Vol 3 Issue 12 Jan 2014 ISSN No : 2230-7850

International Multidisciplinary Research Journal

Indian Streams Research Journal

Welcome to ISRJ

RNI MAHMUL/2011/38595 **ISSN No.2230-7850**

Indian Streams Research Journal is a multidisciplinary research journal, published monthly in English, Hindi & Marathi Language. All research papers submitted to the journal will be double - blind peer reviewed referred by members of the editorial board.Readers will include investigator in universities, research institutes government and industry with research interest in the general subjects.

Address:-Ashok Yakkaldevi 258/34, Raviwar Peth, Solapur - 413 005 Maharashtra, India
Cell : 9595 359 435, Ph No: 02172372010 Email: ayisrj@yahoo.in Website: www.isrj.net

Indian Streams Research Journal
ISSN 2230-7850
Volume-3 | Issue-12 | Jan-2014
Available online at www.isrj.net

ANALYSIS OF *16S RIBOSOMAL RNA* GENE SEQUENCEANDUSED AS A RAPID TOOL FOR IDENTIFICATION OF BACTERIA

Dixit Ojas[1*], Raut Amol[2a] and Choudhary Sameer[2b]

Abstract:-One of the most interesting experiences in introductory microbiology is attempt to identify an unknown microorganism. This work was dependent on identifying the unknown bacteria.16S *rRNA* is the most studied gene for identification of unknown bacteria. The use of 16S *rRN*A gene sequences to study bacterial phylogeny and taxonomy has been by far the most common housekeeping genetic marker used for a number of reasons. These reasons include; its presence in almost all bacteria, the function of the 16S *rRNA* gene over time has not changed, suggesting that random sequence changes are a more accurate measure of time (evolution) and the 16S *rRNA* gene is large enough for informatics purposes. The principle objective of this study is to define scientific parameters to discriminate unknown bacterial cells and develop a reliable and sensitive molecular method to assess the risk of false positive. So to deal the false positive result, a new reverse transcription PCR assay based on used of *mRNA* as a viability marker is implemented. A pair of universal degenerate primer was selected to amplify the 16S *rRNA* gene of unknown bacteria. The molecular identification protocol introduces PCR and cycle sequencing.

Keywords:16S rRNA,cDNA synthesis, Staphylococcus epidermidis.

INTRODUCTION

Sequencing of the *16S ribosomal RNA (rRNA)* gene has served as an important tool for determining phylogenetic relationships between bacteria. The first bacterial *16S rRNA* was sequenced by Ehresmann *et al*. in 1972 for Escherichia coli (Han et al. 2006). Analysis of 16S rRNA gene sequence is the best method to identify bacterial species (Kolbert and Persing, 1999; Drancourt *et al*. 2000). The part of the DNA now most commonly used for taxonomic purposes for bacteria is the 16S *rRNA* gene (Jill *et al*. 2004).

The utilization of microarrays for transcription analysis, cDNA synthesis, Reverse Transcription PCR (RT-PCR) and sequencing of gene of interest requires firstly the extraction of RNA from a biological source of interest(Lopez *et al.* 2007). RNA isolation was done by the TRIzol® reagent(by Life Technologies, Ambion). It is a ready to use reagent, designed to isolate high quality total RNA from the sample. RNA extraction coupled to a RT-PCR based on amplification of 16S rRNA gene. The genomes of an increasing number of organisms have been sequenced; however, the exact structure of many transcripts remains unresolved, particularly their 5' untranslated regions (UTRs). An anchor primer is ligated to the 5' end of the mRNA molecule prior to reverse transcription, which subsequently synthesized cDNA molecules incorporate into their sequence. Therefore, only full-length cDNA molecules are amplified by PCR using the anchor primer and gene-specific primer(Bower et al. 2010). A Sanger dideoxy sequencing reaction is performed on the templates on a solid surface to generate a ladder of DNA fragments randomly terminated by fluorescently-labeled dideoxyribonucleotides. The Sanger sequencing is based on process of DNA template to be sequenced. Strand elongation and termination give a newly synthesized strand of targeting the *16S rRNA* gene *cDNA* sequence.

A new approach to rapid sequence comparison, basic local alignment search tool (BLAST), directly approximates alignments that optimize a measure of local similarity, the maximal segment pair (MSP) score (Altschul *et al.* 1990). Recent mathematical results on the stochastic properties of MSP scores allow an analysis of the performance of this method as well as the statistical significance of alignments it generates. *16S rRNA*gene sequences are searched in database using a dynamic algorithm program called BLAST (Basic local alignment search tool).ClustalW is the multiple sequence alignment (MSA) tool, used for the comparison of sequences. It identifies regions of similarity that may be a consequence of functional, structural or evolutionary relationships between sequences.

The superfamily database contains a library of hidden Markov models representing all proteins of known structure. The database based on the SCOP (Structural Classification of Proteins) 'superfamily' level of protein domain classification which groups together the most distantly related proteins which have a common evolutionary ancestor (Gough *et al*. 2001). Documentation entries describing protein domains, families, functional sites as well as patterns and profiles to identify them, can be searched by Expasy-Prosite. PROSITE is one of these motif descriptor databases. It is an annotated collection of biologically meaningful motif descriptors dedicated to the identification of protein families and domains (Hulo *et al.* 2004). The PROSITE database uses two kinds of motif descriptors, each having its own strengths and weaknesses defining its area of optimum application (Sigrist *et al*., 2002). Prosite used for annotation of domains and features of UniProtKB/ Swiss-Prot entries.

MATERIAL AND METHODS

Bacterial strain/isolate and Growth condition

Bacterialcells were grown in Luria-Bertani(LB) media to exponential phase (OD_{600} 0.2; approximately 2×10^8 cells/ mL).

RNA isolation

TRIzol®reagent (Chomczynski and Sacchi, 1987) used for total RNA isolation, Chloroform, Isopropyl alcohol, 75% ethanol, RNase-free water, was ready prior to use. Harvest 1.5 mL of bacterial culture in 1.5 mL eppendorf and centrifuged at 12,000×g for 10 min. Supernatant was discarded and 0.75 mL TRIzol® reagent was added per 0.25 mL of sample. Cells were lyzed in sample by pippeting up and down several times. Homogenized sample was incubated for 5 min. at room temperature (RT) to permit complete dissociation of the nucleoprotein complex. 0.15 mL of chloroform was added per 0.75 mL of TRIzol® Reagent used for homogenization. Tube was shaken vigorously by hand for 15 sec. and incubated at RT for 2–3 min. Sample was centrifuged at 12,000×g for 15 min. at 4°C. Aqueous phase of the sample was removed by angling the tube at 45° and pipetting the solution out. Avoid drawing any of the interphase or organic layer into the pipette when removing the aqueous phase. Aqueous phase was placed into a new tube. 0.38 mL of 99.9% isopropanol was added to the aqueous phase, per 0.75 mL of TRIzol® Reagent used for homogenization and incubated for 10 min at RT. Centrifuged at 12,000×g for 10 min. at 4°C. Supernatant was removed from the tube, leave only the RNA pellet. Pellet was washed with 0.75 mL of 75% ethanol per 0.75 mL of TRIzol® Reagent used in the initial homogenization. Sample was vortexed briefly, and then centrifuged at 7500×g for 5 min. at 4°C. Wash was discarded, air dried. RNA pellet was resuspended in 20 µL RNase-free water by passing the solution up and down several times through a pipette tip. Sample was incubated in a water bath at 55-60°C for 10-15 min. and stored at 4°C.

First strand cDNA synthesis

RNA sample contaminating DNA was not analyzed further. The presence of target mRNA, 16S rRNA gene was analyzed by RT-PCR. M-MLV RT (Moloney Murine Leukemia Virus Reverse Transcriptase) was used synthesize cDNA from RNA. For master mix preparation; 1.0µL of Oligo dt primer (500µg/mL), 1.0µL 16S rRNA reverse primer (10µM), 1.0µL of dNTPs (10mM), 1.6µL of SMQ (sterile milli Q) water, 9.0µL of RNA sample (1ng-5ng) was added into the PCR tube. Mixture was heated at 65°C for 5 min. and quickly chilled on ice and 4µL of 5X First strand buffer (FS buffer) and 2µL of 0.1 M DTT (Dithiothreitol) was added. Content was mixed and incubated at 37°C for 2 min. 1µL (200 units) of M-MLV RT was added, mixing was done and cDNA was synthesized at; incubated for 50 min. at 37°C, Inactivation the reaction was done by heating at 70°C for 15 min. and sample was stored at 4°C. The cDNA was used as a template for amplification in RT-PCR.

Reverse Transcription PCR amplification of 16S rRNA gene:

16S rRNA gene amplification was proceeding; 2µL of cDNA was used as template for 25µL PCR reaction using 16S rRNA forward and *16S rRNA* primers. For master mix preparation; 16.5µL of SMQ water, 2.5µL of 10× PCR buffer, 1µL of 10µM primer for *16S rRNA* gene amplification reaction, 0.75µL of 50mM $MgCl_2$, 0.5µL of Taq DNA polymerase (5U/µL), 0.5µL of 10mM dNTPs, 2µL of cDNA as template was added into the PCR tubes. Reverse Transcription PCR cycle was carried out; 95°C for 5 min., 35 cycles of 95°C for 30 sec., 57°C for 30 sec., and 72°C for 45sec., finished with an extension at 72°C for 7 min. RT-PCR products were electrophoresed on 1% agarose gel.

Sequencing of gene of interest

1)Purification of PCR product: The complete method for purification can be summarized in four simple steps: Dilution of PCR product, Addition of Binding Buffer, Addition of Wash Buffer and Gel based analysis. For the dilution, 20µL of PCR product and add 80µL of SMQ water was taken. 500µL of PCR binding buffer was added to this eppendorf tube. Content was mixed. Total 600µL of the content was transferred on purification column and allowed to pass through it. Then centrifuged at 12,000 rpm for 1 min. Column was removed and binding buffer was discarded. To this 500µL of PCR washing buffer was

added and centrifuged at 12,000 rpm for 1 min. Wash was discarded. Dry spin of column in tube was carried at 12,000 rpm for 1 min. and Column was transferred in new 1.5mL eppendorf tube. 20µL of SMQ water was added at the center of the column. Incubated at RT for 2 min. and then centrifuged at 12,000 rpm for 1 min. Column was discarded and cap of eppendorf was closed. The sample was ready for quality check (QC). After this step sample was used for cycle sequencing.

2) Amplification by Cycle Sequencing:3 µL of the diluted template DNA was used for further reaction make up. 7.4µL of SMQ water, 1µL of RR (reaction ready) mix, 3.6µL sequencing buffer, mixing of the content was done and distributed in 2 different tubes and 3 µL of PCR product and 1µL of forward and 1µL of reverse primer were added in each tube. Cycle sequencing was carried out for all reaction. 96°C for 1 min., 25 cycles of 96°C for 10 sec., 57°C for 5 sec., 72°C for 4 min.

3)Clean-up process for sequencing: 2.5µL of 125 mM EDTA and 60µL of absolute ethanol to 7µL PCR products were added in a tube. Mixture was incubated at RT for 15 min. Centrifuged at 12,000 rpm for 20 min. at 18°C. Supernatant was slowly discarded and 60µL of 70% ethanol was added for washing. Centrifugation was done at 10,000 rpm for 10 min. The wash procedure was repeated for one more time. Tubes were dried and 10µL Hi-Di formamide was added to each reaction tube. Kept at 95°C for 3 min. (heat shock treatment) quickly chilled on ice. Sample was ready for sequencing. 10µL reaction sample was loaded into individual wells of the sequencer plate and programmed for the sequencing.

***16S rRNA* gene sequence analyses using BLAST and ClustalW**

The BLAST output of both cDNA sequences of *16S rRNA* gene was obtained. Sequences were copied from the Chromas Lite Software, sequence reader. Nucleotide BLAST (nblast) browser from NCBI site was opened. Its web address is http://blast.ncbi.nlm.nih.gov/Blast.cgi. The selected sequences were pasted into the Search box of nblast homepage. Visit the ClustalW home page; http://www.ebi.ac.uk/Tools/msa/clustalw2/ of EMBL-EBI, Which displays the homepage for ClustalW. More than one sequence was uploaded from the result obtained from BLAST output. The maximum similar and closely related sequences obtained in BLASTing our desired cDNA sequence were chosen. MSA of these sequences was carried out to study Phylogenetic relation and comparison of the sequences. These cDNA sequences were pasted in the input area and the data was submitted. Output for the BLAST and ClustalW was displayed within a few seconds.

Analyzing sequence using Expasy-Prosite and SCOP

The Expasy-Prosite output of both cDNA sequences of *S.epidermidis* 16SrRNA gene was obtained. Sequences were copied from the Chromas Lite Software. Expasy-Prosite; homepage was opened and its web address is http://prosite.expasy.org/. The selected sequences were pasted into the scan box of Expasy-Prosite homepage to study conserved regions/domains of the sequences. Superfamily is a database of structural and functional annotation for all proteins and genomes. Copy FASTA formats sequence and paste it in the search box of SCOP homepage its web address is http://supfam.cs.bris.ac.uk/SUPERFAMILY/hmm.html. Output for the Expasy-Prosite and SCOP was displayed within a few seconds.

RESULTS AND DISCUSSION

RNA isolation

The choice of TRIzol® reagent was investigated for the RNA isolation from the bacterial sample. Moreover, the time and speed at which the tubes were shaken were critical, since low speeds and short shaking times favored the isolation of DNA while high speeds and long times favored the isolation of RNA. The quality of total RNA extracted from the bacterial culture (Figure 1) by TRIzol® reagent was monitored by gel electrophoresis on 1% agarose gel. The image show the intact band. The image shows positive result for RNA extraction with a peculiar band. The intensity of the band image also describes about the quantity of the RNA extracted in theoretical terms. It gives the idea of the purity of product and type of contamination present.

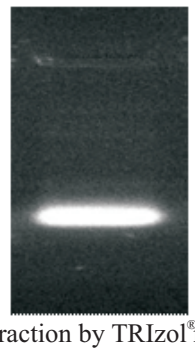


Figure 1 Electrophoresis of RNA extraction by TRIzol® reagent was loaded on 1% agarose gel

Reverse Transcription PCR

The use of ladder was significant for Reverse Transcription PCR interpretation. The comparative result of size of ladder and the amplified product band size was predicted. The position of ladder band and the product band help in estimation of size. Also the primers play a significant role for amplification. A typical electrophoresis result of a sensitive Reverse Transcription PCR is shown here (Figure 2). The following electrophoresis image displays expected Reverse Transcription PCR result that was obtained with *16S rRNA* gene primers in bacterial culture. The product size amplified by the *16S rRNA* gene primer was 700-800 bp. This result also justifies the complementary sequence of the primer used was correct. Thus, I obtained the amplification of *16S rRNA*genefor which the primers were used.

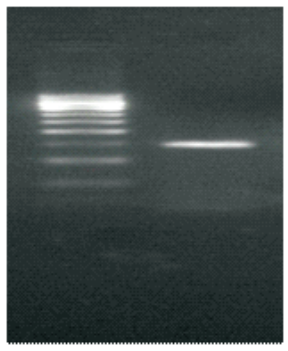


Figure 2 Reverse transcriptions PCR amplification of mRNA of *16S rRNA* gene using the total RNA isolated fromBacterial culture. Lane 1 loaded with ladder sequence of 100-1000 bp and lane 2 loaded with Reverse Transcription PCR product of *16S rRNA* gene

Sequencing

The computer output for a sequencing run consists of chromatogram. A chromatogram (sometimes also called electropherogram) is the visual representation of a DNA sample produced by a sequencing machine. Good sequence generally begins roughly around base 20, and was represented by tall distinct peaks that have little overlap. Poor reactions will result in low or multiple peaks as illustrated below.

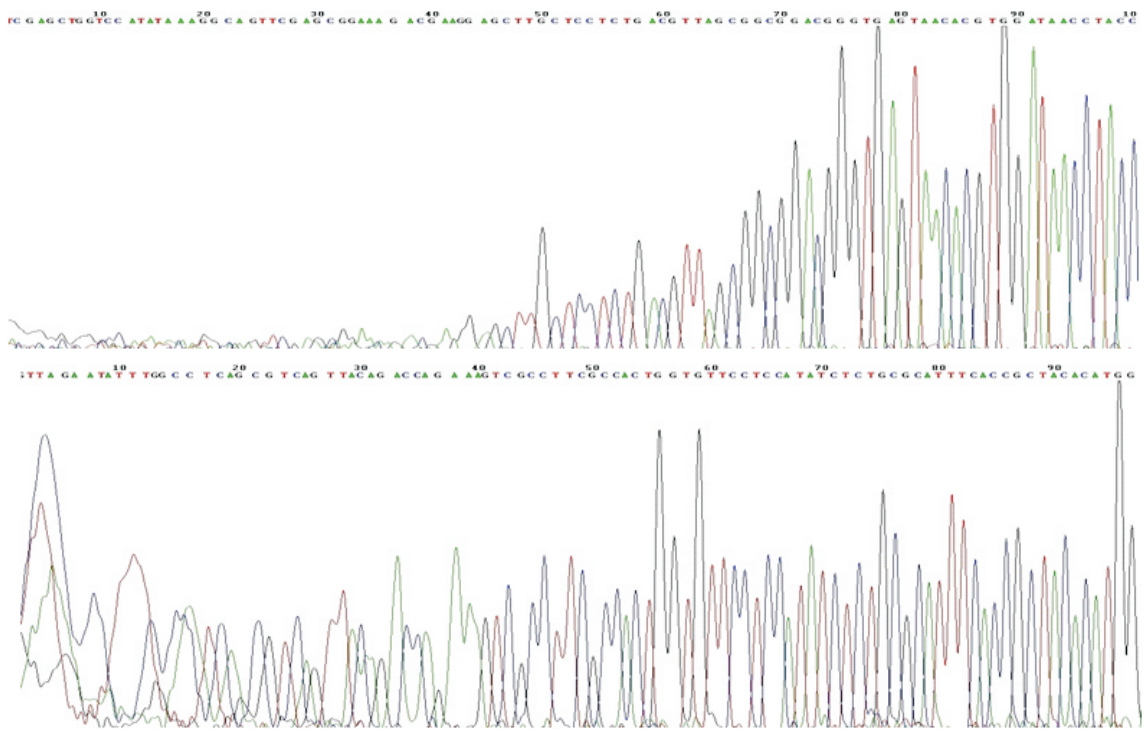


Figure 3 Chromatogram of *16S rRNA* gene primer amplified sequences (a) *16S rRNA* gene forward primer amplified sequence (b) *16S rRNA* gene reverse primer amplified sequence

The sequencing result obtained seems colorful and interesting. Each color signifies one base. Green for Adenine, Red for Thiamine, Black for Guanine and Blue for Cytosine. Also the peak gives idea about the probability of the base at that position. There was direct relationship between the peak and its probability at that position. The output of sequence in Chromas shows excellent chromatogram, with clean, distinct and very low to no background noise. Chromatogram with noisy peaks; which signifies the low quality of the result. The sequence starts off was nice. At the end point chromatogram shows noisy/erratic signal peaks, so may the salt/alcohol contamination was present. The peaks were sharp and distinct but gradually drop in size by the time. There was no GC rich region was observed. If the GC rich region was present it requires high melting temperature to denature. There were usually weak peaks effect occurring within the sequence, which identifies as a lower case letters in the sequence called as miscalls. Weak peaks result from suppression of signal following a strong signal most commonly occurring for G's after A's and often for G's after C's. There is no poly A region was observed. Also, we can observe double peaks at one nucleotide, flanked by clean single-peak sequence. This can indicate that two alleles of the PCR'ed gene were different, and one base was present on one allele, while the second was present on the other allele.

BLAST

BLAST analysis gave the confirmation for my amplified product. The alignment score for different color was mentioned in the colorful bar displayed on the top of query. Graphic summary of the 16S rRNA cDNA sequence obtained using 16S rRNA forward and 16S rRNA reverse primers for bacterial species along with its significant alignment output was displayed. For amplified sequences of 16S rRNA gene gives >=200 alignment score. From this result, the sequence amplifies from the unknown bacteria was Staphylococcus epidermidis.

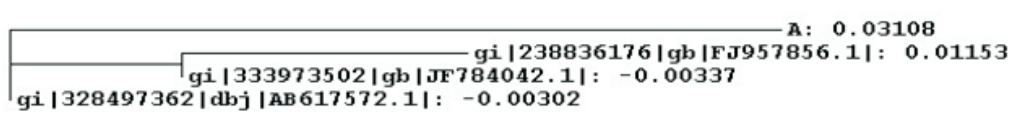
ClustalW

Sequences may be differing in their arrangement. Aligned sequences of nucleotide or amino acid residues were typically represented as rows within a matrix. Gaps were inserted between the residues so that identical or similar characters were aligned in successive columns. If two sequences in an alignment share a common ancestor, mismatches can be interpreted as point mutations and gaps as indels (that is, insertion or deletion mutations) introduced in one or both lineages in the time since they diverged from one another. Each sequence is aligned with our target sequence with its base number. Sequence of interest for *16S rRNA* gene was taken from the BLAST result output (GenBank accession numbers *16S rRNA* forward sequence is FJ957856.1, JF784042.1, AB617572.1 and KC310827.1, KF011975.1, KC443110.1 for reverse sequence). This shows the aligned multiple sequence in a colorful format. '*' indicates identical, '.'indicates semi-conserved substitution and ':' indicates conserved substitution.

```
gi|238836176|gb|FJ957856.1|     --------------------TGCAG-TCGAGCG-AAAGACG-AGGAGCTTG 28
gi|333973502|gb|JF784042.1|     ----------TGCTATACATGCAG-TCGAGCGAACAGACG-AGGAGCTTG 38
gi|328497362|dbj|AB617572.1|    ----TGGCGGTGCTATACATGCAG-TCGAGCGAACAGACG-AGGAGCTTG 44
A                               TCGAGCTGGTCCATATAAAGGCAGTTCGAGCGGAAAGACGAAGGAGCTTG 50
                                         **** ******* *.***** *********

gi|238836176|gb|FJ957856.1|     CTCCTCTGACGTTAGCGGCGGACGGGTGAGTAACACGTGGATAACCTACC 78
gi|333973502|gb|JF784042.1|     CTCCTCTGACGTTAGCGGCGGACGGGTGAGTAACACGTGGATAACCTACC 88
gi|328497362|dbj|AB617572.1|    CTCCTCTGACGTTAGCGGCGGACGGGTGAGTAACACGTGGATAACCTACC 94
A                               CTCCTCTGACGTTAGCGGCGGACGGGTGAGTAACACGTGGATAACCTACC 100
                                **************************************************
```

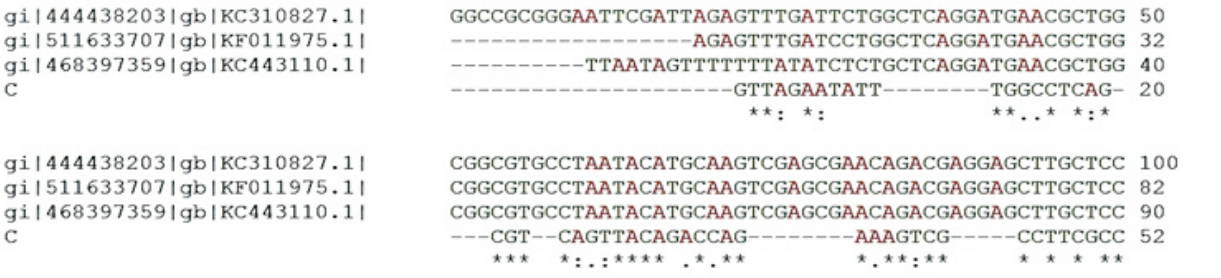
(a)



(b)

Figure 5 ClustalW output of *16S rRNA* forward primer amplified sequence (a) Multiple sequence alignment: Sequences of interest were aligned with *16S rRNA* sequence (b) Phylogenetic tree of *16S rRNA* of S.epidermidis with respect to the closely related sequences available in GenBank

The Multiple sequence alignment in figure 5 is divided into two parts. First part is (figure 5a), Sequence alignment and comparison of sequences and it showed that *16S rRNA* sequence is 100% similar to the related sequences. Because it showed '*' at each position means all are identical with each other. Second part is (figure 5b), Phylogenetic tree. The phylogenetic analysis of this sequences showed that *16S rRNA* sequence has highest similarity with the FJ957856.1 accession number sequence. Because distance between these two sequences is much smaller than the other sequences.



(a)

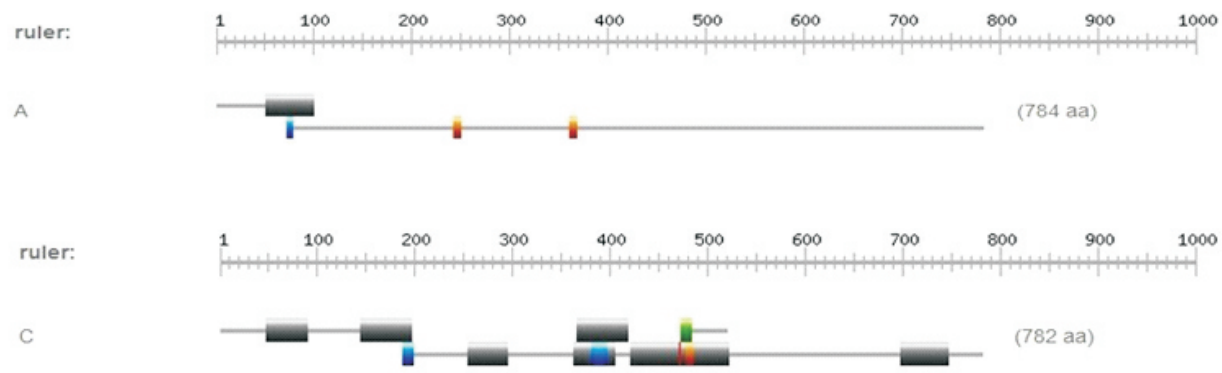
(b)

Figure 6 ClustalW output of *16S rRNA* reverse primer amplified sequence (a) Multiple sequence alignment: Sequences of interest were aligned with *16S rRNA* sequence (b) Phylogenetic tree of *16S rRNA* of S.epidermidis with respect to the closely related sequences available in GenBank

The Multiple sequence alignment in figure 6 is divided into two parts. First part is (figure 6a), Sequence alignment and comparison of sequences and it showed that *16S rRNA* sequence is not 100% similar to the related sequences. Because it showed gaps within the sequence so may be insertion or deletion is present. We can also observe conserved and semi-conserved substitution. Second part is (figure 6b), Phylogenetic tree. The phylogenetic analysis of this sequences showed that *16S rRNA* sequence distantly related to the other sequences. Because distance between these sequences with 16S rRNA sequence is much larger than the other sequences.

Expasy-Prosite

Expasy can access a wide range of resources in many different domains, such as proteomics, genomics, phylogeny/evolution, systems biology, population genetics, Transcriptomics, etc. A sequence logo was a graphical display of a multiple sequence alignment consisting of colour-coded stacks of letters representing amino acids at successive positions. Domain Search pattern for *16S rRNA* gene sequences, which represents graphical view of the domain or family of the protein. For S.epidermidis,*16S rRNA* forward primer amplified sequence contains four conserved regions/domains. *16S rRNA* reverse primer amplified sequence contains twelve conserved regions/domains.



(b)

Figure 4 Expasy-Prosite output of *S.epidermidis 16S rRNA* gene: (a) Conserved regions/domains of *16S rRNA* forward primer amplified sequence (b) Conserved regions/domains of *16S rRNA* reverse primer amplified sequence

Table 1 Conserved regions/domains of *16S rRNA* forward primer amplified sequence

4 Conserved regions/ Domains		
No. of domains	**Name**	**Location (No. of a.a.)**
1	VWFC domain signature	51-100
1	Tubulin subunit alpha, beta and gamma signature	74-80
2	2Fe-2S ferredoxin-type iron-sulfur binding region signature	244-252, 362-370

Table 2 Conserved regions/domains of *16S rRNA* reverse primer amplified sequence

12 Conserved regions/ Domains		
No. of domains	**Name**	**Location (No. of a.a.)**
8	VWFC domain signature	49-91, 145-197, 255-295, 363-405, 366-419, 422-471, 467-522, 698-747
1	4Fe-4S ferredoxin-type iron-sulfur binding region signature	189-200
1	Insulin-like growth factor-binding protein (IGFBP) N terminal domain signature	381-396
1	EGF-like domain signature1	473-484
1	2Fe-2S ferredoxin-type iron-sulfur binding region signature	479-487

Description of the Conserved regions/Domains described below:

1.VWFC domain signature: The VWFC domain is named after the von Willebrand factor (VWF) type C repeat which is found twice in this multidomain protein and thought to participate in oligomerization, but not in the initial dimerization step.
2.Tubulin subunit alpha, beta and gamma signature: Tubulins, the major constituent of microtubules are dimeric proteins which consist of two closely related subunits (α and β). Tubulin binds two molecules of GTP at two different sites (N and E). At the E (Exchangeable) site, GTP is hydrolyzed during incorporation into the microtubule.
3.2Fe-2S ferredoxin-type iron-sulfur binding domain signature: Ferredoxins form a distinct 2Fe-Ferredoxin family. They are proteins of around one hundred amino acids with four conserved cysteine residues to which the 2Fe-2S cluster is ligated. This conserved region is also found as a domain in various metabolic enzymes.
4.4Fe-4S ferredoxin-type iron-sulfur binding domain signature: The 4Fe-4S Ferredoxins, which are found in bacteria and which are thus often referred as 'bacterial-type' Ferredoxins. The structure of these proteins consists of the duplication of a domain of twenty six amino acid residues; each of these domains contains four cysteine residues that bind to a 4Fe-4S center.
5.Insulin-like growth factor-binding protein (IGFBP) N terminal domain signature: The insulin-like growth factors (IGF-I and IGF-II) bind to specific binding proteins in extracellular fluids with high affinity. These IGF-binding proteins (IGFBP) prolong the half-life of the IGFs and have been shown to either inhibit or stimulate the growth promoting effects of the IGFs on cells culture. The N-terminal domain is ~80 residues in length and has an L-like structure (see <PDB:1WQJ>). It can be divided into two subdomains that are connected by a short stretch of amino acids. The two subdomains are perpendicular to each other, creating the "L" shape for the whole N-terminal domain.
6.EGF-like domain signature 1: A common feature of all EGF-like domains is that they are found in the extracellular domain of membrane-bound proteins or in proteins known to be secreted (exception: prostaglandin G/H synthase).

SCOP

Superfamily: A database of sequence superfamily of protein domains: The superfamily database is based on the SCOP classification of protein domains. SCOP is a structural domain based hierarchical classification with several levels including the superfamily level. Proteins grouped together at the superfamily level are defined as having structural, functional and sequence evidence for a common evolutionary ancestor (Gough et al. 2001). We have used the Pfam database that consists of a

.

collection of large number of curated protein domain families with sequences of homologues in each family aligned accurately.

Table 2 Superfamily description of *S.epidermidis* 16S rRNA gene sequence

16S rRNA* gene sequence of *S.epidermidis

The Bubble Protein, family present in the *16S rRNA* gene sequence of *S.epidermidis*, has an E-value for forward sequence and reverse sequence is same, 0.0078. This score is conditional on the domain being a member of the *16S rRNA* Bubble Protein superfamily.

CONCLUSION

After amplification with *16S rRNA* gene primer set ~780 bp product were detected. After sequencing, 784 bp product was obtained for 16S rRNA forward primer and 782 bp for *16S rRNA* reverse primer. There are millions of nucleotide sequences known. These sequences fall into many groups of related sequences known as gene families. Relationships between these sequences are usually discovered by aligning them together and as signing this alignment a score. Pairwise alignment: BLAST, compares *16S rRNA* gene sequence in order to find similarity with homologous sequences of database. When sequence was aligned with *16S rRNA* gene sequences available in GenBank, it gives similarity of 99%to the *16S rRNA* gene sequence. The *16S rRN*A gene sequence is of Staphylococcus epidermidis.By the study of Conserved regions/ Domains using Expasy-Prosite, Forward sequence has less conserved regions/Domains than the reverse sequence. ClustalW output for multiple sequence alignment shows that Forward sequence is more similar than the reverse sequence when compared with GenBank data sequences. In this study, we have identified unknown bacterium based on *16S rRNA*.Analyze the *16S rRNA* sequence to check the conserved regions/ Domains presence and to study its phylogenetic relation with the GenBank data sequences.

CONFLICT OF INTEREST

All authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

REFERENCES

1)Janda, J., Abbott, S., 2007, 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls, Journal of clinical microbiology, 45(9), 2761-2764.
2)Brindha, V., Mathew, A.,2012, Molecular characterization and identification of unknown bacteria from waste water, Indian Journal of innovations and developments, 1(2), 87-91.
3)Clarridge, J., 2004, Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases, Clinical microbiology reviews, 840-862.
4)Scola, B., Michel, G., Raoult, D., 1997, Use of Ampli?cation and Sequencing of the 16S rRNA Gene to Diagnose Mycoplasma pneumoniae Osteomyelitis in a Patient with Hypogammaglobulinemia, Oxford journals, 1161-1163.
5)Mayoral, C., Norona, M., Baroni, M., Giani, R.,Zalazar, F., 2005, Evaluation of a nested PCR assay for Streptococcus pneumoniae detection in pediatric patients with community acquired pneumonia, Revista Argentina de Microbiología, 37, 184-188.
6)Zarain, P., Lopez-tellez, G., Carmen, R., Romero-lopez, C., Martinez-Laguna, Y., Rivera, A., Brambila, E.,2012, Nested-PCR as a tool for the detection and differentiation of gram positive and gram-negative bacteria in patients with sepsis-septic shock, African J of Microbiology Research, 6(21), 4601-4607.
7)Dreier, J., Stormer, M., Kleesiek, K., 2004,Two novel Real-Time Reverse Transcriptase PCR assays for rapid detection of bacterial contamination in platelet concentrates, Journal of Clinical Microbiology, 42, 4759-4764.
8)Leonard, J., Grace, M., Buzard, G., Mullen, M., Barbagallo, C., 1998, Preparation of PCR Products for DNA Sequencing, BioTechniques, 24(2), 314-317.
9)Matz, M., Alieva, N., Chenchik, A., Lukyanov,S., 2003, Amplification of cDNA ends using PCR suppression effect and step-out PCR, Whitney Laboratory, Methods in Molecular Biology, 221, 41-49.
10)Mayoral, C., Schernthaner, J., 2005, Optimization of in Vitro Transcription and Full- Length cDNA Synthesis Using the T4

Bacteriophage Gene 32 Protein, Journal of Biomolecular Techniques, 16(3), 239-247.
11)Puig, M., Jofre, J., Lucena, F., Allard, A., Wadell, G., Girones, R., 1994, Detection of Adenoviruses and Enteroviruses in Polluted Waters by Nested PCR Amplification, Applied and Environmental Microbiology, 60(8), 2963-2970.
12)Yılmaz, M., Ozic, C., Gok, I., 2012, Principles of Nucleic Acid Separation by Agarose Gel Electrophoresis, INTECH, 33-40.
13)Reddy, P., Raju, N., 2012, Gel-Electrophoresis and Its Applications, INTECH, 15-32.
14)Melo, J., Kent, N., Yan, X., Goldman, J., 2013, Controls for reverse transcriptase polymerase chain reaction amplification of BCR-ABL transcripts [letter], Blood: American Society of Hematology, 3984-3986.
15)Monod, J., 2007, The Growth of Bacterial Cultures, Annual Reviews microbiology, 1949(3), 371-394.
16)Fangan, B., Dahlberg, O., Deggerdal, A., Bosnes, M., Larsen, F., 1999, Automated System for Purification of Dye-Terminator Sequencing Products Eliminates Up-Stream Purification of Templates, BioTechniques, 26 (5), 980-983.
17)Hen, Y., Huang, X., 2009, DNA Sequencing By Denaturation: Principle and Thermodynamic Simulations, NIH Public Access Author Manuscript, 384(1), 170-179.
18)Goujon, M., William, H., Weizhong, L., VALENTIN, F.,SQUIZZATO, S., PAERN, J., LOPEZ, R.,2010, A new bioinformatics analysis tools framework at EMBL–EBI, Nucleic acid research, 38, 695-699.
19)Lopez, J., Bohuski, E., 2007, Total RNA Extraction with TRIZOL Reagent and Purification with QIAGEN RNeasy Mini Kit.
20)Altschul, S., Gish, W., Miller, W., Myers, E., Lipman, D., 1990, Basic Local Alignment Search Tool, Journal of Molecular Biology, 403-410.
21)Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettiga, P., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., Lopez, R., Thompson, J., Gibson, T., Higgins, D., 2007, Clustal W and Clustal X version 2.0, Bioinformatics, 23(21), 2947-2948.
22)Kuo-Bin, Li., 2003, ClustalW-MPI: ClustalW analysis using distributed and parallel computing, Bioinformatics, 19(12), 1585-1586.
23)Thompson, J., Higgins, D., Gibson, T., 1994, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acid Research, 22(22), 4673-4680.
24)Sigrist, J., Castro, E., Langendijk-Genevaux, P., Saux, V., Bairoch, A., Hulo, N., 2005, ProRule: a new database containing functional and structural information on PROSITE profiles, Bioinformatics, 21(21), 4060-4066.
25)Sigrist, J., Castro, E., Cerutti, L., Cuche, B., Hulo, N., Bridge, A., Bougueleret, L., Xenarios, I., 2012, New and continuing developments at PROSITE, Nucleic Acid Research, 41, 344-347.
26)Panwar, B., Raghava, G., 2010, Prediction and classification of aminoacyl tRNA synthetases using PROSITE domains, BMC Genomics, 11(507).
27)Schuepbach, T., Pagni, M., Bridge, A., Bougueleret, L., Xenarios, I., Cerutti, L., 2013, pfsearchV3: a code acceleration and heuristic to search PROSITE profiles, Bioinformatics, 29(9), 1215-1217.
28)SOUTHAN, C., 2000, InterPro (The integrated resource of Protein domains and functional sites), Yeast, 17, 327-324.
29)Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Castro, E., Langendijk-Genevaux, P., Pagni, M., Sigrist, J., 2006, The PROSITE database, Nucleic Acid Research, 34, 227-230.
30)Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Castro, E., Langendijk-Genevaux, P., Pagni, M., Sigrist, J., 2002, PROSITE: A documented database using patterns and profiles as motif descriptor, Briefings in Bioinformatics, 3(3), 265-274.

Dixit Ojas