*Journal of Medical Microbiology* (2012), **61**, 1352–1359 DOI 10.1099/jmm.0.044453-0

# Polymorphism in the RD1 locus and its effect on downstream genes among South Indian clinical isolates of *Mycobacterium tuberculosis*

Ahmed Kabir Refaya, Shanmugam Sivakumar, Balaji Sundararaman and Sujatha Narayanan

Department of Immunology, National Institute for Research in Tuberculosis, Chetput, Chennai 600 031, India

Correspondence
Sujatha Narayanan
suja_tha36@yahoo.co.in *or*
sujatha.sujatha36@gmail.com

Received 12 March 2012
Accepted 21 June 2012

RD1, the region of difference between the virulent strains of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG, is the most explored region in terms of mycobacterial virulence and vaccine design. This study found a polymorphic intergenic region between two genes, Rv3870 and Rv3871, in the RD1 region. Sequence analysis revealed a 53 bp repeat element that created a polymorphism among the clinical isolates, reported previously as the mycobacterial interspersed repetitive unit (MIRU) 39 locus. The discriminatory power of this locus was found to be high for EAI strains, as indicated by a Hunter–Gaston diversity index value of 0.58, and low for Beijing (0.26) and CAS (0.29) strains. The presence and variability of MIRU 39 in the intergenic region led us to investigate the functional role of the repeat element by measuring the transcription levels of the downstream genes Rv3871 and Rv3874 by quantitative RT-PCR among the different clades of clinical strains. Higher transcription levels of Rv3871 were observed in strains with four copies of the repeat element in the upstream region, whereas the transcription level of Rv3874 was higher in strains with six copies of the repeat element. These data suggest that changes in transcription levels resulting from insertion of different copy numbers of the repeat element may affect regulation of gene expression in *M. tuberculosis*.

## INTRODUCTION

*Mycobacterium tuberculosis*, one of the most successful pathogens in the world for thousands of years, remains a major public health burden around the globe in the human immunodeficiency virus/AIDS era. Whole-genome hybridization between *Mycobacterium bovis* BCG and virulent clinical strains of *M. tuberculosis* has identified a novel deletion in the genome of the attenuated strain named the region of difference (RD1) (Mahairas *et al.*, 1996; Talbot *et al.*, 1997; Behr *et al.*, 1999). Since the identification of this region, RD1 has become the most investigated and attractive locus for vaccine research, and also for understanding the pathogenesis of mycobacteria. There are nine predicted ORFs (Rv3871–Rv3879c) spanning the 9.5 kb RD1 region (Cole *et al.*, 1998; Behr *et al.*, 1999). Partial deletion of the RD1 region or insertion mutation in the Rv3874 (CFP10) gene, and mutations in Rv3870, Rv3871, Rv3876 and Rv3877 attenuate *M. tuberculosis* and reduce its cytolytic effects on macrophages and its spread within them. Gene disruption and deletion studies have demonstrated that Rv3871 is necessary for the secretion of ESAT6/CFP10 (Hsu *et al.*, 2003), whilst Stanley *et al.* (2003) illustrated that Rv3871 interacts directly with Rv3874 and this interaction is necessary for successful secretion of the latter protein.

Abbreviations: HGDI, Hunter–Gaston diversity index; MIRU, mycobacterial interspersed repetitive unit; VNTR; variable number of tandem repeats.

We have been exploring the genetic polymorphisms of clinical isolates of *M. tuberculosis* using various genetic markers such as IS*6110* RFLP, spoligotyping and deletion microarrays (Narayanan *et al.*, 2008). South Indian *M. tuberculosis* strains have a lower number of IS*6110* copies and have been reported previously to be low in virulence compared with *M. tuberculosis* isolates from London in a guinea pig model (Mitchison *et al.*, 1960). Rao *et al.* (2005) used a single set of primers to amplify the 9.2 kb RD1 region and showed that all clinical strains of Indian origin tested in their study lacked this region. This is technically demanding, and any mutations in the primer-binding site might give false-negative results. Therefore, Soman *et al.* (2007) checked whether the RD1 genes were really absent in clinical strains of Indian origin by using three sets of primers to amplify the coding regions alone. The results of the above two conflicting reports from India tempted us to explore polymorphisms of the RD1 genes in our isolates, which were reported previously to be of low virulence (Mitchison *et al.*, 1960).

 

During screening of the RD1 region, we found a polymorphic intergenic region comprising a 53 bp repeat element between the two genes Rv3870 and Rv3871, which was identified previously as the mycobacterial interspersed repetitive unit (MIRU) 39 locus and was excluded from analysis of 15 loci by mycobacterial interspersed repetitive unit-variable number of tandem repeat (MIRU-VNTR) studies due to its low Hunter–Gaston discriminatory index (HGDI) and low allelic diversity index. The functional role of this repeat element in transcription and translation has been studied in a few pathogenic bacteria. Such repeat elements residing in coding regions may result in both frameshift and non-frameshift mutations, allowing the bacteria to adapt optimally to different environments. Repeat elements have been implicated in the pathogenesis of *M. tuberculosis* and other pathogenic bacteria (Whatmore, 2001; Erwin *et al.*, 2006). For example, polymorphisms in the VNTR 3690 locus of *M. tuberculosis* located in the intergenic region between Rv3304 and Rv3303c resulted in 12.5-fold upregulation of *lpdA* expression (Akhtar *et al.*, 2009). Therefore, we investigated the presence or absence of the RD1 genes in south Indian strains and also studied whether this polymorphism had any effect on the transcription of downstream genes in the RD1 region.

## METHODS

**Reference strains and clinical isolates.** *M. tuberculosis* H37Rv and *M. bovis* BCG were used as reference strains. A total of 407 mycobacterial isolates received from pulmonary tuberculosis patients as part of the model DOTS project from the Tiruvallur district, south India, between 1999 and 2003, were included in this study. Sputum samples were processed using a modified Petroff's method, cultured on Lowenstein–Jensen medium and examined weekly for up to 8 weeks for growth (Canetti *et al.*, 1969). Positive cultures were subjected to identification tests – niacin production, growth in Lowenstein–Jensen medium containing 500 mg *p*-nitrobenzoic acid $l^{-1}$ and catalase production at 68 °C – to identify the organism as *M. tuberculosis* or non-tuberculous mycobacteria.

**Genotyping by spoligotyping.** Genomic DNA from the isolates was extracted using a cetyl trimethylammonium bromide/NaCl method described elsewhere (Baess, 1974). Spoligotyping was performed with a commercially available kit (Ocimum Biosolutions) according to the manufacturer's instructions. The analysis defined 'related strains' as two or more isolates with identical spoligotyping patterns, and these patterns were analysed using Spotclust software (http://tbinsight.cs.rpi.edu/about_spotclust.html). The most common strain in our study region is the EAI strain, which can be defined as isolates lacking spots 29, 30, 31 and 32 by spoligotyping. EAI 3 strains lack spots 2, 3, 37, 38 and 39 in addition to the family characteristic spoligotype, whilst EAI 5 strains lack spot 34 in addition to the family characteristic spoligotype.

**PCR amplification for genes in RD1.** Initially, five sets of primers were designed to amplify the intergenic regions between genes in the RD1 region along with a few hundred bp of the flanking genes. Another set of internal primers for the RD1A region was designed to amplify exactly the intergenic region between Rv3870 and Rv3871 genes, which was named RD-INS. Primer sequences, annealing conditions, expected product sizes and the genomic region they amplified are given in Table 1 and shown in Fig. 1. The PCR mix contained final concentrations of 20 pmol each primer, 250 μM each dNTP, 1× PCR buffer (Invitrogen), 1.5 U *Taq* polymerase (Invitrogen) and 1.5 mM $MgCl_2$, made up to 25 μl with nuclease- and protease-free water (Gibco). Genomic DNA (1–10 ng) from the clinical isolates served as template. PCR amplicons were resolved by 1.5 % agarose gel electrophoresis, except for the RD-INS amplicons, which were resolved in a 2 % agarose gel, and visualized on a UV trans-illuminator after staining with ethidium bromide.

**Automated DNA sequencing.** PCR products were gel purified using Amersham GFX (Wipro GE Healthcare) columns according to the manufacturer's protocol, and the DNA concentration was determined by comparison with known standards following agarose gel electrophoresis. Approximately 200 ng DNA was cycle sequenced using 3 pmol primer with a BigDye Terminator Ready Reaction Mix (Applied Biosystems), and excess unincorporated dye was removed using a DyeEX kit (Qiagen) following the manufacturer's instructions. The final product was vacuum dried, suspended in 13 μl HiDi formamide and sequenced using an ABI PRISM 310 Genetic Analyzer (Applied Biosystems). Sequenced data were analysed using DNA Sequencing Analysis Software (Applied Biosystems). DNA sequences were analysed using GeneTools software (http://www.syngene.com/genetools/).

**RNA isolation and real-time PCR.** All strains were grown to exponential growth phase in Middlebrook 7H9 broth containing albumin/dextrose saline and 0.05 % Tween 80. Trizol reagent (Invitrogen) was added to a cell pellet obtained from 20–30 ml culture and disrupted with 0.1 mm zirconia beads in a mini-bead beater. Total RNA was purified using an RNeasy purification kit (Qiagen). Contaminated DNA in the RNA sample was digested with RNase-free DNase I. The purity of the RNA was determined by measuring the absorbance at 260 and 280 nm. The first-strand cDNA was synthesized from 1 μg total RNA using an Improm Reverse Transcriptase II kit (Promega).

Real-time quantitative RT-PCR (qRT-PCR) for Rv3871 and Rv3874 was performed in an ABI 7500 system (Applied Biosystems) using *Taq*Man assays. 16S rRNA was used for normalization. Each reaction was repeated three times with independent RNA samples. Negative controls consisting of no reverse transcriptase and no template mixtures were run with all reactions. After baseline correction and determination of threshold settings, calculations were carried out using the $2^{-\Delta\Delta C_t}$ method of Livak & Schmittgen (2001), as PCR efficiencies were found to be similar. Results were expressed as fold induction, which denoted the fold change in expression of the gene.

**Statistical analysis.** The HGDI described by Hunter & Gaston (1988) was used as a numerical index for RD1A locus discriminatory power. The HGDI was calculated using the following formula:

$$HGDI = 1 - \left[\frac{1}{N(N-1)}\sum_{j=1}^{s} n_j(n_j-1)\right]$$

where $N$ was the total number of strains in the typing scheme, $s$ was the total number of different patterns in locus RD1A and $n_j$ was the number of strains belonging to the $j$th pattern. A one-way analysis of variance table with Tukey's multiple comparison tests was used to calculate the change in fold induction between each group using GraphPad Prism software.

## RESULTS

PCR conditions for the five sets of primers were standardized using *M. tuberculosis* H37Rv and *M. bovis* BCG

**Table 1.** Primer sequences used in the study and regions they amplify

| Fragment name | Primer or probe name | Sequence (5′→3′) | Annealing conditions (°C/min) | Expected product size (bp) | Genomic coordinates* | Genes amplified |
|---|---|---|---|---|---|---|
| RD1A | RD1AF | CCCCCGCTGCCAACGCTTTT | 60/1 | 810 | 4348245−4349055 | Rv3870–Rv3871 |
| | RD1AR | CGGCCCCGGAAACGTCTACGC | | | | |
| RD1B | RD1BF | CCCCCAAGCGCCGGTTAAGA | 60/1 | 736 | 4350585–4351321 | Rv3872–Rv3873 |
| | RD1BR | TGGCCTGTGTTGACGCGGTTT | | | | |
| RD1C | RD1CF | CGACTGGGACGAAGAGGACGAC | 58/1 | 714 | 4352154–4352868 | Rv3874–Rv3875 |
| | RD1CR | TCGGTCGAAGCCATTGCCTGA | | | | |
| RD1D | RD1DF | CACGGGATCGGGCGAGTTCG | 60/1 | 718 | 4353645–4352928 | Rv3875–Rv3876 |
| | RD1DR | GGCCGGGGACGCAGACG | | | | |
| RD1E | RD1EF | CGGTGTCGTGGTCGTGGCAAGT | 60/1 | 775 | 4354671–4355446 | Rv3876–Rv3877 |
| | RD1ER | GCCCCCACAAAGCGATTCAATG | | | | |
| RDIN | RDINF | CGCGCGCATTCACAGGTTCACC | 60/1 | ~180–420 | 4348667–4348897 | Rv3870 –Rv3871 |
| | RDINR | CGCGATTCAGCAGTGCCGAGC | | | | |
| **qRT-PCR primers** | | | | | | |
| RV3871 | Forward | AGCTTGGACCCGTGCG | | | | |
| | Reverse | ACCCCCGCGGTGATC | | | | |
| | Fluorogenic probe | FAM-CAATAGCCGCGACAC | | | | |
| RV3871 | Forward | TAGACCCAAAAGTCCAGCGG | | | | |
| | Reverse | TGCTTATTGGCTGCTTCTTGGAA | | | | |
| | Fluorogenic probe | FAM-GTCCAAGCAACGTCCC | | | | |

*Coordinates as given by the *M. tuberculosis* H37Rv genome sequence in the TubercuList database (http://genolist.pasteur.fr/TubercuList/).

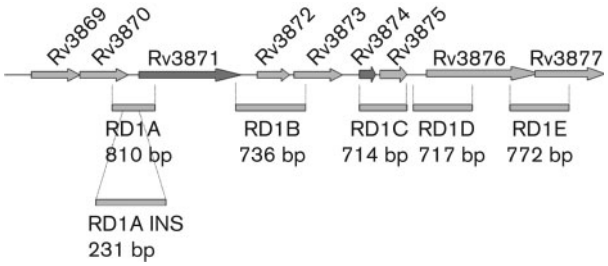


**Fig. 1.** Diagram of the RD1 region. Light grey arrows represent the genes present in the RD1 region, dark grey arrows indicate the genes used for quantitative RT-PCR and lines show the positions of the primers.

genomic DNA as templates. Seventy clinical isolates were selected randomly and analysed for deletions and polymorphisms in the RD1 region. There was no major insertion or deletion polymorphism observed in RD1B, -C, -D or -E, as all the clinical isolates tested generated the expected product sizes of 736, 714, 718 and 775 bp, respectively. The standardized PCR conditions for the primers (RD1-A, -B, -C, -D and -E) are shown in Table 1.

We observed significant polymorphisms in the RD1A region, which comprises part of Rv3870 and Rv3871 and the intergenic region between them in the clinical isolates. The PCR product size varied from approximately 50 to 200 bp, as determined by 1.5 % agarose gel electrophoresis. These products were further sequenced, and multiple alignments were performed using GeneTools software. The intergenic region between Rv3870 and Rv3871 is illustrated as a multiple sequence alignment in Fig. 2. Multiple sequence analysis revealed that there was a 53 bp direct repeat element that occurred with a frequency of one to six repeats. At the 5′ end, 6 bp of the first repeat (CCGTGA) was different from that of the second to the sixth repeats (CAACGG). Single-nucleotide polymorphisms were observed in the repeat elements, and were confirmed by sequencing with forward and reverse primers.

In the case of the RD1A primers, the discriminatory power using 1.5 % agarose gel electrophoresis for the PCR product of ~750–950 bp was comparatively poor and there was a

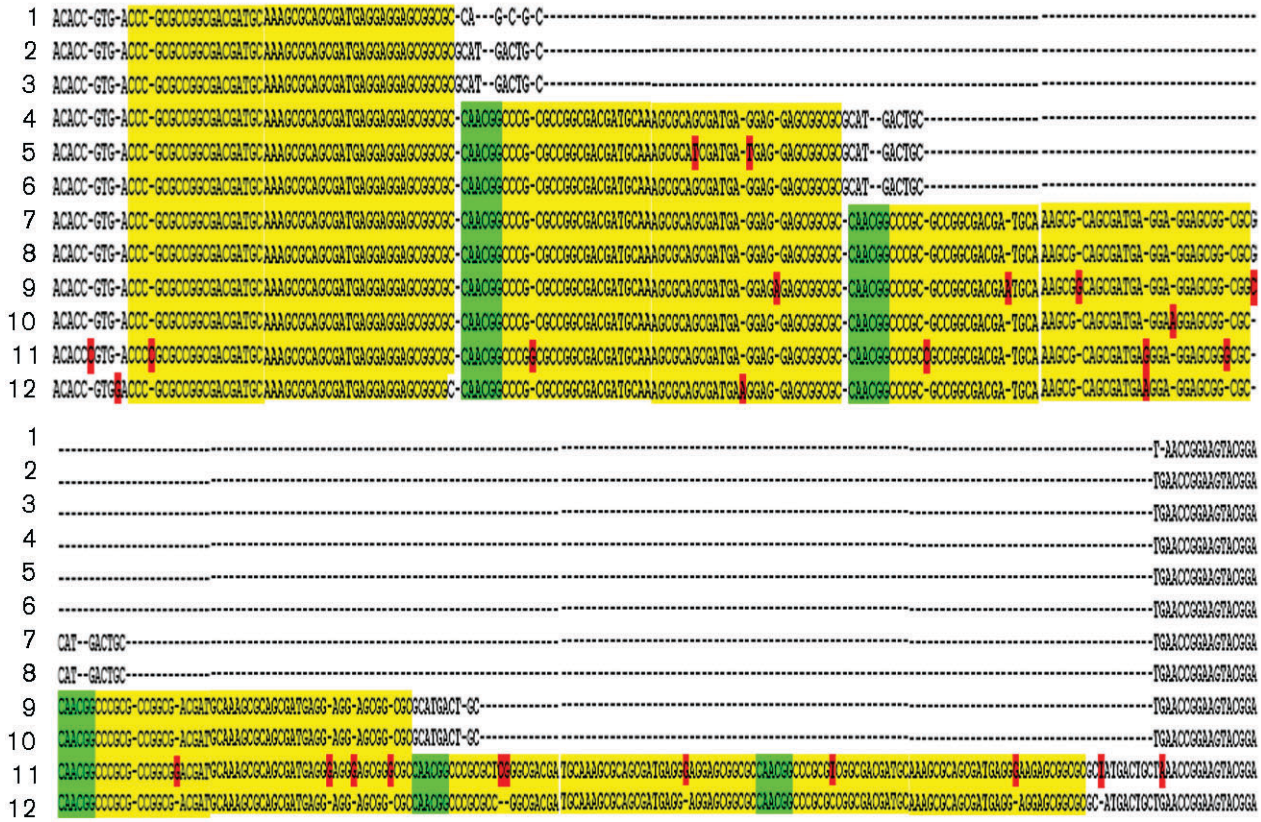

**Fig. 2.** Sequences of PCR products from the RD1A region from various clinical isolates, aligned using GeneTools software as described in Methods. Mismatches are highlighted in red, the 47 bp repeat unit is in yellow and the 6 bp inter-repeat is in green. 1, Consensus sequence; 2 and 3, H1: M0296 and M1346; 4, H37Rv; 5 and 6, H2: M0432 and M0669; 7 and 8, H3: M0030 and M0370; 9 and 10, H4: M0255 and M1520; 11 and 12, H5: M1507 and M1745.

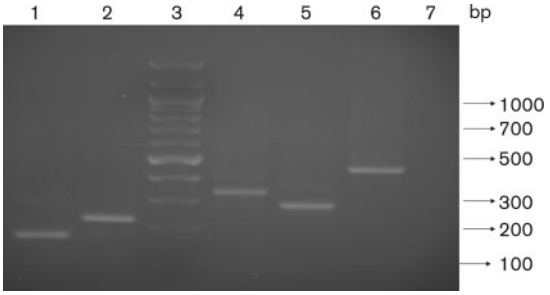


**Fig. 3.** PCR products of selected clinical isolates. Lanes: 1, 2 and 4–6, PCR products of clinical isolates representing the M0296 (H1, ~180 bp), M0432 (H2, ~230 bp), M0030 (H3, ~280 bp), M0255 (H4, ~330 bp) and M1507 (H5, ~420 bp) clusters, respectively; 3, 100 bp marker; 7, negative control.

bias in determining the correct size, as two independent observers read the size of the PCR product differently. As the sequencing results showed, the polymorphism observed was only in the intergenic region of Rv3870 and Rv3871. Hence, a set of primers was designed in the flanking region that would amplify the last 57 bp of Rv3870, the 103 bp intergenic region and the first 70 bp of Rv3871. The standard amplification product for *M. tuberculosis* H37Rv using these primers was 230 bp and was named RD-INS. We repeated the PCR using this set of primers for the 70 test samples, which were already differentiated into four clusters by the RD1A primers. The RD-INS primers differentiated these samples into five groups, namely H1 (~180 bp), H2 (~230 bp), H3 (~280 bp), H4 (~330 bp) and H5 (~420 bp) when resolved in a 2 % agarose gel (Fig. 3). This indicated that there was an ~50 bp deletion (H1) or an ~50 bp insertion (H3), ~100 bp insertion (H4) or ~200 bp insertion (H5) in the intergenic region compared with H37Rv (H2). The insertion of ~200 bp is a novel finding in this study.

We analysed a total of 407 clinical isolates from four major families namely, EAI3 (102), EAI5 (93), Beijing (77) and CAS (84), and 51 samples from the 33, T and LAM families (Table 2) with the RD1-INS primers to look for the frequency of polymorphisms. Almost two-thirds (266/407) of the total tested clinical isolates had a 53 bp insertion ($P<0.0001$), whilst 80 % of the CAS and Beijing isolates had this 53 bp insertion. The clinical isolates belonging to the EAI5 family had a higher percentage of the 53 bp deletion ($P<0.001$, compared with other samples), whilst only the clinical isolates with the EAI3 signature had the 200 bp insertion. There was no significant difference in terms of insertions or deletions between the CAS, Beijing, 33, LAM and T1 families.

The HGDI was calculated for each spoligotype cluster and was found to be 0.48, 0.61, 0.29 and 0.26 for the EAI3, EAI5, CAS and Beijing families, respectively. The HGDI for the EAI family was 0.58.

### Expression of Rv3871 and Rv3874 among the different groups

The expression of Rv3871 and Rv3874 was monitored in six isolates each from the H1, H2, H3 and H4 groups and three isolates from the H5 group by qRT-PCR using gene-specific primers. RNA was extracted from these isolates, and qRT-PCR was performed using 16S rRNA as an internal control for normalization of transcriptional analysis. Expression of Rv3871 was found to be 8- and 12-fold higher in samples containing two and four copies of the repeat element, respectively (Fig. 4a), whilst Rv3874 showed very high expression with a >40-fold increase in samples containing six copies of the repeat element compared with samples having only one copy (Fig. 4b). The results observed were statistically significant ($P<0.05$), suggesting an association between the number of repeat elements and transcription of the downstream genes.

## DISCUSSION

We assessed the presence of RD1 genes in south Indian isolates of *M. tuberculosis* and observed that there were no large sequence alterations or polymorphisms in the coding genes of the RD1 region except for single-nucleotide polymorphisms. In the present study, we observed the presence of a 53 bp tandem repeat in the intergenic region between Rv3870 and Rv3871. The number of repeats of MIRU 39 differed among *M. tuberculosis* isolates. The internal primer used in the study, which amplified the

**Table 2.** Comprehensive results of strain differentiation by the RD-INS primer

Values shown in bold are statistically significant by $\chi^2$ test at $P<0.05$.

| RD-INS | Family | | | | | Total ($n$=407) |
|---|---|---|---|---|---|---|
| | EAI3 ($n$=102) | EAI5 ($n$=93) | CAS ($n$=84) | Beijing ($n$=77) | Others ($n$=51) | |
| H5 | 3 (2.9 %) | 0 (0 %) | 0 (0 %) | 0 (0 %) | 0 (0 %) | 3 (0.7 %) |
| H4 | 11 (10.8 %) | 0 (0 %) | **3 (3.6 %)** | **6 (7.8 %)** | 1 (1.9 %) | 21 (5.2 %) |
| H3 | 75 (73.5 %) | 43 (46.2 %) | **69 (82.1 %)** | **66 (85.7 %)** | 13 (25.6 %) | **266 (65.4 %)** |
| H2 | 13 (12.8 %) | 36 (38.7 %) | 11 (13.1 %) | 5 (6.5 %) | 34 (66.6 %) | 99 (24.3 %) |
| H1 | 0 (0 %) | **14 (15.1 %)** | 1 (1.2 %) | 0 (0 %) | 3 (5.9 %) | 18 (4.4 %) |

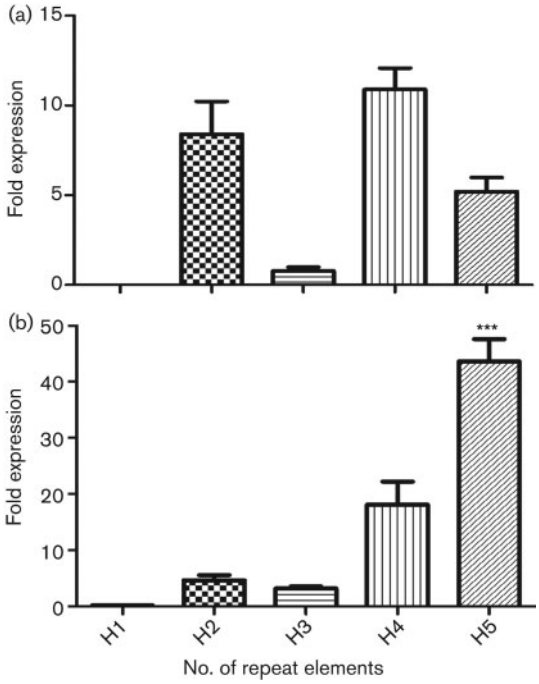


**Fig. 4.** Transcriptional analysis of Rv3871 (a) and Rv3874 (b) expression among *M. tuberculosis* clinical strains with different copy numbers of MIRU 39. (a) *M. tuberculosis* clinical strains with two (H2) and four (H4) copies showed higher expression compared with the others. (b) *M. tuberculosis* clinical strains with four (H4) and six (H5) copies showed higher expression levels compared with the others. ***$P<0.05$.

intergenic region along with short regions of the flanking genes, has an advantage over the other sets of primers that amplify several hundred bases of the flanking genes.

In order to characterize this repeat element, we used BLAST to analyse the sequence in TubercuList and found that the repeat corresponded to the MIRU 39 locus. The MIRU 39 coordinate in the *M. tuberculosis* H37Rv chromosome is between 4348718 and 4348823 (Supply *et al.*, 2000). These authors reported that MIRU 39 has a 53 bp repeat element occurring at a frequency of one to four copies, but we observed this frequency to be one to six copies, which is a novel finding. The standard MIRU-VNTR typing protocol used by these authors included 12 and 15 MIRU loci, which did not include MIRU 39 (Supply *et al.*, 2006; Phyu *et al.*, 2009). The HGDI of the EAI isolates for MIRU 39 was 0.58, which is higher than the previous report of 0.41 (Supply *et al.*, 2006). Thus, MIRU 39 should be included in the MIRU-VNTR typing protocol for the south Indian strains where EAI isolates correspond to 84 % of the total circulating strains (Shanmugam *et al.*, 2011). Our findings also suggest that the MIRU polymorphism may vary among geographically diverse clinical isolates, so region-specific MIRU locus standardization will be necessary to achieve a high discriminatory index as reported previously (Comas *et al.*, 2009). Here, we demonstrated that MIRU 39 occupying the intergenic region between Rv3870 and Rv3871 is present in variable copy numbers ranging from one to six.

It has already been reported that genome-wide insertion is prominent in the Beijing strain (McEvoy *et al.*, 2007). We also observed that 93.5 % of the tested Beijing strains prevalent in south India retained insertions of either 53 or 100 bp compared with the standard strain. Insertion of the 53 bp repeat in the EAI3 family was similar to that of the Beijing and CAS family, whilst the EAI5 strains had fewer insertions and more deletions common among ancient strains, as reported by Gutierrez *et al.* (2006). Two-thirds of the tested strains retained the 53 bp insertion, stressing the functional importance of this repeat element and its effect on the flanking genes.

There are several well-documented examples of the role of VNTRs and point mutations in gene regulation of bacteria, most notably in *Haemophilus influenzae* and *Neisseria* and *Streptococcus* species, in which changes in the copy number of VNTRs results in phase variation and changes in pathogenic properties (van Belkum *et al.*, 1998; Moxon *et al.*, 2006). As many VNTRs occupy the entire region between two genes, it is also possible that they may contain the entire regulatory element, which will consequently be repeated in multiple but variable copies among clinical isolates (Tantivitayakul *et al.*, 2010). Pinto Júnior *et al.* (2007) used PCR amplification of the intergenic region between the *plcB* and *plcC* genes to differentiate *M. tuberculosis* from *M. bovis*. They also tested these primers to detect and differentiate clinical isolates of the *M. tuberculosis* complex. Point-mutation polymorphisms in intergenic regions have been shown to affect drug susceptibility and resistance in *M. tuberculosis*. Sreevatsan *et al.* (1997) demonstrated a polymorphism in the 105 bp *oxyR–ahpC* intergenic region resulting in altered expression of KatG and correlated with resistance to isoniazid (Sreevatsan *et al.*, 1997). Ramaswamy *et al.* (2000) illustrated a polymorphic intergenic region between the *embC* and *embA* genes involved in ethambutol resistance (Ramaswamy *et al.*, 2000).

There is a good correlation between VNTR copy number and transcription of the genes, as cited by Akhtar *et al.* (2009), where there was a 12.5-fold upregulation in the expression of *lpdA* with a flanking region containing four VNTR 3690 repeats compared with a region containing just one repeat (Akhtar *et al.*, 2009). It has also been reported that VNTR 0960c, which occupies the intergenic region between Rv0862c and *errc3*, provided promotion for expression of a downstream *gfp* reporter gene (Tantivitayakul *et al.*, 2010). Similarly, we also observed a significant difference in the gene expression of Rv3871 and Rv3874 among the different groups containing different copy numbers of MIRU 39. As the study by Tantivitayakul *et al.* (2010) revealed, multiple copies might have an effect on the level of expression, albeit in an indirectly proportionate manner. Our study also found differential expression patterns of the two genes with respect to copy number. Thus, the expression level of Rv3871 was

higher in strains containing two and four copies when compared with one copy of the repeat element, whereas Rv3874 expression levels were proportionate to the increase in the number of copies present in the intergenic region.

Thus, mutations and polymorphisms in the intergenic region have been found to influence the expression of flanking genes and consequently their associated functions. From our results, we conclude that the presence of repeat elements has an effect on the transcription levels of Rv3871 and Rv3874, thereby influencing the regulation of gene expression in *M. tuberculosis*.

## ACKNOWLEDGEMENTS

S. B. and A. K. R. are grateful to Department of Biotechnology (DBT) and International Center for Excellence in Research (ICER) and Indian Council of Medical Research (ICMR) for the Research Fellowship. We thank Ms Aarthy for her help during this project. We acknowledge the support from the Department of Bacteriology of National Institute for Research in Tuberculosis (NIRT) for providing samples.

## REFERENCES

**Akhtar, P., Singh, S., Bifani, P., Kaur, S., Srivastava, B. S. & Srivastava, R. (2009).** Variable-number tandem repeat 3690 polymorphism in Indian clinical isolates of *Mycobacterium tuberculosis* and its influence on transcription. *J Med Microbiol* **58**, 798–805.

**Baess, I. (1974).** Isolation and purification of deoxyribonucleic acid from mycobacteria. *Acta Pathol Microbiol Scand B Microbiol Immunol* **82**, 780–784.

**Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S. & Small, P. M. (1999).** Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**, 1520–1523.

**Canetti, G., Fox, W., Khomenko, A., Mahler, H. T., Menon, N. K., Mitchison, D. A., Rist, N. & Smelev, N. A. (1969).** Advances in techniques of testing mycobacterial drug sensitivity, and the use of sensitivity tests in tuberculosis control programmes. *Bull World Health Organ* **41**, 21–43.

**Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S. & other authors (1998).** Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544.

**Comas, I., Homolka, S., Niemann, S. & Gagneux, S. (2009).** Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS ONE* **4**, e7815.

**Erwin, A. L., Bonthuis, P. J., Geelhood, J. L., Nelson, K. L., McCrea, K. W., Gilsdorf, J. R. & Smith, A. L. (2006).** Heterogeneity in tandem octanucleotides within *Haemophilus influenzae* lipopolysaccharide biosynthetic gene *losA* affects serum resistance. *Infect Immun* **74**, 3408–3414.

**Gutierrez, M. C., Ahmed, N., Willery, E., Narayanan, S., Hasnain, S. E., Chauhan, D. S., Katoch, V. M., Vincent, V., Locht, C. & Supply, P. (2006).** Predominance of ancestral lineages of *Mycobacterium tuberculosis* in India. *Emerg Infect Dis* **12**, 1367–1374.

**Hsu, T., Hingley-Wilson, S. M., Chen, B., Chen, M., Dai, A. Z., Morin, P. M., Marks, C. B., Padiyar, J., Goulding, C. & other authors (2003).** The primary mechanism of attenuation of bacillus Calmette–Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proc Natl Acad Sci U S A* **100**, 12420–12425.

**Hunter, P. R. & Gaston, M. A. (1988).** Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J Clin Microbiol* **26**, 2465–2466.

**Livak, K. J. & Schmittgen, T. D. (2001).** Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods* **25**, 402–408.

**Mahairas, G. G., Sabo, P. J., Hickey, M. J., Singh, D. C. & Stover, C. K. (1996).** Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J Bacteriol* **178**, 1274–1282.

**McEvoy, C. R., Falmer, A. A., Gey van Pittius, N. C., Victor, T. C., van Helden, P. D. & Warren, R. M. (2007).** The role of IS6110 in the evolution of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* **87**, 393–404.

**Mitchison, D. A., Wallace, J. G., Bhatia, A. L., Selkon, J. B., Subbaiah, T. V. & Lancaster, M. C. (1960).** A comparison of the virulence in guinea-pigs of South Indian and British tubercle bacilli. *Tubercle* **41**, 1–22.

**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.

**Narayanan, S., Gagneux, S., Hari, L., Tsolaki, A. G., Rajasekhar, S., Narayanan, P. R., Small, P. M., Holmes, S. & Deriemer, K. (2008).** Genomic interrogation of ancestral *Mycobacterium tuberculosis* from south India. *Infect Genet Evol* **8**, 474–483.

**Phyu, S., Stavrum, R., Lwin, T., Svendsen, O. S., Ti, T. & Grewal, H. M. (2009).** Predominance of *Mycobacterium tuberculosis* EAI and Beijing lineages in Yangon, Myanmar. *J Clin Microbiol* **47**, 335–344.

**Pinto Júnior, H., Giuliano Bica, C., Palaci, M., Dietze, R., Basso, L. A. & Santiago Santos, D. (2007).** Using polymerase chain reaction with primers based on the *plcB-plcC* intergenic region to detect *Mycobacterium tuberculosis* in clinical samples. *Braz J Pulmonol* **33**, 437–442.

**Ramaswamy, S. V., Amin, A. G., Göksel, S., Stager, C. E., Dou, S. J., El Sahly, H., Moghazeh, S. L., Kreiswirth, B. N. & Musser, J. M. (2000).** Molecular genetic analysis of nucleotide polymorphisms associated with ethambutol resistance in human isolates of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **44**, 326–336.

**Rao, K. R., Kauser, F., Srinivas, S., Zanetti, S., Sechi, L. A., Ahmed, N. & Hasnain, S. E. (2005).** Analysis of genomic downsizing on the basis of region-of-difference polymorphism profiling of *Mycobacterium tuberculosis* patient isolates reveals geographic partitioning. *J Clin Microbiol* **43**, 5978–5982.

**Shanmugam, S., Selvakumar, N. & Narayanan, S. (2011).** Drug resistance among different genotypes of *Mycobacterium tuberculosis* isolated from patients from Tiruvallur, South India. *Infect Genet Evol* **11**, 980–986.

**Soman, S., Joseph, B. V., Sarojini, S., Kumar, R. A., Katoch, V. M. & Mundayoor, S. (2007).** Presence of region of difference 1 among clinical isolates of *Mycobacterium tuberculosis* from India. *J Clin Microbiol* **45**, 3480–3481.

**Sreevatsan, S., Pan, X., Zhang, Y., Deretic, V. & Musser, J. M. (1997).** Analysis of the *oxyR-ahpC* region in isoniazid-resistant and -susceptible *Mycobacterium tuberculosis* complex organisms recovered from diseased humans and animals in diverse localities. *Antimicrob Agents Chemother* **41**, 600–606.

**Stanley, S. A., Raghavan, S., Hwang, W. W. & Cox, J. S. (2003).** Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proc Natl Acad Sci U S A* **100**, 13001–13006.

**Supply, P., Mazars, E., Lesjean, S., Vincent, V., Gicquel, B. & Locht, C. (2000).** Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Mol Microbiol* **36**, 762–771.

**Supply, P., Allix, C., Lesjean, S., Cardoso-Oelemann, M., Rüsch-Gerdes, S., Willery, E., Savine, E., de Haas, P., van Deutekom, H. & other authors (2006).** Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* **44**, 4498–4510.

**Talbot, E. A., Williams, D. L. & Frothingham, R. (1997).** PCR identification of *Mycobacterium bovis* BCG. *J Clin Microbiol* **35**, 566–569.

**Tantivitayakul, P., Panapruksachat, S., Billamas, P. & Palittapongarnpim, P. (2010).** Variable number of tandem repeat sequences act as regulatory elements in *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* **90**, 311–318.

**van Belkum, A., Scherer, S., van Alphen, L. & Verbrugh, H. (1998).** Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev* **62**, 275–293.

**Whatmore, A. M. (2001).** *Streptococcus pyogenes sclB* encodes a putative hypervariable surface protein with a collagen-like repetitive structure. *Microbiology* **147**, 419–429.