Infection, Genetics and Evolution 9 (2009) 832–839

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Characterization of predominant *Mycobacterium tuberculosis* strains from different subpopulations of India

Jyoti Arora [a], Urvashi Balbir Singh [a,*], Naga Suresh [a], Tanu Rana [a], Chhavi Porwal [a], Amit Kaushik [a], Jitendra Nath Pande [b]

[a] Department of Microbiology, All India Institute of Medical Sciences, New Delhi, India
[b] Sitaram Bharatiya Institute of Science and Research, New Delhi, India

ARTICLE INFO

Article history:
Received 31 October 2008
Received in revised form 27 April 2009
Accepted 3 May 2009
Available online 20 May 2009

Keywords:
Tuberculosis
Spoligotyping
Restriction fragment length polymorphism
Shared type
Clade

ABSTRACT

The predominant strains from India belong to Central-Asian (CAS) and the East-African-Indian (EAI) clade of *Mycobacterium tuberculosis*. The two clades have also been shown to be geographically partitioned. The study of such strains may help to understand the characteristics that make *M. tuberculosis* an effective pathogen and its overrepresentation in certain populations. *M. tuberculosis* isolates characterized by spoligotyping under a population based tuberculosis study covering different regions from the North and South India were further analyzed by restriction fragment length polymorphism (RFLP) and by deletion analysis of *M. tuberculosis* specific deletion region 1 (TbD1). The genetic relationship of the two clades inferred using different genetic markers showed good correlation. In the North where the CAS clade predominates the isolates are characterized by presence of high IS*6110* copy number and absence of TbD1 region whereas in the South where the EAI clade predominates the isolates are characterized by low copy number of IS*6110* and presence of TbD1 region. The ancestral EAI strains were found to be less often associated with drug resistance or young age as compared to the CAS clade. The study highlights that the EAI lineage is well established in India and that CAS may be emerging or more recently introduced to India. The results depict a distinction in the lineage of strains from the North versus South India indicating a need to study if the pathogen has adapted to specific human populations.

1. Introduction

The resurgence of tuberculosis (TB) has renewed interest in understanding the epidemiology and pathogenesis of the disease. Genotyping of clinical isolates in different parts of the world has shown that global epidemiology of TB is propagated by thousands of different genotypes (van Soolingen et al., 1999; Warren et al., 1999). The strains occur at different frequencies, and the relative frequencies in different areas vary between districts, cities, countries and continents (Brudey et al., 2006; Filliol et al., 2002; Filliol et al., 2003). The dynamics of TB epidemic in a given area and time frame may therefore be a factor of the different strains circulating in that region. The knowledge of circulating strains can be used for molecular evolutionary and population genetics studies.

IS*6110*-based typing is the most widely applied genotyping method in the molecular epidemiology of *Mycobacterium tuberculosis*. The IS*6110* insertion sequence is specific to the *M. tuberculosis*, are present in different copy numbers (0–25) and their positions in the *genome* are highly variable among different isolates (van Embden et al., 1993). Spoligotyping targeting the DNA polymorphism at the direct repeat locus (DR locus) of the genome of *M. tuberculosis* complex, allows simultaneous detection and differentiation of *M. tuberculosis* complex strains (Kamerbeek et al., 1997). The DR locus is well conserved and stable enough rendering it specific for detecting *M. tuberculosis* complex strains.

SNPs in the katG and gyrA genes classify *M. tuberculosis* isolates into three principal genetic groups (PGGs). TbD1 is specifically present in a subset of PGG1 strains, but absent in other strains of PGG1, and in PGG2 and PGG3 strains. Based on the presence or absence of an *M. tuberculosis* specific deletion (TbD1), *M. tuberculosis* strains can be segregated into "ancestral" versus "modern" lineages (Brosch et al., 2002). Though the known "modern" *M. tuberculosis* families are very widely prevalent worldwide (Filliol et al., 2002; Filliol et al., 2003; Brudey et al., 2006), ancient principal genetic group 1 (PGG1) clones are responsible for TB in India (Kulkarni et al., 2005; Gutierrez et al., 2006; Singh et al., 2007). Studies in India lack reference to preferential localization of certain strain families in different

* Corresponding author. Tel.: +91 11 26588500/700x6436; fax: +91 11 26588641.
E-mail address: drurvashi@gmail.com (U.B. Singh).

1567-1348/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2009.05.008

subpopulations among different regions of the country. Despite the high burden of TB and various epidemiological studies done in the North and the South regions of India (Das et al., 1995; Narayanan et al., 1997; Radhakrishnan et al., 2001; Siddiqi et al., 2001; Bhanu et al., 2002; Mistry et al., 2002; Kulkarni et al., 2005; Gutierrez et al., 2006; Suresh et al., 2006; Singh et al., 2007; Sharma et al., 2008; Ahmed et al., 2009; Stavrum et al., 2009) there is limited data available pertaining to strains circulating in the country.

Epidemiological data from different studies suggests that difference in transmissibility and virulence among *M. tuberculosis* strains are related to genetic makeup of the organisms (Valway et al., 1998; Caminero et al., 2001; Lopez et al., 2003). Over-representation of clades suggests that they possess biological advantage in specific populations and study of such strains may help to understand the characteristics that make *M. tuberculosis* an effective pathogen.

Based upon our earlier results of spoligotyping on 540 *M. tuberculosis* isolates from five different geographical regions of India, we concluded that different parts of India also differed in regard to predominant spoligotypes, underlining the differences in the introduction and evolution of TB in India, the East-African-Indian (EAI) clade being more commonly found in Southern parts and the (Central-Asian) CAS clade being more predominant in the North (Singh et al., 2007). The present study was undertaken to study genetic composition of the EAI and CAS isolates using different genetic markers.

The study also reflects on the associations between lineages and patient demographic variables.

2. Materials and methods

2.1. Study population

Clinical isolates of *M. tuberculosis* were obtained during the study period of January 2001–September 2003 under a multi-centric (eight centres over the country) project supported by INDIA-CLEN (INCLEN, USA). Samples were processed on the day of arrival by modified Petroff's method and cultured on LJ slants at 37 °C for 6–8 weeks in the respective laboratories. Testing for susceptibility to Rifampin (RIF, 40.0 mg/ml), Ethambutol (EMB, 2.0 mg/ml), Isoniazid (INH, 0.2 mg/ml) and Streptomycin (SM, 4.0 mg/ml) was performed according to the Proportion method (Laszlo et al., 1997). Randomly selected culture isolates (n – 540) from three North Indian cities (Delhi, Pune and Lucknow) and two South Indian cities (Chennai and Trivandrum) were collected at the All India Institute of Medical Sciences (AIIMS), New Delhi, laboratory and stored at −70 °C for further workup. These were well characterized by spoligotyping (Singh et al., 2007) using the standard protocol (Kamerbeek et al., 1997). Of the total isolates spoligotyped, 324 isolates belonged to the two predominant clades, CAS and EAI (153 isolates belonged to CAS clade and 171 isolates belonging to EAI clade). These isolates were further worked up.

2.2. Data analysis and statistics

Demographic data from patients including age, sex, previous history of TB, history of smoking, alcohol and family TB, and hometown were collected at admission in standard questionnaires. The clinical data were analyzed from 320/324 patients with the EAI or CAS clade isolate (the data from four patients was incomplete). Data were analyzed using the EpiInfo v.6.0 program and observed frequencies were compared by means of two-by-two contingency tables. A p value of <0.05 was considered statistically significant. The data and demographic parameters are summarized in Table 1.

Table 1
Clinical and epidemiologic characteristics of patients harbouring isolates belonging to EAI and CAS clade.

Parameters	Number (%) of patients in		p Value
	CAS clade	EAI clade	
Age (y)			
15–45	126 (84)	119 (70)	0.0048
>46	24 (16)	51 (30)	
Gender			
Male	114 (76)	122 (72)	NS
Female	36 (24)	48 (28)	
Birth place			
Village	129 (86)	140 (82)	NS
City	21 (14)	30 (18)	
Previous history of TB			
No previous therapy	67 (45)	119 (70)	0.0003
Failed/defaulter/relapsed	70 (46)	42 (25)	
Cured	13 (9)	9 (5)	
Smoking			
Yes	69 (46)	81 (45)	NS
No	81 (54)	89 (55)	
Alcohol			
Yes	72 (48)	73 (44)	NS
No	78 (52)	97 (56)	
Family TB			
Yes	39 (26)	41 (24)	NS
No	111 (74)	129 (76)	
TB death			
Yes	24 (16)	27 (16)	NS
No	126 (84)	143 (84)	
Drug resistance			
Resistance to one or more drug	69 (46)	56 (33)	0.023
Susceptible to all drugs	81 (54)	114 (66)	
X-ray			
Ext	102 (68)	107 (63)	NS
Lim	48 (32)	63 (37)	
Radiologic findings			
Cav	105 (70)	119 (70)	NS
Noncav	45 (30)	51 (30)	
Sputum smear results			
Negative	12 (8)	12 (7)	NS
1+	52 (40)	56 (33)	
2+ or 3+	78 (52)	102 (60)	

NS, not significant statistically; y, year.

2.3. Restriction fragment length polymorphism (RFLP)

46/153 clinical isolates belonging to CAS clade and 44/171 isolates belonging to EAI clade were selected at random (using random table) from a total of 324 isolates, for RFLP. The distribution of isolates from different cities/geographical regions, their shared-types according to SpolDB4 (Brudey et al., 2006) and the IS*6110* copy number are shown in Table 2.

DNA extraction and IS*6110* RFLP were performed using the recommended international standard protocol (van Embden et al., 1993). The IS*6110* RFLP patterns were analyzed by visual examination as well as computer assisted analysis by use of GelCompar version 3.5 software. (Applied Maths, Belgium). Similarities between RFLP patterns were calculated by using the Dice coefficient, and the dendrogram was produced with the unweighted pair group method using arithmetic averages algorithm (UPGMA).

2.4. TbD1 PCR analysis

The TbD1 region was analyzed using PCR primers designed for TbD1 regions according to the method of Brosch et al. (2002). Primers complementary to the internal sequence as well as those flanking the deleted regions were used. Product was obtained with the internal primers or with the flanking primers depending on the presence or absence of the TbD1 region.