Infection, Genetics and Evolution 12 (2012) 1917–1925

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

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

Factors associated with genotype clustering of *Mycobacterium tuberculosis* isolates in an ethnically diverse region of southern California, United States

Timothy C. Rodwell [a,*], Anokhi J. Kapasi [b], Richard F.W. Barnes [a], Kathleen S. Moser [c]

[a] University of California San Diego School of Medicine, 9500 Gilman Drive, MC 0507, La Jolla, CA 92093-0507, USA
[b] The Degge Group, Ltd., 1616 North Fort Myer Drive, Suite 1430, Arlington, VA 22209, USA
[c] County of San Diego Health and Human Services Agency, Tuberculosis Control and Refugee Health Program, 3851 Rosecrans Street, San Diego, CA 92110, USA

ARTICLE INFO

Article history:
Received 13 March 2012
Received in revised form 24 August 2012
Accepted 27 August 2012
Available online 13 September 2012

Keywords:
Tuberculosis
Molecular epidemiology
Social networks
Genotypes
Spoligotype
MIRU

ABSTRACT

Mycobacterium tuberculosis (*Mtb*) isolates with identical genotypes, found in different patients, are most likely the result of recent transmission. *Mtb* strains with closely related genotypes, called clonal complexes, are most likely derived from one another. We examined *Mtb* genotypes from southern California TB patients from 2005 through 2008 to complete the first comprehensive molecular epidemiology analysis of this complicated and ethnically diverse region. *Mtb* genotypes were characterized with spoligotype and MIRU-12 typing. MIRU-VNTRplus was utilized to assign genotypes to global lineages and complete cluster analyses. Associations between patient characteristics and genotype clustering and clonal complexes were evaluated using logistic regression and frequency analysis. Of 832 *Mtb* isolates analyzed, 480 (58%) fell into 94 strain clusters. The majority of isolates were identified as being in the EA1 (31%), LAM (17%) and Haarlem (15%) lineages, but 13 different lineages were found in this region. TB patients with clustered isolates were more likely to be homeless (AOR 3.44, 95% CI 1.65, 7.18) and male (AOR 1.57, 95% CI 1.17, 2.10). Of the 480 clustered strains, 388 aggregated into six clonal complexes.

Over 45% of reported TB cases were clustered and likely resulted from recent transmission events. Patients with clustered *Mtb* isolates that were grouped into clonal complexes had unique sociodemographic characteristics. These data suggest that TB is being transmitted in relatively insular community networks defined by race/ethnicity and country of origin. The addition of clonal complex analysis to simple cluster analysis provides important public health insights into the local transmission of TB in ethnically diverse regions with diverse *Mtb* genotypes.

1. Introduction

Molecular epidemiology has revolutionized the study of tuberculosis (TB) transmission and the management of TB disease (Allix-Beguec et al., 2008; Bezanahary et al., 2008; David, 2008; Fica et al., 2008; Mathema et al., 2008). The fundamental principal behind the molecular epidemiology of TB is that, unlike other bacteria, *Mtb* replicates clonally and almost never transfers genes horizontally. This together with the fact that *Mtb* underwent a genetic bottleneck recently, makes it relatively monomorphic (Gibson et al., 2008). Detectable genetic variation between different *Mtb* bacilli does evolve over time, however, which means *Mtb* strains with identical genotypes found in different patients are most likely the result of recent transmission (Allix-Beguec et al., 2008; Bezanahary et al., 2008; David, 2008; Fica et al., 2008; Mathema et al., 2008). Furthermore, *Mtb* strains with genotypes shown to be closely related by phylogenetic analysis (defined as differing by only one MIRU locus in this study) were probably derived from one another, giving rise to the notion of a clonal complex of *Mtb* strains that may, but not necessarily do, share a recent transmission event as well (Allix-Beguec et al., 2008; Weniger et al., 2010).

From a TB management perspective, characterizing *Mtb* strains at the genomic level enables TB programs to track the transmission of specific *Mtb* strains, follow epidemics, and identify new outbreaks. Evaluating molecular epidemiology from a population genetics perspective also enables public health researchers to determine which patient factors might be associated with clusters of related strains and how interventions might be formulated to prevent those strains from being transmitted in the future (Behr and Mostowy, 2007; Garzelli and Rindi, 2012).

In the US, the national genotyping program determines *Mtb* genotypes primarily on the basis of spacer oligonucleotide typing (spoligotyping) (Kamerbeek et al., 1997) and mycobacterial interspersed repetitive unit (MIRU) typing (Mazars et al., 2001). In this

* Corresponding author. Address: Division of Global Public Health, Department of Medicine, University of California San Diego, 9500 Gilman Drive, MC 0507, San Diego, CA 92093-0507, USA. Tel.: +1 858 822 3018; fax: +1 858 534 7566.
E-mail addresses: trodwell@ucsd.edu (T.C. Rodwell), anokhik@gmail.com (A.J. Kapasi), rgarfein@ucsd.edu, r1barnes@ucsd.edu (R.F.W. Barnes), Kathleen.Moser@sdcounty.ca.gov (K.S. Moser).

1567-1348/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2012.08.022

study, we examined *Mtb* spoligotype and MIRU genotypes from San Diego County from 2005 through 2008, in order to complete the first comprehensive molecular epidemiology analysis of TB of this region.

San Diego County is unique in the US. Over 70% of TB cases in San Diego County occur in foreign-born patients of multiple nationalities, and TB incidence is almost double the US national average (San Diego Health and Human Service Agency (HHSA), 2011). In order to better understand the factors associated with transmission of TB in this complicated and ethnically diverse region, we completed, (i) an analysis of *Mtb* strain lineages to determine the diversity of strains present; (ii) an analysis of clustered vs. non-clustered strains to determine the proportion of recent transmission vs. reactivated TB cases (Small et al., 1994); and (iii) an analysis of clonal complexes within the clustered cases to determine how the recent transmission cases might be related to one another genetically and epidemiologically (Allix-Beguec et al., 2008; Weniger et al., 2010).

2. Materials and methods

2.1. Study population

San Diego is the second largest county in California with a diverse population of over three million people (United States Census Bureau, 2011) and one of the highest TB case rates in the US (6.0/100,000 population) (Centers for Disease Control and Prevention (CDC), 2011). Located in the southernmost portion of the state, it shares one of the world's busiest land border-crossings with Mexico (Lange et al., 1999), and has a high rate of immigration from countries such as Vietnam and Philippines, where TB is endemic (International Community Foundation (ICF), 2010). California law (Health and Safety Code Title 17 §2505) requires that all verified cases of TB be documented and reported to the US National TB Control Program. Approximately 200–300 incident TB cases are reported in San Diego County annually (San Diego Health and Human Service Agency (HHSA), 2011).

2.2. Data source

Socio-demographic and clinical data, as well as pathogen genotypes for TB patients, included in this study were obtained from the San Diego Report of a Verified Case of Tuberculosis (RVCT) database supplemented with locally collected variables. RVCT data were collected based on US national TB surveillance guidelines and all TB cases evaluated in the study met the national surveillance TB case definition (laboratory or clinical evidence of disease caused by *Mycobacterium tuberculosis* complex) (Centers for Disease Control and Prevention (CDC), 1997). Pathogen genotypes and TB surveillance variables from all culture-positive cases of TB reported to the San Diego County TB program from 2005 through 2008 were included in this study. The study protocol was approved by an Institutional Review Board at the University of California, San Diego.

2.3. TB culture and species identification

Since the early 1990s, the San Diego County Public Health Laboratory obtained an *Mtb* isolate from over 80% of all TB patients for pathogen species identification and drug susceptibility testing (DST) (Rodwell et al., 2008). All TB isolates from patient specimens were initially identified as *Mtb* complex on the basis of the AccuProbe hybridization protection assay (GenProbe, San Diego, CA, USA). Due to the fact that TB from *Mycobacterium bovis* is also prevalent in Southern California (Rodwell et al., 2010b; Rodwell et al., 2008), cultured isolates were further identified as either *Mtb* or *M. bovis* on the basis of culture morphologic findings, the results of the niacin strip test, the nitrate reduction test (Grange et al., 1996) or their spoligotype (Streicher et al., 2007). TB cases identified as *M. bovis* strains were excluded from further analysis.

2.4. Strain genotyping

Mtb genotypes were characterized using spoligotyping (Kamerbeek et al., 1997) and MIRU-12 typing (Mazars et al., 2001). Spoligotyping and MIRU were performed and reported by the Microbial Diseases Laboratory at the California Department of Public Health according to Centers for Disease Control and Prevention (CDC) guidelines (Centers for Disease Control and Prevention (CDC), 2004). Spoligotyping was completed with Luminex-based methods which detect 43 specific spacer sequences in the direct repeat locus (Cowan et al., 2004). Each spoligotype was converted from the 43-digit binary sequence to the octal format for analysis (Dale et al., 2001). MIRU-12 was completed using procedures described by Cowan et al. (2002) to determine the number of repeat sequences at the 12 MIRU loci: 02, 04, 10, 16, 20, 23, 24, 26, 27, 31, 39, and 40 (Mazars et al., 2001). We acknowledge that MIRU-15 or MIRU-24 typing has recently been adopted by most nations (including the US) to replace MIRU-12 genotyping as MIRU-12 typing plus spoligotyping has been shown to overestimate isolate clustering due to insufficient variation at the MIRU-12 loci (Supply et al., 2006). As this analysis spans the years 2005 through 2008, MIRU-24 data was not available. We discuss the potential effect of MIRU-12 genotyping on our estimates in the discussion section.

2.5. Determination of lineage and lineage evaluation

Genotype information was uploaded to the MIRU-VNTRplus web application (http://www.miru-vntrplus.org) for lineage assignment and cluster analysis (Allix-Beguec et al., 2008; Weniger et al., 2010). A strain lineage was identified for each isolate using the "similarity search" module, which compared the combined spoligotype and MIRU genotypes against a collection of 186 reference strains representing major global *Mtb* complex lineages (Allix-Beguec et al., 2008). A 4-step progressive approach was employed for phylogenetic lineage assignment. First, the lineage matches were based on categorical distance measures using MIRU and spoligotype sequences. In the "tree-based identification" module set to the neighbor-joining (NJ) algorithm (Saitou and Nei, 1987), any isolates still missing lineages were then analyzed for categorical distance matches to the nearest subtree using MIRU sequence alone, then MIRU and spoligotype sequences for further remaining isolates, and finally, spoligotype sequences alone for the last unassigned isolates. Characterized lineages were matched with the global repository of TB lineages (spolDB4) and strains with no match in spolDB4 were considered "orphan" strains (Brudey et al., 2006).

2.6. Cluster definition

TB cases were defined as "clustered" when $\geqslant 2$ *Mtb* isolates from different patients evaluated during the study period had identical spoligotype and MIRU-12 genotypes. Strains not found to have matching genotypes with any other isolates in the study period were classified as "singletons." Clustered cases were assumed to have resulted from recently acquired, person-to-person infections while singleton cases were assumed to most likely have originated from reactivated latent tuberculosis infections acquired distantly in time and space (Mathema et al., 2008; Small et al., 1994). Patients with the earliest report date within a cluster were assumed to be the index cases, making the number of assumed