Infection, Genetics and Evolution 35 (2015) 82–88

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Phylogenetic assignment of *Mycobacterium tuberculosis* Beijing clinical isolates in Japan by maximum *a posteriori* estimation

Junji Seto [a,*,1], Takayuki Wada [b,*,1], Tomotada Iwamoto [c], Aki Tamaru [d], Shinji Maeda [e], Kaori Yamamoto [f,g], Atsushi Hase [f], Koichi Murakami [h], Eriko Maeda [h], Akira Oishi [h], Yuji Migita [i], Taro Yamamoto [b,g], Tadayuki Ahiko [a]

[a] Department of Microbiology, Yamagata Prefectural Institute of Public Health, 1-6-6 Toka-machi, Yamagata-shi, Yamagata 990-0031, Japan
[b] Department of International Health, Institute of Tropical Medicine, Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan
[c] Department of Microbiology, Kobe Institute of Health, 4-6 Minatojima-nakamachi, Chuo-ku, Kobe 650-0046, Japan
[d] Department of Microbiology, Osaka Prefectural Institute of Public Health, 1-3-69 Nakamichi, Higashinari-ku, Osaka 537-0025, Japan
[e] School of Pharmacy, Hokkaido Pharmaceutical University, 7-15-4-1 Maeda, Teine-ku, Sapporo, Hokkaido 006-8590, Japan
[f] Department of Microbiology, Osaka City Institute of Public Health and Environmental Sciences, 8-34 Tojo-cho, Tennoji-ku, Osaka 543-0026, Japan
[g] Department of International Health, Nagasaki University Graduate School of Biomedical Sciences, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan
[h] Department of Health Science, Fukuoka Institute of Health and Environmental Sciences, 39 Mukaizano, Dazaifu, Fukuoka 818-0135, Japan
[i] Department of Microbiology, Nagasaki Prefectural Institute for Environmental Research and Public Health, 2-1306-11 Ikeda, Ohmura, Nagasaki 856-0026, Japan

ARTICLE INFO

Article history:
Received 23 June 2015
Received in revised form 13 July 2015
Accepted 24 July 2015
Available online 26 July 2015

Keywords:
Phylogeny
Genotype
Maximum *a posteriori*
Mycobacterium tuberculosis
VNTR

ABSTRACT

Intra-species phylogeny of *Mycobacterium tuberculosis* has been regarded as a clue to estimate its potential risk to develop drug-resistance and various epidemiological tendencies. Genotypic characterization of variable number of tandem repeats (VNTR), a standard tool to ascertain transmission routes, has been improving as a public health effort, but determining phylogenetic information from those efforts alone is difficult. We present a platform based on maximum *a posteriori* (MAP) estimation to estimate phylogenetic information for *M. tuberculosis* clinical isolates from individual profiles of VNTR types. This study used 1245 *M. tuberculosis* clinical isolates obtained throughout Japan for construction of an MAP estimation formula. Two MAP estimation formulae, classification of Beijing family and other lineages, and classification of five Beijing sublineages (ST11/26, STK, ST3, and ST25/19 belonging to the ancient Beijing subfamily and modern Beijing subfamily), were created based on 24 loci VNTR ($24_{Beijing}$-VNTR) profiles and phylogenetic information of the isolates. Recursive estimation based on the formulae showed high concordance with their authentic phylogeny by multi-locus sequence typing (MLST) of the isolates. The formulae might further support phylogenetic estimation of the Beijing lineage *M. tuberculosis* from the VNTR genotype with various geographic backgrounds. These results suggest that MAP estimation can function as a reliable probabilistic process to append phylogenetic information to VNTR genotypes of *M. tuberculosis* independently, which might improve the usage of genotyping data for control, understanding, prevention, and treatment of TB.

1. Introduction

Tuberculosis (TB) caused by *Mycobacterium tuberculosis* is an important public global health issue. An estimated 9.0 million people developed TB in 2013; 1.5 million died from the disease according to the World Health Organization (WHO, 2014). Comparative genomic analyses have revealed that clinical strains of *M. tuberculosis* are divisible into four major phylogenetic lineages: Indo-Oceanic, East Asian, East-African – Indian, and Euro-American (Gagneux and Small, 2007). The Beijing family, of East Asian lineage, is predominant in countries of eastern Asia such as China, Korea, and Japan (Hill et al., 2012; Mokrousov, 2013). The phylogenetic family also tends to extend to other areas worldwide such as Russia, South Africa, and southern Asian countries (Cowley et al., 2008; Hanekom et al., 2011; Mokrousov, 2013). Some epidemiological results suggest that the Beijing family possesses more virulent phenotypes than other lineages (non-Beijing lineages), with higher frequency of relapse, transmissibility, and acquisition of drug resistance (Anh et al., 2000; Hanekom et al., 2007; Huyen et al., 2013; Niemann et al., 2010; Wada et al., 2009a).

* Corresponding authors.
E-mail addresses: setoj@pref.yamagata.jp (J. Seto), twada@nagasaki-u.ac.jp (T. Wada).
[1] J.S. and T.W. contributed equally to the work.

http://dx.doi.org/10.1016/j.meegid.2015.07.029
1567-1348/

The Beijing family is divisible into two phylogenetic subgroups, ancient and modern subfamilies, by IS*6110* insertion in Rv3128c, designated as the NTF region (Mokrousov et al., 2005). These two subfamilies are classifiable further into refined types of sequences (STs) using the patterns of 10 single-nucleotide polymorphisms (SNPs) identified by genomic comparison among referential strains (Filliol et al., 2006; Iwamoto et al., 2008; Nakanishi et al., 2013). Such detailed subdivisions reflect the phylogeographical variation of the Beijing family in countries of eastern Asia (Kang et al., 2010; Wada et al., 2009b). The most outstanding features of the genetic population structure are found in Japan, located at the far eastern edge of Asia. In contrast to the worldwide predominance of the modern subfamily (Merker et al., 2015), the ancient subfamily consisting of several STs accounts for the major part of *M. tuberculosis* clinical isolates in the country (Iwamoto et al., 2009; Wada et al., 2009b). Recent epidemiological studies have revealed that Beijing sublineages might be related to some features such as the frequency of multidrug-resistant (MDR), the difference of sublineage distribution in each age group, and higher transmissibility (Iwamoto et al., 2009, 2008; Wada et al., 2009a). The Beijing family might possess high virulence as noted above. In addition, findings related to Beijing sublineages indicate that the phylogenetic subdivision of Beijing family in clinical isolates might be useful to support public health countermeasures in Japan (Iwamoto et al., 2009).

Variable number of tandem repeats (VNTR), a popular genotyping method, can identify clinical strains for investigation of transmission routes among TB patients (De Beer et al., 2014; Niemann et al., 2010; Niobe-Eyangoh et al., 2004; van Deutekom et al., 2005; Wada et al., 2009a). The method relies on polymorphism of the number of repetitive units located at various positions in the *M. tuberculosis* genome. At this time, a subset of 15 loci mycobacterial interspersed repetitive units (MIRU)-VNTR (Supply et al., 2006) has been proposed as an international standard. In spite of its reasonable output for global comparison, the discriminatory power of the subset is reportedly insufficient to discriminate clinical isolates belonging to the Beijing family (Allix-Beguec et al., 2014). To overcome such defects for practical use as a genotyping tool, other subsets of VNTR have been proposed: Japan Anti-Tuberculosis Association (JATA)-(12)-VNTR, comprising 8 loci in 15-MIRU-VNTR and 4 additional loci (Maeda et al., 2008; Murase et al., 2008); $24_{Beijing}$-VNTR, comprising 15-MIRU-VNTR and 9 additional loci (Iwamoto et al., 2012); and 28 loci newly VNTR, comprising 24-MIRU-VNTR and 4 hypervariable loci (Allix-Beguec et al., 2014).

The VNTR profiles of *M. tuberculosis* not only provide genotypic discrimination of clinical isolates; they also correspond with phylogenetic classification in clustering analysis. Supply et al. reported that *M. tuberculosis* lineages are distinguished by a minimum spanning tree (MST) of 24-MIRU-VNTR profiles of cosmopolitan isolates (Supply et al., 2006). Moreover, previous reports describe that VNTR clustering of detailed sublineages of the Beijing family was observed according to the phylogenetic classification (Kang et al., 2010; Maeda et al., 2010; Merker et al., 2015). These findings indicate that the phylogeny of *M. tuberculosis* can be estimated properly from VNTR genotypes that were originally improved to trace transmission routes.

This study investigated a statistical pipeline to assign phylogenetic classification based on the VNTR genotypes. A simple statistical method introduced herein, maximum *a posteriori* (MAP) estimation, can estimate the *M. tuberculosis* phylogeny of clinical isolates solely from their VNTR profiles, especially for the Beijing family. The platform will enhance information derived from genotypic data of clinical strains, which might better elucidate the state of *M. tuberculosis* prevalence in the area.

2. Materials and methods

2.1. Bacterial strains

For this study, 1245 clinical isolates from seven surveillance studies were screened: Yamagata Pref. (339 isolates obtained during 2009–2013); Tokyo metropolitan area (230 isolates obtained during 2004–2006); a southern part of Osaka Pref. (76 isolates obtained during 2009–2010); Osaka City (98 isolates obtained from a high incidence area during 2009–2010); Kobe City (221 isolates obtained in 2009); Fukuoka Pref. (66 isolates obtained in 2012); and Okinawa Pref. (215 isolates obtained during January 2005 – March 2007). They were collected systematically from their respective areas by the local government institutes for public health. The isolate collections were partly correspondent with those described in our previous works (Seto et al., 2013; Wada et al., 2015; Wada and Maeda, 2013). The cultured bacilli were suspended and boiled in H_2O. Then supernatants of the heat-killed suspensions were used as PCR templates for genotyping.

2.2. VNTR analysis

Genotypic data of $24_{Beijing}$-VNTR (Iwamoto et al., 2012) of all isolates were collected for this study. The subset includes two other subsets of VNTR: 15-MIRU-VNTR (Supply et al., 2006) and JATA-(12)-VNTR (Maeda et al., 2008; Murase et al., 2008) (Table S1). All profiles are presented in Table S2. Their numbers of repeats for each locus were calculated from the sizes of PCR products in agreement with published allelic tables (Iwamoto et al., 2007). The amplified PCR fragment sizes were confirmed using various electrophoresis methods along with equipment facilities of the respective institutes: agarose gel electrophoresis, a capillary electrophoresis system (SV1210; Hitachi High-Technologies Corp., Tokyo, Japan), or a genetic analyzer (ABI Prism 310; Applied Biosystems, Foster City, CA) (Iwamoto et al., 2007; Wada et al., 2007). All institutes had shared common DNA templates of *M. tuberculosis* clinical strains for mutual quality control to calculate the exact alleles.

2.3. Clustering analysis

Clustering analyses of $24_{Beijing}$-VNTR genotypes of *M. tuberculosis* isolates were conducted with the MST algorithm using software (Bionumerics ver. 4.6; Applied Maths Inc., Saint-Martens-Latem, Belgium). To simplify the calculations, 44 of 1245 genotypes including multiple repeat numbers in the 24 loci were omitted. Consequently, 1201 genotypes were subjected to clustering analyses. The reconstruction rules were the following. A categorical coefficient was selected. The priority rule was set such that the type with the highest number of single-locus variants and double-locus variants would be linked first. Creation of hypothetical types was not allowed.

2.4. Phylogenetic classification

Classification between the Beijing family and non-Beijing lineages was determined based on at least one genotyping method: spoligotyping patterns (van Soolingen et al., 1995), RD207 deletion (Hillemann et al., 2006; Warren et al., 2004), and a definitive nucleotide substitution in Rv0679c (Nakajima et al., 2013), according to each report. Isolates belonging to the Beijing family were classified further into STs defined by Filliol's 10 SNPs (Filliol et al., 2006) with modification to avoid mis-assignments caused by ambiguous homoplasy (Nakanishi et al., 2013; Wada et al., 2012). According to previous studies, STs of the Beijing family were