

Regulatory elements in the 5' region of 16SrRNA gene of *Bacillus* sp. strain SJ-101

Braj R. Singh, Abdulaziz A. Al-Khedhairi, Saud A. Alarifi, Javed Musarrat*

DNA Research Chair Program, College of Science, King Saud University, P.O. Box 2455, Riyadh-11451, Saudi Arabia; Javed Musarrat - Email: musarratj1@yahoo.com; Tel: +966-4677249; Fax: +966-4675514; *Corresponding author

Received February 25, 2009; revised April 01, 2009; accepted April 18, 2009; published June 13, 2009

Abstract:

Advancement in bioinformatics with the development of computational tools has enabled the *in-silico* prediction and identification of transcription regulatory factors and other genetic elements with great ease. In this study, computational analysis of sequence homology of 546 bp 5' region of 16SrRNA gene of *Bacillus* sp. strain SJ-101 resulted in identification of promoter-like sequences within the *rrn* gene. Using BPROM tool, the regulatory motifs like -35 and -10 boxes were mapped at 392 and 411 positions, respectively. Furthermore, the *cis*-acting elements as the binding sites for transcription factors (TF) *cpxR* and *argR* were identified at positions 413 and 416 at the upstream of an open reading frame (ORF). The probable functions of the putative TFs were predicted through the Uni-Prot/Swiss-Prot protein database. Search for the Shine-Dalgarno sequence (SD) found the presence of highly conserved SD sequence (AATACC), and a short 42 bp coding sequence/ORF bounded with characteristic transcription start site (AAC) and a stop codon (TGA) at positions 426 and 465 downstream to the promoter elements. A 13 amino acid long translation product of a short ORF has exhibited 100% homology with protein sequences of *Bacillus* spp., while showing some degree of polymorphism with other reference strains. The comparative homology of the small protein exhibited maximum similarity with Prolyl-4 hydroxylase of *Chlamydomonas reinhardtii* with 4.11 ZSCORE. The highly conserved regulatory elements and the putative ORF predicted within the 16SrRNA gene may help understand the role of relatively unexplored short ORFs within *rrn* operon, and their functional products in genetic regulatory mechanisms in eubacteria.

Keywords: 16SrRNA gene; transcription factors; promoter region; *in-silico* analyses

Background:

Bacillus sp. strain SJ-101 isolated from soil, is a nickel (Ni)-tolerant strain with intrinsic potential of plant growth promotion, Ni biosorption and bioaccumulation [1-2]. This non-pathogenic and culturable microorganism is amenable to reverse genetics for functional analysis and regarded as a model system for numerous industrial, medical, and ecological applications. Overall 88% of *Bacillus* genome has been predicted to be either translated into protein or transcribed into stable structural RNA [3]. The involvement of rRNA in the transcription regulation and translation process apart from its major role in the organization of ribosome structure is quite intriguing [4-6]. Although, the ribosomal RNA (rRNA) genes are among the most actively transcribed genes in eubacterial cells, the ubiquity of rRNA in living systems plays an important unique role as a general probe of evolutionary history [7]. It is established that the ribosomal gene rich region of both the prokaryotic and eukaryotic genomes comprise of sequences that are conserved during evolution, interspersed among divergent regions. The very efficient and coordinated transcription of rRNA molecules ensures the delicately balanced constitution of the protein biosynthesis machinery. In addition to the well known infrastructural RNA types, such as tRNA, rRNA, and snRNA, the RNA molecule performs multifarious biologic functions as micro-(miRNA), small interfering-(siRNA), Piwi interacting-(piRNA) and small modulatory-(sm RNA) [8].

The structure of transcription regulatory regions and their specific recognition by combinations of transcription factors (TF) is critical to genetic expression. Lately, the development in computational techniques and availability of whole genome sequences have promoted the *in-silico* prediction and identification of the putative *cis* and/or *trans* acting elements involved in transcriptional regulation

of functional genes. Recently, several regulatory elements within 16S and 23S spacer region have been identified in eubacteria using *in-silico* tools [9-10]. Also, the *in vitro* translation studies in *E. coli* have demonstrated the region within 16SRNA gene encoding for small proteins, and sequence elements characteristics of prokaryotic promoters have been reported [11]. This has prompted us to conduct the computational analysis of 5' region of 16SrRNA gene for the (i) identification of regulatory elements such as -10 and -35 sequences, transcription factors and their corresponding binding sites using BPROM tool, (ii) presence of SD sequence and putative ORF with transcription start site and stop codon, and (iii) functionality of ORF through translational tool using ExpASY Proteomics server. This study has, therefore, integrated the pre-existing biological knowledge to predict the putative *cis*- and *trans*- acting regulatory elements and a coding sequence within the *rrn* gene.

Methodology:

PCR amplification and sequencing of 16S rRNA gene

Genomic DNA from freshly grown culture of strain SJ-101 was isolated and purified by a cetyltrimethylammonium bromide (CTAB) miniprep procedure [12]. Total genomic DNA (50 ng) was used as a template for amplification of 16SrRNA gene employing primers, rD1 (5'-AGAGTTTGATCCTGGCTCAG-3') and rD1 (5'-AAGGAGGTGATCCAGCC-3') complementary to the 5' and 3' regions of eubacterial 16S rRNA genes, respectively. The amplicon was gel purified using a Gel Extraction kit (Qiagen, USA) and sub-cloned into pGEMT-Easy vector (Promega, USA). The selected clone was subjected to sequencing of 16SrRNA gene fragment with SP6 and T7 sequencing primers using ABI prism 3730 sequencer.

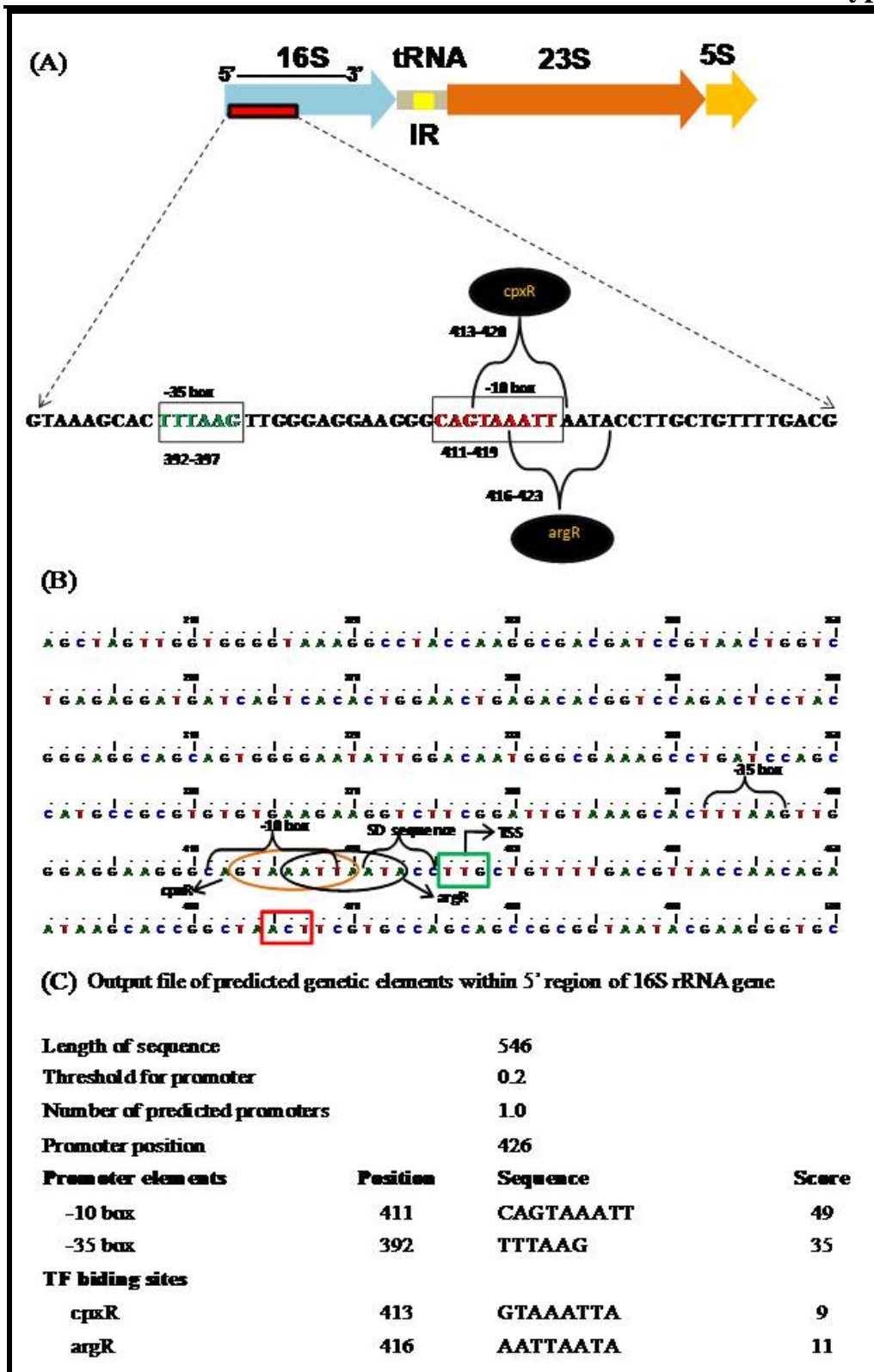


Figure 1: (A): Structural organization of ribosomal operon depicting distal 5' region containing regulatory motifs; (B): nucleotide sequence of functional ORF; and (C): output file indicating the predicted promoter elements and transcription factor binding sites within 5' region of 16S rRNA gene of strain SJ-101.

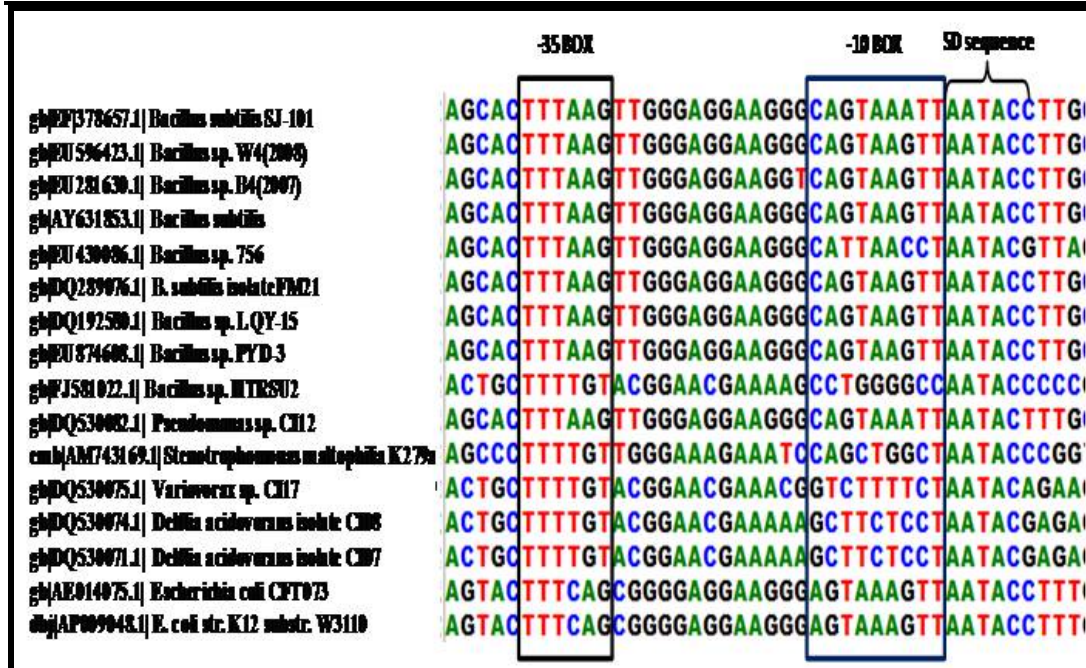


Figure 2: Multiple clustalW alignment of -35 box, -10 box, and Shine-Dalgarno (SD) sequence upstream of ORF identified within of 5' region of 16SrRNA gene of strain SJ-101.

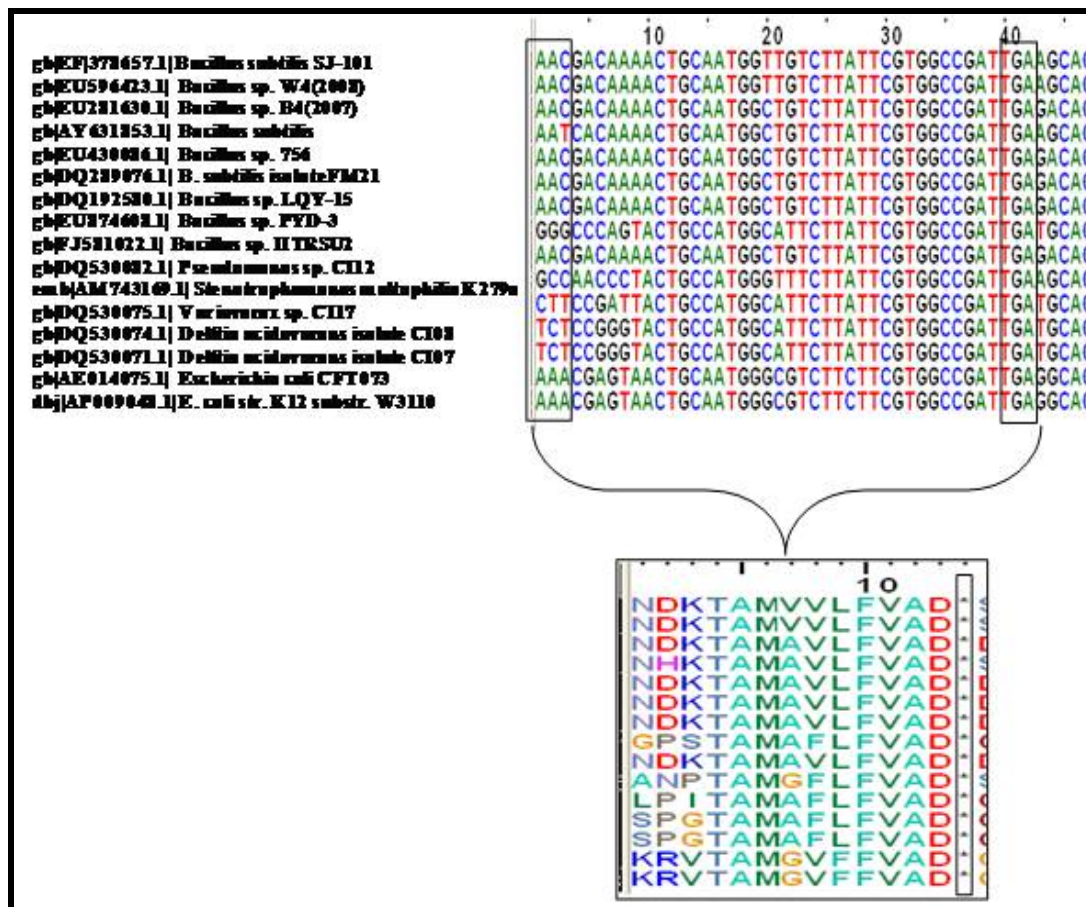


Figure 3: Multiple sequence alignments of putative ORF and its corresponding short peptide identified within 16SrRNA gene of strain SJ-101.

***In-silico* prediction of regulatory motifs**

The Blastn analysis of a partial 564 bp sequence of 5' region obtained from the annotated 1436 bp 16S rRNA gene (Accession No. EF378657) of strain SJ-101 was performed using the sequences available in NCBI Genbank. The homologous 16S rRNA gene sequences were subjected to the multiple sequence alignments [13]. The selected 5' region of 16S rRNA gene of strain SJ-101 was probed through computational tools viz. BPPROM [14] and UniProt/SwissProt [15] for prediction of the putative promoter elements (-10 box and -35 box), and transcription factor binding sites as well as their functions. Translation and homology search of putative ORF was performed using the translation tool through ExPASy Proteomics server [16] and FUGUE v2.s.07 server [17], respectively.

Discussion:

The 16SrRNA gene of bacterial strain SJ-101 has been amplified to obtain a product of 1436 bp using universal primers fD1 and rD1. The amplicon was cloned in pGEMT-Easy vector and sequenced from both the 5' and 3' ends using T7 and SP6 sequencing primers, respectively. Computational analysis of 16SrRNA gene fragment for prediction of TF and other vital regulatory motifs suggested the presence of *cis*-acting sites of promoter sequence, and transcription factors binding sites within 5' region of the gene. Sequence analysis through bioinformatics tools revealed the genetic map of promoter elements viz. the -35 box (TTACGG) and -10 box (TGCTACAAT) at positions 392 and 397, respectively (Figure 1). Multiple sequence alignment of these regulatory motifs suggested their highly conserved status (Figure. 2). Furthermore, the transcription factor (TF) binding sites and their probable functions have been predicted by comparing the entries in protein database UniProt/Swiss-Prot, which has high quality annotations. The three putative TF binding sites have been identified at positions 413 and 416 for cpxR and argR factors, respectively (Figure 1 B and C). The details of each TF binding sites and their probable functions with Swiss-Prot ID are summarized in Table 1 (supplementary material). The results corroborate with the observations of Berg *et al.* [11], who have reported the *in vivo* translation of a segment of *E. coli* *rrnB* 16SrRNA gene. Brosius *et al.* [18] have also suggested the presence of ribosome binding site (Shine-Dalgarno [SD] sequence), an ATG start site codon and a 252 bp (84 codon) ORF starting at position 1187 in the *rrnB* sequence. *In vitro* translation has also been reported for the general region of *rrnB* gene of *E. coli* [19] and homologous region with 16S gene of *Caulobacter crescentus* [20]. Our results of *in-silico* analysis of *rrn* 16S gene in *Bacillus* sp. strain SJ-101, predicting the regulatory elements at the 5' region of the 16SrRNA gene are in accordance with earlier reported *in vitro* translation studies. These regulatory elements within highly conserved 16SrRNA gene may be responsible for synthesis of some small RNA (sRNA) molecules, as suggested by the *in vitro* and *in vivo* translation techniques in many eubacteria. However, the significance of sRNAs is still elusive. Recently, Morita and Aiba [21] suggested that sRNA not only act through base-pairing mechanism but also serve as an mRNA template for small functional proteins to deal with metabolic stress. Nevertheless, many such 'RNA within RNA' (sRNAs) molecules warrant further

investigations in order to understand their cryptic functions.

In this investigation, the presence of transcription start site (AAC) and stop codon (TGA) at positions 426 and 465, respectively revealed the existence of a conserved short (42 bp) ORF within the *rrn* gene. The Blastn analysis of 5' region of 16SrRNA gene sequence showed the homology of the short ORF with members of the genus *Bacillus*, *Pseudomonas*, *Stenotrophomonas*, *Variovorax*, *Delftia* and *Escherichia* (Figure 3). The presence of a canonical ribosome binding site (Shine-Dalgarno sequence) (SD) upstream of transcription start site as shown in Figures 1 and 2 provide a strong evidence for a coding sequence within 5' region of *rrn* gene. It has been confirmed through translation of the ORF using translation tool at ExPASy Proteomics server, which has resulted in a product of a 13 amino acid long peptide. It is interesting to know as to whether the protein encoded by the predicted ORF exhibits any similarity to other proteins available in database. Homology search of the short protein encoded by the ORF within 5' region of 16SrRNA gene of *Bacillus* sp. strain SJ-101 against Fold library HOMSTRAD database through FUGUE v2.s.07 showed maximum similarity with Prolyl-4 hydroxylase of *Chlamydomonas reinhardtii* with 4.11 ZSCORE within 95% confidence limit (Table 2 in supplementary material). Thus, in spite of apparent structural differences, the ORF within the ribosomal RNA showed a high degree of conservation, which suggests some important role of the putative ORF in genetic regulatory mechanisms in eubacteria.

Conclusion:

It is concluded that the 5' region of 16SrRNA of *Bacillus* sp. strain SJ-101 contains a functional ORF with essential regulatory motifs. The *in-silico* translation of a 42 bp ORF decoded a short peptide, which has exhibited maximum similarity with Prolyl-4-hydroxylase of *Chlamydomonas reinhardtii* with 4.11 ZSCORE. It would be interesting for molecular biologists to further experimentally probe the functional role of such a short peptide being encoded within the *rrn* 16S gene. This computational study has integrated the existing biological information for prediction of regulatory elements, TF binding sites, transcription start sites, and coding regions within specified region of 16SrRNA gene. The molecular information unfolded in this study will be important for understanding the role of the novel short coding RNAs and corresponding proteins originating from *rrn* operon, in the genetic regulatory network. Nevertheless, careful integration between the *in-silico* analysis and wet-lab biological approach is crucial for elucidation of the intricate gene regulatory mechanisms.

Acknowledgement:

Financial support through the DNA Research Chair program, King Saud University, Riyadh, KSA is greatly acknowledged.

References:

- [1] S. Zaidi and J. Musarrat, J. Environ. Sci. Health, 39: 681-691 (2004)
- [2] S. Zaidi *et al.*, Chemosphere, 64 (6): 991-997 (2006)
- [3] J. S. Ki, *et al.*, J. Microbiol. Methods, (2009)

- [4] C. R. Woese, *et al.*, *Microbiol. Rev.*, 47(4): 621-669 (1983)
- [5] H. F. Noller, *Annu. Rev. Biochem.*, 53: 119-162 (1984)
- [6] R. Brimacombe, *et al.*, *Prog. Nucleic Acid Res. Mol. Biol.*, 28: 1-48 (1983)
- [7] M.T. La Duc, *et al.*, *J. Microbiol. Methods*, 56: 383-394 (2004)
- [8] F. F. Costa, *Gene*, 386 (1-2): 1-10 (2007)
- [9] N. Dwivedi, *et al.*, *Bioinformatics*, 2 (8): 363-368 (2008)
- [10] S. Srivastava, *et al.*, *Bioinformatics*, 3(4): 173-176 (2008)
- [11] K. K. Berg, *et al.*, *J. Bacteriol.*, 169(4):1691-1701 (1987)
- [12] F. M. Ausubel *et al.*, (Eds.), *Short Protocols in Molecular Biology*, John Wiley & Sons, New York, 2-4 (1995)
- [13] J. D. Thompson, *et al.*, *Nucleic Acids Res.*, 22: 4673-4680 (1994)
- [14] <http://www.softberry.com/berry.html>
- [15] <http://www.uniprot.org>
- [16] <http://www.expasy.ch/tools/dna>
- [17] <http://tardis.nibio.go.jp/fugue>
- [18] J. Brosius, *et al.*, *Proc. Natl. Acad. Sci. USA*, 75: 4801-4805 (1978)
- [19] D. Stüber and H. Bujard, *Proc. Natl. Acad. Sci. USA*, 78: 167-171 (1981)
- [20] K. Amemiya, *et al.*, *J. Mol. Biol.*, 187: 1-14 (1986)
- [21] T. Morita and H. Aiba, *Proc. Natl. Acad. Sci. USA*, 104: 20149-20150 (2007)

Edited by P. Kanguane

Citation: Singh *et al.*, *Bioinformatics* 3(9): 375-380 (2009)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material

TF	TF binding site	Position	Function	Score	Swiss-Prot ID
cpxR	GTAAATTA	413 - 420	Member of the two-component regulatory system cpxA/cpxR. This system combats a variety of extracytoplasmic protein-mediated toxicities. It performs this function by increasing the synthesis of the periplasmic protease, degP as well as that of cpxP protein.	9	P0AE88 (CPXR_ECOLI)
argR	AATTAATA	416 - 423	Represses the synthesis of biosynthetic enzymes and activates the arginine catabolism. Controls the transcription of the two operons rocABC and rocDEF.	11	P17893 (ARGR_BACSU)

Table 1: Putative transcription factors binding sites within the 5' region of 16SrRNA gene of *Bacillus* sp. strain SJ-101.

Profile Hit	ZSCORE	Confidence	Protein in fold library
hs2jiga	4.11	Likely	Prolyl-4 hydroxylase - <i>Chlamydomonas reinhardtii</i>
hsd1cpo_2	2.51	Guess	Cloroperoxidase
hs1skha	2.51	Guess	Major prion protein 2 - <i>Bos taurus</i>
hsd1b8ja	2.43	Guess	Alkaline phosphatase
alk_phosphatase	2.40	Guess	Alkaline phosphatase - <i>E. coli</i>
hs1o75a	2.35	Guess	47 kda membrane antigen
hs3c01d	4.11	Guess	Surface presentation of antigens protein spas <i>Salmonella typhimurium</i>

Table 2: Homology of the short protein encoded by the ORF within the 5' region of 16SrRNA gene of *Bacillus* sp. strain SJ-101 using Fold library HOMSTRAD database through FUGUE v2.s.07.