ClpB can be used as an Environmental Marker to Identify Novel Bacteria from the Environment

Soumitra Rajagopal

Microbial Diversity Course- 2002

University of Nebraska- Lincoln

Corresponding address,

Department of Biological Sciences

University of Nebraska- Lincoln

E151, Beadle Center, 1900 Vine st.

Lincoln, NE-68588-0666

Abstract

The present study focuses on using ClpB as an environmental marker to identify novel bacteria from the environment. Seven different isolates were obtained from enrichments carried out using soil samples from the Sippiwissett salt marsh and marine environments. Isolates were identified by 16S rDNA sequencing and phylogenetic analysis. An attempt to amplify *clpB* gene sequences from these isolates was made unsuccessfully. The intended phylogenetic comparison of the isolates based on their 16S rDNA sequences and *clpB* was not made as intended.

Introduction

The soil microflora has an enormous amount of diversity. More than 10^9 microorganisms from a vast variety of bacterial species can be present in one gram of soil (Ovreas et al., 2002). Equally, diverse populations can also be found in marine environments, but their cell densities may differ based on whether they are free living or in symbiotic association with a marine animal. Only about 65 genomes have been completely sequenced and about 1% of the total microbial flora can be obtained in pure culture (Ovreas et al., 2002). Thus the microbial diversity of microorganisms in nature and particularly in the soil has been relatively unexplored. This is partially due to the complexity and variability of different ecosystems in which microorganisms thrive. The complexities arise due to genetic differences within different taxa and within species and the metabolic interactions between different species. Variability also exists in terms of cell number and relative abundances of different organisms within a particular ecological niche.

Various methods are employed to study genetic diversity in a population. Enrichment and isolation of microorganisms is one of the most important methods in this regard. A vast majority of the microbial population however, is non-culturable which makes isolation procedures difficult. The recent development of methods for studying the diversity of unculturable microrganisms using non-culture techniques has made it easier. These methods call for a combination of microscopy and nucleic acid analysis. One of such important methods involves using 16S rDNA sequencing analysis. It is used widely to help identify unknown bacteria from the environment. A second approach involving the use of highly conserved single copy genes such as EF-Tu, DNA gyrase, ribosomal proteins (and possibly ClpB) etc. are now being used as a basis for culture independent molecular diversity studies.

In this study an attempt is made at using *clpB* gene as an environmental marker to study phylogenetic relationships and identify novel organisms from the environment. The ClpB protease belongs to the HSP100/Clp family of proteins. The Clp family of proteases help in removing denatured/misfolded proteins from the cytoplasm that occur as a result of environmental stress such as heat, osmotic shock, detergent shock etc. The Clp system comprises of the ClpP proteolytic subunit flanked by the ClpA or ClpX ATP binding chaperones (Gottesman ., 1996). These chaperones direct the substrate to the proteolytic subunit for degradation. ClpB is also an ATPase and homologous to ClpA, but unlike ClpA or ClpX does not directly associated with ClpP in an ATP dependent fashion (Lindquist et al., 1996). One function is to prevent protein aggregation and subsequent denaturation during stress (Lindquist et al., 1996). ClpB is a heat shock protein (Squires et al., 1991). Mutants of *clpB* have a slower growth rate at 44^0C and decreased cell

survival at 50^0C and are sensitive to 0.5% SDS (Rajagopal et al., 2002). Proteins homologous to ClpB have been found in all three kingdoms. These include eukaryotic organisms such as *Drosophila melanogaster*, *Caenorhabdits elegans* and the archaebacterium *Methanosarcina acetovorans* (Squires et al., 1991) and prokaryotes like *Escherichia coli*, *Bacillus anthracis* (Fouet et al., 1999), *Mycobacterium leprae*, *Streptococcus mutans* (Squires et al., 1991) etc. The Clp proteins have different sizes but all have two very highly conserved regions comprising 234 and 192 amino acids in common which encode for the nucleoside triphosphate binding domains and are separated by a highly variable spacer sequence (Squires et al., 1991). The spacer sequence helps classify the Clp ATPases from one another.

Degenerate DNA oligonucleotide probes encoding for the two ATP binding domains and *E.coli* specific probes were designed to amplify *clpB* like sequences from the natural isolates by PCR. An attempt was also made at directly amplifying environmental DNA from soil and 'red berry' like structures found in the Sippiwissett salt marsh using these primers.

Materials and Methods

Enrichment and Isolation- Marine soil was used to isolate p- hydroxyl benzoate bacteria (Seq1). Propionic acid bacteria (Seq 5) were isolated from swiss cheese. Lactic acid bacteria (Seq 6) were isolated from sauerkraut preparations. Luminescent bacteria (Seq 2) were isolated from marine water sample. Cyanobacteria (Seq 3) and purple non sulfur bacteria (Seq 4&8) were isolated from soil samples from the school street marsh. These

were enriched for and isolated using procedures described in the 2002 MBL Microbial Diversity handbook.

16S rDNA amplification- ribosomal DNA sequences were amplified using eubacterial primers SDBact 0008F20 (~30-40 pmol) [5'agagtttgatcctggctcag 3'] and SDBact 1492R19 (~30-40pmol)[5'ggttaaccttgttacgactt 3']. PCR reactions were set up as described in the handbook. PCR thermal cycle profile was, 25 cycles of denaturation at 95^0C, annealing at 55^0C for 30 seconds and a final extension at 72^0C for 5 minutes. 5 ul of the amplified product in each case was run on a 1% agarose gel. 0.5X TBE was used as a running buffer. Eubacterial control DNA was provided as a template for the positive control. No DNA was added in case for the negative control.

Amplification of ClpB sequences from the isolates- template DNA was prepared from the isolates Seq1-3 & Seq5-8 using the MolBio Ultra Clean Soil DNA isolation kit. A scoop of cells were scraped off from a plate using a toothpick and used as cell material for DNA extraction. Two PCR primers were used to amplify a 1.4 kb and a 2.5 kb product respectively. Degenerate primers atp-1, 5'- ti ggi gaa ttc ggi gti ggi aa(ag) aci gci at(atc) g-3' and atp-2, 5'- gi (ag)tg gaa ttc at(ag) (at)a(ct) tci gac at(ag) tc--3' used to amplify *clpB* from *Synechoccocus* PCC 7942 (Clarke et al., 1996). However primer atp-2 was modified and the sequence was as follows, 5'-gi (ag)tg gaa ttc atg aa(ct) tci gac at(ag) tc-3'. PCR thermal cycle profile was as follows, 30 cycles of denaturation at 95^0C for 1 minute, annealing for 1 min at 50^0C, and extension for 1 minute at 72^0C.Additionally,temperature gradient of +/- 7^0C, a range of template DNA

concentrations from 25ng-200ng, and a primer concentration range of 50 and 100 pmol were also tried for amplification. Primers obg0 and obg14 specific towards the *clpB* gene from *E. coli* were used for analysis: obg0- 5'- cgc tca ctt ttc agg ca- 3' and obg14-5'- ccg act gcc tcc ctg tgc tg- 3' were used to amplify a 2.5 kb base pairs. PCR thermal cycle profile was as follows, 35 cycles of denaturation at 95^0C for 1 minute, annealing at 50^0C and extension for 1 minute at 72^0C.Additionally temperature gradient of 50^0C +/- 10^0C and template DNA concentrations of 25ng and 50ng were also used. 5 ul of PCR product was run on 1% agarose gel. 0.5X TBE was used as running buffer. PCR with template DNA from *E.coli* and eubacterial primers was carried out in parallel under the same conditions and used as positive control. A no DNA control was also carried out.

Phylogenetic Analysis- PCR products were purified using a Qiagen PCR purification kit. 300-500ng of the PCR products were sent for sequencing to University of Connecticut-Mass. Sequences were then analyzed using the ARB software analysis computer package.

PCR analysis of microbial populations from DNA extracted from Sippewessett soil- DNA extraction from soil samples and red berries was carried out using the Mol Bio Ultra Clean soil DNA kit. 25ng of the template DNA was then used for PCR amplification using 16S rDNA primers and ClpB primers obg0/obg14 and atp-1/atp-2 primer sets was also done. The PCR thermal cycle protocol as described earlier for the 16S rDNA analysis was used in these set of experiments. PCR products were run on 1% agarose gel. 0.5X TBE was used as running buffer.

Results

Enrichment and Isolation- all the enrichments were isolated in pure culture as determined by microscopy (Fig.1, 40X magnification)

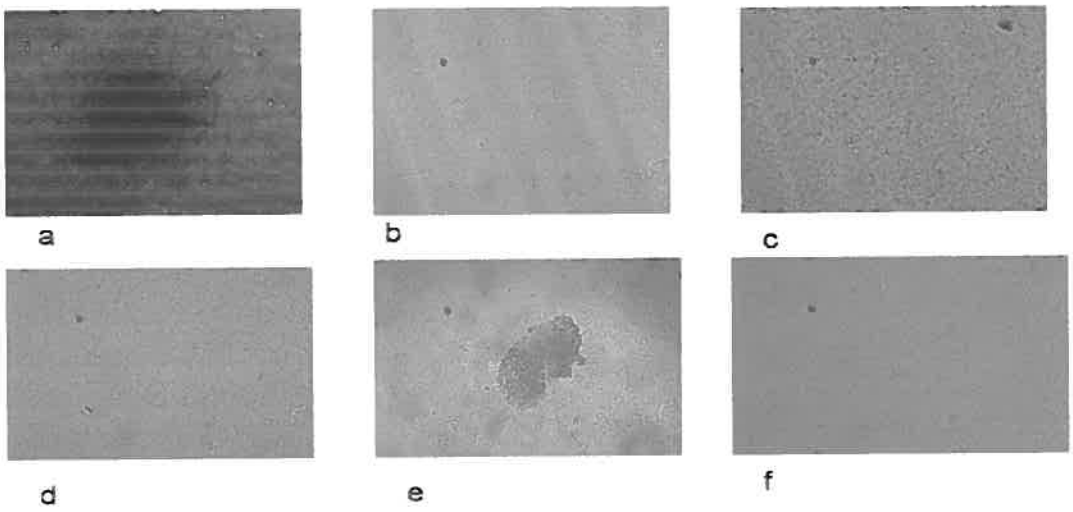

Fig-1 Isolates in pure culture, a- cyanobacteria (Seq 5), b-Lactic acid bacteria (Seq-3), c- Luminescent Bacteria (Seq-2), d- propionic acid bacteria (Seq-6), e- purple-non sulfur bacteria (Seq-4 & Seq-8) and f- p-hydroxy benzoate degrading bacteria (Seq-1).

16S rDNA amplifications- amplification of the 16S rDNA sequences from all the seven isolates yielded a single fragment of ~1500 base pairs (Fig 2). The PCR products were analyzed further by sequencing.

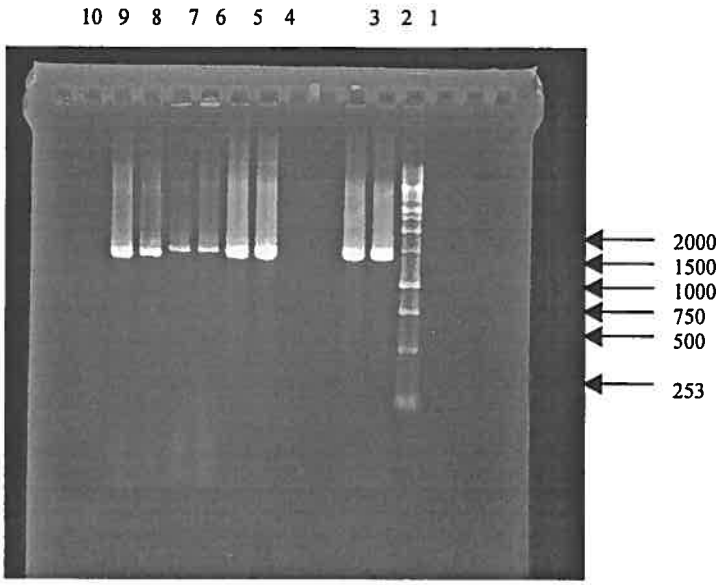

Fig 2. 16S rDNA amplification of 'Seq' isolates. Lane-1 1 Kb DNA ladder, Lanes- 2,3 Seq 1& Seq 2, Lane- 4 Seq 3, Lanes 5-8, Seq 5-8, Lane 9 positive control, Lane 10- negative control.

Phylogenetic analysis- *Thermotoga maritima* was used as outgroup to identify the various isolates by phylogeny. Using ARB, a phylogenetic tree was constructed by neighbor joining program (Fig.3) for the various sequences. Seq-1 was found to be closely related to *Pseudomonas putida*, Seq-2 was related to the *Vibrio* spp. , Seq-4 & 8 were closely related to *Rubrivivax gelatinosus*, Seq-5 was closely related to the chloroplast sequences of *Euglena gracilis*, Seq-3 was related to *Lactococcus* fructosus and Seq-6 was related to *Propionibacterium freudenrechii*.

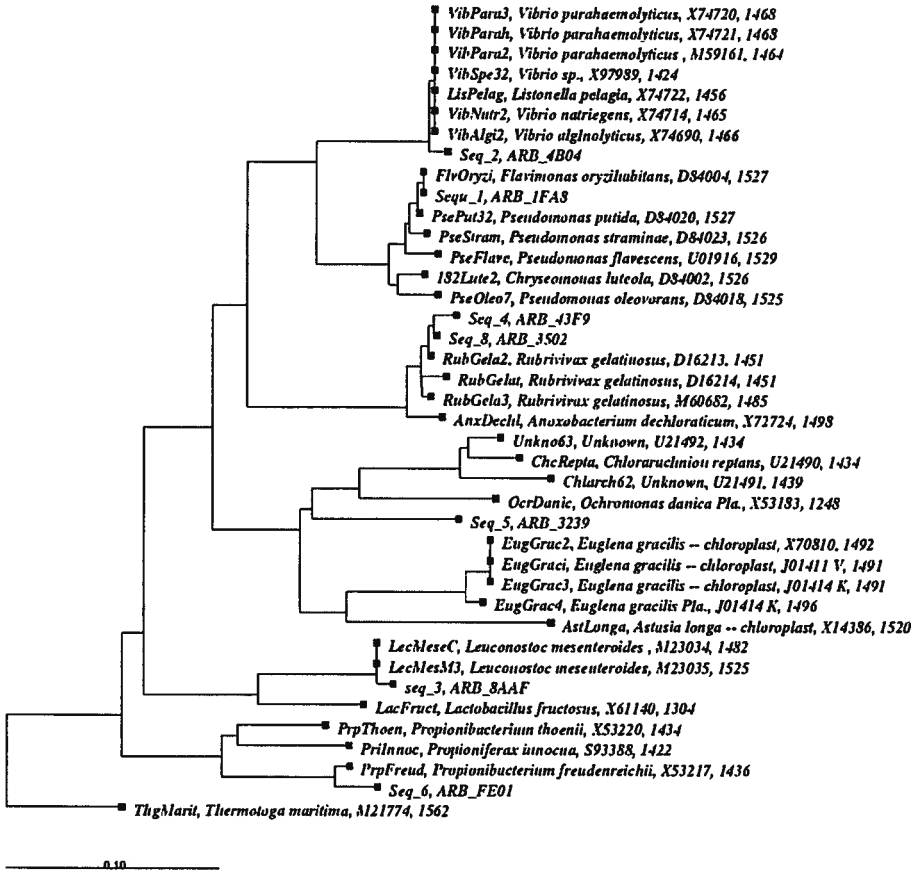

Fig-3 Phylogenetic analysis of 16S rDNA sequences amplified by PCR from various isolates obtained in this study.

clpB gene PCR amplification- primer sets atp-1/atp-2 and obg0/obg14 were not able to amplify clpB gene sequences from the various 'Seq' isolates(Fig 4). These primer sets were also not able to amplify *clpB* gene from *E.coli* which was used as an intended positive control. The 16S rDNA was amplified in each case using the eubacterial primers,

were also not able to amplify *clpB* gene from *E.coli* which was used as an intended positive control. The 16S rDNA was amplified in each case using the eubacterial primers, which was later used as the positive control. Phylogenetic analysis could not be carried out due to lack of sequence information for *clpB*.

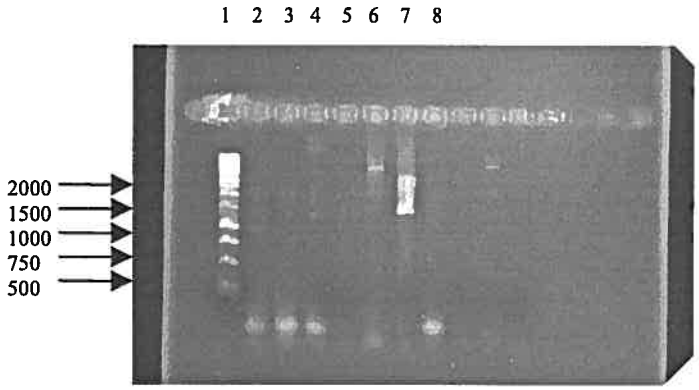

Fig-4 Lane-1 1 kb DNA ladder, Lane-2 E.coli; Lane-3, Seq-1; Lane-4, Seq-2; Lane-5 Seq-3; Lane-6, Seq-4,; Lane-7, Positive control; Lane-8, Negative control.

Direct PCR amplification of environmental DNA – DNA extracted from soil and red berry sample was used as template and the 16S rDNA eubacterial primers, atp-1/atp-2 and obg0/obg14 primer sets. A 1500 base pair PCR product was obtained for both soil and red berry template DNA using the eubacterial primers. In case for the primer atp-1/atp-2, two PCR products of 1 and 1.2 kb were obtained for the red berry template DNA only. No amplification was obtained for the soil DNA. In case of the *E.coli* specific probes, again a very strong signal at 1.5 kb was obtained for the template DNA from red berry and no amplification was observed from the soil template DNA (Fig 5).

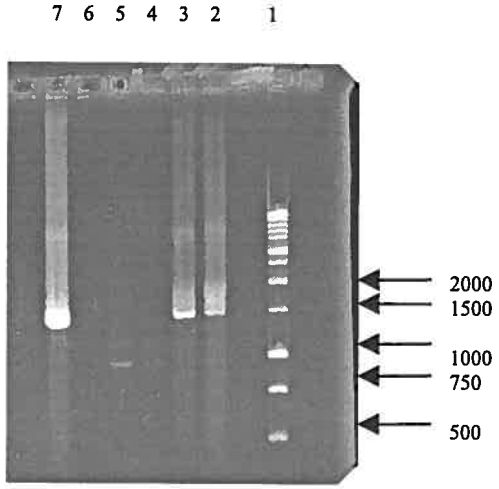

Fig-5 Lane-1, 1 Kb DNA ladder, Lanes-2,3 PCR products using eubacterial primers for soil and red berry template DNA, Lanes 4,5 PCR products using atp-1/atp-2 primers for soil and red berry template DNA, Lanes 6,7 PCR product using obg0/obg14 primers for soil and red berry template DNA.

ARDRA analysis of PCR products from soil and red berry-__ restriction patterns obtained were different for both the PCR products obtained from berry DNA for the 16S and the *clpB* primers

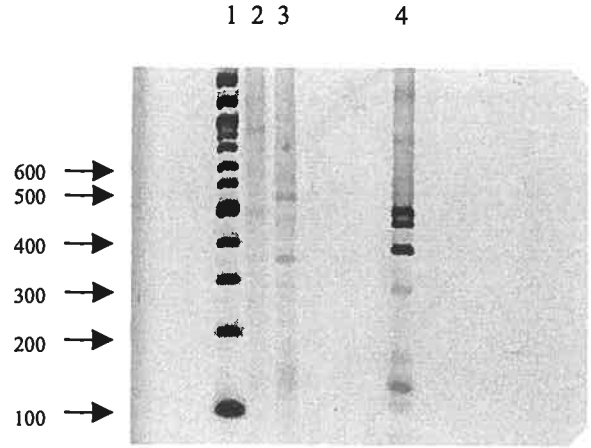

Fig-6 ARDRA analysis of soil and berry PCR product using 16S and obg0/obg14 primers. Lane-1, 100 bp DNA ladder; Lane-2 & 3, Soil DNA & red berry DNA

respectively (amplified using 16S rDNA primers) and Lane-4, red berry DNA using clpB specific primers obg0/obg14 primers.

Discussion

Gene sequences that code for translation products are increasingly being used as a marker for phylogenetic analysis because they have several advantages: a/ they help to fine tune a phylogenetic tree especially at branch endpoints. 16S rDNA sequences generate variability at these positions. Also, the 16S phylogenetic trees are broad, whereas trees based on a functional gene are more specific b/ they can be used as additional proof for the existing trees made by 16S sequence analysis and c/ can be used to determine if lateral gene transfer has occurred for a particular function encoded by the gene of interest.. This determines its evolutionary course. It is also possible that the candidate gene has not 'moved around' considerably and thus in this case is like 16S. It can be used to check the correctness of a tree made using 16S sequences. The criteria for using *clpB* as a phylogenetic marker were : a/ it's a single copy gene, b/ its precise function is not very well understood, c/ but it is very widely distributed and highly conserved in nature. Is it possible that the sequences of this gene among various organisms are not remarkably different because the gene codes for yet another unknown function that is absolutely essential for the organism and hence it cannot afford to lose it by lateral gene transfer or any other mechanism?

All the16S products obtained from purified isolates were correctly identified by sequencing and phylogenetic analysis for ex., Seq 1 which was isolated from a para hydroxy benzoate enrichment was identified as a close relative of *Pseudomonas putida*

which is known for its ability to degrade a variety of xenobiotic compounds. Similarly, Seq-2 was identified as related to the marine *Vibrio* spp. These isolates were purified from enrichments for bioluminescent bacteria etc.

The amplification of *clpB* sequences from *E. coli* and the purified isolates were not successful. There may be number of reasons for this failure. The degenerate primers atp-1/atp-2 were designed to target the nucleoside triphosphate binding domains which are so highly conserved in all the large family of Clp proteins (Clarke et al., 1996). It is possible that the two bases that were picked at random for primer atp-2 (they were picked to reduce the cost of primer synthesis) were absolutely essential for amplification although the annealing temperature was lowered to 45^0C to allow for degeneracy to occur. However it may also be true that the primers are not suitable for amplifying *clpB*. An assortment of tools were tried to optimize the PCR experiment including using a temperature gradient, varying the template and primer concentrations, varying concentrations of Mg^{+2} ions in the PCR reaction. All proved to be unsuccessful. Next primers specific for *clpB* from *E.coli* also were not able to amplify the gene from *E.coli*. The same types of troubleshooting options were also tried for these reactions. The 16S rDNA gene product could be amplified using *E.coli* template DNA. This meant that the PCR experiment itself was not faulty. Therefore further optimization and better primer design for amplifying *clpB* is needed. Alternatively, it is possible that *clpB* is not universally conserved and hence there is no amplification, but this cannot be proven until the *clpB* gene from *E.coli* can be amplified as a positive control.

A 'zoo blot' made using probes for *clpB* sequence showed amplification of DNA from plants, archaea, proteobacteria, lower eukaryotes, protozoa etc.(Cathy

Squires, personal communication). It is reassuring ClpB is found in such a wide variety of organisms and it is very tempting to suggest that the enriched isolates have *clpB* and a high degree of similarity in their gene sequences, especially in the nucleoside triphosphate domains. If this is true then, *clpB* is a good candidate for a gene that has not undergone a great deal of lateral gene transfer. It has been suggested that the genes that code for primary metabolic functions undergo a great deal of lateral gene transfers (Wagner et al., 2001). Mutations in *clpB* in *E.coli* are not lethal and are temperature sensitive (Squires et al., 1991). If they are dispensable then why are these proteins very well conserved in such a wide variety of organisms? Functional genes are used to construct phylogenetic trees and help identify a specific class of bacteria, like *dsrAB* operon in the dissimilatory sulfate reducing bacteria (Wagner et al., 2001). Constructing a phylogenetic tree using *clpB* sequences may help understand the other functions of the ClpB product. One of those functions maybe vital to the survival of the cell. This in turn may explain its conservation in such a wide variety of organisms.

Finally, attempt to directly amplify *clpB* like sequences from DNA, extracted from the soil and red berry was marginally successful. A 1 kb and a 1.2 kb fragment were amplified from red berry DNA using the atp-1/atp-2 primers. The PCR product needs to be characterized further by sequence analysis. Also a PCR product of 1.6 kb was amplified from red berry DNA using primers obg0/obg14. This product was further analyzed by ARDRA and 4 restriction fragments of sizes 150 bp, 400 bp and 500 bp were obtained. Their profiles were not similar to the ARDRA profiles of the 16S rDNA amplified sequences. The ARDRA pattern suggests that a single organism may have been amplified using the *E.coli* specific probes from red berry DNA. Could clpB

now be used to amplify specific class of organisms? This would also shed more light on its function in nature.

In summary it is possible that ClpB exists in the various isolates used in this study and further have the same common consensus sequences and putative function/s. An ideal next step would be to collect these sequences and compare it to the sequences already present in the database, compare their phylogenetic profiles with the 16S rDNA phylogeny for the same isolates and look for lateral gene transfer phenomenon or more importantly no occurrence at all.

References

Clarke, A. K., and Erikkson, J.M. 1996. The heat shock protein ClpB mediates the development of thermotolerance in the cyanobacterium *Synechococcus* sp. Strain PCC 7942. J. Bacteriol. **178:** 4839-4846

Fouet, A., Mock, M., and Namy, O. 1999. Co- existence of *clpB* and *clpC* in the Bacillaceae. FEMS. **173:** 297-302

Gottesman, S. 1996. Proteases and their targets in *Escherichia coli.* Annu. Rev. Genet. **30**:465-506

Lindquist, S., Singer, M.A., Glover, J.R., and Schirmer, E.C. 1996. HSP100/Clp proteins: a common mechanism explains diverse functions. TIBS. **21**:289-296

Microbial Diversity Experiment HandBook. 2002

Ovreas, L., and Torsvik, V. 2002. Microbial diversity and function in soil: from genes to ecosystems. Curr. Opin. Microbiol. **5:** 240-245

Rajagopal, S., Narsimhan, S., and Nickerson, K.W. 2002. SDS hypersensitivity of *clpP* and *clpB* mutants of Escherichoa coli. Appl. Environ. Microbiol. **68**:4117-4121

Squires, C., Ross, B.M., Pedersen, S., and Squires, C. 1991. ClpB is the Escherichia coli Heat Shock Protein F84.1. J. Bacteriol. **173**: 4254-4262

Wagner, M., Stahl, A.D., Blackall, L.A., Abicht, H., Fishbain, S., Hugenholtz, P., Roger, A.J., Friedrich M., and Klein, M. 2001. Multiple lateral transfers of dissimilatory sulfite reductase genes between major lineages of sulfate reducing prokaryotes. J. Bacteriol. **183:**6028-6035