblast (Russian Federation), 13.2 % in Karakalpakstan (Uzbekistan), 12.2 % in Estonia, 10.4 % in Liaoning Province (China), 9.4 % in Lithuania, 9.3 % in Latvia, 7.8 % in Henan Province (China), 6.6 % in Ecuador. Among previously treated cases with data available from 66 settings (8,405 patients) the median prevalence of resistance to at least one drug (any resistance) was 18.4 %, including INH, 14.4 %; SM, 11.4 %; RFP, 8.7 %; and EMB, 3.5 %. The median prevalence of MDR-TB was 7.0 %. The highest prevalence were 58.3 % in Oman and 56.4 % in Kazakhstan. Countries of the former Soviet Union had a median prevalence of resistance to the four drugs of 30 %, compared with 1.3% in all other settings. However, these data should be interpreted with caution given the small number of subjects tested in some settings.

In the present study, we have used conventional phenotypic methods including the proportion method, the resistance ratio method and the absolute concentration method (Canetti et al., 1963;

Canetti et al., 1969; Kent and Kubica, 1985) to test drug sensitivity on a total of 2018 clinical *M. tuberculosis* strains isolated from patients with tuberculosis collected from 10 provinces in China.

In China the prevalence of drug resistance is the highest in the world. WHO has confirmed that China is one of the countries of high burden of tuberculosis and estimated that a third of the world's cases of MDR tuberculosis are in China, even though the country has only 17% of the global burden of tuberculosis. Unfortunately this tends has not been controlled well. According to the results of the present study, the prevalence is increasing very much; the total rate of drug resistance (44.70%), prevalence of resistance to single drug 15.51% and prevalence of MDR-TB (29.19%) are much higher than that of the 2000 Nationwide Random Survey for the Epidemiology of TB in China (except Taiwan, Hongkong and Macao) (Tuberculosis, 2002) and also much higher than that in some recent reports (Aziz et al.; Khue et al., 2007; van Klingeren et al., 2007). So the drug resistance has reached a most dangerous level, and a lot of work must be done for drug resistant TB control and prevention.

There are large differences from one province to another for resistance to single or multiple drugs. This ranges from about 15% MDR strains in GuangXi province to more than 50% in Xizang province. The reason for this might be related to the strain genotype or to difference in access to the drugs.

4.1.2. Genotypic resistance test

At present more and more research work on drug resistance of *M. tuberculosis* has been reported. The drug targets and mechanisms of resistance to first-line and several second-line antituberculosis agents have been deciphered (Maus et al., 2005) (Ramaswamy and Musser, 1998) (Rengarajan et al., 2004). Generally *M. tuberculosis* acquires drug resistance by small deletions, or insertions in specific chromosomal loci. This specificity of *M. tuberculosis* drug resistance, coupled with fast and efficient DNA sequencing methods, makes studying drug resistance highly amenable for molecular epidemiologic investigations. In general molecular epidemiologic studies on drug resistance identify genotype-specific mutations, association of specific mutations with phenotypic resistance, prevalence of specific mutations in a population and patient risk factors (e.g., HIV) for acquiring resistance.

Rifampicin (RFP) is one of the most potent antituberculosis drugs; therefore, resistance to RFP often results in high clinical relapse rate, particularly if RFP resistance is associated with resistance

to other antituberculosis drugs, such as Isoniazid (INH). Moreover, more than 90% of RFP-resistant isolates are also resistant to Isoniazid; therefore, detection of RFP resistance could also identify MDR strains. At present there are many research works reported on the mechanism of RFP-resistance (Kazue et al., 1999; Cavusoglu et al., 2002; Srivastava et al., 2004). It has been shown that more than 95% of RFP-resistant *M. tuberculosis* strains have a mutation within the 81-bp "hot-spot region" (Cluster I in Figure 16, codons from 507 to 533) of the RNA polymerase B subunit (*rpo*B) gene. It has been found that there are differences in the mutation rate in the "hot-spot region" in different countries and regions (Morgan et al., 2005). In this report the *rpo*B gene mutations observed are consistent with results from other studies in China and in other countries. The codons 531, 526, 516 and 511 sites are the most common positions of mutation, and the proportion of them are 53.61%, 27.71%, 9.64%, and 6.02%, respectively. Meanwhile, mutations also exist in some sporadic sites, such as 515, 513, 510, 509, 512, 514, 522 and so on. This result is similar to that recently reported by Afanas'ev MV *et al.* (Afanas'ev et al., 2007).

There are combinative mutations and deletion and insertion of basic group. Moreover, reverse dot blot hybridization based on PCR can be used to detect gene multi-mutations at one time, and its results can be read directly with naked eyes easily. The coincidence of RDH and DNA sequencing can achieve 97.67% and there is no significant difference between the two methods by Chi-square statistics test ($P>0.05$). This method is so rapid, sensitive, and simple that it can be applied for screening the RFP-resistance of MTB clinical isolates. Furthermore, following the technical development of gene chips, RDH has extensive perspective of clinical rapid diagnosis for RFP-resistance.

The mechanism of MTB resistance to INH has been demonstrated primarily with the development of molecular biotechnological methods. It was verified that INH-resistant MTB strains have a mutation principally in one of the following genes: *kat*G, pre-*inh*A, *inh*A, *ndh* and *oxy*R-*ahp*C genes (Zhang et al., 1993; Banerjee and Sugantino, 1998; Miesel et al., 1998; Fressatti-Cardoso et al., 2004).

To investigate the mutations in drug resistance genes correlated with INH (*kat*G, pre-*inh*A, *inh*A, *ndh*, *oxy*R-*ahp*C) and RFP (*rpo*B) resistance in *M. tuberculosis*, and to identify the characteristic and the distribution of the mutations, 300 clinical isolates of *M. tuberculosis* from patients with tuberculosis in Xizang, Hunan, Henan, Sichuan, Fujian, Anhui, and Shanxi provinces

of China were analysed by DNA sequencing and oligonucleotide probe reverse dot blot hybridization (RDH).

In our study by means of DNA sequencing analysis, we found that the rate of mutation in the "hot spot mutation area" of the *kat*G gene was 91.06% (163/179). There were 11 kinds of mutation occurring and 7 of which were not previously reported. For 131 strains (73.18%) the mutation occurred at codon 315 including 4 kinds of mutation types, but the main mutation kind was AGC (Ser) to ACC (Thr), which occurred in 123 strains (93.89%). However the mutation at codon 463 showed no correlation with INH-resistance. For 18 (10.06%) INH-resistance strains mutations occurred in *oxy*R-*ahp*C gene, of which 7 (3.39%) had a -9C to T mutation, 3 strains had a -5G to A mutation, 3 strains had a -14C to T change; while no mutation was found in the sensitive strains. Only 1 strain (0.56%) in INH-resistance strains showed a mutation in the *inh*A gene while none was found in the sensitive strains. For 3 (1.68%) INH-resistance strains a mutation in *ndh* was found (Glu191Gly), while no change was observed in the sensitive strains. Concerning pre-*inh*A gene, in 179 INH-resistance strains, there were 12 strains showing a change of 15C to T. Other mutations, such as -19G to A, -8T to C and 13G to C, were also found in some stains while no change was found in the sensitive strains. There was no mutation detected in *oxyR* and *ahpC* genes in any of the strains.

This research further confirmed the relationship between *M. tuberculosis* INH resistance and *kat*G, pre-*inh*A, *inh*A, *ndh*, *oxy*R-*ahp*C gene mutations. At the same time, it provides the foundation for the probe design for the next step, and simultaneously also showed that there are other mechanisms participating in the INH-resistance. This will be further investigated in the future.

4.2. Genetic diversity study of *M. tuberculosis* in China

M. tuberculosis, is part of the *M. tuberculosis* complex (MTBC), a group of closely-related slow-growing mycobacteria that include *M. tuberculosis*, *M. bovis*, *M. africanum*, *M. microti* and *M. canettii*. Until recently, it was assumed that cattle transmitted the disease to man, as the host range of *M. bovis* was much broader than that of *M. tuberculosis*. However, the generally low level of genetic variation seen in *M. tuberculosis* indicates that the total population resulted from clonal expansion following an evolutionary bottleneck, estimated to have occurred between 15

000 and 35 000 years ago (Sreevatsan et al., 1997; Brosch et al., 2002; Fabre et al., 2004; Gutierrez et al., 2005; Filliol et al., 2006).

Traditional typing methods, including phage-typing, drug sensitivity typing, cell protein electrophoresis, biochemical diversity analysis, multi-loci enzyme electrophoresis show some limitations. With the development of molecular biological techniques, the molecular genotyping methods have become the major tools in the epidemiological study of tuberculosis.

Since 1980, specific identification methods based on genetic polymorphism have gradually expanded, for example, restriction fragment length polymorphism (RFLP), the analysis of DNA fingerprint pattern, pulsed field gel electrophoresis (PFGE), the DNA polymorphism of random-PCR, DNA sequencing and gene chip, and others. Both spacer oligotyping (Spoligotyping), and multiple loci variable number tandem repeat analysis (MLVA) developed recently are based on polymerase chain reaction (PCR). These methods are easy and fast.

In this study from 2004 to 2006, 2751 *M. tuberculosis* isolates collected from 14 provinces were isolated from sputum samples and stored for future investigation. Patient information was obtained for each sample in order to perform epidemiological studies. The repartition of the samples between the two genders is correct but repartition according to age is still not perfect. The number of cases among young children (under 15 years old) is low thanks to the effect of BCG vaccination. The number of cases among old people (more than 70 years old) is also low. In the future, efforts will be made in recovering more samples from these two age groups.

At present time, 1937 isolates from 12 provinces (Beijing, Henan, Sichuan, Hunan, Zhejiang, Fujian, Jilin, Shannxi, Gansu, Xinjiang, Xizang, and Guangxi) have been studied for their genetic diversity. Molecular typing identification and genetic polymorphism of the strains were determined by looking at different polymorphic markers, DR loci (Spoligotyping), RD (Genetic deletion, Large sequence polymorphisms (LSPs), or/and tandem repeats (MLVA).

4.2.1. Geographical distribution of *M. tuberculosis* Beijing family in 12 provinces

Since 1995 van soolingen D. *et al*. reported that the Beijing genotype of *M. tuberculosis* was first recognized as highly prevalent strain in East Asia (van Soolingen et al., 1995), and in the past decade, molecular epidemiological data collected in several countries have revealed that the

genotype is widespread around the world (Bifani et al., 2002; Glynn et al., 2002; Filliol et al., 2003). Many molecular epidemiological investigations have indicated the worldwide occurrence of Beijing/W Strains of *M. tuberculosis* (Figure 44) (Glynn et al., 2002).

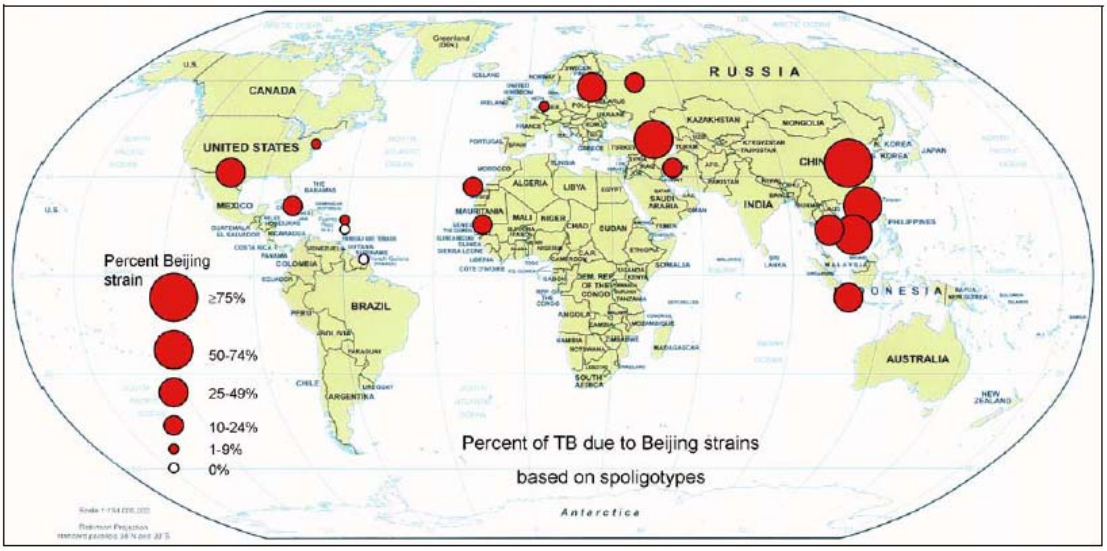


Figure 44 Percentage of tuberculosis due to Beijing strains. Data from studies based on spoligotyping (Glynn et al., 2002)

In China, Li WM, *et al.* analysed a total of 408 *M. tuberculosis* by spoligotyping in a study of "Molecular epidemiology of *Mycobacterium tuberculosis* in China: a nationwide random survey in 2000". Of the *M. tuberculosis* strains, 64.9% (265/408) were of the Beijing genotype. Using the χ^2 test, no statistically significant differences were observed in the proportion of Beijing genotype strains in patients of different sex, age or living in different areas of the country (Li et al., 2005).

In the present study, by means of spoligotyping, separation of strains into two main groups was achieved: those belonging to the Beijing family and the non-Beijing family group. The mean percentage of Beijing family strains was 75.68% (1,466/1,928), with large differences in the 12 provinces (55.3% ~ 93.4%) (Table 24), whereas in other Asian countries they represent 50% or less. According to this study, geographical contribution of *M. tuberculosis* Beijing family in China could be divided roughly into four types. The results show that the highest number of Beijing family strains are observed in the Northern China including Xizang (Tibet), the second was in center of China, e. g. between Yellow River and Yangtze River, the lowest was in southern China. The Xinjiang province in the North East apparently shows a specific distribution. There are probably

multiples reasons for this repartition linked to geographical, climatic and human differences. The very high percentage of Beijing family strains in certain part of China strongly support the hypothesis that this family might have originated in China. This might be further supported by the existence of ancestral members of this family as well as the high genetic diversity observed by MLVA.

4.2.2. The origin of the Beijing family

Large-sequence polymorphisms (LSP) are thought to mainly occur as a result of genomic deletions and rearrangements rather than through recombination following horizontal transfer (Brosch et al., 2001). The deletions are irreversible and often unique events and therefore are solid markers for constructing phylogenies (Brosch et al., 2002; Goguet de la Salmoniere et al., 2004; Tsolaki et al., 2004). One LSP (RD105) was seen in all Beijing/W strains and serve as a useful marker for the identification of this family of strains in addition to the specific spoligotype. Additional LSPs (RD142, RD150, and RD181) further divided this family into monophyletic subgroups, demonstrating a deep population structure (Tsolaki et al., 2005).

We have sought to identify ancestral Beijing strains by testing the presence/absence of the RD181 region in all the Beijing strains of the present study. The percentage of ancestral strains varies in the different provinces and interestingly the higher percentage is found in the GuangXi province, the province with the lower percentage of Beijing-family strains. This is also in this province that some isolates (accounting for 3.47% of isolates) show a spoligotype and a MLVA profile suggesting that they are phylogenetically related to Beijing-family strains . Very importantly this group of strain possess the RD105 deletion which almost certainly place them has candidate ancestors of the Beijing lineage.

In the work by Mokrousov et al. (Mokrousov et al., 2002) a distinction was made between typical (96% of the isolate in Northwestern Russia) and atypical Beijing strains on the basis of IS*6110*-RFLP analysis. Additional differences were found for the Rv3135 PPE-family gene and for the number of IS*1547* copies suggesting that atypical strains were evolutionary older. We will test whether the strains that are not deleted for RD180 possess the characteristics of atypical strains. Similarly the group of putative Beijing ancestor will be analysed in depth using other genetic markers to confirm its proximity with the Beijing-family members.

4.2.3. The age distribution of ancestral Beijing-family strains

It was important to know whether the ancestral Beijing strains might have been responsible for a large proportion of infection in the past before the BCG vaccination was generalized. For this we looked whether there was a difference in age-groups for the distribution of ancestral Beijing strains (Figure 45 and Table 30).

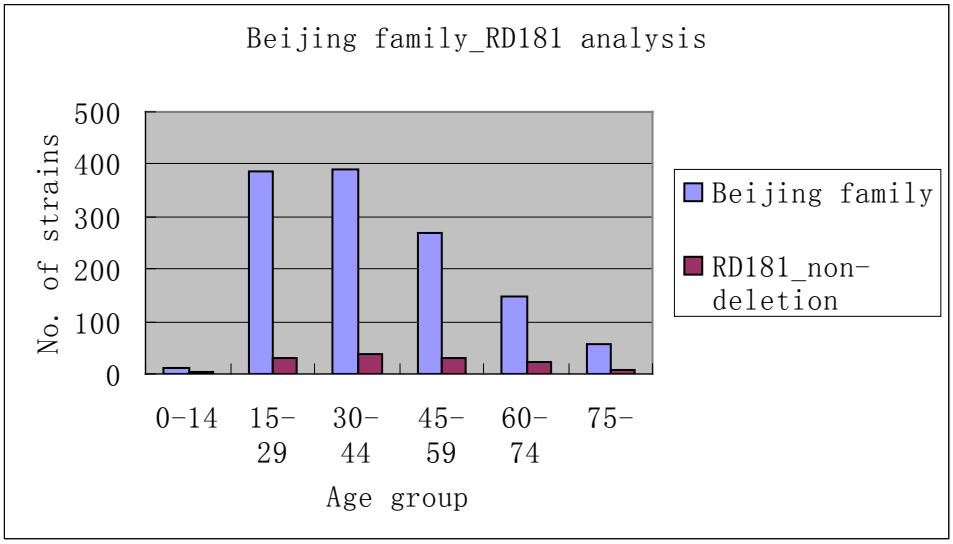


Figure 45 Distribution of Beijing family strains and of ancestral Beijing

Table 30 Percentages of ancestral Beijing family strains in different age groups

	age groups						
	0-14	15-29	30-44	45-59	60-74	75	Total
Beijing family	12	386	392	270	149	55	1264
RD181-Pos	3	32	37	32	21	8	133
Total	15	418	429	302	170	63	1397
% ancestral	20%	7,60%	8,60%	10,60%	12,30%	12.7%	9,50%

There is apparently a higher proportion of ancestral strains in the 45 to 74 years old people as compared to young adults whereas the high level of ancestral strains in children needs to be verified on a larger population.

4.2.4. Genetic diversity of Beijing-family isolates as assessed by MLVA

The presence/absence of the RD142 and RD150 deletions will be analysed in the future in order to get more information on the diversity of the Beijing family.

Using MLVA15, the large majority of Beijing isolates fall into a large homogenous group in which only one or two alleles differ between each sample. There are also some groups of isolates with identical genotypes. About 10% of isolates present more diversity, some of them possessing several rare alleles. A preliminary study using 6 additional VNTRs showed that additional informativity could be obtained although it is not clear whether this will help in understanding the relationship between the different isolates. The method of clustering currently used (categorical coefficient and UPGMA) does not take into account the rate of evolution at each VNTR. More studies need to be done to learn how to produce phylogenetic analyses in this quite homogenous family. The knowledge of other genetic markers such as the Beijing-family deletions and specific SNPs will help in constructing a phylogeny.

4.2.5. Molecular epidemiological Characteristics of non-Beijing isolates

The analysis of the non-Beijing isolates by MLVA15 and spoligotyping allowed their distribution into several large families. By comparison with isolates from different part of the world, we identified three families found in every provinces of which two seem to be over-represented in China (8.16% and 9.21% of the total strains respectively). The spoligotypes of strains from the 3 main groups correspond to the ill-defined T/H clade and most correspond to previously identified spoligotypes. However there is not clear clustering when they are placed in a MLVA dendrogram. The H-like spoligotypes apparently fall into the China 2 family (groupIII in Figure 41). The molecular biological characteristics and epidemiological significance (e.g. relationship with drug resistance, etc.) of these families will be investigated in the future.

There is obvious genetic diversity in the *M. tuberculosis* isolated from different provinces (municipality and autonomous regions) of China. Though there are three main groups including Beijing family, China 2 and China 3 in every province (municipality and autonomous regions), the

prevalence is very different in these provinces (Table 30). We found some more significant provinces (autonomous regions) where there were more polymorphisms in *M. tuberculosis* genotypes. There were six main groups in Xinjiang autonomous region, and in addition to the main prevalent genotypes (Beijing family, China 2 and China 3), we found a cluster of isolates belonging to the CAS (Central-Asian) lineage which was previously found mainly in India and in the Indian subcontinent (Bhanu et al., 2002; Singh et al., 2004). Singh UB *et al* recently reported the genetic biodiversity of 540 *M. tuberculosis* isolates in India evaluated by spoligotyping. The results showed that the most predominant clades among tuberculosis isolates were Central Asian (CAS) and East-African Indian (EAI) with shared-types (ST) ST26 and ST11 alone being responsible for 34% of all TB cases. In the North, the CAS family predominated (Singh et al., 2007). Concerning the CAS strains found in Xinjiang region of China additional investigation (e.g. epidemiological, clinical, molecular epidemiological, etc.) need to be done to learn more about their diversity. We can consider that CAS strains were transported by trade, tour, migration, etc. from India and in the Indian subcontinent. Among 1937 strains from 12 provinces (municipality and autonomous regions) of China analysed in this study, no CAS strains were found in the other regions.

Another remarkable observation concerns the Guangxi autonomous region located in southwest China where there was the highest degree of polymorphism in *M. tuberculosis*. Of 202 *M. tuberculosis* clinical isolates there were seven groups. The main prevalent genotypes were also Beijing family, China 2 and China 3, but the percentage of genotypes was obviously different from other part of China (Table 30). The percentage of Beijing family was the lowest (55.3%), China 3 was highest (16.83%) and China 2 was the second (14.85%) only little lower than that in Sichuan province (16.83%). The discovery of a new genotype found almost exclusively in this region and possibly representing an ancestor to the Beijing family, is of high importance as discussed above. Flores *et al.* have described strains originating from Vietnam, Laos and China which may belong to the same group (Flores et al., 2007).

Our present hypothesis is that the Southern part of China is where the ancestral Beijing strain first emerged after the first *M. tuberculosis* strains were brought from Africa. This lineage evolved to produce a more modern Beijing lineage, probably showing a higher infectiosity and which spread over the whole country. Subsequently, additional, modern lineages, corresponding to the T/H clades arrived again from the South and East coast and harbors. These clones have started spreading and were founders of two new groups called here China2 and China3. These might now be

progressively replacing the Beijing lineage. In remote places like the Northern part of China or Tibet, the Beijing lineage is still most highly prevalent. On the contrary in the Southern part other lineages are gaining importance. Of course, this view, which is the opposite of the common view of an emerging Beijing lineage, needs to be further tested by more extended typing, across more provinces, together with more focus on some provinces.

4.3. Association of *M. tuberculosis* Beijing genotype with BCG vaccination and drug-resistance

In the past decade, a lot of attention has been given to the emergence of the Beijing family and its spreading worldwide. In particular the association of *M. tuberculosis* Beijing genotype with BCG vaccination and drug-resistance is of major interest to a large number of research groups. The large epidemiological study that we are performing in China may help to shed some light on these problems.

4.3.1. Role of BCG vaccination in the dissemination of the Beijing family

According to the study of van Soolingen *et al*, the results indicated that strains of the Beijing family recently expanded from a single ancestor which had a selective advantage. So they speculated that long-term *Mycobacterium bovis* BCG vaccination may be one of the selective forces implicated in the successful spread of the Beijing genotype (van Soolingen et al., 1995). It was hypothesized that the *Mycobacterium bovis* BCG-induced immunological defense may not protect against infection by strains of the Beijing family (van Soolingen et al., 1995). However in other studies from Vietnam and Indonesia the results showed that there was no association with BCG vaccination and the percentage of Beijing-family strains (Anh et al., 2000; van Crevel et al., 2001).

In Xizang (Tibet) municipality it was possible to check this hypothesis because some of the population is not vaccinated. We found that there was no significant difference in the two groups whether BCG was used or not. In Beijing family strains, there were 87 strains (46.28%) from BCG vaccinated patients, while 101 strains (53.72%) were from non-vaccinated patients. There was no significant difference in statistics (χ^2=0.012，P>0.05). This strongly suggests that in China there might be no relationship between the Beijing family of *M. tuberculosis* and BCG vaccination.

4.3.2. The role of Drug-resistance in the dissemination of the Beijing family of *M. tuberculosis.*

It has been hypothesized that strains of the Beijing family might become drug-resistant with a high efficiency, possibly through the existence of mutator genes, thus implying that drug-resistance was a factor enhancing the emergence of this family (Rad et al., 2003). Many outbreaks of multidrug-resistant (MDR) TB, with poor response to treatment and high disease and death rates, have been reported. For example, in the United States, large outbreaks of MDR TB were caused by strains of the W family, which form a minor subgroup of the *M. tuberculosis* Beijing family (Bifani et al., 1999; Bifani et al., 2001; Bifani et al., 2002). Some results showed that the drug-resistant prevalence of Beijing family strains was higher than that of non-Beijing family strains (Anh et al., 2000; Kruuner et al., 2001), whereas some reported that there was no significance between the two genotypes (van Crevel et al., 2001; Laserson et al., 2000), and Chan MY, *et al.* reported that the drug-resistant prevalence of Beijing family strains was lower than that of non-Beijing family strains (Chan et al., 2001).

In the present study it was possible to check this hypothesis with the results of drug susceptibility testing on 1286 *M. tuberculosis* clinical isolates from 10 provinces of China.

Overall, the prevalence of single drug-resistance and MDR in 12 provinces (municipality and autonomous regions) was 12.69% and 34.12% in Beijing family strains and 9.77% and 32.18% in non-Beijing family strains respectively, which is statistically not different. Therefore we believe from the results of our study that there is no correlation between Beijing family and drug resistance. Therefore it is important to understand why there is such a high number of drug-resistant strains in China. It might be that the other genotypes, China 2 and China 3 present also a high capacity to become resistant as suggested for the Beijing lineage. This can be tested experimentally or by analysing the characteristics of the mutator genes. Another probably important reason, is that antibiotics were not always used properly and this will require further control on the way treatments are delivered in particular in rural area.

4.4. Future work

In this study, we have described some valuable results, such as the polymorphism of Chinese *M. tuberculosis* genotypes, different geographical distribution of the Beijing family, association of *M.*

tuberculosis Beijing genotype with BCG vaccination and drug-resistance, new genotypes revealed in China, etc., but it is only the beginning of the ambitious epidemiology project that we wish to achieve. We have investigated strains in 12 provinces (municipality and autonomous regions) representing less than half the Chinese administrative regions, and there are some limitations, for example, in the sampling of strains from some provinces. Therefore a lot of work still need to be done in the future.

Molecular epidemiological investigation on *M. tuberculosis* in the remaining provinces (municipality and autonomous regions) by means of different genotyping methods described here, will be performed. As shown in the present study we will explore the characteristics of new families, the characteristics and distribution of Beijing genotypes, the association of *M. tuberculosis* Beijing genotype with BCG vaccination and drug-resistance.

We will select some particular regions to perform studies in more detail, and on a larger scale, such as Xinjiang and Guangxi provinces in which more diversity was observed.

We need also to compare our results to those of other studies worldwide to better define whether we see in China emerging families of strains or whether these were previously encountered elsewhere. Our own database of about 600 genotypes does not cover the entire world.

In the recent past genotyping methods for studying *M. tuberculosis* have been developing fast. For example, by using new spacer oligonucleotides, differentiation and interpretability of spoligotyping has been improved (van der Zanden et al., 2002). Automated typing based on VNTRs combining analysis of multiplex PCRs on a fluorescence-based DNA analyzer with computerized automation of the genotyping is being used. As website are being set up for the analysis of *M. tuberculosis* MIRU-VNTR genotypes via the Internet, this opens the way for global epidemiological surveillance of tuberculosis and should lead to novel insights into the evolutionary and population genetics of this major pathogen (Supply et al., 2001).

Finally, will are presently finishing the genome sequencing of two clinical strains of Beijing family, one is drug-resistant to four drugs (INH, FFP, EMB and SM), the other is sensitive to the four drugs. Using this data, we will develop studies of genetic markers for Chinese *M. tuberculosis* strains and new methods, especially automated typing method.

Additional studies will be performed on the new main genetic groups found in China (China 2, China3) and on the putative Beijing ancestor. In particular we will test in the Beijing CDC the virulence of these strains using macrophage cultures, and the immune response of mice. We will

experimentally analyse whether these lineages have a high propensity to become drug-resistant. It will probably be of interest to sequence the genome of a member of each of these lineages.

We still need to analyse the patient information data in order to know whether there is a link between the genotypes we have characterised, the patient situation and the course of the disease. This is still not achieved because we need to recover more information and to enter the data in the database presently running with BioNumerics in Orsay. This will require more exchanges between the two participating laboratories to develop this part of the study, mostly using bio-informatics and statistics.

4.5. Conclusion

The present study was conducted in order to understand the situation of drug-resistance and the genotype distribution of *M. tuberculosis* in China, by performing a large collaborative project between the Chinese CDC in Beijing and the IGM in Université Paris-Sud Orsay. The project started in 2003, was helped by an EEC contribution for exchanges between the two laboratories. The results described in this thesis is only a part of what needs to be done to answer all the questions that remain, concerning the emergence of virulent *M. tuberculosis* lineages and the way to fight them.

We hope in the future to identify *M. tuberculosis* characteristics that will help in better define a vaccine and treatments against tuberculosis, in China and in the rest of the world.

ANNEXES

ANNEX 1 Approval from the Ethics Committee of National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention.

ANNEX 2 Patient informed consent form.

ANNEX 3 Subject information sheet

ANNEX 4 Epidemiological questionnaire. Table of Examination Results of Patient with TB and its Samples Preservation

ANNEX 5 VNTR markers for MLVA

ANNEX 6 Spoligotyping: the 43 sequences of oligo probes

Annex 1: approval from the Ethics committee of ICDC

中国疾病预防控制中心传染病预防控制所

伦理审查委员会

CHINESE CENTER FOR DISEASE CONTROL AND PREVENTION

INSTITUTE FOR COMMUNICABLE DISEASE CONTROL AND PREVENTION

ETHICAL REVIEW COMMITTEE

Approval Notice 0/200

PRINCIPAL INVESTIGATOR OF PROJECT:

TITLE OF PROJECT:

INSTITUTE: Chinese Center for Disease Control and Prevention,

Institute for Communicable Disease Control and Prevention

DEPARTMENT/DIVISION: National Reference Laboratory of Tuberculosis

FUNDING AGENCY :

DATE SUBMITTED:

DATE FOR WHICH REVIEWED:

DATED APPROVED:

The Ethical Review Committee of Institute for Communicable Disease Control and Prevention (ICDC), Chinese Center for Disease Control and Prevention (China CDC) has reviewed the proposed use of human subjects in the above-mentioned project. It is recognized that the right and the welfare of the subject are adequately protected; the potential risks are outweighed by potential benefits.

SIGNATURE _____

Xu Jian Guo

Chair, ICDC, China CDC Ethical Review Committee

DATE

Annex 2: Patient informed consent form

I, undersigned, (name, forename, date of birth, address) certify that I benefit from health care.

Doctor (name, forename, address, telephone), investigator physician, has offered me to participate in a research protocol organised by Name of the principal investigator Wan Kanglin on Epidemiology of Tuberculosis

He/she clearly indicated that I am free to accept or refuse to participate into this research.

I received a written information note presenting the details of implementation of this medical study and explaining the following elements

-Aim of the Research
-Methods utilised
-Duration
-Constraint
-Possible risk
-Enrolment into a computerised database
-Possible verification of the data

I had the possibility to ask all the questions that I consider useful for good comprehension of the information note, and to receive clear and precise answers.

I had sufficient time before making a decision.

I received the results of previous medical examination by Doctor (name, Forename).

I freely and willingly accept to participate into the research described at the conditions described above.

My consent do not offload the organizers of the research of their responsibilities and I keep all my rights.

I know that I can at any moment withdraw from this research without bearing any responsibility. I will in such an event inform Doctor (name of the investigator).

I was informed that, in agreement with the rules applying to the protection of individuals with regard to the processing of personal data and on the free movement of such data, the research was approved by the Ethical Committee on the (Date).

I have been informed of my right to access and modify my personal information.

All the data and information applying to me will REMAIN STRICTLY CONFIDENTIAL.

I give permission to consult this information only to people chosen by Doctor (name of the investigator) and eventually to an health official.

I will be able at any time to ask complementary information to Doctor (name and telephone of the investigator).

Date and Place

Signature of the Investigator	Signature of the patient
(or the Doctor representing him/her)	(or the legal representing person)

Annex 3: Subject information sheet

Title of the Research: Epidemiology of Tuberculosis

Name and address of the principal investigator:

Wan Kanglin

National Reference Laboratory of Tuberculosis,
National Institute for Communicable Disease Control & Prevention,
Chinese Center for Disease Control & Prevention,
P. O. Box 5, Changping, Beijing 102206
The People's Republic of China

Tel: 0086 10 61739459 Fax: 0086 10 61730233

Name, address, telephone of the investigator:

Objective of the research:

The aim of the research is to study *Mycobacterium tuberculosis*, the bacteria that infect the population and cause Tuberculosis in China. It is also intended to analyse the effect of BCG vaccination and antibiotics treatment on the spreading of particular bacterial strains. Analysis of bacteria infecting young people will tell which strains have recently caused tuberculosis whereas study of older people will tell which strains were present in the past. This research should help improving the prevention and treatment of tuberculosis.

Type of research:

The research is Without Direct Individual Benefice

Methods

From a sputum sample or any other sample collected as part of the usual practice for the diagnosis of tuberculosis, a bacterial culture will be made. The bacteria will be tested for drug resistance and its genetic material purified and analysed. The bacteria will be kept frozen to allow future research. The data on the patient health status and on the bacteria will be stored in a computerized database. The information collected from several hundred patients and the bacteria will be compared.

Length of the study

2 years

Constraint, obligations and restrictions

There will be no supplementary physical constraint, obligations and restrictions to those normally required for diagnosis and treatment of tuberculosis. Questions on the health status of the patient and history of treatment will be asked (see examination form).

Risks

None

Opinion of the Ethical Review Committee

The research has received approval on (Date)

Process of encoding or anonymisation used

The name of the patient will be known only to the investigator and will not be included in the database which will be produced in the course of this research. Each isolate will receive a coded number. The database will contain only a code and there will be no way to trace back the identity of the patients from the information contained in the database.

Scientists will never communicate about the patient information when this can be linked with the concerned patient.

Annexe 4: Epidemiological Questionnaire. Table of Examination Results of Patient with TB and Its Samples Preservation

Province_____ **No:**_____

Name		Sex		Date of Birth	
Profession		Address of Residence			
Dwelling Place			Phone		
Time of First Diagnosis		Diagnosis Place			
Diagnosis	Tuberculosis： Ⅰ; Ⅱ; Ⅲ; **Inanition**： Yes; No; or Others: _____				
Complication:	No **Yes**（Diabetes, Tumour: _____, or Other _____ ）				
Bacterium Culture	**Sputum Smear:** First time (+, -), Second time (+, -), Third time (+, -) Sputum Culture: +, - Bacterium identity:_____ **Other Sample Culture:** CSF, Thoractic Fluid, LFA +, - **Bacterium identity:**_____ Test of Drug Resistance:				

	RFP: S, R: +, ++, +++ **INH:** S, R: +, ++, +++ **EMB:** S, R: +, ++, +++ **PZA:** S, R: +, ++, +++ **SM:** S, R: +, ++, +++ **PAS:** S, R: +, ++, +++ Date of the Drug Resistance Test Results Reported:
Test of Antigen /Antibody	**Method:** **Reagents:** **Sample:** Serum, CSF, Thoractic Fluid, **Results:** Negative, Positive
DNA Tested（PCR）	**Sample:** Sputum, CSF, Thoractic Fluid, LF A, Tissue **Results:** Negative, Positive
Kind of Case	New, Relapse, Failing of Firstly Treated, Chronic Bacteria Ejector
History of Treatment	No **Yes**（Months of accumulative total: , Regulation, Irregulation）
Treated	**RFP:** ___Months **INH:** ___Months **SM:** ___Months **EMB:** ___Months **PZA:** ___Months **PAS:** ___Months
Contagium （ Contact with TB Patient）	**No**, **Yes**（Father, Mother, Brother, Sister, Husband, Wife, Colleague, Schoolfellow, Others）
General instances	Date of Collection: _____ Collector: _____ Date of Examination: _____ Laboratorian: _____

	Date of Report: _____ Result of Reporter: _____	
Sample Preserved	Strain, Serum, Sputum, CSF, Thoractic Fluid, LF A, Tissue	
Signature of the Patient's Agreement		
Remark:		

注：**INH:** Isoniazid; **RFP:** RifaRFPicin; **SM:** Streptomycin; **EMB:** Ethambutol; **PZA:** Pyrazinamide; **PAS:** p-aminosalicylic acid; **CSF:** cerebrospinal fluid; **LFA:** Lavement Fluid from Alveolus; **S:** Sensitivity, **R:** Resistance

Annexe 5: VNTR markers for MLVA15

Locus name	Minisatellite markers	Reference	TR location on H37Rv genome	Expected length in H37Rv (copy number)	Expected length in CDC1551 (copy number)	Expected length in M bovis AF2122 (copy number)
H37Rv_1955_57 bp	Mtub21	(Le Flèche et al., 2002)	1955580	206 (2)	263 (3)	263 (3)
H37Rv_2401_58 bp	Mtub30	T(Le Flèche et al., 2002)	2401815	319 (2)	435 (4)	435 (4)
H37Rv_3690_58 bp*	Mtub39	(Le Flèche et al., 2002)	3690947	341 (2.6)*	397 (3.6)	341 (2.6)
H37Rv_0580_77 bp	MIRU4 (ETR-D)	(Frothingham and Meeker-O'Connell, 1998)	580546	353 (3.3)	330 (3)	483 (5)
H37Rv_0959_53 bp	MIRU10	(Supply et al., 2000)	959868	643 (3)	750 (5)	590 (2)
H37Rv_1644_53 bp	MIRU16	(Supply et al., 2000)	1644026	671 (2)	724 (3)	671 (2)
H37Rv_2531_53 bp	MIRU23	(Supply et al., 2000)	2531560	873 (6)	820 (5)	767 (4)
H37Rv_2996_51 bp	MIRU26	(Supply et al., 2000)	2996002	614 (3)	716 (5)	716 (5)
H37Rv_3006_53 bp	MIRU27 (QUB-5)	(Supply et al., 2000)	3006875	657 (3)	657 (3)	657 (3)
H37Rv_3192_53 bp	MIRU31 (ETR-E)	(Frothingham and Meeker-O'Connell, 1998)	3192168	651 (3)	651 (3)	651 (3)
H37Rv_4348_53 bp	MIRU39	(Supply et al., 2000)	4348401	646 (2)	646 (2)	646 (2)
H37Rv_0802_54 bp	MIRU 40	(Supply et al., 2000)	802194	199 (1)	415 (5)	253 (2)
H37Rv_2165_75 pb	ETR-A	(Frothingham and Meeker-O'Connell, 1998)	2165223	397 (3)	322 (2)	847 (9)
H37Rv_2461_57 bp	ETR-B	(Frothingham and Meeker-O'Connell, 1998)	2461279	292 (3)	235 (2)	406 (5)
H37Rv_0577_58 bp	ETR-C	(Supply et al., 2000)	577172	346 (4)	288 (3)	404 (5)

Polymorphic minisatellite markers for the *M. tuberculosis* complex. The markers are listed according to their position in the H37Rv genome..

Annex 6: Spoligotyping, the 43 sequences of oligo probes

The 43 sequences of oligo probe

Serial Number [a]	Genome spacer no.	Sequences of Oligo Probe [b]
1	2	ATAGAGGGTCGCCGGTTCTGGATCA
2	3	CCTCATAATTGGGCGACAGCTTTTG
3	4	CCGTGCTTCCAGTGATCGCCTTCTA
4	12	ACGTCATACGCCGACCAATCATCAG
5	13	TTTTCTGACCACTTGTGCGGGATTA
6	14	CGTCGTCATTTCCGGCTTCAATTTC
7	15	GAGGAGAGCGAGTACTCGGGGCTGC
8	18	CGTGAAACCGCCCCCAGCCTCGCCG
9	19	ACTCGGAATCCCATGTGCTGACAGC
10	20	TCGACACCCGCTCTAGTTGACTTCC
11	21	GTGAGCAACGGCGGCGGCAACCTGG
12	22	ATATCTGCTGCCCGCCCGGGGAGAT
13	23	GACCATCATTGCCATTCCCTCTCCC
14	24	GGTGTGATGCGGATGGTCGGCTCGG
15	25	CTTGAATAACGCGCAGTGAATTTCG
16	26	CGAGTTCCCGTCAGCGTCGTAAATC
17	27	GCGCCGGCCCGCGCGGATGACTCCG
18	28	CATGGACCCGGGCGAGCTGCAGATG
19	29	TAACTGGCTTGGCGCTGATCCTGGT
20	30	TTGACCTCGCCAGGAGAGAAGATCA
21	31	TCGATGTCGATGTCCCAATCGTCGA
22	32	ACCGCAGACGGCACGATTGAGACAA
23	33	AGCATCGCTGATGCGGTCCAGCTCG
24	34	CCGCCTGCTGGGTGAGACGTGCTCG
25	35	GATCAGCGACCACCGCACCCTGTCA
26	36	CTTCAGCACCACCATCATCCGGCGC
27	37	GGATTCGTGATCTCTTCCCGCGGAT
28	38	TGCCCCGGCGTTTAGCGATCACAAC
29	39	AAATACAGGCTCCACGACACGACCA
30	40	GGTTGCCCCGCGCCCTTTTCCAGCC

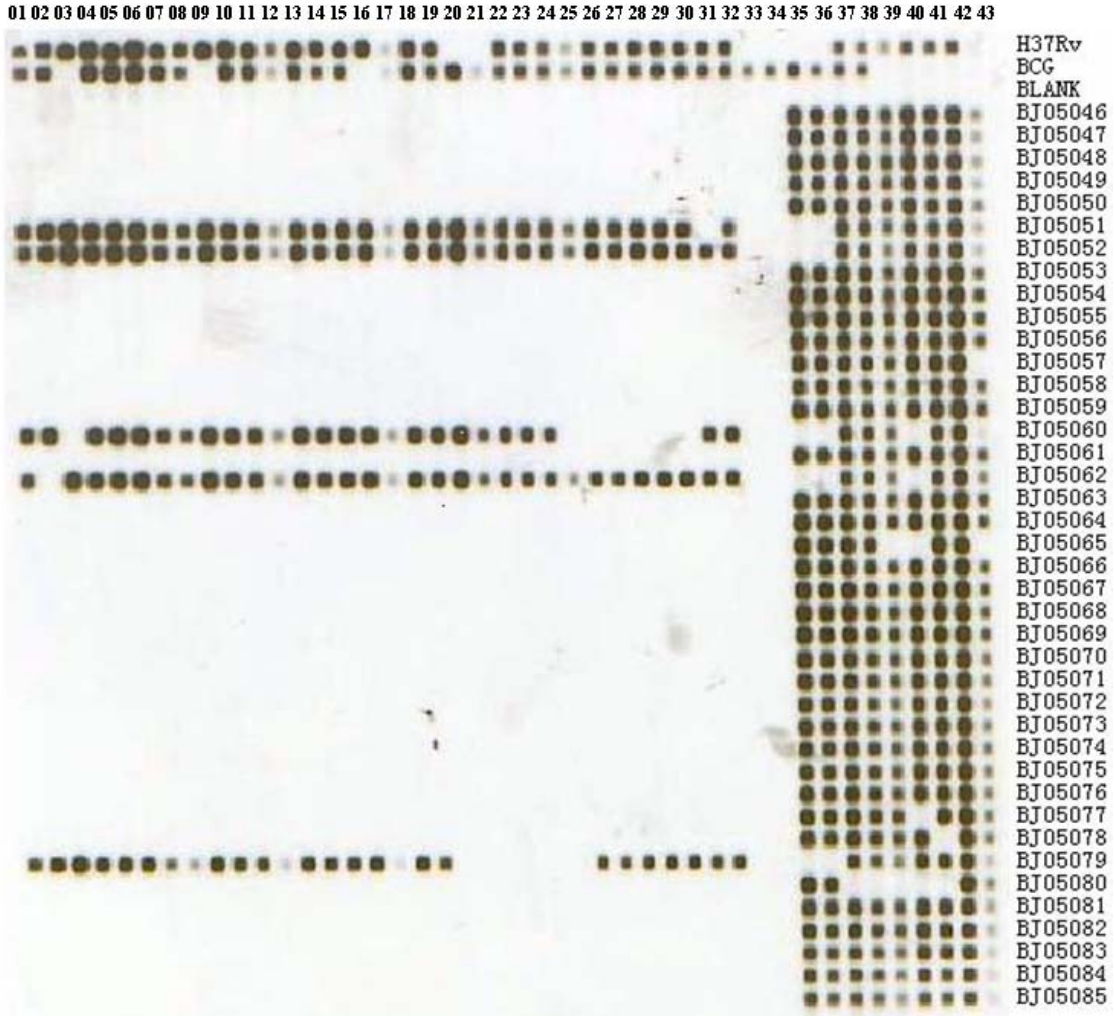
31	41	TCAGACAGGTTCGCGTCGATCAAGT
32	42	GACCAAATAGGTATCGGCGTGTTCA
33	43	GACATGACGGCGGTGCCGCACTTGA
34	44	AAGTCACCTCGCCCACACCGTCGAA
35	46	TCCGTACGCTCGAAACGCTTCCAAC
36	47	CGAAATCCAGCACCACATCCGCAGC
37	51	CGCGAAC TCGTCCACAGTCCCCCTT
38	52	CGTGGATGGCGGATGCGTTGTGCGC
39	53	GACGATGGCCAGTAAATCGGCGTGG
40	62	CGCCATCTGTGCCTCATACAGGTCC
41	63	GGAGCTTTCCGGCTTCTATCAGGTA
42	64	ATGGTGGGACATGGACGAGCGCGAC
43	65	CGCAGAATCGCACCGGGTGCGGGAG

[a] Data from reference (van der Zanden et al., 2002)

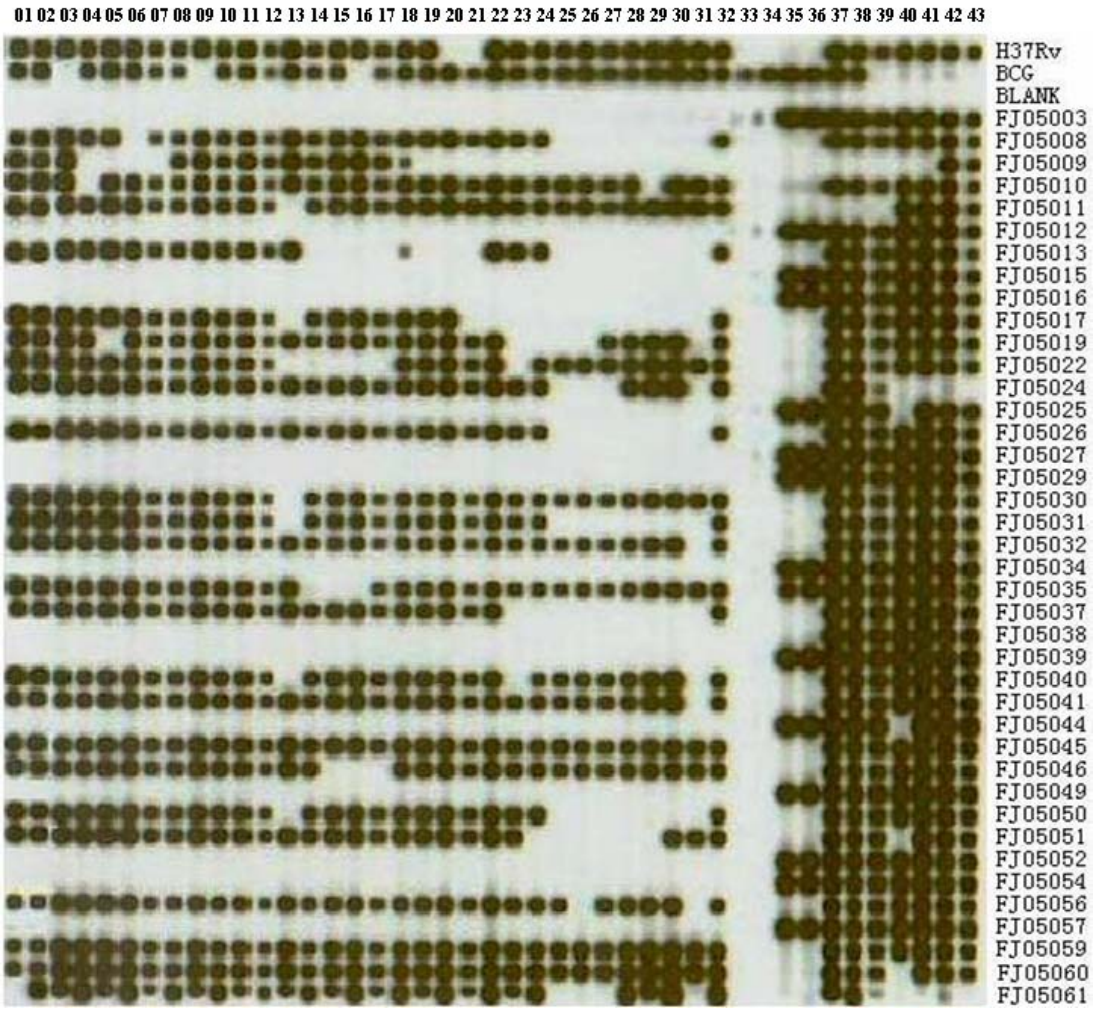
[b] Sequences used for 5' amino acid-linked oligonucleotides in spoligotyping

Examples of spoligotyping results

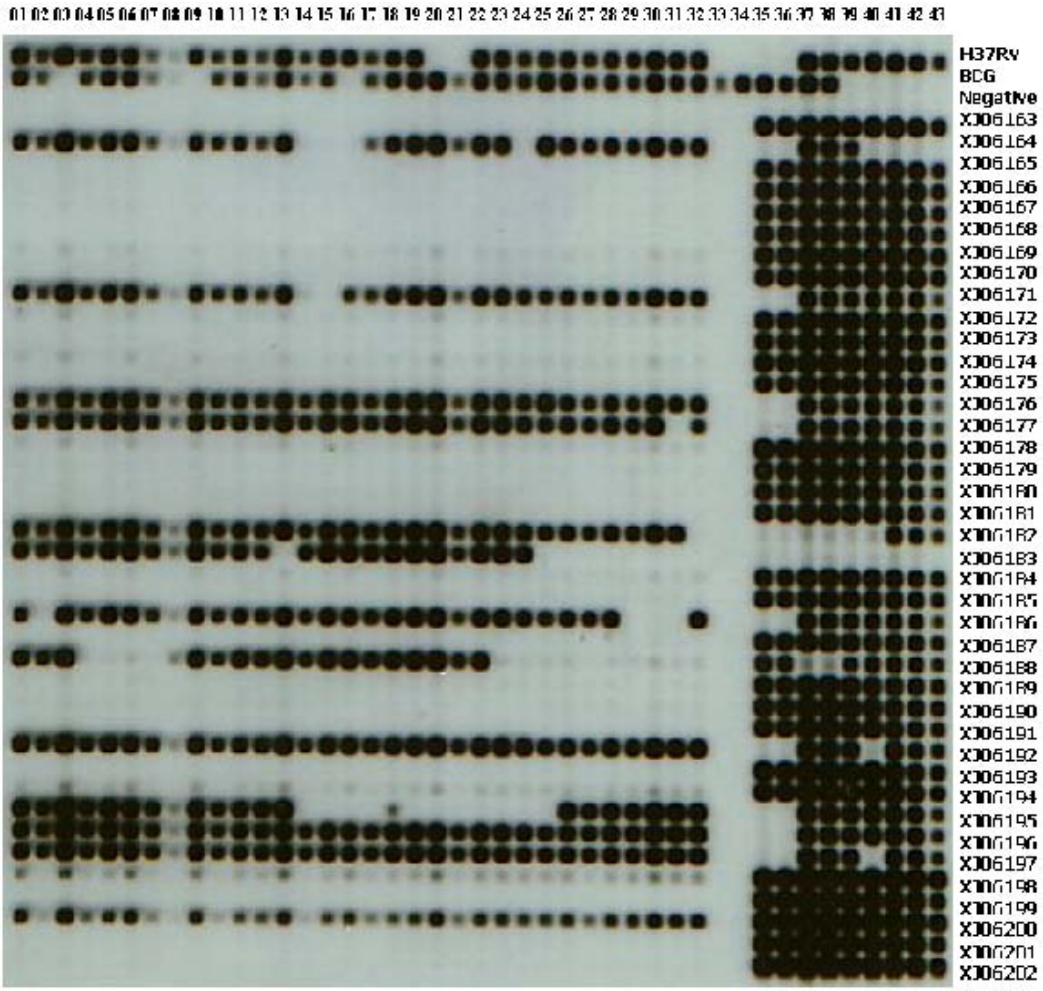
Genotype identification of Beijing province *M. tuberculosis* isolates with Spoligotyping



Genotype identification of FuJian *M. tuberculosis* isolates with Spoligotyping



Genotype identification of XinJiang *M. tuberculosis* isolates with Spoligotyping



Spoligotypes of non-Beijing strains

Strain ID	Spoligo_group	Strain ID	Spoligo_group	Strain ID	Spoligo_group
BJ05019	Modern_variant (?)	XZ06030	T1_variant (334)	SC06019	T2 (52)
BJ05026	Undesignated (?)	XZ06032	T1_variant (334)	SC06023	T1_variant (334)
BJ05043	T2 (52)	XZ06035	T1_variant (334)	SC06037	T1 (53)
BJ05051	T1_variant (50)	XZ06047	Modern_variant (?)	SC06041	T1 (53)
BJ05052	H3 (53)	XZ06055	T1 (53)	SC06043	Modern_variant (?)
BJ05062	T2_variant (1302)	XZ06063	Modern_variant (?)	SC06055	Modern_variant (?)
BJ05079	LAM_variant (1688)	XZ06074	T1 (53)	SC06057	T4 (40)
BJ05090	Modern_variant (?)	XZ06092	T1 (53)	SC06060	T1_variant (913)
FJ05008	Haarlem_variant (?)	XZ06106	T1_variant (196)	SC06061	T1_variant (913)
FJ05011	T3_variant (?)	XZ06107	Modern_variant (?)	SC06066	T2 (52)
FJ05019	H3_variant (?)	XZ06151	T1 (53)	SC06075	T1 (53)
FJ05023	T2 (52)	XZ06157	MANU2 (54)	SC06076	T2_variant (853)
FJ05024	Haarlem_variant (?)	XZ06188	LAM_variant (?)	SC06077	LAM04_variant (?)
FJ05033	H3_variant (?)	XZ06200	T1 (53)	SC06081	T2_variant (1302)
FJ05035	MANU2_variant (?)	XZ06206	T1 (53)	SC06084	T1 (53)
FJ05037	Haarlem_variant (946)	XZ06210	CAS1_Delhi (26)	SC06088	T1 (53)
FJ05040	H3_variant (?)	Shx05092	Modern_variant (?)	SC06090	LAM_variant (803)
FJ05043	Haarlem_variant (?)	Shx05117	Modern_variant (?)	SC06100	T2_variant (?)
FJ05047	T1 (53)	Shx05124	Modern_variant (172)	SC06102	T2_variant (1302)
FJ05051	Modern (?)	Shx05129	T3_variant (504)	SC06103	Modern_variant (?)
FJ05056	H3_variant (121)	Shx05133	Modern_variant (131)	SC06106	T1 (53)
FJ05059	T1 (53)	Shx05168	Modern_variant (?)	SC06112	T2_variant (1302)
FJ05060	T2 (52)	Shx05177	T2_variant (?)	SC06113	Modern_variant (?)
FJ05063	T1 (53)	Shx05178	Modern_variant (172)	SC06125	Modern_variant (?)
FJ05070	H3_variant (390)	Shx05206	Modern_variant (?)	SC06129	Modern_variant (?)
FJ05074	Haarlem_variant (?)	Shx05207	T2 (52)	SC06135	H4 (127)
FJ05082	T2 (52)	Shx05211	T1 (53)	SC06147	T2_variant (120)
FJ05083	H3_variant (293)	Shx05218	T1 (53)	SC06155	T2 (52)
FJ05098	Haarlem_variant (?)	Shx05260	Undesignated (46) ?	SC06156	Modern_variant (?)
FJ05104	T1_variant (118)	Shx05290	Modrern_variant (956) ?	SC06166	T2_variant (?)
FJ05105	Undesignate (1226)	Shx05308	Modern_variant (?)	SC06184	Modrern_variant (?)
FJ05107	H3 (53)	JL06001	T1 (53)	SC06194	T1 (53)
FJ05109	MANU2 (54)	JL06005	BOV1 (482)	HN06001	H3 (50)
FJ05120	T1_variant (118)	JL06012	T3_variant (?)	HN06004	MANU2 (54)
FJ05121	Haarlem_variant (?)	JL06014	T2_variant (1332)	HN06016	T1 (53)

Strain ID	Spoligo_group	Strain ID	Spoligo_group	Strain ID	Spoligo_group
FJ05125	Modern (?)	JL06046	T2_variant (242)	HN06018	T1 (53)
FJ05132	T1 (53)	JL06051	T2_variant (853)	HN06028	T1 (53)
FJ05135	H3 (53)	JL06052	T2 (52)	HN06033	T3-variant (?)
FJ05143	Haarlem_variant (742)	JL06053	Modern (?)	HN06034	H3 (50)
FJ05157	T2 (52)	JL06056	T2_variant (515)	HN06040	T3 (37)
FJ05158	T1 (53)	JL06109	T1 (53)	HN06042	T1 (53)
FJ05163	H3 (53)	JL06110	T1 (53)	HN06050	T5-RUS1-variant (?)
FJ05166	Modern_variant (?)	JL06131	T1 (53)	HN06051	T-variant (?)
FJ05169	T2_variant (?)	JL06139	T1-variant (462)	HN06053	H4-variant (?)
FJ05175	H3 (53)	JL06143	T2 (52)	HN06054	H4-variant (?)
FJ05181	H3 (53)	JL06150	T2 (52)	HN06060	BOV1-variant (?)
FJ05183	T3 (37)	JL06151	MANU2 (54)	HN06062	T1 (53)
FJ05185	Modern_variant (1926)	JL06170	T1 (53)	HN06064	T3 (37)
FJ05188	Undesignated (?)	JL06172	T2 (52)	HN06067	T-variant (393)
FJ05191	Modern (?)	JL06174	T3_variant (?)	HN06072	H3-variant (168)
FJ05194	MANU2_variant (?)	JL06183	Undesignated (?)	HN06073	T3-variant (?)
JL06238	T1 (53)	JL06185	T1 (53)	HN06074	T3-variant (?)
XJ06002	Modern_variant (?)	JL06218	T1_variant (?)	HN06077	H3 (50)
XJ06014	Undesignated (1314)	JL06231	T1 (53)	HN06083	T-variant (804)
XJ06034	MANU2_variant (1096)	JL06236	T1 (53)	HN06089	T2 (52)
XJ06072	MANU2_variant (1096)	JL06237	T1 (53)	HN06093	T1 (53)
XJ06086	Modern_variant (86)	JL06254	T1 (53)	HN06094	T2 (52)
XJ06090	H4 (127)	JL06281	T1 (53)	HN06110	T4-variant (?)
XJ06095	Undesignated (616)	JL06304	T3_variant (1547)	GX06112	Undesignated (623)
XJ06115	MANU2_variant (?)	JL06313	T3_variant (1547)	GX06160	Modern (?)
XJ06183	Modern_variant (56)	JL06328	T3_variant (?)	XZ06158	T1 (53)
GX06001	T1_variant (1166)	XJ06001	Modern_variant (?)	Shx05125	Modern_variant (?)
GX06002	Undesignated (246)	XJ06013	Modern_variant (?)	Shx05130	T1 (53)
GX06005	T1 (53)	XJ06017	T2_variant (?)	Shx05239	T1 (53)
GX06007	T1 (53)	XJ06018	Modern_variant (357)	Shx05296	MANU2 (54) ?
GX06008	T1 (53)	XJ06021	T1 (53)	Shx05306	T1 (53) ?
GX06013	T3 (37)	XJ06025	T1_variant (240)	Shx05314	MANU2 (54) ?
GX06014	Modern (?)	XJ06033	Undesingated (?)	Shx05324	MANU2 (54) ?
GX06017	T3_variant (?)	XJ06036	Modern_variant (1570)	Shx05372	MANU2 (54)
GX06018	Modern (?)	XJ06038	LAM_variant (803)	JL06340	T3_variant (?)
GX06019	T1_variant (50)	XJ06039	H4 (127)	JL06342	T1 (53)
GX06020	T5_variant (535)	XJ06040	Undesingated (27)	JL06359	LAM09 (49)

Strain ID	Spoligo_group	Strain ID	Spoligo_group	Strain ID	Spoligo_group
GX06021	T3_variant (?)	XJ06042	H_variant (35)	FJ05009	Modern_variant (1777)
GX06027	T1_variant (50)	XJ06047	T2 (52)	FJ05010	Modern_variant (?)
GX06030	Undesignated (246)	XJ06051	T1_variant (?)	FJ05013	Haarlem_variant (?)
GX06035	Modern (358)	XJ06052	T2 (52)	FJ05014	Haarlem_variant (?)
GX06040	Undesignated (246)	XJ06056	H4_variant (?)	FJ05017	Haarlem_variant (?)
GX06041	Modern (?)	XJ06059	Modern_variant (?)	FJ05018	MANU2_variant (1634)
GX06042	T1 (53)	XJ06060	Undesignated (?)	FJ05022	Modern (?)
GX06043	Modern (956)	XJ06063	H4 (127)	FJ05026	Haarlem_variant (742)
GX06049	T1 (53)	XJ06071	Haarlem_variant (?)	FJ05030	T3 (37)
GX06050	T2_variant (515)	XJ06075	Haarlem_variant (?)	FJ05031	Haarlem_variant (?)
GX06051	T1_variant (462)	XJ06081	Modern_variant (?)	FJ05032	H3 (53)
GX06053	MANU2 (54)	XJ06083	LAM_variant (?)	FJ05041	H3 (53)
GX06056	T2_variant (?)	XJ06089	H4 (127)	FJ05045	T1 (53)
GX06057	Modern (?)	XJ06108	H4_variant (?)	FJ05046	Modern (?)
GX06060	T1 (53)	XJ06109	Modern_variant (?)	FJ05050	Haarlem_variant (?)
GX06061	Modern (?)	XJ06116	Modern_variant (56)	FJ05053	LAM_variant (1107)
GX06063	Modern (358)	XJ06129	CAS1_variant (?)	FJ05055	H3 (53)
GX06066	T1_variant (50)	XJ06132	Modern_variant (?)	FJ05061	Haarlem_variant (?)
GX06067	Modern_undesignated (51)	XJ06142	T2_variant (?)	FJ05062	Haarlem_variant (?)
GX06069	T1 (53)	XJ06149	Modern_variant (?)	FJ05073	Modern (?)
GX06072	T1 (53)	XJ06150	Modern_variant (?)	FJ05075	Modern (?)
GX06074	Modern (358)	XJ06153	Undesingated (?)	FJ05078	H3 (53)
GX06075	T1 (53)	XJ06155	H4_variant (?)	FJ05080	Haarlem_variant (?)
GX06077	T1_variant (393)	XJ06156	T1 (53)	FJ05088	Haarlem_variant (742)
GX06081	T1_variant (86)	XJ06158	Undesingated (?)	GS05023	T2_variant (?)
GX06086	MANU2 (54)	XJ06159	Modern_variant (?)	FJ05091	Haarlem_variant (?)
GX06087	Modern (?)	XJ06160	Modern_variant (?)	FJ05108	H3 (53)
GX06088	Modern (61)	XJ06161	T2_variant (?)	FJ05110	MANU2_variant (?)
GX06089	Modern (?)	XJ06164	Modern_variant (?)	FJ05117	H3_variant (?)
GX06090	MANU2 (54)	XJ06171	Modern_variant (?)	FJ05123	T1_variant (118)
GX06092	Modern (?)	XJ06176	T1 (53)	FJ05127	T1_variant (462)
GX06094	T1 (53)	XJ06177	H3 (57)	FJ05133	Haarlem_variant (?)
GX06095	T1 (53)	XJ06182	Modern_variant (?)	FJ05142	T2_variant (?)
GX06097	Modern (?)	XJ06186	H4 (127)	FJ05150	Undesignated (523)
GX06098	T2 (52)	XJ06192	T2 (52)	FJ05167	MANU2 (54)
GX06101	T1 (53)	XJ06195	LAM_variant (264)	FJ05172	MANU2 (54)
GX06102	T1_variant (196)	XJ06196	T1 (53)	FJ05173	H3 (53)

Strain ID	Spoligo_group	Strain ID	Spoligo_group	Strain ID	Spoligo_group
GX06105	Modern (?)	XJ06197	T2 (52)	FJ05178	H3 (53)
GX06107	Modern_undesignated (51)	GS06006	MANU2 (54)	FJ05189	H3 (53)
GX06109	T1 (53)	GS06007	T1 (53)	FJ05198	T2 (52)
GX06113	Modern_undesignated (180)	GS06087	MANU2_variant (1096)	FJ05199	Haarlem_variant (275)
GX06114	T1 (53)	GS06088	T1_variant (334)	GS05041	MANU2 (54)
GX06116	T3_variant (?)	GS05040	T1 (53)	GS05043	MANU2 (54)
GX06117	T1_variant (50)	GS05056	T2_variant (1023)	GS05050	BOV1_variant (?)
GX06118	T3_variant (504)	GS05071	H37Rv (451)	GS05070	MANU2 (54)
GX06120	T1 (53)	GS05132	H4 (127)	GS05115	MANU2 (54)
GX06121	Undesignated (?)	GS05160	T1_variant (462)	GS05155	H4 (127)
GX06124	T1 (53)	GS05163	H4 (127)	GS05180	MANU2_variant (?)
GX06126	T3_variant (?)	GS05178	H37Rv (451)	GS05187	H37Rv (451)
GX06129	MANU2 (54)	ZJ06001	H3_variant (293)	GS05192	Undesignated (?)
GX06130	Modern_undesignated (956)	ZJ06005	T2_variant (?)	GS06036	Haarlem_variant (?)
GX06131	T1_variant (1129)	ZJ06008	Haarlem_variant (?)	GS06049	T3 (37)
GX06132	T1 (53)	ZJ06011	Modern (?)	GS06051	T3 (37)
GX06134	Modern (?)	ZJ06012	T2 (52)	GS06053	T3 (37)
GX06135	T2_variant (1332)	ZJ06014	T1 (53)	GS06096	Modern_variant (?)
GX06137	T1_variant (44)	ZJ06024	Modern (?)	HN06032	T1 (53)
GX06140	T3_variant (?)	ZJ06028	T1 (53)	HN06071	H3-variant (168)
GX06145	Undesignated (?)	ZJ06036	T2_variant (153)	HN06101	T-variant (?)
GX06151	T1 (53)	ZJ06037	T2_variant (?)	SC06110	T1 (53)
GX06158	T1 (53)	ZJ06054	T2 (52)	SC06215	MANU2 (54)
GX06161	T1 (53)	ZJ06057	H3 (57)	SC06217	MANU2 (54)
GX06162	Undesignated (523)	ZJ06062	T1 (53)	SC06227	MANU2 (54)
GX06165	T1 (53)	ZJ06063	T1 (53)	SC06230	MANU2 (54)
GX06168	T1_variant (50)	ZJ06064	T1 (53)	XJ06019	MANU2_variant (1096)
GX06173	Modern (?)	ZJ06065	T1 (53)	XJ06026	MANU2_variant (?)
GX06175	T3_variant (73)	ZJ06066	T1 (53)	XJ06055	H4_variant (?)
GX06176	Modern (?)	ZJ06067	T2_variant (73)	XJ06080	Modern_variant (?)
GX06187	Modern (?)	ZJ06068	T2_variant (73)	XJ06106	MANU2_variant (?)
GX06189	Modern (358)	ZJ06069	T2_variant (73)	XJ06110	MANU2_variant (?)
GX06190	T1 (53)	ZJ06070	T2_variant (73)	XJ06114	Modern_variant (?)
GX06191	T1_variant (713)	ZJ06074	LAM_variant (?)	XJ06130	Modern_variant (?)
GX06192	T1 (53)	ZJ06082	T1 (53)	XJ06145	Undesingated (?)
GX06200	Modern (742)	ZJ06084	T1 (53)	XJ06188	Undesignated (1198)
GX06203	Undesignated (?)	ZJ06085	T1_variant (7)	XJ06200	MANU2_variant (?)

Strain ID	Spoligo_group	Strain ID	Spoligo_group	Strain ID	Spoligo_group
GX06204	Modern (?)	ZJ06086	T1 (53)	ZJ06019	MANU2 (54)
GX06205	T2_variant (853)	ZJ06094	MANU2 (54)	ZJ06025	MANU2_variant (?)
GX06207	T1_variant (7)	ZJ06098	Undesignated (623)	ZJ06026	MANU2 (54)
GX06208	T1 (53)	ZJ06101	T1_variant (7)	ZJ06029	MANU2_variant (?)
XZ06003	CAS1_Delhi (26)	SC06003	T1 (53)	ZJ06030	MANU2 (54)
XZ06008	T1 (53)	SC06005	Undesignated (532)	ZJ06058	MANU2 (54)
XZ06009	T1 (53)	SC06006	MANU2_variant (?)	BJ05123	T1_variant (7)
XZ06023	CAS1_Delhi (26)	SC06016	T1 (53)		

5. PUBLICATIONS

5.1. Publication N°1 (Translation from Chinese) Study on the mutation characteristics of *M. tuberculosis* Rifampicin-resistance gene (*rpoB)*

LIU Jinghua, LIU Zhiguang, ZHAO Xiuqin et al. State Key Laboratory for Infectious Diseases Prevention and Control, National Institute for Communicable Disease Control & Prevention, Chinese Center for Disease Control & Prevention, P. O. Box 5, Changping, Beijing 102206 P. R. China. ZHANG Lishui, HUANG Mingxiang, WANG Lin. *Fuzhou Pulmonary Hospital of Fujian Province.

#**Correspongding author: WAN Kanglin,** *wankanglin@icdc.cn*

Objective: To elucidate the characters of *ropB* gene mutation in rifampicin-resistant clinical isolates of *Mycobacterium tuberculosis*.

Methods: A 286bp DNA fragment of *rpoB* gene including the 81bp hot-spot region (rifampicin resistance determination region, RRDR) was analyzed by PCR-SSCP. Then the 286bp DNA fragment of each strain proved to have mutation by PCR-SSCP was sequenced.

***M. tuberculosis* isolates.** The 110 clinical *M. tuberculosis* isolates obtained from Fuzhou Pulmonary hospital of Fujian province used in this study were isolated from Jan. to Dec. 2004.

Results

26 strains were susceptible to all tested drugs, 11 strains were susceptible to rifampicin but resistant to other drugs and 73 strains were resistant to rifampicin. BACTECTM MGITTM 960 system has been used to test the resistance to the primary anti-tuberculosis drugs. The drug susceptibility patterns of the isolates are shown in Table1.

Table1 Drug susceptibility patterns of *M. tuberculosis* strains

M. tuberculosis isolates		No. of strains	RIF	INH	EMB	SM
Drug-Resistant (84)	Rifampicin-resistant (73)	38	R	R	R	R
		5	R	R	R	S
		18	R	R	S	R
		9	R	R	S	S
		1	R	S	R	R
		1	R	S	S	S
	Rifampicin-susceptible and other drugs-resistant (11)	9	S	R	S	R
		1	S	R	R	R
		1	S	R	R	S
Drug-susceptible(26)		26	S	S	S	S

R:drug-resistant；S:drug-susceptible

PCR-SSCP analysis

286bp *rpoB* gene fragments of 110 clinical isolates were amplified successfully (Fig1). Mutations were detected rapidly by SSCP method (Fig2). Of all the studied strains, 47 (64.38%) Rifampicin-resistant isolates, 2 (18.18%) Rifampicin-susceptible and other drugs-resistant isolates, 1 (3.85%) drug-susceptible isolate showed PCR-SSCP pattern different from that of H37Rv (Table2).

Table2 The PCR-SSCP results of the *rpoB* gene

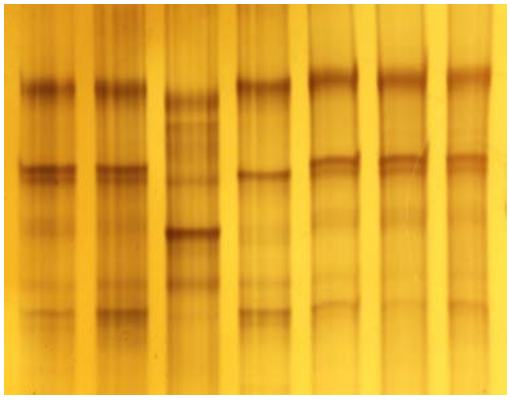
M. tuberculosis isolates	No. of strains	No. of strains with mutation (PCR-SSCP)	Mutation rate
Rifampicin-resistant	73	47	64.38%
Rifampicin-susceptible and other drugs-resistant	11	2	18.18%
Drug-susceptible	26	1	3.85%
Total	110	50	45.45%

M 1 2 3 4 5

0

Fig1. Electrophoresis-photogram of PCR product of *rpoB* gene

M. PCR Marker; 1. H37Rv; 2-4. Clinical isolates, 5.Negative Control



M 1 2 3 M 4 5

Fig 2. results of *rpoB* PCR-SSCP

M. H37Rv; 1-5. clinical isolates; isolate 2 and 3 are different from the H37Rv.

2. DNA sequence analysis

DNA sequence analysis were carried out on the strains showing PCR-SSCP patterns different from that of H37Rv. All the strains showed missense mutations.

Rifampicin-susceptible isolates

In the drug-susceptible strains, 1 had a single-point mutation of His526→Asn (CAC→AAC). Among the Rifampicin-susceptible and other drugs-resistant strains, one showed a single-point mutation of Leu511→Pro (CTG→CCG), and another showed double mutations, mutation of His526→Asn (CAC→AAC) combined with mutation of Pro_{535}→Leu (CCC→CTC).

Rifampicin-resistant isolates

DNA sequence analysis of 47 Rifampicin-resistant strains revealed 20 different kinds of missense mutations, affecting 10 amino acids within a 286-bp region of the *rpoB* gene (Table3). 36 (76.6%) strains had a single-point mutation, and the highest frequency of mutation was observed in the codon of Ser531 (61.1%). Another codon, His526, also showed higher mutation frequency (25.0%) than the others. Among the resistant strains, 11 (23.4%) showed double mutations, 11 different combined mutation types. Most of them are point mutations except in 2 strains; one was an insertion and another was a deletion. In this study, no silent mutation was found.

Table 3 Mutations in *rpoB* gene of 47 refampin-resistant strains detected by DR method

Serial number	Codon(s) (E. coli) [a]	ORF codon(s) (M.tuberculosis) [b]	Base change(s)	Amino acid change	No. of isolates	Companent Ratio(%)	Mutation Rate(%)
1	513	438	CAA→CCA	Gln→Pro	1	2.1	1.4
2	516	441	GAC→TAC	Asp→Tyr	3	6.5	4.1
3	522	447	TCG→TTG	Ser→Leu	1	2.1	1.4
4	526	451	CAC→GAC	His→Asp	3	6.5	4.1
5	526	451	CAC→AAC	His→Asn	1	2.1	1.4
6	526	451	CAC→CGC	His→Arg	4	8.6	5.5
7	526	451	CAC→CTC	His→Leu	1	2.1	1.4
8	531	456	TCG→TTG	Ser→Leu	22	46.9	30.1
9	511 516	436 441	CTG→CCG GAC→GGC	Leu→Pro Asp→Gly	1	2.1	1.4
10	511 516	436 441	CTG→CGG GAC→GGC	Leu→Arg Asp→Gly	1	2.1	1.4
11	491 526	416 451	ACG→A_G CAC→CGC	 His→Arg	1	2.1	1.4
12	491 526	416 451	ACG→CAG CAC→GAC	Thr→Gln His→Asp	1	2.1	1.4
13	530	455	CTG→CGG	Leu→Arg	1	2.1	1.4

	531	456	TCG→TTG	Ser→Leu			
	508	433	ACC→CCC	Thr→Pro			
14					1	2.1	1.4
	526	451	CAC→TAC	His→Tyr			
	516	441	GAC→GAA	Asp→Glu			
15					1	2.1	1.4
	526	451	CAC→AAC	His→Asn			
	490	415	CAG→CAC	Gln→His			
16					1	2.1	1.4
	531	456	TCG→TTG	Ser→Leu			
	511	436	CTG→CCG	Leu→Pro			
17					1	2.1	1.4
	516	441	GAC→TAC	Asp→Tyr			
	490-491	415－416					
	insert C	insert C					
18					1	2.1	1.4
	526	451	CAC→AAC	His→Asn			
	535	460	CCC→CTC	Pro→Leu			
	490	415	CAG→CAC	Gln→His			
19	516	441	GAC→TAC	Asp→Tyr	1	2.1	1.4
	526	451	CAC→GAC	His→Asp			

Note: a Codons are given relative to *E. coli rpoB.*

b *M. tuberculosis rpoB* codon numbering

Conclusions: Mutations among the 81bp RRDR of *rpoB* gene are the main reason of *Mycobacterium tuberculosis* resistance to rifampicin. The Ser_{531} and His_{526} codons are the most common positions of mutations.

5.2. Publication N°2 (Translation from Chinese) Study on genetic diversity of 65 *Mycobacterium tuberculosis* strains from three provinces of China based on 7 variable number of tandem repeats

LIU Jinghua, WAN Kanglin [#], LIU Zhiguang, et al. National Institute of Commmunicable Disease Control & Prevention, Chinese Center of Disease Control & Prevention, P. O. Box 5, Changping, Beijing 102206 P. R. China

[#]**Corresponsding author: WAN Kanglin, E-mail:** Wan Kanglin，National Institute of Commmunicable Disease Control & Prevention, Chinese Center of Disease Control & Prevention, P. O. Box 5, Changping, Beijing 102206, P. R. China Tel:010-61739459

E-mail: klwan1217@163.com

Objective: To study the characters of genetic diversity of *Mycobacterium tuberculosis* strains in China by VNTR method.

Methods: 11 tandem repeat loci in the total genome were analyzed by PCR and agarose gel electrophoresis method. The polymorphism of DNA fingerprinting was analyzed using Gel-Pro analyzer 3.1 software and BioNumerics 3.0 software.

M. tuberculosis strains

Six-five *M. tuberculosis* strains were isolated from Anhui TB institute (29), Hunan TB institute (17), Jiangsu TB institute (7) and ICDC, China CDC (12) respectively. The detail information was shown in Table1.

Table 1 Strain and origin

Strain number	Original sample	Place of sampling	Time of isolation (year)	Place of isolation
AH113	Sputum	Anhui province	2003	Anhui TB Institute
AH312	Sputum	Anhui province	2003	Anhui TB Institute
AH327	Sputum	Anhui province	2003	Anhui TB Institute
AH348	Sputum	Anhui province	2003	Anhui TB Institute
AH361	Sputum	Anhui province	2003	Anhui TB Institute
AH372	Sputum	Anhui province	2003	Anhui TB Institute
AH379	Sputum	Anhui province	2003	Anhui TB Institute
AH416	Sputum	Anhui province	2003	Anhui TB Institute
AH423	Sputum	Anhui province	2003	Anhui TB Institute
AH444	Sputum	Anhui province	2003	Anhui TB Institute
AH455	Sputum	Anhui province	2003	Anhui TB Institute
AH456	Sputum	Anhui province	2003	Anhui TB Institute
AH463	Sputum	Anhui province	2003	Anhui TB Institute
AH469	Sputum	Anhui province	2003	Anhui TB Institute
AH495	Sputum	Anhui province	2003	Anhui TB Institute
AH500	Sputum	Anhui province	2003	Anhui TB Institute
AH501	Sputum	Anhui province	2003	Anhui TB Institute
AH505	Sputum	Anhui province	2003	Anhui TB Institute
AH506	Sputum	Anhui province	2003	Anhui TB Institute
AH507	Sputum	Anhui province	2003	Anhui TB Institute
AH515	Sputum	Anhui province	2003	Anhui TB Institute
AH575	Sputum	Anhui province	2003	Anhui TB Institute
AH607	Sputum	Anhui province	2003	Anhui TB Institute
AH608	Sputum	Anhui province	2003	Anhui TB Institute
AH617	Sputum	Anhui province	2003	Anhui TB Institute
AH635	Sputum	Anhui province	2003	Anhui TB Institute
AH643	Sputum	Anhui province	2003	Anhui TB Institute
AH651	Sputum	Anhui province	2003	Anhui TB Institute
AH663	Sputum	Anhui province	2003	Anhui TB Institute
HN2050	Sputum	Hunan province	2002	Hunan TB Institute
HN2073	Sputum	Hunan province	2002	Hunan TB Institute
HN2074	Sputum	Hunan province	2002	Hunan TB Institute
HN2075	Sputum	Hunan province	2002	Hunan TB Institute
HN2076	Sputum	Hunan province	2002	Hunan TB Institute

Strain number	Original sample	Place of sampling	Time of isolation (year)	Place of isolation
HN2079	Sputum	Hunan province	2002	Hunan TB Institute
HN2085	Sputum	Hunan province	2002	Hunan TB Institute
HN2119	Sputum	Hunan province	2002	Hunan TB Institute
HN2147	Sputum	Hunan province	2002	Hunan TB Institute
HN2157	Sputum	Hunan province	2002	Hunan TB Institute
HN2176	Sputum	Hunan province	2002	Hunan TB Institute
HN2221	Sputum	Hunan province	2002	Hunan TB Institute
HN2247	Sputum	Hunan province	2002	Hunan TB Institute
HN2264	Sputum	Hunan province	2002	Hunan TB Institute
HN2287	Sputum	Hunan province	2002	Hunan TB Institute
HN2327	Sputum	Hunan province	2002	Hunan TB Institute
HN2328	Sputum	Hunan province	2002	Hunan TB Institute
JS01-077	Sputum	Jiangsu province	2002	Jiangsu TB Institute
JS01-167	Sputum	Jiangsu province	2002	Jiangsu TB Institute
JS03-216	Sputum	Jiangsu province	2002	Jiangsu TB Institute
JS08-065	Sputum	Jiangsu province	2002	Jiangsu TB Institute
JS08-374	Sputum	Jiangsu province	2002	Jiangsu TB Institute
JS09-130	Sputum	Jiangsu province	2002	Jiangsu TB Institute
JS15-355	Sputum	Jiangsu province	2002	Jiangsu TB Institute
NRL-JS03010	Sputum	Xuzhou, Jiangsu	2003	ICDC, China CDC
NRL-JS03012	Sputum	Xuzhou, Jiangsu	2003	ICDC, China CDC
NRL-JS03013	Sputum	Xuzhou, Jiangsu	2003	ICDC, China CDC
NRL-JS03014	Sputum	Xuzhou, Jiangsu	2003	ICDC, China CDC
NRL-JS03015	Sputum	Xuzhou, Jiangsu	2003	ICDC, China CDC
NRL-JS03016	Sputum	Xuzhou, Jiangsu	2003	ICDC, China CDC
NRL-JS03025	Sputum	Xuzhou, Jiangsu	2003	ICDC, China CDC
NRL-JS03034	Sputum	Xuzhou, Jiangsu	2003	ICDC, China CDC
NRL-JS03043	Sputum	Xuzhou, Jiangsu	2003	ICDC, China CDC
NRL-JS03045	Sputum	Xuzhou, Jiangsu	2003	ICDC, China CDC
NRL-JS03050	Sputum	Xuzhou, Jiangsu	2003	ICDC, China CDC
NRL-JS03053	Sputum	Xuzhou, Jiangsu	2003	ICDC, China CDC

Results:

7 VNTR loci incorporating 6 ETR loci (ETR-A, ETR-B, ETR-C, ETR-D, ETR-E, ETR-F) and 1 MPTR locus MPTR-A) were selected to genotype the 65 *M. tuberculosis* clinical strains (Table2). These loci showed obvious length polymorphism among the studied strains (Fig1, Fig2). Finally, twenty-six different genotypes were defined and the clustering analysis results showed that most of the clinical strains (69.2%) were divided into two main groups (Fig3).

Conclusions:

In China, perhaps there exist some main epidemic groups among *M. tuberculosis* strains. VNTR method is a powerful, quick and low cost assay. It can be widely used in epidemiological study.

Table 2 Primer sequence and fragment size of each VNTR unit

位点名称	引 物 序 列 (5'-3')	片段大小(bp)
ETR-A	AAATCGGTCCCATCACCTTCTTAT	75
	CGAAGCCTGGGGTGCCCGCGATTT	
ETR-B	GCGAACACCAGGACAGCATCATG	57
	GGCATGCCGGTGATCGAGTGG	
ETR-C	GTGAGTCGCTGCAGAACCTGCAG	58
	GGCGTCTTGACCTCCACGAGTG	
ETR-D	CAGGTCACAACGAGAGGAAGAGC	77
	GCGGATCGGCCAGCGACTCCTC	
ETR-E	CTTCGGCGTCGAAGAGAGCCTC	53
	CGGAACGCTGGTCACCACCTAAG	
ETR-F	CTCGGTGATGGTCCGGCCGGTCAC	79
	GGAAGTGCTCGACAACGCCATGCC	
MPTR-A	GGTTACCACTTCGATGCGTCTGCG	15
	AGCCGCCGAAACCCATC	

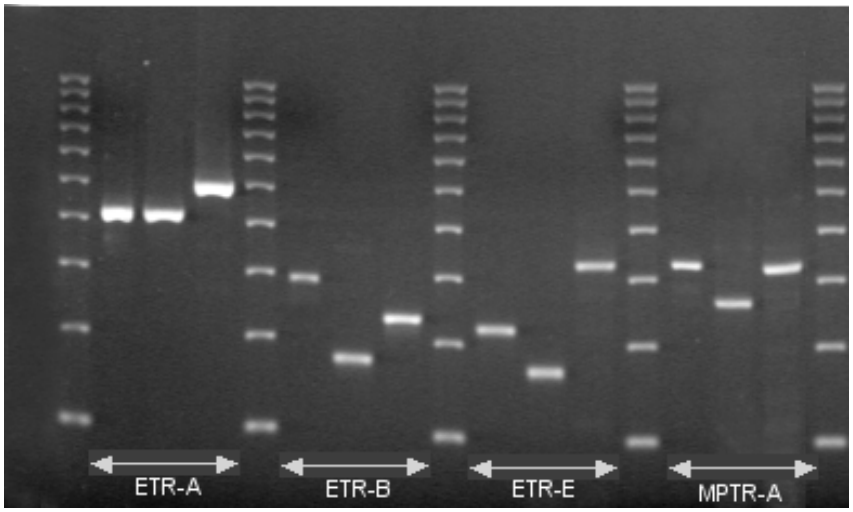


Fig 1. Length polymorphisms at four VNTR loci in the studied strains

From left to right: ETR-A ETR-B ETR-E MPTR-A

M：100bp DNA ladder；1：H37Rv；2：NRL-JS03013；3：NRL-JS01012

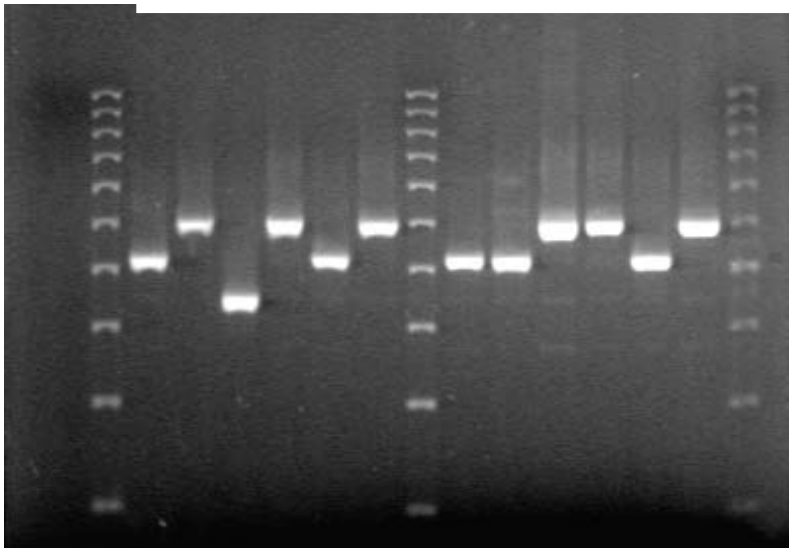


Fig 2. Length polymorphisms at the ETR-A locus in the studied isolates.

M：100bp DNA ladder (Bio-rad)；1、7：H37Rv；2：HN2075；3：HN2119；4：HN2327；5：HN2157；6：HN2074； 8：HN2079；9：HN2147；10：HN2076；11： HN2221；12：HN2085

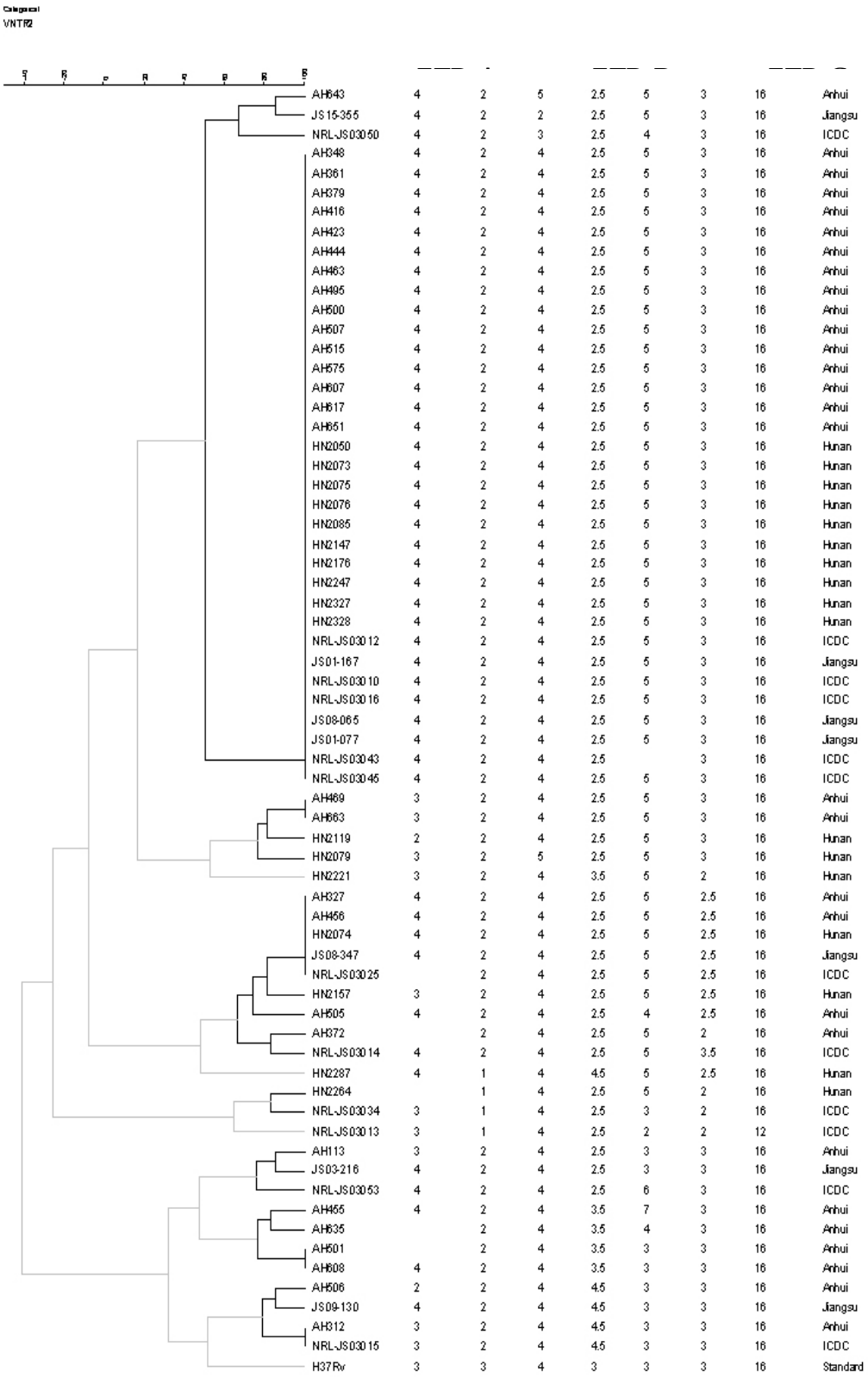


Fig 3. Dendrogram deduced from the clustering analysis of the 65 isolates

Publication N°3 *Mycobacterium tuberculosis* diversity in 5 Chinese provinces as assessed by MLVA and spoligotyping.

Kanglin WAN [1,2],Yolande HAUCK[1], Sabrina IVOL[1], Philippe LE FLECHE[1,3], Yuan-Yuan ZHAN [2], Zhi-Guang LIU [2], Xiu-Qin ZHAO [2], Qin WANG [5], Yun-Hong TAN [6], Wei-Guo XU [7], Gilles VERGNAUD[1,3], Christine POURCEL[1*]

1 Génome, Polymorphisme et Minisatellites (GPMS), Institut de Génétique et Microbiologie, Bat. 400, Université Paris XI, 91405 Orsay cedex, France

2 Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention , Beijing, P. R. China.

3 Centre d'Etude du Bouchet, BP3, 91710 Vert le Petit, France

4 Anhui Provincial Institute of Tuberculosis Prevention and Cure, P. R. China

5 Hunan Anti-TB Institute, P. R. China

6 Jiangsu Provincial Center for Diseases Prevention and Control, P. R. China

* Corresponding author. Mailing address : GPMS, Institut de Génétique et Microbiologie, Bat. 400, Université Paris XI, 91405 Orsay cedex, France.

Phone : (33) 1 69 15 30 01.

Fax: (33) 1 69 15 66 78

E-mail: Christine.Pourcel@igmors.u-psud.fr

Running title: Diversity of Chinese *Mycobacterium tuberculosis* strains

Legends

TABLE 1. Strains origin and characteristics.

TABLE 2. Allelic numbers and diversity of each VNTR locus in the 259 strains and in the two major Chinese groups.

Figure 1: Minimum spanning tree showing the clustering of 259 M. tuberculosis isolates by MLVA.

A minimum spanning tree was constructed using the genotyping data obtained by MLVA21 (the data will be made accessible from a Web page: http://bacterial-genotyping.igmors.u-psud.fr). The two new Chinese groups are clustered.

Figure 2: Dendrogram showing the clustering of 60 strains by MLVA.

Clustering was performed using the MLVA21 genotyping data from 36 non-Beijing Chinese isolates, 8 Beijing isolates and a selection of isolates belonging to the main families. The spoligotype is shown next to each isolate but is not used in the clustering.

Figure 3: Dendrogram showing the clustering of 78 Beijing strains by MLVA.

The following markers were typed ETRA, ETRB, ETRC, ETRD, ETRE, MIRU02, MIRU10, MIRU16, MIRU23, MIRU26, MIRU27, MIRU39, MIRU40, Mtub01, Mtub02, Mtub21, Mtub29, Mtub30, Mtub39, Qub11a. Missing data are due to the presence of several PCR products.

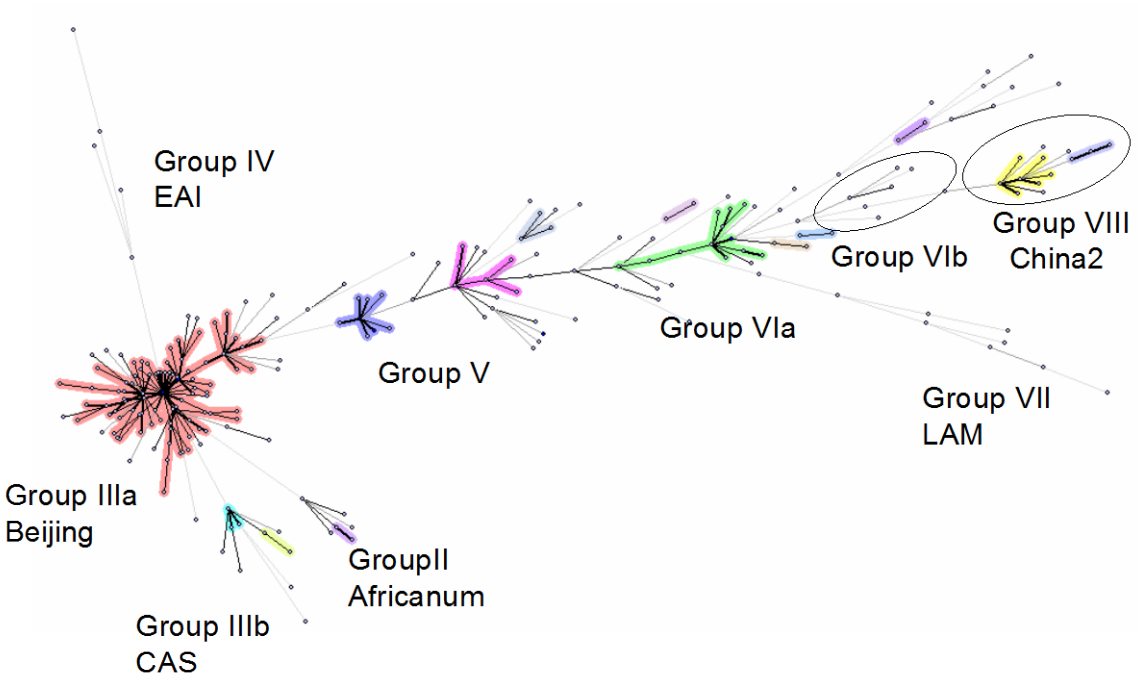
TABLE 1 104 Chinese isolates

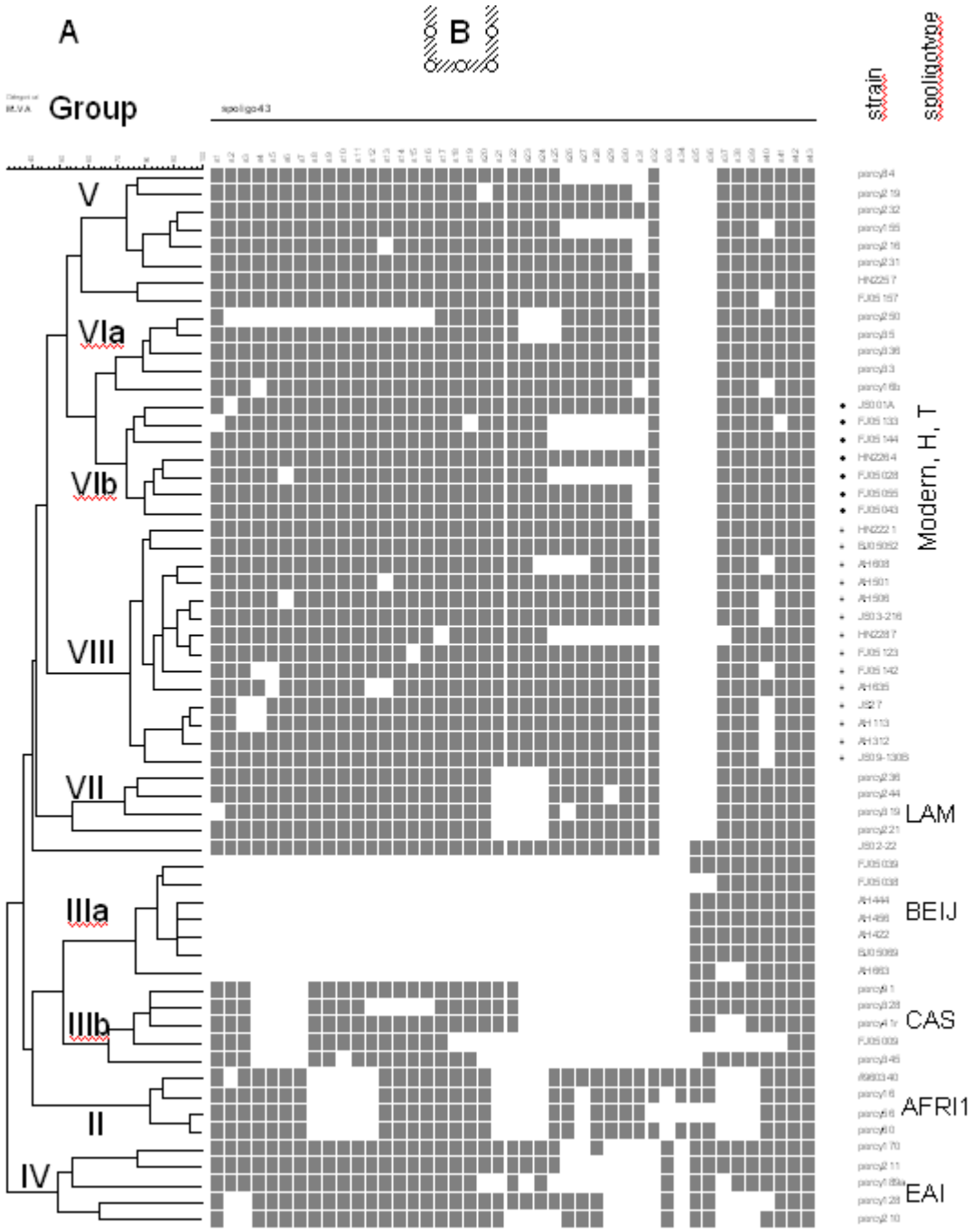
Strain	Province	Year	Group	Strain	Province	Year	Group
JS01-077	JiangSu	2000	IIIa/BEIJ	**AH608**	AnHui	2003	VIII/China2
JS08-347	JiangSu	2002	IIIa/BEIJ	**AH651**	AnHui	2003	IIIa/BEIJ
JS27	JiangSu	2002	VIII/China2	**AH423**	AnHui	2003	IIIa/BEIJ
JS02-022	JiangSu	2002	DM	**AH444**	AnHui	2003	IIIa/BEIJ
JS08-065	JiangSu	2002	IIIa/BEIJ	**AH455**	AnHui	2003	IIIa/BEIJ
JS03-216	JiangSu	2002	VIII/China2	**AH456**	AnHui	2003	IIIa/BEIJ
JS01-167	JiangSu	2002	IIIa/BEIJ	**AH463**	AnHui	2003	IIIa/BEIJ
JS09-130A	JiangSu	2002	IIIa/BEIJ	**AH469**	AnHui	2003	IIIa/BEIJ
JS09-130B	JiangSu	2002	VIII/China2	**AH495**	AnHui	2003	IIIa/BEIJ
JS15-355	JiangSu	2002	IIIa/BEIJ	**AH500**	AnHui	2003	IIIa/BEIJ
JS054B	JiangSu	2003	IIIa/BEIJ	**AH501**	AnHui	2003	VIII/China2
JS024B	JiangSu	2003	IIIa/BEIJ	**AH505**	AnHui	2003	IIIa/BEIJ
JS001A	JiangSu	2003	VIb	**BJ05008**	Beijing		IIIa/BEIJ
JS005B	JiangSu	2003	IIIa/BEIJ	**BJ05015**	Beijing		IIIa/BEIJ
JS012C	JiangSu	2003	IIIa/BEIJ	**BJ05023**	Beijing		IIIa/BEIJ
JS012A	JiangSu	2003	IIIa/BEIJ	**BJ05030**	Beijing		IIIa/BEIJ
HN2058	Hunan	2001	IIIa/BEIJ	**BJ05036**	Beijing		IIIa/BEIJ
HN2119	Hunan	2002	IIIa/BEIJ	**BJ05039**	Beijing		IIIa/BEIJ
HN2247	Hunan	2002	IIIa/BEIJ	**BJ05042**	Beijing		IIIa/BEIJ
HN2327	Hunan	2002	IIIa/BEIJ	**BJ05045**	Beijing		IIIa/BEIJ
HN2328	Hunan	2002	IIIa/BEIJ	**BJ05049**	Beijing		IIIa/BEIJ
HN2147	Hunan	2002	IIIa/BEIJ	**BJ05052**	Beijing		VIII/China2
HN2074	Hunan	2002	IIIa/BEIJ	**BJ05056**	Beijing		IIIa/BEIJ
HN2264	Hunan	2002	VIb	**BJ05057**	Beijing		IIIa/BEIJ
HN2221	Hunan	2002	VIII/China2	**BJ05069**	Beijing		IIIa/BEIJ
HN2075	Hunan	2002	IIIa/BEIJ	**BJ05071**	Beijing		IIIa/BEIJ
HN2257	Hunan	2002	V	**BJ05083**	Beijing		IIIa/BEIJ
HN2250	Hunan	2002	IIIa/BEIJ	**BJ05094**	Beijing		IIIa/BEIJ
HN2176	Hunan	2002	IIIa/BEIJ	**BJ05104**	Beijing		IIIa/BEIJ
HN2073	Hunan	2002	IIIa/BEIJ	**BJ05105**	Beijing		IIIa/BEIJ
HN2085	Hunan	2002	IIIa/BEIJ	**BJ05107**	Beijing		IIIa/BEIJ
HN2076	Hunan	2002	IIIa/BEIJ	**BJ05113**	Beijing		IIIa/BEIJ
HN2079	Hunan	2002	DM	**FJ05006**	Fujian		IIIa/BEIJ
HN2287	Hunan	2003	VIII/China2	**FJ05009**	Fujian		IIIb
AH113	AnHui	2003	VIII/China2	**FJ05028**	Fujian		VIb
AH312	AnHui	2003	VIII/China2	**FJ05036**	Fujian		IIIa/BEIJ
AH327	AnHui	2003	IIIa/BEIJ	**FJ05038**	Fujian		IIIa/BEIJ
AH348	AnHui	2003	IIIa/BEIJ	**FJ05039**	Fujian		IIIa/BEIJ
AH361	AnHui	2003	IIIa/BEIJ	**FJ05043**	Fujian		VIb
AH372	AnHui	2003	IIIa/BEIJ	**FJ05055**	Fujian		VIb
AH379	AnHui	2003	IIIa/BEIJ	**FJ05115**	Fujian		IIIa/BEIJ
AH416	AnHui	2003	IIIa/BEIJ	**FJ05123**	Fujian		VIII
AH422	AnHui	2003	IIIa/BEIJ	**FJ05133**	Fujian		VIb
AH663	AnHui	2003	IIIa/BEIJ	**FJ05142**	Fujian		VIII
AH506	AnHui	2003	VIII/China2	**FJ05144**	Fujian		VIb
AH575	AnHui	2003	IIIa/BEIJ	**FJ05146**	Fujian		IIIa/BEIJ
AH643	AnHui	2003	IIIa/BEIJ	**FJ05151**	Fujian		IIIa/BEIJ
AH507	AnHui	2003	IIIa/BEIJ	**FJ05157**	Fujian		V
AH515	AnHui	2003	IIIa/BEIJ	**FJ05162**	Fujian		IIIa/BEIJ
AH617	AnHui	2003	IIIa/BEIJ	**FJ05164**	Fujian		IIIa/BEIJ
AH635	AnHui	2003	VIII/China2	**FJ05190**	Fujian		IIIa/BEIJ
AH607	AnHui	2003	IIIa/BEIJ	**FJ05197**	Fujian		IIIa/BEIJ

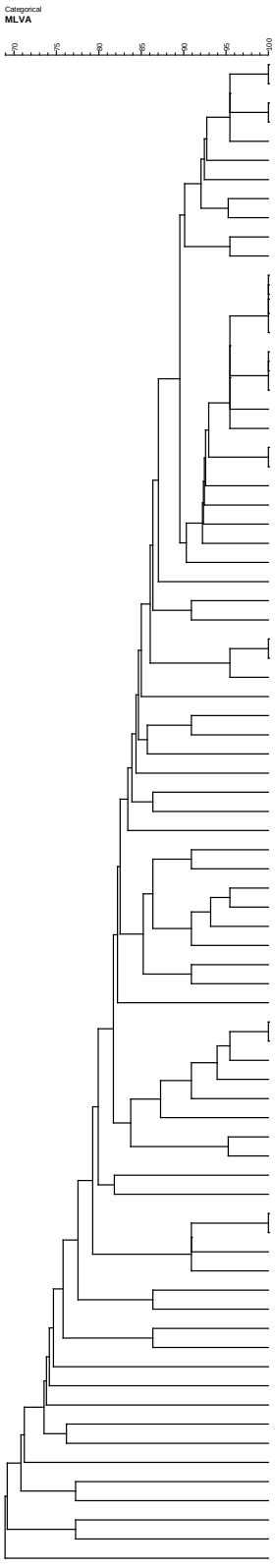
TABLE 2

VNTR locus[a]	Total strains (256)		BEIJ (119)		China2 (13)	
	No. of alleles	HGI	No. of alleles	HGI	No. of alleles	HGI
ETRA-2165	7	0.551	3	0.127	3	0.603
ETRB-2461	5	0.188	2	0.017	2	0.154
ETRC-0577	4	0.392	3	0.082	1	0.000
ETRD-0580	8	0.209	4	0.098	2	0.538
ETRE-3192	7	0.658	6	0.246	3	0.295
MIRU02-0154	3	0.068	2	0.017	1	0.000
MIRU10-0959	8	0.554	4	0.187	1	0.000
MIRU16-1644	5	0.415	4	0.159	3	0.564
MIRU23-2531	7	0.321	4	0.186	2	0.282
MIRU26-2996	9	0.758	7	0.509	4	0.718
MIRU27-3006	4	0.181	3	0.033	1	0.000
MIRU39-4348	5	0.524	4	0.269	2	0.154
MIRU40-0802	7	0.576	4	0.260	2	0.154
Mtub01-0024	4	0.463	3	0.066	1	0.000
Mtub12-1121	4	0.192	3	0.098	1	0.000
Mtub21-1955	8	0.779	7	0.587	1	0.000
Mtub29-2347	3	0.225	3	0.066	1	0.000
Mtub30-2401	5	0.516	3	0.066	1	0.000
Mtub39-3690	18	0.482	7	0.220	1	0.000
Mtub02-0079	13	0.794	6	0.340	2	0.154
Qub11a-2163	17	0.845	15	0.702	3	0.295

[a] name and position on the genome (kb)




A
B
Group
spoligo43
strain
spoligotype
V
VIa
VIb
VIII
VII
IIIa
IIIb
II
IV
Modern, H, T
LAM
BEIJ
CAS
AFRI1
EAI



AH607	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	4.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
AH444	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	4.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
AH361	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
AH372	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
AH379	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	2.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
AH507	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	7.0
AH422	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	4.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
AH617	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
BJ05008	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	4.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
AH456	4.0	2.0	4.0	3.0	5.0	2.0	2.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
JS024B	4.0	2.0	4.0	3.0	5.0	2.0	2.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
HN2073	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
BJ05039	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
BJ05069	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
FJ05036	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
JS01-0.	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	6.0
JS01-1.	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	6.0
BJ05030	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	6.0
AH463	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	7.0
JS005B	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	26.0
HN2147	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	29.0	4.0	5.0	4.0	4.0	4.0	9.0
BJ05107	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	29.0	4.0	5.0	4.0	4.0	4.0	9.0
HN2075	4.0	2.0	4.0	3.0	4.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
BJ05094	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	2.0	4.0	4.0	4.0	9.0
BJ05104	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	1.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
FJ05115	4.0	2.0	4.0	3.0	5.0	2.0	3.0	2.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
BJ05042	4.0	2.0	4.0	3.0	5.0	2.0	3.0	1.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	6.0
FJ05146	4.0	2.0	5.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	10.0
AH575	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	5.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	6.0
AH651	4.0	2.0	4.0	3.0	5.0	2.0	3.0	2.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	6.0
AH469	3.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	6.0
BJ05057	3.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	6.0
FJ05151	3.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	2.0	4.0	6.0
AH495	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	3.0	5.0	4.0	4.0	4.0	28.0
BJ05113	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	5.0	3.0	2.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
FJ05197	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	2.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	2.0	9.0
HN2076	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0		4.0	5.0	4.0	2.0	2.0	9.0
JS15-3.	4.0	2.0	2.0	3.0	5.0	1.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
BJ05036	4.0	2.0	4.0	3.0	6.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	2.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	6.0
BJ05049	4.0	2.0	4.0	3.0	6.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	26.5	4.0	5.0	4.0	4.0	4.0	9.0
AH423	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	2.0	9.0	28.0	4.0	5.0	4.0	2.0	4.0	9.0
BJ05045	4.0	2.0	4.0	3.0	5.0	2.0	4.0	3.0	5.0	5.0	3.0	3.0	3.0	9.0	28.0	4.0	4.0	4.0	4.0	3.0	9.0
FJ05162	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	5.0	3.0	3.0	3.0	9.0	28.0	4.0	4.0	4.0	4.0	5.0	9.0
HN2250	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	4.0	4.0	4.0	4.0	9.0
FJ05190	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	4.0	3.0	3.0	3.0	9.0	28.0	4.0	4.0	4.0	4.0	4.0	9.0
FJ05039	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	4.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	4.0	4.0	4.0	4.0	9.0
BJ05083	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	6.0	6.0	3.0	3.0	3.0	9.0	28.0	4.0	4.0	4.0	4.0	4.0	9.0
HN2058	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	3.0	4.0	4.0	3.0	9.0
FJ05038	4.0	2.0	4.0	3.0	5.0	2.0	3.0	4.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	3.0	4.0	4.0	4.0	9.0
BJ05015	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	6.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	5.0	6.0
HN2327	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	4.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
HN2328	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	4.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
HN2247	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	4.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	4.0	4.0	4.0	4.0	9.0
AH348	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	4.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	6.0
HN2119	2.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	4.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	4.0	4.0	4.0	4.0	9.0
AH327	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	4.0	7.0	3.0	3.0	3.0	9.0	29.0	4.0	4.0	4.0	4.0	4.0	27.0
AH500	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	9.0	3.0	3.0	3.0	9.0	28.0	4.0	4.0	4.0	4.0	4.0	7.0
BJ05071	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	9.0	3.0	3.0	3.0	9.0	28.0	4.0	4.0	4.0		4.0	9.0
AH455	4.0	2.0	4.0	4.0	7.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
BJ05056	4.0	2.0	4.0	2.0	4.0	2.0	3.0	3.0	5.0	4.0	3.0	3.0	3.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	9.0
JS012A	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	4.0	7.0	3.0	3.0	3.0	9.0	26.0	4.0	4.0	4.0	4.0	4.0	8.0
JS012C	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	4.0	7.0	3.0	3.0	3.0	9.0	26.0	4.0	4.0	4.0	4.0	4.0	8.0
JS054B	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	26.0	4.0	4.0	4.0	4.0	4.0	10.0
HN2085	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	26.0	4.0	3.0	4.0	4.0	4.0	8.0
AH416	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	3.0	3.0	3.0	3.0	9.0	29.0	4.0	4.0	4.0	4.0	4.0	28.0
AH515	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	2.0	3.0	9.0	29.0	4.0	5.0	4.0	4.0	4.0	28.0
JS09-1.	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	4.0	9.0	29.0	4.0	4.0	4.0	4.0	8.0	9.0
FJ05006	4.0	3.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	4.0	9.0	29.0	4.0	6.0	4.0	4.0	4.0	9.0
JS08-0.	4.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	4.0	3.0	3.0	7.0	28.0	4.0	4.0	4.0	4.0	4.0	25.0
JS08-3.	4.0	2.0	4.0	2.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	3.0	9.0	27.0	3.0	4.0	4.0	4.0	4.0	9.0
AH663	3.0	2.0	4.0	3.0	5.0	2.0	3.0	3.0	4.0	7.0	3.0	3.0	2.0	9.0	28.0	4.0	4.0	4.0	4.0	4.0	9.0
AH505	4.0	2.0	4.0	3.0	4.0	2.0	3.0	3.0	5.0	6.0	3.0	3.0	1.0	9.0	28.0	4.0	4.0	4.0	4.0	4.0	8.0
BJ05105	4.0	2.0	4.0	3.0	4.0	2.0	3.0	3.0	5.0	5.0	3.0	3.0	3.0	9.0	28.0	4.0	3.0	2.0	4.0	4.0	
HN2176	4.0	2.0	4.0	3.0	7.0	2.0	3.0	2.0	5.0	6.0	3.0	3.0	3.0	10.0	28.0	4.0	4.0	4.0	4.0	4.0	9.0
HN2079	3.0	2.0	5.0	3.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	2.0	2.0	9.0	28.0	4.0	5.0	4.0	4.0	4.0	6.0
BJ05023	3.0	2.0	4.0	1.0	5.0	2.0	3.0	3.0	5.0	7.0	3.0	3.0	4.0	9.0	28.0	4.0	5.0	4.0	4.0	3.0	6.0
AH643	4.0	2.0	5.0	3.0	5.0	2.0	4.0	4.0	5.0	5.0	3.0	3.0	2.0	9.0	28.0	4.0	4.0	4.0	4.0	4.0	8.0
FJ05164	4.0	2.0	5.0	3.0	5.0	2.0	3.0	4.0	5.0	3.0	3.0	3.0	3.0	9.0	28.0	4.0	4.0	4.0	4.0	4.0	9.0
HN2074	4.0	2.0	4.0	3.0	5.0	2.0	1.0	3.0	5.0	9.0	3.0	3.0	3.0	9.0	29.0	4.0	4.0	4.0	4.0	5.0	3.0

6. REFERENCES

Afanas'ev, M.V., Ikryannikova, L.N., Il'ina, E.N., Sidorenko, S.V., Kuz'min, A.V., Larionova, E.E., Smirnova, T.G., Chernousova, L.N., Kamaev, E.Y., Skorniakov, S.N., Kinsht, V.N., Cherednichenko, A.G. Govorun, V.M., (2007). Molecular characteristics of rifampicin- and isoniazid-resistant *Mycobacterium tuberculosis* isolates from the Russian Federation. J Antimicrob Chemother 59, 1057-64.

Alland, D., Kalkut, G.E., Moss, A.R., McAdam, R.A., Hahn, J.A., Bosworth, W., Drucker, E. Bloom, B.R., (1994). Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. N Engl J Med 330, 1710-6.

Anh, D.D., Borgdorff, M.W., Van, L.N., Lan, N.T., van Gorkom, T., Kremer, K. van Soolingen, D., (2000). *Mycobacterium tuberculosis* Beijing genotype emerging in Vietnam. Emerg Infect Dis 6, 302-5.

Arnold, C., Thorne, N., Underwood, A., Baster, K. Gharbia, S., (2006). Evolution of short sequence repeats in *Mycobacterium tuberculosis*. FEMS Microbiol Lett 256, 340-6.

Aziz, M.A., Wright, A., Laszlo, A., De Muynck, A., F., P., Van Deun, A., Wells, C., Nunn, P., Blanc, L. Raviglione, M., Epidemiology of antituberculosis drug resistance (the Global Project on Anti-tuberculosis Drug Resistance Surveillance): an updated analysis. WHO/International Union Against Tuberculosis And Lung Disease Global Project on Anti-tuberculosis Drug Resistance Surveillance. Stop TB Department, WHO, Geneva, Switzerland.

Baker, L., Brown, T., Maiden, M.C. Drobniewski, F., (2004). Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. Emerg Infect Dis 10, 1568-77.

Banerjee, A. Sugantino, M., (1998). The mabA gene from the inhA opeeron of *Mycobacterium tuberculosis* encodes a 3-katoacyl reductase that fails to confer isoniazid resistance. Microbiology 144, 2697-2707.

Bauer, J., Andersen, A.B., Kremer, K. Miorner, H., (1999). Usefulness of spoligotyping To discriminate IS6110 low-copy-number *Mycobacterium tuberculosis* complex strains cultured in Denmark. J Clin Microbiol 37, 2602-6.

Bhanu, N.V., van Soolingen, D., van Embden, J.D., Dar, L., Pandey, R.M. Seth, P., (2002). Predominace of a novel *Mycobacterium tuberculosis* genotype in the Delhi region of India. Tuberculosis (Edinb) 82, 105-12.

Bifani, P., Mathema, B., Campo, M., Moghazeh, S., Nivin, B., Shashkina, E., Driscoll, J., Munsiff, S.S., Frothingham, R. Kreiswirth, B.N., (2001). Molecular identification of streptomycin monoresistant *Mycobacterium tuberculosis* related to multidrug-resistant W strain. Emerg Infect Dis 7, 842-8.

Bifani, P.J., Mathema, B., Kurepina, N.E. Kreiswirth, B.N., (2002). Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. Trends Microbiol 10, 45-52.

Bifani, P.J., Mathema, B., Liu, Z., Moghazeh, S.L., Shopsin, B., Tempalski, B., Driscol, J., Frothingham, R., Musser, J.M., Alcabes, P. Kreiswirth, B.N., (1999). Identification of a W variant outbreak of *Mycobacterium tuberculosis* via population-based molecular epidemiology. Jama 282, 2321-7.

Brosch, R., Gordon, S., Garnier, T., Eiglmeier, K., Frigui, W., Valenti, P., Dos Santos, S., Duthoy, S., Lacroix, C., Garcia-Pelayo, C. al, e., (2007). Genome plasticity of BCG and impact on vaccine efficacy. Proc Natl Acad Sci U S A 104, 5596-5601.

Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L.M., Pym, A.S., Samper, S., van Soolingen, D. Cole, S.T., (2002). A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc Natl Acad Sci U S A 99, 3684-9.

Brosch, R., Pym, A.S., Gordon, S.V. Cole, S.T., (2001). The evolution of mycobacterial pathogenicity: clues from comparative genomics. Trends Microbiol 9, 452-8.

Brudey, K., Driscoll, J.R., Rigouts, L., Prodinger, W.M., Gori, A., Al-Hajoj, S.A., Allix, C., Aristimuno, L., Arora, J., Baumanis, V., Binder, L., Cafrune, P., Cataldi, A., Cheong, S., Diel, R., Ellermeier, C., Evans, J.T., Fauville-Dufaux, M., Ferdinand, S., Garcia de Viedma, D., Garzelli, C., Gazzola, L., Gomes, H.M., Guttierez, M.C., Hawkey, P.M., van Helden, P.D., Kadival, G.V., Kreiswirth, B.N., Kremer, K., Kubin, M., Kulkarni, S.P., Liens, B., Lillebaek, T., Ho, M.L., Martin, C., Martin, C., Mokrousov, I., Narvskaia, O.,

Ngeow, Y.F., Naumann, L., Niemann, S., Parwati, I., Rahim, Z., Rasolofo-Razanamparany, V., Rasolonavalona, T., Rossetti, M.L., Rusch-Gerdes, S., Sajduda, A., Samper, S., Shemyakin, I.G., Singh, U.B., Somoskovi, A., Skuce, R.A., van Soolingen, D., Streicher, E.M., Suffys, P.N., Tortoli, E., Tracevska, T., Vincent, V., Victor, T.C., Warren, R.M., Yap, S.F., Zaman, K., Portaels, F., Rastogi, N. Sola, C., (2006). *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol 6, 23.

Canetti, G., Fox, W., Khomenko, A. al., e., (1969). Advances in techniques of testing mycobacterial drug sensitivity, and the use of sensitivity tests in tuberculosis control programmes. Bull World Health Organ 41, 21-43.

Canetti, G., Froman, S., Grosset, J. al., e., (1963). Mycobacteria: laboratory methods for testing drug sensitivity and resistance. Bull World Health Organ 29, 565-578.

Cave, M.D., Eisenach, K.D., Templeton, G., Salfinger, M., Mazurek, G., Bates, J.H. Crawford, J.T., (1994). Stability of DNA fingerprint pattern produced with IS6110 in strains of *Mycobacterium tuberculosis*. J Clin Microbiol 32, 262-6.

Cavusoglu, C., Hilmioglu, S., Guneri, S. Bilgic, A., (2002). Characterization of rpoB mutations in rifampin-resistant clinical isolates of *Mycobacterium tuberculosis* from Turkey by DNA sequencing and line probe assay. J Clin Microbiol 40, 4435-4438.

Chan, M.Y., Borgdorff, M., Yip, C.W. al, e., (2001). Seventy percent of the *Mycobacterium tubercuosis* isolates in Hong kong represent the Beijing genotype. Epidemiol Infect 127, 169-171.

Chandler, M., and J. Mahillon, (2002). Insertion sequences revisited. American Society for Microbiology, Washington, D.C.

Chin, P.J., Chiu, C.C. Jou, R., (2007). Identification of Beijing lineage *Mycobacterium tuberculosis* with combined mycobacterial interspersed repetitive unit loci 26, 31, and ETR-A. J Clin Microbiol. 45, 1022-3.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S. Barrell, B.G., (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393, 537-44.

Cowan, L.S., Mosher, L., Diem, L., Massey, J.P. Crawford, J.T., (2002). Variable-number tandem repeat typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS*6110* by using mycobacterial interspersed repetitive units. J Clin Microbiol 40, 1592-602.

Cruciani, F., Santolamazza, P., Shen, P., Macaulay, V., Moral, P., Olckers, A., Modiano, D., Holmes, S., Destro-Bisol, G., Coia, V. al., e., (2002). A back migration
from Asia to sub-Saharan Africa is supported by high resolution analysis of human Y-chromosome haplotypes. Am J Hum Genet 70, 1197-1214.

Dale, J.W., Brittain, D., Cataldi, A.A., Cousins, D., Crawford, J.T., Driscoll, J., Heersma, H., Lillebaek, T., Quitugua, T., Rastogi, N., Skuce, R.A., Sola, C., Van Soolingen, D. Vincent, V., (2001). Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: recommendations for standardised nomenclature. Int J Tuberc Lung Dis 5, 216-9.

De Beenhouwer, H., Lhiang, Z., Jannes, G. al., e., (1995). Rapid detection of rifampicin resistance insputum and biopsy specimens from tuberculosis patients by PCR and line probe assay. Tuber Lung Dis 76, 425-30.

Driscoll, J.R., McGarry, M.A. Taber, H.W., (1999). DNA typing of a nonviable culture of *Mycobacterium tuberculosis* in a homeless shelter outbreak. J Clin Microbiol 37, 274-5.

Espasa, M., Gonzalez-Martin, J., Alcaide, F. al, e., (2005). Direct detection in clinical samples ofmultiple gene mutations causing resistance of *Mycobacterium tuberculosis* to isoniazid and rifampicin using fluorogenic probes. J Antimicrob Chemother 55, 860-5.

Fabre, M., Koeck, J.-L., Le Fleche, P., Simon, F., Herve, V., Vergnaud, G. Pourcel, C., (2004). High genetic diversity revealed by variable-number tandem repeat genotyping and analysis of hsp65 gene polymorphism in a large collection of "*Mycobacterium canettii*"

strains indicates that the *M. tuberculosis* complex is a recently emerged clone of "*M. canettii*". J Clin Microbiol 42, 3248-55.
Feil, E.J., Li, B.C., Aanensen, D.M., Hanage, W.P. Spratt, B.G., (2004). eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol 186, 1518-1530.
Filliol, I., Driscoll, J.R., van Soolingen, D., Kreiswirth, B.N., Kremer, K., Valetudie, G., Dang, D.A., Barlow, R., Banerjee, D., Bifani, P.J., Brudey, K., Cataldi, A., Cooksey, R.C., Cousins, D.V., Dale, J.W., Dellagostin, O.A., Drobniewski, F., Engelmann, G., Ferdinand, S., Gascoyne-Binzi, D., Gordon, M., Gutierrez, M.C., Haas, W.H., Heersma, H., Kassa-Kelembho, E., Ho, M.L., Makristathis, A., Mammina, C., Martin, G., Mostrom, P., Mokrousov, I., Narbonne, V., Narvskaya, O., Nastasi, A., Niobe-Eyangoh, S.N., Pape, J.W., Rasolofo-Razanamparany, V., Ridell, M., Rossetti, M.L., Stauffer, F., Suffys, P.N., Takiff, H., Texier-Maugein, J., Vincent, V., de Waard, J.H., Sola, C. Rastogi, N., (2003). Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. J Clin Microbiol 41, 1963-70.
Filliol, I., Motiwala, A.S., Cavatore, M., Qi, W., Hernando Hazbon, M., Bobadilla Del Valle, M., Fyfe, J., Garcia-Garcia, L., Rastogi, N., Sola, C., Zozio, T., Guerrero, M.I., Leon, C.I., Crabtree, J., Angiuoli, S., Eisenach, K.D., Durmaz, R., Joloba, M.L., Rendon, A., Sifuentes-Osornio, J., Ponce de Leon, A., Cave, M.D., Fleischmann, R., Whittam, T.S. Alland, D., (2006). Global Phylogeny of *Mycobacterium tuberculosis* Based on Single Nucleotide Polymorphism (SNP) Analysis: Insights into Tuberculosis Evolution, Phylogenetic Accuracy of Other DNA Fingerprinting Systems, and Recommendations for a Minimal Standard SNP Set. J Bacteriol 188, 759-772.
Filliol, I., Sola, C. Rastogi, N., (2000). Detection of a previously unamplified spacer within the DR locus of *Mycobacterium tuberculosis*: epidemiological implications. J Clin Microbiol 38, 1231-4.
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A., Bishai, W., Jacobs Jr, W.R., Jr., Venter, J.C. Fraser, C.M., (2002). Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. J Bacteriol 184, 5479-90.
Flores, L., Van, T., Narayanan, S., Deriemer, K., Kato-Maeda, M. Gagneux, S., (2007). Large Sequence Polymorphisms Classify Mycobacterium tuberculosis Strains with Ancestral Spoligotyping Patterns. J Clin Microbiol 45, 3393-5.
Fomukong, N., Beggs, M., el Hajj, H., Templeton, G., Eisenach, K. Cave, M.D., (1997). Differences in the prevalence of IS6110 insertion sites in *Mycobacterium tuberculosis* strains: low and high copy number of IS6110. Tuber Lung Dis 78, 109-16.
Fressatti-Cardoso, R., Cooksey, R.C., Morlock, G.P. al, e., (2004). Screening and Characterization of Mutations in Isoniazid-Resistant *Mycobacterium tuberculosis* Isolates Obtained in Brazil. Antimicrobial Agents And Chemotherapy 3373-3387.
Frothingham, R. Meeker-O'Connell, W.A., (1998). Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology 144, 1189-1196.
Garcia de Viedma, D., (2003). Rapid detection of resistance in *Mycobacterium tuberculosis*: a review discussing molecular approaches. Clin Microbiol Infect 9, 349-59.
Garnier, T., Eiglmeier, K., Camus, J., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsempe, C. al, e., (2003). The complete genome sequence of *Mycobacterium bovis*. Proc Natl Acad Sci U S A 3, 3.
Glynn, J.R., Whiteley, J., Bifani, P.J., Kremer, K. van Soolingen, D., (2002). Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. Emerg Infect Dis 8, 843-9.
Goguet de la Salmoniere, Y.O., Kim, C.C., Tsolaki, A.G., Pym, A.S., Siegrist, M.S. Small, P.M., (2004). High-throughput method for detecting genomic-deletion polymorphisms. J Clin Microbiol 42, 2913-8.
Goloubeva, V., Lecocq, M., Lassowsky, P., Matthys, F., Portaels, F. Bastian, I., (2001). Evaluation of mycobacteria growth indicator tube for direct and indirect drug

susceptibility testing of *Mycobacterium tuberculosis* from respiratory specimens in a Siberian prison hospital. J Clin Microbiol 39, 1501-5.
Gordon, S.V., Brosch, R., Billault, A., Garnier, T., Eiglmeier, K. Cole, S.T., (1999). Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. Mol Microbiol 32, 643-655.
Goyal, M., Young, D., Zhang, Y., Jenkins, P.A. Shaw, R.J., (1994). PCR amplification of variable sequence upstream of katG gene to subdivide strains of *Mycobacterium tuberculosis* complex. J Clin Microbiol 32, 3070-1.
Groenen, P.M., Bunschoten, A.E., van Soolingen, D. van Embden, J.D., (1993). Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis;* application for strain differentiation by a novel typing method. Mol Microbiol 10, 1057-65.
Gryadunov, D., Mikhailovich, V., Lapa, S., Roudinskii, N., Donnikov, M., Pan'kov, S., Markova, O., Kuz'min, A., Chernousova, L., Skotnikova, O., Moroz, A., Zasedatelev, A. Mirzabekov, A., (2005). Evaluation of hybridisation on oligonucleotide microarrays for analysis of drug-resistant *Mycobacterium tuberculosis*. Clin Microbiol Infect 11, 531-9.
Guinn, K.M., Hickey, M.J., Mathur, S.K., Zakel, K.L., Grotzke, J.E., Lewinsohn, D.M., Smith, S. Sherman, D.R., (2004). Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium tuberculosis*. Mol Microbiol 51, 359-370.
Gutacker, M.M., Mathema, B., Soini, H., Shashkina, E., Kreiswirth, B.N., Graviss, E.A. Musser, J.M., (2006a). Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. J Infect Dis 193, 121-128.
Gutacker, M.M., Mathema, B., Soini, H., Shashkina, E., Kreiswirth, B.N., Graviss, E.A. Musser, J.M., (2006b). Single-nucleotide polymorphism-based population genetic analysis of Mycobacterium tuberculosis strains from 4 geographic sites. J Infect Dis 193, 121-8.
Gutacker, M.M., Smoot, J.C., Migliaccio, C.A., Ricklefs, S.M., Hua, S., Cousins, D.V., Graviss, E.A., Shashkina, E., Kreiswirth, B.N. Musser, J.M., (2002). Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. Genetics 162, 1533-1543.
Gutierrez, M.C., Brisse, S., Brosch, R., Fabre, M., Omais, B., Marmiesse, M., Supply, P. Vincent, V., (2005). Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. PLoS Pathog 1, e5.
Hazbon, M.H., (2004). Recent advances in molecular methods for early diagnosis of tuberculosis and drug-resistant tuberculosis. Biomedica 24 Supp 1, 149-62.
Heep, M., Brandstatter, B., Rieger, U., Lehn, N., Richter, E., Rusch-Gerdes, S. Niemann, S., (2001). Frequency of rpoB mutations inside and outside the cluster I region in rifampin-resistant clinical *Mycobacterium tuberculosis* isolates. J Clin Microbiol 39, 107-10.
Heersma, H.F., Kremer, K. van Embden, J.D., (1998). Computer analysis of IS6110 RFLP patterns of *Mycobacterium tuberculosis*. Methods Mol Biol 101, 395-422.
Heifets, L., (2000). Conventional methods for antimicrobial susceptibility testing of *Mycobacterium tuberculosis*. Academic Publishers, Dordrecht, The Netherlands.
Heifets, L.B., (1996). Clinical mycobacteriology. Drug susceptibility testing. Clin Lab Med 16, 641-56.
Hermans, P.W., van Soolingen, D., Dale, J.W., Schuitema, A.R., McAdam, R.A., Catty, D. van Embden, J.D., (1990). Insertion element IS986 from *Mycobacterium tuberculosis*: a useful tool for diagnosis and epidemiology of tuberculosis. J Clin Microbiol 28, 2051-8.
Hermans, P.W., van Soolingen, D. van Embden, J.D., (1992). Characterization of a major polymorphic tandem repeat in *Mycobacterium tuberculosis* and its potential use in the epidemiology of *Mycobacterium kansasii* and *Mycobacterium gordonae*. J Bacteriol 174, 4157-65.
Hillemann, D., Weizenegger, M., Kubica, T., Richter, E. Niemann, S., (2005). Use of the genotype MTBDR assay for rapid detection of rifampin and isoniazid resistance in *Mycobacterium tuberculosis* complex isolates. J Clin Microbiol 43, 3699-703.
Hirsh, A., Tsolaki, E., A. G., DeRiemer, K., Feldman, M.W. Small, P.M., (2004). Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. Proc. Natl. Acad. Sci. USA 101, 4871-4876.

Hunter, P.R. Gaston, M.A., (1988). Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol 26, 2465-6.
Idigoras, P., Beristain, X., Iturzaeta, A., Vicente, D. Perez-Trallero, E., (2000). Comparison of the automated nonradiometric Bactec MGIT 960 system with Lowenstein-Jensen, Coletsos, and Middlebrook 7H11 solid media for recovery of mycobacteria. Eur J Clin Microbiol Infect Dis 19, 350-4.
Jansen, R., Embden, J.D., Gaastra, W. Schouls, L.M., (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol 43, 1565-75.
Jiao, W.W., Mokrousov, I., Sun, G.Z., Li, M., Liu, J.W., Narvskaya, O. Shen, A.D., (2007). Molecular characteristics of rifampin and isoniazid resistant *Mycobacterium tuberculosis* strains from Beijing, China. Chin Med J (Engl) 120, 814-9.
Johansen, I., Thomsen, V., Marjamaki, M., Sosnovskaja, A. Lundgren, B., (2004). Rapid, automated, nonradiometric susceptibility testing of *Mycobacterium tuberculosis* complex to four first-line antituberculous drugs used in standard short-course chemotherapy. Diagn Microbiol Infect Dis 50, 103-7.
Kam, K., Yip, C.W., Tse, L.W., Wong, K.L., Lam, T.K., Kremer, K., Au, B.K. van Soolingen, D., (2005). Utility of mycobacterial interspersed repetitive unit typing for differentiating multidrug-resistant *Mycobacterium tuberculosis* isolates of the Beijing family. J Clin Microbiol. 43(1), 306-313.
Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., Bunschoten, A., Molhuizen, H., Shaw, R., Goyal, M. van Embden, J., (1997). Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 35, 907-14.
Kaufhold, A., Podbielski, A., Baumgarten, G., Blokpoel, M., Top, J. Schouls, L., (1994). Rapid Typing of group A streptococci by the use of DNA amplification and non-radioactive allele-specific oligonucleotide probes. FEMS Microbiol Lett 19-26.
Kazue, H., Chiyojia., A. Mitsuyoshi, T., (1999). Mutations in the rpoB gene of rifampin-resistant *Mycobacterium tuberculosis* strains isolated mostly in Asian countries and their rapid detection by line probe assay. J Clin Microbiol 378, 2663-2666.
Kent, P.T. Kubica, G.P., (1985). Public Health Mycobacteriology. A guide for the Level III Laboratory. Atlanta, GA: CDC.
Khue, P.M., Truffot-Pernot, C., Texier-Maugein, J., Jarlier, V. Robert, J., (2007). Ten years of prospective surveillance of *Mycobacterium tuberculosis* drug resistance in France (1995-2004). Eur Respir J.
Kim, S.Y., Park, Y.J., Song, E., Jang, H., Kim, C., Yoo, J. Kang, S.J., (2006). Evaluation of the CombiChip Mycobacteria Drug-Resistance detection DNA chip for identifying mutations associated with resistance to isoniazid and rifampin in *Mycobacterium tuberculosis*. Diagn Microbiol Infect Dis 54, 203-10.
Koch, R., (1882). Die Aetiologie der Tuberculose. Klinische Wochenschr. 19, 221-230.
Kovalev, S.Y., Kamaev, E.Y., Kravchenko, M.A., Kurepina, N.E. Skorniakov, S.N., (2005). Genetic analysis of *Mycobacterium tuberculosis* strains isolated in Ural region, Russian Federation, by MIRU-VNTR genotyping. Int J Tuberc Lung Dis 9, 746-52.
Kremer, K., Arnold, C., Cataldi, A., Gutierrez, M.C., Haas, W.H., Panaiotov, S., Skuce, R.A., Supply, P., van der Zanden, A.G. van Soolingen, D., (2005a). Discriminatory power and reproducibility of novel DNA typing methods for *Mycobacterium tuberculosis* complex strains. J Clin Microbiol 43, 5628-38.
Kremer, K., Au, B.K., Yip, P.C., Skuce, R., Supply, P., Kam, K.M. van Soolingen, D., (2005b). Use of Variable-Number Tandem-Repeat Typing To Differentiate *Mycobacterium tuberculosis* Beijing Family Isolates from Hong Kong and Comparison with IS6110 Restriction Fragment Length Polymorphism Typing and Spoligotyping. J Clin Microbiol 43, 314-20.
Kremer, K., van Soolingen, D., Frothingham, R., Haas, W.H., Hermans, P.W., Martin, C., Palittapongarnpim, P., Plikaytis, B.B., Riley, L.W., Yakrus, M.A., Musser, J.M. van Embden, J.D., (1999). Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol 37, 2607-18.

Kruuner, A., Hoffner, S.E., Sillastu, H., Danilovits, M., Levina, K., Svenson, S.B., Ghebremichael, S., Koivula, T. Kallenius, G., (2001). Spread of drug-resistant pulmonary tuberculosis in Estonia. J Clin Microbiol 39, 3339-45.

Kurepina, N., Likhoshvay, E., Shashkina, E., Mathema, B., Kremer, K., van Soolingen, D., Bifani, P. Kreiswirth, B.N., (2005). Targeted hybridization of IS6110 fingerprints identifies the W-Beijing *Mycobacterium tuberculosis* strains among clinical isolates. J Clin Microbiol 43, 2148-54.

Kurepina, N.E., Sreevatsan, S., Plikaytis, B.B., Bifani, P.J., Connell, N.D., Donnelly, R.J., van Sooligen, D., Musser, J.M. Kreiswirth, B.N., (1998). Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS6110 elements in *Mycobacterium tuberculosis*: non-random integration in the dnaA-dnaN region. Tuber Lung Dis 79, 31-42.

Laserson, K.F., Osorio, L., Sheppard, J.D. al, e., (2000). Clinical and programmatic mismanagement rather than community outbreak as the cause of chronic, drug-resistant tuberculosis in Buenaventura, Colombia, 1998. Int J Tuberc Lung Dis 673-683.

Le Flèche, P., Fabre, M., Denoeud, F., Koeck, J.L. Vergnaud, G., (2002). High resolution, on-line identification of strains from the *Mycobacterium tuberculosis* complex based on tandem repeat typing. BMC Microbiol 2, 37.

Le Flèche, P., Hauck, Y., Onteniente, L., Prieur, A., Denoeud, F., Ramisse, V., Sylvestre, P., Benson, G., Ramisse, F. Vergnaud, G., (2001). A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. BMC Microbiol 1, 2.

Lee, A.S., Teo, A.S. Wong, S.Y., (2001). Novel mutations in ndh in isoniazid-resistant *Mycobacterium tuberculosis* isolates. Antimicrob Agents Chemother 45, 2157-9.

Legrand, E., Filliol, I., Sola, C. Rastogi, N., (2001). Use of spoligotyping to study the evolution of the direct repeat locus by IS6110 transposition in *Mycobacterium tuberculosis*. J Clin Microbiol 39, 1595-9.

Lewis, K.N., Liao, R., Guinn, K.M., Hickey, M.J., Smith, S., Behr, M.A. Sherman, D.R., (2003). Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guerin attenuation. J Infect Dis 187, 117-123.

Li, W.M., Wang, S.M., Li, C.Y., Liu, Y.H., Shen, G.M., Zhang, X.X., Niu, T.G., Gao, Q., van Soolingen, D., Kremer, K. Duanmu, H.J., (2005). Molecular epidemiology of *Mycobacterium tuberculosis* in China: a nationwide random survey in 2000. Int J Tuberc Lung Dis 9, 1314-9.

Li, X., Zhang, Y., Shen, X., Shen, G., Gui, X., Sun, B., Mei, J., Deriemer, K., Small, P.M. Gao, Q., (2007). Transmission of drug-resistant tuberculosis among treated patients in Shanghai, China. J Infect Dis. 195, 864-869.

Liu, J., Pourcel, C., Zhao, X., Wan, K., Liu, Z., Zhang, Y., Wang, Q., Tan, Y. Xu, W., (2004). Study on genetic diversity of 65 *Mycobacterium tuberculosis* strains isolated in China based on 7 variable number of tandem DNA repeats. Chin J Microbiol Immunol 24, 733-737.

Liu, J.H., Zhang, L.S., Liu, Z.G., Huang, M.X., Zhao, X.Q., Wang, L. Wan, K.L., (2006). Study on the characteristics of mutation on *Mycobacterium tuberculosis* rifampicin-resistance gene. Zhonghua Liu Xing Bing Xue Za Zhi 27, 973-976.

Lopez, B., Aguilar, D., Orozco, H., Burger, M., Espitia, C., Ritacco, V., Barrera, L., Kremer, K., Hernandez-Pando, R., Huygen, K. van Soolingen, D., (2003). A marked difference in pathogenesis and immune response induced by different *Mycobacterium tuberculosis* genotypes. Clin Exp Immunol 133, 30-7.

Ma, X., Wang, H., Deng, Y., Liu, Z., Xu, Y., X., P., Musser, J.M. Graviss, E.A., (2006). rpoB Gene mutations and molecular characterization of rifampin-resistant *Mycobacterium tuberculosis* isolates from Shandong Province, China. J Clin Microbiol 44, 3409-3412.

Magdalena, J., Vachee, A., Supply, P. Locht, C., (1998). Identification of a new DNA region specific for members of *Mycobacterium tuberculosis* complex. J Clin Microbiol 36, 937-43.

Mahairas, G.G., Sabo, P.J., Hickey, M.J., Singh, D.C. Stover, C.K., (1996). Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. J Bacteriol 178, 1274-1282.

Marmiesse, M., Brodin, P., Buchrieser, C., Gutierrez, C., Simoes, N., Vincent, V., Glaser, P., Cole, S.T. Brosch, R., (2004). Macro-array and bioinformatic analyses reveal mycobacterial

'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. Microbiology 150, 483-96.
Mathema, B., Kurepina, N.E., Bifani, P.J. Kreiswirth, B.N., (2006). Molecular epidemiology of tuberculosis: current insights. Clin Microbiol Rev 19, 658-85.
Maus, C.E., Plikaytis, B.B. Shinnick, T.M., (2005). Mutation of tlyA confers capreomycin resistance in *Mycobacterium tuberculosis*. Antimicrob Agents Chemother 49, 571-7.
Mazars, E., Lesjean, S., Banuls, A.L., Gilbert, M., Vincent, V., Gicquel, B., Tibayrenc, M., Locht, C. Supply, P., (2001). High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. Proc Natl Acad Sci U S A 98, 1901-6.
McHugh, T.D. Gillespie, S.H., (1998). Nonrandom association of IS6110 and *Mycobacterium tuberculosis*: implications for molecular epidemiological studies. J Clin Microbiol 36, 1410-3.
Miesel, L., Weisbord, T.R., Amarcinkeviciene, J. al, e., (1998). NADH dehydrogenase defects confer isoniazid resistance and conditional lethalty in *Mycobacterium Smegmatis*. Bacteriol. 2459-2467.
Mokrousov, I., Jiao, W.W., Valcheva, V., Vyazovaya, A., Otten, T., Ly, H.M., Lan, N.N., Limeschenko, E., Markova, N., Vyshnevskiy, B., Shen, A.D. Narvskaya, O., (2006). Rapid detection of the *Mycobacterium tuberculosis* Beijing genotype and its ancient and modern sublineages by IS6110-based inverse PCR. J Clin Microbiol 44, 2851-6.
Mokrousov, I., Ly, H.M., Otten, T., Lan, N.N., Vyshnevskyi, B., Hoffner, S. Narvskaya, O., (2005). Origin and primary dispersal of the *Mycobacterium tuberculosis* Beijing genotype: clues from human phylogeography. Genome Res 15, 1357-64.
Mokrousov, I., Narvskaya, O., Otten, T., Vyazovaya, A., Limeschenko, E., Steklova, L. Vyshnevskyi, B., (2002). Phylogenetic reconstruction within *Mycobacterium tuberculosis* Beijing genotype in northwestern Russia. Res Microbiol 153, 629-37.
Morcillo, N., Zumarraga, M., Alito, A., Dolmann, A., Schouls, L., Cataldi, A., Kremer, K. van Soolingen, D., (2002). A low cost, home-made, reverse-line blot hybridisation assay for rapid detection of rifampicin resistance in *Mycobacterium tuberculosis*. Int J Tuberc Lung Dis 6, 959-65.
Morgan, M., Kalantri, S., Flores, L. Pai, M., (2005). A commercial line probe assay for the rapid detection of rifampicin resistance in *Mycobacterium tuberculosis*: a systematic review and meta-analysis. BMC Infect Dis 5, 62.
Mostowy, S., Cousins, D., Brinkman, J., Aranaz, A. Behr, M.A., (2002). Genomic deletions suggest a phylogeny for the *Mycobacterium tuberculosis* complex. J Infect Dis 186, 74-80.
Musser, J.M., (1995). Antimicrobial agent resistance in mycobacteria: molecular genetic insights. Clin Microbiol Rev 8, 496-514.
Namwat, W., Luangsuk, P. Palittapongarnpim, P., (1998). The genetic diversity of *Mycobacterium tuberculosis* strains in Thailand studied by amplification of DNA segments containing direct repetitive sequences. Int J Tuberc Lung Dis 2, 153-9.
Nikolayevskyy, V.V., Brown, T.J., Bazhora, Y.I., Asmolov, A.A., Balabanova, Y.M. Drobniewski, F.A., (2007). Molecular epidemiology and prevalence of mutations conferring rifampicin and isoniazid resistance in *Mycobacterium tuberculosis* strains from the southern Ukraine. Clin Microbiol Infect. 13(2), 129-138.
Oota, H., Kitano, T., Jin, F., Yuasa, I., Wang, L., Ueda, S., Saitou, N. Stoneking, M., (2002). Extreme mtDNA homogeneity in continental Asian populations. Am J Phys Anthropol 118, 146-53.
Palomino, J.C., (2005). Nonconventional and new methods in the diagnosis of tuberculosis: feasibility and applicability in the field. Eur Respir J 26, 339-50.
Palomino, J.C., (2006). Newer diagnostics for tuberculosis and multi-drug resistant tuberculosis. Curr Opin Pulm Med 12, 172-8.
Pfyffer, G., Bonato, D., Ebrahimzadeh, A. al, e., (1999). Multicenter laboratory validation of susceptibility testing of *Mycobacterium tuberculosis* against classical second-line and newer antimicrobial drugs by using the radiometric BACTEC 460 technique and the proportion method with solid media. J Clin Microbiol 37, 3179-86.
Pfyffer, G.E., Welscher, H.M., Kissling, P., Cieslak, C., Casal, M.J., Gutierrez, J. Rusch-Gerdes, S., (1997). Comparison of the Mycobacteria Growth Indicator Tube (MGIT) with radiometric and solid culture for recovery of acid-fast bacilli. J Clin Microbiol 35, 364-8.

Plikaytis, B.B., Marden, J.L., Crawford, J.T., Woodley, C.L., Butler, W.R. Shinnick, T.M., (1994). Multiplex PCR assay specific for the multidrug-resistant strain W of *Mycobacterium tuberculosis*. J Clin Microbiol 32, 1542-1546.

Qian, L., Van Embden, J.D., Van Der Zanden, A.G., Weltevreden, E.F., Duanmu, H. Douglas, J.T., (1999). Retrospective analysis of the Beijing family of *Mycobacterium tuberculosis* in preserved lung tissues. J Clin Microbiol 37, 471-4.

Rad, M.E., Bifani, P., Martin, C., Kremer, K., Samper, S., Rauzier, J., Kreiswirth, B., Blazquez, J., Jouan, M., van Soolingen, D. Gicquel, B., (2003). Mutations in putative mutator genes of *Mycobacterium tuberculosis* strains of the W-Beijing family. Emerg Infect Dis 9, 838-45.

Ramaswamy, S. Musser, J.M., (1998). Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. Tuber Lung Dis 79, 3-29.

Raviglione, M.C., (2003). The TB epidemic from 1992 to 2002. Tuberculosis (Edinb) 83, 4-14.

Reed, M.B., Domenech, P., Manca, C., Su, H., Barczak, A.K., Kreiswirth, B.N., Kaplan, G. Barry, C.E., 3rd, (2004). A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. Nature 431, 84-7.

Reed, M.B., Gagneux, S., Deriemer, K., Small, P.M. Barry, C.E., (2007). 3rd: The W-Beijing lineage of *Mycobacterium tuberculosis* overproduces triglycerides and has the DosR dormancy regulon constitutively upregulated. J Bacteriol 189, 2583-2589.

Rengarajan, J., Sassetti, C.M., Naroditskaya, V., Sloutsky, A., Bloom, B.R. Rubin, E.J., (2004). The folate pathway is a target for resistance to the drug para-aminosalicylic acid (PAS) in mycobacteria. Mol Microbiol 53, 275-82.

Rieder, H.L., Chonde, M., Myking, H., Urbanczik, R., Laszlo, A., Kim, S.-J., Van Deun, A. Trébucq, A., (1998). The Public Health Service National Tuberculosis Reference Laboratory and the National Laboratory Network Minimum Requirements, Role and Operation in a Low-Income Country International Union Against Tuberculosis and Lung Disease. 68, boulevard Saint-Michel, 75006 Paris, France.

Roberts, G.D., Goodman, N.L., Heifets, L., Larsh, H.W., Lindner, T.H., McClatchy, J.K., McGinnis, M.R., Siddiqi, S.H. Wright, P., (1983). Evaluation of the BACTEC radiometric method for recovery of mycobacteria and drug susceptibility testing of *Mycobacterium tuberculosis* from acid-fast smear-positive specimens. J Clin Microbiol 18, 689-96.

Roring, S., Scott, A.N., Glyn Hewinson, R., Neill, S.D. Skuce, R.A., (2004). Evaluation of variable number tandem repeat (VNTR) loci in molecular typing of *Mycobacterium bovis* isolates from Ireland. Vet Microbiol 101, 65-73.

Ruiz, M., Torres, M., Llanos, A., Arroyo, A., Palomares, J. Aznar, J., (2004). Direct detection of rifampin- and isoniazid-resistant *Mycobacterium tuberculosis* in auramine-rhodaminepositive sputum specimens by real-time PCR. J Clin Microbiol 42, 1585-9.

Rusch-Gerdes, S., Pfyffer, G.E., Casal, M., Chadwick, M. Siddiqi, S., (2006). Multicenter laboratory validation of the BACTEC MGIT 960 technique for testing susceptibilities of *Mycobacterium tuberculosis* to classical second-line drugs and newer antimicrobials. J Clin Microbiol 44, 688-92.

Sajduda, A., Brzostek, A., Poplawska, M., Augustynowicz-Kopec, E., Zwolska, Z., Niemann, S., Dziadek, J. Hillemann, D., (2004). Molecular characterization of rifampin- and isoniazid-resistant *Mycobacterium tuberculosis* strains isolated in Poland. J Clin Microbiol 42, 2425-31.

Seibert, F.B., (1934). The isolation and properties of the purified protein derivative of tuberculin. American Review of Tuberculosis 30, 713.

Shamputa, I.C., Rigouts And, L. Portaels, F., (2004). Molecular genetic methods for diagnosis and antibiotic resistance detection of mycobacteria from clinical specimens. Apmis 112, 728-52.

Singh, U.B., Arora, J., Suresh, N., Pant, H., Rana, T., Sola, C., Rastogi, N. Pande, J.N., (2007). Genetic biodiversity of *Mycobacterium tuberculosis* isolates from patients with pulmonary tuberculosis in India. Infect Genet Evol 7, 441-8.

Singh, U.B., Suresh, N., Bhanu, N.V., Arora, J., Pant, H., Sinha, S., Aggarwal, R.C., Singh, S., Pande, J.N., Sola, C., Rastogi, N. Seth, P., (2004). Predominant tuberculosis spoligotypes, Delhi, India. Emerg Infect Dis 10, 1138-42.

Skuce, R.A., McCorry, T.P., McCarroll, J.F., Roring, S.M., Scott, A.N., Brittain, D., Hughes, S.L., Hewinson, R.G. Neill, S.D., (2002). Discrimination of *Mycobacterium tuberculosis* complex bacteria using novel VNTR-PCR targets. Microbiology 148, 519-28.

Small, P.M., Hopewell, P.C., Singh, S.P., Paz, A., Parsonnet, J., Ruston, D.C., Schecter, G.F., Daley, C.L. Schoolnik, G.K., (1994). The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N Engl J Med 330, 1703-9.

Smith, I., (2003). *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. Clin Microbiol Rev 16, 463-96.

Smith, N.H., Dale, J., Inwald, J., Palmer, S., Gordon, S.V., Hewinson, R.G. Smith, J.M., (2003). The population structure of *Mycobacterium bovis* in Great Britain: clonal expansion. Proc Natl Acad Sci U S A 100, 15271-5.

Smittipat, N., Billamas, P., Palittapongarnpim, M., Thong-On, A., Temu, M.M., Thanakijcharoen, P., Karnkawinpong, O. Palittapongarnpim, P., (2005). Polymorphism of variable-number tandem repeats at multiple loci in *Mycobacterium tuberculosis*. J Clin Microbiol 43, 5034-43.

Smittipat, N. Palittapongarnpim, P., (2000). Identification of possible loci of variable number of tandem repeats in *Mycobacterium tuberculosis*. Tuber Lung Dis 80, 69-74.

Sola, C., Filliol, I., Legrand, E., Lesjean, S., Locht, C., Supply, P. Rastogi, N., (2003). Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. Infect Genet Evol 3, 125-33.

Sreevatsan, S., Pan, X., Stockbauer, K.E., Connell, N.D., Kreiswirth, B.N., Whittam, T.S. Musser, J.M., (1997). Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci U S A 94, 9869-74.

Srivastava, K., Das, R., Jakhmola, P. al., e., (2004). Correlation of mutations detected by INNO-LIPA with levels of rifampin resistance in *Mycobacterium tuberculosis*. Indian J Med Res 8, 100-105.

Suffys, P.N., Ivens de Araujo, M.E., Rossetti, M.L., Zahab, A., Barroso, E.W., Barreto, A.M., Campos, E., van Soolingen, D., Kremer, K., Heersma, H. Degrave, W.M., (2000). Usefulness of IS6110-restriction fragment length polymorphism typing of Brazilian strains of *Mycobacterium tuberculosis* and comparison with an international fingerprint database. Res Microbiol 151, 343-51.

Sun, Y.J., Lee, A.S., Wong, S.Y., Heersma, H., Kremer, K., van Soolingen, D. Paton, N.I., (2007). Genotype and phenotype relationships and transmission analysis of drug-resistant tuberculosis in Singapore. Int J Tuberc Lung Dis. 11(4), 436-442.

Supply, P., Allix, C., Lesjean, S., Cardoso-Oelemann, M., Rusch-Gerdes, S., Willery, E., Savine, E., de Haas, P., van Deutekom, H., Roring, S., Bifani, P., Kurepina, N., Kreiswirth, B., Sola, C., Rastogi, N., Vatin, V., Gutierrez, M.C., Fauville, M., Niemann, S., Skuce, R., Kremer, K., Locht, C. van Soolingen, D., (2006). Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J Clin Microbiol 44, 4498-4510.

Supply, P., Lesjean, S., Savine, E., Kremer, K., van Soolingen, D. Locht, C., (2001). Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. J Clin Microbiol 39, 3563-71.

Supply, P., Mazars, E., Lesjean, S., Vincent, V., Gicquel, B. Locht, C., (2000). Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. Mol Microbiol 36, 762-71.

Telenti, A., Honore, N., Bernasconi, C. al., e., (1997). Genotypic assessment of isoniazid rifampin resistance in *Mycobacterium tuberculosis*: a blind study at reference laboratory level. J Clin Microbiol 35, 719-723.

Thierry, D., Cave, M.D., Eisenach, K.D., Crawford, J.T., Bates, J.H., Gicquel, B. Guesdon, J.L., (1990). IS6110, an IS-like element of *Mycobacterium tuberculosis* complex. Nucleic Acids Res 18, 188.

Tsolaki, A.G., Gagneux, S., Pym, A.S., Goguet de la Salmoniere, Y.O., Kreiswirth, B.N., Van Soolingen, D. Small, P.M., (2005). Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. J Clin Microbiol 43, 3185-91.

Tsolaki, A.G., Hirsh, A.E., DeRiemer, K., Enciso, J.A., Wong, M.Z., Hannan, M., Goguet de la Salmoniere, Y.O., Aman, K., Kato-Maeda, M. Small, P.M., (2004). Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. Proc Natl Acad Sci U S A 101, 4865-70.

Tuberculosis, N.T.S.G.o.t.E.S.S.f., (2002). Report on fourth national epidemiological sampling survey of tuberculosis. Chin J. Tuberc Respir Dis 25 (1), 3-7.

van Crevel, R., Nelwan, R.H., de Lenne, W., Veeraragu, Y., van der Zanden, A.G., Amin, Z., van der Meer, J.W. van Soolingen, D., (2001). *Mycobacterium tuberculosis* Beijing genotype strains associated with febrile response to treatment. Emerg Infect Dis 7, 880-3.

van der Zanden, A.G., Hoentjen, A.H., Heilmann, F.G., Weltevreden, E.F., Schouls, L.M. van Embden, J.D., (1998). Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* complex in paraffin wax embedded tissues and in stained microscopic preparations. Mol Pathol 51, 209-14.

van der Zanden, A.G., Kremer, K., Schouls, L.M., Caimi, K., Cataldi, A., Hulleman, A., Nagelkerke, N.J. van Soolingen, D., (2002). Improvement of differentiation and interpretability of spoligotyping for *Mycobacterium tuberculosis* complex isolates by introduction of new spacer oligonucleotides. J Clin Microbiol 40, 4628-39.

van Embden, J.D., Cave, M.D., Crawford, J.T., Dale, J.W., Eisenach, K.D., Gicquel, B., Hermans, P., Martin, C., McAdam, R., Shinnick, T.M. et al., (1993). Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 31, 406-9.

van Embden, J.D., van Gorkom, T., Kremer, K., Jansen, R., van Der Zeijst, B.A. Schouls, L.M., (2000). Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. J Bacteriol 182, 2393-401.

van Klingeren, B., Dessens-Kroon, M., van der Laan, T., Kremer, K. van Soolingen, D., (2007). Drug susceptibility testing of *Mycobacterium tuberculosis* complex using a high throughput, reproducible, absolute concentration method. J Clin Microbiol.

van Soolingen, D., de Haas, P.E., Hermans, P.W., Groenen, P.M. van Embden, J.D., (1993). Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. J Clin Microbiol 31, 1987-95.

van Soolingen, D., Hermans, P.W., de Haas, P.E., Soll, D.R. van Embden, J.D., (1991). Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. J Clin Microbiol 29, 2578-86.

van Soolingen, D., Qian, L., de Haas, P.E., Douglas, J.T., Traore, H., Portaels, F., Qing, H.Z., Enkhsaikan, D., Nymadawa, P. van Embden, J.D., (1995). Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J Clin Microbiol 33, 3234-8.

Veen, J., (1992). Microepidemics of tuberculosis: the stone-in-the-pond principle. Tubercle Lung Dis 73, 73-76.

Vergnaud, G. Pourcel, C.: Multiple Locus VNTR (Variable Number of Tandem Repeat) Analysis. In: Stackebrandt, E. (Ed.), Molecular Identification, Systematics, and Population Structure of Prokaryotes. Springer-Verlag, Berlin Heidelberg, 2006, pp. 83-104.

WHO, (2004a). Anti-tuberculosis drug resistance in the world. Report No. 3. Prevalence and trends. WHO/HTM/TB/2004.

WHO, (2004b). Global tuberculosis control; surveillance, planning, financing. WHO/HTM/TB 2004. 331.

WHO, (2007). Extensively drug-resistant tuberculosis.

XDR, (2006). Extensively drug-resistant tuberculosis (XDR-TB): recommendations for prevention and control. Wkly Epidemiol Rec 81, 430-2.

Yang, Z.H., Ijaz, K., Bates, J.H., Eisenach, K.D. Cave, M.D., (2000). Spoligotyping and polymorphic GC-rich repetitive sequence fingerprinting of *Mycobacterium tuberculosis* strains having few copies of IS6110. J Clin Microbiol 38, 3572-6.

Zhang, Y., Garbe, T., Young, D. al, e., (1993). Transformation with katG restoresisoniazid-sensitivity in *Mycobacterium Tuberculosis* isolates resisant to a range of drugcontrations. Mol Microbiol 521-524.
Zhang, Y., Vilcheze, C. Jacobs, W.R.J., (2005). Mechanisms of drug resistance in *Mycobacterium tuberculosis*. American Society for Microbiology, Washington, D.C.