e-ISSN:2320-3528
p-ISSN:2347-2286

Research & Reviews: Journal of Microbiology and Biotechnology

Identification And Validation Of Highly Specific And Stable Diagnostic Targets In *Mycobacterium tuberculosis*

Guangyu Xu[1,2#], Yue Shi[1#], Xiaoyu Sun[1], Chengguo Wei[1], Guoqing Wang[1*], Fan Li[1*]

[1]Key Laboratory of Zoonosis, Ministry of Education, College of Basic Medicine, Jilin University, Jilin, PR, China

[2]Pharmaceutical College, Beihua University, Jilin city, Jilin, PR, China

[#]These two authors contributed equally to this paper

Research Article

Received date: 03/08/2015
Accepted date: 19/10/2015
Published date: 09/11/2015

***For Correspondence**

Guangyu Xu, Key Laboratory of Zoonosis, Ministry of Education, College of Basic Medicine, Jilin University, 126 Xinmin Street, Changchun, Jilin, PR China.

E-mail: xuguangyu2005@163.com

Keywords: *Mycobacterium tuberculosis*, BLAST, Diagnostic targets, PCR

ABSTRACT

Tuberculosis is the most widespread infectious respiratory disease with the largest incidence. To control the spread of tuberculosis, identifying rapid and accurate early diagnostic methods is of great importance. The polymerase chain reaction (PCR) is a simple, rapid, sensitive, and specific gene diagnostic technology, but there are still some problems in selecting appropriate diagnostic targets. Modern computational biology and bioinformatics provide powerful tools to screen new diagnostic targets for tuberculosis. In this article, we compared all genes of *Mycobacterium tuberculosis* H37Rv to all microbial genes that have been sequenced by genome-wide high-throughput scanning technology, and identified 12 genes that are only present in *M. tuberculosis*. Three of these genes, Rv1513, Rv1976, and Rv3738, can be used for early diagnostic screening. The specificity of using the three genes for diagnostic screening was validated on clinical tuberculosis specimens using the PCR method. We also performed PCR to investigate whether the three diagnostic genes could be detected in asthma, pneumonia, and lung cancer samples and the results were negative. These results demonstrate that the three genes can be used as targets for the early diagnosis of tuberculosis.

INTRODUCTION

Tuberculosis (TB) is the most widespread respiratory infectious disease, and it has the largest incidence. The number of active pulmonary TB patients is approximately 20 million worldwide. Each year, there are 8–10 million TB cases, and 2 million people die from the disease [1,2]. Although TB was largely eliminated in the 1950s, a recent survey showed a reemergence in the incidence of TB during the past 20 years [3]. The reemergence of TB is due to its multiple drug resistance, HIV infections, and lower immunity in the elderly population [4]. In addition, the early symptoms of TB patients are similar to the symptoms of asthma, pneumonia, lung cancer, and other diseases, which can delay the diagnosis of TB and lead to mass infections [5]. Therefore, to control the spread of TB more effectively, determining rapid and accurate early diagnostic methods is of great importance.

Polymerase chain reaction (PCR) is a simple, rapid, sensitive, and specific gene diagnostic technology [6]. However, there may be problems in the process of DNA amplification if the specificity of the diagnostic target is weak. In these cases, false positive and false negative results may occur [7,8].

Modern computational biology and bioinformatics provide powerful tools to screen new diagnostic targets for *Mycobacterium tuberculosis* [9,10]. In this study, our goal was to compare all genes of *M. tuberculosis* H37Rv to all microbial genes that have been sequenced by genome-wide high-throughput scanning technology to identify genes specific to *M. tuberculosis*. After identifying targets for the early diagnosis of tuberculosis, we aimed to validate the specificity of the target genes on clinical *M. tuberculosis* specimens using the PCR.

e-ISSN:2320-3528
p-ISSN:2347-2286

METHODOLOGY

Data collection

The genome sequences of all microorganisms and the standard *M. tuberculosis* strains H37Rv, H37Ra, CDC1551, F11, and KZN1435 (National Center for Biotechnology Information [NCBI] numbers NC_000962, NC_009525, NC_002755, NC_009565, and NC_012943, respectively) were collected and downloaded from GenBank to a local disk.

Screening for genes specific to *M. tuberculosis* H37Rv using genome-wide high-throughput scanning technology

The latest version of the Basic Local Alignment Search Tool (BLAST) for Win-32 was downloaded (ftp://ftp.ncbi.nlm.nih.gov/blast/). The blast/bin directory was entered at the command line after decompression and installation. The makeblastdb was produced and executed with the following command: -in D:\blast\DB\in.fasta; -dbtype nucl; -out. D:\blast\DB\in. D:\blast\DB\in.fasta was the source file directory, and D:\blast\DB\in was the target database directory. Finally, the sequence comparison was made to all microorganisms by entering each H37Rv gene. Only the results under 0.00001 were displayed and stored in D:\blast\DB\out.txt.

Bacterial strains

The 15 bacterial strains were from The First Bethune Hospital of Jilin University. All of the strains were isolated from clinical sputum and verified to be *M. tuberculosis* by bacterial culture and serological experiments.

Extraction of genomic DNA from clinical sputum samples

DNA was isolated from sputum samples. The sputum samples were decontaminated, and then the DNA was isolated according to the previously published method [11].

PCR primer design

Primers were designed based on gene sequence information from the GenBank database, and Primer Premier 5 was adopted in software-aided design. The conserved sequences of the gene coding regions (CDS) were selected for primer design. The pre-amplification fragments covered most of the main amino acids sites of the CDS. The synthesis process was completed by Sangon Biotech, LLC.

PCR conditions

The following standard PCR conditions were used: 5 min at 94°C for pre-denaturation, 30 s at 94°C for denaturation, 30 s at 55°C for annealing, and 60 s at 72°C for extending. The 30 cycles were followed by a final extension at 72°C for 10 min.

Ethical consideration

Ethical approval to conduct the study was obtained from institutional ethic review committee before the commencement of the study. Patients were required to sign a written informed consent for the study and for tuberculosis testing.

RESULTS

Candidate diagnostic targets

We compared all genes of *M. tuberculosis* H37Rv to all other microbial (not include *Mycobacterium* species) genes that have been sequenced using BLAST (E-value<e-5, sequence identity>95%). We screened out the genes that were only present in the standard strain H37Rv, but did not exist in other microorganisms. These 12 candidate genes have the potential to be diagnostic targets for TB **(Table 1).**

Table 1. Candidate diagnostic target genes of H37Rv.

Gene	Function
Rv1512	nucleotide-sugar epimerase epiA
Rv1513	hypothetical protein
Rv1516c	sugar transferase
Rv1973	MCE associated membrane protein
Rv1974	hypothetical protein
Rv2646	Integrase
Rv2655c	phiRv2 prophage protein
Rv2659c	phiRv2 prophage integrase
Rv3119	molybdenum cofactor biosynthesis protein E
Rv3120	hypothetical protein
Rv3617	epoxide hydrolase
Rv3738c	PPE family protein

e-ISSN:2320-3528
p-ISSN:2347-2286

Screening by bibliometric method

The candidate diagnostic genes may have been identified previously by other methods, so we performed searches using the bibliometric method [12]. We found that Rv1512, Rv1516, Rv1973, Rv2646, and Rv3617 have previously been used as diagnostic targets **(Table 2).** Therefore, we excluded these five genes from further analysis.

Table 2. Genes that have been previously used as diagnostic targets.

Gene	Function	Reference
Rv1512	nucleotide-sugar epimerase epiA	[15]
Rv1516c	sugar transferase	[16]
Rv1973	MCE associated membrane protein	[17]
Rv2646	Integrase	[18]
Rv3617	epoxide hydrolase	[19]

Screening for genetic stability

The stability of a diagnostic target is an important factor, and one criterion for stability is whether the flanking sequences contain transposons, integrons, or other movable genetic elements. Therefore, we analyzed the transposons, integrons, and other movable genetic elements in the flanking sequences within 1500 bp of the candidate diagnostic targets. Many transposons were present within 1500 bp of Rv2655 and Rv2659c, so these genes were very unstable for use as diagnostic targets. There were anomalous regions and labile regions within 1500 bp of Rv3119, Rv312, and Rv3617, so these genes were not suitable as diagnostic targets. **Table 3** shows details about the flanking sequences around the candidate diagnostic genes. From this analysis, Rv1513, Rv1974, and Rv3738 had conservative flanking sequences.

Table 3. Flanking sequences of candidate diagnostic targets in H37Rv.

Gene	1500 bp flanking region		
	tRNA	Transposase	phage
Rv1513			
Rv1974			
Rv3738c			
Rv2655c		+	+
Rv2659c		+	+
Rv3119		+	
Rv3120		+	

Homology screening

Because we only scanned the genome of *M. tuberculosis* H37Rv, there remained questions about whether these diagnostic genes were also specific for other *M. tuberculosis* strains. We compared the diagnostic genes to the genomes of four other strains of *M. tuberculosis* (H37Ra, CDC1551, F11, and KZN1435), and the results showed that the genes were highly conservative **(Figure 1).** The names of the three conserved genes in the five *M. tuberculosis* standard strains are shown in **Table 4.**

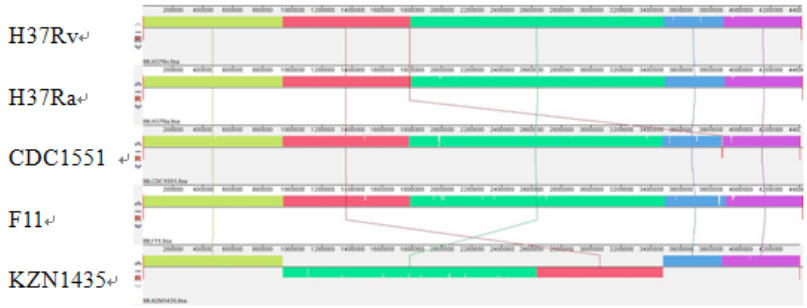

Figure 1. Comparison of the genomic sequences of five *M. tuberculosis* strains (In the standard color scheme, the region of sequence covered by a colored block is entirely collinear and homologous among the genomes).

Table 4. Comparison of genes name of five *M. tuberculosis* strains.

H37Rv	H37Ra	CDC1551	F11	KZN1435
Rv1513	MRA_1523	MT1561	TBFG_11545	TBMG_02464
Rv1973	MRA_1981	MT2022	TBFG_12001	TBMG_02018
Rv3738c	MRA_3778	MT3848	TBFG_13772	TBMG_03785

PCR validation of the diagnostic targets in clinical strains of *M. tuberculosis*

We screened the three candidate diagnostic targets: Rv1513, Rv1973, and Rv3738, by using BLAST, the bibliometrics method, and the stability filter method. The three genes were specific and stable for use as new diagnostic targets, but still needed

e-ISSN:2320-3528
p-ISSN:2347-2286

to be verified by PCR. We extracted DNA from 15 clinical *M. tuberculosis* strains and performed PCR using the primers shown in **Table 5** for the three candidate diagnostic targets. The PCR results shown in **Figure 2** demonstrate that the positive rate of detecting the three genes in clinical isolates was 100%, and there was good sensitivity and accuracy.

Table 5. PCR primer sequences.

Gene	Sequence (5′-3′)	Expected length (bp)
Rv1513	TCCGTCTTGCCGATGTT ATCGGCCTGCAGCATTC	360
Rv1974	GTGGGTGCGCTGCTATGC GGCGTTGGCGAAGTTGT	117
Rv3738	CCGAGTTGGCCGCTAATCA CAGCACCGCCGTTGAACA	216

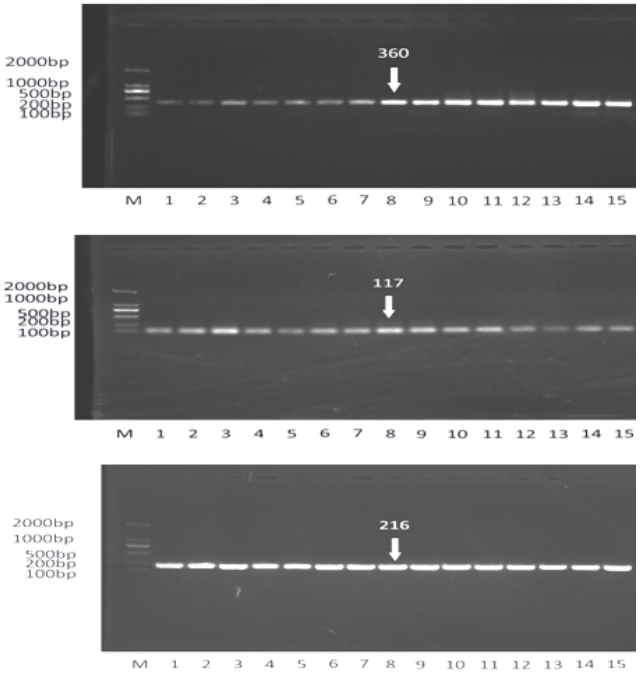

Figure 2. PCR results for Rv1513 (top), Rv1974 (middle), and Rv3738 (bottom) in 15 clinical *M. tuberculosis* strains.

Validation that PCR for TB diagnostic targets are negative in asthma, pneumonia, and lung cancer samples

Early signs of asthma, pneumonia, and lung cancer are similar to TB. Patients with these disorders present coughs and lung shadows in computed tomography scans, making it difficult to differentiate TB from these other diseases. To test the specificity of the three diagnostic genes, we performed PCR to detect the diagnostic targets in asthma, pneumonia, and lung cancer samples. The PCR results in the asthma, pneumonia, and lung cancer samples were all negative.

DISCUSSION

In this study, we compared all genes of *M. tuberculosis* H37Rv to all microbial genes that been sequenced by genome-wide high-throughput scanning technology, and identified 12 genes that are only present in *M. tuberculosis*. Three of these genes, Rv1513, Rv1976, and Rv3738, can be used for early diagnostic screening. The specificity of using these three genes for diagnostic screening was validated on clinical *M. tuberculosis* specimens using the PCR method. Additionally, when these genes were used as diagnostic targets for asthma, pneumonia, and lung cancer, which have symptoms similar to TB symptoms, the results were negative, indicating that the genes are specific for TB. Therefore, these three genes have a high potential for use as early diagnostic targets of TB.

Of the 12 genes we identified by genome-wide high-throughput scanning technology, five have already been used in PCR diagnostic methods. These genes account for 42% of the genes identified in our screen, indicating that genome-wide high-throughput scanning filters out diagnostic targets. These five genes have been used as tuberculosis diagnostic targets only on a small scale and are not gold standards. Although the genes are specific for *M. tuberculosis*, the stability of these genes was not screened, and this omission may lead to a certain amount of false positives in PCR assays for *M. tuberculosis* detection.

We verified that there was a 100% positive rate of detecting the three candidate target genes by PCR in some *M. tuberculosis* clinical strains. The PCR results showed good sensitivity and accuracy, indicating that the genes may be used as early diagnostic targets of *M. tuberculosis* following validation and clinical testing in large scale samples. We also verified that there was a 100%

negative rate of detecting the three candidate target genes by PCR in asthma, pneumonia, and lung cancer samples because these disease have similar symptoms to TB. This result demonstrated that the genes we identified by high-throughput screening exist only in *M. tuberculosis*, and do not exist in other pathogens or non-tuberculosis pathogens.

Diagnostic targets must have two characteristics: strong specificity and stability. The genes filtered out by genome-wide high-throughput scanning technology have very strong specificity. Besides screening for specificity, we also screened for the stability of the diagnostic targets in this study. To some extent, this screening may eliminate false positives when detecting *M. tuberculosis*. We screened the flanking sequences of the genes for transposons, integrons, and other movable genetic elements to determine the stability of the three target genes. Rv1513, Rv1973, and Rv3738 have conservative flanking sequences, and thus can be used as diagnostic targets.

Rv3738 belongs to the Pro-Pro-Glu (PPE)/Pro-Glu (PE) family, which is a specific protein family of Mycobacterium that is involved in the pathogenicity and infection of *M. tuberculosis*[13,14]. Genes of the PE/PPE family were once used as diagnostic targets, but because these genes were screened only for specificity and not for stability, these targets are prone to false positive or false negative results[7,8]. In this study, the Rv3738 gene was screened for specificity by whole genome scanning technology and for stability, so it can be used appropriately as a diagnostic target. The genic functions of Rv1513 and Rv1973 are unknown. We speculate that they may be specific and necessary genes in the *M. tuberculosis* family.

CONCLUSION

In this study, 12 candidate diagnostic genes were filtered out using bioinformatics. Following data mining and screening for stability, we identified Rv1513, Rv1973, and Rv3738 as new, highly specific diagnostic targets for *M. tuberculosis*. According to PCR testing, these three diagnostic targets are only present in *M. tuberculosis* and are not present in other organisms or non-tuberculosis diseases, so the genes can be used as diagnostic targets for TB. This simple and rapid diagnostic method can be applied in the early diagnosis of TB and to prevent large-scale transmission of the disease.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

GX participated in study design, literature search, data analysis, manuscript writing, editing and submission of the manuscript. YS, XS, CW, GW and FL participated in data analysis, manuscript writing and editing. All authors read and approved the final manuscript.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (81401712).

REFERENCES

1. Glaziou P, et al. Global epidemiology of tuberculosis. Semin Respir Crit Care Med (2013); 34: 3-16.
2. Pontali E, et al. Drug-resistant tuberculosis. Curr Opin Pulm Med (2013); 19: 266-272.
3. Zar HJ and Udwadia ZF. Advances in tuberculosis 2011-2012. Thorax (2013); 68: 283-287.
4. Rajasekaran A. Tuberculosis is still a scourge of mankind--why? J Indian Med Assoc (2013); 110: 323-324.
5. Porcel JM, et al. Year in review 2012: lung cancer, respiratory infections, tuberculosis, pleural diseases, bronchoscopic intervention and imaging. Respirology (2013); 18: 573-583.
6. Pholwat S, et al. Digital PCR to detect and quantify heteroresistance in drug resistant *Mycobacterium tuberculosis*. PLoS One (2013); 8: e57238.
7. Gomez PD, et al. [Effectiveness of smears and cultures in gastric aspirate samples in the diagnosis of tuberculosis]. An Esp Pediatr (2000); 53: 405-411.
8. Strumpf IJ, et al. The reliability of gastric smears by auramine-rhodamine staining technique for the diagnosis of tuberculosis. Am Rev Respir Dis (1976); 114: 971-976.
9. Wang TC and Chen FC. The evolutionary landscape of the *Mycobacterium tuberculosis* genome. Gene (2013); 518: 187-193.
10. Bouwman AS, et al. Genotype of a historic strain of *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A (2012); 109: 18511-18516.
11. Silva JL, et al. Plasmid-based controls to detect rpoB mutations in *Mycobacterium tuberculosis* by quantitative polymerase chain reaction-high-resolution melting. Mem Inst Oswaldo Cruz (2013); 108: 106-109.

12. Ramos JM, et al. A bibliometric analysis of tuberculosis research indexed in PubMed, 1997-2006. Int J Tuberc Lung Dis (2008); 12: 1461-1468.

13. Lew JM, et al. TubercuList--10 years after. Tuberculosis (2010); 91: 1-7.

14. Camus JC, et al. Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. Microbiology (2002); 148: 2967-2973.

15. Cao YT. Cloning, Pr-okaryotic Expression and Specificity Identification of *Mycobacterium tuberculosis* Rv1512 Gene. J Clin Pulmo Med (2011); 16: 1887-1889.

16. Al-Khodari NY, et al. Identification, diagnostic potential, and natural expression of immunodominant seroreactive peptides encoded by five *Mycobacterium tuberculosis*-specific genomic regions. Clin Vaccine Immunol. (2010); 18: 477-482.

17. Song H, et al. Identification of outer membrane proteins of *Mycobacterium tuberculosis*. Tuberculosis. (2008); 88: 526-544.

18. Ding ZX, et al. Functional Study of Region of Differences in *Mycobacterium tuberculosis* and its Application. J Clin Pulmo Med (2012); 17: 505-507.

19. Biswal BK, et al. Cloning, expression, purification, crystallization and preliminary X-ray studies of epoxide hydrolases A and B from *Mycobacterium tuberculosis*. Acta Crystallogr Sect F Struct Biol Cryst Commun (2006); 62: 136-138.