Tuberculosis 95 (2015) S177–S189

Contents lists available at ScienceDirect

Tuberculosis

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

Tuberculosis – A global emergency: Tools and methods to monitor, understand, and control the epidemic with specific example of the Beijing lineage

David Couvin, Nalin Rastogi*

WHO Supranational Tuberculosis Reference Laboratory, Tuberculosis and Mycobacteria Unit, Institut Pasteur de la Guadeloupe, Abymes, Guadeloupe, France

Keywords:
Mycobacterium tuberculosis
Beijing genotype
Epidemiology
Spoligotyping
MIRU-VNTR
Drug-resistance

S U M M A R Y

We argue in favor of a concerted and coordinated response to stop tuberculosis (TB) by monitoring global TB spread, drug-resistance surveillance and populations at risk using available molecular and web tools to identify circulating clones of *Mycobacterium tuberculosis* complex (MTBC). We took specific example of the Beijing lineage associated with worldwide emergence of both multiple, and extensively drug resistant (MDR/XDR)-TB. The study dataset (n = 10,850 isolates, 92 countries of patient origin) was extracted from our multimarker SITVIT2 database on MTBC genotyping (n = 111,635 isolates, 169 countries of patient origin). Epidemiological and demographic information in conjunction with spoligotyping (n = 10,850), MIRU-VNTR minisatellites (n = 2896), and drug resistance (n = 2846) data was mapped at macro-geographical (United Nations subregions) and country level, followed by statistical, bioinformatical, and phylogenetical analysis. The global male/female sex ratio was 1.96, the highest being 4.93 in Russia vs. range of 0.8–1.13 observed in Central America, Caribbean, Eastern Africa and Northern Europe ($p < 0.0001$). The major patient age-group was 21–40 yrs worldwide except Japan (with majority of patients >60 yrs). Younger patients were more common in South America, South Asia, and Western Africa since 25–33% of TB cases due to Beijing genotype occurred in the age group 0–20 yrs. A continuous progression in the proportion of MDR and XDR strains is visible worldwide since 2003 and 2009 respectively. Pansusceptible TB mainly concerned older patients >60 yrs (44%) whereas Drug resistant, MDR and XDR-TB concerned patients preferentially aged 21–40 yrs (between 52 and 58%). Although the proportion of SIT1 pattern vs. other patterns was very high (93%); the proportion of MDR was highest for an emerging genotype SIT190 ($p < 0.0001$). Lastly, proportion of pansusceptible strains was highest in Japan, while MDR/XDR strains were most common in Russia and Northern Europe. We underline remarkable macro/micro-geographical cleavages in phylogenetic and epidemiologic diversity of Beijing genotype, with phylogeographical specificity of certain genotypes.

1. Introduction

Tuberculosis (TB) remains a major global health problem and one of the most important infectious diseases affecting mankind today. In 2012, an estimated 8.6 million people developed TB and 1.3 million died from the disease; poverty, malnutrition, overcrowding and immunosuppression are some risk factors which promote TB expansion [1]. A concerted and coordinated response to monitor and assess global TB spread, drug-resistance surveillance and populations at risk is urgently needed for global TB control. In the last decade, innovations in the molecular epidemiology of tuberculosis have helped define the global distribution of *Mycobacterium tuberculosis* lineages, to monitor the international spread of high-risk strains, to explore the evolutionary features of the bacterium, as well as to discriminate between events of recent transmission from those due to reactivation and to differentiate between recurrences due to reactivations vs. exogenous reinfections [2].

Much of these advances in micropopulation and macropopulation analysis of *M. tuberculosis* global epidemiology were made thanks to two major advances: PCR-based large-scale,

* Corresponding author. Institut Pasteur de Guadeloupe, Morne Jolivière, BP484, F97183 Abymes Cedex, Guadeloupe, France. Tel.: +33 590 590 897661; fax: +33 590 590 896941.
E-mail address: nrastogi@pasteur-guadeloupe.fr (N. Rastogi).

http://dx.doi.org/10.1016/j.tube.2015.02.023
1472-9792/

high-throughput genotyping of *M. tuberculosis* and publicly available huge international databases in conjunction with web-tools [reviewed in [2-4]]. Although most of the earlier molecular epidemiological studies were performed using IS*6110*-RFLP [5], this labor-intensive methodology was replaced with alternative PCR-based strategies in the last decade. The most commonly used methodology for TB genotyping today essentially uses a combination of 2 methods – spoligotyping based on the polymorphism of the direct repeat (DR) locus [6] and MIRU-VNTR minisatellites used in 12–, 15–, or 24–loci formats [7,8], and constitutes the gold-standard for optimal TB surveillance at national, regional, and global levels by "universal genotyping" of patient isolates. A query in PubMed on October 30th 2013 made for "IS*6110*-RFLP AND tuberculosis" returned 210 publications vs. 801 for "spoligotyping AND tuberculosis", 546 using the term "MIRU OR VNTR OR MLVA OR MIRU-VNTR AND tuberculosis", and 259 publications looking into the association of last 2 queries. In summary, the above mentioned PCR-based genotyping methods in conjunction with classical epidemiological investigations show adequate resolution for tracing TB transmission and predicting diverse strain lineages, and are used widely for TB surveillance worldwide today [9–12].

Furthermore, the numerical format of the spoligotype patterns and MIRU-VNTRs made it possible that the information generated is easily stored in huge international databases and made available to the international research community along with new web tools to study TB epidemiology and compare genotyping information worldwide; some examples include recently published SITVITWEB ([13]; http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE) and its earlier version SpolDB4 [14]; MIRU-VNTRplus ([15]; http://www.miru-vntrplus.org/MIRU/index.faces); and TB-Lineage ([16]; http://tbinsight.cs.rpi.edu/run_tb_lineage.html).

Working on tools and methods to monitor, understand, and control the TB epidemic for last many years at Institut Pasteur de la Guadeloupe, we recently undertook the construction of a newer database named SITVIT2; the version under development already contains spoligotyping data on 111,635 *M. tuberculosis* complex clinical isolates from 169 countries of patient origin – almost twice as compared to the SITVITWEB developed and released publicly in 2012 (which incorporated spoligotyping data on 62,582 clinical isolates corresponding to 153 countries of patient origin [13]). Addition of new tools to the database and supplementary information allows global mapping of combined search results on *M. tuberculosis* genotypes, epidemiological and demographic information, and drug resistance.

In this paper, we evaluate the usefulness of such an approach by studying the specific example of the *M. tuberculosis* Beijing lineage which represents about 50% of all TB strains in East Asia and at least 13% of strains worldwide [17]. First reported in East Asian countries where they are highly endemic [18,19] and later found to be disseminated throughout the world [20,21], Beijing family strains are predominant in the former USSR countries, especially in Russia [20–22]. With an intrinsic advantage over other *M. tuberculosis* genotypes in terms of virulence (i.e., transmission, progression from latent to active tuberculosis, acquisition of drug resistance, or disease chronicity), Beijing lineage strains are notably associated with multiple drug resistant (MDR) and extensively drug resistant (XDR) TB [17,23,24]. It has been suggested that the study of Beijing genotype may reflect hidden patterns of human migrations [25,26]; undoubtedly, the rapid global dissemination of this most successful clone of *M. tuberculosis* complex (MTBC) makes it a major cause of concern in public health. In this context, this paper will describe global mapping of *M. tuberculosis* Beijing lineage strains by analyzing its population structure and available demographic, epidemiologic, and drug-resistance data in the SITVIT2 database.

2. Materials and methods

2.1. Brief description of the database

The detailed description of the SITVIT2 database, which is an updated version of the recently released SITVITWEB database [13], will be the subject of a separate publication (refer to Table 1 for a summary of SITVITWEB vs. SITVIT2 databases). At the time of this study, SITVIT2 contained a total of 111,635 MTBC clinical isolates from 169 countries of patient origin. In this database, Spoligotype International Type (SIT) and MIRU International Type (MIT) designate identical patterns shared by 2 or more patient isolates, whereas "orphan" designates patterns reported for a single isolate that does not correspond to any of the patterns recorded in the repository of the SITVIT2 database. Furthermore, strains are classified in major phylogenetic clades assigned according to signatures provided earlier, which includes various MTBC members (AFRI, *Mycobacterium africanum*; BOV, *Mycobacterium bovis*; CANETTII, *Mycobacterium canettii*; MICROTI, *Mycobacterium microti*; PINI, *Mycobacterium pinnipedii*), as well as for *M. tuberculosis* sensu stricto, i.e., the Beijing clade, the Central-Asian (CAS) clade, the East-African-Indian (EAI) clade, the Haarlem/Ural clades, the Latin-American-Mediterranean (LAM) clade, the Cameroon and Turkey lineages, the "Manu" family, the IS*6110*-low banding X clade, and the ill-defined T clade. Note that some spoligotypes previously classified as H3/H4 sublineages within Haarlem family were recently relabeled "Ural" [27]; these include patterns belonging to H4 sublineage that were relabeled "Ural-2", and some patterns previously classified as H3 sublineage but with an additional specific signature (presence of spacer 2, absence of spacers 29 to 31, and 33–36), that are now relabeled "Ural-1". With their definitive reclassification pending, we refer to these as H4/Ural-2 and H3/Ural-1 in SITVIT2 database. Furthermore, two LAM sublineages were recently raised to independent lineage level: LAM10-CAM as

Table 1
A summarized representation of the SITVITWEB and SITVIT2 databases and the corresponding major phylogenetic lineages of the *M. tuberculosis* complex (MTBC).

Major lineages*	SITVITWEB* (n = 62,582)		SITVIT2 (n = 111,635)	
	Nb	%	Nb	%
Beijing	6159	9.84	10,850	9.72
AFRI	695	1.11	965	0.86
BOV	6486	10.36	25,741	23.06
CANETTII	12	0.02	12	0.01
CAS	2480	3.96	4362	3.91
EAI	4674	7.47	6617	5.93
Haarlem/Ural	7058	11.28	10,580	9.48
LAM	8042	12.85	12,245	10.97
Cameroon (previously LAM10-CAM)	650	1.04	1095	0.98
Turkey (previously LAM7-TUR)	370	0.59	593	0.53
Manu	675	1.08	1064	0.95
MICROTI	29	0.05	29	0.03
PINI	152	0.24	159	0.14
S	1151	1.84	1606	1.44
T	12,038	19.24	17,947	16.08
X	4088	6.53	4683	4.19

* The strains are classified in major phylogenetic clades assigned according to signatures provided earlier [13]; which includes various MTBC members (AFRI, *M. africanum*; BOV, *M. bovis*; CANETTII, *M. canettii*; MICROTI, *M. microti*; PINI, *M. pinnipedii*), as well as for lineages/sub-lineages of *M. tuberculosis* sensu stricto (note that the sublineages are not shown), i.e., the Beijing clade, the Central-Asian (CAS) clade, the East-African-Indian (EAI) clade, the Haarlem/Ural clades, the Latin-American-Mediterranean (LAM) clade, the Cameroon and Turkey lineages, the "Manu" family, the IS*6110*-low banding X clade, and the ill-defined T clade. Note that the sublineages are not shown.