Infection, Genetics and Evolution 55 (2017) 244–250

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Conservation in gene encoding *Mycobacterium tuberculosis* antigen Rv2660 and a high predicted population coverage of H56 multistage vaccine in South Africa

Angy P. Perez-Martinez[a,1], Edison Ong[b,1], Lixin Zhang[a,2], Carl F. Marrs[a], Yongqun He[c,d,e], Zhenhua Yang[a,*]

[a] Department of Epidemiology, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109-2029, United States
[b] Department of Computational Medicine and Bioinformatics, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109-2029, United States
[c] Unit for Laboratory Animal Medicine, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109-2029, United States
[d] Department of Microbiology and Immunology Department, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109-2029, United States
[e] Center of Computational Medicine and Bioinformatics, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109-2029, United States

ARTICLE INFO

Keywords:
Tuberculosis
Subunit vaccines
Rv2660c
H56
Epitope prediction
Population coverage

ABSTRACT

H56/AERAS-456 + IC31 (H56), composed of two early secretion proteins, Ag85B and ESAT-6, and a latency associated protein, Rv2660, and the IC31 Intercell adjuvant, is a new fusion subunit vaccine candidate designed to induce immunity against both new infection and reactivation of latent tuberculosis infection. Efficacy of subunit vaccines may be affected by the diversity of vaccine antigens among clinical strains and the extent of recognition by the diverse HLA molecules in the recipient population. Although a previous study showed the conservative nature of Ag85B- and ESAT-6-encoding genes, genetic diversity of *Rv2660c* that encodes RV2660 is largely unknown. The population coverage of H56 as a whole yet remains to be assessed. The present study was conducted to address these important knowledge gaps. DNA sequence analysis of *Rv2660c* found no variation among 83 of the 84 investigated clinical strains belonging to four genetic lineages. H56 was predicted to have as high as 99.6% population coverage in the South Africa population using the Immune Epitope Database (IEDB) Population Coverage Tool. Further comparison of H56 population coverage between South African Blacks and Caucasians based on the phenotypic frequencies of binding MHC Class I and Class II supertype alleles found that all of the nine MHC-I and six of eight MHC-II human leukocyte antigen (HLA) supertype alleles analyzed were significantly differentially expressed between the two subpopulations. This finding suggests the presence of race-specific functional binding motifs of MHC-I and MHC-II HLA alleles, which, in turn, highlights the importance of including diverse populations in vaccine clinical evaluation. In conclusion, H56 vaccine is predicted to have a promising population coverage in South Africa; this study demonstrates the utility of integrating comparative genomics and bioinformatics in bridging animal and clinical studies of novel TB vaccines.

1. Introduction

In 2015 tuberculosis (TB) caused by *Mycobacterium tuberculosis* infection claimed about 1.4 million lives (World Health Organization, 2016). TB continues to be a significant global health problem due to the lack of a broadly effective vaccine (Izzo, 2017), along with other challenges currently facing global TB control. The widely used live-attenuated *Mycobacterium bovis* Bacillus Calmette-Guerin vaccine (BCG) has proven to be effective against severe form of childhood TB. However, it provides little protection against pulmonary TB in adults, the most prevalent and contagious form of the disease (Brewer and Colditz, 1995; Fine, 1995), nor does it protect against the reactivation of latent TB infection (LTBI) (Fine, 1995) that is estimated to be present in one third of the world's population (World Health Organization, 2016). The

* Corresponding author at: Department of Epidemiology, School of Public Health, University of Michigan, 1415 Washington Heights, SPHII/5124, Ann Arbor, MI 48109-2029, United States.
E-mail addresses: apperez@umich.edu (A.P. Perez-Martinez), edong@umich.edu (E. Ong), lzhang@epi.msu.edu (L. Zhang), cfmarrs@umich.edu (C.F. Marrs), yongqunh@umich.edu (Y. He), zhenhua@umich.edu (Z. Yang).
[1] These authors contributed equally to this work as the first authors.
[2] Present Address: Department of Epidemiology and Biostatistics, and Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, 909 Fee Road, East Lansing, MI 48824, United States.

http://dx.doi.org/10.1016/j.meegid.2017.09.023
Received 13 March 2017; Received in revised form 11 September 2017; Accepted 19 September 2017
Available online 21 September 2017
1567-1348/

ideal new TB vaccine would, therefore, be expected to induce immunity against not only new infection but also reactivation from LTBI.

H56/AERAS-456 + IC31 (H56) developed at the Statens Serum Institute, Copenhagen, Denmark, is a new vaccine candidate designed with this goal. H56 is a fusion subunit vaccine composed of two early secretion proteins, Ag85B and ESAT-6, a latency associated protein, Rv2660, and the IC31 Intercell adjuvant (Aagaard et al., 2011).

H56 was shown to surpass BCG protection levels and elicit both CD4 + and CD8 + immune responses in latently infected animal models (Lin et al., 2012). Currently, H56 is undergoing phase I/IIa clinical trials in South Africa (ClinicalTrials.gov, 2017).

To ensure a vaccine's global efficacy, the antigens contained in the vaccine should ideally be conserved across different genetic lineages of *M. tuberculosis* clinical strains (Scarselli et al., 2005). Although the *M. tuberculosis* genome was initially thought to be highly conserved, many studies have elucidated its variability (Herbert et al., 2007; Kong et al., 2006; Talarico et al., 2005). DNA polymorphisms have been found in genes encoding novel TB vaccine candidates, such as PPE18 and PepA (Herbert et al., 2007) that constitute the fusion protein vaccine candidate Mtb72f (Skeiky et al., 2004). Such genetic variations among *M. tuberculosis* strains can result in amino acid sequence changes, impacting the recognition of potentially immunogenic epitopes.

As demonstrated by the recent failure of the MVA85A vaccine candidate (Ndiaye et al., 2015), there is no guarantee that the protective effect for any vaccine tested in animals is reproducible in humans, given the large genetic variability worldwide. An effective TB vaccine must also be capable of inducing and maintaining powerful CD4 + and CD8 + T-cell immunity in the greatest proportion of the targeted population of diverse human leukocyte antigen (HLA) alleles (Kaufmann, 2013). HLA genes code for the major histocompatibility complex (MHC) cell surface glycoproteins that present foreign epitopes to the appropriate T-cells. MHC-I molecules present epitopes to CD8 + T-cells, while MHC-II present to CD4 + T-cells for subsequent activation. MHC molecules are extremely diverse; polymorphisms are concentrated in the peptide-binding groove, resulting in a wide array of epitope binding specificities across individuals (Shiina et al., 2009; Spurgin and Richardson, 2009). Furthermore, the frequency of HLA allele expression varies across different ethnicities which may limit a vaccine's population coverage (Shiina et al., 2009; Spurgin and Richardson, 2009). The success of a subunit vaccine can be compromised by pathogen and host genetic variation that negatively impact the antigen recognition and presentation necessary for establishing protective immunity.

A previous study demonstrated the conservative nature of Ag85B and ESAT-6 encoding genes (Davila et al., 2010), but information on the genetic diversity of *Rv2660c* that encodes Rv2660 is limited. A study that sequenced 22 *M. tuberculosis* strains found two strains with a single non-synonymous SNP (Reddy et al., 2009). Although immunogenicity of Rv2660 has been demonstrated in animal models (Lin et al., 2012) and humans (Govender et al., 2010), neither epitope prediction using bioinformatics tools nor epitope mapping based on *ex vivo* assays has been conducted so far. Furthermore, H56 as a whole has not been evaluated for its population coverage. To address these important knowledge gaps, we investigated DNA sequence variation of *M. tuberculosis Rv2660c* among 84 clinical strains representing four major genetic lineages. In addition, we estimated the population coverage of H56 based on epitope predictions in the HLA genetic background of South African population, which has been one of the primary populations for new TB vaccine clinical evaluation.

2. Materials and methods

2.1. M. tuberculosis strains

The study sample included 84 clinical isolates representing 84 different strains defined using a combination of IS*6110* DNA fingerprinting and spoligotyping, well-established and validated *M. tuberculosis* strain definition systems for TB molecular epidemiology studies (Barnes and Cave, 2003). The 84 strains contained four different genetic lineages of *M. tuberculosis*: Indo-Oceanic (10, 11.9%), East-Asian (16, 19.0%), East-African Indian (3, 3.6%), Euro-American (46, 54.8%). Genetic lineages were not defined for nine (10.7%) of the 84 strains. Lineage classification was based on spoligotype family, as described previously (Kato-Maeda et al., 2011). Spoligotype family was assigned according to spoligotype octal code using the SITVIT2 proprietary database of Institute Pasteur de la Guadeloupe (Demay et al., 2012).

2.2. Assessment of genetic variation in Rv2660c

The *Rv2660c* gene was amplified using Platinum TaqPCRx DNA Polymerase Kit (Invitrogen, Carlsbad, CA) and primers *Rv2660c*-F4 (5′-TTGAAGGTTTTGGGGGCGAT-3′) and *Rv2660c*-R4 (5′-CCATCAGTGCCCTCGTTGAT-3′) that were located 269 bp upstream and 276 bp downstream of *Rv2660c*, respectively. In addition, another reverse primer, *Rv2660c*-R6 (5′ CTCAACGACCGTTTGCAGC-3′), located 114 bp downstream of *Rv2660c*, was used for the successful amplification of *Rv2660c* of the strains that contained sequence variation in *Rv2660c*-R4 primer region. The PCR master mix content was the same as previously described (Davila et al., 2010), except for an addition of 5 ul of 100% DMSO. The thermocycling program used for the *Rv2660c* amplification was: 1 min at 94 °C, 33 cycles at 94 °C for 30 s, 58 °C for 30 s, and 70 °C for 3 min and a final cycle at 72 °C for 10 min. PCR products were purified for DNA sequencing using PureLink PCR Micro Kit (Invitrogen Carlsbad, CA), following the manufacturer's instructions. The same PCR primers were also used for DNA sequencing of the PCR products from different study strains. DNA sequences of the study strains were compared to that of *M. tuberculosis* laboratory reference strain H37R using Muscle in MEGA 2.5 (http://www.megasoftware.net/).

2.3. HLA allele selection

Previous studies have indicated that both CD4 + and CD8 + cellular immune responses are necessary for *M. tuberculosis* infection immunologic control (Lewinsohn et al., 2011; Blythe et al., 2007). In order to assess population coverage of H56 in South Africa, we, therefore, selected MHC-I and MHC-II HLA gene alleles for epitope prediction. The selected HLA alleles included all the MHC-I and MHC-II HLA gene alleles that were identified by a previous study on nine Black and two Caucasian South African ethnic/linguistic groups through high resolution genotyping of a cohort of 302 South Africans (Paximadis et al., 2012). In addition, any HLA alleles not reported by Paximadis et al. but found in the Allele Frequency Net Database (AFND) (www.allelefrequencies.net) were also included to ensure covering the most complete genetic background of the study population (Gonzalez-Galarza et al., 2015). Specifically, we included 88 HLA-A, 240 HLA-B, 41 HLA-C and 40 HLA-DRB1 alleles. We included both MHC-I and MHC-II HLA alleles in our study because both CD4 + and CD8 + T-cells play an important role in TB immunity (Blythe et al., 2007; Lindestam-Arlehamn et al., 2014). The selected alleles were represented by nine Class I supertype alleles (A*0101, A*0201, A*0301, A*2601, B*0702, B*1501, B*2705, B*4001, and B*5801) and eight MHC Class II HLA-DRB1 supertype alleles (DRB1*0101, *0301, *0401, *0701, *0801, *1101, *1301, and *1501). Supertype alleles are a set of HLA alleles that share overlapping peptide binding specificity and represent the primary functional binding motifs of the majority of Class I alleles and almost all HLA-DRB1 alleles (Sette and Sidney, 1999; Lund et al., 2004).

2.4. Epitope prediction

T-cell epitope predictions were performed using the Immune Epitope Database (IEDB) standalone software, MHC-I version 2.17