Infection, Genetics and Evolution 43 (2016) 15–21

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Research paper

MIRU-VNTR genotype diversity and indications of homoplasy in *M. avium* strains isolated from humans and slaughter pigs in Latvia

Adrija Kalvisa [a,b,1], Constantinos Tsirogiannis [c], Ivars Silamikelis [a], Girts Skenders [d], Lonija Broka [d], Agris Zirnitis [e], Inta Jansone [a], Renate Ranka [a,b,*]

[a] Latvian Biomedical Research and Study Centre (LV BMC), Riga, Latvia
[b] Riga Stradins University (RSU), Riga, Latvia
[c] Center for Massive Data Algorithmics (MADALGO), Aarhus University, Aarhus, Denmark
[d] Riga East University Hospital, Tuberculosis and Lung Diseases Center, Latvia
[e] Department of Veterinary Medicine, Latvia University of Agriculture, Jelgava, Latvia

ARTICLE INFO

Article history:
Received 23 December 2015
Received in revised form 6 May 2016
Accepted 8 May 2016
Available online 11 May 2016

Keywords:
Non-tuberculous mycobacteria
MIRU-VNTR
IS1245 RFLP
Homoplasy

ABSTRACT

Diseases which are caused by non-tuberculous mycobacteria (NTM) are an increasing problem in the developed countries. In Latvia, one of the most clinically important members of NTM is *Mycobacterium avium* (*M. avium*), an opportunistic pathogen which has been isolated from several lung disease patients and tissue samples of slaughter pigs. This study was designed to characterize the genetic diversity of the *M. avium* isolates in Latvia and to compare the distribution of genotypic patterns among humans and pigs. Eleven (Hall and Salipante, 2010) clinical *M. avium* samples, isolated from patients of Center of Tuberculosis and Lung Diseases (years 2003–2010), and 32 isolates from pig necrotic mesenterial lymph nodes in different regions (years 2003–2007) were analyzed. The majority (42 of 43) of samples were identified as *M. avium* subsp. *hominissuis*; one porcine isolate belonged to *M. avium* subsp. *avium*. MIRU-VNTR genotyping revealed 13 distinct genotypes, among which nine genotype patterns, including *M. avium* subsp. *avium* isolate, were newly identified. IS1245 RFLP fingerprinting of 25 *M. avium* subsp. *hominissuis* samples yielded 17 different IS1245 RFLP patterns, allowing an efficient discrimination of isolates. Clusters of identical RFLP profiles were observed within host species, geographical locations and time frame of several years. Additional in silico analysis on simulated MIRU-VNTR genotype population datasets showed that the MIRU-VNTR pattern similarity could partly arise due to probabilistic increase of acquiring homoplasy among subpopulations, thus the similar MIRU-VNTR profiles of *M. avium* strains even in close geographical proximity should be interpreted with caution.

1. Introduction

Non-tuberculous mycobacteria (NTM) increase the infection burden, targeting the elderly population and the immunocompromised individuals (Cook, 2010; Karakousis et al., 2004; Sexton and Harrison, 2008; Weiss and Glassroth, 2012). NTM diseases are diverse and hard to treat due to their innate drug resistance to major antimycobacterial drugs (van Ingen et al., 2012). *Mycobacterium avium* (*M. avium*) is one of the clinically important members of NTM — it is one of the most abundant among clinical isolates (Cook, 2010; Thomson and Yew, 2009; Kendall and Winthrop, 2013; Uchiya et al., 2013). The genotypic analysis of *M. avium* presents several difficulties: *M. avium* is genetically monomorphic when compared to other bacteria; *M. avium* has few markers applicable for detecting of genetic variability; and *M. avium* grows slowly, making it difficult to obtain sufficient amounts of DNA for genotyping in clinical settings (Achtman, 2008; Falkinham, 2002; Grant et al., 2003). Even so, repeats-based approaches, such as eight loci Mycobacterial Interspersed Repetitive Units-Variable Number of Tandem Repeats (MIRU-VNTR) are useful for investigating the genetic diversity of these microorganisms (Thibault et al., 2008; Thibault et al., 2007). In recent years, multilocus sequence typing analysis (MLST) was proposed as definite and well-discriminatory method for the characterization of *M. avium* strains (Kolb et al., 2014; Kim et al., 2016; Vluggen et al., 2016). However, the analysis of VNTR loci is a simple and reliable PCR-based technique, allowing a numerical and reproductive digitalization of typing data (Leão et al., 2014). A large collection of *M. avium* genotype profiles of this type have been published, and a database of MIRU-VNTR genotypes is available (Thibault et al., 2007).

High level of size homoplasy remains the major drawback for using repetitive sequences solely as markers for epidemiological studies

* Corresponding author at: Latvian Biomedical Research and Study Centre, Ratsupites Str. 1, Riga LV-1067, Latvia.
E-mail address: renate_r@biomed.lu.lv (R. Ranka).
[1] Present address: Department of Biochemistry and Molecular Biology, Southern Denmark University, Odense, Denmark.

http://dx.doi.org/10.1016/j.meegid.2016.05.013
1567-1348/

(Estoup et al., 2002). Two genotypes are considered homoplasic if they are identical by state but not by descent. Homoplasic profiles cannot be distinguished as coming from different ancestral lineages, and result in assuming a false connection between isolates. Some studies showed that including extra genotyping loci improve the discrimination when using repeats-based genotyping methods (Wei et al., 2015), especially in genotyping of human pathogen *Mycobacterium tuberculosis* (Christianson et al., 2010; Alonso-Rodríguez et al., 2008). However, *M. avium* does not follow the human-to-human infection route: *M. avium* as constitutive pathogen is a subject to wider variety of unknown selection pressure factors (Falkinham, 2002). Therefore, we argue that increasing the number of loci may not reduce the number of homoplasic genotypes among closely related subpopulations of *M. avium*. Surprisingly, until now there has been no investigations how the level of relatedness between individuals may influence the number of homoplasic genotypes that appear in *M. avium* subpopulation descending from the same ancestral lineage.

In this study, (Achtman, 2008) we provide for the first time the analysis of genetic diversity of the *M. avium* isolates in Latvia from human and porcine subjects. As the output of these method may contain many homoplasic genotypes, (Alonso-Rodríguez et al., 2008) we further exploited a mathematical modelling to evaluate the probability of homoplasy. For this purpose, genotypes made by simulated phylogenies that represented the two-step evolution process of repetitive sequence-based molecular markers were used. A new analytical tool – mean phylogenetic distance – was used to measure the level of homoplasy between these genotypes. Based on this, we introduce an intuitive approach to evaluate whether there is a difference in probability of developing homoplasic genotypes in varying levels of relatedness within different subpopulations mimicking MIRU-VNTR genotypes.

2. Materials and methods

2.1. Mycobacterial isolates

In total, forty-three (43) human and pig *M. avium* samples were analyzed in this study. Clinical human *M. avium* samples were isolated from patients of Center of Tuberculosis and Lung Diseases, Riga East University Hospital (years 2003–2010). Only culture positive isolates were included. The exclusion criterion was the presence of mixed infection. In total, eleven (Hall and Salipante, 2010) clinical *M. avium* samples from the same number of patients were available for this study. Thirty-two (Tsirogiannis et al., 2012) porcine *M. avium* samples were isolated from pig necrotic mesenterial lymph nodes in seven different regions of Latvia (Dobele, 4002; Jēkabpils, 1100; Rēzekne, 2100; Rīga, 0100; Talsi, 8802; Tukums, 9002; Valmiera, 2500) in years 2003–2007 (Table S1).

2.2. Genotyping

The *M. avium* subspecies were detected by sequencing the *RpoB* gene fragment as described by Ben Salah et al. (2008). Further, all the isolates were genotyped by the 8-loci MIRU-VNTR method as described by Thibault et al. (2007).

Further, 25 porcine *M. avium* subsp. *hominissuis* isolates were additionally analyzed using the IS1245 RFLP fingerprint method (van Soolingen et al., 1998). The IS1245 RFLP was applied only on a part of samples because it was impossible to meet the requirements of DNA amount and quality for IS1245 RFLP among the rest of the samples.

Phylogenetic analysis of IS1245 RFLP patterns, as well as of combined MIRU-VNTR and IS1245 RFLP patterns, was performed on Manhattan distance matrix by using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) method. The alignment and minimum spanning tree was created using BioNumerics v5.10 (Applied Maths, Austin, TX).

2.3. Simulating clonal expansion of MIRU-VNTR genotypes

Due to the existence of homoplasy, bacteria that were observed to have the same genotype may not be in fact clonally related. This is especially the case for the *Mycobacterium* species, which are considered to be highly clonal (Supply et al., 2003). In this study, a probability of homoplasy was evaluated by a simulated model of clonal expansion of phylogenetic trees (i.e. clonal expansion trees). Mathematical model based on the general laws of VNTR genotype changes, which are discussed by Vogler et al. (2006, 2007) and implemented by Hall and Salipante (2010), was used. This model produced a binary phylogenetic tree where a genotype is assigned to each node. The leaves of the tree represented the final (i.e. present) population. Each of these leaves is considered as unique by descent. As a consequence, each pair of leaves containing identical genotypes was considered as homoplasic.

The construction of a clonal expansion tree consisted of two stages. In the first stage, the underlying tree topology was constructed (i.e. the tree nodes and how these nodes are linked to each other). This was done using a randomised process that created a binary tree. The average branch length, that is the average weight value of the links that connect nodes in the resulting tree, was two units. In the second stage, genotypes were assigned to the nodes of the constructed tree. These simulated genotypes resembled the ones that we acquired for the forty-three experimentally analyzed *M. avium* isolates by using the MIRU-VNTR method. Each simulated genotype consisted of k loci, and each locus stored an integer value, the allele. The integer value on each locus could be any number from the set $S = (0,1,2, \ldots,12)$, the values that were also observed for the *M. avium* isolates.

The genotypes were assigned to the nodes of the clonal expansion tree in the following way: initially, we set manually the genotype at the root node of the tree. Then, for every child node v of the root the genotype of v was created by copying the genotype of the root and then changing the values of one or more loci. The loci whose values got changed were selected by a randomised process, and the initial value of the locus was incremented or decremented by one unit. In any case, the new value should always be one of the values in the set S. If the initial value of a locus was “0” then no change could happen to this locus; this value was called the absorbing allele based on previous observations on *M. avium* genotypes. After computing the genotypes for the child nodes of the root, the same process was repeated recursively for the rest of the nodes of the tree, until all the nodes were assigned with a genotype.

By this model, several clonal expansion trees were constructed, each tree having exactly 500 leaf nodes. More specifically, we produced five groups of trees to simulate five different MIRU-VNRT loci numbers ($k = 8, 10, 12, 14, 16$); each group consisting of 1000 trees. We chose the minimum number of loci to be equal to eight to be comparable to the eight-loci MIRU-VNTR set described by Thibault et al. (2007). We set the maximum number of loci to 16 to compare with the sixteen-loci MATR-VNTR set used by Inagaki et al. (2009). For all the clonal expansion trees we set every locus in the genotype of the root to be equal to “2”, because two is the minimum amount of repeats that are necessary for generating variance in a locus (Supply et al., 2000).

2.4. The measurement of relatedness between homoplasic individuals in simulated data

To analyse the structure of simulated clonal expansion trees, we used two parameters: the mean phylogenetic distance (MPD) and the probability of homoplasy (PH). In this section, we describe in detail these two concepts.

For any pair of leaves in a clonal expansion tree, we defined the phylogenetic distance as the sum of the branch lengths that appears on the path that connects these two leaves. This parameter describes the level of relatedness between two individuals in a simulated population. For any subset S of leaves in a clonal expansion tree, MPD is equal to the