(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 June 2015 (11.06.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/083056 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/IB2014/066469

(22) International Filing Date:
1 December 2014 (01.12.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
5572/CHE/2013 3 December 2013 (03.12.2013) IN

(71) Applicants: **KUSUMA SCHOOL OF BIOLOGICAL SCIENCES** [IN/IN]; Indian Institute Of Technology, Delhi, Hauz Khas, New Delhi 110016 (IN). **EUROPEAN MOLECULAR BIOLOGY LABORATORY** [DE/DE]; Meyerhofstrasse 1, 69117 Heidelberg (DE).

(72) Inventors: **RAHMAN, Syed Asad**; 7 Primrose Lane, Impington, Cambridge, Cambridge CB24 9PX (GB). **AHMAD, Javeed**; Rm No. 318, RPC-1, Boys Hostel, AIIMS, Ansari Nagar, New Delhi 110029 (IN). **EHTESHAM, Nasreen Z**; 13, West Avenue, IIT Campus, Hauz Khas, New Delhi 110016 (IN). **HASNAIN, Seyed E**; 13, West Avenue, IIT Campus, Hauz Khas, New Delhi 110016 (IN).

(74) Agent: **SANTHANAM, Kausalya**; 128/129 Phase I, Royal Enclave, Srirampura, Bangalore Indian (IN).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to the identity of the inventor (Rule 4.17(i))*

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

— *of inventorship (Rule 4.17(iv))*

Published:

— *with declaration under Article 17(2)(a); without abstract; title not checked by the International Searching Authority*

(54) Title: GENETIC MARKERS FOR DIAGNOSIS OF TUBERCULOSIS CAUSED BY MYCOBACTERIUM TUBERCULOSIS

(57) Abstract:

WO 2015/083056 A2

GENETIC MARKERS FOR DIAGNOSIS OF TUBERCULOSIS CAUSED BY *MYCOBACTERIUM TUBERCULOSIS*

Related Application

This application is related to and takes priority from the Indian Provisional Application 5572/CHE/2013 filed on December 3, 2013 and is incorporated herein in its entirety.

Field of the Invention

The application is related to novel signature sequences for diagnosis of *Mycobacterium tuberculosis* in clinical samples. These signature sequences have the ability to differentiate *Mycobacterium tuberculosis* DNA from other mycobacterial species by PCR with 100% specificity and very high sensitivity.

Background of the Invention

Tuberculosis (TB) is a major global health problem with an alarming rate of mortality associated with it. It is one of the leading infectious diseases caused by bacteria taking one human life every 15-20 seconds globally. Estimates of 2011 reveal that there are almost 9 million new cases and 1.4 million TB deaths (Global Tuberculosis Report 2012, WHO 2013). The disease is caused by *Mycobacterium tuberculosis*, a member of the genus Mycobacterium, while in a few cases *Mycobacterium bovis* has been reported to be the causal organism. More than 100 mycobacterium species are known and among them only a few are pathogenic for humans.

Conventional diagnostic methods include examination of sputum smear under a microscope for acid-fast mycobacteria and radiological reading of lungs. However, in most cases the sputum smear examination turns out to be negative for the bacteria due to early stages of infection and lung changes are not readily visible on an x-rays until several months into the infection.

Diagnosis of mycobacteria-related disease poses difficulties for several reasons which have been recognized by researchers and clinicians over the years. Firstly, these bacteria are in small numbers and are already at a contagious stage when they become detectable by conventional methods. Pulmonary disease caused by different mycobacteria are not readily detectable

clinically or radiologically. Detection of organism in culture achieves 100% specificity but the growth of mycobacteria in culture takes about 3-6 weeks (Bates et al, Am. Respir.134, 415-417, 1986) thus making the process time-consuming. In addition, repeated cultures may be required to ensure success.

Several molecular tests for tuberculosis have been developed in the past. These include the Gen-probe 'Amplified mycobacterium direct test' by Abbe et al (J. Clin. Mincrobiol. 31, 3270, 1993), ligase chain reaction (LCR) (J. Clin. Microbiol. 35, 2424, 1997), PCR based Roche Amplicor TB test (J. Clin. Microbiol. 33, 1832, 1995) and IS6110 based detection (J. Clin. Microbiol. 28, 2668, 1990).

US5168039 discloses the IS6110 based detection of *M. tuberculosis* wherein a repetitive DNA segment specific for members of *M. tuberculosis* complex is used for the diagnosis. While IS6110 based detection system has been shown to have high level of specificity, there are also reports on false positive detections of 3 to 20% making it unreliable (J. Clin. Microbiol. 32, 277, 1994). In addition, lack of IS6110 sequence in some *M. tuberculosis* strains may also limit the use of the same for diagnostic tests routinely (Tuber. Lung Dis. 76, 550, 1995). US7638309 provides detection of mycobacteria in clinical specimens in the intergenic region between methyl mycolic acid synthase genes mmaA1 and mmaA2 and the flanking region in mmaA1 and mmaA2 genes.

Thus, it appears that there is a paucity of simple, rapid and reliable tests that can specifically detect *M. tuberculosis* during early stages of the disease. The present invention has identified 'signature sequences' that can differentiate *M. tuberculosis* from a large number of other mycobacterial DNA. These 'signature sequences' are used in detection of early disease in clinical samples of patients.

Summary of the Invention

The invention provides novel signature sequences for diagnosis of Mycobacterium species (sps) in clinical samples with 100% specificity and a very high degree of sensitivity.

In one aspect, the invention provides a nucleotide sequence capable of selectively detecting pathogenic Mycobacterium sps using oligonucleotide primers corresponding to the signature sequence selected from SEQ ID NO: 1, 2, 3 or 4.

In another aspect the invention provides a method of detecting pathogenic mycobacterium sps in a clinical sample, said method comprising the steps of:

a. removal of contaminants from the clinical sample by conventional methods;

b. extraction of genomic DNA from the contaminant-free clinical sample;

c. designing a set of specific oligonucleotide primers capable of specifically detecting SEQ ID NO: 1, 2, 3 or 4 for use in RT-PCR;

d. analyzing PCR product by electrophoresis or specific probe nucleotide sequence complementary to SEQ ID NO: 1, 2, 3 or 4.

The set of oligonucleotide primers of the invention are selected from

(i) 5' ATGCAGGTTGCGACTGTACACCCGG 3'

3' GGCCGCTCTTGTTCTTCGTCGGAT 5'

(ii) 5' GTGTTTGCGTTGAGTAATAATCTGAACCGTGT 3'

3' AGCCAATTCCAGCACGATGTCGCC 5'

(iii) 5' ATGCAGGTTGCGACTGTACACCCGG 3'

3' GGCCGCTCTTGTTCTTCGTCGGAT 5'

(iv) 5' TGTACCGCGTGCCCGACGATTTG 3'

3' ACAGGCAGCTAACAGGGCGTCGG 5'

(v) a set of oligonucleotide primers complementary to (i), (ii), (iii) or (iv)

or

(vi) a set of oligonucleotide primers comprising of sequence containing any 10 consecutive bases from one of the sequences selected from SEQ ID NO: 1, 2, 3 or 4.

In yet another aspect, the invention provides a kit for the detection of pathogenic mycobacterium sps in clinical samples, said kit comprising set of oligonucleotide primers selected from

(i) 5' ATGCAGGTTGCGACTGTACACCCGG 3'

3' GGCCGCTCTTGTTCTTCGTCGGAT 5'

(ii) 5' GTGTTTGCGTTGAGTAATAATCTGAACCGTGT 3'

3' AGCCAATTCCAGCACGATGTCGCC 5'

(iii) 5' ATGCAGGTTGCGACTGTACACCCGG 3'

3' GGCCGCTCTTGTTCTTCGTCGGAT 5'

(iv) 5' TGTACCGCGTGCCCGACGATTTG 3'

3' ACAGGCAGCTAACAGGGCGTCGG 5'

(v) a set of oligonucleotide primers complementary to (i), (ii), (iii) or (iv)

or

(vi) a set of oligonucleotide primers comprising of sequence containing any 10 consecutive bases from one of the sequences selected from SEQ ID NO: 1, 2, 3 or 4.

Furthermore, the invention provides a method of detecting pathogenic mycobacterium sps in a clinical sample wherein the sample is isolated from individuals vaccinated against tuberculosis. It also provides a method of detecting pathogenic mycobacterium sps in a clinical sample wherein the sample is isolated from individuals treated against tuberculosis.

Brief Description of Figures

Fig. 1A: Amplification of SS1 at low and varied DNA template concentrations.

Fig. 1B: Amplification of SS2 at low and varied DNA template concentrations.

Fig. 1C: Amplification of SS3 at low and varied DNA template concentrations.

Fig. 1D: Amplification of SS4 at low and varied DNA template concentrations.

Fig. 2: Amplification of signature sequences SS1, SS2, SS3 and SS4 from patient sputum samples.

Fig. 3: Amplification of signature sequences SS1 and SS3 from patient blood samples.

Details of the Invention

The present invention relates to detection of pathogenic Mycobacterium species using signature sequences SEQ ID NO: 1, 2, 3 or 4 with a high degree of sensitivity and 100% specificity.

In one embodiment, the invention provides novel DNA diagnostic markers for specific detection of *Mycobacterium tuberculosis* which causes tuberculosis.

A three-pronged approach was carried out to identify novel DNA diagnostic marker for detection of pathogenic mycobacterium sps, especially, *Mycobacterium tuberculosis*. First step provides an *in-silico* approach to identify and shortlist potential sequences of Mycobacterium, unique and exclusive to pathogenic mycobacterium sps, especially *Mycobacterium tuberculosis*. The criteria used for selection of the potential sequences are presented below which involves comparative proteomic analysis of 13 mycobacterium species:

i. Strict pathogens (the most virulent pathogens) such as *Mycobacterium tuberculosis*, *Mycobacterium laprae*, *Mycobacterium ulcernus* and *Mycobacterium bovis*.

ii. Opportunistic pathogens, which belong to Non Tuberculous Mycobacteria (NTM) group, can cause pulmonary and other disseminated infections in immune compromised individuals (Infect. Genet. Evol. 12, 832, 2012; Appl. Environ. Microbiol. 79, 825, 2013). *Mycobacterium marinum, Mycobacterium avium, Mycobacterium intracellulare, Mycobacterium aviumpara tuberculosis and Mycobacterium abscessus,* cause opportunistic pulmonary infection in human whereas *Mycobacterium avium* subspecies *paratuberculosis* (MAP), the third member of MAC is the suspected causative agent of Crohn's disease in human (Appl. Environ. Microbiol. 79, 825, 2013; Crit. Rev. Microbiol. 38, 52, 2012).

iii. Non-pathogenic group includes *Mycobacterium gilvum*, *Mycobacterium vanbaalenii,Mycobacterium smegmatis*and *Mycobacterium indicus pranii*, which does not cause disseminated infections even in immune compromised individuals (BMC Microbiol. 10, 237, 2010; Infect. Genet. Evol., 12, 853, 2012; Br. J. Exp. Pathol. 52, 627, 1971).

The following bioinformatics process flow resulted in the identification of potential gene markers of the invention.

a) Perform all against all NCBI BLAST (J Mol Biol., 215, 403, 1990) on the protein sequences from the selected genomes.

b) Perform all against all NCBI BLAST ((J Mol Biol., 215, 403, 1990) on the nucleotide sequences from the selected genomes.

c) Identify Protein Domains from Pfam (Comp. Genomics, 396, 43, 2007) and GENE-3D/CATH (Nucleic Acids Research, 40, D465, 2012) using InterPro (Nucleic Acids Research, 40, D306, 2012).

d) Classify sequences into three categories namely i) CLASS 1 and ii) CLASS 2 as potential hits and iii) rest as non-hits.

The above process flow resulted in the classification of the potential hits into class 1 and class 2. As per the present invention, class 1 are genes unique to the organism of interest based on the fact that they do not share protein domain and protein sequence identity of more than 20% and nucleotide sequence identity of more than 35% with any other organism in the selected organism list. In another embodiment, class 2 genes are those that do not share protein domain and protein sequence identity of more than 20% and nucleotide sequence identity of more than 35% with any other organism in the selected organism list.

Table 1 provides potential candidate genes carrying 'signature sequences' based on the bioinformatics process flow.

Table 1: Potential "signature sequences" carrying candidate genes

H37Rv Gene Identifiers	Class	H37Rv protein description
Rv1507A	1	Hypothetical protein
Rv1509	1	Hypothetical protein
Rv2645	1	Hypothetical protein
Rv2653c	1	Possible PhiRv2 prophage protein

Rv2654c	1	Possible PhiRv2 prophage protein
Rv2658c	1	Possible prophage protein
Rv0064A	2	Possible antitoxin VapB1
Rv0078B	2	Hypothetical protein
Rv0397A	2	Hypothetical protein
Rv0456B	2	Possible antitoxin MazE1
Rv0959A	2	Possible antitoxin VapB9
Rv1366A	2	Hypothetical protein
Rv1954A	2	Hypothetical protein
Rv1991A	2	Antitoxin MazE6
Rv2142A	2	Possible antitoxin ParD2
Rv2231A	2	Possible toxin VapC16
Rv2231B	2	Possible antitoxin VapB16
Rv2274A	2	Possible antitoxin MazE8
Rv2395A	2	Acid and phagosome regulated protein A AprA
Rv2395B	2	Acid and phagosome regulated protein B AprB
Rv2862A	2	Possible antitoxin VapB23
Rv3190A	2	Hypothetical protein

Rv3344c	2	PE-PGRS family protein PE_PGRS49] [partial=5']
Rv3512	2	PE-PGRS family protein PE_PGRS56] [partial=5']
Rv3599c	2	Hypothetical short protein

In one aspect, the invention functionally characterizes the potential 'signature sequences (SS)'-carrying candidate genes based on functional information retrieved from Tuberculist (Tuberculosis (Edinb) 91, 7, 2011) and TB database (Nucleic Acids Research, 37, D499, 2009). Accordingly, the potential signature sequences can be functionally characterized into the following groups:

a. 9 (Rv0064A, Rv0456B, Rv0959A, Rv1991A, Rv2142A, **Rv2231A**, Rv2231B, Rv2274A and Rv2862A) fell into the toxin-antitoxin category.
b. 3 (Rv2653c, Rv2654c, Rv2658c) are possible prophages.
c. 2 (Rv3344c and Rv3512) belong to PE_PGRS family of proteins
d. 2 (Rv2645 and Rv2653c) are deleted (partially or completely) in one or more clinical isolates eliminating their use as diagnostic markers.
e. 9 (**Rv1507A, Rv1509**, Rv0078B, Rv2645, Rv0397A, Rv1366A, **Rv1954A**, Rv3599c, Rv3190A) are hypothetical proteins.
f. 2 (Rv2395A, Rv2395B) are acid and phagosome regulated proteins.

Based on the *in-silico* analysis, two **Class 1 genes (Rv1507A and Rv1509)** and two **Class 2 genes (Rv1954A and Rv2231A)** with homologs in *Mycobacterium bovis BCG* were selected as potential candidates (Table 2).

Table 2: Nucleotide Sequences of Potential "signature sequences" carrying genes

Gene Name	Sequence	Description	Prediction Class

Rv1507A hypothetical protein	>Rv1507A ATGCAATCAGGTCAAAATATCCTCGCCAAGGT ATGTAATTTGATTGAACAATCGCGACTTTCTTC AACGCGGTGTCTCCAATTTAGAATAACAAATA CGTCGCGCCCGCGACAGCTCCGCTGGAGCGA GTTCAAGCGATTCTGCGACATATTCAATATGG TGCTCGGGAAGGCCAGGATGGGCCGCGACCC GGGGCGTCCGGTGCGCGATGAACGTCGCATC GTCTCCTGTGAGATAATTGCATCCGATCATAT AGGGCTGGCTGCGGCTAGGTTGCTGGCAAAA AGATATCGCGGCCGATCCGTTTCTGGTTTTGT CTTGATGATCAAATCCGCTTCCGTTCACGAGA TCGATTCCTGGTCTTCCCCCAGCGTCGCGATG TCGATAGGTGTCGCGCTTTGTTCGTACCCGCA CTACGCGGCGGCGAGAACCTCGCCACCGAAT CGGGATTGGGGGGAGGATACCACTCGGTCGA GGCCCGTCACCGGCCTTCTAGCGGGTTG	Gene length: 504bp, Protein length: 167aa	**CLASS 1**
Rv1509 Essential hypothetical protein	>Rv1509 GTGTTTGCGTTGAGTAATAATCTGAACCGTGT GAACGCATGCATGGATGGATTCCTTGCCCGTA TCCGCTCACATGTTGATGCGCACGCGCCAGAA TTGCGTTCACTGTTCGATACGATGGCGGCCGA GGCCCGATTTGCACGCGACTGGCTGTCCGAG GACCTCGCGCGGTTGCCTGTCGGTGCAGCATT GCTGGAAGTGGGCGGGGGGGTACTTCTGCTC AGCTGTCAACTGGCGGCGGAGGGATTTGACA TCACCGCCATCGAGCCGACGGGTGAAGGTTTT GGCAAGTTCAGACAGCTTGGCGACATCGTGC TGGAATTGGCTGCAGCACGACCCACCATCGC	Gene length: 882bp, Protein length: 293aa	**CLASS 1**

	GCCATGCAAGGCGGAAGACTTTATTTCCGAG AAGCGGTTCGACTTCGCCTTCTCGCTGAATGT GATGGAGCACATCGACCTTCCGGATGAGGCA GTCAGGCGGGTATCGGAAGTGCTGAAACCGG GGGCCAGTTACCACTTCCTGTGCCCGAATTAC GTATTCCCGTACGAACCGCATTTCAATATCCCA ACATTCTTCACCAAAGAGCTGACATGCCGGGT GATGCGACATCGCATCGAGGGCAATACGGGC ATGGATGACCCGAAGGGAGTCTGGCGTTCGC TCAACTGGATTACGGTTCCCAAGGTGAAACGC TTTGCGGCGAAGGATGCGACGCTGACCTTGC GCTTCCACCGTGCAATGTTGGTATGGATGCTG GAACGCGCGCTGACGGATAAGGAATTCGCTG GTCGCCGGGCACAATGGATGGTCGCTGCTATT CGCTCGGCGGTGAAATTGCGTGTGCATCATCT GGCAGGCTATGTTCCCGCTACGCTGCAGCCCA TCATGGATGTGCGGCTAACGAAGAGGTAA		
Rv1954a Hypothetical protein	>gi|448814763:2201231-2201623 Mycobacterium tuberculosis H37Rv complete genome TGGTATAAGCTGGTTTTAGACGAAAAGGACC CCACCTCGGGGTCTGATGGCCAGGGGCAGGG TCGTGTGCATTGGGGATGCAGGTTGCGACTG TACACCCGGCGTGTTCCGCGCGACAGCGGGT GGGATGCCGGTGCTGGTGGTCATCGAGTCTG GGACAGGAGGTGATCAGATGGCTCGTAAAGC TACGTCCCCGGGTAAGCCGGCTCCGACGTCG GGACAGTATCGCCCGGTTGGCGGTGGCAACG AGGTGACCGTTCCGAAGGGACACCGTCTGCC	Gene length: 303bp, Protein length: 100bp	**CLASS 2**

	TCCCTCGCCCAAGCCCGGTCAGAAGTGGGTG AACGTCGATCCGACGAAGAACAAGAGCGGCC GCGGCTGAGCTTGTGCCGTCGGGATGGGTGT CGCACCGTCTCGGCGGGTCGC		
Rv2231A	>gi|448814763:c2506224-2505671 Mycobacterium tuberculosis H37Rv complete genome GCCGCGGCGAGCCGGTAGCAAAGCTTGTGCC GCTGCATCCTCATGAGACTCGGCGGTTAGGCA TTGACCATGGCGTGTACCGCGTGCCCGACGAT TTGGACGCTCCGTTGTCAGACGACGTGCTCGA ACGCTTTCACCGGTGAAGCGCTACCTCATCGA CACCCACGTTTGGCTGCGGATGCCGTCAACGA AACACGGGCGATTGTTCAGGACGTCCGCAAC AGCATTCTCTTGTCGGCCGCCAGTGCCTGGGA GATCGCGATCAACTACCGCCTCGGCAAGCTCC CGCCGCCCGAGCCATCGGCCTCTTACGTGCCC GATCGAATGCGCCGCTGCGGCACGTCGCCGC TGTCAGTTGACCACGCACACACTGCGCACCGC AGAGCTTCCGGATCACCATCGACATCCATTCG ACCGTGTGCTCATCGCCCAGGCACAGCTGCTT GGCCTGACGATCATCACCGCCGACGCCCTGTT AGCTGCCTGTGATGTCGCGGTTGTCGCCGCGT AGACAACGCGTCGGCGGTGCTCTGGATTCTTG GCCCGCACACCG	Gene length: 426bp, Protein length: 141aa	**CLASS 2**

In a most preferred aspect, the signature sequences were designed keeping in view the diagnostic tool of RT PCR. These were short sequences amenable for PCR amplification from selected genes. The specific signature sequences, SS1 (Rv1507A), SS2 (Rv1509), SS3 (RV1954A)

and SS4 (Rv2231A) of the invention are provided below. Homology search using NCBI nucleotide BLAST against the genus Mycobacterium was conducted on these signature sequences to confirm their uniqueness.

SS1 (Rv1507A): SEQ ID NO: 1

>Rv1507A

ATGCAATCAGGTCAAAATATCCTCGCCAAGGTATGTAATTTGATTGAACAATCGCGACTTTCTTCAACGC
GGTGTCTCCAATTTAGAATAACAAATACGTCGCGCCCGCGACAGCTCCGCTGGAGCGAGTTCAAGCGATT
CTGCGACATATTCAATATGGTGCTCGGGAAGGCCAGGATGGGCCGCGACCCGGGGCGTCCGGTGCGCGAT
GAACGTCGCATCGTCTCCTG

SS2 (Rv1509): SEQ ID NO: 2

>gi|448814763:1700212-1701093 Mycobacterium tuberculosis H37Rv complete genome

GTGTTTGCGTTGAGTAATAATCTGAACCGTGTGAACGCATGCATGGATGGATTCCTTGCCCGTATCCGCT
CACATGTTGATGCGCACGCGCCAGAATTGCGTTCACTGTTCGATACGATGGCGGCCGAGGCCCGATTTGC
ACGCGACTGGCTGTCCGAGGACCTCGCGCGGTTGCCTGTCGGTGCAGCATTGCTGGAAGTGGGCGGGGGG
GTACTTCTGCTCAGCTGTCAACTGGCGGCGGAGGGATTTGACATCACCGCCATCGAGCCGACGGGTGAAG
GTTTTGGCAAGTTCAGACAGCTTGGCGACATCGTGCTGGAATTGGCTGCA

SS3 (RV1954A): SEQ ID NO: 3

>gi|448814763:2201277-2201579 Mycobacterium tuberculosis H37Rv complete genome

ATGGCCAGGGGCAGGGTCGTGTGCATTGGGGATGCAGGTTGCGACTGTACACCCGGCGTGTTCCGCGCGA
CAGCGGGTGGGATGCCGGTGCTGGTGGTCATCGAGTCTGGGACAGGAGGTGATCAGATGGCTCGTAAAGC
TACGTCCCCGGGTAAGCCGGCTCCGACGTCGGGACAGTATCGCCCGGTTGGCGGTGGCAACGAGGTGACC
GTTCCGAAGGGACACCGTCTGCCTCCCTCGCCCAAGCCCGGTCAGAAGTGGGTGAACGTCGATCCGACGA

SS4 (Rv2231A): SEQ ID NO: 4

>gi|448814763:c2506161-2505736 Mycobacterium tuberculosis H37Rv complete genome

TTGACCATGGCGTGTACCGCGTGCCCGACGATTTGGACGCTCCGTTGTCAGACGACGTGCTCGAACGCTT
TCACCGGTGAAGCGCTACCTCATCGACACCCACGTTTGGCTGCGGATGCCGTCAACGAAACACGGGCGAT

TGTTCAGGACGTCCGCAACAGCATTCTCTTGTCGGCCGCCAGTGCCTGGGAGATCGCGATCAACTACCGC
CTCGGCAAGCTCCCGCCGCCCGAGCCATCGGCCTCTTACGTGCCCGATCGAATGCGCCGCTGCGGCACGT
CGCCGCTGTCAGTTGACCACGCACACACTGCGCACCGCAGAGCTTCCGGATCACCATCGACATCCATTCG
ACCGTGTGCTCATCGCCCAGGCACAGCTGCTTGGCCTGA

For the purposes of PCR validation, the signature sequences SS1, SS2, SS3 and SS4 were selected and oligonucleotide primers were designed to generate corresponding specific PCR amplification products. Table 3 provides the set of designed oligonucleotide primers.

Table 3: Signature sequences SS1, SS2, SS3 and SS4 and respective oligonucleotide primers

Signature Sequences (SS)	Sequence	Description	Prediction Class
SS1 from Rv1507A	>Rv1507A ATGCAATCAGGTCAAAATATCCTCGCC AAGGTATGTAATTTGATTGAACAATCG CGACTTTCTTCAACGCGGTGTCTCCAAT TTAGAATAACAAATACGTCGCGCCCGC GACAGCTCCGCTGGAGCGAGTTCAAGC GATTCTGCGACATATTCAATATGGTGCT CGGGAAGGCCAGGATGGGCCGCGACC CGGGGCGTCCGGTGCGCGATGAACGT CGCATCGTCTCCTG	NZE_Rv1954A_F ATGCAGGTTGCGACTGTA CACCCGG Length = 25, Tm = 58.6, %G+C = 60 NZE_Rv1954A_R GGCCGCTCTTGTTCTTCGT CGGAT Length = 24, Tm = 57.4, %G+C = 58.3 Amplicon Size = **~280 bp**	**CLASS 1**
SS2 from	>gi|448814763:1700212-1701093 Mycobacterium tuberculosis H37Rv	NZE_Rv1509_F GTGTTTGCGTTGAGTAAT	**CLASS 1**

Rv1509	complete genome GTGTTTGCGTTGAGTAATAATCTGAACC GTGTGAACGCATGCATGGATGGATTCC TTGCCCGTATCCGCTCACATGTTGATGC GCACGCGCCAGAATTGCGTTCACTGTT CGATACGATGGCGGCCGAGGCCCGATT TGCACGCGACTGGCTGTCCGAGGACCT CGCGCGGTTGCCTGTCGGTGCAGCATT GCTGGAAGTGGGCGGGGGGGTACTTC TGCTCAGCTGTCAACTGGCGGCGGAGG GATTTGACATCACCGCCATCGAGCCGA CGGGTGAAGGTTTTGGCAAGTTCAGAC AGCTTGGCGACATCGTGCTGGAATTGG CTGCA	AATCTGAACCGTGT Length = 32, Tm = 57.5%G+C = 41 NZE_Rv1509_R AGCCAATTCCAGCACGAT GTCGCC Length = 24, Tm = 58.8, %G+C = 58.3 Amplicon Size = ~**330bp**	
SS3 from **Rv1954A**	>gi|448814763:2201277-2201579 Mycobacterium tuberculosis H37Rv complete genome ATGGCCAGGGGCAGGGTCGTGTGCATT GGGGATGCAGGTTGCGACTGTACACCC GGCGTGTTCCGCGCGACAGCGGGTGG GATGCCGGTGCTGGTGGTCATCGAGTC TGGGACAGGAGGTGATCAGATGGCTC GTAAAGCTACGTCCCCGGGTAAGCCGG CTCCGACGTCGGGACAGTATCGCCCGG TTGGCGGTGGCAACGAGGTGACCGTTC CGAAGGGACACCGTCTGCCTCCCTCGC CCAAGCCCGGTCAGAAGTGGGTGAAC	NZE_ Rv1954A_F ATGCAGGTTGCGACTGTA CACCCGG Length = 25, Tm = 58.6, %G+C = 60 NZE_ Rv1954A_R GGCCGCTCTTGTTCTTCGT CGGAT Length = 24, Tm = 57.4, %G+C = 58.3	**CLASS 2**

	GTCGATCCGACGA	Amplicon Size = **~280 bp**	
SS4 from Rv2231A	>gi|448814763:c2506161-2505736 Mycobacterium tuberculosis H37Rv complete genome TTGACCATGGCGTGTACCGCGTGCCCG ACGATTTGGACGCTCCGTTGTCAGACG ACGTGCTCGAACGCTTTCACCGGTGAA GCGCTACCTCATCGACACCCACGTTTGG CTGCGGATGCCGTCAACGAAACACGGG CGATTGTTCAGGACGTCCGCAACAGCA TTCTCTTGTCGGCCGCCAGTGCCTGGG AGATCGCGATCAACTACCGCCTCGGCA AGCTCCCGCCGCCCGAGCCATCGGCCT CTTACGTGCCCGATCGAATGCGCCGCT GCGGCACGTCGCCGCTGTCAGTTGACC ACGCACACACTGCGCACCGCAGAGCTT CCGGATCACCATCGACATCCATTCGACC GTGTGCTCATCGCCCAGGCACAGCTGC TTGGCCTGAC	NZE_Rv2231A_F TGTACCGCGTGCCCGACG ATTTG Length= 23, Tm = 59.1, %G+C = 61 NZE_Rv2231A_R ACAGGCAGCTAACAGGG CGTCGG Length = 23, Tm = 57.1, %G+C = 65 Amplicon Size = ~**390 bp**	**CLASS 2**

In a preferred embodiment, pathogenic mycobacterium sps can be detected with 100 % specificity following PCR using DNA isolated from clinical samples from patients who presented with clinical symptoms of the disease. In another embodiment, pathogenic mycobacterium sps is also detected using the above method in clinical samples isolated from individuals vaccinated against tuberculosis. In yet another embodiment, pathogenic mycobacterium sps is also detected using the above method in clinical samples isolated from individuals treated against tuberculosis.

Pathogenic mycobacterium sps, as provided in the invention, includes *Mycobacterium tuberculosis* and *Mycobacterium bovis*. More specifically, pathogenic mycobacterium sps represents *Mycobacterium tuberculosis*, the TB causing bacterium.

Clinical samples, as meant here, includes specimens such as blood, sputum, cerebrospinal fluid, gastric lavage, tissue biopsies and the likes thereof. PCR product can be easily visualized by any conventional method that can be readily recognized by a person skilled in the art such as electrophoresis.

Following Examples serve as a tool to illustrate the invention. However, it should in no way be considered to be limiting the invention.

Example 1

Determination of specificity and sensitivity of signature sequences

Genomic DNA for PCR amplification

Genomic DNA of *Mycobacterium tuberculosis* and 13 other mycobacterial species were used for testing the specificity of signature sequences using PCR. These include, *M. avium* subspecies *paratuberculosis*, *M. smegmatis* (ATCC19420), *M. vaccae*, *M. marinum* (ATCC927), *M. chelonae* (ATCC14472), *M. flavescens* (ATCC14474), *M. fortuitum* (ATCC6481), *M. kansasii*(ATCC12478), *M. bovis* (ATCC27294), *M. bovis* (BCG), *M. avium* (ATCC25291), *M. gastri*, *M. indicuspranii*.

PCR Reaction

The PCR reaction mixture (50μl) consisted of 10xtaqPCR buffer, 0.5mmolMgCl2, 0.4mmol dNTP, 10 pmol forward and reverse primers respectively, 4% DMSO and 1Utaq DNA polymerase.The reaction conditions were the following: 95°C for 5minutes, followed by 35 cycles of 95°C for 30seconds, annealing temperature 50°C for 30seconds, 72° C for 1minute and finally 72°C for 10 minutes. All PCR products were electrophoresed on 2% agarose gel with ethidium bromide staining.

The "signature sequences" were tested for their ability to differentiate *Mycobacterium tuberculosis* DNA from a large number of other mycobacterial DNA in PCR using primers complementary to these "signature sequences" as shown in Table 3. For this purpose, chromosomal DNA extracted from 13 mycobacterium species including human genomic DNA were tested by mycobacteruim genus-specific primers of the 'signature sequences'. SS1 and SS2 were negative for all 13 mycobacterium species tested whereas SS3 and SS4 show positive PCR results only when *M. bovis* BCG genomic DNA was used as template.

Table 4 summarizes the specificity data resulting from PCR using specific primers of signature sequences SS1, SS2, SS3 and SS4.

Table 4: Specific amplification of signature sequences from *Mycobacterium tuberculosis* and *M. bovis* BCG

Specimen	SS1	SS2	SS3	SS4
M. tuberculosis	+	+	+	+
M.bovis BCG	-	-	+	+
M. avium	-	-	-	-
M. smegmatis	-	-	-	-
M. vaccae	-	-	-	-
M. avium	-	-	-	-
M. chelonae	-	-	-	-
M. flori	-	-	-	-
M. fortuitum	-	-	-	-
M. kansasi	-	-	-	-

M. bovis	-	-	-	-
M. marinum	-	-	-	-
M. gastri	-	-	-	-
MIP	-	-	-	-
MAP	-	-	-	-
M. leprae				
Human Genome	-	-	-	-

Furthermore, sensitivity analysis revealed that the signature sequences were highly sensitive in being able to detect <1 ng (100pg) DNA as shown in Fig. 1A, 1B, 1C and 1D for the four primers of the signature sequences SS1, SS2, SS3 and SS4 respectively.

Example 2

EVALUATION OF MYCOBACTERIUM TUBERCULOSIS-SPECIFIC PRIMER PAIR USING CLINICAL SAMPLES

A) Amplification of signature sequences from patient sputum samples

Sputum samples were processed by the Universal Sample Processing (USP) method for DNA extraction as described by Chakravorty et al (J Clin Microbiol 43, 4357, 2005). DNA was isolated from the USP sediments by boiling in the presence of five volumes of solution containing 10% Chelex-100 resin, 0.03% triton X-100, and 0.3% Tween 20. The isolated DNA was stored at 20˚c and used for PCR assay.

PCR reaction was carried out using specific primers as given in Table 3.

The results show amplification of signature sequences in patient sputum sample (Fig. 2) demonstrating the diagnostic utility of the signature sequences for detecting pathogenic *Mycobacterium tuberculosis*.

B) Amplification of signature sequences from patient blood samples

DNA from blood samples of tuberculosis patients were isolated as per the protocol described in van Helden et al (Isolation of DNA from Mycobacterium tuberculosis, Paul D. van Helden, Thomas C. Victor, Robin M. Warren, and Eileen G. van Helden)

The results show amplification of SS1 and SS3 as seen in Fig. 3.

Sequence Listing

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 4

(2) INFORMATION FOR SEQ ID NO: 1

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 230 base pairs

(B) TYPE: Nucleic Acid

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1, Rv1507A

1 ATGCAATCAG GTCAAAATAT CCTCGCCAAG GTATGTAATT TGATTGAACA ATCGCGACTT
61 TCTTCAACGC GGTGTCTCCA ATTTAGAATA ACAAATACGT CGCGCCCGCG ACAGCTCCGC
121 TGGAGCGAGT TCAAGCGATT CTGCGACATA TTCAATATGG TGCTCGGGAA GGCCAGGATG
181 GGCCGCGACC CGGGGCGTCC GGTGCGCGAT GAACGTCGCA TCGTCTCCTG

(2) INFORMATION FOR SEQ ID NO: 2

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 330 base pairs

(B) TYPE: Nucleic Acid

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2, Rv1509

1 GTGTTTGCGT TGAGTAATAA TCTGAACCGT GTGAACGCAT GCATGGATGG ATTCCTTGCC

61 CGTATCCGCT CACATGTTGA TGCGCACGCG CCAGAATTGC GTTCACTGTT CGATACGATG
121 GCGGCCGAGG CCCGATTTGC ACGCGACTGG CTGTCCGAGG ACCTCGCGCG GTTGCCTGTC
181 GGTGCAGCAT TGCTGGAAGT GGGCGGGGGG GTACTTCTGC TCAGCTGTCA ACTGGCGGCG
241 GAGGGATTTG ACATCACCGC CATCGAGCCG ACGGGTGAAG GTTTTGGCAA GTTCAGACAG
301 CTTGGCGACA TCGTGCTGGA ATTGGCTGCA

(2) INFORMATION FOR SEQ ID NO: 3

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 280 base pairs

(B) TYPE: Nucleic Acid

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3, RV1954A

1 ATGGCCAGGG GCAGGGTCGT GTGCATTGGG GATGCAGGTT GCGACTGTAC ACCCGGCGTG
61 TTCCGCGCGA CAGCGGGTGG GATGCCGGTG CTGGTGGTCA TCGAGTCTGG GACAGGAGGT
121 GATCAGATGG CTCGTAAAGC TACGTCCCCG GGTAAGCCGG CTCCGACGTC GGGACAGTAT
181 CGCCCGGTTG GCGGTGGCAA CGAGGTGACC GTTCCGAAGG GACACCGTCT GCCTCCCTCG
241 CCCAAGCCCG GTCAGAAGTG GGTGAACGTC GATCCGACGA

(2) INFORMATION FOR SEQ ID NO: 4

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 389 base pairs

(B) TYPE: Nucleic Acid

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4, Rv2231A

1 TTGACCATGG CGTGTACCGC GTGCCCGACG ATTTGGACGC TCCGTTGTCA GACGACGTGC
61 TCGAACGCTT TCACCGGTGA AGCGCTACCT CATCGACACC CACGTTTGGC TGCGGATGCC
121 GTCAACGAAA CACGGGCGAT TGTTCAGGAC GTCCGCAACA GCATTCTCTT GTCGGCCGCC
181 AGTGCCTGGG AGATCGCGAT CAACTACCGC CTCGGCAAGC TCCCGCCGCC CGAGCCATCG
241 GCCTCTTACG TGCCCGATCG AATGCGCCGC TGCGGCACGT CGCCGCTGTC AGTTGACCAC
301 GCACACACTG CGCACCGCAG AGCTTCCGGA TCACCATCGA CATCCATTCG ACCGTGTGCT
361 CATCGCCCAG GCACAGCTGC TTGGCCTGA

Claims

We claim:

1. A nucleotide sequence capable of selectively detecting pathogenic Mycobacterium sps using oligonucleotide primers corresponding to the signature sequence selected from SEQ ID NO: 1, 2, 3 or 4.

2. A method of detecting pathogenic mycobacterium sps in a clinical sample, said method comprising the steps of:

a. removal of contaminants from the clinical sample by conventional methods;

b. extraction of genomic DNA from the contaminant-free clinical sample;

c. designing a set of specific oligonucleotide primers capable of specifically detecting SEQ ID NO: 1, 2, 3 or 4 for use in RT-PCR;

d. analyzing PCR product by electrophoresis or specific probe nucleotide sequence complementary to SEQ ID NO: 1, 2, 3 or 4.

3. The method of claim 2 wherein the set of oligonucleotide primers are selected from

(i) 5' ATGCAGGTTGCGACTGTACACCCGG 3'

3' GGCCGCTCTTGTTCTTCGTCGGAT 5'

(ii) 5' GTGTTTGCGTTGAGTAATAATCTGAACCGTGT 3'

3' AGCCAATTCCAGCACGATGTCGCC 5'

(iii) 5' ATGCAGGTTGCGACTGTACACCCGG 3'

3' GGCCGCTCTTGTTCTTCGTCGGAT 5'

(iv) 5' TGTACCGCGTGCCCGACGATTTG 3'

3' ACAGGCAGCTAACAGGGCGTCGG 5'

(v) a set of oligonucleotide primers complementary to (i), (ii), (iii) or (iv)

or

(vi) a set of oligonucleotide primers comprising of sequence containing any 10 consecutive bases from one of the sequences selected from SEQ ID NO: 1, 2, 3 or 4.

4. A kit for detection of pathogenic mycobacterium sps in clinical samples, said kit comprising set of oligonucleotide primers selected from

(i) 5' ATGCAGGTTGCGACTGTACACCCGG 3'

3' GGCCGCTCTTGTTCTTCGTCGGAT 5'

(ii) 5' GTGTTTGCGTTGAGTAATAATCTGAACCGTGT 3'

3' AGCCAATTCCAGCACGATGTCGCC 5'

(iii) 5' ATGCAGGTTGCGACTGTACACCCGG 3'

3' GGCCGCTCTTGTTCTTCGTCGGAT 5'

(iv) 5' TGTACCGCGTGCCCGACGATTTG 3'

3' ACAGGCAGCTAACAGGGCGTCGG 5'

(v) a set of oligonucleotide primers complementary to (i), (ii), (iii) or (iv)

or

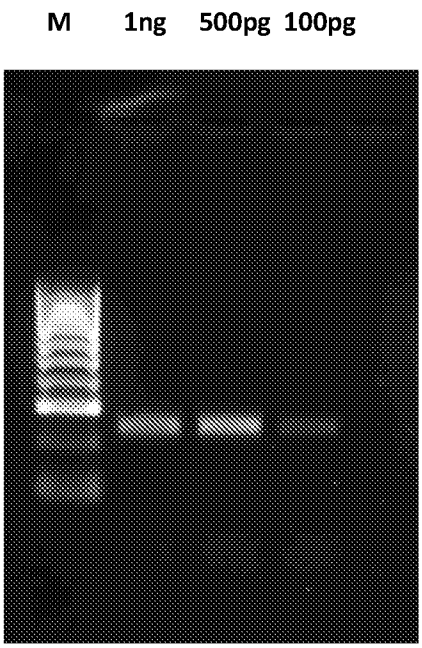
(vi) a set of oligonucleotide primers comprising of sequence containing any 10 consecutive bases from one of the sequences selected from SEQ ID NO: 1, 2, 3 or 4.

5. A method of detecting pathogenic mycobacterium sps in a clinical sample as in claim 2 wherein the sample is isolated from individuals vaccinated against tuberculosis.

6. A method of detecting pathogenic mycobacterium sps in a clinical sample as in claim 2 wherein the sample is isolated from individuals treated against tuberculosis.

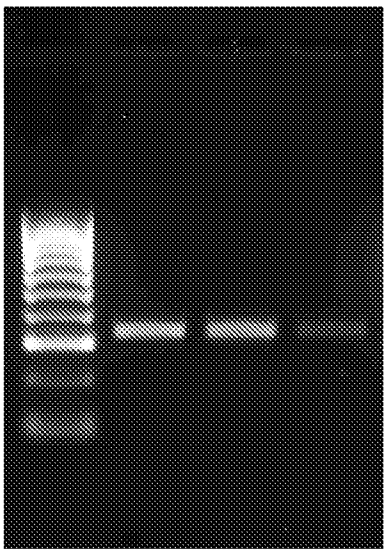
Figures

Fig. 1A



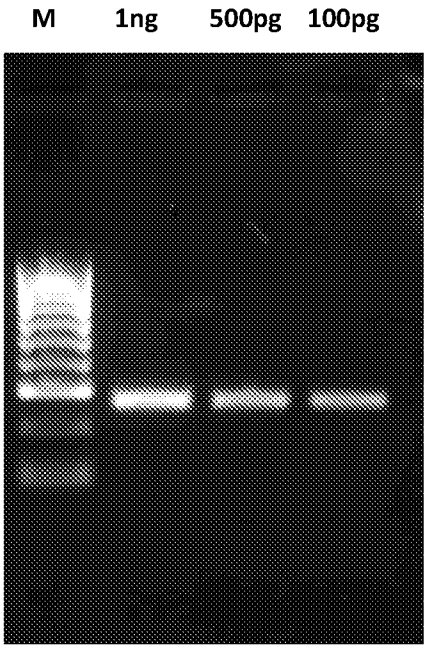
M: 100bp DNA ladder

Fig. 1B



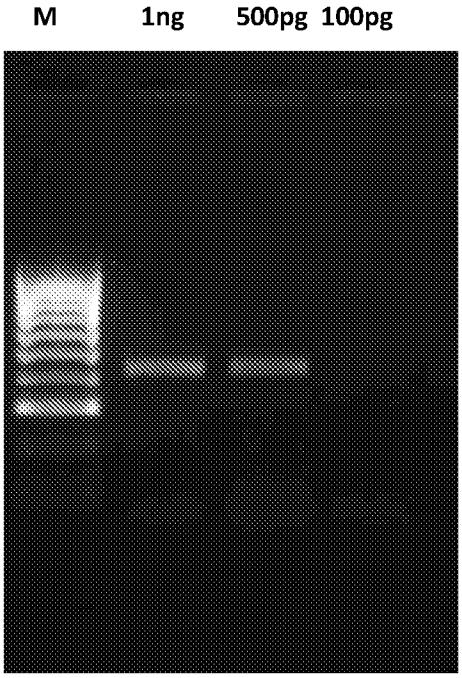
M: 100bp DNA ladder

Fig. 1C



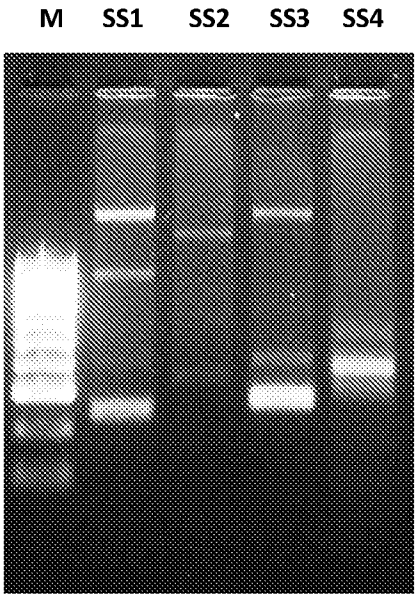
M: 100bp DNA ladder

Fig. 1D



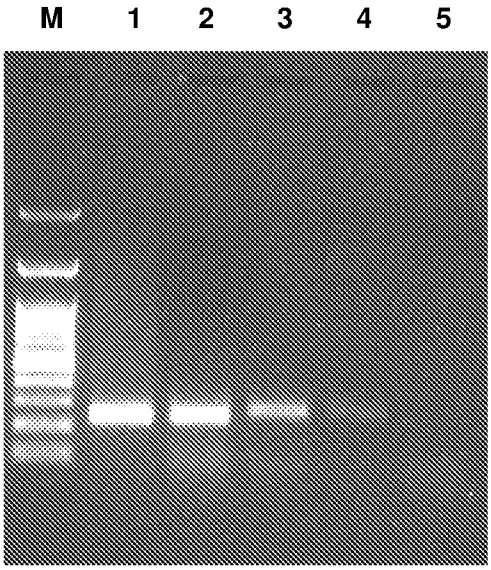
M: 100bp DNA ladder

Fig. 2



M: 100bp DNA ladder

Fig. 3



Lane M: 100bp DNA Ladder

Lane 1: Positive control (genomic DNA as template)

Lane 2: Rv1507A (patient1)

Lane 3: Rv1954A (patient2)

Lane 4: Rv1954a (Patient3)

Lane 5: Rv1509 (patient1)

PATENT COOPERATION TREATY

PCT

DECLARATION OF NON-ESTABLISHMENT OF INTERNATIONAL SEARCH REPORT

(PCT Article 17(2)(a), Rules 13ter.1(c) and Rule 39)

Applicant's or agent's file reference	IMPORTANT DECLARATION	Date of mailing *(day/month/year)*
1101PCT2014		24 April 2015 (24-04-2015)
International application No.	International filing date *(day/month/year)*	(Earliest) Priority date *(day/month/year)*
PCT/IB2014/066469	1 December 2014 (01-12-2014)	3 December 2013 (03-12-2013)
International Patent Classification (IPC) or both national classification and IPC		
C12Q1/68		
Applicant		
KUSUMA SCHOOL OF BIOLOGICAL SCIENCES		

This International Searching Authority hereby declares, according to Article 17(2)(a), that **no international search report will be established** on the international application for the reasons indicated below

1. ☐ The subject matter of the international application relates to:
   - a. ☐ scientific theories.
   - b. ☐ mathematical theories
   - c. ☐ plant varieties.
   - d. ☐ animal varieties.
   - e. ☐ essentially biological processes for the production of plants and animals, other than microbiological processes and the products of such processes.
   - f. ☐ schemes, rules or methods of doing business.
   - g. ☐ schemes, rules or methods of performing purely mental acts.
   - h. ☐ schemes, rules or methods of playing games.
   - i. ☐ methods for treatment of the human body by surgery or therapy.
   - j. ☐ methods for treatment of the animal body by surgery or therapy.
   - k. ☐ diagnostic methods practised on the human or animal body.
   - l. ☐ mere presentations of information.
   - m. ☐ computer programs for which this International Searching Authority is not equipped to search prior art.

2. ☐ The failure of the following parts of the international application to comply with prescribed requirements prevents a meaningful search from being carried out:
   ☐ the description ☐ the claims ☐ the drawings

3. ☒ The failure of the nucleotide and/or amino acid sequence listing to comply with the standard provided for in Annex C of the Administrative Instructions prevents a meaningful search from being carried out:
   ☒ the written form has not been furnished or does not comply with the standard.
   ☐ the computer readable form has not been furnished or does not comply with the standard.

4. Further comments:

Name and mailing address of the International Searching Authority	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL-2280 HV Rijswijk Tel. (+31-70) 340-2040 Fax: (+31-70) 340-3016	SMAJDA, Iveta Tel: +49 (0)89 2399-7942

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 203

Present application refers to nucleic acid sequences. These are provided as part of the description in typed form (see pages 19 and 20). With notification of 06.02.2015, the Applicant was invited to provide a sequence listing in electronic form to enable the ISA to perform a search for the subject-matter as claimed (Rule 13ter.1(a) PCT).
The Applicant failed to file the requested sequence listing in electronic form.
Following Rule 13ter.1(d) PCT, the ISA is in case of failure to provide the requested sequence listing only required to perform a search to the extent as a meaningful search can be performed without the said sequence listing.
In the present case, the ISA considers that all claims on file refer to sequence as identified by SEQ ID NO:1-4. Moreover, considering the description of the present application (page 2, lines 22-26), it would appear that said nucleic acid sequences represent the core element of the invention.
Taken together, the ISA considers that in the absence of a sequence listing in electronic form no meaningful search can be performed (Rule 13ter.1(d) PCT).

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guidelines C-IV, 7.2), should the problems which led to the Article 17(2) declaration be overcome.