Tuberculosis 115 (2019) 121–125

Contents lists available at ScienceDirect

Tuberculosis

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

Detection of tuberculosis laboratory cross-contamination using whole-genome sequencing

Jie Wu[a,f,*], Chongguang Yang[b,c], Liping Lu[d], Wanqin Dai[e]

[a] Department of Hospital Infection Control, Shanghai First Maternity and Infant Hospital, Tongji University School of Medicine, Shanghai, China
[b] Key Laboratory of Medical Molecular Virology, Ministry of Education and Health, School of Basic Medical Sciences, Fudan University, Shanghai, China
[c] Department of Epidemiology of Microbial Diseases, Yale School of Public Health, Yale University, New Haven, CT, USA
[d] Department of Tuberculosis Control, Songjiang District Center for Disease Control and Prevention, Shanghai, China
[e] Clinical Laboratory, Shanghai Songjiang District Central Hospital, Shanghai, China
[f] Department of Tuberculosis Control, Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China

ARTICLE INFO

Keywords:
Mycobacterium tuberculosis
Cross-contamination
Whole-genome sequencing

ABSTRACT

Objective: Detection of tuberculosis laboratory cross-contamination using whole-genome sequencing.
Methods: A total of 22 *M. tuberculosis* strains with high genotypic homology from one hospital were collected during the drug resistance surveillance. Genome sequencing and epidemiological investigation were conducted to determine the occurrence of cross-contamination.
Results: The pair wise comparison between the genomes in each cluster indicated that 15 (71.4%) of 21 strains with available genomic data had no SNP differences with at least one other strain within the same cluster. The analysis of the specimen collection time found that, among the 16 strains collected on the same day, 14 (87.5%) of them had no SNP differences with one another strain; meanwhile, among the strains within the same cluster whose SNP distance was 0, 93.3% (14/15) of them had the same collection time, suggesting that these findings were most likely caused by cross contamination.
Conclusion: A high proportion of *M. tuberculosis* strains with genotypic homology from the single institute that shared the same process time period was most likely caused by the cross contamination. Whole genome sequencing analysis can help to determine the occurrence of cross contamination.

1. Introduction

The World Health Organization (WHO) has reported that about one third of the global population is infected with *Mycobacterium tuberculosis*. In 2014, there were about 9,600,000 new patients with tuberculosis worldwide, and 1,500,000 patients died of this disease. Currently, tuberculosis and AIDS are responsible for the highest number of annual infectious disease-related deaths [1]. Tuberculosis is predicted to remain one of the most severe diseases in the coming 10 years [2].

Currently, the main method of laboratory diagnosis for tuberculosis is the use of a microbiological test, such as a sputum smear, sputum solid culture, or liquid culture system (e.g., MGIT 960). The accuracy of the test results can be influenced by various factors, including operation proficiency, specimen status, and reagent quality [3]. The results of these tests are used to determine the correct medication and to evaluate the disease transmission and prognosis. Therefore, quality control of tuberculosis diagnosis tests is important. To improve the test quality across laboratories, standard policies have been released; these guidelines propose a series of quality control protocols for the layout, equipment, culture medium, and reagents used in *M. tuberculosis* laboratories [3,4]. However, there is no universally accepted protocol for the quality control of tuberculosis laboratory test results.

Cross contamination among different specimens is the most frequent problem when conducting a sputum culture test. Cross contamination has been proven to cause false positives, which lead to clinical misdiagnosis and the prescription of incorrect medication [5]; however, few reports in China have addressed this issue. Recently, tuberculosis surveillance uncovered 22 *M. tuberculosis* strains from a single hospital in Shanghai that all had high genotype homology. To confirm the relationship among these strains and to differentiate which results may have been due to cross contamination, we conducted a molecular epidemiologic investigation.

* Corresponding author. Shanghai First Maternity and Infant Hospital, Tongji University School of Medicine, Shanghai, China.
E-mail address: wujie@51mch.com (J. Wu).

https://doi.org/10.1016/j.tube.2018.10.012
Received 26 July 2018; Received in revised form 24 October 2018; Accepted 24 October 2018

Table 1
9 basic loci and 3 hypervariable loci.

Loci	PCR primer sequence(5′-3′)	Flanking region Size + repeat Unit size(bp)	Size of amplicons (bp) and repeat unit size(bp) × no.in H37Rv
QUB-11b	CGTAAGGGGGATGCGGGAAATAGG	67 + 69	412
	CGAAGTGAATGGTGGCAT		69 × 5 + 10
QUB-18	ATCGTCAGCTGCGGAATAGT	182 + 78	621
	AATACCGGGGATATCGGTTC		78 × 5 + 49
MIRU26	GCGGATAGGTCTACCGTCGAAATC TCCGGGTCATACAGCATGATCA	243 + 48	387
			48 × 3
QUB-26	AACGCTCAGCTGTCGGAT	129 + 111	708
	GGCCAGGTCCTTCCCGAT		111 × 5 + 24
Mtub21	AGATCCCAGTTGTCGTCGTC	92 + 57	206
	CAACATCGCCTGGTTCTGTA		57 × 2
MIRU31	CGTCGAAGAGAGCCTCATCAATCAT	108 + 52	264
	AACCTGCTGACCGATGGCAATATC		52 × 3
Mtub04	GTCCAGGTTGCAAGAGATGG	137 + 51	269
	GGCATCCTCAACAACGGTAG		51 × 2 + 30
VNTR2372	ACCTCCGTTCCGATAATC	172 + 57	298
	CAGCTTTCAGCCTCCACA		57 × 2 + 12
Miru40	GGGTTGCTGGATGACAACGTGT	354 + 54	408
	GGGTGATCTCGGCGAAATCAGATA		54 × 1
VNTR 3820	TGCGCGGTGAATGAGACG	247 + 57	444
	ACCTTCATCCTTGGCGAC		57 × 3 + 26
VNTR3232	CCCCAGCCTTACGACTGA	191 + 56	407
	GTCGGGCTTGGTGAAGG		56 × 3 + 48
VNTR4120	GTTCACCGGAGCCAACC	310 + 57	447
	GAGGTGGTTTCGTGGTCG		57 × 2 + 23

2. Materials and methods

2.1. Specimen source

The sample collection was based on the *Diagnosis for Pulmonary Tuberculosis* (WS288-2008), which was released by the Ministry of Health of the People's Republic of China [6]. A total of 22 strains with high genotype homology from one hospital were collected during the surveillance of drug resistance of tuberculosis.

2.2. Drug sensitivity test

The 1% standard proportioning method recommended by WHO was applied when performing the drug-susceptibility test on solid culture medium. The drug concentrations used for each test were as follows: isoniazid 0.2 mg/L, streptomycin 4 mg/L, rifampin 40.0 mg/L, and ethambutol 2 mg/L [7].

2.3. Epidemiological survey

All of the patients in this study signed an informed consent form. Following this, our trained investigators interviewed the patients to collect their epidemiological information, including name, age, address, and disease history. After enquiring about the close contacts of patients, the investigators recorded the locations of activity of the patients. In this study, the people who had direct contact with the patients or those occupying a particular public space together with the patients were considered to have a probable epidemiologic relationship(and vice versa).

2.4. Variable number of tandem repeat (VNTR)genotyping

In this study, we used the 9 + 3 loci VNTR method for the genotyping of *M. tuberculosis* strains, as this method has been validated to be suitable for typing the prevalent *M. tuberculosis*strains in China [8]. PCR amplification combined with gel electrophoresis was used in the detection for VNTR genotyping.

2.5. Genomic DNA extraction

All of the strains were inactivated in a water bath at 85 °C for 30 min in a Biosafety Level 2 Laboratory (with negative pressure), then suspended in DNA extraction liquid and boiled for 10 min in water. After cooling on ice for 2–3 min, cell lysates were centrifuged for 10 min (12,000 rpm). The supernatant was used for genotyping assays. Genomic DNA for whole genome sequencing and analysis was extracted by cetyl trimethylammonium bromide (CTAB) method.

2.6. Whole genome sequencing analysis

We constructed genomic libraries with an insert fragment length ~300 bp following the standard procedures of paired-end sequencing (Illumina). Whole genome sequencing was performed on an Illumina HiSeq 2500 platform, and the average sequencing depth was about 100-fold. The low-quality reads were trimmed by *Sickle* software (https://github.com/ucdavis-bioinformatics/sickle). *Bowtie 2* software (v2.3.1) was used to map all of the reads to the*M. tuberculosis* H37Rv genome (NC_000962.2), and remove multiple overlapping regions. *SAMtools* (v1.4) was used to exclude the nucleotides with sequencing quality and mapping quality below 30. Then, *VarScan* (v.2.3.9) was used to detect fixed mutations with a depth deeper than 10% that of the average sequencing depth and a frequency higher than 75%. According to the PPE/PE-PGRS gene list provided by NCBI, the fixed mutations located in these genes were filtered before the following analysis.

3. Results

The VNTR (9 + 3) genotyping results of 22 tuberculosis strains isolated from a single hospital formed a total of seven clusters (Table 1). This study collected basic patient information from the patients whose samples were the sources of these strains, and assessed the epidemiology as well as the possible cross contamination within the seven clusters.

3.1. Basic information and epidemiologic survey results

The basic information survey revealed that, except for specimens

Table 2
VNTRgenotyping results of the 22 *M. tuberculosis*strains.

strain no.	cluster.100%match	QUB11b	QUB18	Mtub21	Miru26	QUB26	Mtub04	Miru31	Miru40	VNTR2372	VNTR3820	VNTR4120	VNTR3232
1	2	3	3	3	3	7	1	3	3	2	2	5	10
2	2	3	3	3	3	7	1	3	3	2	2	5	10
3	2	3	3	3	3	7	1	3	3	2	2	5	10
4	40	6	10	4	7	6	4	5	3	3	16	20	17
5	40	6	10	4	7	6	4	5	3	3	16	NA	17
6	40	6	10	4	7	6	4	5	3	3	NA	9,20	17
7	120	6	8	5	7	8	4	5	3	3	12	7	15
8	120	6	8	5	7	8	4	5	3	3	12	7	15
9	120	6	8	5	7	8	4	5	3	3	12	7	15
10	193	5	7	5	7	8	3	4	3	3	14	10	13
11	193	5	7	5	7	8	3	4	3	3	14	10	13
12	193	5	7	5	7	8	3	4	3	3	14	10	13
13	202	5	8	5	2	8	6	5	3	3	14	11	13
14	202	5	8	5	2	8	6	5	3	3	14	11	13
15	202	5	8	5	2	8	6	5	3	3	14	11	13
16	205	6	7	5	7	8	4	5	3	3	14	6	15
17	205	6	7	5	7	8	4	5	3	3	14	6	15
18	205	6	7	5	7	8	4	5	3	3	14	6	15
19	230	4	10	4	7	8	3	5	3	3	9	8	8
20	230	4	10	4	7	8	3	5	3	3	9	8	8
21	230	4	10	4	7	8	3	5	3	3	9	8	8
22	230	4	10	4	7	8	3	5	3	3	9	8	8

No. 6, 10, 15, 18, 19, and 22, all of the specimens were tested by sputum smear or sputum culture on the same day as one another (Table 3).

To investigate the transmission relationship among cases, we performed an epidemiologic investigation on all 22 tuberculosis patients. The collected information included the name, age, address, and disease history of the patient. We also enquired about their close contacts and a range of daily activities. The results revealed that there were no confirmed or possible epidemiologic relationships among the 22 patients (Table 2).

3.2. Whole genome sequencing analysis of the M. tuberculosis strains

Whole genome sequencing data were obtained for 21 of the 22 strains; we were unable to obtain the data for strain No. 19 from cluster No. 230. Pairwise comparisons between the genomes in each cluster indicated that there were 15 strains (71.4%) that had no single nucleotide polymorphisms (SNPs) with at least one other strain within their cluster (Fig. 1). The remaining six strains had 2–71 SNPs compared with other strains in the same cluster, and five of them had only 2–4 SNPs. The differences were not uniformly distributed among the clusters. For example, cluster No. 2 in the first group and cluster No. 120 in the third group had no SNPs with one another, which indicates likely cross contamination. Cluster No. 205 in the sixth group had at least 3–4 SNPs from other strains, which could exclude the possibility of cross contamination in this sample. Other clusters showed varying results.

An analysis of the specimen collection time indicated that, among the 16 strains collected on the same day, 14 of them had no SNPs with one another (87.5%). Among the strains within the same cluster whoseSNP distance was 0, 93.3% (14/15) of them had the same collection time, suggesting that this result was most likely caused by cross contamination (Fig. 1. Genetic distance of isolate pairs by duration time).

Table 3
Test information of 22 strains.

strain no.	group	cluster.	Smear time	smear results	Culture time	Culture results	SNPs
1	1	2	2013.12.19	positive	2013.12.19	positive	0
2			2013.12.19	positive	2013.12.19	positive	0
3			2013.12.19	positive	2013.12.19	positive	0
4	2	40	2013.7.9	positive	2013.7.9	positive	0
5			2013.7.9	negative	2013.7.9	positive	0
6			2010.9.14	positive	2010.9.15	positive	71
7	3	120	2013.5.22	positive	2013.5.22	positive	0
8			2013.5.22	negative	2013.5.22	positive	0
9			2013.5.22	negative	2013.5.22	positive	0
10	4	193	2014.3.17	negative	2014.3.17	positive	2
11			2014.6.12	positive	2014.6.12	positive	0
12			2014.6.12	negative	2014.6.12	positive	0
13	5	202	2011.6.9	positive	2011.6.8	positive	0
14			2011.6.9	negative	2011.6.8	positive	0
15			2013.10.24	positive	2013.10.24	positive	4
16	6	205	2014.9.2	negative	2014.9.2	positive	3
17			2014.9.2	positive	2014.9.2	positive	4
18			2014.9.4	negative	2014.9.4	positive	3
19	7	230	2010.8.26	positive	2010.8.26	positive	none
20			2010.8.16	positive	2010.8.13	positive	0
21			2010.8.16	negative	2010.8.13	positive	0
22			2010.7.26	negative	2010.7.26	positive	0

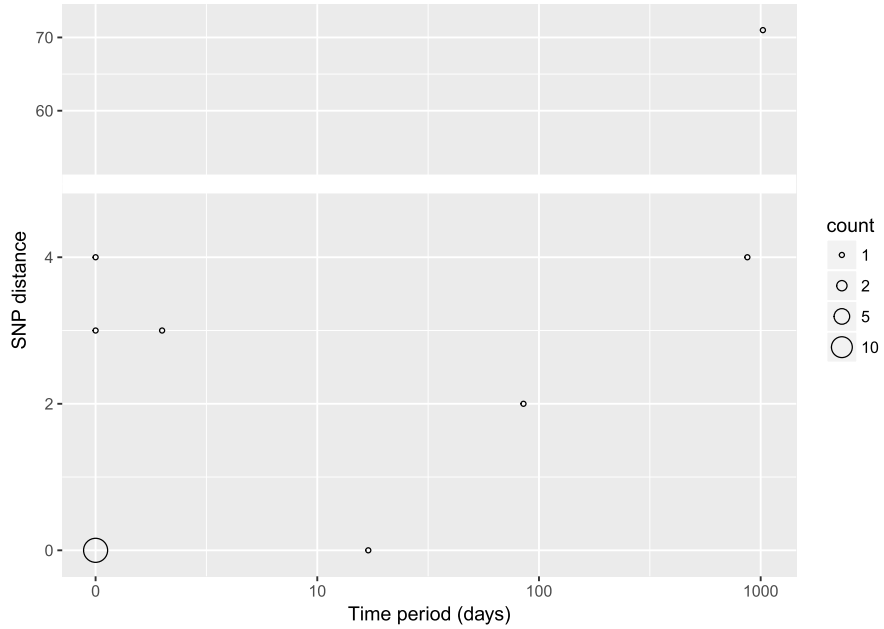

Fig. 1. The distribution of genomic SNPs and the time period of the pairwise comparison within each cluster. The size of circle represents the count of the pairs.

4. Discussion

A measures index based on national criteria is used to guarantee the work quality and improve the test capability in tuberculosis laboratories in each country. For example, some *Mycobacterium* laboratories in the USA follow the Capability Demonstration Program provided by the College of American Pathologists or participate in one of two self-evaluation projects from the Centers for Disease Control and Prevention, which include AFB smear, nucleic acid amplification, growth detection, identification, and drug sensitivity tests [9,10]. In China, the National Tuberculosis Reference Laboratory of the Chinese Center for Disease Control and Prevention regularly conducts external quality assessments (EQAs) of sputum smears, drug sensitivity test proficiency, and molecular biology assays [11,12]. These series of quality control measures effectively improve the test quality of the laboratories [13,14]. However, there are no clear standards for the definition or detection of cross contamination.

To date, there have been only a few reports of cross contamination occurrence in tuberculosis laboratories. Previous studies indicate that the incidence of cross contamination in the test cases was about 1%–10% [15–18]. Generally, if different clinical strains isolated at the same time from the same laboratory testing location are detected as having the same genotype, cross contamination is suspected, although additional evidence is needed to prove its occurrence. For example, in 1995, strains from three tuberculosis patients in Alabama, USA, from non-overlapping locations were identified by a single laboratory as having the same genotype, suggesting cross contamination may have occurred in that laboratory [16]. A similar event was reported for a tuberculosis laboratory in the northwest of Iran [17]. For laboratories in countries and regions with a low tuberculosis incidence, multiple positive culture cases appearing within a short period of time may also be caused by cross contamination; this possibility could be confirmed using a genotyping method [18].

Here, we used whole genome sequencing to analyze the cross contamination occurrence in tuberculosis strains cultured from specimens collected at the same medical institution. The results reveal that 66.7% (14/21) of the strains from one hospital with reported genotype homology had cross contamination; specifically, some of the specimens that shared the same processing time had no difference in genotype, which strongly indicates that cross contamination has occurred. Thus, if the genotypes of strains isolated from samples that were processed together in the same batch appear to be the same, the possibility of cross contamination occurrence during operation should be considered, and corresponding intervention measures should be taken.

In the second, fourth, and fifth groups, one specimen from each group had a SNP from two other specimens in the same group. Notably, these samples were not processed at the same time as the others in the group, which suggests that using whole genome sequencing to detect cross contamination in laboratories is reliable. Data from the sixth group showed that although three specimens had the same or close processing times, they had 3–4 SNPs based on whole genome sequencing. These results suggested that the similarities among these strains were unlikely caused by cross contamination but instead might have been caused by a common infection source, and the small SNP differences among these isolates may have been generated during transmission or be due to host diversity within the strain population. In the seventh group, specimen No. 19 did not undergo whole genome sequencing, but the other three strains in this group were found to lack any SNPs, even though the operation time for specimen No. 22 was different from that of the other two specimens. However, specimens No. 20–22 were assessed in a drug sensitivity test in the same batch on the same day, so the cross contamination in these samples might have occurred during that period.

The laboratory processes performed on each sputum specimen include vortexing, digesting by a NaCl—NaOH solution, neutralizing by phosphate-buffered saline (PBS), centrifuging, and discarding the supernatant. If these procedures are not performed properly, cross contamination can easily occur. For example, a sputum smear and culture are commonly performed in the same biosafety cabinet, and no sterilization is typically conducted between these two operations. Additionally, when a sputum specimen is processed by the digesting solution, the dropper dispensing this reagent may come in contact with multiple samples. Furthermore, the aerosol generated when adding liquid using a dropper, centrifuging, or concentrating could easily cause cross contamination. Thus, having standard operating procedures that prevent cross contamination is of great importance in clinical laboratories. For instance, to limit the chances of cross contamination, specimens that are negative based on the results of a sputum smear are processed first, before those with positive results.

Traditional methods are unable to identify cross contamination

among tuberculosis specimens, but the application of molecular techniques has made the detection of cross contamination possible. Here, the VNTR genotyping technique was initially used to screen the clustered tuberculosis strains, but no epidemiologic relationship was identified among them. Whole genome sequencing was subsequently performed on 21 of the strains. Among the 16 strains that shared the same operation time, 14 (87.5%) of them had no SNPs, which suggested the occurrence of cross contamination. The molecular epidemiologic study of tuberculosis often excludes the possibility of transmission among genotypically-clustered strains that were processed at the same time, identifying it as likely cross contamination [9]. Currently, the most common genotyping method for cross contamination detection in tuberculosis samples is VNTR. However, because only part of the genetic information in the genome is utilized by the VNTR method, it cannot be used to judge the overall genetic relationship between different strains. Whole genome sequencing greatly improves the reliability of differentiation among strains. Therefore, for emergency public health events, such as an epidemic outbreak that requires etiologic testing within a short period of time, whole genome sequencing could be applied to the identification of highly homologous strains, in addition to being used for surveillance of potential cross contamination.

This is a retrospective study, so it is unlikely to trace back all possible sources of cross contamination during sample processing. The use of a prospective experimental design and appropriate controls is needed to clarify this issue. Furthermore, there is still limited report to evaluate whether the slight genome differences among strains from specimens processed at the same time are due to cross contamination or pathogen heterogeneity; this distinction needs further exploration.

Acknowledgement

CGY receives funding from the Robert E. Leet and Clara Guthrie Patterson Trust Fellowship Award and from the MIDAS Center for Communicable Disease Dynamics (grant U54 GM088558).

References

[1] WHO. Global tuberculosis report. World Health Organization; 2015.
[2] Murray CJ, Salomon JA. Modeling the impact of global tuberculosis control strategies. Proc Natl Acad Sci U S A 1998;95(23):13881–6.
[3] STOP TB. Mycobacteriology laboratory manual. A publication of the Global Laboratory Initiative a Working Group of the Stop TB Partnership; 2014.
[4] Tuberculosis laboratory biosafety manual. World Health Organization.
[5] Fitzpatrick L, Braden C, Cronin W, et al. Investigation of Laboratory cross-contamination of *Mycobacterium tuberculosis* cultures. Clin Infect Dis 2004;38(6):e52–4.
[6] People's Medical Publishing House. Diagnostic criteria for tuberculosis. Ministry of health of the People's Republic of China. 2008.
[7] Chinese Anti-tuberculosis Association. The laboratory science procedure of diagnostic bacteriology in tuberculosis. 2006. p. 46–51.
[8] Yang C, Shen X, Peng Y, et al. Transmission of *Mycobacterium tuberculosis* in China: a population-based molecular epidemiologic study. Clin Infect Dis 2013. https://doi.org/10.1093/cid/civ255.
[9] Drobniewski FA, Hoffner S, Rusch-Gerdes S, et al. Recommended standards for modern tuberculosis laboratory services in Europe. Eur Respir J 2006;28(5):903–9.
[10] Drobniewski FA, Caws M, Gibson A, et al. Modern laboratory diagnosis of tuberculosis. Lancet Infect Dis 2003;3(3):141–7.
[11] Jiang Guanglu, Shang Mei, Song Yuanyuan, et al. Analysis of first round external quality assessment for MTB drug sensitivity test within disease control laboratories in China. Chin. J. Antituberculosis 2010;32:806–10.
[12] Van Deun A, Wright A, Zignol M, et al. Drug susceptibility testing proficiency in the network of supranational tuberculosis reference laboratories. Int J Tuberc Lung Dis 2011;15:116–24.
[13] Wang Sumin, Liu Yuhong, Jiang Guanglu, et al. Evaluation of quality control system for direct smear microscopy in National Tuberculosis Program in China. Chin. J. Antituberculosis 2000;22:209–10.
[14] Peng Y, Yang C, Li X, Luo T, Li F, Gao Q. Multiple samples improve the sensitivity for detection of mixed Mycobacterium infections. Tuberculosis 2013;93(5):548–50.
[15] Dunlap NE, Harris RH, Benjamin Jr. WH, Harden JW, Hafner D. Laboratory contamination of *Mycobacterium tuberculosis* cultures. Am J Respir Crit Care Med 1995;152(5 Pt 1):1702–4.
[16] Asgharzadeh M, Kafil HS, Ghorghanlu S, Pourostadi M. Laboratory cross-contamination of *Mycobacterium tuberculosis* in northwest of Iran. Egypt J Chest Dis Tuberc 2015;64:665–9.
[17] Fitzpatrick L, Braden C, Cronin W, et al. Investigation of laboratory cross-contamination of *mycobacterium tuberculosis* cultures. Clin Infect Dis 2004;38(6):E52–4.
[18] de CRM, Soini H, Roscanni GC, Jaques M, Villares MC, Musser JM. Extensive cross-contamination of specimens with *Mycobacterium tuberculosis* in a reference laboratory. J Clin Microbiol Apr. 1999:916–9.