16S rRNA gene sequence analyses of the metagenome derived from waters of river Mahananda at Siliguri: An approach to understand bacterial diversity

Shriparna Mukherjee1,2, Dhananjay Kumar1, Ashis K Nanda3 and Ranadhir Chakraborty1*

1Omics Laboratory, Department of Biotechnology, University of North Bengal, Siliguri 734 013, India
2Department of Botany, P D Women’s College, Jalpaiguri 735 101, India
3Department of Chemistry, University of North Bengal, Siliguri 734 013, India

Bacterial community of the sub-Himalayan river Mahananda at Siliguri, India was characterized by amplification and sequencing of 16S rRNA genes. Analysis of almost complete 115 sequences revealed a total of 43 operational taxonomic units (OTUs). Statistical analysis revealed that coverage of the clone library was 81.73%. All sequences had >85% similarity to 16S rRNA gene sequences retrieved from the DNA databases. The sequences largely constituted of the class Proteobacteria (10.43% Alpha subdivision, 50.43% Beta subdivision, 9.56% Gamma subdivision & 14.7% Epsilon subdivision), phylum Bacteroidetes (8.69%) and Firmicutes (5.21%), besides single representative belonging to the candidate division OP 11 (0.86%). 31.3% of the total clones showed high similarity to uncultured bacteria, suggesting the usefulness of this technique in uncovering microbial diversity in river water. Abundance of Comamonads (61.76%) among Betaproteobacteria indicated transition to eutrophic state, confirming borderline healthy state of Mahananda River at Siliguri. The most dominant genus of this clone library was Arcobacter, whose presence in environmental waters being associated with fecal pollution has been confirmed.

Keywords: Amplified ribosomal DNA restriction analysis (ARDRA), bacterial community, clone library, Mahananda river, phylogeny, ribosomal RNA gene

Introduction

The study of bacterial diversity of water sources is basically dependent on analysis of diversity in 16S rRNA gene sequences derived from environmental DNA1 or culture-dependent analysis of isolates2. Both approaches have revealed complementary and often non-overlapping features of diversity. There are several studies that have provided valuable data from various freshwater and transitional ecosystems2-6.

The river Mahananda flows through the fastest growing city, Siliguri in the plains of northern West Bengal, India. With the growth in population, sewage load into this river is on rise. This is likely to have influence on the bacterial composition of the river. The present paper describes the culture independent microbial diversity analysis of Mahananda river water at Siliguri. Our data revealed the occurrence of unique bacterial 16S rRNA gene sequences in addition to the sequences that were commonly reported elsewhere from other freshwater environments.

Materials and Methods

Water Sampling, Analyses and Site Climate

The river Mahananda flows through the middle of the Siliguri city. The maximum river gorge below the main Mahananda bridge is about 50 m wide (varies from 10-50 m in different seasons). Water samples were collected from both the banks and mid stream of the river. For sampling, sterilized water bottles were used. Samples were brought to the laboratory in ice box. The average pH of the water samples was 7.3. Other physico-chemical analyses were performed in accordance with the standard methods7. The average DO (dissolved oxygen), BOD (biological oxygen demand), TDS (total dissolved solid) and TPP (total particulate phosphorus) of river water at Siliguri was 4.5, 2.8, 104, and 0.84 mg/L (average values of triplicate analysis), respectively; and chloride ion to an average concentration of 6.5 mg/L. The charge balance between cations and anions was within acceptable limits. HCO3- was the most dominant ion, followed by Ca2+, Mg2+, Cl, NO3-, NO2-, F, SO42-, Na+, and K+. Anions like Cl, NO3- and SO42- were the major inorganic components. Prior to the total DNA isolation, all the water samples were mixed to homogeneity to make a single composite sample for DNA isolation.
Concentration by Filtration and Nucleic Acid Extraction

Water samples (250 mL) were filtered separately by a vacuum pump through cellulose nitrate membrane filter (pore size, 0.2 µm; diam 47 mm). The samples were kept at ambient water temperature during filtration. Pre-filtration was not done for limiting the size fraction of the retained cells. After filtration, membrane filters were stored at –20°C until further analysis. DNA was extracted following the methods of Bostrom et al8, with minor modifications.

Clone Library Construction

PCR was performed in 50 µL reaction mixtures [40 µM deoxynucleoside triphosphates, 12 pmol of each primer, 1 U of Taq DNA polymerase (Promega), 1× PCR buffer (Promega)] for each DNA sample, using universal bacterial primer 27f (5′-AGAGTTTGATCCTGGCTCAG-3′) and 1492r (5′-TACGGTTACCTTGTTACGACTT-3′). PCR amplification began with 3 min denaturation at 94°C; this was followed by 20 to 25 cycles of 94°C for 30 sec, 58°C for 30 sec and 72°C for 1.5 min. The final cycle was extended at 72°C for 7 min. During exponential production of PCR product, the cycles were stopped. Repeated PCR of the resultant product of the initial reaction was done to obtain optimum amount for cloning. PCR products from successive reactions were combined, concentrated and purified using DNA purification kit, Gen Elute™ PCR Clean-Up kit (Sigma Aldrich USA) in accordance with the manufacturer’s instructions. Purified PCR products were subsequently ligated into the pGEM-T easy vector system II (Promega) and used to transform Escherichia coli JM109 competent cells (Promega) as per manufacturer’s instructions. Positive clones were selected and stored on agar plates.

ARDRA Method

One hundred and fifteen clones from the clone library were chosen for amplified ribosomal DNA restriction analysis (ARDRA). The plasmid inserts were PCR amplified by using vector-specific primers SP6 (5′-GATTTAGGTGACACTATAG-3′) and T7 (5′-TAAATCGACTCACTATAGGG-3′). All the PCR assays were carried out in a total volume of 50 µL mixtures containing 40 µM deoxynucleoside triphosphates, 12 pmol of each primer, 1 U of Taq DNA polymerase (Promega), 1× PCR buffer (Promega). One cycle of PCR was run at 94°C for 5 min, followed by 25 cycles each comprised of 94°C for 30 sec, 47°C for 30 sec and 72°C for 2 min. A final extension was performed at 72°C for 7 min. PCR products were digested with HaeIII in accordance with the manufacturer’s instruction, electrophoresed on a 2% (w/v, solution of agarose in TAE buffer) agarose gel, and band sizes were determined by using 100 bp DNA ladder as size standard.

Sequencing

The clones corresponding to the unique banding pattern were used for sequencing. Nucleotide sequencing was performed with the ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction kit and the reactions were analyzed on an ABI PRISM 377 DNA sequencer. The recombinant plasmids were sequenced with T7 and SP6 primers from both ends and the middle portion was sequenced with 530F primer. Only the sequences with overlapping reads were included in further analysis.

Phylogenetic and Statistical Analysis

CHIMERA_CHECK program of the Ribosomal Database Project II (RDP II) was used to detect the presence of possible chimeric sequences9. In order to find closely related sequence(s) and to obtain a preliminary phylogenetic affiliation of the clones, submissions of unaligned sequences were made to the sequence match program of the RDP and to the advanced BLAST search program of NCBI. Sequence alignment and comparison was performed using CLUTALX (v 1.83)10. BIOEDIT sequence editor was used for minor modifications in the alignment. Subsequent phylogenetic analyses were conducted with sequences devoid of ambiguity and gaps in nucleotide positions. Phylogenetic trees were constructed using MEGA (v 3.1) by the neighbor-joining algorithm and the Jukes-Cantor distance estimation method with bootstrap analyses for 1000 replicates11. The clone coverage, i.e., the representation of the phylotypes was calculated from the equation, C = (1-n1/N) [where n1 is the number of single clones, and N is the total number of clones in the clone library]12.

Nucleotide Sequence Accession Number

The nucleotide sequence accession numbers for sequences determined in this study are as follows: FN428747 to FN428789.

Results

Analyses of 16S rRNA Gene Clone Library of Mahananda River Water DNA

The purified PCR products were used for construction of 16S rRNA gene clone library of river
Sequence Analysis of 16S rRNA Gene Clones

Representative clones of all OTUs derived from ARDRA were selected for 16S rRNA gene sequence analysis. A set comprising of 2-3 clones, each of five randomly selected OTUs, were sequenced to confirm the uniformity of the representative clones in the same OTU. Finally, one randomly selected clone from each OUT out of the rest and unique representative(s) of different OTUs were sequenced. Altogether 43 clones, representative of the different groups, were sequenced (nearly full-length sequencing of 16S rRNA gene, ~1500 nucleotides). The per cent similarities of the sequences fell in the range of 85 to 99% with the 16S rRNA gene sequences in the database. Of the total 115 cloned sequences of the library, 36 sequences (31.3%) clustered with sequences derived from uncultured environmental clones. The sources of the isolation of the matched uncultured environmental clones were fresh water (15 sequences), waste water (11 sequences), estuary (9 sequences) and sediment (1 sequence).

Composition of Mahananda Clone Library

The sequenced clones fell into four major lineages of the domain bacteria: Proteobacteria (Alpha, 10.43%; Beta, 50.43%; Gamma, 9.56%; & Epsilon, 14.7%), Bacteroidetes (8.69 %), Firmicutes (5.21%) and candidate division OP11 (0.86%).

In the present study, Betaproteobacteria represented the most dominant subclass of Proteobacteria comprising of 19 OTUs (consisting of 58 clones). Phylogenetic analysis showed that 13 OTUs (with 34 clones) clustered with cultured members of the Betaproteobacteria. The sequences of cultured Betaproteobacteria were restricted to four families, Comamonadaceae, Neisseriaceae, Rhodocyclaceae and Methylphilaceae. The members affiliated to Comamonadaceae (61.76% of 34 clones) predominated. The per cent distribution of the sequences of representative genera under Comamonadaceae was: Polaromonas (28.57%), Hydrogenophaga (23.8%), Acidovorax (23.8%), Macromonas (14.28%) and Comamonas (9.52%). One frequently encountered group, consisting of six clones, was apparently affiliated with the genus P. vacuolata but not supported by high bootstrap value. MRA3019 and MRM2033 were most similar to the sequences of Hydrogenophaga sp. that were isolated from activated sludge. About 32.5% of the sequences were affiliated to the family Neisseriaceae comprising of genera Chromobacterium, Aquaspirillum and Mitsuaria. 6 OTUs constituted of 24 clones clustered with uncultured betaproteobacterial clones, majority of them (62.5%) match best with sequences from other fresh water systems.

The sequences related to Epsilonproteobacteria, the second most abundant fraction, comprised 6 OTUs consisting of 17 clones. Highest phylogenetic similarities of all the representatives were recorded with Arcobacter spp. (85 to 99% similarity) which was validated by 88-100% bootstrap confidence value on the phylogenetic tree (Fig. 1). Three sequences clustered with uncultured Arcobacter sp., of which two were isolated from Danshui river estuary of Northern Taiwan and the other one from an anaerobic waste water treatment reactor. The third most abundant subclass Alphaproteobacteria was represented by only 3 OTUs with 12 clones (Fig. 1). Representative clones of Gammaproteobacteria exhibited similarity with three different genera, Aeromonas (4.34%), Acinetobacter (1.73%) and Pantoea (0.86%).

Except for Proteobacteria, the remaining OTUs belonged to Bacteroidetes (Fig. 2), Firmicutes (Fig. 2) and candidate division OP11 (not shown in the phylogenetic tree). Four OTUs clustered with members of Bacteroidetes. Four OTUs were designated to Firmicutes, of which 2 OTUs with four clones matched maximally with Anaerovibrio glycerini. Only one sequence, MRM2031 (FN428777) shared similarity to the members of OP11 class (not shown in phylogenetic tree).

Discussion

The present study is an attempt to understand bacterial diversity (by clone library method) of a city-waste polluted river. In principle, universal 16S rRNA gene libraries provide a snapshot of the relative proportions of phylogenetic types in a community. Attempts were taken to minimize the amplification of contaminant genes as well as over amplification of rare genes by reducing the number of PCR cycles and
Fig. 1 Contd…
Fig. 1—Neighbor-joining tree showing positions of proteobacterial sequences in river water clone library including reference sequences obtained from GenBank. Bold numbers are clones obtained in this study. *Methanothermococcteus okinawensis* is used as an outgroup. GenBank accession numbers are in parentheses.
also by terminating the amplification abruptly (in a stage when PCR product was still accumulating exponentially).

The dominant group in clone library was related to *Proteobacteria*, which was consistent with other studies. However, the proportions of subclasses of *Proteobacteria* were different in clone libraries from different fresh water habitats. The dominance of betaproteobacterial phyotypes in Mahananda water corresponds to the reports of earlier studies on freshwater environments. The 16S rRNA gene clone library has revealed affiliation of 61.76% of betaproteobacterial members to Comamonadaceae. Overriding representation of phyotypes assigned to Comamonadaceae was indicative of eutrophic nature of river water at Siliguri. The genera represented in the study by Beta-group members of the *Proteobacteria*, i.e., *Hydrogenophaga*, *Comamonas* and *Acidovorax*, were found to cluster with some sequences of the clone library under study (Fig. 1).

All these genera are well known for the utilization of a wide spectrum of carbon sources. One OTU, MRA1008 branched with *Zoogloea ramigera* showing 97% similarity (Fig. 1). The occurrence of sequence affiliated to species capable of reducing nitrate in eutrophic environments and normally associated with a terrestrial habitat or a more nutrient rich environment (*Z. ramigera*) suggests a particle associated origin for these cell types, consistent with the fact that we used an unfiltered sample.

*Arcobacter* was identified as the most dominant genus within the subclass *Epsilonproteobacteria*. Representatives of *Arcobacter* spp. were considered only transient residents in the aquatic environment. Presence of *Arcobacter* spp. in fresh water samples have been found significantly associated with bacterial indicators of fecal pollution. In the present study, alphaproteobacterial members represent the third most abundant subclass. *Novosphingobium* sp. served as the closest described 16S relative of the predominant OTU (6.95%) of this subclass. Other clones exhibited similarity with *Rhodobacter* spp., which is an anoxygenic phototrophic bacterium. *Gammaproteobacteria* are often considered as of little importance in freshwater ecosystems.

*Gammaproteobacteria* were identified as members of the freshwater gamma I cluster, but they have shown high degree of similarity to species observed in other freshwater systems like Swedish lakes and effluent from a bioremediation site. In a previous study on bacterioplanktons of a lake, 16 clusters were recovered, of which only one was affiliated with the Gamma subclass of *Proteobacteria*. Increased abundances of *Gammaproteobacteria* as compared to *Betaproteobacteria* has been noted in river Aliakmon, Greece.

Members of the phylum *Bacteroidetes* are seemingly omnipresent in aquatic environment. They are second only to the *Proteobacteria* in terms of aquatic abundance. Members of phylum *Bacteroidetes* are involved in cycling organic carbon, particularly utilization of high molecular mass dissolved organic matters in nutrient-rich aquatic habitats. Sequences of the cultured bacteria that matched maximally with Mahananda clone sequences were *Cloacibacterium normanense*, originally isolated.
from treated municipal waste water\textsuperscript{28}; \textit{Flavobacterium indicum} from hot spring water \textsuperscript{29} and \textit{Flectobacillus major} from fresh water bodies \textsuperscript{30}. The members of the family \textit{Flavobacteriaceae} are thought to play a role in the breakdown of complex organic matters and degradation of several biopolymers in diverse aquatic habitat where they reside\textsuperscript{31}. Phyotypes detected under \textit{Firmicutes} were mostly affiliated to \textit{A. glycerini}, followed by \textit{Acidaminococcus fermentans} and \textit{Bacillus psychrodurans}. \textit{A. glycerini} was originally isolated from anoxic freshwater sediments\textsuperscript{32}. Only one sequence was found to share similarity with the members from candidate division OP11.

The present study provides valuable baseline data, which are qualitative but not quantitative. Further phylogenetic studies along with the culture-dependent data would enable to understand the bacterial diversity and community structure in an anthropologically stressed river like Mahananda in more details. The presence of relatively high numbers of nonaffiliated bacterial genera identifies this microenvironment as microbial hotspots. The results presented in the present study can be considered a fair-enough starting point where-from in future we shall be demonstrating the responses of the bacterial community to changes of factors affecting river water quality. The community response could be monitored by assessing changes in the phylogenetic diversity, functional gene composition or global mRNA expression (metatranscriptomics). Such data are important with respect to our understanding of the bacterial pool that can serve as a potential source for novel genes and gene products.

Acknowledgement

The present research work was supported by the Department of Biotechnology, Government of India, New Delhi through a major research grant (BT/PR-7164/BCE/448/2006) and fellowship to DK. SM was supported by Research Associateship (Ack. No. 313165/2K7/1) from Council of Scientific and Industrial Research (CSIR), New Delhi.

References


