Original Article

A new Multilocus Sequence Analysis Scheme for *Mycobacterium tuberculosis*[*]

LU Bing[‡], DONG Hai Yan[‡], ZHAO Xiu Qin[‡], LIU Zhi Guang[‡], LIU Hai Can, ZHANG Yuan Yuan, JIANG Yi, and WAN Kang Lin[#]

State Key Laboratory for Infectious Diseases Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, & National Tuberculosis Reference Laboratory Beijing 102206, China

Abstract

Objective Tuberculosis remains one of the most serious infectious diseases in the world. In this study, a scheme of *Mycobacterium tuberculosis* (*M. tuberculosis*) multilocus sequence analysis (MLSA) was established for the phylogenetic and epidemiology analysis.

Methods To establish the scheme of *M. tuberculosis* MLSA, the genome of H37Rv, CCDC5079 and CCDC5180 were compared, and some variable genes were chosen to be the MLSA typing scheme. 44 *M. tuberculosis* clinical isolates were typed by MLSA, IS6110-RFLP, and soligotyping, to evaluate the MLSA methods.

Results After comparison of the genome, seven high discrimination gene loci (*recX*, *rpsL*, *rmlC*, *rpmG1*, *mprA*, *gcvH*, *ideR*) were chosen to be the MLSA typing scheme finally. 11 variable SNP sites of those seven genes were found among the 44 *M. tuberculosis* isolate strains and 11 sequence types (STs) were identified. Based on the Hunter-Gaston Index (HGI), MLSA typing was not as good for discrimination at the strain level as IS6110-RFLP, but the HGI was much better than that of spoligotyping. In addition, the MEGA analysis result of MLSA data was similar to spoligotyping/PGG lineage, showing a strong phylogenetic signal in the modern strains of *M. tuberculosis*. The MLSA data analysis by eBURST revealed that 4 sequence types (ST) came into a main cluster, showing the major clonal complexes in those 44 strains.

Conclusion MLSA genotyping not only can be used for molecular typing, but also is an ideal method for the phylogenetic analysis for *M. tuberculosis*.

Key words: *M. tuberculosis*; Multilocus sequence analysis; Genotyping

Biomed Environ Sci, 2012; 25(6):620-629 *doi: 10.3967/0895-3988.2012.06.003* *ISSN:0895-3988*

www.besjournal.com(full text) *CN: 11-2816/Q*

INTRODUCTION

China has a heavy disease burden of tuberculosis. 2009 saw an estimated tuberculosis incidence of 96/100 000 population and a mortality rate of 12/100 000 population[1].

Identification of the relationships between different clinical strains in different regions has great significance to public health and remains a high

[*]This research was supported by the fund of State Key Laboratory for Infectious Diseases Prevention and Control (2011SKLID208), and the project "Transmission Mode of Tuberculosis" of National Key Program of Mega Infectious Diseases (2008ZX100/03-010).

[#]Corresponding author, WAN Kang Lin, Tel: 86-10-58900776; Fax: 86-10-58900776; E-mail: wankanglin@icdc.cn

[‡]LU Bing, DONG Hai Yan, ZHAO Xiu Qin, LIU Zhi Guang contributed equally to this work.

Biographical note for the first author: LU Bing, female, born in 1974, Ph.D from Chinese Center for Disease Control and Prevention, associate professor, majoring in pathogen biology.

Received : November 23, 2011; Accepted: April 30, 2012

priority for epidemiology research. DNA fingerprinting of *Mycobacterium tuberculosis* (*M. tuberculosis*) is a powerful aid in the molecular epidemiologic management and transmission research. It was ever used for tuberculosis case finding and contact tracing[2-5]. The most common method of DNA fingerprinting was IS6110-RFLP, spoliogotyping and variable number tandem repeats (VNTR). IS6110-RFLP typing was the most widely used method for differentiating *M. tuberculosis* strains[6-8]. However, RFLP analysis required growth of mycobacterial colonies, which consumed much time in epidemiology research. Spoligotyping, a method based on polymerase chain reaction (PCR) amplification of repetitive genetic elements in the direct repeat region[9], could get the result within one day, but the discrimination of spligotyping appeared to be not enough for *M. tuberculosis*, especially for Beijing family strains[10], which was prevalent in Far-East-Asia and accounted for 11% in the fourth international spoligotyping database (SpolDB4)[11]. VNTR is a promising, relatively easy and rapid method; however, the discriminatory power of the VNTR loci are often varied, and the methods were not fully evaluated in areas where Beijing genotype strains predominated[12].

Multilocus sequence typing (MLST) was a typing technique for strain characterization that indexed variation in multiple housekeeping genes[13]. It was proposed as a portable, universal, and definitive method for characterizing bacteria[14]. MLST could help discover the source of epidemic and judge a recurrence or a new infection. MLST could type isolates directly from clinical materials[15]. In addition, MLST data, or the nucleotide sequence from internal fragments of housekeeping genes, was suitable for comparing between Laboratories and evolutionary, phylogenetic or population genetic studies[14,16]. Furthermore, because the loci of MLST were all functional protein encoded genes, it was imaginable to understand the relationship between MLST genotyping and phynotyping. Although MLST had these advantages, some researchers considered that MLST scheme was not suitable for *M. tuberculosis* as housekeeping genes from different isolates were always identical[15]. The highly conserved *M. tuberculosis* genome made the phylogenetic analysis by multilocus sequencing of housekeeping genes unimformative[17]. But Hershberg et al. used sequence of as many as 89 housekeeping and non-housekeeping genes to study the *M. tuberculosis* phylogeny[18]. Abadia et al. used seven 3R (replication, repair and recombination functions) genes to accurately classify *M. tuberculosis* complex isolates at the sublineage level by single nucleotide polymorphisms[19]. These researchers got successive results for comprehensive understanding of the phylogenetical relation among *M. tuberculosis* isolates.

In order to explore whether non-housekeeping genes can be used for the MLST method for *M. tuberculosis*, we conducted a study to evaluate the discrimination power of every locus and chose high discrimination gene loci to establish the scheme of *M. tuberculosis* multilocus sequence genotype method. In this study, we compared the genome of H37Rv with that of two clinical isolates from China and chose variable genes for typing *M. tuberculosis*. Generally speaking, less than 450 bp fragment of genes should be used[15], but with the developed sequencing technique, the sequencing result of less than 700 bp fragment was satisfactory. We chose the gene loci which had SNP and the length under 700 bp. Then we calculated the discrimination power measured by Hunter-Gaston Index (HGI) of every gene loci. Since MLST achieves high levels of discrimination when using only seven loci[15], we chose seven top HGI value gene loci to establish the scheme of *M. tuberculosis* multilocus sequence genotype.

M. tuberculosis multilocus sequence genotype method was assigned directly by nucleotide sequencing of variable genes such as virulence related gene, antibody encoding genes, and drug resistance related genes[20], and it was different from routine MLST typing which relied on the housekeeping genes. Then we named *M. tuberculosis* multilocus sequence genotype method as the multilocus sequence analysis (MLSA).

MATERIALS AND METHODS

Genome Comparing and Gene Loci

We compared two genomes (CCDC5079, CCDC5180) with that of H37Rv (GenBank: NC 000962 at NCBI) separately to find the high discrimination gene loci of MLSA scheme. The two strains, CCDC5079, CCDC5180, were isolated from patients with secondary pulmonary tuberculosis in China in 2005, and we got the genome sequence of these two strains by classical Sanger genome sequencing.

MLSA were assigned directly by nucleotide

sequencing. Generally speaking, less than 450 bp fragment of genes was used[15], but with the developed sequencing technique, the sequencing result of less than 700 bp fragment was satisfactory. The criteria with which we chose the genes include[15]: (1) SNP located in gene; (2) the length of the gene under 700 bp; and (3) encoding functional protein.

Isolates

44 strains isolates were selected from pulmonary tuberculosis patients from 2002 to 2006 in eight provinces of China. The patients were diagnosed by the clinical symptoms, the X ray and sputum stain tests in the local lung disease hospitals or tuberculosis hospitals. Those patients were proven to have no contact on the basis of standard epidemiological investigations.

Every gene locus of 44 strains of *M. tuberculosis* was sequenced, and the discrimination power of every locus was calculated. To evaluate the MLSA scheme, those 44 strains were also typed by IS6110-RFLP and spoligotyping.

Nucleotide Sequencing of Gene Fragments

The PCR primers to amplify the gene loci from chromosomal DNA of the 44 *M. tuberculosis* were listed in Table 1.

2×Taq PCR MasterMix was added with 5 μL digested-ligated DNA and 1 μL primers to form a final volume of 40 μL. The PCR parameters were: initial denaturation at 94 °C for 10 min followed by 30 cycles of 94 °C for 30 s, 61 °C for 30 s, 72 °C for 45 s, and a final extension of 10 min at 72 °C. Sequencing of both strands of the amplified fragments was achieved by the 3730 genomic analysis system (ABI), and packed by Sequencher and DNAstar software.

Discrimination Power of Each Fragment

The sequences of every gene loci of 44 strains were aligned by DNAStar software. Owing to the single nucleotide polymorphism (SNP), we got different typing results by every gene locus. The discriminatory power of each locus for all 44 isolates was measured with Hunter-Gaston Index (HGI). HGI calculated the probability of different genotypes in any two unrelated isolates by using the following formula[21]:

$$HGI = 1 - \frac{1}{N(N-1)} \sum_{j=1}^{s} n_j(n_j - 1)$$

Where D is the numerical index of discrimination, N is the total number of strains in the typing scheme, s is the total number of different strain types, and n_j is the number of strains belonging to the j type.

Because MLSA achieved high levels of discrimination when using only seven loci[15], seven higher HGI gene loci were chosen for the scheme of tuberculosis MLSA method.

MLSA Data Analysis

A MLSA typing system was developed based on the seven high HGI gene loci. MLSA was performed with each of the strains as described previously[22-23]. For each gene, the sequences from the 44 strains were compared; distinct alleles were identified; and combination of the seven MLSA allelic numbers was considered as a unique multilocus sequence type (ST). The relatedness between each ST was shown as dendrogram, and constructed by BURST algorithm[24-25] with the arithmetic averages UPGMA (unweighted pair group cluster method with arithmetic mean) method.

The other analyses of those isolates were performed with the MEGA4 software[20]. First, the sequences of the seven gene loci were concatenated to produce a single sequence for each isolate. Second, a cluster map was constructed with the neighbor-joining tree method by MEGA4 software.

DNA Fingerprinting

Spoligotyping Spoligotyping which relies on the highly polymorphic DR locus in the *M. tuberculosis* was finished according to the method described by Kamerbeek et al.[26], the amplified biotin-labeled DR locus was hybridized to a set of 43 different DR spacers in a Miniblotter. The hybridized signals were detected by chemiluminescence and were visualized as profiles of discrete dots.

IS6110-RFLP IS6110-RFLP was performed by the standard methods described previously[27-28]. Chromosomal DNA of mycobacterium isolates was digested with *Pvu* II, and after being mixed with marker DNA, the fragments were separated by overnight electrophoresis. The fragments were transferred to nylon membranes, and hybridized with peroxidase-labeled IS6110 probe. After washing with washing buffer, the membranes were rehybridized by peroxidase-labeled maker DNA.

Data Analysis The results of spoligotyping and IS6110-RFLP were all analyses made by Bionumerics software Version 5.0.

Table 1. The Information and Hunter-Gaston Index of 16 Gene Loci

Gene	Product	PrimerF (5'-3')	PrimerR (5'-3')	Length	Hunter-Gaston Index	CI	Alleles Number
recX	RecA regulator RecX	ttagacgcgtcggcgctccc	atgacggtgtcctgcccgcc	525	0.502	0.483 - 0.521	2
rpsL	30S ribosomal protein S12	atgccaaccatccagcagct	tcagcccttctctttcttag	375	0.458	0.390 - 0.526	3
rmlC	dTDP-4-keto-6-deoxyglucose epimerase	atgaaagcacgcgaactcga	ctaggtgccgcgcatctccc	609	0.426	0.373 - 0.479	2
rpmG1	50S ribosomal protein L33	tcagcgttcctcgcgaaagt	atggcgcgcaccgacatccg	165	0.426	0.373 - 0.479	2
mprA	Mycobacterial persistence regulator MRPA	gtgtccgtgcgaattcttgt	tcagggtggtgtttcacgta	693	0.411	0.368 - 0.518	4
gcvH	glycine cleavage system protein H	gtgagcgatatcccgtccga	tcactcggtcagtgtgccgc	405	0.369	0.298 - 0.440	3
ideR	Iron-dependent repressor and activator ider	atgaacgagttggttgatac	tcagactttctcgaccttga	693	0.359	0.294 - 0.425	2
mpt53	Soluble secreted antigen MPT53 precursor	tcaggacgtcagcgcagcca	atgagtcttcgcctggtgtc	522	0.304	0.245 - 0.419	2
tatB	sec-independent translocase	gtgttcgccaacatcggttg	ctaggtggcatcgctgtcaa	396	0.304	0.234 - 0.375	2
pncA	Pyrazinamidase (PZase)	tcaggagctgcaaaccaact	atgcgggcgttgatcatcgt	561	0.296	0.234 - 0.375	8
rplO	50S ribosomal protein L15	gtgacgctcaagctgcatga	ctagagctcggtggctgaac	441	0.09	0.210 - 0.382	3
ribC	riboflavin synthase subunit alpha	atgttcaccggaattgttga	tcagccggcgctccgcatta	606	0.09	0.035 - 0.145	3
rplM	50S ribosomal protein L13	tcattgcgccacctgcttga	gtgcccacgtacgcgcccaa	444	0.09	0.035 - 0.145	3
apt	Adenine phosphoribosyltransferase APT	tcacaggcggctcaggctgt	gtgtgccatggcggtacatg	672	0.045	0.035 - 0.145	2
rplQ	50S ribosomal protein L17	ctaattctcgggggcctctg	atgcccaagcctaccaaggg	543	0.045	0.005 - 0.086	2
hbhA	Iron-regulated heparin binding hemagglutinin HBHA (ADHESIN)	atggctgaaaactcgaacat	ctacttctgggtgaccttct	600	0.045	0.005 - 0.086	2

RESULTS

Screen Analysis of Appropriate Gene Loci

There were 2332 SNP discovered between the genomes of *M. tuberculosis* strains H37Rv and CCDC5079, and 2251 SNP between H37Rv and CCDC5180.

16 gene loci (*recX*, *rpsL*, *rmlC, rpmG1, mprA*, *gcvH*, *ideR*, *mpt53*, *tatB*, *pncA, rplO*, *ribC*, *rplM*, *apt*, *rplQ,* and *hbhA*) were found to satisfy the initial inclusion rules. The information such as encoded product, location in the whole H37Rv genome of 16 genes, is showed in Table 1.

Identification of Seven Gene Loci for the Scheme

HGI of diversity for the genotyping by each of the 16 genes was calculated, which, together with the confidence interval (CI) and grouping number of each gene, was also showed in Table 1. The gene *recX* has the highest HGI (0.502) while *hbhA* has the lowest one (0.045). The top 7 genes (*recX*, *rpsL*, *rmlC, rpmG1, mprA*, *gcvH*, *ideR*) were chosen to be the gene loci for MLSA scheme. In addition, the seven genes were mapped on a physical genome map and identified as unlinked (Figure 1). There has no overlapped area among these seven genes. The number of alleles of these seven genes ranged from 2 to 4 (Table 2).

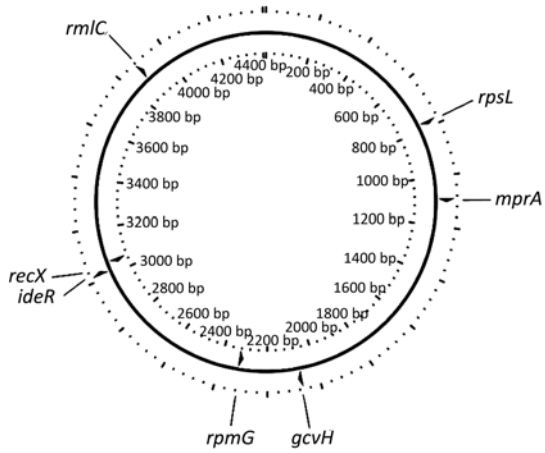

Figure 1. Chromosomal locations of seven gene loci. The locations are drawn on the physical map of strain *Mycobacterium tuberculosis* H37Rv, complete genome-1. 4411532.

Table 2. Variable Nucleotide Sites and Alleles Identified at Each Locus

Alleles	*recX*	*rpsL*		*rmlC*	*rpmG*1	*mprA*			*ideR*	*gcvH*	
	175*	128	263	345	137	*208*	*425*	*627*	457	239	362
1	G	A	A	A	G	G	C	C	C	A	C
2	C	G	A	C	T	G	T	C	A	A	A
3		A	G			G	T	T		C	C
4						A	C	C			

Note. *The number of column name represented the position of the gene locus.

MLSA Data Analysis

MLSA typing was performed by the sequence of seven high HGI gene loci. We found 11 variable SNP sites among the 44 *M. tuberculosis* isolate strains, as displayed in Table 2. All the strains were characterized by the combination of seven MLSA allelic numbers, or allelic profiles. Each unique allelic profile was assigned a sequence type (ST) number. There were 11 STs identified from 44 strains, which are showed in Table 3.

eBURST Analysis eBURST used the STs as input and divided those strains into groups in which all the STs in the same group were sharing ≥6 out of 7 loci, resulting in non-overlapping groups or clonal complexes[29]. All the STs determined by MLSA with the BURST algorithm (Figure 2), and 4 STs (including ST6, 9, 10, 11) gathered together in the center of the graph and formed the main cluster. ST9 was the founder of the main cluster and had 3 single locus variants (SLVs) and 1 double locus variant (DLVs).

MEGA Analysis Seven gene loci of every strain were linked to be a single sequence and analyzed. The typing analysis of 44 isolates was performed and the dendrogram map (based on Neighbor Joining Method) is showed in Figure 3. The HGI of MLSA was 0.815 (CI: 0.778-0.852).

Table 3. 44 Isolates Descriptions and MLSA Results

Code	Origin*	*recX*	*rmlC*	*rpmG1*	*mprA*	*rpsL*	*gcvH*	*ideR*	ST
15	Guangxi	1	1	1	1	1	1	1	ST1
16	Guangxi	1	1	1	1	1	1	1	ST1
24	Xinjiang	1	1	1	1	1	1	1	ST 1
27	Guangxi	1	1	1	1	1	1	1	ST 1
34	Xizang	1	1	1	1	1	1	1	ST 1
37	Sichuan	1	1	1	1	1	1	1	ST 1
43	Sichuan	1	1	1	1	1	1	1	ST 1
49Rv	H37Rv	1	1	1	1	1	1	1	ST 1
33	Xizang	1	1	1	2	1	1	1	ST 2
30	Xinjiang	1	1	1	2	1	2	2	ST 3
20	Henan	1	1	1	3	1	2	2	ST 4
41	Sichuan	1	1	1	3	1	2	2	ST 4
8	Fujian	1	1	1	4	1	3	1	ST 5
10	Gansu	1	2	2	3	1	2	2	ST 6
36	Xizang	1	2	2	3	1	2	2	ST 6
13	Gansu	1	2	2	3	2	2	2	ST 7
31	Xizang	1	2	2	3	2	2	2	ST 7
40	Sichuan	1	2	2	3	2	2	2	ST 7
32	Xizang	1	2	2	3	3	2	2	ST 8
1	Beijing	2	2	2	3	1	2	2	ST 9
2	Beijing	2	2	2	3	1	2	2	ST 9
4	Beijing	2	2	2	3	1	2	2	ST 9
5	Fujian	2	2	2	3	1	2	2	ST 9
6	Fujian	2	2	2	3	1	2	2	ST 9
7	Fujian	2	2	2	3	1	2	2	ST 9
9	Gansu	2	2	2	3	1	2	2	ST 9
12	Gansu	2	2	2	3	1	2	2	ST 9
14	Guangxi	2	2	2	3	1	2	2	ST 9
17	Guangxi	2	2	2	3	1	2	2	ST 9
22	Henan	2	2	2	3	1	2	2	ST 9
25	Xinjiang	2	2	2	3	1	2	2	ST 9
26	Xinjiang	2	2	2	3	1	2	2	ST 9
29	Xinjiang	2	2	2	3	1	2	2	ST 9
35	Xizang	2	2	2	3	1	2	2	ST 9
45	Sichuan	2	2	2	3	1	2	2	ST 9
3	Beijing	2	2	2	3	2	2	2	ST 10
11	Gansu	2	2	2	3	2	2	2	ST 10
21	Henan	2	2	2	3	2	2	2	ST 10
23	Henan	2	2	2	3	2	2	2	ST 10
46	Sichuan	2	2	2	3	2	2	2	ST 10
47	Sichuan	2	2	2	3	2	2	2	ST 10
48	Sichuan	2	2	2	3	2	2	2	ST 10
19	Henan	2	2	2	3	3	2	2	ST 11
39	Sichuan	2	2	2	3	3	2	2	ST 11

Note. *Origin: the province of China.

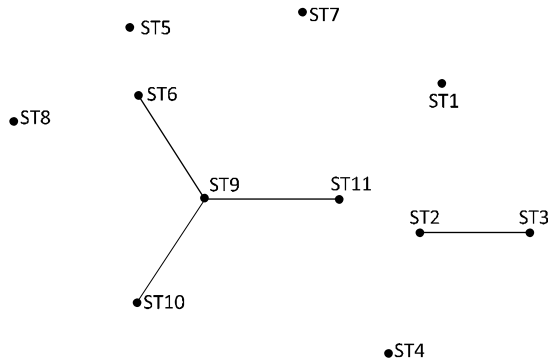

Figure 2. Clonal groupings among 44 isolate. Allelic profiles were analyzed by eBURST and groups were defined as sets of relations sharing identicalness at six of seven loci. ST9 was the putative founder of the clonal complex. Single locus variants were joined by lines.

Evaluation of the MLSA Scheme

Spoligootyping genotype showed that one main cluster of 44 isolates was Beijing family strains (36 strains) (Figure 3), which had the particularly spoligotype signature (absent in Spacer 1-34, and present in Spacer 35-43)[30], and the others included T1, U, CAS and *M. bovis*. From the perspective of phylogenetic, spoligotyping differentiated the strains into PGG1-3 lineage[31-32]. Beijing and CAS families belonged to PGG1 group and others to PGG3 except *M. bovis* strains (FJ07111). PGG1 and PGG3 groups were also exactly divided by the MLSA analysis (Figure 3). HGI of spoligotyping was 0.569 (CI: 0.483-0.654), and that of MLSA was much higher.

44 strains were genotyped by IS6110-RFLP and differentiated 44 distinct IS6110-RFLP patterns among these strains (Figure 3). HGI of IS6110-RFLP was 1 (CI: 0.996-1.000).

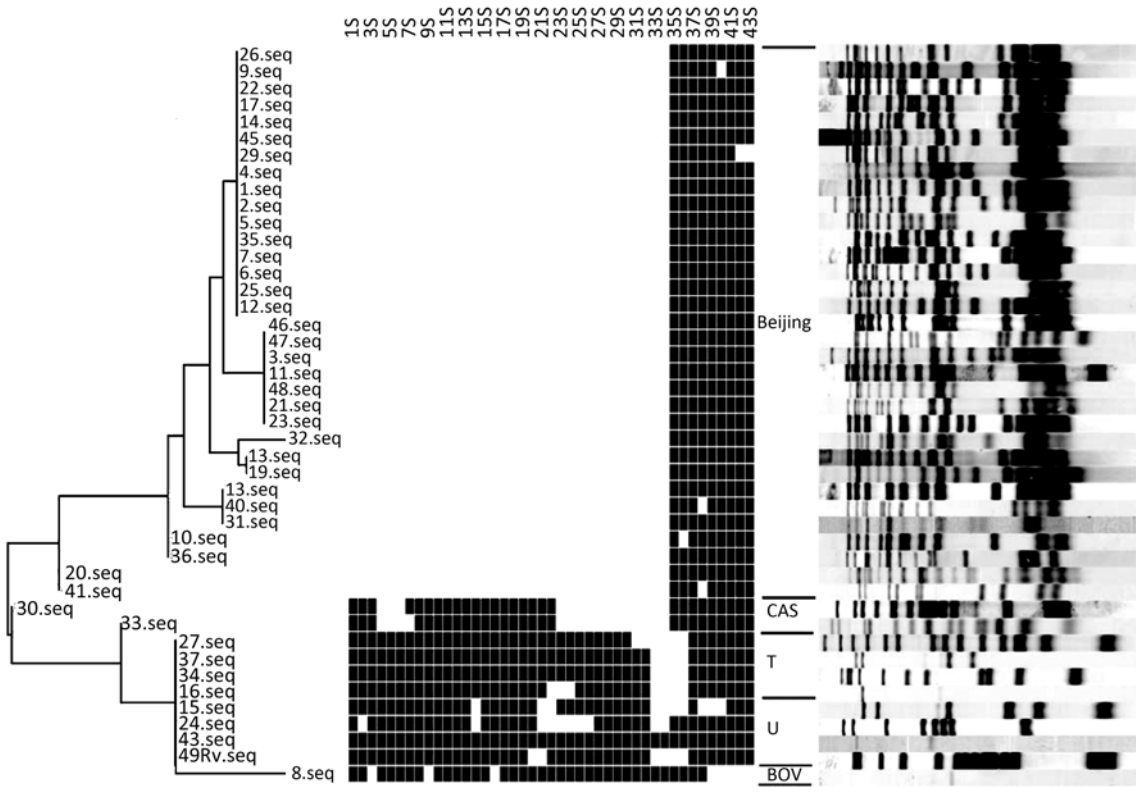

Figure 3. Dendrogams of 44 isolates typed by MLSA. **Left column**. Multilocus sequence analysis of closely related species. Evolution tree constructed by using MEGA4, showing clusters among 44 isolates of *M. tuberculosis*. The tree was built by using seven genes. **Middle column**. The spoligotyping pattern was showed from Space 1 to Space 43 (left to right); the right part showed the spoligotyping lineage. **Right column**. IS*6110*-RFLP pattern. The IS*6110*-RFLP pattern was shown from high to low molecular weight (left to right).

DISCUSSION

The MLSA Scheme

In this study, high discrimination power was the important guideline, so we obtained the seven gene loci with higher HGI: *recX*, *rpsL*, *rmlC, rpmG1, mprA*, *gcvH*, and *ideR.* The functions of some gene loci were associated with the special characters, such as pathogen virulence (*ideR*, *mprA*), and bacterium growth (*rmlC*).

The iron-dependent regulator (IdeR), encoded by *ideR*, is a well-characterized global regulator responsible for maintaining iron homeostasis in *Mycobacterium tuberculosis*. Iron plays a central role in enzymic reactions involved in electron transport, nucleic acid synthesis and oxidative stress defense. Therefore, iron is critical for most living systems. These suggest that ideR is an essential gene in tuberculosis[33].

MprA (encoded by *mprA*), together with MprB, is an intact two-component signal-transducing pair, and plays important roles in *M. tuberculosis* physiology and pathogenesis. The system is found to be necessary in *M. tuberculosis* for establishment and maintenance of persistent infection in a tissue- and stage-specific fashion[34-35].

rmlC gene encoding dTDP-4-keto-6-deoxyglucose epimerase, is essential for mycobacterial cell-wall synthesis, and the production RmlC is structure-related protein, which is considered to be the most promising drug target[36].

Those 'specialist' phenotypes, which include clinically relevant properties such as virulence, are more often encoded on accessory genes or elements, can be passed 'horizontally' between unrelated strains by homologous recombination or the dissemination of mobile elements and rapidly disseminated throughout a population and equally rapidly lost. But *M. tuberculosis* is not, as *M. tuberculosis* is a strictly clonal organism, with no evidence of lateral gene transfer[20]. Those genes of *M. tuberculosis* could be inherited into the descents steadily, and the stability was very valuable for the MLSA method.

Hershberg et al. used sequence of 89 genes to study the *M. tuberculosis* phylogeny[18], and discovered a total of 370 SNP in 107 strains. Compared with those results, the diversity of seven gene loci in this study was the same. It has been recognized that the nature selection for increased transmissibility might drive pathogens toward increased virulence[37-38], and thus, it is reasonable to believe that some virulence-related genes have good variation for typing clinical isolates of *M. tuberculosis*.

Hershberg et al. considered that the evolution characteristics of tuberculosis bacteria could be identified by SNP, and much of the diversity had functional consequences and could affect the efficacy of new tuberculosis diagnosis, drugs, and vaccines. On this point, we agree with Hershberg, but to establish a method for epidemiological research, the cost of sequencing 89 genes was too expensive to accept. Therefore, we chose seven highly diverse gene sequences to be retained in the present MLSA scheme.

An epidemiologist or medical microbiologist is always asked about the questions: Can molecular genotyping data reveal the danger posed by an isolate? What is fundamental to the design of *M. tuberculosis* MLSA is the emphasis on those genes encoded by 'specialist' phenotypes. Our findings indicate that it is possible to understand the relationship between MLSA genotypes and clinical characters or epidemiological data.

MLSA, IS6110-RFLP, and Spoligotyping

Although more than thirty MLSA schemes (mostly for pathogenic bacteria) have been published and made available on the internet, this is a new method for *M. tuberculosis*. IS6110-RFLP and spoligotyping have been used for *M. tuberculosis* genotyping in epidemiologic studies and to indicate the relationship among strains for decades[27]. Compared with the HGI of IS*6110*-RFLP and spoligotyping, MLSA typing was not good for discrimination at the strain level as IS6110-RFLP, but the HGI was much better than that of spoligotyping. Thus, it is valuable to keep on the research of the MLSA scheme.

In addition, it is important to note that all MLSA-based trees were clustered into their spoligotyping/PGG lineages, indicating that the MLSA study gave a strong phylogenetic signal in the modern strains of *M. tuberculosis*. The limitation of spoligotyping lies in its failure in determining whether the loss of multiple adjacent spacers is due to a single or multiple evolutionary events. Comas et al. used the complete DNA sequences of 89 coding genes from a global strain collection, and showed that the multilocus sequence data were highly congruent and statistically robust, and suitable for the phylogenetic relationship analysis[39].

To be honest, the number of the isolates studied in this research was rather small, and Beijing family strains were much excessive, which was the main character of *M. tuberculosis* distribution in China[40]. So this new method needs to be validated by large samples in the future.

Utility of MLSA Typing for Epidemiology Research

MLST genotyping provided many new insights into phylogenetic relationships among members of the *M.tuberculosis* complex, and it was also an ideal method to perform the population analysis in large-scale epidemiology studies[13]. It helped epidemiologists much in surveillance and case control. According to a typical MLST allelic profile, a characterized isolate of Spanish clones of penicillin-resistant and multiple-antibiotic-resistant *S. pneumoniae* was identified and raised the advantages of combining MLST with more discriminate genes to improve characterization[41].

M.tuberculosis MLSA provides data that can also be used to probe the population and evolutionary biology of the species[15]. In this study, we got 11 STs from 44 strains and some STs were clustered in the middle of the map clearly (Figure 2), showing the major clonal complexes in those 44 strains. ST9 (*recX*-2, *rmlC*-2, *rpmG1*-2, *mprA*-3, *rpsL*-1, *gcvH*-2, and *ideR*-2) was identified to be the potential ancestral type. With this approach, epidemiologists can easily describe strain-specific variation in clinical phenotypes of *M. tuberculosis* and perform case control during laboratory-based efforts. However, it must be noticed that future molecular epidemiological research needs larger-scale data, and the 44 strains MLSA analysis was primarily a methodology study. Therefore, continued research is needed to perfect this method and obtain more knowledge of MLSA data distributed in China.

ACKNOWLEDGEMENTS

We would like to express our thanks to DU Peng Cheng and CHEN Chen for their data analyses, to LIU Mei and LI Zhao Na for their isolate genotyping, and to members of provincial CDCs for their strain collection.

REFERENCES

1. World Health Organization. http://www.who.int/gho/mdg/diseases/tuberculosis/en/index.html.
2. Edlin BR, Tokars JI, Grieco MH, et al. An outbreak of multidrug-resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. N Engl J Med, 1992; 326, 1514-21.
3. Jereb JA, Burwen DR, Dooley SW, et al. Nosocomial outbreak of tuberculosis in a renal transplant unit: application of a new technique for restriction fragment length polymorphism analysis of Mycobacterium tuberculosis isolates. J Infect Dis, 1993; 168, 1219-24.
4. Daley CL, Small PM, Schecter GF, et al. An outbreak of tuberculosis with accelerated progression among persons infected with the human immunodeficiency virus. An analysis using restriction-fragment-length polymorphisms. N Engl J Med, 1992; 326, 231-5.
5. Genewein A, Telenti A, Bernasconi C, et al. Molecular approach to identifying route of transmission of tuberculosis in the community. Lancet, 1993; 342, 841-4.
6. Beck-Sague C, Dooley SW, Hutton MD, et al. Hospital outbreak of multidrug-resistant Mycobacterium tuberculosis infections. Factors in transmission to staff and HIV-infected patients. JAMA, 1992; 268, 1280-6.
7. Pearson ML, Jereb JA, Frieden TR, et al. Nosocomial transmission of multidrug-resistant Mycobacterium tuberculosis. A risk to patients and health care workers. Ann Intern Med, 1992; 117, 191-6.
8. Alland D, Kalkut GE, Moss AR, et al. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. N Engl J Med, 1994; 330, 1710-6.
9. Brosch R, Gordon SV, Marmiesse M, et al. A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci USA, 2002; 99, 3684-9.
10. Cronin WA, Golub JE, Magder LS, et al. Epidemiologic usefulness of spoligotyping for secondary typing of Mycobacterium tuberculosis isolates with low copy numbers of IS6110. J Clin Microbiol, 2001; 39, 3709-11.
11. Brudey K, Driscoll JR, Rigouts L, et al. Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol, 2006; 6, 23.
12. Supply P, Allix C, Lesjean S, et al. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of Mycobacterium tuberculosis. J Clin Microbiol, 2006; 44, 4498-510.
13. Maiden MC. Multilocus sequence typing of bacteria. Annu Rev Microbiol, 2006; 60, 561-88.
14. Maiden MC, Bygraves JA, Feil E, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci USA, 1998; 95, 3140-5.
15. Spratt BG. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. Curr Opin Microbiol, 1999; 2, 312-6.
16. Enright MCand Spratt BG. A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. Microbiology, 1998; 144 (Pt 11), 3049-60.
17. Filliol I, Motiwala AS, Cavatore M, et al. Global phylogeny of Mycobacterium tuberculosis based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. J Bacteriol, 2006; 188, 759-72.
18. Hershberg R, Lipatov M, Small PM, et al. High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. PLoS Biol, 2008; 6, e311.

19. Abadia E, Zhang J, Dos Vultos T, et al. Resolving lineage assignation on Mycobacterium tuberculosis clinical isolates classified by spoligotyping with a new high-throughput 3R SNPs based method. Infect Genet Evol, 2010; 10, 1066-74.
20. Baker L, Brown T, Maiden MC, et al. Silent nucleotide polymorphisms and a phylogeny for Mycobacterium tuberculosis. Emerg Infect Dis, 2004; 10, 1568-77.
21. Hunter PRand Gaston MA. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol, 1988; 26, 2465-6.
22. Bougnoux ME, Morand Sand D'enfert C. Usefulness of multilocus sequence typing for characterization of clinical isolates of Candida albicans. J Clin Microbiol, 2002; 40, 1290-7.
23. Tavanti A, Gow NA, Senesi S, et al. Optimization and validation of multilocus sequence typing for Candida albicans. J Clin Microbiol, 2003; 41, 3765-76.
24. Robles JC, Koreen L, Park S, et al. Multilocus sequence typing is a reliable alternative method to DNA fingerprinting for discriminating among strains of Candida albicans. J Clin Microbiol, 2004; 42, 2480-8.
25. Sullivan CB, Diggle MAand Clarke SC. Multilocus sequence typing: Data analysis in clinical microbiology and public health. Mol Biotechnol, 2005; 29, 245-54.
26. Kamerbeek J, Schouls L, Kolk A, et al. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol, 1997; 35, 907-14.
27. Van Embden JD, Cave MD, Crawford JT, et al. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol, 1993; 31, 406-9.
28. Van Soolingen D, Haas PEWDand Kremer K. Restriction fragment length polymorphism typing of mycobacteria. in Mycobacterium tuberculosis protocols (eds Parisch, Tand Stoker, NG) 165-203 (Humana Press, 2001).
29. Aanensen DMand Spratt BG. The multilocus sequence typing network: mlst.net. Nucleic Acids Res, 2005; 33, W728-33.
30. Van Soolingen D, Qian L, De Haas PE et al. Predominance of a single genotype of Mycobacterium tuberculosis in countries of east Asia. J Clin Microbiol, 1995; 33, 3234-8.
31. Thorne N, Borrell S, Evans J, et al. IS6110-based global phylogeny of Mycobacterium tuberculosis. Infect Genet Evol, 2011; 11, 132-8.
32. Sreevatsan S, Pan X, Stockbauer KE, et al. Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci USA, 1997; 94, 9869-74.
33. Janagama HK, Senthilkumar TM, Bannantine JP et al. Identification and functional characterization of the iron-dependent regulator (IdeR) of Mycobacterium avium subsp. paratuberculosis. Microbiology, 2009; 155, 3683-90.
34. Zahrt TC, Wozniak C, Jones D, et al. Functional analysis of the Mycobacterium tuberculosis MprAB two-component signal transduction system. Infect Immun, 2003; 71, 6962-70.
35. He Hand Zahrt TC. Identification and characterization of a regulatory sequence recognized by Mycobacterium tuberculosis persistence regulator MprA. J Bacteriol, 2005; 187, 202-12.
36. Kantardjieff KA, Kim CY, Naranjo C, et al. Mycobacterium tuberculosis RmlC epimerase (Rv3465): a promising drug-target structure in the rhamnose pathway. Acta Crystallogr D Biol Crystallogr, 2004; 60, 895-902.
37. Aguilar D, Hanekom M, Mata D, et al. Mycobacterium tuberculosis strains with the Beijing genotype demonstrate variability in virulence associated with transmission. Tuberculosis (Edinb), 2010; 90, 319-25.
38. Lipsitch Mand Moxon ER. Virulence and transmissibility of pathogens: what is the relationship? Trends Microbiol, 1997; 5, 31-7.
39. Comas I, Homolka S, Niemann S, et al. Genotyping of genetically monomorphic bacteria: DNA sequencing in Mycobacterium tuberculosis highlights the limitations of current methodologies. PLoS One, 2009; 4, e7815.
40. Dong H, Liu Z, Lv B, et al. Spoligotypes of Mycobacterium tuberculosis from different Provinces of China. J Clin Microbiol, 2010; 48, 4102-6.
41. Maiden MC and Stuart JM. Carriage of serogroup C meningococci 1 year after meningococcal C conjugate polysaccharide vaccination. Lancet, 2002; 359, 1829-31.