Tuberculosis xxx (2016) 1–7

Contents lists available at ScienceDirect

Tuberculosis

journal homepage: http://intl.elsevierhealth.com/journals/tube

IMMUNOLOGICAL ASPECTS

Genetic diversity of immune-related antigens in Region of Difference 2 of *Mycobacterium tuberculosis* strains

Yi Jiang [a, 1], Haican Liu [a, 1], Xuezhi Wang [a, 1], Shiqi Xiao [b], Machao Li [a], Guilian Li [a], Lili Zhao [a], Xiuqin Zhao [a], Xiangfeng Dou [c], Kanglin Wan [a, *]

[a] *State Key Laboratory for Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Chinese Center for Disease Control and Prevention, Beijing, China*
[b] *State Key Laboratory of Agrobiotechnology, Department of Microbiology and Immunology, College of Biological Sciences, China Agricultural University, Beijing 100193, China*
[c] *Beijing Center for Diseases Prevention and Control, Beijing 100013, China*

ARTICLE INFO

Article history:
Received 26 August 2015
Received in revised form
23 April 2016
Accepted 1 May 2016

SUMMARY

Region of Difference 2 (RD2) was lost during the ongoing propagation of BCG between 1927 and 1931, a time that coincides with reports of the ongoing attenuation of the vaccine. Some data demonstrate that RD2 plays a role in mycobacterial virulence, and that its deletion from *Mycobacterium tuberculosis* leads to a decrease in bacterial growth in both a macrophage and a murine model. Human T-cell epitopes of *M. tuberculosis* are evolutionarily hyperconserved and thus it was deduced that *M. tuberculosis* lacks antigenic variation and immune evasion. However, two antigens, Rv1986 and MPT64, encoded by RD2 harbored more than one amino acid changes. In this study, we used same set of clinical *M. tuberculosis* complex (MTBC) isolates from China, amplified the five genes containing T and B cell epitopes other than MPT64 encoded by RD2, and compared the sequences. It turned out that proteins in RD2 region, especially Rv1980c, Rv1985 and Rv1986 may be a special region that undergo antigenic variation in response to host immune pressure and may be involved in ongoing immune evasion. The dN/dS value of all six genes (including MPT64) were 2.33, much higher than 1, which means T cell antigens in RD2 region appeared to be under diversifying selection. Our data support the view that RD2 regions tend to be more variable than we expected to evade host immunity and the immune-related antigens in RD2 were more variable than we expected, especially in T-cell epitope regions.

1. Introduction

Tuberculosis (TB) is one of the most important problems of public health in the world as it is still a serious infectious disease of worldwide prevalence. About one third of the world population have been infected with *Mycobacterium tuberculosis*, 9.3 million people develop active TB, and 1.8 million people die of TB each year [1]. The current efforts to reduce the global problem have been focused on improved diagnosis and effective vaccines. Biochemical, immunological, and molecular biological characterization of *M. tuberculosis* has led to the identification of several antigens which may be useful in the development of improved diagnostic methods and/or vaccines [2].

In 2010, Inaki Comas et al. reported that human T-cell epitopes of *M. tuberculosis* are evolutionarily hyperconserved and thus deduced that *M. tuberculosis* is lack of antigenic variation and immune evasion [3]. In their study, Rv1986 is one of few antigens that harbored more than one amino acid change. In our previous study [4], we found that polymorphisms of the *mpt64* gene in the MTBC may be the reason for changes in the antigen produced, which may in turn cause altering of related functions, thereby allowing immune evasion. Rv1986 and MPT64 both are located on RD2 region, which encodes 11 ORFs from Rv1978 to Rv1988. In the Immune Epitope Database (IEDB) established by Ernst, J.D. et al. [5], there were seven antigens (Rv1979c, Rv1980c, Rv1983, Rv1984c, Rv1985c, Rv1986 and Rv1987) encoded by RD2 that harbored T- and/or B-cell epitopes. As the epitopes in IEDB were tested by at least one immune assay in lab, the seven proteins are deemed to be

* Corresponding author. National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, P.O. Box 5, Changping, Beijing 102206, China. Tel./fax: +86 10 58900779.
E-mail address: wankanglin@icdc.cn (K. Wan).
[1] Yi Jiang, Haican Liu and Xuezhi Wang contributed equally to this study.

http://dx.doi.org/10.1016/j.tube.2016.05.002
1472-9792/

related to immune reaction. Herein, we aimed to amplify the remaining genes, sans MPT64(Rv1980c) which was reported in our earlier study and Rv1983 due to technical reasons, by using the same set of strains as in the earlier study and compared the sequences of the six genes (five amplified here and including Rv1980c from the earlier study) to evaluate the impact of sequence change to immune recognition.

2. Materials and methods

2.1. 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.

2.2. Strains and DNA preparation

We used same set of clinical isolates in our previous study. Considering the predominance of the Beijing family strains in China, we chose about half of the Beijing family strains (92) and half non-Beijing family strains (88). We randomly selected the 92 Beijing family strains from 1738 Beijing strains among 2346 strains. The remaining 88 strains were selected from 608 non-Beijing family isolates. Further, we attempted to purposely include strains representing different spoligotypes that were isolated from different places [4].

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

2.3. Primers design

As Rv1983 is one of Pro-Glu proteins contains polymorphic GC-rich sequences (PE-PGRS family protein) belonging to Pro-Glu/Pro-Pro-Glu (PE/PPE) genes, we excluded it from this study for technical reason. The nucleotide sequences of the primers (from the 5′ to 3′ end) used in this study were designed with DNAstar software according to H37Rv genome sequence and showed in Table 1:

2.4. Polymerase chain reactions

The PCRs 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 5 min at 94 °C was followed by 35 cycles of denaturation at 94 °C for 45 s, annealing at 60 °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 PCR products.

Table 1
Primers used in this study for PCR amplification.

Locus tag or gene product	Length (bp)	Primer
Rv1979c	1794	5′-AGCGCACTACGTGCCTACAG-3′F
		5′-CTCGATGCTGGCCTAGACTC-3′R
Rv1984c	904	5′-CCTAAGGCTACCGTTCTGAC-3′F
		5′-GCTCTACGCGTTCAACACAA-3′R
Rv1985c	1359	5′-TGAGTTCACGCAATCGACAC-3′F
		5′-GCTCGCAGCTCACCTTCAAT-3′R
Rv1986	1691	5′-GTTTACCGCAATGGTGATCC-3′F
		5′-CATGTGTGCAAGACGGATTC-3′R
Rv1987	896	5′-CCGCATCAACTCCGAGATGAC-3′F
		5′-ACATACCGTTCGGCAGGAGG-3′R

2.5. Sequence and data analysis

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

The sequences were first aligned by ClustalW [6] software with the corresponding gene sequences from *M. tuberculosis* H37Rv genome to determine the gene region, and then this region was split out by a personalized Perl script. The sequences were compared and sliced by Bioedit software.

We submitted our sequence data to Genebank. The GenBank accession numbers of the sequence data is as followed: file Rv1979c.sqn:KX116040 – KX116211, file Rv1980c.sqn:KX116212 – KX116383, file Rv1984c.sqn:KX116384 – KX116555, file Rv1985c.sqn:KX116556 – KX116727, file Rv1986.sqn:KX116728 – KX116899, file Rv1987.sqn: KX116900 – KX117071.

3. Results

3.1. Mutation, insertion and deletion in gene sequences

We excluded strains for failure in the DNA extraction, poor quality of DNA template, multiple infections or double peaks in sequencing. Then we analyzed the results of 172 strains for six proteins.

All 172 strains presented relative PCR products of six proteins. The mutations, insertion and deletion in six proteins are presented in Table 2. There were six nonsynonymous mutations and two synonymous mutations in Rv1979c. Four strains, i.e. FJ05490, FJ07042, FJ05199 and FJ07031 contained the same synonymous mutation on 481R of Rv1979c. For Rv1980c, eight strains had a 63-bp deletion, whereas 4 strains showed a single-base mutation; all of these mutations were nonsynonymous [4]. Rv1984 had two nonsynonymous mutations among 172 strains. GX06043 and GX06130 both had a same mutation on 198Q and the mutation caused a stop codon. Single-base deletion/insertion(s) occurred in Rv1985c and Rv1986, resulting in frameshift. A synonymous change in Rv1985c and a nonsynonymous change in Rv1986, both displayed a higher level of incidence among the strains. There were two nonsynonymous mutations and one synonymous in Rv1987. AH03019 and AH03029 had a same nonsynonymous mutation of I26T in Rv1987.

3.2. Changes in protein level

The synonymous mutations of 359L and 481R in Rv1979c resulted in no AA changes. Other nonsynonymous mutations in Rv1979c caused AA change. The 63 bp deletion and 4 non-synonymous mutations in Rv1980c caused AA changes. Strain XJ06112 had a single-base insertion in Rv1980c, but this did not result in an amino acid change, as the C insertion occurred in the codon for the last amino acid (Ala) at the carboxyl terminus, in the second base of the "GCC." [4] For Rv1984, the nonsense mutation at 198Q results in a premature termination. A frameshift mutation at

Please cite this article in press as: Jiang Y, et al., Genetic diversity of immune-related antigens in Region of Difference 2 of *Mycobacterium tuberculosis* strains, Tuberculosis (2016), http://dx.doi.org/10.1016/j.tube.2016.05.002