Sequence Homology and Expression Profile of Genes Associated with DNA Repair Pathways in *Mycobacterium leprae*

Mukul Sharma1, Sundeep Chaitanya Vedithi2, Madhusmita Das3, Anindya Roy1, Mannam Ebenezer3
1Department of Biotechnology, Indian Institute of Technology, Hyderabad, Telangana, 2Schieffelin Institute of Health Research and Leprosy Center, Vellore, Tamil Nadu, India, 3Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA, UK

Abstract

**Background:** Survival of *Mycobacterium leprae*, the causative bacteria for leprosy, in the human host is dependent to an extent on the ways in which its genome integrity is retained. DNA repair mechanisms protect bacterial DNA from damage induced by various stress factors. The current study is aimed at understanding the sequence and functional annotation of DNA repair genes in *M. leprae*. **Methods:** The genome of *M. leprae* was annotated using sequence alignment tools to identify DNA repair genes that have homologs in *Mycobacterium tuberculosis* and *Escherichia coli*. A set of 96 genes known to be involved in DNA repair mechanisms in *E. coli* and Mycobacteriaceae were chosen as a reference. Among these, 61 were identified in *M. leprae* based on sequence similarity and domain architecture. The 61 were classified into 36 characterized gene products (59%), 11 hypothetical proteins (18%), and 14 pseudogenes (23%). All these genes have homologs in *M. tuberculosis* and 49 (80.32%) in *E. coli*. A set of 12 genes which are absent in *E. coli* were present in *M. leprae* and in Mycobacteriaceae. These 61 genes were further investigated for their expression profiles in the whole transcriptome microarray data of *M. leprae* which was obtained from the signal intensities of 60bp probes, tiling the entire genome with 10bp overlaps. **Results:** It was noted that transcripts corresponding to all the 61 genes were identified in the transcriptome data with varying expression levels ranging from 0.18 to 2.47 fold (normalized with 16S rRNA). The mRNA expression levels of a representative set of seven genes (four annotated and three hypothetical protein coding genes) were analyzed using quantitative Polymerase Chain Reaction (qPCR) assays with RNA extracted from skin biopsies of 10 newly diagnosed, untreated leprosy cases. It was noted that RNA expression levels were higher for genes involved in homologous recombination whereas the genes with a low level of expression were involved in the direct repair pathway. **Conclusion:** This study provided preliminary information on the potential DNA repair pathways that are extant in *M. leprae* and the associated genes.

**Keywords:** DNA repair, gene expression, homology, *Mycobacterium leprae*, phylogeny, transcriptome

**Introduction**

Stability and integrity of genetic information is crucial to cell survival and multiplication. Both prokaryotes and eukaryotes contain a repertoire of DNA repair pathways that are crucial to protecting the DNA from a myriad of harming errors which can be caused by various external and intracellular factors. Environmental agents such as chemicals, ultraviolet light and ionizing radiation, as well as errors in DNA metabolism, challenge the chemical structure and stability of the genome. These etiological factors lead to a variety of alterations in the normal DNA structure such as single- and double-strand breaks, chemically modified bases, abasic sites, inter- and intra-strand cross-links, and base-pairing mismatches. Given this diversity of threats and their effects, it is not surprising that there is a corresponding diversity in DNA repair pathways.[1] The diversity in functions and complexity of DNA repair pathways is better understood by comparing the mechanisms of action of each of the pathways. Most of what is thought for bacterial DNA repair mechanisms is derived from research in *Escherichia coli* (*E. coli*). However, genome sequencing has revealed many

**Access this article online**

**Quick Response Code:**

Website: www.ijmyco.org

DOI: 10.4103/ijmy.ijmy_111_17

**Address for correspondence:** Dr. Madhusmita Das, Molecular Biology Laboratory, Schieffelin Institute of Health Research and Leprosy Center, Karigiri, Vellore - 632 106, Tamil Nadu, India. E-mail: madhusmitadas21@gmail.com

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Sharma M, Vedithi SC, Das M, Roy A, Ebenezer M. Sequence homology and expression profile of genes associated with DNA repair pathways in *Mycobacterium leprae*. Int J Mycobacteriol 2017;6:365-78.
genes with unknown capabilities, and clear variations improve questions about the ubiquity of similar DNA repair pathways in the bacterial kingdom. For instance, many species of bacteria, including E. coli, lack an end joining pathway and depend on non-homologous recombination to repair double stranded breaks and alternatively on non-homologous end joining mechanisms (NHEJ).[2] Proteins associated with NHEJ were identified in a number of bacteria, some of which include Bacillus subtilis, Mycobacterium tuberculosis, Mycobacterium smegmatis,[3-6] and Mycobacterium marinum.[7] Bacteria utilize a remarkably compact version of NHEJ wherein all the required activities are contained in only two proteins: a Ku homodimer and a multifunctional ligase/polymerase/nuclease LigD.[8]

Originating from the family of Mycobacteriaceae, the genus Mycobacteria consists of pathogens known to cause serious diseases in humans, including tuberculosis and leprosy. The etiological agent of leprosy is Mycobacterium leprae. This bacteria has never been successfully grown on an artificial cell culture medium.[9] Instead, it has been grown in mouse foot pads and in armadillos. Armadillos develop infection and manifest disease. M. leprae also has the longest doubling time of 14 days.[10] Due to the absence of an axenic culture medium for propagation, studying cellular processes, especially those belonging to DNA repair pathways is often challenging. In general, the genes involved in DNA repair mechanisms are a part of the core metabolism and Possess similarity with E. coli and other Mycobacterial genomes, however intriguing minor differences suggest biological diversity in bacterial responses to DNA damage.

In this study, the genes in M. leprae that possess a probable role in DNA repair pathways, were identified and annotated using computational and laboratory tools. Initially, a bioinformatics approach was employed to analyze and describe the open reading frames (ORFs) in the genome of M. leprae. Specific homologues and orthologs of genes corresponding to DNA repair pathways in E. coli and M. tuberculosis were identified from the public databases. Most of the genes indicated a range of similarity and identity with orthologs in the genome of M. tuberculosis. However, M. leprae does not possess genes of the typical mismatch repair (MMR) system that are found in most of the other bacteria. Although M. leprae and E. coli belong to separate phylogenetic groups, many of their DNA repair genes possess substantial similarity. However, some of the vital DNA Repair genes that are present in E. coli, are absent in M. leprae.[11] Conversely, some of the functionally related genes that are present in M. leprae, are absent in E. coli.

Methods

Sequence annotation to identify DNA repair genes in M. leprae genome

The putative ORFs of M. leprae were compared with known DNA repair related genes obtained from public databases using the “BlastP” and DELTA-Blast search over Genbank non-redundant (nr) database of proteins. In a few precise cases, potential DNA repair genes in M. leprae genome were identified both by sequence similarity searches (using seed sequence orthologs from other organisms) and keyword searches. The candidate genes that are associated with DNA repair pathways are therefore confirmed by sequence similarity searches and domain analysis using CDD Blast on a Conserved Domain Database (National Centre for Biotechnology Information (NCBI)).

Sequence phylogeny analysis

Sequence similarities and evolutionary relatedness of all the probable DNA repair genes in M. leprae which are identified by above methods, were further analyzed by searching for orthologous and paralogous sequences in KEGG SSDB database using Smith–Waterman (SW) scoring matrix.[12] Phylogenetic trees were generated for a group of hypothetical protein orthologs and paralogs present in Mycobacteria family. Protein sequences were aligned using “MUSCLE” (multiple sequence alignment program)[13] and manually adjusted with “Bio-Edit” (http://www.mbio.ncsu.edu/bioedit/bioedit.html). The maximum likelihood phylogenies with 100 bootstrap replicates were performed with PhyML[14] using the “Phylogeny.fr.”[15]

Identification of ribosome binding sites and promoters

Nucleotide sequences of putative promoter regions for selected hypothetical proteins were obtained from publicly available databases. For all open-reading frames, 200 nucleotides upstream of the translation initiation site were considered while mapping promoters. Ribosome binding sites (RBS) and promoter sequences were predicted for a common motif by DNA alignments using MUSCLE.[13]

Insights from whole transcriptome microarray experiments

To determine the activity of the DNA repair genes, expression levels of these genes were analyzed in the transcriptome of M. leprae (whole RNA extracted from human skin biopsies of newly diagnosed untreated leprosy cases) using unpublished data on whole transcriptome experiments conducted by Chaitanya et al. (Schieflin Institute of Health Research and Leprosy Center, Karigiri) (GEO dataset: GSE85948 private series). Differential gene expressions in terms of signal intensities of the DNA repair genes in the microarray experiment were normalized with that of 16SrRNA, which is most commonly used housekeeping gene to measure the basal level of mRNA expressions in prokaryotes.[16,17] The median intensity value of 16SrRNA as noted from the experiments is 8.051386 and this value was used to calculate the expression folds.

Quantitative polymerase chain reaction (qPCR) experiments

Source of Mycobacterium leprae RNA

M. leprae RNA was obtained from the skin biopsies of active leprosy patients. A total of 10 newly diagnosed untreated leprosy cases from the Dermatology Outpatient Department.
of “Schieffelin Institute of Health–Research and Leprosy Centre”, Karigiri, Tamil Nadu, India, were enrolled in the study following the institutional ethical guidelines. An informed and written consent for participation was obtained from all the subjects before enrolling in the study, following the ethical guidelines as laid down by the Indian Council of Medical Research. All the procedures conducted in the study were in accordance with the guidelines of the institutional ethical committee and with the ethical standards as laid down in the 1964 declaration of Helsinki and its later amendments or comparable ethical standards. The excisional skin biopsy samples were collected in RNA later (Catalog No: R0901, Sigma-Aldrich) in aseptic conditions, by a clinician and were sent to Molecular Biology laboratory for RNA extraction and quantitative polymerase chain reaction (qPCR) experiments.

RNA extraction

RNA extraction was performed using RNeasy Blood and Tissue Kit (Catalog No: 74104; Qiagen Inc., USA) according to manufacturer’s protocol. Aseptically, 2 mm × 2 mm size skin tissues were cut from the actual biopsy sample and were minced/grinded thoroughly using manual glass homogenizer. Alternatively, the tissues (up to 30 mg) were disrupted in Buffer RLT and homogenized using Tissuelyser LT(Catalog No.: 69980, Qiagen Inc., USA). Ethanol was added to the lysate to promote selective binding of RNA to the RNeasy membranes. The sample was then applied to the RNeasy Mini spin column. The contaminants were washed twice and high-quality RNA was eluted in RNase-free water. Genomic DNA contamination was removed by performing DNase treatment (Catalog No: EN0521, Thermo Fischer Scientific). To rule out the presence of DNA contamination in the RNA samples, a PCR was set up as reported earlier [18] without reverse transcription reaction. P2 and P3 primers for 16S rRNA gene of M. leprae directly from the RNA samples were used in the PCR amplifications. qPCR experiments to determine/confirm the expression levels. The contaminants were washed twice and high-quality RNA was eluted in RNase-free water. Genomic DNA contamination was removed by performing DNase treatment (Catalog No: EN0521, Thermo Fischer Scientific). To rule out the presence of DNA contamination in the RNA samples, a PCR was set up as reported earlier [18] without reverse transcription reaction. P2 and P3 primers for 16S rRNA gene of M. leprae directly from the RNA samples without reverse transcription reaction. P2 and P3 primers as reported earlier [18] were used in the PCR amplifications. Complementary DNA (cDNA) was constructed from 1 µg of total RNA from each of the sample using high-capacity cDNA reverse transcription kit (Catalog No.: 4368814, Applied Biosystems).

Quantitative polymerase chain reaction

Based on the expression levels of the DNA repair genes identified from the transcriptome data, genes corresponding to a set of 4 highly expressed and annotated proteins and 3 highly expressed hypothetical proteins were selected for qPCR experiments to determine/confirm the expression levels. cDNA corresponding to these 10 transcripts was amplified on a Rotor Gene-Q qPCR machine (Qiagen Inc., USA, Serial Number: R0414139) using respective primers [Table 1] and by following reaction conditions. A volume of 20 µl reaction mix containing 10 µl of QuantiNova SYBR Green PCR Master Mix (Qiagen, Cat No: 208054), 0.25 µM (0.5 µl) concentration of each of the forward and reverse primers for respective genes, 7 µl of nuclease free distilled water and 2 µl of cDNA (containing approximately 200 ng) were cycled in Rotor-Gene Q. Cycling conditions include one cycle of hold at 95°C for 2 min (initial denaturation and activation of enzyme) followed by 40 cycle of 95°C for 10 s, annealing at 60°C for 15 s and elongation at 72°C for 20 s. Fluorescence was acquired on green channel during the annealing step. This was followed by a melting step which involves an increase in temperature from 72°C to 95°C at a rate of 1°C/s. Melting curve analysis was performed to determine the integrity of the amplification and to rule out primer-dimer formation.

Analysis of quantitative polymerase chain reaction data

The mRNA expression levels were normalized using 16SrRNA as a reference. The threshold fluorescence values were normalized to those of 16SrRNA threshold fluorescence (Ct) values. The mRNA expression levels were calculated after determining the primer efficacy for all the targets using Pfaffl Method [16] by a standard curve with a 7-fold dilution of M. leprae DNA from 500 pg/reaction to 7.813 pg/reaction. Melting curve analysis was performed to determine the integrity of the amplification and to rule out primer-dimer formation. PCR for 16SrRNA PCR was performed as reported earlier [17].

Table 1: Primer sequences for Seven DNA repair genes which chosen for gene expression analysis

| Serial number | Name of the gene | Primer sequence | Annealing temperature (°C) | Amplicon size (bp) |
|---------------|------------------|-----------------|---------------------------|--------------------|
| 1             | RecN/ML1360      | Forward 5'-GACTGTACTGACCGGCCGAAA-3' | 60 | 116 |
|               |                  | Reverse 5'-CAGCACGGTTAGCTCCCTGAT-3' |               |                  |
| 2             | DnaJ1/ML2494c    | Forward 5'-CACCCTGACCACTCCCGTTA-3' | 60 | 120 |
|               |                  | Reverse 5'-AGGATACCGCCCATCTGAGGT-3' |               |                  |
| 3             | ML1105           | Forward 5'-GGTTGTGGTCGAGCTACGTT-3' | 60 | 119 |
|               |                  | Reverse 5'-TACAACACCCTTGCTGAACC-3' |               |                  |
| 4             | ML0603           | Forward 5'-GTGGAAGCGTGTGTTGCTCTG-3' | 60 | 108 |
|               |                  | Reverse 5'-CTGTGATAACGCTGTAACCGC-3' |               |                  |
| 5             | ML0202           | Forward 5'-CCTGTGCTAGCGGAACTATGAC-3' | 60 | 120 |
|               |                  | Reverse 5'-GCCATCTGCAAATCCCGCAG-3' |               |                  |
| 6             | RuvA/ML0482      | Forward 5'-ATAGTGATGTCGCTCCTCGTG-3' | 60 | 85  |
|               |                  | Reverse 5'-ACCTGTGTCGTAACCTCCAG-3' |               |                  |
| 7             | RecA/ML0987      | Forward 5'-AACCTTCTGCCAAAACTCGTG-3' | 60 | 114 |
|               |                  | Reverse 5'-CCGAATGTTGCCCCATTAGCG-3' |               |                  |
RESULTS

Genomic sequence annotations

A set of 96 DNA repair genes in the genome of *E. coli* and *M. tuberculosis* were considered as a reference and searched for their presence in *M. leprae* [Table 2, supplementary data]. This approach was adopted to identify the conserved nature of the DNA repair genes in Mycobacteriaceae and conversely, to identify the unique DNA repair genes in *M. leprae*. BLAST search on the protein database revealed the presence of 61 genes in the genome of *M. leprae* whose products detect orthologous DNA repair genes in *E. coli* and *M. tuberculosis*. Genbank annotations of the 61 genes identify 36 as characterized gene products (59%), 11 as hypothetical proteins (18%), and 14 as pseudogenes (23%). All these genes have orthologs in *M. tuberculosis* and 49 (80.32%) in *E. coli*. A set of 12 genes which are absent in *E. coli*, are present in *M. leprae* and Mycobacteriaceae. These include DNA ligases, DNA helicase II (*uvrD*), DNA helicase *RecC*3, Error-prone DNA polymerase DnaE2, DNA MMR protein *mutT*, and uracil DNA glycosylases. Functional annotation of all these proteins in DNA repair mechanisms is presented in Table 2.

Sequence comparison and phylogenetic analysis

A set of 11 hypothetical genes, namely *ML1105, ML1889, ML0202, ML0190, ML0603, ML2157, ML1351, ML1682, ML2698, ML1683*, and *ML1175* which are identified in the above approach were further searched for homologs across the prokaryotic databases using KEGG SSDB search with SW scoring matrix. This was performed to identify the functional characteristics of the hypothetical proteins in relevance to DNA repair and to decipher the evolutionary relatedness with homologs in other bacteria. Multiple sequence alignment of these proteins with MUSCLE indicated that many of Mycobacteriaceae family members contain the conserved residues. All the close homologs that had high sequence identities are hypothetical proteins themselves and are identified as entities of Mycobacteriaceae family. These were selected to build a phylogenetic tree. The phylogenetic profiles were bootstrapped 100 times before constructing the trees. All the phylogenetic trees confirmed a close relationship between the 11 hypothetical proteins and proteins from the Mycobacteriaceae family. Hence, these hypothetical proteins are well conserved and might possess a functional role. Some of the closely related species matches include *M. haemophilum, M. tuberculosis, M. marinum*, and *M. kansasii*.

Annotation of ribosome binding sites and promoters

To identify the expression characteristics of the 11 hypothetical protein coding genes mentioned in sections above, presence of RBS and promoter like sequences in the 5’ UTR were determined by multiple sequence alignment with promoter-like regions of other Mycobacterial homologs. A representative set of alignments for two hypothetical proteins with their transcription initiation sites, Shine – Dalgarno (SD) sequence and translational start points were aligned to their homologs in Mycobacteria [Figure 1]. Some of the hypothetical proteins demonstrate low similarities with their Mycobacterial counterparts. Although Mycobacterial promoters, for the most part, comprise of some indistinguishable segments from established bacterial promoters and occur upstream of and/or lie between the coding areas of two adjoining gene fragments; some much diverse promoter sequences concurrently exist, which direct the sequence interpretation and transcription in *M. leprae*. To check whether these hypothetical protein coding genes express in *M. leprae*, despite lacking canonical promoter regions, a set of 3 hypothetical proteins that indicated low similarity with their homologs in other mycobacteria, were chosen and qPCR was performed to identify gene expression.

Gene expression profiles from the *Mycobacterium leprae* whole transcriptome microarray

Transcriptome data were analyzed for 61 genes identified from the sequence based homology searches above and it was noted that transcripts corresponding to all the 61 genes were detected from the transcriptome data. A set of 60 nt length probes tiling every 10 nt and complementary to the transcripts of each of the 61 DNA repair genes in *M. leprae* (with mean signal-to-noise ratio cut-off value of ≥2), were analyzed. The signal intensities of each of the transcript was normalized with that of 16SrRNA whose median signal intensity was 8.051386. The fold-change in average gene expression levels was obtained by dividing the 16SrRNA signal intensity value with that of the expressed DNA repair genes followed by logarithmic transformation. It was noted that *ML1335c* demonstrated highest signal intensity and it was annotated as a pseudogene in *M. leprae* having seven stop codons. These observations correlate with the earlier findings on higher expression of pseudogenes and their implications in *M. leprae*. It was noted that *RecN* which is primarily involved in homologous recombination process was overexpressed in the current experimental sample. However, the other genes contributing to this pathway are moderately expressed. The least expressed gene is *RuvA*, which has a signal intensity that is nearly equal to that of 16SrRNA. A heatmap indicating expression levels of all the 61 DNA repair genes is represented in Figure 2.

Determination of gene expressions of a representative set of seven DNA repair genes by quantitative polymerase chain reaction

The gene expression profiles of 3 hypothetical protein coding genes (*ML1105, ML0202*, and *ML0603*) and 4 regular DNA repair genes (*RecN, DNAJ1, RuvA*, and *RecA*) from untreated patients’ sample were analyzed using qPCR. qPCR assays were based on target-specific primers and a master mix containing SYBR Green I fluorescent dye that intercalates with double-stranded DNA (dsDNA/cDNA) that was generated during each progressive cycle of the PCR and emits a fluorescence signal which is quantitatively measured to track the amplification of cDNA. There is a quantitative relationship between the amount of starting template and the PCR product at the exponential phase of the PCR.
Table 2: Comparison of DNA repair genes of *Mycobacterium leprae* with *Escherichia coli* and *Mycobacteriaceae* family with focus on *Mycobacterium tuberculosis*

| Name of protein (E. coli) | GI accession | Uniprot Id | Mycobacteriaceae | M. leprae (TN strain) | M. tuberculosis | Gene name in M. leprae | Gene name in M. leprae | NCBI gene ID | GI accession | Function/alternative name |
|--------------------------|--------------|------------|------------------|----------------------|----------------|------------------------|------------------------|--------------|--------------|----------------------------|
| **Base excision repair** |              |            |                  |                      |                |                        |                        |              |              |                             |
| Adenine DNA Glycosylase  |              |            |                  |                      |                |                        |                        |              |              | Adenine DNA glycosylase     |
| *MutY*                   | 16130862     | P17802     | Present          | Present              | Present        | Rv3589                 | Present                | ML1920       | 910168       | Adenine DNA glycosylase     |
| *UdgB*                   |              | Absent     | Present          | Present              | Present        | Rv1259                 | Present                | ML1105       | 910195       | Family 5 UDG               |
| *Ung*                    | 148149       | P12295     | Present          | Present              | Present        | Rv2976c                | Present                | ML1675c      | 910041       | UDG                        |
| *Mug*                    | 1789449      | P0A9H1     | Present          | Present              | Present        | Rv0276c                | Present                | ML1931       | 910101       | Exodeoxyribonuclease III   |
| **Nucleotide excision repair** |            |            |                  |                      |                |                        |                        |              |              |                             |
| *UvrA*                   | 16131884     | P0A698     | Present          | Present              | Present        | Rv1638                 | Present                | ML1392       | 910525       | ATPase and DNA damage recognition protein of nucleotide excision repair excinuclease UvrABC |
| *UvrB*                   | 67474768     | P0A8F8     | Present          | Present              | Present        | Rv1633                 | Present                | ML0483       | 909232       | UvrABC system protein B, excinuclease ABC subunit B |
| *UvrC*                   | 189038049    | A1AC65     | Present          | Present              | Present        | Rv1420                 | Present                | ML0481       | 909230       | UvrABC system protein C, excinuclease ABC subunit C |
| *Mfd*                    | 1787357      | P30958     | Present          | Present              | Present        | Rv1020                 | Present                | ML0252       | 908750       | Transcription-repair coupling factor |
| *UvrD*                   | 148212       | P03018     | Present          | Present              | Present        | Rv0949 uvrD1           | Present                | ML0153       | 908505       | DNA helicase II            |
| *UvrD2*                  |              | Absent     | Present          | Present              | Present        | Rv3198c                | Present                | ML0637       | 909420       | DNA helicase II paralog     |
| *erd3*                   |              | Absent     | Present          | Present              | Present        | Rv0816c                | Present                | ML2157       | 908199       | Helicase (eukaryotic)       |
| **Mismatch repair**      |              |            |                  |                      |                |                        |                        |              |              |                             |
| *MutH*                   | 730086, 42065 | P06722     | Absent           | Absent               | Absent         | Absent                 | Absent                 | DNA mismatch repair protein mutH, methyl-directed mismatch repair protein |
| *MutL*                   | 42067        | P23367     | Present          | Absent               | Absent         | Absent                 | Absent                 | MutL protein |
| *MutS*                   | 17017340     | P23909     | Present          | Absent               | Absent         | Absent                 | Absent                 | MutS         |              |                             |

*Contd...*
| Name of protein (E. coli) | GI accession | Uniprot Id | Mycobacteriaceae | M. tuberculosis | Gene name in M. tuberculosis | M. leprae (TN strain) | Gene name in M. leprae | NCBI gene ID | GI accession | Function/alternative name |
|--------------------------|--------------|------------|------------------|----------------|-----------------------------|-----------------------|-----------------------|--------------|--------------|--------------------------|
| vsr                      | 16129906     | P09184     | Absent           | Absent         | Absent                      | ML1940                | ML1941                | 910082       | NP_302308    | Very short patch repair  |
| sbcB                     | 16129952     | P04995     | Absent           | Absent         | Absent                      |                      |                      |              |              |                           |
| XseA                     | 148275       | P04994     | Present          | Present        | Rv1108c                     | Present               | ML1940                | 910082       | NP_302308    | Exonuclease VII large subunit |
| XseB                     | 89107292     | P0A8G9     | Present          | Present        | Rv1107c                     | Present               | ML1941                | 910018       | NP_302309    | Exonuclease VII small subunit |
| RadA                     | 16132206     | P24554     | Present          | Present        | Rv3585                      | Present               | ML0318c               | 910137       |              | Homologous recombination |
| RecA                     | 37362719     | P0A7G6     | Present          | Present        | Rv2737c                     | Present               | ML0987                | 910009       | NP_301732    | RecA          |
| RecB                     | 16130724     | P08394     | Present          | Present        | Rv0630c                     | Absent                |                      |              |              | Exonuclease V (RecBCD complex), beta subunit |
| RecC                     | 16130726     | P07648     | Present          | Present        | Rv0631c                     | Absent                |                      |              |              | Exonuclease V (RecBCD complex), gamma chain |
| RecD                     | 16130723     | P04993     | Present          | Present        | Rv0629c                     | Absent                |                      |              |              | Exonuclease V (RecBCD complex), alpha chain |
| RecE                     | 147536       | P15032     | Absent           | Absent         | Absent                      |                      |                      |              |              | RecE          |
| RecF                     | 147539       | P0A7H0     | Present          | Present        | Rv0003                      | Present               | ML0003                | 910262       | NP_301131    | RecF          |
| RecG                     | 42669        | P24230     | Present          | Present        | Rv2973c                     | Present               | ML1671c               | 910037       | NP_302148    | RecG          |
| RecJ                     | 887842       | P21893     | Present          | Absent         | Absent                      |                      |                      |              |              | Single-stranded DNA-specific exonuclease |
| RecN                     | 42693        | P05824     | Present          | Present        | Rv1696                      | Present               | ML1360                | 910489       | NP_301970    | Unnamed protein product |
| RecO                     | 499369       | P0A7H3     | Present          | Present        | Rv2362c                     | Present               | ML0633                | 909415       | NP_301524    | RecO          |
| RecQ                     | 147559       | P15043     | Present          | Absent         | Absent                      |                      |                      |              |              | RecQ          |
| RecR                     | 16128456     | P0A7H6     | Present          | Present        | Rv3715c                     | Present               | ML2329c               | 908695       | NP_302515    | Gap repair protein |
| RecT                     | 397681       | P33228     | Present          | Absent         | Absent                      |                      |                      |              |              | Unknown       |
| RecX                     | 16130605     | P33596     | Present          | Present        | Rv2736c                     | Present               | ML0988                | 910283       | NP_301733    | Regulatory protein for RecA |
| RuvA                     | 581226       | P0A809     | Present          | Present        | Rv2593c                     | Present               | ML0482                | 909231       | NP_301422    | Unnamed protein product |
| RuvB                     | 42903        | P0A812     | Present          | Present        | Rv2592c                     | Present               | ML0483                | 909232       | NP_301423    | Unnamed protein product |
| RuvC                     | 42175        | P0A814     | Present          | Present        | Rv2594c                     | Present               | ML0481                | 909230       | NP_301421    | RuvC          |
| Ku                       | Absent       | Present    | Present          | Present        | Rv0937c                     | Present               | ML2092                | 908303       |              |                           |
| LigB                     | Present      | Present    | Present          | Present        | Rv3062                      | Present               | ML1747                | 910746       |              |                           |
| LigC                     | Absent       | Present    | Present          | Present        | Rv3731                      | Absent                |                      |              |              |                           |
| LigD                     | Absent       | Present    | Present          | Present        | Rv0938                      | Absent                |                      |              |              |                           |
| Nonhomologous end joining|             |            |                  |                |                            |                      |                      |              |              |                           |
| Ku                       | Absent       | Present    | Present          | Present        | Rv0937c                     | Present               | ML2092                | 908303       |              |                           |
| LigB                     | Present      | Present    | Present          | Present        | Rv3062                      | Present               | ML1747                | 910746       |              |                           |
| LigC                     | Absent       | Present    | Present          | Present        | Rv3731                      | Absent                |                      |              |              |                           |
| LigD                     | Absent       | Present    | Present          | Present        | Rv0938                      | Absent                |                      |              |              |                           |
| Translesion synthesis    |              |            |                  |                |                            |                      |                      |              |              |                           |
| UmuC                     | 84060801     | P04152     | Present          | Absent         | Absent                      | Absent                | Absent                |              | UmuC         |                           |
Table 2: Contd...

| Name of protein (E. coli) | GI accession | Uniprot Id | Mycobacteriaceae | M. tuberculosis | Gene name in M. tuberculosis | M. leprae (TN strain) | Gene name in M. leprae | NCBI gene ID | GI accession | Function/alternative name |
|--------------------------|--------------|------------|------------------|----------------|-----------------------------|----------------------|----------------------|-------------|--------------|---------------------------|
| UmuD                     | 85376582     | P0AG11     | Present          | Absent         | Absent                      | ML1197 pseudogene    | ML1197 pseudogene    | 910299      |             | UmuD                      |
| DinB                     | 16128217     | Q47155     | Present          | Present        | Rv1537 DinX                 | ML1739 pseudogene    |                      | 910754      |             | DNA polymerase IV           |
| DinP                     | Absent       | Present    | Present          | Present        | Rv3056                      |                      |                      |             |             |                           |
| Direct repair            |              |            |                  |                |                             |                      |                      |             |             |                           |
| Ogt                      | 84027827, 16129296 | P0A FH0 | Present          | Present        | Rv1316e                     | ML1151c              | 910248              | NP_301845   |             | 6-O-methylguanine-DNA methyltransferase, O-6-methylguanine-DNA-alkyltransferase, Methylated-DNA-protein-cysteine methyltransferase |
| AlkB                     | 405945       | P05050     | Present          | Present        | Rv1000c                     | ML0190 hypothetical protein | 908584              | NP_301263   |             | AlkB                      |
| Ada                      | 461468       | P06134     | Present (abseccus) | Absent         | Absent                      |                      |                      |             |             | Regulatory protein ada, regulatory protein of adaptive response, contains: Methylated-DNA-protein-cysteine methyltransferase |
| Phr                      | 6006450      | P00914     | Present          | Absent         | Absent                      |                      |                      |             |             | CPD photolyase, DNA photolyase |
| Nucleotide pool          |              |            |                  |                |                             |                      |                      |             |             |                           |
| MutT                     | 16128092     | P08337     | Present          | Present        | Rv2985                      | ML1682 hypothetical protein | 910049              | NP_302156   | 8-oxo-dGTP diphosphatase |
| MutT2                    | Absent       | Present    | Present          | Present        | Rv1160                      | ML1503 pseudogene    | 909535              |             | MutT2                     |
| MutT3                    | Absent       | Present    | Present          | Present        | Rv0413                      | ML0301 pseudogene    | 908862              |             | MutT3                     |
| MutT4                    | Absent       | Present    | Present          | Present        | Rv3908                      | ML2698 hypothetical protein | 908257              | NP_302721   | Possible mutT4, mutator protein |
| RdgB                     | 16130855     | P52061     | Present          | Present        | Rv1341                      | ML1175 hypothetical protein | 910273              | NP_301857   | dITP/XTP pyrophosphatase |
| MaxG                     | 16130688     | P0AEY3     | Present          | Present        | Rv1021                      | ML0253 pseudogene    | 908752              |             | MaxG                      |
| Dut                      | 16131511     | P06968     | Present          | Present        | Rv2697c                     | ML1028               | 910096              | NP_301761   | Deoxyuridine 5'-triphosphate nucleotidohydrolase |

Contd...
| Name of protein (E. coli) | GI accession | Uniprot Id | Mycobacteriaceae | M. tuberculosis | Gene name in M. tuberculosis (M. leprae (TN strain)) | M. leprae (Gene name in M. leprae) | NCBI gene ID | GI accession | Function/alternative name |
|--------------------------|--------------|------------|------------------|----------------|-------------------------------------------------|---------------------------------|-------------|--------------|---------------------------|
| **Regulatory**            |              |            |                  |                |                                                 |                                 |             |              |                           |
| *LigA*                   | 91211748     | P15042     | Present          | Present        | Rv3014c  | Present                                        | ML1705c          | 910789       | NP_302174    | NAD-dependent DNA ligase  |
| *LexA*                   | 67467382     | P0A7C3     | Present          | Present        | Rv2720   | Present                                        | ML1003c          | 910052       | NP_301742    | LexA repressor          |
| *PolI*                   | 147312, 42461 | P00582     | Present          | Present        | Rv1629   | Present                                        | ML1381c          | 910507       | NP_301982    | DNA polymerase I         |
| *Ssb*                    | 16131885     | P0AGE0     | Present          | Present        | Rv0054   | Present                                        | ML2684c          | 908269       | NP_302712    | Single-stranded DNA-binding protein |
| **Other potential DNA repair genes** |            |            |                  |                |                                                 |                                 |             |              |                           |
| *AidB*                   | 1264215      | P33224     | Present          | Absent         | Absent   | Absent                                         |                   |             |                           | Protein AidB               |
| *Dun*                    | 16131265     | P0AE8      | Present          | Absent         | Absent   | DNA adenine methylase                          |                   |             |                           | DNA dammage-inducible SOS response protein |
| *DinF*                   | 89110765     | P28303     | Present          | Present        | Rv2836c  | Absent                                         |                   |             |                           | ATP-dependent DNA helicase |
| *DinFl* like             | Absent       | Absent     | Present          | Present        | Rv2090   | Present                                        | ML1335 pseudogene | 910459       |                           | DNA damage-inducible SOS response protein |
| *DinG*                   | 89107650     | P27296     | Present          | Present        | Rv1329c  | Absent                                         |                   |             |                           | predicted antitoxin of YafQ-DinJ toxin-antitoxin system |
| *DinI*                   | 89107907     | P0ABRI     | Absent           | Absent         | Absent   | DNA damage-inducible protein I                 |                   |             |                           |                                   |
| *DinJ*                   | 89107101     | Q47150     | Absent           | Absent         | Absent   | DNA damage-inducible protein I                 |                   |             |                           |                                   |
| *DnaE*                   | 146663       | P10443     | Present          | Present        | Rv1547   | Present                                        | ML1207           | 910310       | NP_301875    | DNA polymerase III holoenzyme, alpha subunit |
| *DnaE2*                  | Absent       | Absent     | Present          | Present        | Rv3370   | Present                                        | ML0416 pseudogene | 909155       |                           |                                   |
| *DnaF*                   | 16128009     | P08622     | Present          | Present        | Rv0352   | Present                                        | ML2494           | 908467       | NP_302611    | Chaperone protein DnaF    |
| *DnaN*                   | 145761       | P0A988     | Present          | Present        | Rv0002   | Present                                        | ML0002           | 908144       | NP_301130    | DNA polymerase III beta-subunit |
| *DnaQ*                   | 147679       | P03007     | Present          | Present        | Rv3711c  | Present                                        | ML2325c pseudogene | 908707       |                           | DNA polymerase III epsilon subunit |
| *DnaT*                   | 16132183     | P0A82      | Absent           | Absent         | Absent   | DNA biosynthesis protein (primosomal protein I) |                   |             |                           |                                   |
| *DnaX*                   | 118808       | P06710     | Present          | Present        | Rv3721c  | Present                                        | ML2335c          | 908684       | NP_302521    | DNA polymerase III subunit tau; contains: DNA polymerase III subunit gamma |
| *Fpg (MutM)*             | 146545       | P05523     | Present          | Present        | Rv2924c  | Present                                        | ML1658c          | 910013       | NP_302139    | Formamidopyrimidine-DNA glycosylase Helicase IV |
| *HelD*                   | 146328       | P15038     | Present          | Absent         | Absent   | DNA polymerase III delta subunit               |                   |             |                           |                                   |
| *HolA*                   | 145729       | P28630     | Present          | Present        | Rv2413   | DNA polymerase III delta subunit               |                   |             |                           |                                   |
| *HolB*                   | 145799       | P28631     | Present          | Present        | Rv3644c  | DNA polymerase III delta prime subunit         |                   |             |                           |                                   |
| *HolC*                   | 145537       | P28905     | Absent           | Absent         | Absent   | DNA polymerase III chi subunit                 |                   |             |                           |                                   |
| *HolD*                   | 147387       | P28632     | Absent           | Absent         | Absent   | DNA polymerase III psi subunit                 |                   |             |                           |                                   |
| *HolE*                   | 145787       | P0ABS8     | Absent           | Absent         | Absent   | DNA polymerase III theta subunit               |                   |             |                           |                                   |

Contd...
Table 2: Comparative analysis of expression levels of all the seven genes

| Name of protein (E. coli) | Uniprot Id | Gene name in M. leprae | Gene name in M. tuberculosis (TN strain) | M. leprae | M. tuberculosis | NCBI accession | GI accession |
|--------------------------|------------|------------------------|-----------------------------------------|----------|----------------|----------------|-------------|
| HupA                     | B99124     | ML1351                 | Rv0944                                  | Present  | Present        | NC_000965     | 910050      |
| HupB                     | B99125     | ML1352                 | Rv0945                                  | Absent   | Absent         | NC_000965     | 910050      |
| MPg                      | P0ACF4     | Mpg                    | Rv1688                                  | Present  | Present        | NP_301965     | 909513      |
| Nei2                     | P0ACF5     | Nei2                   | Rv1689                                  | Present  | Present        | NP_301965     | 909513      |
| Nfi                      | P0ACF6     | Nfi                    | Rv1690                                  | Absent   | Absent         | NP_301965     | 909513      |
| Nth                      | P0ACF7     | Nth                    | Rv1691                                  | Present  | Present        | NP_301965     | 909513      |
| PolII                    | P0ACF8     | PolII                  | Rv1692                                  | Absent   | Absent         | NP_301965     | 909513      |

**Sharma, et al.: DNA repair in M. leprae**

**Endonuclease V**

| HU subunit alpha | HU subunit beta | DNA polymerase II |
|------------------|-----------------|-------------------|
| Absent           | Absent          | Absent            |

**DNA polymerase II**

| ML1483 | P0ACF4 | Present | Present | Present |
|--------|--------|---------|---------|---------|

**Uniprot**

| 90111673 | 3-methyladenine DNA glycosylase | Absent | Present | Rv3674c | Absent |
|----------|---------------------------------|--------|---------|---------|-------|
| 16129591 | DNA glycosylase and AP lyase     | Present| Present | Rv2968c | Absent |

**M. tuberculosis** (Br4923 strain) was used to investigate the putative genes and pathways detected in the genome of *M. leprae*. The presence and absence of DNA repair genes are discussed and predictions are made considering the particular aspects of the *M. leprae* among other known DNA repair pathways. Sequence annotations of DNA repair genes in *M. leprae* with insights from their orthologs in *E. coli* and *M. tuberculosis* enabled identification of potential DNA repair pathways. DNA repair genes were stratified based on their function in the following mechanisms: base excision repair, nucleotide excision repair (NER), MMR, recombination.

**Standard curves to determine the amplification efficiency of the selected genes in quantitative polymerase chain reaction**

Before testing on clinical samples, pure stocks of bacterial reference DNA of *M. leprae* (Br4923 strain) was used to construct standard graphs. These graphs were developed to validate the assays, identify lower detection limit and determine error rates in the qPCR experiments. Standard curves were constructed by estimating threshold cycle values for seven 10-fold serial dilutions of purified *M. leprae* DNA ranging from 0.5 ng to 7.8 pg for each qPCR assay [Table 3 and Figure 3]. Optimal fluorescence thresholds were chosen based on the common practice that it should be positioned on the lower half of the fluorescence accumulation curves plot from the 10-fold dilutions and was used both to calculate the Ct for standard curve fitting and Ct for all the 10 clinical samples in the study.

**Relative abundance of DNA repair gene transcripts in *M. leprae* RNA from clinical isolates**

qPCR of *16S* rRNA served as a positive control, imparting incremental sensitivity over assays based on the detection of a single or multiple copies of genomic sequences, since each cell contains 1000–10,000 copies of rRNA. Real time PCR was performed in duplicates for each of the 10 skin biopsies. The mRNA expression levels of all the 10 genes in clinical isolates from newly diagnosed untreated leprosy cases reveal a range of threshold fluorescence values. The average Ct values for all the 10 samples for each of the gene was represented in Figure 4.

Comparative analysis of expression levels of all the seven genes using qPCR and microarray data suggested that *RecA*, ML0202 and ML0603 indicated substantial correlation. Rest of the genes in the analysis revealed a poor correlation with observations from microarray data [Figure 5]. *RuvA* indicated increased expression in qPCR and low intensities in microarray data. One of the possible reasons for this observation could be due to the selection of leprosy cases which are all highly bacillated providing high quantities of bacterial RNA. *RecA*, ML0202 and ML0603 indicated similar expressions in both qPCR and microarray data which suggests that ML0202 and ML0603 may have a significant functional role in the DNA repair pathways. The mean Ct values of each of the genes along with the normalized (delta Ct) values are represented in Table 4 and the microarray fold changes for the same set of genes has been represented in Table 5.

**Discussion**

The relevance of this comparative analysis is to provide the basis for investigating the putative genes and pathways detected in the genome of *M. leprae*. The presence and absence of DNA repair genes are discussed and predictions are made considering the particular aspects of the *M. leprae* among other known DNA repair pathways. Sequence annotations of DNA repair genes in *M. leprae* with insights from their orthologs in *E. coli* and *M. tuberculosis* enabled identification of potential DNA repair pathways. DNA repair genes were stratified based on their function in the following mechanisms: base excision repair, nucleotide excision repair (NER), MMR, recombination.
repair, NHEJ, translesion synthesis (TLS), direct reversal, nucleotide pool, regulatory and other related processes.

**Base excision repair**

One of the primary mechanisms for the repair of alkylated bases is BER, which is initiated by one of the 3-methyladenine DNA glycosylases, *tagA* or *alkA*. A homolog of the *tagA* gene is present in *M. leprae* which includes 10 stop codons, splitting the corresponding locus into many reading frames and has been annotated as a pseudogene-(ML0190). A gene encoding “3-methyladenine DNA glycosylase” is also present in Mycobacteria and possess conserved regions throughout the Mycobacterial species. In *M. leprae*, it has been annotated as a hypothetical protein ML1351. Although no functional studies have been reported, the conservation of this gene across various species suggests its indispensable role. One of the most common and stable oxidation products in DNA is 8-oxo-7,8-dihydroguanine (8-oxo-G), having a propensity to mispairing with adenine. Both modified bases act as substrates for the formamidopyrimidine-DNA glycosylase, known as *fpg* or *mutM*. The *fpg* gene has been shown to be involved in the repair of DNA lesions induced by hydrogen peroxide in *E. coli*. *M. tuberculosis* (H 37Rv) has four genes of the *fpg/nei* family of DNA glycosylases: Rv2924c annotated as *fpg* (ML1658 in *M. leprae*), Rv3297 annotated as *nei*, Rv0944 (ML0148 in *M. leprae*) annotated as a possible *fpg*, and Rv2464c (ML1483 in *M. leprae*) annotated as a possible DNA glycosylase. Homologs of all four of these

![Figure 1: Promoter-like sequences upstream of transcribed Mycobacterium leprae hypothetical proteins ML1683 and ML0190: It shows representative alignments of promoter-like sequences for Mycobacterium leprae genes and their mycobacterial homologs which are within 200 nt upstream of the translational start point. Panel A and B represent the ML0190 and ML1683 upstream promoter-like regions containing -35 and -10 regions and initiation site (i) in relationship to their ribosomal binding sites and translational start codons (Start), respectively.](image-url)
genes are found in the other Mycobacterial genomes and in *M. leprae* the loci corresponding to *Rv0944* and *Rv2464c* contain pseudogenes, whereas there is no equivalent of *Rv3297*. Endonuclease III (*Nth*) excises oxidative pyrimidines. A homolog of *Nth* is present in both the mycobacterial genomes and it is named as *ML2301c* in *M. leprae*.

Adenine can be incorporated rather than the cognate cytosine opposite 8-oxo-G during DNA replication, leading to G.C and T.A transversions. To contract this, the adenine DNA glycosylase (*mutY*) excises the mismatched pair, which also includes nucleotides on the complementary strand. The *mutY* gene in *M. leprae* is *ML1920* which has homologs that are identified in other Mycobacterial genomes as noted in earlier studies.[23] Uracil can also be found in DNA either because of misincorporation or deamination of cytosine. The archetypal family-1 Uracil DNA glycosylases (ung) are specific to uracil in DNA and excise it from both double-stranded (ds) and single-stranded (ss) substrates.[24] The homologs of *udgB* from *E. coli* and *M. tuberculosis* are present in *M. leprae* as *ung* and ML1105. The second step in BER is the cleavage of sugar-phosphate backbone by an apurinic/apyrimidinic endonuclease. In *E. coli*, endonuclease IV (*Nfo*) and exonuclease III (*XthA*) produce a single-strand (ss) break at abasic sites by attacking the phosphodiester bond 5’ to the site of base loss, leaving 3’OH groups. Homologs of *Nfo* have been identified in many Mycobacterial species and in *M. leprae*, it is annotated as hypothetical protein (*ML1889*). Similarly, *XthA* is also present in all Mycobacterial species except *M. leprae* where a corresponding pseudogene (*ML1931*) is found.

**Nucleotide excision repair**

This system recognizes the distortion in the double helix caused by lesions which can recognize a larger variety of base modifications. Removal of lesions from the intact oligonucleotide forms is facilitated by the sequential action of nucleases and helicases, followed by DNA polymerization and ligation by DNA ligase.[25] It includes proteins *uvrA*, *uvrB*, the nuclease *uvrC*, the helicase *uvrD* and the dsDNA translocase *Mfd*. Homologs of *uvrA*, *uvrB* and *uvrC* are present in all the Mycobacterial genomes including *M. leprae*, suggesting that this pathway of DNA repair is important to Mycobacteria. Despite the canonical *uvr* genes, an additional protein involved in the incision step of NER has been identified in *E. coli*, termed *cho* having sequence similarity with the N-terminal portion of *uvrC* and containing the domain for

**Figure 2:** Heat-map of significant expression level changes in genes associated with DNA-repair

**Figure 3:** Standard graph of 16srRNA gene of *Mycobacterium leprae*

**Table 3:** Standard curves parameters and results for quantitative polymerase chain reaction assays of *Mycobacterium leprae* DNA

| Concentration (ng/reaction) | RecN | Ogt | DnaJ1 | RuvA | RecA | ML1105 | ML1889 | ML0202 | ML0190 | ML0603 |
|-----------------------------|------|-----|-------|------|------|--------|--------|--------|--------|--------|
| 0.50000000                  | 14.45| 16.44| 13.75 | 18.34| 13.14| 15.13  | 15.90  | 13.86  | 18.14  | 16.56  |
| 0.25000000                  | 14.80| 16.23| 14.06 | 18.53| 13.17| 14.74  | 15.66  | 13.96  | 17.50  | 16.71  |
| 0.12500000                  | 15.34| 17.51| 15.11 | 19.72| 14.01| 15.88  | 16.84  | 14.87  | 19.37  | 17.84  |
| 0.06250000                  | 16.53| 18.11| 16.05 | 20.75| 15.53| 17.38  | 17.78  | 15.92  | 19.83  | 18.67  |
| 0.03125000                  | 17.74| 19.97| 16.76 | 21.93| 16.21| 18.08  | 18.47  | 16.79  | 21.00  | 19.88  |
| 0.01562500                  | 18.48| 20.80| 17.91 | 23.05| 17.28| 19.51  | 19.87  | 18.11  | 22.38  | 21.21  |
| 0.00781250                  | 20.39| 21.72| 18.90 | 23.80| 18.11| 20.75  | 21.58  | 19.45  | -      | 21.62  |
the 3’ incision. The sequence of this protein is conserved throughout the Mycobacterial species, except *M. leprae*, where the corresponding locus is a pseudogene (ML0884c). Transcription-coupled repair is a sub-pathway of NER that selectively removes lesions from the transcribed strands, mediated by the transcription-repair coupling factor (mfd). Homologs of mfd have been identified in *M. leprae* (ML0252); however, the actual function is yet to be deciphered. In *M. leprae*, there are two homologs of *uvrD*, annotated as *uvrD1* and *uvrD2*. While their role is not experimentally determined, their orthologs in *M. tuberculosis* interact with Ku, a component of the NHEJ pathway of DNA repair, stimulating the helicase activity. Thus, it may be that *uvrD1* is involved in multiple DNA repair pathways in Mycobacteria. While most of the Mycobacterial genomes have homologs for superfamily II helicases known in eukaryotes, the *M. leprae* gene ML2157 encodes for ERCC3, a 3’-5’ helicase and is reported as the first example of this gene in prokaryotes.[26]

Mismatch repair

The *mutS/mutL* complex recognizes DNA replicative errors or misalignments and will perform an excision of the section containing the mismatch.[27] *M. leprae* lacks a system for MMR, as *mutS*, *mutL* or *mutH* could not be identified and not even their homologs. The exonucleases recJ or exol (encoded by sbcB or xonA) are also absent in *M. leprae*. This indicates that Mycobacteria may possess alternative control over homologous recombination, possibly involving a *recA*-mediated strand transfer. *E. coli* and related enteric bacteria also possess a system known as very short patch repair that targets mismatched T.G base pairs arising from deamination of 5-methylcytosine, especially within motifs recognized by DNA cytosine methyltransferase. Repair is initiated by the Vsr protein which nicks the DNA immediately upstream of the mismatch pair, followed by synthesis of a short stretch (<10 nucleotides) of DNA by DNA polymerase I and ligation.[28] Both these genes are absent in *M. leprae*.

Homologous recombination

Recombination repair maintains genome integrity. In *E. coli*, two pathways, the RecBCD and RecFOR recruit RecA to single stranded DNA and provoke the repair of double stranded breaks or repair post replication daughter strand gaps respectively breaks or of postreplication daughter strand gap, respectively.[29] RecA plays a central role in recombination repair and homologous recombination by promoting homologous pairing and DNA strand exchange using ATP, involving the formation of a nucleoprotein filament.[30] In some Mycobacteria like *M. tuberculosis*, recA is encoded by an elongated gene containing an intein which is made active by protein splicing[31-33] and similar observations were noted in *M. leprae*. *M. leprae-recA* intein binds to cognate DNA and

| Table 4: Summary of the qPCR results for selected DNA repair genes |
|---------------------------------------------------------------|
| **Gene name**   | **Mean Ct values** | **Delta Ct (Ct of target gene - Ct of reference gene)** |
| 16srRNA (reference gene) | 25.95 | - |
| recN | 23.05 | -2.9 |
| dnaJ1 | 19.68 | -6.27 |
| ruvA | 26.07 | 0.12 |
| recA | 25.50 | -0.45 |
| ML1105 | 22.41 | -3.54 |
| ML0202 | 25.92 | -0.03 |
| ML0603 | 24.83 | -1.12 |

Ct: Cycle threshold

| Table 5: Summary of the gene expressions from microarrays |
|---------------------------------------------------------------|
| **Gene name**   | **Mean expression values** | **Fold difference (gene/16srRNA)** | **Log2 values** |
| 16srRNA (reference gene) | 8.051386 | | |
| recN | 41.5385775 | 5.159183462 | 2.367143 |
| dnaJ1 | 35.746767 | 4.43927751 | 2.150504 |
| ruvA | 9.104483 | 1.130796983 | 0.773399 |
| recA | 15.478265 | 1.922434845 | 0.9429347 |
| ML1105 | 28.226898 | 3.505843342 | 1.809762 |
| ML0202 | 21.6997795 | 2.695160746 | 1.430371 |
| ML0603 | 16.3059385 | 2.02523373 | 1.018088 |

Figure 4: Mean Ct values of 4 DNA repair genes and 3 hypothetical protein coding genes along with 16srRNA
displays endonuclease activity in the presence of alternative
divalent cations like Mg\textsuperscript{2+} or Mn\textsuperscript{2+}.\cite{34} In \textit{E. coli}, several
pathways exist for the initial processing of dsDNA breaks
to single stranded substrates for recombination, each featuring
the action of exonucleases and helicases. \textit{M. leprae} possesses
neither of these systems, but it does possess homologs of
an archaeal exonuclease (\textit{ML1155}) and helicase (\textit{ML1312})
belonging to the rec\textit{B} family of exonuclease/helicases\cite{34} in
addition to \textit{ML2157} and exonuclease (\textit{sbcD}\textit{ML1119}, \textit{xseAB})
which can perform the break-processing function. Ruv\textit{ABC}
and Rec\textit{G} complete the process of recombination by Rec\textit{A}.
The Ruv\textit{AB} complex or the helicase Rec\textit{G} catalyze branch
migration of Holliday junctions formed by the crossing over
of strands from two DNA duplexes, and Rec\textit{V} resolves this
structure to allow separation of the DNA helices.\cite{35} Homologs
of each of Ruv\textit{A}, Ruv\textit{B}, Ruv\textit{C}, and Rec\textit{G} are present in \textit{M. leprae}.

The functions of Rec\textit{N} and Rec \textit{X} has not been elucidated to a
substantial level in Mycobacteria and hence, their role in the
repairing the double stranded breaks in \textit{M. leprae} is unknown.
\textit{M. leprae} does not possess homologs of Rec\textit{E} and Rec\textit{T} genes.
Homologs of Rad\textit{A} are present in all the Mycobacterial
species except in \textit{M. ulcerans} and \textit{M. leprae} consists of it in
the form of a pseudogene (\textit{ML0318c}).

\textbf{Non-homologous end-joining}
NHEJ also operates in some prokaryotes, including
Mycobacteria,\cite{36} but only Ku and ligase proteins are required.\cite{9}
Ku homologues are present in all the Mycobacterial species,
with the single exception of \textit{M. leprae} where it is present as a
pseudogene (\textit{ML2092}). Many Mycobacteria encode at
least three different ATP-dependent ligases, known as Lig\textit{B},
Lig\textit{C} and Lig\textit{D}; expect in \textit{M. leprae}, in which these genes
are annotated as pseudogenes \textit{ML1747} for Lig\textit{B} and \textit{ML2090}
for Lig\textit{D}. Lig\textit{C} is absent in \textit{M. leprae}.

\textbf{Translesion synthesis}
In \textit{M. leprae}, genes related to TLS are present as pseudogenes.
Din\textit{B}, Din\textit{P} and dna\textit{E2} coding genes are annotated as
pseudogenes \textit{ML1197}, \textit{ML1739}, and \textit{ML0416}, whereas other genes
\textit{umuC}, \textit{umuD}, and \textit{polB} are absent.

\textbf{SOS Repair systems}
The genes \textit{umuC} and \textit{umuD} form a complex Umu\textit{C}/Umu\textit{D2},
known as DNA polymerase \textit{V},\cite{37} which is responsible for
the induced mutagenesis through the SOS repair in \textit{E. coli}.
However, these polymerases are absent in \textit{M. leprae}.
The SOS inducible and error prone DNA polymerase \textit{IV (dinB)}
is involved in TLS in \textit{E. coli},\cite{38} and thought to be doing the
same regulatory function in \textit{M. leprae}. The SOS induced
mutagenesis in \textit{M. leprae} has been proven to be promoted by
enzymes encoded by operon including a second subunit of
Dna\textit{E} (the catalytic subunit of DNA polymerase III) called
Dna\textit{E2}.

The principal motivation for this study was to identify all the
DNA repair genes present in the \textit{M. leprae} genome, identify
their expression from available microarray data and validate a
representative set (especially the hypothetical proteins) using
qPCR assay. Overall, 100\% of the DNA repair genes were
found to be transcribed as noted in microarrays. Different
DNA repair pathways of \textit{M. leprae} exhibited different levels
of RNA expression. RNA expression was relatively higher for
genes involved in the homologous recombination, whereas,
the genes with a low level of expression were involved in the
direct repair pathway. There were some differences in the levels
of RNA expression detected by microarray and qPCR. The
level of expression of hypothetical proteins involved in direct
repair pathway detected by microarray were higher than the
level from the same genes detected by qPCR, when compared
to \textit{16S}rRNA expression. This discrepancy might reflect the
difference in the target length for both methods as well as the
difference in the length of transcribed RNA.

The presence of promoter-like sequences in the 5'UTR of
transcribed \textit{M. leprae} hypothetical genes with translational
start codons was investigated, using alignment of promoter like
regions with that of Mycobacterial homologs. These promoters
aligned very well with that of other Mycobacterial homologs and
showed relationship to their-35 and-10 box, initiation site, RBS,
and translational start codon. Although the results of this study
indicate that some hypothetical proteins (supplementary data)
having weak RBS sequences, some of the hypothetical genes
like ML0190, ML1683 have intact ribosome-binding sequences
of similar strength to the orthologs of Mycobacteriaeae.
In addition, phylogenetic analysis also revealed that these
hypothetical proteins from \textit{M. leprae} are well conserved and
might possess a functional role.

Functional annotation of most of the above-mentioned gene
products using experimental approaches is vital to elucidate
the DNA repair mechanisms in \textit{M. leprae}. Understanding and
targeting the DNA repair processes in \textit{M. leprae} can be
an important strategy for the development of potential future
therapeutics for leprosy as they are essential for the survival
at different stages of infections. During leprosy infection,
different sets of genes play a vital role in maintaining the
stability of the Mycobacterial genome; therefore, an improved
understanding of the role of DNA repair in the pathogenesis
of Mycobacteria may uncover the great possibility for the
effective treatment against leprosy. Nonetheless, the majority
of the \textit{in silico} work should be confirmed experimentally,
this work provides a profile of those genes responsible for
the maintenance of genome stability, contributing to the
understanding of the mechanisms of genome protection and
mutagenesis in \textit{M. leprae}. It also provides a useful framework
for further investigations on the functions of these genes with
the confirmation of their presence in microarray and qPCR
experiments.

\textbf{Acknowledgement}
Authors would like to thank the scientific staff and students of the
Department of Biotechnology, Indian Institute of Technology
Hyderabad – who contributed in the Bioinformatics analysis.
Our special thanks to all the research staff of the branch of laboratories and the directorate of SIH-R&LC Karigiri for providing access to microarray data and infrastructure to conduct all the scientific experiments.

**Financial support and sponsorship**
Nil.

**Conflicts of interest**
There are no conflicts of interest.

**REFERENCES**

1. Eisen JA, Hanawalt PC. A phylogenomic study of DNA repair genes, proteins, and processes. Mutat Res 1999;435:171-213.
2. Chayot R, Montagne B, Mazel D, Ricchetti M. An end-joining repair mechanism in *Escherichia coli*. Proc Natl Acad Sci U S A 2010;107:2141-6.
3. Shuman S, Glickman MS. Bacterial DNA repair by non-homologous end joining. Nat Rev Microbiol 2007;5:852-61.
4. Gong C, Martins A, Bongiorno P, Glickman M, Shuman S. Biochemical and genetic analysis of the four DNA ligases of mycobacteria. J Biol Chem 2004;279:20594-606.
5. Gong C, Bongiorno P, Martins A, Stephanou NC, Zhu H, Shuman S, et al. Mechanism of nonhomologous end-joining in mycobacteria: A low-fidelity repair system driven by Ku, ligase D and ligase C. Nat Struct Mol Biol 2005;12:304-12.
6. Aravind L, Koonin EV. Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. Genome Res 2001;11:1365-74.
7. Wright DG, Castore R, Shi R, Mallick A, Ennis DG, Harrison L, et al. *Mycobacterium tuberculosis* and *Mycobacterium marinum* non-homologous end-joining proteins can function together to join DNA ends in *Escherichia coli*. Mutagenesis 2017;32:245-56.
8. Della M, Palmbos PL, Tseng HM, Tonkin LM, Daley JM, Topper LM, et al. Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. Science 2004;306:683-5.
9. McMurray DN. Mycobacteria and Nocardia. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 33. Available from: https://www.ncbi.nlm.nih.gov/books/NBK7812/.
10. Shepard CC. The first decade in experimental leprosy. Bull World Health Organ 1971;44:821-7.
11. Vissa VD, Brennan PJ. The genome of *Mycobacterium leprae*: A minimal edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 33. Available from: https://www.ncbi.nlm.nih.gov/books/NBK7812/.
12. Smith EF, Waterman MS. Identification of common molecular substructures. J Mol Biol 1981;147:195-7.
13. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32:1792-7.
14. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O, et al. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst Biol 2010;59:307-21.
15. Duretta A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 2008;36:W465-9.
16. Pfiff MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 2001;29:e45.
17. Cox RA, Kempell K, Fairclough L, Colston MJ. The 16S ribosomal RNA of *Mycobacterium leprae* contains a unique sequence which can be used for identification by the polymerase chain reaction. J Med Microbiol 1991;35:284-90.
18. Phetsukrisi B, Rudeeaneksin J, Supakul P, Wachapong S, Mahotarn K, Brennan PJ, et al. A simplified reverse transcriptase PCR for rapid detection of *Mycobacterium leprae* in skin specimens. FEMS Immunol Med Microbiol 2006;48:319-28.
19. Williams DL, Slayden RA, Amin A, Martinez AN, Pittman TL, Mira A, et al. Implications of high level pseudogene transcription in *Mycobacterium leprae*. BMC Genomics 2009:10:397.
20. Demple B, Harrison L. Repair of oxidative damage to DNA: Enzymology and biology. Annu Rev Biochem 1994;63:915-48.
21. Gros L, Saraparbaev MK, Laval J. Enzymology of the repair of free radicals-induced DNA damage. Oncogene 2002;21:8905-25.
22. Asad NR, de Almeida CE, Asad LM, Felzenszwalb IA, Leitão AC. Fpg and uvrA proteins participate in the repair of DNA lesions induced by hydrogen peroxide in low iron level in *Escherichia coli*. Biochimie 1995;77:262-4.
23. Kurthkoti K, Varshney U. Base excision and nucleotide excision repair pathways in mycobacteria. Tuberculosis (Edinb) 2011;91:533-43.
24. Pearl LH. Structure and function in the uracil-DNA glycosylase superfamily. Mutat Res 2000;460:165-81.
25. Truglio JJ, Croteau DL, Van Houten B, Kisker C. Prokaryotic nucleotide excision repair: The UvrABC system. Chem Rev 2006;106:233-52.
26. Poterszman A, Lamour V, Egly JM, Moras D, Thierry JC, Poch O, et al. A eukaryotic XBP/ERCC3-like helicase in *Mycobacterium leprae*? Trends Biochem Sci 1997;22:418-9.
27. Balasingham SV, Zegeye ED, Homberst H, Rossi ML, Laerdahl JK, Bohr VA, et al. Enzymatic activities and DNA substrate specificity of *Mycobacterium tuberculosis* helicase XBP. PLoS One 2012;7:e36960.
28. Bhagwat AS, Lieb M. Cooperation and competition in mismatch repair: Very short-patch repair and methyl-directed mismatch repair in *Escherichia coli*. Mol Microbiol 2002;44:1421-8.
29. Morimatsu K, Kowalczykowski SC. RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: A universal step of recombinational repair. Mol Cell 2003;11:1347-57.
30. Kowalczykowski SC, Dixon DA, Eggleston AK, Lauder SD, Rehrauer WM. Biochemistry of homologous recombination in *Escherichia coli*. Microbiol Rev 1994;58:401-65.
31. Saves I, Lancelle MA, Daffé M, Masson JM. Inteins invading mycobacterial RecA proteins. FEBS Lett 2000;480:221-5.
32. Davis EO, Thangaraj HS, Brooks PC, Colston MJ. Evidence of selection for protein introns in the recAs of pathogenic mycobacteria. EMBO J 1994;13:699-703.
33. Davis EO, Jenner PJ, Brooks PC, Colston MJ, Sedgwick SG. Protein splicing in the maturation of *M. tuberculosis* RecA protein: A mechanism for tolerating a novel class of intervening sequence. Cell 1992;71:201-10.
34. Singh P, Tripathi P, Silva GH, Pingoud A, Muniyappa K. Characterization of *Mycobacterium leprae* RecA intein, a LAGLIDADG homing endonuclease, reveals a unique mode of DNA binding, helical distortion, and cleavage compared with a canonical LAGLIDADG homing endonuclease. J Biol Chem 2009;284:25912-28.
35. McGlynn P, Lloyd RG. Recombinational repair and restart of damaged replication forks. Nat Rev Mol Cell Biol 2002;3:859-70.
36. De Mot R, Schoofs G, Vanderleyden J. A putative regulatory gene of *Escherichia coli* with an impaired replicative polymerase. J Bacteriol 2000;182:6742-50.