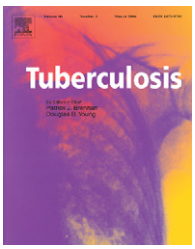
Tuberculosis (2007) **87**, 426–436

Available at www.sciencedirect.com

ScienceDirect

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



Genome analysis shows a common evolutionary origin for the dominant strains of *Mycobacterium tuberculosis* in a UK South Asian community ☆

M. Carmen Menéndez[a,1], Roger S. Buxton[a,*], Jason T. Evans[b,e], Deborah Gascoyne-Binzi[c], Rachael E.L. Barlow[c], Jason Hinds[d], Peter M. Hawkey[b,e], Dr. M.J. Colston[a,✠]

[a]*Division of Mycobacterial Research, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK*
[b]*Division of Immunity and Infection, University of Birmingham, The Medical School, Edgbaston, Birmingham B15 2TT, UK*
[c]*Department of Microbiology, Leeds General Infirmary, Leeds LS1 3EX, UK*
[d]*Bacterial Microarray Group, Division of Cellular and Molecular Medicine, St. George's, University of London, Crammer Terrace, London SW17 0RE, UK*
[e]*Health Protection Agency – West Midlands Laboratory, Birmingham, Heartlands Hospital, Birmingham B9 5SS, UK*

Received 8 September 2006; received in revised form 27 April 2007; accepted 3 May 2007

KEYWORDS
Mycobacterium tuberculosis;
PE/PPE;
Polymorphism;
VNTR 42235;
South Asian community

Summary
We have investigated the *Mycobacterium tuberculosis* strain types present in the South Asian population of the UK, in which tuberculosis is particularly prevalent. In contrast to the widespread Beijing strains which have the variable number tandem repeats (VNTR) profile 42435, isolates with the VNTR profile 42235, jointly with 02335 or 42234 profiles, appear more frequently in tuberculosis patients of South Asian ethnic origin (SA-strains) in the UK than in any other ethnic group. Using microarray-based comparative genomics to distinguish total or partially deleted genes, we found that three of the common deleted regions in the SA-strains were identical to some deleted genes in the strain CH, which caused an outbreak among South Asian patients in Leicester in 2001 but were different from genomic deletions found in Beijing/W strains. Analysis of some of the deleted regions revealed differences in comparison to the strain CH including the polymorphism in some of

☆Accession numbers: The nucleotide sequences of deleted regions of SA-strains are in the EMBL Data Bank with the accession numbers: AJ878456, AJ878457, AJ878458, AJ878459, AJ878460, AJ878461, AJ879166, AJ879167, AJ879168, AJ879169, AJ879170, AJ879171, AJ879172, AJ879173, AJ879174, AJ879175, AJ879176, AJ879177, AJ879178, AJ879179, AJ879180, and AJ879181.

*Corresponding author. Tel.: +44 20 8816 2225; fax: +44 20 8906 4477.
E-mail address: rbuxton@nimr.mrc.ac.uk (R.S. Buxton).
[1]Present address: Departamento de Medicina Preventiva, Facultad de Medicina, Universidad Autonoma de Madrid, Arzobispo Morcillo, s/n. 28029 Madrid, Spain.
✠Deceased 20th February 2003.

1472-9792/$ - see front matter
doi:10.1016/j.tube.2007.05.017

the PE/PPE and Esat-6 genes, which may be responsible for the diversity of antigenic variation or differences in the activation of the host immune response. Interrupted genes or the replacement by insertion elements was confirmed in some of the deleted genomic regions. Our results are consistent with the hypothesis that the SA-strains may present common features, implying a common origin for this group of strains.

Introduction

According to the World Health Organization report[1] one-third of the world's population is infected by *Mycobacterium tuberculosis*. The bacillus causes 1.6 million deaths each year and more than 8 million new cases annually, with the majority in South East Asia (3 million cases each year). In the UK, despite a low national prevalence of tuberculosis, the incidence is much higher in cities with a large South Asian community.[2] A large outbreak of TB was recently reported in 2001 among people of South Asian origin in a school in Leicester, a city with a high overall incidence of tuberculosis.[3] Amongst people of South Asian ethnic origin a dominant group has been described.[4] This epidemiological profile can be identified by characteristic Exact Tandem Repeat profiles (42235, 02235 and 42234).[4] The variable number tandem repeats (VNTR) profile 42235 represented 23% of patient isolates found in Leeds and Bradford, cities in West Yorkshire, UK, and 37% of patient isolates collected in Rawalpindi, Pakistan.[4] It is essential to emphasise that this particular group of strains, which we have called South Asian strains (SA-strains), were not related epidemiologically to those in the outbreak in Leicester despite most of the strains having the identical VNTR profile 42235.[3]

Recent work in San Francisco suggests that patients tend to become infected by strains of *M. tuberculosis* that have similar genotypes to those associated with their region of birth.[5] The susceptibility to intracellular infections in people of South Asian ethnicity cannot be the only factor in the specific host–pathogen interaction with these strains as the combination of host factors with microbial determinants is also likely to play an important role in strain specificity with certain patient populations.

The molecular basis of pathogenicity, virulence and transmissibility in *M. tuberculosis* is not well known. The study of genetic variability within natural populations of pathogens can provide insight into their evolution and pathogenesis with comparative genomics a powerful tool providing important data that can be used to control transmission, and complements the more extensive studies of variation resulting from insertion sequences such as IS*6110*.[6,7]

Many new deleted regions have been found in the genome of different clinical strains of various members of the *M. tuberculosis* complex. Comparison of the genomes of *M. tuberculosis* H37Rv and *M. bovis* Bacilli Calmette-Guérin Pasteur (BCG) identified 14 sequences present in *M. tuberculosis* H37Rv but absent in *M. bovis* BCG. These were called regions of difference (RD1-14).[8] Similarly, 6 regions were identified, that were absent from the *M. tuberculosis* H37Rv genome relative to other members of the *M. tuberculosis* complex: H37Rv relative deletions (RvD1-5) and *M. tuberculosis* specific deletion 1 (TbD1).[8,9] Evolutionary studies have been undertaken to characterise genomic deletions and determine the probable evolution in a large number of different clinical isolates of *M. tuberculosis*[9–11] and some of the genomic deletions of strains of *M. tuberculosis* studied were suggested as useful markers for defining the Beijing/W family of strains[11] and different global lineages of *M. tuberculosis*.[12]

The aim of our study was to investigate the genomic characteristics in a set of strains which have been reported predominantly amongst the South Asian community in the UK, in contrast to other strains such as those of the Beijing family which have spread very widely around the world among ethnically mixed populations. We have used microarray-based comparative genomics to analyse the distribution of the deleted regions around the genome of six clinical strains of *M. tuberculosis* defined as belonging to the South Asian group by associated VNTR profiles (42235, 02235 and 42234[4]) in comparison to the previously reported CH strain. Examples of clades that are also common but have been isolated from ethnically diverse patients were also included.[13,14] Our results indicate that strains of the South Asian group have a common evolutionary origin similar to the strain CH but distinct from members of the Beijing/W clade. The SA-strains appear to be included in the East-African-Indian lineage, defined by the RD750 deletion, which is frequently found in Southeast Asia.[12] Recent results of Newton and collegues[15] highlights the immunological relevance of this deleted region in the CH strain.

Materials and methods

Strains and growth conditions

Clinical strains 8088, 9375 and 9866 together with isolate 6947 were isolated from South Asian patients in Leeds and Bradford and were from the collection of D. Gascoyne-Binzi (Leeds Teaching Hospitals). The other clinical strains (0135, 2566, 3242) from South Asian patients and 2 strains of ''Haarlem family'', 1339 and 7009, were from the collection of P.M. Hawkey (University of Birmingham). The VNTR profiles of these clinical strains are shown in Table 1. The South Indian clinical isolate (TMC120) ATCC 35811 was also included in this study with *M. tuberculosis* H37Rv used as the reference strain.

All the strains were grown at 37 °C in Dubos medium containing 0.05% Tween and supplemented with 0.04% (v/v) Dubos medium Albumin and 0.2% (v/v) glycerol.

DNA isolation and hybridization

Genomic DNA extraction[16] and microarray hybridization procedures were performed as previously described,[17,18]