Indian Journal of Biotechnology
Vol 15, January 2016, pp 64-68

Rare codons caused poor expression of mammalian cell entry (Mce1A) gene cloned from *Mycobacterium leprae* in *Escherichia coli*

Ramachandran Sarojini Santhosh[1,2]* and Kuppamuthu Dharmalingam[1,3]

[1]Department of Genetic Engineering, School of Biotechnology, Madurai Kamaraj University, Palkalainagar Madurai 625 021, India

[2]School of Chemical and Biotechnology, SASTRA University, Thanjavur 613 401, India

[3]Department of Proteomics, Aravind Medical Research Foundation, Aravind Eye care System, Madurai 625 020, India

Received 26 October 2014; revised 16 February 2015; accepted 4 April 2015

Mammalian cell entry gene of mycobacterium helps its entry into epithelial and natural target cells. In order to carry out the functional assay of *Mycobacterium leprae* Mce1A, the whole gene was cloned in *Escherichia coli*. An overexpression vector carrying *M. leprae mce1A* gene, under IPTG inducible T5 promoter, was cloned in *E. coli* for expression of encoded protein with N-terminal 6xHis-tag. In the absence of IPTG, *E. coli* cells carrying the *mce1A* gene grew normally but induction of gene expression led to inhibition of cell growth. Western blot analysis using anti-His HRP conjugate showed full length Mce1A protein expressed in low amount. Deletion of N-terminal region having adjacent arginine rare codons resulted in overexpression of truncated protein as inclusion bodies without inhibiting cell division with size reduction of recombinant protein. However the full length protein poorly expressed without size reduction.

Keywords: Codon usage, Mce1A protein, *Mycobacterium leprae*, overexpression, rare codons

Introduction

Intracellular pathogens that replicate in macrophages have several mechanisms to enter and survive inside the phagocytic cell. An encoded protein that helps in the survival as well as efficient entry of mycobacteria into target cells is mammalian cell entry (Mce1A) protein[1]. Several groups have studied the role of Mce1A protein of *Mycobacterium tuberculosis* in considerable detail. Four *mce* loci are organized as operons in *M. tuberculosis* chromosome as unlinked clusters with six *mce* genes (*mce1A-F*) in each operon. *M. tuberculosis mce1A* gene was expressed in *Escherichia coli* as transcriptional fusion protein and it was shown that the amino acids in the N-terminal region spanning the positions 106-163 is involved in the enhanced survival of *M. tuberculosis* in the macrophages[2]. *M. leprae* genome has only one such functional *mce* cluster[3]. Two other *mce* loci of *M. leprae* occur as pseudogenes at different nucleotide positions [*mce1* (2622745-2623859); *mce4* (455770-456328)]. Maintenance of one potentially functional locus (3092446-3100675) in the genome comprising six *mce* genes implies that the function of this operon is essential for the survival and possibly determines the virulence of *M. leprae*. We have demonstrated that this cluster is transcribed as an operon in *M. leprae* as well as in the heterologous host *M. smegmatis*[4]. *M. leprae* being an uncultivable organism, the analysis of its gene functions were carried out in heterologous hosts[4]. The functional assay using truncated *M. leprae* Mce1A shows that the protein promoted the uptake of latex beads into human epithelial, dermal fibroblasts and keratinocytes[5].

In the present report, we have examined the expression of full length *M. leprae mce1A* gene in *E. coli*. The gene was fused to the promoter of pQE31 expression vector and six histidine residues tag allowed purification using NiNTA metal affinity columns, and detection of the protein using anti-His antibodies.

Materials and Methods

Strains and Plasmids

E. coli strains M15 pREP4 (Qiagen Inc., USA) and XL1 Blue MRF' (Stratagene, USA) were used for the expression of *M. leprae* protein and propagation of plasmids, respectively. QIA expression system

*Author for correspondence:
Tel: +91-4362-304000-2321; Fax: +91-4362-264120.
Mobile: +919443455218
santhosh@biotech.sastra.edu; rs_santhosh@yahoo.co.in
†For Supplementry Data, see website: www.nopr.niscair.res.in, or www.niscair.res.in

(Qiagen Inc., USA) was used as described by the manufacturers. Standard DNA manipulations and protein protocols were carried out.

Cloning of *mce1A* Gene in Bacterial Expression Vector

Human leprosy biopsy samples were used for total mRNA isolation of bacilli and random primers were used for the preparation of cDNA using reverse transcriptase. The coding region of *M. leprae* Mce1A (MLMce1A) was amplified from the cDNA using forward (5′TTTGACAAGCTTGCCGATGAC3′) and reverse (5′TACCGGAAGCTTTCATGGGTT3′) primers. PCR product (1326) was precipitated and cloned into pGEM-T vector (Promega). The cloned fragment was released as a *Hind*III fragment from the recombinant plasmid and cloned into pQE31 vector (Qiagen Inc., USA) in the correct reading frame. The resulting construct pQE31MCE$_{1\text{-}441}$ has the *mce1A* gene under the control of bacteriophage T5 promoter. Calculated mol wt of the native Mce1A protein is 47.2 kDa. Additional 28 amino acids were added to the N-terminal end of Mce1A from cloning vectors increasing the mol wt of the recombinant protein to 50.3 kDa. Plasmid pQE31MCE$_{1\text{-}441}$ (Fig. 1A) was digested with *Bam*HI and self-ligated to create pQE31MCE$_{164\text{-}441}$, deleting 179 amino acids. This construct has amino acids spanning 164-441 of MLMce1A and the expected size of the construct was 30.8 kDa. Plasmid constructs were confirmed by restriction digestion and sequence analysis.

Sequence analysis using T5 promoter primers confirmed the open reading frame (ORF). Automated sequencing was performed in the FIST (Funds for Improvement of Science and Technology Infrastructure) facility of the School of Biotechnology, Madurai Kamaraj University, India using ABI Prism 377 Genetic analyzer, Applied Biosystems, USA. Sequencing data was assembled and edited using BioEdit[6].

Transformation and Expression of Proteins in *E. coli*

The constructs, pQE31MCE$_{1\text{-}441}$ and pQE31MCE$_{164\text{-}441}$, were transformed into chemically competent M15 (pREP4) for protein expression. In these transformants, *lac* repressor protein from plasmid pREP4 prevents leaky expression of the protein since the promoter for protein expression is regulated by *lac* operon. Single colony from antibiotic plate was grown overnight at 37°C in LB medium (10 g L^{-1} bactotryptone, 5 g L^{-1} yeast extract and 10 g L^{-1} NaCl, pH 7.0) containing ampicillin (100 mg L^{-1}) and kanamycin (25 mg L^{-1}).

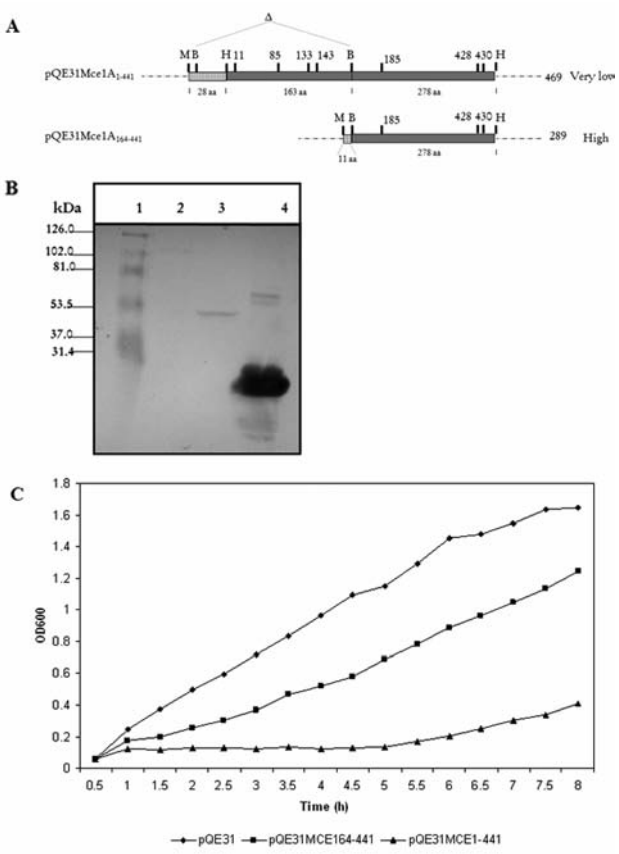


Fig. 1 (A-C)—(A). Schematic representation of the constructs pQE31MCE$_{1\text{-}441}$ and pQE31MCE$_{164\text{-}441}$. The region between two *Bam*HI sites of pQE31MCE$_{1\text{-}441}$ was deleted to construct pQE31MCE$_{164\text{-}441}$ by self-ligation without any change in the reading frame. [Arginine rare codons are marked as 11, 85, 133, 143 and 185. Proline rare codons are marked as 428 and 430. Dotted line represents pQE31 vector; M: First amino acid from the vector, B: *Bam*HI and H: *Hind*III. Gridded rectangles represent the amino acids added from the vector along with 6× histidine tag. Dark rectangles represent the *M. leprae* Mce1A.] (B). Western blot analysis of recombinant *E. coli* lysates using anti-His HRP conjugate. [Lane 1, Prestained marker; 2, Control (50 µg); 3, MLMce1A (50 µg); & 4, MCE$_{164\text{-}441}$ (25 µg)] (C). Growth curve analysis of *E. coli* cells carrying cloned *mce1A* and deletion derivative.

Localization of Protein in *E. coli*

A fresh single colony was inoculated in LB medium (2 mL) containing ampicillin (100 mg L^{-1}) and kanamycin (25 mg L^{-1}) and grown at 37°C for 8 h. Subcultured (5%) in LB (100 mL) and grown to OD$_{600}$ of 0.4. To 50 mL of this culture, IPTG (final concentration, 1 mM) was added, while the other half was kept uninduced. 1 mL of the induced culture was withdrawn at 1 h intervals and cell pellets were stored at 4°C for SDS-PAGE analysis.

Both induced and uninduced cultures were incubated at 37°C for 12 h. The cells from 12 h

culture were pelleted and washed with TE buffer. The cells in TE buffer was sonicated and centrifuged at 10,000 rpm for 4 min at 4°C. The pellet (P, insoluble protein) was resuspended in TE buffer. The supernatant (soluble protein) was again centrifuged at 60,000 rpm for 1 h at 4°C. The supernatant (C, cytoplasmic protein) was collected in an eppendorf and stored at −20°C. The pellet was washed thrice with TE buffer. The pellet was resuspeded in 1 mL of TE buffer with 20 µL of 30% sarkosyl solution and incubated at 23°C for 20 min. After incubation, it was centrifuged at 60,000 rpm for 1 h at 4°C. The supernatant solution (I, inner membrane) was stored at −20°C. The pellet (O, outer membrane) was resuspended in 300 µL of 0.1% SDS. The samples were stored after adding PMSF at −70°C.

SDS-PAGE and Western blotting

E. coli total lysates were separated using 12% polyacrylamide gels and stained with Coomassie brilliant blue R250 (CBR-250). For Western blot analysis, separated proteins were transferred onto a nitrocellulose membrane (Sigma Chemical, USA). Blots were incubated with anti-His HRP conjugated monoclonal antibody (Qiagen Inc., USA). 3,3′-Diaminobenzidine tetra hydrochloride (Sigma, USA) was used for visualization as described by the suppliers. Prestained markers (Sigma) were used for mol wt calculation.

Purification of Recombinant Protein Expressed in *E. coli*

Bacterial cells from 250 mL culture were pelleted and resuspended cells in 25 mL buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl) were sonicated. Inclusion bodies from the lysate were pelleted by centrifugation at 10,000× g for 15 min. The pellet was dissolved in 10 mL of buffer A (100 mM NaH_2PO_4, 10 mM Tris Cl, 6 M guanidinium hydrochloride, pH 8.0) and spun at 10,000×g for 15 min. Chelating sepharose (Amersham Biosciences, Hongkong) charged with nickel ions (Ni^{2+}) was prepared according to the supplier. Clear supernatant was mixed with sepharose and incubated for 1 h at room temperature. The resin was collected by centrifugation (5 min, 800× g). Non-specific proteins were washed off with 5 volume of denaturation buffer (100 mM NaH_2PO_4, 10 mM Tris Cl, 8 M urea) pH 6.3. The recombinant protein was eluted in two steps using 5 mL of denaturation buffer pH 5.9, followed by 10 mL of denaturation buffer pH 4.5. The supernatants obtained from these fractions were analyzed by SDS-PAGE.

Mass Spectrophotometry

Coomassie brilliant blue G250 stained protein band was manually cut using a 2 mm diameter spot cutter. The gel pieces were cut into a size of 1 mm bits and washed twice with water, followed by two washes in 25 mM ammonium bicarbonate in 50% acetonitrile. The gel pieces were dehydrated in 100% acetonitrile for 15 min and dried completely under vacuum. Reduction and alkylation were done using iodoacetamide and dithiothreitol (DTT). In gel, digestion was performed using trypsin (Promega). After overnight digestion, peptides were extracted using acidified acetonitrile, followed by 100% acetonitrile. Total extraction was accomplished by sonication in a sonic water bath. Peptides were spotted using sandwitch method on stainless steel target plate. A modified HCCA matrix was used to suppress the HCCA adducts and to increase the ionization efficiency. Mass spectra were recorded in Kratos Axime CFR Plus instrument equipped with nitrogen laser and a CID tube.

Results

Expression of MLMce1A

The expression of Mce1A protein, 5 h after IPTG induction, was examined by Western blot analysis using anti-His HRP conjugate. A protein band of size 50.3 kDa (Fig. 1B, lane 3) was detected from the cells carrying $pQE31MCE_{1\text{-}441}$, but not in cells carrying the plasmid vector (Fig. 1B, lane 2). The amount of recombinant MLMce1A protein present in 50 µg of total lysate was very low.

The cells carrying *mce1A* failed to grow after the addition of IPTG (Fig. 1C). The cell count did not increase because cells failed to divide but were viable while plating. Fig. 1A shows the positions of five arginine codons AGA and AGG, which were rarely used in *E. coli*. The sequence of *mce* gene of *M. leprae* was compared with that of *M. tuberculosis* (Suppl Fig. S1). *M. tuberculosis mce1A* have only three arginine rare codons and amino acid position 11 is common in Mce1A of *M. leprae* and *M. tuberculosis.*

Removal of Rare Codons Leads to Overexpression of Protein as Inclusion Bodies

In order to examine the effect of rare codons on expression of *mce1A* gene, a deletion derivative—deleted N-terminal 179 amino acids from $pQE31MCE_{1\text{-}441}$—was constructed without the change

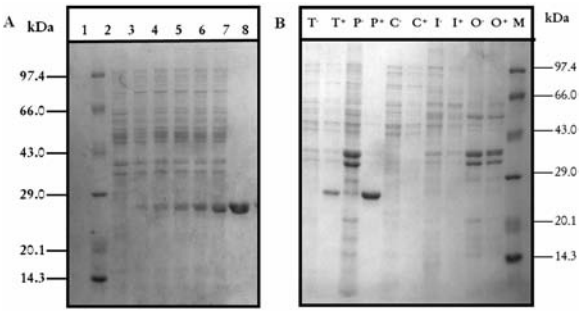
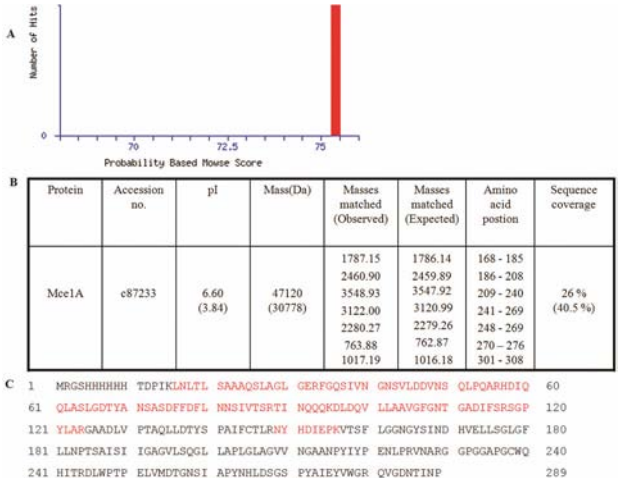


Fig. 2 (A & B)—Expression of cloned Mce1A$_{164-441}$ in *E. coli*: (A). Total lysates (30 µg) from different time intervals were electrophorsed in a 12% SDS polyacrylamide gel. [Lanes: 1, Protein marker standards; 2, Uninduced (12 h); 3, 4, 5, 6 & 7, Total lysates at time points 1, 2, 3, 4 & 12 h, respectively; & 8, Purified Mce1A$_{164-441}$.] (B). Expression of Mce1A$_{164-441}$ in *E. coli* M15 (pREP4, pQE31MCE$_{164-441}$). Cells grown in IPTG free (-) and IPTG containing (+) media were fractionated as described under Materials and Methods. Proteins from different fractions (30 µg) were electrophoresed in a 12% SDS polyacrylamide gel. [Lanes: T, Total lysates; P, Insoluble fractions; C, Cytoplasmic fractions; I, Inner membrane fractions; O, Outer membrane fractions; & M, Protein marker standards.]



B

Protein	Accession no.	pI	Mass(Da)	Masses matched (Observed)	Masses matched (Expected)	Amino acid postion	Sequence coverage
Mce1A	e87233	6.60 (3.84)	47120 (30778)	1787.15 2460.90 3548.93 3122.00 2280.27 763.88 1017.19	1786.14 2459.89 3547.92 3120.99 2279.26 762.87 1016.18	168 - 185 186 - 208 209 - 240 241 - 269 248 - 269 270 – 276 301 - 308	26 % (40.5 %)

C

```
1   MRGSHHHHHH TDPIKLNLTL SAAAQSLAGL GERFGQSIVN GNSVLDDVNS QLPQARHDIQ  60
61  QLASLGDTYA NSASDFFDFL NNSIVTSRTI NQQQKDLDQV LLAAVGFGNT GADIFSRSGP 120
121 YLARGAADLV PTAQLLDTYS PAIFCTLRNY HDIEPKVTSF LGGNGYSIND HVELLSGLGF 180
181 LLNPTSAISI IGAGVLSQGL LAPLGLAGVV NGAANPYIYP ENLPRVNARG GPGGAPGCWQ 240
241 HITRDLWPTP ELVMDTGNSI APYNHLDSGS PYAIEYVWGR QVGDNTINP             289
```

Fig. 3 (A-C)—Mass spectroscopic analysis of Mce1A having amino acids from 164 to 441: (A). Probability Based Mowse Score Ions score is -10*Log(P), where P is the probability. Protein scores greater than 61 are significant ($p<0.05$). (B). Mce1A protein identified with peptide mass finger printing. The corresponding values of Mce1A$_{164-441}$ is shown in paranthesis. (C). Deduced amino acid sequence of Mce1A$_{164-441}$. The sequence covered by the peptide masses are shown in red.

in reading frame. Results in Fig. 2A and B show that, under identical conditions, this construct was highly expressed in *E. coli* as inclusion bodies (27 kDa) and the protein expressed was detectable even after 12 h of induction. Over expression of protein as inclusion bodies was evident in Fig. 2B. It was also shown that cells carrying the deletion derivative grew after IPTG induction (Fig. 1C). In *M. leprae* Mce1A$_{1-441}$, 80% of the arginine rare codons were encoded in the first 163 amino acids. When this region was deleted, the protein was overexpressed as inclusion bodies (27kDa), but it was lesser in size compared to the expected size (30.8 kDa). Two proline rare codons were present at 428 and 430 amino acid positions that might have caused the incomplete synthesis of polypeptide due to the less supply of rare tRNAs. The reduced size of inclusion body indicate a translation arrest may be taking place at proline residues (Fig. 1A, 428th & 430th amino acids), estimated size upto these proline residues (27 kDa) was conforming to the mol wt calculated from SDS-PAGE. The expression of full length MLMce1A was poor even after cloning them in *E. coli* host having the rare tRNAs (data not shown).

Purification of Mce Proteins

MLMce1A could not be purified from the total protein due to its low level of expression. However, we could purify the truncated version since these formed inclusion bodies. The inclusion bodies solubilized in 6 M guanidinium hydrochloride was allowed to bind Ni^{2+} charged chelating sepharose. The bound protein was eluted from the sepharose after washing as described under Materials and Methods. The eluted proteins migrated as a single band in SDS-PAGE gel (Fig. 2A, lane 8). Mass spectrometric analysis showed the analyzed band to be Mce1A protein (Fig. 3).

Discussion

It has been shown earlier that arginine codons AGA and AGG are used infrequently in *E. coli*. Presence of clustered rare codons is shown to have profound effect on the expression of heterologous proteins[7]. The movement of ribosomes along mRNA slows down at rare codon positions due to less availability of isoaccepting tRNAs that leads to the inhibition of protein synthesis and cell division. It is more critical during the log phase of growth because major tRNA species that cognize preferred codon seems to increase in abundance, while most of the minor tRNA species that cognize rare codons decrease in number[8,9]. This occurs as a part of gene regulation at translational level to conserve cell resources. The gene *dnaG* (primase), involved in lagging strand synthesis of DNA during replication, has rare codons and is poorly expressed, although the two neighboring genes *rpsU* and *rpoD* are highly expressed being in the same operon[10-12]. It also implies that the induced gene will be competing with *dnaG* gene for rare

tRNAs leading to the blockage of DNA replication and hence arrest of cell growth. The extent of heterogeneity among genes in *M. tuberculosis* is rather lower compared to *E. coli* and *Bacillus subtilis* based on codon usage, which reflects weaker selection during evolution in *M. tuberculosis* due to its long generation time[13]. *M. leprae* accumulated more rare codons in genes (†Suppl Figs S1 & S2) even though less in GC content compared to *M. tuberculosis*. Translation mediated selection of genes is less in *M. leprae* due to the slow growth. Unlike *M. leprae mce1A* gene, only three rare codons are present in the *mce1A* of *M. tuberculosis*. This implies that Mce1A may be translated at a lesser rate in *M. leprae* as compared to *M. tuberculosis.* Accumulation of rare codons ensures high fidelity in translation and for an intracellular pathogen it may help in survival inside host. Therefore, in a highly host restrictive bacteria like *M. leprae*, five rare codons for arginine amino acid in *M. leprae* Mce1A virulent gene need an intense study[14-15].

Acknowledgement

RSS acknowledges the Council of Scientific and Industrial Research, New Delhi, India for the award of Senior Research Fellowship. He also thanks the Vice Chancellor, SASTRA University, Thanjavur, India for providing infrastructural facilities through Research and Modernisation fund.

References

1 Arruda S, Bomfim G, Knights R, Huima-Byron T & Riley L W, Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells, *Science*, 261 (1993) 1454-1457.

2 Chitale S, Ehrt S, Kawamura I, Fujimura T, Shimono N *et al*, Recombinant *Mycobacterium tuberculosis* protein associated with mammalian cell entry, *Cell Microbiol*, 3 (2001) 247-254.

3 Cole S T, Eiglmeier K, Parkhill J, James K D, Thomson N R *et al,* Massive gene decay in the leprosy bacillus, *Nature* (*Lond*), 409 (2001) 1007-1011.

4 Santhosh R S, Pandian S K, Lini N, Shabaana A K, Nagavardhini A *et al*, Cloning of *mce1* locus of *Mycobacterium leprae* in *Mycobacterium smegmatis* mc^2 155 SMR5 and evaluation of expression of *mce1* genes in *M. smegmatis* and *M. leprae*, *FEMS Immunol Med Microbiol*, 45 (2005) 291-302.

5 Sato N, Fujimura T, Masuzawa M, Yogi Y, Matsuoka M *et al*, Recombinant *Mycobacterium leprae* protein associated with entry into mammalian cells of respiratory and skin components, *J Dermatol Sci*, 46 (2007) 101-110.

6 Hall T A, BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, *Nucleic Acids Symp Ser*, 41 (1999) 95-98.

7 Kane J F, Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*, *Curr Opin Biotechnol*, 6 (1995) 494-500.

8 Emilsson V & Kurland C G, Growth rate dependence of transfer RNA abundance in *Escherichia coli*, *EMBO J*, 9 (1990) 4359-4366.

9 Emilsson V, Naslund A K & Kurland C G, Growth rate dependent accumulation of 12 tRNA species in *Escherichia coli*, *J Mol Biol*, 230 (1993) 483-491.

10 Konigsberg W & Godson G N, Evidence for use of rare codons in the *dnaG* gene and other regulatory genes of *Escherichia coli, Proc Natl Acad Sci USA*, 80 (1983) 687-691.

11 Smiley B L, Lupski J R, Svec P S, McMacken R & Godson G N, Sequences of the *Escherichia coli dnaG* primase gene and regulation of its expression, *Proc Natl Acad Sci USA*, 79 (1982) 4550-4554.

12 Lupski J R, Smiley B L & Godson G N, Regulation of the *rpsU-dnaG-rpoD* macromolecular synthesis operon and the initiation of DNA replication in *Escherichia coli* K-12, *Mol Gen Genet*, 189 (1983) 48-57.

13 Andersson S G E & Sharp P M, Codon usage in the *Mycobacterium tuberculosis* complex, *Microbiology*, 142 (1996) 915-925.

14 Dutta C & Paul S, Microbial lifestyle and genome signatures, *Curr Genomics*, 13 (2012) 153-162.

15 Zhou J-H, Ding Y-Z, He Y, Chu Y-F, Zhao P *et al*, The effect of multiple evolutionary selections on synonymous codon usage of genes in the *Mycoplasma bovis* genome, *PLoS One*, 9 (2014) e108949.