Infection, Genetics and Evolution 11 (2011) 1164–1167

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Short communication

How dormant is *Mycobacterium tuberculosis* during latency? A study integrating genomics and molecular epidemiology

Zhenhua Yang [a,*], Mariana Rosenthal [a], Noah A. Rosenberg [b,c,d], Sarah Talarico [a,1], Lixin Zhang [a], Carl Marrs [a], Vibeke Østergaard Thomsen [e], Troels Lillebaek [f], Aase B. Andersen [g]

[a] *Department of Epidemiology, School of Public Health, University of Michigan, M5124 SPH II, 1415 Washington Heights, University of Michigan, Ann Arbor, MI 48109, USA*
[b] *Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA*
[c] *Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA*
[d] *Life Sciences Institute, University of Michigan, Ann Arbor, MI 48109, USA*
[e] *International Reference Laboratory of Mycobacteriology, Statens Serum Institut, 5, Artillerivej, DK-2300 Copenhagen S, Denmark*
[f] *Department of Infectious Diseases, Copenhagen University Hospital, 9 Blegdamsvej, DK-2100 Copenhagen Ø, Denmark*
[g] *Odense University Hospital, 29 S. Boulevard, DK-5000 Odense C, Denmark*

ARTICLE INFO

Article history:
Received 8 November 2010
Received in revised form 2 February 2011
Accepted 3 February 2011
Available online 18 February 2011

Keywords:
Latent tuberculosis infection
Genomics
Molecular epidemiology

ABSTRACT

Mycobacterium tuberculosis may survive for decades in the human body in a state termed latent tuberculosis infection (LTBI). We investigated the occurrence during LTBI of insertion/deletion events in a selected set of mononucleotide simple sequence repeats, DNA sequence changes in four *M. tuberculosis* genes, and large sequence variations in 4750 *M. tuberculosis* open reading frames. We studied 13 paired *M. tuberculosis* clinical isolates, with each pair representing a reactivation of LTBI more than three decades after primary infection. Absence of sequence variations between paired isolates in nearly all investigated loci suggests a low likelihood of bacterial replication during LTBI.

1. Introduction

Tuberculosis (TB) remains a leading infectious cause of mortality and morbidity worldwide. As one-third of the world's population is estimated to be latently infected with *Mycobacterium tuberculosis*, people with latent tuberculosis infection (LTBI) represent a significant reservoir for future disease reactivation.

Molecular epidemiological studies have provided evidence of endogenous reactivation of *M. tuberculosis* after more than three decades of latent infection (Lillebaek et al., 2002). For immunocompetent individuals, the lifetime risk of reactivation is estimated to be ~10% on average; however, for persons co-infected with HIV, the risk is much higher, ~10% per year (Nahid and Daley, 2006). With the increasing rate of HIV and *M. tuberculosis* co-infection in many countries where TB is prevalent, development of methods for curing LTBI and preventing LTBI from reactivation remains a major challenge to global TB control. The importance of studying LTBI is particularly great in light of the recent global spread of multi-drug resistant (MDR) and extensively drug resistant (XDR) TB infection. A proportion of the people newly infected with *M. tuberculosis* potentially have latent TB infection caused by MDR or XDR strains, representing both a near- and a long-term challenge to global control of TB.

The development of effective drugs for eradication of LTBI requires a better understanding of the biological basis of LTBI. Previous studies have generated two alternative hypotheses about the modes for *M. tuberculosis* persistence in humans. One hypothesis is that during latency, *M. tuberculosis* enters a very slow replicating or non-replicating dormant state in which the bacilli are insensitive to killing by the host immune system and anti-TB drugs. The other hypothesis is that during latency, *M. tuberculosis* replicates but is killed by the host immune system at a rate roughly equal to its replication rate. The former hypothesis was generated by studies using the Cornell mouse model (McCune et al., 1966) and its variants (Scanga et al., 1999), and also by the Wayne *in vitro* stationary phase culture model of latency (Wayne, 1977). It has been further supported by more recent studies using colony-forming unit (CFU) counting and quantitative real-time PCR to monitor the dynamics of *M. tuberculosis* in chronically infected mice (Munoz-Elias et al., 2005). However, despite the widespread notion that LTBI is in a dormant static equilibrium, with very slow or no replication, other evidence has supported the alternative hypothesis of substantial continued replication *in vivo* (Gill et al., 2009). Distinguishing between these different modes of

---

[*] Corresponding author. Tel.: +1 734 763 4296; fax: +1 734 764 3192.
E-mail address: zhenhua@umich.edu (Z. Yang).
[1] Present address: Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington.

1567-1348/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2011.02.002

M. tuberculosis persistence would provide insight into the ability of different strategies to successfully eliminate the latent bacteria, because drugs targeting actively replicating bacterial cells would have limited effect on bacterial cells in static phase. In this study, we took a novel approach that integrates molecular epidemiology with genomics to gain insight into the biological basis of human LTBI.

2. Materials and methods

We investigated 13 *M. tuberculosis* isolates from the 1960s and 13 isolates from 1990s, respectively, all sampled in Denmark. The patients in the 1990s were most likely infected during the 1960s. Subsequently, they developed TB due to reactivation of *M. tuberculosis* after three decades of latent infection. The transmission between these 'modern' and 'historic' cases was confirmed based on matching molecular typing patterns of their *M. tuberculosis* isolates and epidemiological investigation, as previously reported (Lillebaek et al., 2002, 2003). For each of the 13 reactivation cases, the estimated duration of LTBI was between 30 and 39 years investigation. The study was approved by the Health Sciences and Behavioral Sciences Institutional Review Board of the University of Michigan, the Danish National Ethical Committee, and the Danish Data Protection Agency.

To assist in understanding the effect of host immune selective pressure on latency and reactivation, we sought information about the patients' HIV status. Because HIV was not known in the 1960s, it is likely that the first set of patients were HIV sero-negative. Although TB patients were not routinely screened for HIV infection during the 1990s, HIV infection suspects were tested for HIV sero positivity, and none of the study patients from the 1990s was known to be HIV sero-positive. They had no clinical signs of immune deficiency, and their mean age at diagnosis was 62.9 years (Lillebaek et al., 2003). Further, in Denmark, only 2% of TB patients are known to be HIV sero-positive, and HIV sero-positive TB patients are mostly younger persons from high-incidence areas in Africa or risk groups in Denmark, rather than older Danes.

Genomic DNA was extracted from Lowenstein–Jensen cultures of the study isolates using standard procedures (Murray and Thompson, 1980). We investigated the occurrence of insertion/deletion (indel) events in a selected set of mononucleotide simple sequence repeats (SSRs). SSRs, also known as microsatellites, are short DNA sequence stretches in which motifs of one to six bases are tandemly repeated. SSRs gain or lose repeats due to DNA replication slippage, which occurs at very high rates compared to other types of mutations (Ellegren, 2004). Sreenu et al. (2007) reported that SSRs are distributed throughout mycobacterial genomes at an average rate of 220–230 SSR tracts per kb. Because mycobacteria lack the post-replicative DNA mismatch repair system that would normally correct for strand slippage errors during replication (Cole et al., 1998), the occurrence of indels in their SSRs may provide insight into the question of whether or not *M. tuberculosis* replicates during latency.

Ten intergenic regions, likely subject to less selective pressure than coding regions, were PCR-amplified and DNA sequenced in order to detect indels among their mononucleotide SSRs (Table 1). A total of 752 mononucleotide SSRs were present in these ten regions: 1(A)7, 2(A)6, 4(A)5, 7(A)4, 25(A)3, 108(A)2, 1(T)7, 4(T)4, 22(T)3, 87(T)2, 1(G)7, 3(G)6, 5(G)5, 14(G)4, 41(G)3, 171(G)2, 1(C)9, 5(C)6, 4(C)5, 14(C)4, 46(C)3, and 186(C)2. In this list, the notation X(Y)Z indicates that among the amplified regions, there are X non-overlapping appearances of the sequence 'Y...Y', where Y is repeated *Z* times. Isolates were PCR-amplified using the BD Advantage^{TM} GC 2 PCR kit (BD Biosciences Clontech, Palo Alto, CA). Each standard 50 μl reaction consisted of 10 μl of 5X reaction buffer, 5 μl of GC Melt, 2 μl of each primer (at 10 μM each), 1 μl of 50× deoxyribonucleoside triphosphate mixture, 1 μl of 50× BD Advantage^{TM} 2 Polymerase mixture, 2 μl (20 ng) of DNA template, and 27 μl of PCR-grade water. The same thermocycling program was used for all PCR amplifications. This program included: 94 °C (3 min), 30 cycles of 94 °C (30 s), 64 °C (30 s), and 72 °C (1 min), and a final cycle of 72 °C (10 min). The PCR products were examined by 1.0% (w/v) agarose gel electrophoresis performed in 1× Tris–borate–EDTA buffer to determine their size. PCR products were purified for DNA sequencing using QIAquick® PCR Purification kit following the manufacturer's instructions (QIAGEN, Inc., Valencia, CA).

We also investigated DNA sequence changes in four *M. tuberculosis* genes by PCR and DNA sequencing, as described previously (Hebert et al., 2007; Talarico et al., 2007). These four genes included genes coding for three proposed antigenic proteins,

Table 1

Genomic locations and lengths of the ten intergenic regions of the *M. tuberculosis* H37Rv strain selected for SSR polymorphism investigation and the primers for their PCR amplification and DNA sequencing.

Investigated non-coding regions (NCR)	NCR length (bp)	Primer names[a]	Primer sequences	Product length (bp)
854156–854264	109	pks5_papA4_F	agctaccggcgtaacacgtgtcc (23 bp)	601
		pks5_papA4_R	ttgatccgtccagtgaaacctgca (24 bp)	
1728408–1728950	543	Rv2293c_Rv2294_F	agacagtgccgcaaaggcg (19 bp)	364
		Rv2293c_Rv2294_R	cgcagttgctcgagcgttagc (21 bp)	
2565031–2565324	294	Rv3660c_Rv3661_F	gatcggtcagcatcgccaacac (22 bp)	394
		Rv3660c_Rv3661_R	attgcgtccggtacttgcgtgac (23 bp)	
4099146–4099643	498	Rv0759cRv0760c_F	accggactggggttgtcgatcgtcg (25 bp)	1120
		Rv0759cRv0760c_R	acatggtgaccactgcgtacaagctggagc (30 bp)	
2634097–2634525	429	PPE38_PPE39_F	tgatctccggcggcaaccacga (22 bp)	513
		PPE38_PPE39_R	tcagcggcaatgggctttcg (20 bp)	
4340025–4340266	242	Rv3863_Rv3864_F	tccgtcagatcaattgaggtcg (22 bp)	290
		Rv3863_Rv3864_R	tcgtcttgcaaagaccgcta (20 bp)	
53243–53660	418	Rv0049_ponA_F	tggtcaagtcatacgtcctgg (21 bp)	572
		Rv0049_ponA_R	acgatcaggtaggccatcgtgaa (23 bp)	
1914875–1915524	650	tyrS_IprJ_F	cgctggttagtgctacgtcgtgga (24 bp)	656
		tyrS_IprJ_R	tggccaacaagagtcacgttcac (23 bp)	
4336079–4336773	695	gltB_Rv3860_F	cttaggcgtcatacccacctaacc (24 bp)	695
		gltB_Rv3860_R	tgacaggcgatcgattgtcg (20 bp)	
809644–809943	300	Rv0712_Rv0713_F	acgatcagccagaccctcaagg (22 bp)	478
		Rv0712_Rv0713_R	cacctgttggtgggtcgcta (20 bp)	

[a] The custom primers were made by Invitrogen (Invitrogen, Inc., Carslbad, CA) and are designated by the names of the genes upstream and downstream from the investigated non-coding genomic regions. The letter 'F' stands for forward and the letter 'R' stands for reverse.

PE_PGRS26, PE_PGRS33, and PPE18, and the *rpf*B gene encoding a proposed resuscitation-promoting factor protein that has a role in the reactivation of LTBI (Mukamolova et al., 2002).

High numbers of genetic variants in the PE_PGRS33 and PE_PGRS26 genes have been found in the natural population of *M. tuberculosis* isolates; such allelic diversity among *M. tuberculosis* isolates indicates that these genes could serve as a source of antigenic variation for the pathogen and might have clinical and epidemiological consequences (Talarico et al., 2007).

The genomic variability of the PPE gene family and its possible role as a major source of antigenic variation has also been well-documented. PPE18 is a component of the new recombinant subunit tuberculosis vaccine Mtb72F. Substantial numbers of polymorphisms have been found in this gene among clinical isolates (Hebert et al., 2007).

In addition to the SSR and genic analyses, we also considered genome-wide differences in a specific pair of isolates. One pair among the 13 pairs of isolates was chosen for a microarray-based comparative genomic hybridization to detect the differences in genomic content between the two isolates. The pair of isolates was obtained from a father and a son, in the 1960s and 1990s, respectively, representing reactivation after three decades of latency (Lillebaek et al., 2002). The competitive hybridization of genomic DNA samples from the two isolates labeled with different fluorescence dyes was performed using a *M. tuberculosis* microarray (obtained from NIAID Pathogen Functional Genomics Resource Center). A total of 4750 oligonucleotide probes were on the microarray, representing 4127 open reading frames (ORFs) from the genome of *M. tuberculosis* strain H37Rv and 623 unique ORFs from *M. tuberculosis* strain CDC1551.

3. Results

In the SSR analysis of the 13 pairs of clinical isolates of *M. tuberculosis* representing the same strains before and after more than three decades of LTBI, while different pairs of isolates had different tract lengths of mononucleotide SSRs, of the 752 SSRs analyzed, there was only one SSR [(C) 9] in which an insertion was found in one of the 13 reactivation isolates, compared with its paired epidemiologically linked 'historic' isolate. No deletions were observed in the reactivation isolates.

The investigation of DNA sequence changes in the four *M. tuberculosis* genes produced a similar result. For 12 of the 13 pairs of isolates, we detected no DNA sequence differences in the four genes and their adjacent regions between the two isolates of the pair. However, in the same isolate that was found to have an insertion in SSR [(C) 9], a 9 bp insertion and a synonymous single-base pair change were found in the PE_PGRS33 gene. None of the 4750 oligonucleotide probes on the microarray slide showed differential hybridization to the genomic DNA samples of the father–son isolate pair.

4. Discussion

In this study, we attempted to address a long-standing question about whether *M. tuberculosis* replicates during human LTBI. We used an innovative approach that integrates genomics with molecular epidemiological data. Our unique collection of *M. tuberculosis* clinical isolates, well-characterized by population-based molecular epidemiological studies (Lillebaek et al., 2002, 2003), provided a unique opportunity for studying LTBI using a genomics approach. This study represents the first exploration of *M. tuberculosis* genomic content changes during human LTBI spanning three decades or more. Our results that almost no DNA sequence changes were detected during more than three decades of LTBI provide strong suggestive evidence that LTBI involves relatively little replication.

Despite the well-documented DNA polymorphisms in the PE_PGRS33, PE_PGRS26, and PPE18 genes among clinical isolates, we found no difference from the historic isolates in 12 of the 13 reactivation cases. This finding is intriguing as genes hypothesized to be involved in host–pathogen interaction are expected to be under more host selection pressure and therefore are more likely to undergo genetic changes during LTBI. The absence of DNA sequence changes in the PE_PGRS and PPE18 genes may suggest that limited bacterial replication occurred or that the pathogen has little interaction with the host immune system during LTBI, thereby supporting the hypothesis that *M. tuberculosis* remains truly dormant during LTBI. This possibility is further supported by our finding of relatively few changes in the 752 mononucleotide SSRs analyzed, despite the fact that mononucleotide SSRs are known to be less stable than other types of SSRs (Moxon et al., 2006).

We note, however, that little is known about baseline mutation rates in the *M. tuberculosis* loci that we have investigated. Thus, due to the absence of data from serial subcultures of *M. tuberculosis* that would allow us to define the relationship between replication rates of *M. tuberculosis* and genomic sequence changes, our study cannot provide a formal quantitative argument for differentiating the predictions of the two hypotheses. Nevertheless, the view that replication is rare during LTBI is supported by the observation in a previous investigation of identical restriction fragment length polymorphism patterns for the rapidly evolving IS6110 marker between each pair among the study isolates (Lillebaek et al., 2002, 2003).

Another limitation of our study is the under-sampling of the genome of the study isolates. In each isolate pair, we investigated 14 genomic regions that included a total of 12,355 base pairs/sites at which mutation could have occurred, and considering all isolate pairs, we observed only one SSR change and a 9 bp insertion and a synonymous single-base pair change in the PE_PGRS33 gene. Because the total sample of sequences for each pair only accounted for approximately 0.3% of the total *M. tuberculosis* genome, it is possible that mutation patterns in this portion of the genome are not representative. A genome-wide, high-fidelity sequencing approach to uncovering genomic sequence changes during human LTBI will not only help to address our current study question about the occurrence of replication during LTBI, it will also set the stage for future work on LTBI pathogenesis and *M. tuberculosis* virulence factor discovery, thereby contributing to the ultimate development of new vaccines and drugs for tuberculosis control.

Conflict of interest

No conflict of interest related to this article for all authors.

Acknowledgments

This study was supported in part by the National Institutes of Health grant NIH-R01-AI151975 and in part by a research grant from the University of Michigan. Mariana Rosenthal was supported by the Rackham Merit Fellowship, University of Michigan. The authors thank Pia Kristiansen for her assistance in recovering the study isolates and Dong Yang for her assistance in the maintenance of the strain collection used for the study.

References

Cole, S.T., Brosch, R., Parkhill, J., et al., 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393, 537–544.

Ellegren, H., 2004. Microsatellites: simple sequences with complex evolution. Nat. Rev. Genet. 5, 435–445.

Gill, W.P., Harik, N.S., Whiddon, M.R., Liao, R.P., Mittler, J.E., Sherman, D.R., 2009. A replication clock for *Mycobacterium tuberculosis*. Nat. Med. 15, 211–214.

Hebert, A.M., Talarico, S., Yang, D., et al., 2007. DNA polymorphisms in the pepA and PPE18 genes among clinical strains of *Mycobacterium tuberculosis*: implications for vaccine efficacy. Infect. Immun. 75, 5798–5805.

Lillebaek, T., Dirksen, A., Baess, I., Strunge, B., Thomsen, V.O., Andersen, A.B., 2002. Molecular evidence of endogenous reactivation of *Mycobacterium tuberculosis* after 33 years of latent infection. J. Infect. Dis. 185, 401–404.

Lillebaek, T., Dirksen, A., Vynnycky, E., Baess, I., Thomsen, V.O., Andersen, A.B., 2003. Stability of DNA patterns and evidence of *Mycobacterium tuberculosis* reactivation occurring decades after the initial infection. J. Infect. Dis. 188, 1032–1039.

McCune, R.M., Feldmann, F.M., Lambert, H.P., McDermott, W., 1966. Microbial persistence I. The capacity of *tubercle bacilli* to survive sterilization in mouse tissues. J. Exp. Med. 123, 445–468.

Moxon, R., Bayliss, C., Hood, D., 2006. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. Annu. Rev. Genet. 40, 307–333.

Munoz-Elias, E.J., Timm, J., Botha, T., Chan, W.T., Gomez, J.E., McKinney, J.D., 2005. Replication dynamics of *Mycobacterium tuberculosis* in chronically infected mice. Infect. Immun. 73, 546–551.

Mukamolova, G.V., Turapov, O.A., Young, D.I., Kaprelyants, A.S., Kell, D.B., Young, M., 2002. A family of autocrine growth factors in *Mycobacterium tuberculosis*. Mol. Microbiol. 46, 623–635.

Murray, M.G., Thompson, W.F., 1980. Rapid isolation of high molecular-weight plant DNA. Nucleic Acids Res. 8, 4321–4325.

Nahid, P., Daley, C.L., 2006. Prevention of tuberculosis in HIV-infected patients. Curr. Opin. Infect. Dis. 19, 189–193.

Scanga, C.A., Mohan, V.P., Joseph, H., Yu, K., Chan, J., Flynn, J.L., 1999. Reactivation of latent tuberculosis: variations on the Cornell Murine model. Infect. Immun. 67, 4531–4538.

Sreenu, V.B., Kumar, P., Nagaraju, J., Nagarajam, H.A., 2007. Simple sequence repeats in mycobacterial genomes. J. Biosci. 32, 3–15.

Talarico, S., Cave, M.D., Foxman, B., et al., 2007. Association of *Mycobacterium tuberculosis* PE PGRS33 polymorphism with clinical and epidemiological characteristics. Tuberculosis 87, 338–346.

Wayne, L.G., 1977. Synchronized replication of *Mycobacterium tuberculosis*. Infect. Immun. 17, 528–530.