Int J Clin Exp Pathol 2015;8(6):6891-6900
www.ijcep.com /ISSN:1936-2625/IJCEP0008867

*Original Article*

# Conserved hypothetical protein Rv1977 in *Mycobacterium tuberculosis* strains contains sequence polymorphisms and might be involved in ongoing immune evasion

Yi Jiang[1,2*], Haican Liu[1,2*], Xuezhi Wang[1,2*], Guilian Li[1,2*], Yan Qiu[1,2*], Xiangfeng Dou[3], Kanglin Wan[1,2]

[1]*State Key Laboratory for Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, P. R. China;* [2]*Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Hangzhou 310003, China;* [3]*Beijing Center for Diseases Prevention and Control, Beijing 100013, China.* [*]*Equal contributors.*

Received April 8, 2015; Accepted May 26, 2015; Epub June 1, 2015; Published June 15, 2015

**Abstract:** Host immune pressure and associated parasite immune evasion are key features of host-pathogen co-evolution. A previous study showed that human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved and thus it was deduced that *M. tuberculosis* lacks antigenic variation and immune evasion. Here, we selected 151 clinical *Mycobacterium tuberculosis* isolates from China, amplified gene encoding Rv1977 and compared the sequences. The results showed that Rv1977, a conserved hypothetical protein, is not conserved in *M. tuberculosis* strains and there are polymorphisms existed in the protein. Some mutations, especially one frameshift mutation, occurred in the antigen Rv1977, which is uncommon in *M.tb* strains and may lead to the protein function altering. Mutations and deletion in the gene all affect one of three T cell epitopes and the changed T cell epitope contained more than one variable position, which may suggest ongoing immune evasion.

**Keywords:** Rv1977, *Mycobacterium tuberculosis*, immune evasion

## Introduction

Host-pathogen coevolution is characterized by reciprocal adaptive changes in interacting species [1]. Host immune pressure and associated parasite immune evasion are key features of this process, often referred to as an 'evolutionary arms race' [2, 3] *M. tuberculosis* is the most successful pathogen with multiple mechanisms to subvert host immune response, resulting in insidious disease. Studies in human pathogenic viruses, bacteria and protozoa have revealed that genes encoding antigens tend to be highly variable as a consequence of diversifying selection to evade host immunity [4-7]. In 2010, Inaki Comas et al. reported that human T cell epitopes of M. tuberculosis are evolutionarily hyperconserved by sequencing the genomes of 21 strains representative of the global diversity and six major lineages of the *M. tuberculosis* complex (MTBC) [8]. In their study, only five epitopes, contained in esxH, pstS1, and Rv1986, harbored more than one variable position. Among 78 proteins included in the study, there were eight proteins belong to conserved hypotheticals according to Tuberculist (http://tuberculist.epfl.ch/), i.e. Rv1158c, Rv1461, Rv1977, Rv2182c, Rv3207c, Rv3333c, Rv3378c and Rv3714c. We compared the sequences of these two 8 antigens among 25 whole-genome-sequenced *M. tuberculosis* complex (MTBC) strains whose sequences were downloaded from NCBI website. We found that Rv1977 contained a number of polymorphisms and one strain even had frameshift, which are uncommon in *M. tuberculosis* strains. Also, some changes were located on T cell epitopes of the antigen. All these give us a hint that Rv1977 in *M. tuberculosis* strains might have antigenic variation and might be involved in ongoing immune evasion.

Rv1977 is a hypothetical protein in *M. tuberculosis* genome and its function is unknown and

**Table 1.** 25 MTBC strains whose data were obtained from NCBI website

| ID No. | Strain name |
|---|---|
| 01 | *Mycobacterium canettii CIPT 140070017* |
| 02 | *Mycobacterium canettii CIPT 140070010* |
| 03 | *Mycobacterium tuberculosis F11* |
| 04 | *Mycobacterium canettii CIPT 140010059* |
| 05 | *Mycobacterium tuberculosis CCDC5180* |
| 06 | *Mycobacterium canettii CIPT 140070008* |
| 07 | *Mycobacterium tuberculosis CCDC5079* |
| 08 | *Mycobacterium tuberculosis H37Ra* |
| 09 | *Mycobacterium tuberculosis KZN 4207* |
| 10 | *Mycobacterium tuberculosis RGTB423* |
| 11 | *Mycobacterium tuberculosis H37Rv* uid170532 |
| 12 | *Mycobacterium tuberculosis H37Rv* uid57777 |
| 13 | *Mycobacterium tuberculosis str. Erdman* |
| 14 | *Mycobacterium tuberculosis str. Beijing/NITR203* |
| 15 | *Mycobacterium tuberculosis CDC1551* |
| 16 | *Mycobacterium tuberculosis str. Haarlem* |
| 17 | *Mycobacterium tuberculosis CAS/NITR204* |
| 18 | *Mycobacterium canettii CIPT 140060008* |
| 19 | *Mycobacterium tuberculosis 7199-99* |
| 20 | *Mycobacterium tuberculosis KZN 1435* |
| 21 | *Mycobacterium tuberculosis KZN 605* |
| 22 | *Mycobacterium tuberculosis UT205* |
| 23 | *Mycobacterium tuberculosis CCDC5079* |
| 24 | *Mycobacterium tuberculosis CTRI-2* |
| 25 | *Mycobacterium tuberculosis RGTB327* |

rarely studied, while it harbors three T cell epitopes [9], which suggest it may play some roles in reaction between *M. tuberculosis* and human T cells. Here, we selected 151 clinical *M. tuberculosis* isolates in China; amplified genes of the antigen Rv1977 and compared the sequences. The results showed that some mutations, especially one frameshift mutation, occurred in the antigen Rv1977. In addition, mutations and deletion in the gene all affect one of three T cell epitopes and the changed T cell epitope contained more than one variable position, which may suggest ongoing immune evasion.

## Materials and methods

### *Ethics statement*

The study obtained approval from the Ethics Committee of National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention. The patients with TB included in the present research protocol were given a subject information sheet and they all gave written informed consent to participate in the study.

### *Strains and DNA preparation*

Firstly, we compared the sequences of 8 conserved hypothetical antigens (i.e. Rv1158c, Rv1461, Rv1977, Rv2182c, Rv3207c, Rv3333c, Rv3378c and Rv3714c) among 25 whole-genome-sequenced *M. tuberculosis* complex (MTBC) strains whose sequences were downloaded from NCBI website. We found that Rv1977, one of these 8 antigens, contained a number of polymorphisms, and one strain even had frameshift, (**Table 1**; **Figure 1**) which are uncommon in *M. tuberculosis* strains. Also, some changes were located on T cell epitopes of the antigen. All these give us a hint that Rv1977 in *M. tuberculosis* strains might have antigenic variation and might be involved in ongoing immune evasion.

Then we chose 151 clinical *M. tuberculosis* to clarify this hypothesis. 151 strains were selected from 2346 MTBC strains isolated in Beijing Municipality and 12 provinces and autonomous regions, China, which were genotyped by Spoligotyping in a previous study [10]. Strains belonging to all major and rare spoligotypes in China were included. Considering the predominance of the Beijing family strains in China, we chose about half of the Beijing family strains (77 strains) and half non-Beijing family strains (74 strains). We randomly selected the 77 Beijing family strains from 1738 Beijing strains among 2346 strains. The remaining 81 strains were selected from 608 non-Beijing family isolates. Furthermore, we attempted to include strains representing different spoligotypes that were isolated from different places. **Table 2** shows the numbers of strains used in this study that were obtained from different provinces in China. The spoligotype patterns of 151 stains were showed in **Table 3**.

The strains were cultured using a standard Löwenstein-Jensen medium method, heat inactivated and then used directly in polymerase chain reactions (PCRs).

### *Primers*

The nucleotide sequences of the primers (from the 5' to 3' end) used in this study were

```
Strains      10        20        30        40        50        60        70        80        90        100       110       120
    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
01  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
02  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ LLKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
03  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
04  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
05  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
06  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYMMQSPIA DAFTIGMGKP FTVITSGLYD
07  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
08  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
09  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
10  MSQTPATTRK TFPEISSSSV GAPRRPDRPF RAAPAQRLRP DLEADVGDVA GTAAPAAVPG QRGTGRAAQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
11  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
12  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
13  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
14  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
15  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
16  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
17  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
18  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
19  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
20  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
21  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
22  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
23  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
24  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD
25  MSQTPATTRK TFPEISSRAW EHPADRTALS ALRRLKGFDQ ILKLMSGMLR ERQHRLLYLA SAARVGPRQF ADLDALLDEC VDVLDASAKP ELYVMQSPIA DAFTIGMGKP FTVITSGLYD

          130       140       150       160       170       180       190       200       210       220       230       240
    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
01  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
02  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
03  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
04  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
05  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
06  LVTHDEMRFV MGHELGHALS GHAVYRTMTM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
07  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
08  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
09  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
10  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
11  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
12  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
13  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
14  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
15  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
16  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
17  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
18  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
19  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
20  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
21  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
22  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
23  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
24  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
25  LVTHDEMRFV MGHELGHALS GHAVYRTMMM HLLRLARSFG VLPVGGWALR AIVAALLEWQ RKSELSGDRA GLLCAQDLDT ALRVEMKLAG GCRLDKLDSE AFLAQAREYE TSGDMRDGVL
```

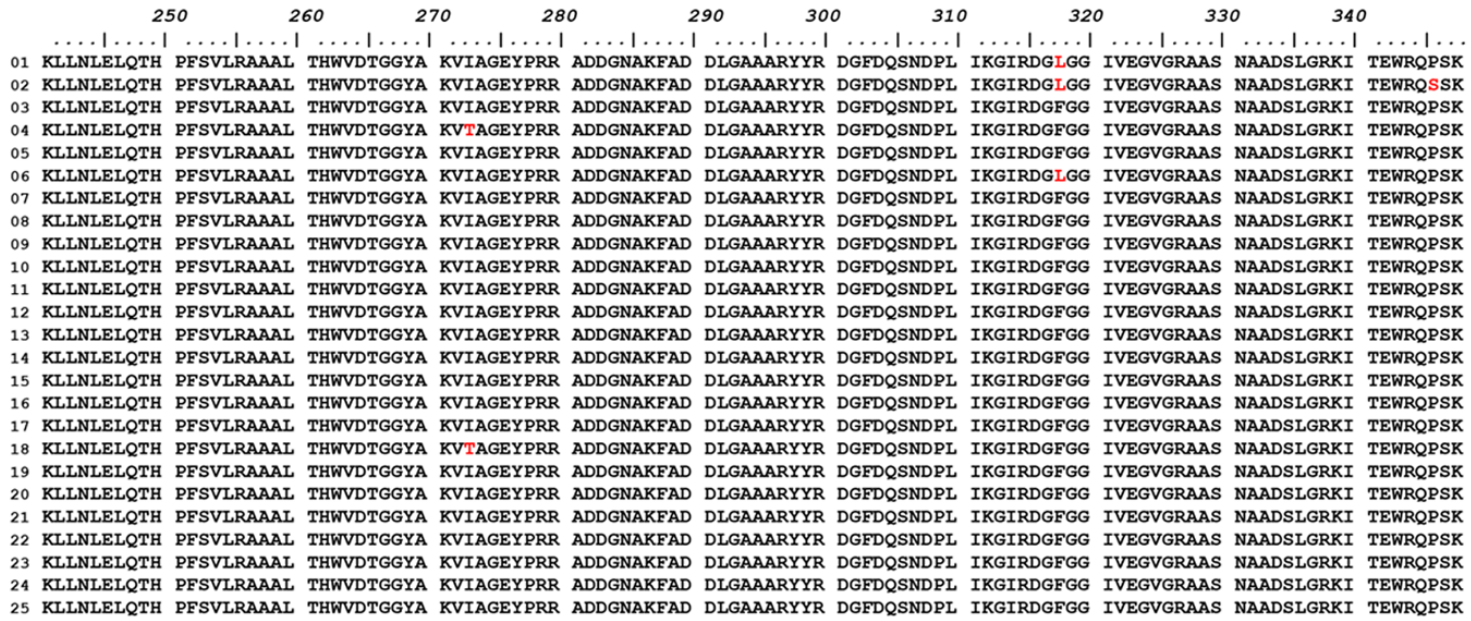


**Figure 1.** AA sequence alignment for antigen Rv1977 of 25 whole-genome-sequenced MTBC strains. AA changes were marked in red. Shading indicates T cell epitope areas. Strain 10 had a frameshift.

**Table 2.** No. of strains in different provinces of China

| Places | No. of isolates |
|---|---|
| Anhui Province | 10 |
| Shannxi Province | 15 |
| Beijing Municipality | 9 |
| Fujian Province | 28 |
| Gansu Province | 10 |
| Guangxi Zhuang Autonomous Region | 22 |
| Sichuan Province | 1 |
| Henan Province | 12 |
| Hunan Province | 5 |
| Xizang (Tibet) Autonomous Region | 11 |
| Xinjiang Uygur Autonomous Region | 9 |
| Jilin Province | 10 |
| Zhejiang Province | 9 |

**Table 3.** No. of strains of each Spoligotype pattern

| Spoligotypes | No. of strains |
|---|---|
| Beijing | 77 |
| T | 12 |
| U | 24 |
| MANU | 6 |
| Haarlem | 5 |
| EAI | 1 |
| LAM | 2 |
| S | 1 |
| CAS | 4 |
| H37Rv family | 1 |
| new | 18 |

designed with DNASTAR software according to H37Rv genome sequence and were as follows: 5'-CGCCGTATTCTGAAGACC-3' and 5'-GTGTTC-GCAATGCTATGAG-3'.

*PCR*

The PCR were performed in a total volume of 20 μl. The PCR mix contained 10 μl PCR buffer, 100 nM each primer, 200 μM each of the four dNTPs and 0.5 U DNA Taq Polymerase (Takara). An initial denaturation of 5min at 94°C was followed by 35 cycles of denaturation at 94°C for 45 s, annealing at 55°C for 45 s and extension at 72°C for 1 min, followed by a final extension at 72°C for 10 min. Negative controls (reagents only, no DNA) were included each time when the PCR was performed. The positive control was 500 pg DNA from *M. tuberculosis* H37Rv. The presence and size of each PCR product were determined by electrophoresis on 2% agarose gel in Tris-boric acid-EDTA buffer followed by staining with ethidium bromide. We performed all of the PCRs at least twice to validate the reproducibility. The variants were confirmed by sequencing of the new PCR products.

*Sequence*

The sequences of the PCR products were determined by ABI 3730xl DNA Analyzer.

*Data analysis*

The sequences were first aligned by ClustalW [11] software with the *Rv1977* gene sequence from *M. tuberculosis* H37Rv genome to determine the Rv1977 region, and then this region was split out by a personalized Perl script. The sequence compare and translate were carried out by Bioedit software. The mutated protein structures were predicted by Phyre2 software online (http://www.sbg.bio.ic.ac.uk/phyre2). Values of dN were calculated by MEGA5.

## Results

*Mutations and deletion in gene sequences*

All 151 clinical strains we chose presented relative PCR products of antigen Rv1977. Among the 151 *M. tuberculosis* strains, five isolates presented polymorphism in the gene sequence of this protein (**Figure 2**; **Table 4**). There were three nonsynonymous mutations and one single base deletion. GX06187, HuN06099 and GS05113 all had a different nonsynonymous mutation. Two strains, FJ05406 and FJ06051, presented same one single base deletion at position 204.

*Changes in protein level*

**Table 5** showed the AA change and position in antigen Rv1977 among 151 strains. All changes resulted in AA change. Three nonsynonymous mutations led to AA change of the protein. The deletion in FJ05406 and FJ06051 resulted in premature termination, located at position 68 of the AA (**Figure 2**; **Table 4**).

After we predicted protein structures by Phyre2 software online, we found that AA64-251 constitute the domain (**Figure 3**). The deletion in

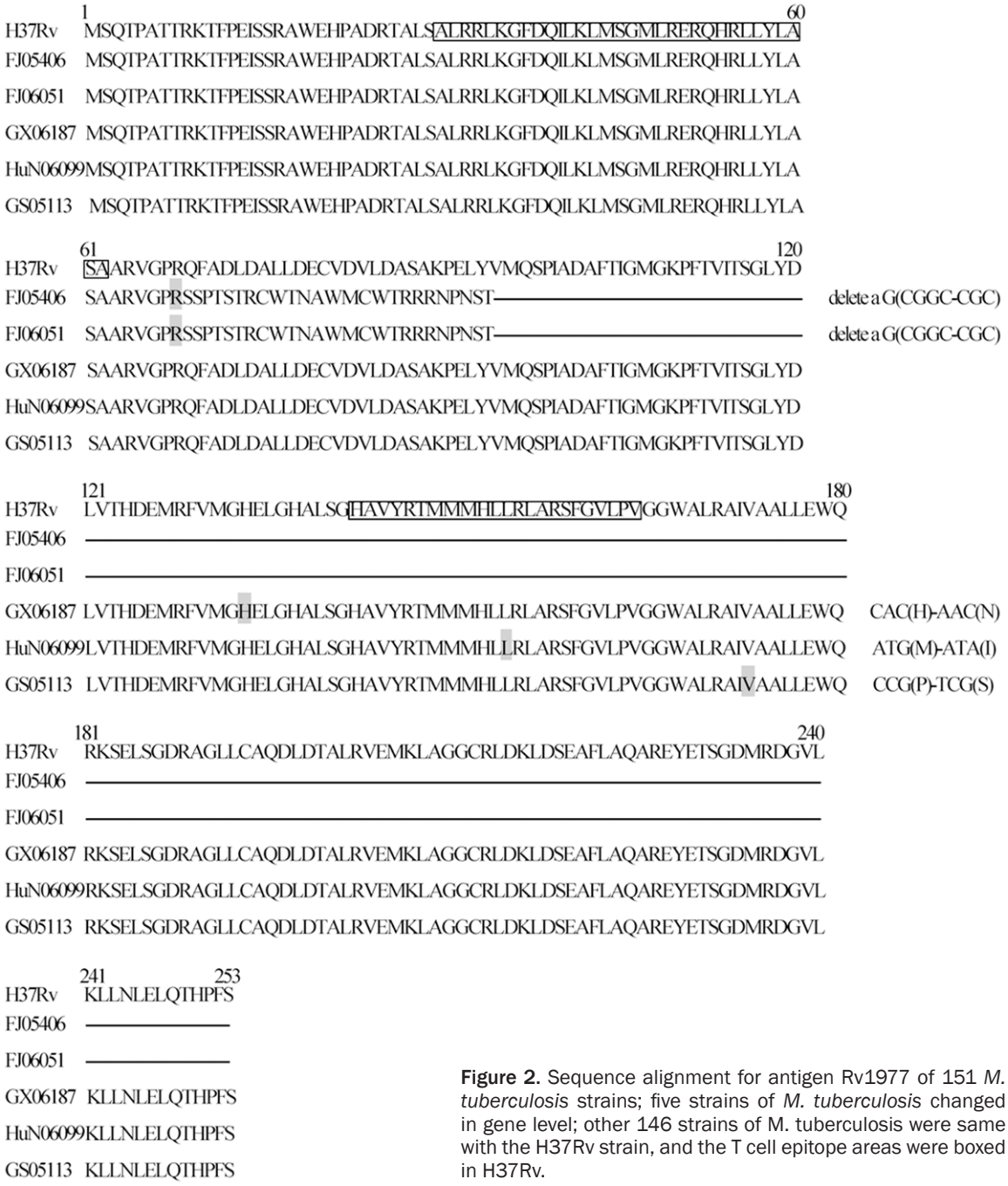


**Figure 2.** Sequence alignment for antigen Rv1977 of 151 *M. tuberculosis* strains; five strains of *M. tuberculosis* changed in gene level; other 146 strains of M. tuberculosis were same with the H37Rv strain, and the T cell epitope areas were boxed in H37Rv.

**Table 4.** Changes in antigen Rv1977 among 151 clinical strains*

| Isolates | Base change | AA change | Spoligotypes |
|---|---|---|---|
| FJ05406 | Delete a G at position 204 | Frameshift | EAI |
| FJ06051 | | | New |
| GX06187 | C397A | H133N | New |
| HuN06099 | G444A | M148I | U |
| GS05113 | C487T | P163S | Beijing |

*: Use the CDS of Rv1977 of *M. tuberculosis* H37Rv strain as the reference sequence.

FJ05406 and FJ06051 led to dramatic change in the protein structure. Three nonsynonymous mutations were all located on the domain region. Two of three mutations, i.e. H133N in GX06187 and M148I in HuN06099 were changed between same nature AA, while P163S in GS05113 were changed between amino acids with different properties.

**Table 5.** AA changes of the T cell epitopes included in antigen Rv1977*,#

| IEDB_ID | Peptide sequence | AA changes |
|---|---|---|
| 21306 | GMLRERQHRLLYLASA | No change |
| 23584 | HAVYRTMMMHLLRLARSFGVLPV | ATG (M)-ATA (I); CCG(P)-TCG (S); Frameshift |
| 2867 | ALRRLKGFDQILKLMSGMLR | No change |

*: Use the CDS of Rv1977 of *M. tuberculosis* H37Rv strain as the reference sequence; #: Underlined AA indicates locations of amino acid changes.

*Changes in T cell epitopes*

There are three human T cell epitopes in the antigen Rv1977 according to the Immune Epitopes Database (IEDB) [9]. Three mutations and deletion all affect one of three T cell epitopes in antigen Rv1977, i.e. IEDB ID numbers 23584. The changed epitope contained more than one variable position (**Table 5**).

**Table 6** showed the distribution nonsynonymous SNPs in Rv1977. As there are no synonymous SNPs in the gene, we only calculated dN values of epitope regions, non-epitope regions and the whole gene. The dN value of epitope region was higher than non-epitope region, which means the former has accumulated significantly more amino acid changes than the latter. In addition, deletion are strong indicators of immune selection, which affected both epitope region and non-epitope region.

## Discussion

In this study, we chose 151 clinical *M. tuberculosis* strains in China which originated from a very large geographical area and had different spoligotyping patterns; the data provided by them could therefore be representative of genetic diversity that might be present within China, at least to some extent.

Rv1977 is a conserved hypothetical protein in *M. tuberculosis* genome and its function is unknown and rarely studied, while it harbors three T cell epitopes [9], which suggest it may play some roles in reaction between *M. tuberculosis* and human T cells. Among 151 *M. tuberculosis* strains, five strains presented polymorphism in gene level and showed AA change of the protein. Especially, one frameshift occurred in it, which led to premature termination of the protein code. This means Rv1977, a conserved hypothetical protein, is not conserved in *M. tuberculosis* strains and there are polymorphisms existed in the protein.

It is reported that T cell epitopes of *M. tuberculosis* are evolutionarily hyperconserved and thus it was deduced that *M. tuberculosis* lacks of antigenic variation and immune evasion [8]. However, our previous studies showed that PstS1 [12], MPT64 [13], Rv3878 [14], Rv3878 [15], Rv2945c and Rv0309 [16] also harbored higher polymorphisms of T cell epitopes, which suggesting their role in diversifying selection to evade host immunity. In this study, there were polymorphisms existed in Rv1977 which might show variable as a consequence of diversifying selection to evade host immunity. One of three T cell epitopes in antigen Rv1977 contained more than one variable position and presented AA changes. Two of three nonsynonymous mutations and one single base deletion caused alterations on the corresponding T cell epitopes. Initial comparative sequencing of *M. tuberculosis* revealed very low sequence diversity compared with other bacteria. Insertion and deletion is uncommon in genes of the pathogen and usually a strong indicator of selection, as they intrigue alteration of amino acid code which could affect the structure and even function of the protein. This gives us a hint that protein Rv1977 may be a specific antigen that undergo antigenic variation in response to host immune pressure and that it may be involved in diversifying selection to evade host immunity. The dN value of epitope region was higher than non-epitope region, which means the former has accumulated significantly more amino acid changes than the latter. In addition, deletion are strong indicators of immune selection, which affected both epitope region and non-epitope region. However, further investigations are needed to determine whether the observed changes are due to immune pressure, other selection pressure(s) and mere random genetic drift.

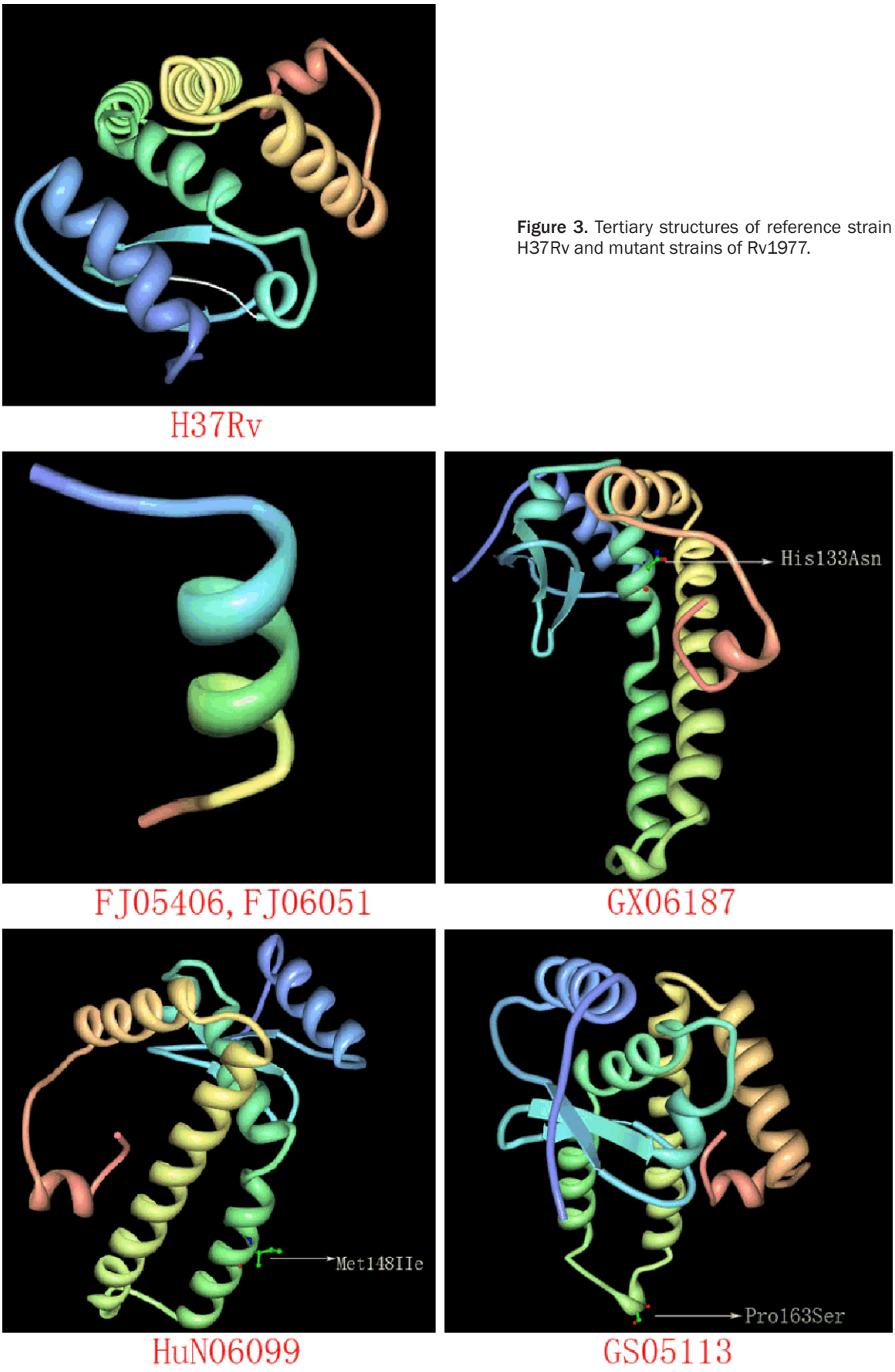


**Figure 3.** Tertiary structures of reference strain H37Rv and mutant strains of Rv1977.

**Table 6.** Distribution of synonymous and nonsynonymous SNPs in *Rv1977**

| | Length(bp) | Nonsyn SNPs | dN |
|---|---|---|---|
| Epitope region | 165 | 2 | 0.00022 |
| Non-epitope region | 594 | 1 | 0.00003 |
| All | 759 | 3 | 0.00007 |

*: H37Rv was used as reference to base the change in allele for the SNPs

Although the function of Rv1977 is currently unclear, we propose that the two strains with G deletion interact differently with human T cells and may represent a special type of MTBC strain that merits further investigation. Moreover, strains from other countries and areas should be collected to prove whether sequence variants in Rv1977 are ubiquitous.

In conclusion, there are polymorphisms existed in antigen Rv1977 and one of three T cell epitopes changed, which may reflect that the antigen are involved in diversifying selection to evade host immunity.

## Acknowledgements

This work was financially supported by the projects 81401647 of Natural Science Foundation of China, 2013ZX10003006, 2013ZX-10003002-001 and 2013ZX10004-101 of Chinese National Key Program of Mega Infectious Diseases.

## Disclosure of conflict of interest

None.

**Address correspondence to:** Dr. Kanglin Wan, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, P.O., Box 5, Changping, Beijing 102206, People's Republic of China. Tel: 0086 10 58900776; Fax: 0086 10 58900574; E-mail: wankanglin@icdc.cn

## References

[1] Woolhouse ME, Webster JP, Domingo E, Charlesworth B, Levin BR. Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nat Genet 2002; 32: 569-77.

[2] Dawkins R, Krebs JR. Arms races between and within species. Proc R Soc Lond B Biol Sci 1979; 205: 489-511.

[3] Brunham RC, Plummer FA, Stephens RS. Bacterial antigenic variation, host immune response, and pathogen-host coevolution. Infect Immun 1993; 61: 2273-2276.

[4] Farci P, Shimoda A, Coiana A, Diaz G, Peddis G, Melpolder JC, Strazzera A, Chien DY, Munoz SJ, Balestrieri A, Purcell RH, Alter HJ. The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. Science 2000; 288: 339-344.

[5] Urwin R, Russell JE, Thompson EA, Holmes EC, Feavers IM, Maiden MC. Distribution of surface protein variants among hyperinvasive meningococci: implications for vaccine design. Infect Immun 2004; 72: 5955-5962.

[6] Jeffares DC, Pain A, Berry A, Cox AV, Stalker J, Ingle CE, Thomas A, Quail MA, Siebenthall K, Uhlemann AC, Kyes S, Krishna S, Newbold C, Dermitzakis ET, Berriman M. Genome variation and evolution of the malaria parasite Plasmodium falciparum. Nat Genet 2007; 39: 120-125.

[7] Kawashima Y, Pfafferott K, Frater J, Matthews P, Payne R, Addo M, Gatanaga H, Fujiwara M, Hachiya A, Koizumi H, Kuse N, Oka S, Duda A, Prendergast A, Crawford H, Leslie A, Brumme Z, Brumme C, Allen T, Brander C, Kaslow R, Tang J, Hunter E, Allen S, Mulenga J, Branch S, Roach T, John M, Mallal S, Ogwu A, Shapiro R, Prado JG, Fidler S, Weber J, Pybus OG, Klenerman P, Ndung'u T, Phillips R, Heckerman D, Harrigan PR, Walker BD, Takiguchi M, Goulder P. Adaptation of HIV-1 to human leukocyte antigen class I. Nature 2009; 458: 641-645.

[8] Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, Ernst JD, Gagneux S. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. Nat Genet 2010; 42: 498-503.

[9] Ernst JD, Lewinsohn DM, Behar S, Blythe M, Schlesinger LS, Kornfeld H, Sette A. Meeting Report: NIH Workshop on the Tuberculosis Immune Epitope Database. Tuberculosis (Edinb) 2008; 88: 366-370.

[10] Dong H, Liu Z, Lv B, Zhang Y, Liu J, Zhao X, Liu J, Wan K. Spoligotypes of *Mycobacterium tuberculosis* from different Provinces of China. J Clin Microbiol 2010;48: 4102-4106.

[11] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG. Clustal W and Clustal X version 2.0. Bioinformatics 2007; 23:2947-2948.

[12] Liu H, Jiang Y, Dou X, Wang H, Zhao X, Zhang W, Wan L, Zhang Z, Chen C, Wan K. pstS1 polymorphisms of Mycobacterium tuberculosis strains may reflect ongoing immune evasion. Tuberculosis (Edinb) 2013; 93: 475-481.

[13] Jiang Y, Liu H, Wang H, Dou X, Zhao X, Bai Y, Wan L, Li G, Zhang W, Chen C, Wan K. Polymorphism of antigen MPT64 in Mycobacterium tuberculosis strains. J Clin Microbiol 2013; 51: 1558-1562.

[14] Jiang Y, Wan L, Zhang Z, Liu H, Pang H, Zhang W, Zhao X, Wang H, Li G, Chen C, Kan B, Wan K. Conserved alanine rich protein Rv3878 in Mycobacterium tuberculosis contains sequence polymorphisms. Tuberculosis (Edinb) 2014; 94: 245-251.

[15] Jiang Y, Liu H, Qiu Y, Li G, Dou X, Wan K. Polymorphisms of FtsK/SpoIIIE protein in Mycobacterium tuberculosis complex strains may affect both protein function and host immune reaction. Int J Clin Exp Med 2014; 7: 5385-5393.

[16] Jiang Y, Dou X, Zhang W, Liu H, Zhao X, Wang H, Lian L, Yu Q, Zhang J, Li G, Chen C, Wan K. Genetic diversity of antigens Rv2945c and Rv0309 in Mycobacterium tuberculosis strains may reflect ongoing immune evasion. FEMS Microbiol Lett 2013; 347: 77-82.