ORIGINAL ARTICLE

SAMPLING BACTERIAL BIODIVERSITY FROM A HIGHLY CONTAMINATED STREAM FLOWING THROUGH A DENSELY POPULATED URBAN AREA IN KARACHI

Syed Faaiz Enam, Humaira Qureshi*, Sohail A. Qureshi*

Department of Biological & Biomedical Sciences, Aga Khan University Hospital, Karachi,
***Department of Biology, LUMS School of Science and Engineering, SSE Complex, DHA Lahore, Pakistan**

Background: Few studies have attempted to understand the complexity of microbial populations in Pakistan where infectious diseases are prevalent. This study was undertaken to assess bacterial biodiversity in *Nehr-e-Khayyam* a heavily polluted stream connected to the Arabian Gulf, which runs through a densely populated urban area in Karachi, Pakistan. **Methods:** Employing a universal pair of oligonucleotides capable of amplifying species-specific segments of 16S rRNA gene from all Eubacteria, we generated a library of PCR products using total DNA purified from the collected sample, cloned the amplifers into pGEM-T-Easy and sequenced each recombinant clone. The obtained DNA sequences were subjected to bio-informatic analyses. **Results**: A total of 71 recombinant clones were obtained from the amplified 16S rDNA products and sequenced. Bioinformatics analyses revealed that 54 (out of 71) were unique sequences from which 42 shared >97% and 12 shared <97% homology to their database counterparts. One sequence originated from the plastid DNA of eukaryote *Pyramimonas disomata*. From the remaining 53 sequences, 45 were Proteo-bacteria and 8 Fermicute in origin. Among 71 sequences, Alpha-, Beta- and Gamma-proteobacteria species constituted ~86% of Proteo-bacteria identified in the sample while only ~13% were Fermicutes. **Conclusions:** The microbial niche in *Nehr-e-Khayyam* is occupied predominantly by heterotrophic Proteo-bacterial and Firmicute strains, some of which are known human pathogens.

Keywords: Bacteria, biodiversity, 16S rDNA, Karachi

INTRODUCTION

Prokaryotic micro-organisms belonging to domains Bacteria and Archaea constitute a major fraction of Earth's biota and are found in diverse habitats ranging from very mild to extremely harsh. Although classical approaches for identifying and characterising bacteria relied largely on staining and microscopic techniques, introduction of DNA-based methods in the mid-1980s revolutionized the field of microbial ecology, enabling biologists to not only establish robust relationships between members within and across the bacterial and archaeal domains but to also identify low abundance, hitherto unknown species of bacteria from different environments without the need to cultivate them. Besides PCR, new methods for studying microbial biodiversity in a high throughout manner have also been introduced.[1–5] To date, a number of genetic markers have been employed for studying bacterial phylogeny[6–9] but 16S rDNA has emerged as the ideal 'biosignature' because it is present in all eubacterial genomes and some of its regions are highly conserved across bacterial phyla whereas others are specie-specific. Moreover, extensive use of 16S rDNA sequences for typing bacteria has generated an impressive database which is useful for determining whether a queried sequence originated from a new species of bacteria or from one already discovered.[10] The 16S rDNA genotyping method makes use of universal primers that anneal to conserved regions within the 16S rRNA gene and can amplify the intervening variable regions using total genomic DNA isolated from environmental samples. The resulting library of DNA amplifiers, the complexity of which mirrors different types of bacteria present in the sample, is then segregated by cloning and the sequence obtained from each recombinant plasmid screened against the 16S rDNA database. This approach has played an instrumental role in identifying thousands of novel uncultured prokaryotic micro organisms and in enhancing our understanding of the degree of microbial complexity that exists in different ecological niches across our planet. The 16S rDNA strategy has also been used for studying bacterial populations in clinical samples.[11,12]

Pakistan is home to several rivers, hundreds of lakes and hot springs, some of the highest mountains in the world, numerous mines and 100 Km of Arabian Sea coastline that span across the southern part of the country. However, all of its major cities are heavily polluted and have hundreds of stationary as well as running aqueous reservoirs that receive unprocessed sewerage and industrial waste, and contain decomposed garbage. In Karachi, the largest southern city, it is estimated that approximately 70% of the water supplied to the city is returned as sewage of which only 60% is processed. The unprocessed sewerage is funnelled into the Arabian Sea through Lyari and Malir rivers and five streams (*nallahs*) among which *Nehr-e-Khayyam* is very prominent. Ranging from 18–45 meters in width, the pungent smelling *Nehr-e-Khayyam* comprises the

initial 1.1 km stretch of an 11.7 Km stream that collects garbage as well as unprocessed sewage from nearby areas and flows into the Arabian Sea.

Although numerous studies have been carried out on samples from different regions of the world, very little microbial biodiversity-related information is available from samples collected from man-made, natural or polluted sites within Pakistan.[13,14] To learn more about the different types of micro organisms inhabiting the most polluted areas across the country, we employed the 16S rDNA genotyping approach for studying bacterial biodiversity in *Nehr-e-Khayam*. Our results reveal that this heavily polluted body of water is unusually rich in Proteo-bacterial and Firmicute bacterial strains some of which are human pathogens.

MATERIAL AND METHODS

Approximately 50 ml samples were collected from *Nehr-e-Khayyam*, filtered and centrifuged at 12,000 rpm for 10 minutes at room temperature. Supernatant was discarded and pellet either processed immediately or stored at -20 °C. Each pellet was suspended in 400 µl of STE buffer (100 mM NaCl, 10 mM Tris-Cl; *p*H 8, and 1 mM EDTA) and cells lysed with 50 µl of 10% SDS and incubated in boiling water for 5 minutes. After centrifugation at 12,000 rpm for 5 minutes, 400 µl of supernatant was collected and mixed with 40 µl of 5 M NaCl. DNA was precipitated with 880 µl of ice-cold 95% ethanol and tubes stored at -20 °C for 30–60 minutes. DNA precipitates were harvested by centrifugation at 12,000 rpm for 10 minutes and washed with 500 µl of 75% ethanol twice and dried. DNA was suspended in 20 µl of TE (10 mM Tris-Cl; *p*H 8 and 1 mM EDTA) and stored at -20 °C.

DNA was amplified using forward (5′-AACTGGAGGAAGGTGGGGAT-3′) and reverse (5′-AGGAGGTGATCCAACCGCA-3′) 16S rDNA universal primers with Taq DNA polymerase. A typical 25 µl reaction contained 2 µl (~50 ηg) of total genomic DNA, 0.5 µg each of forward and reverse primers, 200 µM dNTPs, 1× Mg^{2+} containing buffer and 0.5 units of Taq DNA polymerase (Promega). Thermocycling conditions for the PCR were as follows: 94 °C for 5 minutes followed by 30 cycles of 94 °C, 55 °C and 72 °C each of 30 seconds duration, and a final extension at 72 °C for 5 minutes. For analysis, one tenth of the PCR reaction was electrophoresed alongside DNA markers on 1.2% agarose gel prepared and run in 1× TBE. DNA was visualized by ethidium bromide staining.

Amplifers (~370 bp) were gel purified with Gel Extraction kit (Qiagen). Purified PCR products were ligated into pGEM-T-Easy (Promega) and products used to transform chemically competent XL1-BLUE cells. Transformed cells were plated on LB/agar plates containing ampicillin, colonies grown overnight at 37 °C in 5 ml LB/ampicillin and plasmid DNA isolated using the Plasmid Miniprep kit (Qiagen). Recombinant plasmids were screened by EcoRI digestion and subjected to PCR-based DNA sequencing with BigDye using T7 sequencing primer (5′-TAATACGACTCACTATAGGG-3′).

To screen GenBank database, DNA sequences were submitted via BLAST using default parameters. Sequence alignment was carried out with CLUSTAL-X using the BioEdit (version 5.0.9) software package.

RESULTS

To gauge bacterial complexity in a heavily polluted sewer running through an urbanized region of Karachi, we selected *Nehr-e-Khayyam* as a source of our sample. Prior to DNA purification, pellets from the 10ml samples were washed twice with PBS to remove traces of contaminating DNA. Approximately one-tenth of the purified genomic DNA (~50 ηg) was employed as template for PCR amplification of a region of 16S rRNA gene with a pair of universal primers. The employed primers are eubacteria-specific and incapable of amplifying 16S rDNA from any archaeal species. To ascertain that all amplifers were produced from the purified DNA and not from contaminants, a no template DNA reaction was also included in the experiment as negative control which failed to produce any product (data not shown). Sample-derived PCR products were gel purified and subsequently cloned into pGEM-T-Easy. All 105 ampicillin-resistant colonies that appeared on the plate were screened from which 71 were found to be recombinant. DNA sequence of each of the cloned fragments was then determined with the T7 primer which anneals 58 bp upstream of pGEM-T-Easy cloning site. Sequencing data showed that lengths of cloned inserts ranged in size from 349–372 bp and 17 sequences appeared more than once. Our final bioinformatic analyses were therefore carried out on 54 clones.

Each of the 54 sequences was subjected to basic local alignment search tool (BLAST) against information in GenBank database using default parameters. Sequences sharing >97% homology to the prototype strain in the database were scored as known entities. Based on this homology threshold, the identities of 42 sequences were deciphered confidently. Among 42, only 1 sequence was 100% homologous to uncultured bacterium clone SJTU_A1-7_60 whereas 13, 20 and 8 sequences shared 99%, 98% and 97% sequence homology to pre-existing sequences in the database, respectively. Moreover, 41 of 42 sequences represented known bacteria except for one 371 bp sequence which was 98% homologous to chloroplast 16S rDNA of *Pyramimonas disomata*. Further sequence and phylogenetic analyses of 41 clones showed that they belonged to Proteo-bacteria and Firmicutes lineages. From the 41 positively identified sequences, 34 were

Proteo-bacterial in origin among from which Alpha-, Beta-, and Gamma proteo-bacteria occupied the largest proportion (20%, 29% and 27%, respectively) whereas Delta- and Epsilon proteo-bacteria constituted only 5% and 2.5%, respectively. Some of the identified Proteo-bacterial strains were pathogens. For example *Bordatella bronchiseptica* RB50, *Aeromonas hydrophilia*, *Escherichia coli* O157-H7, *Klebsiella pneumoniae*, *Pseudomonas mendocina* are well known disease causing agents found in our sample along with *Arcobacter* sp clone DS012, a virulent member of Epsilon proteo-bacteria. Additionally, 7 out of 53 sequences were derived from Firmicutes. *Trichococcus flocculiformis*, *Anaerosinus glycerin*i and *Lachnospira pectinoschiza* were among the cultured Fermicutes whereas three have not been cultured. One sequence showed considerable homology with the 16S rDNA of unidentified bacterium DNA (z94800.10). With the exception of Clostridium species which may potentially be virulent, none of the others were pathogenic. Table-1 lists the identities of 42 sequences together with length of sequence spanning forward and reverse universal primers, percent identify to their closest matching hits, and each bacterial strain's ability cause disease (virulence).

Among the 12 sequences which shared <97% homology to their counterparts, 7 were 95–96% homologous and 4 in which the homology to their database counterparts ranged between 91–94%. The lowest level of homology among the 54 sequences was 89% which was between a 370 bp sequence and *Pseudomonas putida* strain QR2, however it carried a statistically significant E-value of 3.00E-135. Within this set, 11 sequences appeared to be Proteo-bacterial and 1 Firmicute in origin. Notably, one sequence was 93% homologous to the *Neisseria meningitiditis* whereas another one similar to *Pseudomonas grimonti* a potentially harmful strain. Listed in Table-2 are all the strains which showed <97% homology against 16S rDNA sequences obtained from the *Nehr-e-Khayyam* sample.

Among the 54 unique sequences, there were several which appeared more than once. For example, sequences originating from *Rhodobacterales bacterium CB1006, Uncultured bacterium clone E11_10.3_1*, *Pseudomonas mendocina ymp*, *Uncultured bacterium clone SSW4Au* and *Uncultured bacterium clone MFC_PBS_Ac_036*[2] strains were found in duplicate. Similarly, from the library of 71 sequences, 3 corresponded to *Uncultured bacterium JM9_B4*, and 4 to *Pseudidiomarina sp QA*. Most abundant was the sequence derived from *Leptothrix cholodnii SP-6* which appeared 8 times.

DISCUSSION

In this study, we have gauged the level of bacterial biodiversity in a sample collected from *Nehr-e-Khayyam*, a heavily polluted stream that flows through a highly populated urban area of Karachi. Using 16S rDNA as a molecular marker, we were able to identify 54 unique sequences (from 71) in our sample with levels of homology ranging from 89–100%. Using the 97% homology cut-off which is typically used for such studies, we found that 41 sequences derived from previously identified strains of bacteria whereas one originated from the plastid DNA of alga *Pyramimonas disomata*. Despite being somewhat diverged from their prototypical strains, the statistically significant E-values associated with the 12 sequences also allowed us to identity, with some confidence, the phylum they belonged to. Strikingly, we found that approximately 86% (61/71) of bacterial biomass in our sample was Proteobacterial while ~13% (9/71) were Firmicutes. Alpha-, Beta- and Gamma proteo-bacterial species altogether constituted about 82% (58/71) of identified Proteo-bacteria. Although *Nehr-e-Khayyam* represents is highly contaminated with sewerage, industrial waste and decomposed garbage it is intriguing that within our sample we failed to identify classes of bacteria other than Proteo-bacteria and Fermicutes.

The sequences identified in this work are likely to have arisen from intact cells since the pellet obtained after centrifugation of collected sample was washed twice to rid of any contaminating nucleic acids. Among the identified bacterial strains there were 29 which have been cultured and 24 that were uncultured. Within the 71 sequence library we found 46 unique, 5 duplicate, and three sequences that were present in 3, 4 and 8 copies. Assuming complete cell lysis of all bacteria in the sample, unbiased amplification of all 16S rDNAs with universal primers along with equivalent cloning efficiency of amplifers, the number of times a sequence is found in a library is likely to mirror the relative abundance of its associated bacterial strain. Based on this, it can be assumed that in the collected sample the number of *Leptothrix cholodnii SP-6* cells is twice as many as *Pseudidiomarina sp QA* cells and eight-times greater than *Aeromonas hydrophilia* cells. While the pungent smell of *Nehr-e-Khayyam* suggests that it contains methanogens and perhaps other archaeons, we failed to obtain any archaea derived sequences suggesting that the employed 16S rDNA universal primers are indeed highly specific for eubacteria.

Given the diversity associated with domain Eubacteria, we are surprised to note that no bacterial strains outside the Proteo-bacteria and Fermicute phyla were found in this study. This result is unlikely to be a consequence of the procedure that was used for cell lysis and instead reflects the real distribution of bacterial species in *Nehr-e-Khayyam*. The presence of Firmicutes in our sample was expected in light of the fact that members belonging to this phylum are capable of inhabiting harsh conditions. Firmicutes strains, however,

represented only a small fraction of bacterial population in the collected sample. Moreover, bacterial strains such as *Alteromonas macleodii, Comamonas testosteroni CNB-2, Dechloromonas sp. JJ, Desulfomicrobium baculatum DSM 4028, Janthinobacterium sp. A1-13 gene, Pseudomonas putida strain QR2 16S, Thauera sp. MZ1T,* and *Uncultured hydrogenophaga* which are known to exist industrial waste sites were also found in our sample.[15]

CONCLUSION

Taken together, our study indicates that *Nehr-e-Khayyam* is occupied with bacteria that belong exclusively to the Proteo-bacteria and Firmicute lineages and that some of the strains inhabiting it are known pathogens. Although other reports have been published on samples collected from man-made and uranium rich-industrial sites within Pakistan[13,14], this is the first study in which the complexity of bacterial population was assessed in a heavily contaminated stream in a culture independent manner. A larger scale study is needed to assess the complexity of microbial populations in samples collected from other contaminated and natural water samples as well as from hot springs and lakes found in Pakistan's mountainous Northern areas.

Table-1: Bacterial strains sharing >97% homology to the 16S rDNA sequences

Species	E-value	% Identity	Length	Virulent
α-Proteobacteria (8)				
*Rhodobacterales bacterium CB1006 [2]	7.00E-163	98	351	No
Roseibacterium elongatum	5.00E-164	97	351	No
Roseobacter denitrificans	7.00E-163	97	351	No
Uncultured bacterium clone 44	5.00E-171	98	351	No
*Uncultured bacterium clone E11_10.3_1 [2]	1.00E-166	97	353	No
Uncultured bacterium clone E12_10.1_2	7.00E-169	98	351	No
Uncultured Rhodobacter sp. clone MBfR_FP53	2.00E-164	97	351	No
Uncultured Rhodobacteraceae bacterium	2.00E-168	98	351	No
β-Proteobacteria (12)				
Acidovorax sp. BSB421	0.00E+00	98	370	No
Aquaspirillum sp. 411	1.00E-179	99	371	No
Bordetella bronchiseptica strain RB50	0.00E+00	98	371	Yes
Comamonas nitrativorans strain 23310	0.00E+00	98	370	No
Comamonas testosteroni CNB-2	0.00E+00	99	371	No
Dechloromonas sp. JJ	0.00E+00	98	371	No
Janthinobacterium sp. A1-13 gene	0.00E+00	98	371	No
*Leptothrix cholodnii SP-6 [8]	0.00E+00	99	371	No
Thauera sp. MZ1T	0.00E+00	99	370	No
Uncultured bacterium clone EtOHMFC-23	5.00E-177	97	371	No
Uncultured beta proteobacterium clone CB31B10	0.00E+00	98	371	No
Uncultured Hydrogenophaga	0.00E+00	98	369	No
γ-Proteobacteria (11)				
Aeromonas hydrophila	0.00E+00	99	371	Yes
Alteromonas macleodii	0.00E+00	99	370	No
Escherichia coli O157:H7	0.00E+00	99	368	Yes
Klebsiella pneumonia	0.00E+00	99	371	Yes
*Pseudomonas mendocina ymp [2]	0.00E+00	99	371	Yes
Uncultured bacterium clone 69-7G	4.00E-178	97	370	No
*Uncultured bacterium clone SSW4Au [2]	2.00E-169	99	372	No
Uncultured bacterium gene (AB105442.1)	3.00E-179	98	370	No
Uncultured bacterium gene (AB494273.1)	2.00E-168	98	371	No
*Uncultured bacterium JM9_B4 [3]	0.00E+00	99	371	No
Uncultured gamma proteobacterium clone CB11H02	4.00E-178	98	369	No
δ- Proteobacteria (2)				
Desulfomicrobium baculatum DSM 4028	0.00E+00	99	371	No
Uncultured bacterium clone JMYB36-58	1.00E-165	98	370	No
ε-Proteobacteria (1)				
Uncultured Arcobacter sp. clone DS012	2.00E-176	98	366	Yes
Firmicutes (7)				
Trichococcus flocculiformis	0.00E+00	98	371	No
Anaerosinus glycerini strain DSM 5192	1.00E-177	99	372	No
Lachnospira pectinoschiza	3.00E-180	98	369	No
*Uncultured bacterium clone MFC_PBS_Ac_036 [2]	8.00E-175	98	371	No
Uncultured bacterium clone SJTU_A1_7_60	1.00E-172	100	370	No
Uncultured Clostridium sp. clone 16IIISN	8.00E-162	97	370	Possibly
Unidentified bacterium DNA (Z94008.1)	6.00E-176	97	370	No
Chlorphyta prasinophyceae (eukaryote)				
Pyramimonas disomata plastid	3.00E-179	98	371	No

Key: Numbers in brackets denote the number of times a sequence and corresponding strain (asterisked) was found in library

Table-2: Bacterial strains sharing <97% homology to the 16s rDNA sequences

SPECIES	E-value	% Identity	Length	Virulent	Classification
Arsenophonus nasoniae	4.00E-165	95	370	No	γ-Proteobacterium
Dinoroseobacter shibae	9.00E-155	95	349	No	α-Proteobacterium
Neisseria meningitidis alpha14	9.00E-155	93	371	Yes	β-Proteobacterium
Neptuniibacter caesariensis	9.00E-168	95	371	No	γ-Proteobacterium
*Pseudidiomarina sp. QA8 [4]	1.00E-171	96	371	No	γ- Proteobacterium
Pseudomonas grimontii	2.00E-170	96	371	Possibly	γ-Proteobacterium
Pseudomonas putida strain QR2 16S	3.00E-135	89	370	No	γ- Proteobacterium
Rhodobacter sp. TCRI 3 gene	4.00E-159	96	350	No	α-Proteobacterium
Uncultured bacterium clone G12_10.4_2	9.00E-168	96	369	No	γ-Proteobacterium
Uncultured bacterium clone IA-11	6.00E-151	92	371	No	Firmicute
Uncultured beta proteobacterium gene (AB478649.1)	4.00E-146	91	371	No	β-Proteobacterium
Uncultured Comamonas sp. clone DS091	3.00E-161	94	370	No	β-Proteobacterium

Key: Numbers in brackets denote the number of times a sequence and corresponding strain (asterisked) was found in library

REFERENCES

1. DeSantis TZ, Brodie EL, Moberg JP, Zubieta IX, Piceno YM, Andersen GL.. High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. Microb Ecol 2007;53:371–83.
2. Humblot C, Guyot JP. Pyrosequencing of tagged 16S rRNA gene amplicons for rapid deciphering of the microbiomes of fermented foods such as pearl millet slurries. Appl Environ Microbiol 2009;75:4354–61.
3. Leclerc M, Delgenes JP, Godon JJ. Diversity of the archaeal community in 44 anaerobic digesters as determined by single strand conformation polymorphism analysis and 16S rDNA sequencing. Environ Microbiol 2004;6:809–19
4. Thies FL, Konig W, Konig B. Rapid characterization of the normal and disturbed vaginal microbiota by application of 16S rRNA gene terminal RFLP fingerprinting. J Med Microbiol 2007;56:755–61.
5. Tzeneva VA, Heilig HG, van Vliet WA, Akkermans AD, de Vos WM, Smidt H. 16S rRNA targeted DGGE fingerprinting of microbial communities. Methods Mol Biol 2008;410:335–49.
6. Huang WM. Bacterial diversity based on type II DNA topoisomerase genes. Annu Rev Genet 1996;30:79–107.
7. Meyer B, Kuever J. Phylogenetic diversity and spatial distribution of the microbial community associated with the Caribbean deep-water sponge Polymastia cf. corticata by 16S rRNA, aprA, and amoA gene analysis. Microb Ecol 2008;56:306–21.
8. Santillana N, Ramirez-Bahena MH, Garcia-Fraile P, Velazquez E, Zuniga D. Phylogenetic diversity based on rrs, atpD, recA genes and 16S-23S intergenic sequence analyses of rhizobial strains isolated from Vicia faba and Pisum sativum in Peru. Arch Microbiol 2008;189:239–47.
9. Yin H, Cao L, Qiu G, Wang D, Kellogg L, Zhou J, *et al.* Molecular diversity of 16S rRNA and gyrB genes in copper mines. Arch Microbiol 2008;189:101–10.
10. Tringe SG, Hugenholtz P. A renaissance for the pioneering 16S rRNA gene. Curr Opin Microbiol 2008;11:442–6.
11. Kim TK, Thomas SM, Ho M, Sharma S, Reich CI, Frank JA, *et al.* Heterogeneity of vaginal microbial communities within individuals. J Clin Microbiol 2009;47:1181–9.
12. Ventura M, Turroni F, Canchaya C, Vaughan EE, O'Toole PW, van Sinderen D.. Microbial diversity in the human intestine and novel insights from metagenomics. Front Biosci 2009;14:3214–21.
13. Ghauri MA, Khalid AM, Grant S, Grant WD, Heaphy S. Phylogenetic analysis of bacterial isolates from man-made high-pH, high-salt environments and identification of gene-cassette-associated open reading frames. Curr Microbiol 2006;52:487–92
14. Ghauri MA, Khalid AM, Grant S, Heaphy S, Grant WD. Phylogenetic analysis of different isolates of Sulfobacillus spp. isolated from uranium-rich environments and recovery of genes using integron-specific primers. Extremophiles 2003;7:341–5
15. Maukonen J, Saarela M. Microbial communities in industrial environment. Curr Opin Microbiol 2009;12:238–43.

Address for Correspondence:

Prof. Sohail A. Qureshi, Department of Biology, LUMS School of Science & Engineering, SSE Complex, Room 9-419A, DHA, Lahore-54792, Pakistan. **Tel:** +92-42-35608349, **Cell:** +92-300-2690346
Email: sohail.qureshi@lums.edu.pk