Journal of Applied Biology & Biotechnology Vol. 12(2), pp. 295-303, Mar-Apr, 2024
Available online at http://www.jabonline.in
DOI: 10.7324/JABB.2024.171556

Microbial diversity of river Ganga at Haridwar (Uttarakhand) through metagenomic approaches

Annapurna Katara, Sumit Chand[#], Harshvardhan Chaudhary, Harish Chandra*, Ramesh Chand Dubey

Department of Botany and Microbiology, Gurukula Kangri (Deemed to be University), Haridwar, Uttrakhand, India.

[#]contributed equally

ARTICLE INFO

Article history:
Received on: October 01, 2023
Accepted on: January 08, 2024
Available online: Febuary 20, 2024

Key words:
Metagenomics,
Taxonomic classification,
Microbial diversity,
Phylogenetic analysis.

ABSTRACT

People revere the Ganga for providing the ecosystem and environment, which are vital for giving life and sustaining it. It's spiritual, mythological, socio-cultural, and historical importance to Indians adds to its significance. This river also supports a very diverse array of life forms, including plants, animals, and microorganisms. The current investigation involves the metagenomic analysis of collected samples taken from three separate sampling points along the length of the river. By employing the CCMetagen and MG-RAST web servers, metagenomic sequence data was analyzed to comprehend microbial diversity. While *Methano regulaboonei*, *Methanosae tathermophila*, and *Methanosarcina acetivoran* are the most prevalent archeans, the most abundant bacteria are *Novosphingobium aromaticivorans*, *Erythrobacter litoralis*, and *Sphingopyxis alaskensis*. *Malassezia globosa*, *Ustilago maydis*, and *Neosartorya fumigata* are the fungi with the highest number of processed reads, whereas the top three viruses are *Rhizobium* phage 16-3, *Pseudomonas* phage 73, and Phage phiJL001. The identified bacterial species are very diverse, as deduced from the phylogenetic analysis. This brought forward the alpha diversity of the water sample under study and enabled us to categorize it according to their relative taxonomic abundances among various levels of the hierarchy, including classes, orders, and families. The taxonomic richness and evenness at different hierarchical levels were also obtained from the data. Such a study will bring forward an in-depth analysis of the microbial diversity of the river Ganges.

1. INTRODUCTION

The Ganga originates from the Gaumukh ice cave in the Gangotri glacier system. Its primary headstream, Bhagirathi, confluences with a superior stream from Badrinath, i.e., Alaknanda in Devprayag, Uttarakhand, forming the Ganga [1,2]. The river flows 2525 km through the plains of Northern India, forming a drainage basin covering 8,61,404 sq. km and comprising 26.2% of the country's total land area [3,4]. The Ganges, a large basin area with abundant water, contributes to the country's food security by providing agriculturally fertile land and a significant portion of crop production [5,6]. The British bacteriologist Ernest Hankin first noted the antibacterial properties of Ganga water in 1896 against *Vibrio cholera*. According to studies, Ganga water does not support the growth of many infectious microorganisms. In another study, the National Environmental Engineering and Research Institute also reported that, as compared to Yamuna and Narmada water, Ganga water contains a substantially greater proportion of bacteriophages. The higher Himalayan stretch of the Ganga had a greatly elevated bacteriophage population. These findings indicate that the Ganga River has a greater bactericidal capacity [7,8]. Though this self-cleansing property of Ganga water is primarily attributed to a large diversity of bacteriophages that are natural predators to many strains of disease-causing bacteria that also cause foul smells and deteriorate the quality of water [9,10].

A recent metagenomic study explored the viral habitats of Ganges, focusing primarily on phage identification and assessment of their diversity. It provided insights into the abundance and diversity of phages, microbial communities, and antibiotic resistance genes in river sediment [11]. Very little is known about the types of microbes that are prevalent in freshwater habitats, except species from the division cyanobacteria, which are rather suitable for cultivation and morphological differentiation. It was thus concluded that there was no apparent distinction between soil bacteria and aquatic bacteria based on culture methods and that bacteria found in groundwater, spring water, and streams are also present in soil [12]. These findings cast doubt on the idea that there is a distinct freshwater bacterial flora. Rivers are diligently explored for novel microbiota because they serve as reservoirs for various microbiomes. Due to their numerous advantages for humans and for every individual in the ecological tier system, these microorganisms are considered significant [13].

*Corresponding Author:
Dr. Harish Chandra, Department of Botany and Microbiology, Gurukula Kangri (Deemed to be University), Haridwar, 249 404, Uttarakhand, India.
E-mail: hreesh5 @ gmail.com

In another metagenomic investigation, a microbial community with bioremediation potential in sediment samples from the Ganga and Yamuna rivers was found, where urea ABC transporter, UrtA, UrtD, UrtE, zinc/cadmium/mercury/lead-transporting ATPase, and other protein domains played crucial roles [14]. With the continual inflow of microorganisms from soil, plant, or animal residues, the river ecology interacts and converges with many other ecosystems. Rivers have extraordinarily high levels of microbial diversity because they are continually importing microbial communities from other ecosystems [15-17]. Another metagenomic study indicated a correlation between the composition of the microbiome and pollution, revealing a high abundance of microbial communities with implications for human health. The Ganga river, particularly upstream, showed a highly dynamic microbial structure, highlighting the functional capabilities of these communities [18]. To examine the genetic make-up, functional potential, and ecological dynamics of microbial communities in various contexts, modern metagenomic research applies cutting-edge sequencing technology, bioinformatics tools, and analytical methods. These studies use the most recent advancements in computational biology and genetics to better comprehend the intricate microbial ecosystems [19,20]. The lives of humans, animals, and plants are influenced by the experimental identification and characterization of microorganisms in their niche. However, the procedure is laborious and expensive. Contrarily, as computational approaches are complementary to experimental procedures, employing the NGS in river microbiomes in a specific habitat may be characterized and annotated more quickly [21]. Several previous studies have been done on biological diversity in the river Ganga, with some emphasis on the diversity of diatoms, phytoplankton, green algae, and blue-green algae [22-24]. Due to anthropogenic activity, many disease-causing bacteria have also been reported in the River Ganga, adding to its microbial diversity [25,26].

The river is also a repository for many bacteriophages, fungi, and non-pathogenic species of bacteria [27]. To the best of our knowledge, few significant studies on the metagenomic profile of the sediments from the Ganga River have been conducted so far. This study will assist those involved with developing proactive plans to address the health concerns of those dependent on the river, directly or indirectly, and thus maintain the health of the aquatic ecosystem. The microbial diversity in the Ganga's flowing water is also thoroughly examined in this study.

2. MATERIALS AND METHODS

2.1. Sample Collection

Water samples were collected from three distinct sites along the river Ganga in Haridwar, Uttarakhand, India; these were then pooled together for further assessments. Before subsequent analysis, the sample was transported at 4°C in 50 ml sterile centrifuge tubes (HiMedia Pvt. Ltd., Mumbai). The sites are primarily chosen based on their relative higher anthropogenic activities.

The sites include Har-ki-Pauri (N 29°57'21.1" E 78°10'15.2"), Chandi Devi Ghat-Eastern Ganga Canal Road (N 29°56'35.2" E 78°1011.6), and Daksheshwer Mahadev Temple, Haridwar (N 29°55'17.5" E 78°08'46.1").

All samples were collected on the same day. Figure 1 depicts all three locations together with satellite imagery and GPS-included pictures showing the actual locations of the sample collection.

2.2. Sample Storage and DNA Extraction

Before undergoing DNA extraction, the sample was kept at −20°C. The Xploregene kit was employed to extract the metagenomic DNA from this sample, and the manufacturer's instructions were strictly followed. Before being used for the next procedures, the samples containing extracted DNA underwent a GEL check and a nanodrop. The nanodrop readings of 260/280 at a ~ value of 1.8–2 are used to determine the quality of DNA. A substantial quantity of extracted DNA of high quality was used to create the metagenomic library.

2.3. Metagenomic Library Preparation

The DNA from the extracted samples was fragmented into lengths of 600 bp using the KAPPA fragmentation process. After the samples were successfully broken down, they were subjected to end repair and A-tailing using a blend of the enzymes Hypa peep plus ERAT. Subsequently, the adaptor was introduced and ligated to the end-repaired DNA fragments using DNA ligase. Illumina primers were used to carry out library amplification on the adapter-ligated samples.40 ng of extracted DNA is used for amplification, along with 10 pM of each primer. Initial denaturation was performed at 98°C. Further steps involved four complete cycles involving denaturation at 98°C for 15 s, followed by annealing at 60°C for 30 s, and extension at 72°C for 30 s. The final extension was executed at 72°C for 1 min and then stored at 4°C before further analysis. Libraries were purified using Ampure beads and quantitated using the Qubit dsDNA High Sensitivity Assay Kit [28].

2.4. Metagenome Sequencing and Quality Assessment

Raw sample reads were obtained from the popular high-throughput sequencing platform Sanger/Illumina. Its version 1.9 was used to obtain raw sample readings from this popular high-throughput sequencing platform. The quality assessment of the raw fast reads thus obtained from the sample was performed using FastQC v. 0.11.9 using default parameters [29]. After quality evaluation, the raw fastq readings are processed using Fastpv.0.20.1 (parameters: --trim_front1 3 --trim_front2 3 --length_required 50 --correction -- trim_poly_g --qualified_quality_phred 30) [30], and then the reads are once again evaluated for quality using FastQCv.0.11.9.

2.5. Bioinformatics Analysis and Sequence Annotation

For metagenomic sequences, MG-RAST software was used with default parameters as it provided automated quality checking, annotation, comparative analysis, and archiving services [31]. First, reads with incorrect lengths were omitted, and low-quality reads were trimmed out using SolexaQA. The program identified the gene sequences using the machine learning approach (FragGeneScan). To evaluate the probable functions and annotation of the genes, MG-RAST created clusters of proteins with a 90% similarity level using the UCLUST operation in QIIME. The longest sequence from each cluster was picked up for analysis of the identity of the sequence. These sequences were then computed against various databases consisting of ribosomal RNA databases and protein databases providing information on the functional hierarchy, for example, RefSeq, IMG, TrEMBL, KEGG, GenBank, Swiss Prot, PATRIC, eggnog, KO, GO, NOG, and COG. The processed reads were also

Figure 1: Sample collection sites along with satellite and GPS images; (a) Har ki Pauri, (b) Chandi Devi Ghat, (c) Daksheshwer Mahadev temple.

classified metagenomically using CCmetagen v.1.2.5 and visualized with Krona (CCMetagen-1.2) [32]. The OTU tables thus obtained were used to deduce a phylogenetic tree using NCBI's common tree file [33] and were plotted using Interactive Tree of Life (iTOL) [34]. Barplots were plotted using Phyloseq R packages [35], and a heatmap was plotted using Ampvis 2 R packages [36].

3. RESULTS AND DISCUSSIONS

3.1. Generation of Sequences

To detect microbial biodiversity, samples from three different sites were pooled and subsequently analyzed using high-throughput next-generation sequencing. The total number of reads for the raw data and the trimmed data, i.e., 8522567 and 82429791, respectively, as well as their read orientation (R1 and R2), average GC% of 58, and length of base pair reads for raw data were found to be 159. A similar finding was also reported by the author, in which the GC% was 53 for samples of sediments collected from the Ganga River at Haridwar, and a striking resemblance was also found between the distribution pattern of GC content and the presence of bacterial diversity grouped at the OTU level [18]. After trimming, a read length of 50–156 was further processed for analysis [Table 1]. For the analysis of taxonomic classification, all the high-quality reads we gained from the sample were taken into consideration.

3.2. Taxonomical classification of metagenome

Based on the taxonomical classification and functional analysis using MG-RAST and CC Metagen, the diversity was categorized on three levels, i.e., class, order, and family. In the manner of their relative abundance, the top 30 classes are shown in Figure 2a and comprises *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, Actinobacteria (class), *Deltaproteobacteria*, Deinococci, Flavobacteria, *Sphingobacteria*, Cytophagia, Bacilli, Planctomycetacia, *Clostridia*, *Bacteroidia*, *Verrucomicrobiae*, Anaerolineae, *Methanomicrobia*, and *Chloroflexi* (class), to name a few. Previous studies on the microbial flora from the sediments also indicate Proteobacteria as the most prevalent bacterial flora, followed by Actinobacteria, *Firmicutes*, and *Deinococcus-Thermus* [27]. Similar studies also revealed that Proteobacteria, *Planctomycetes*, Actinobacteria, Bacteroidetes, *Verrucomicrobia*, *Firmicutes*, and *Cyanobacteria*. were highly abundant in Ganga water [36,37]. In a study conducted in Kanpur (India), *Pseudomonadota*, *Actinomycetota*, Bacillota, Nitrospirota, Gemmatimonadota, *Cyanobacteria*, and Acidobacteriota were among the other numerous phyla that were represented [39].

Table 1: Total numbers of reads for the raw data and the trimmed data.

Sample name	Read orientation	Raw data			Trimmed reads		
		Average GC (%)	Length (bp)	Total sequences	Average GC (%)	Length (bp)	Total sequences
Water	R1	58	159	85222567	58	50-156	82429791
	R2	58	159	85222567	58	50-156	82429791

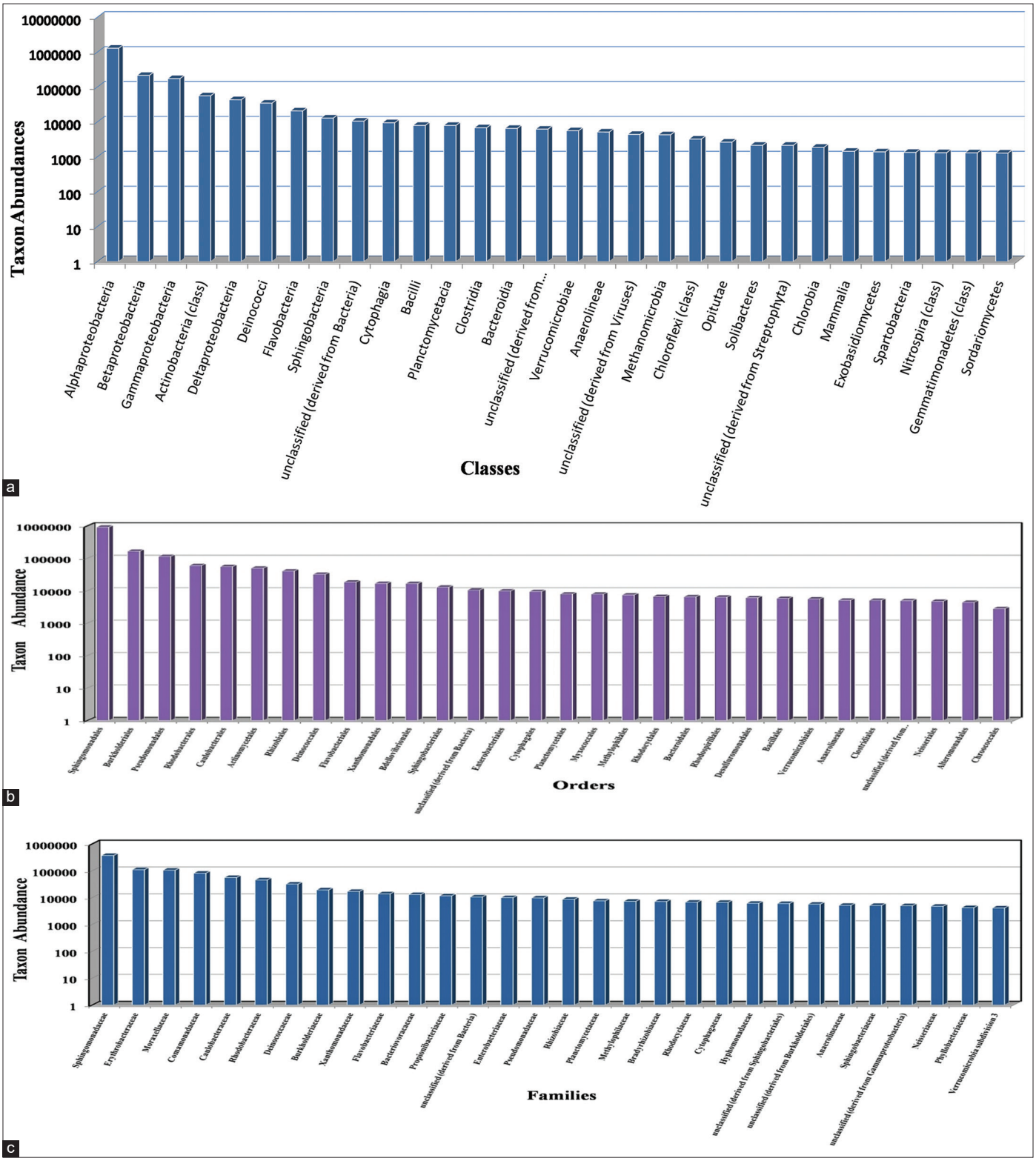

Figure 2: (a) Top 30 classes based on their relative abundance. (b) Top 30 orders based on their relative abundance. (c) Top 30 families based of their relative abundances.

Among orders the most prevalent ones involve Sphingomonadales, *Burkholderiales*,*Pseudomonadales*,*Rhodobacterales*,Caulobacterales, *Actinomycetales*, *Rhizobiales*, Deinococcales, *Flavobacteriales*, *Xanthomonadales*, Bdellovibrionales, *Sphingobacteriales*, Enterobacteriales, *Cytophagales*, Planctomycetales, *Myxococcales*, Methylophilales, Rhodocyclales, Bacteroidales, and *Rhodospirillales* as represented by the bar graph in Figure 2b.

In the case of families [Figure 2c], the prominent ones with the greatest number of reads comprises *Sphingomonadaceae*, *Erythrobacteraceae*, *Moraxellaceae*, *Comamonadaceae*, and Caulobacteraceae. *Rhodobacteraceae*, Deinococcaceae, *Burkholderiaceae*, *Xanthomonadaceae*, *Flavobacteriaceae*, *Bacteriovoracaceae*, *Propionibacteriaceae*, *Enterobacteriaceae*, Pseudomonadaceae, *Rhizobiaceae*, Planctomycetaceae, Methylophilaceae, *Bradyrhizobiaceae*, and *Rhodocyclaceae*. *Comamonadaceae* was the most prevalent family in the three sediment samples from the Ganga stretch near Kanpur, India. Other prominent families were *Rhodocyclaceae*, *Burkholderiaceae*, and *Sphingomonadaceae* [38,39].

3.3. Microbial Diversity at a Glance

Employing a contigLCA algorithm to determine the distribution of taxa and identify a single consensus taxonomic entity for each feature on every distinct sequence has brought forward the diversity at the domain level. The sample contains 118 more sequences in addition to 98.88% processed reads from bacteria (2,366,875), 0.58% from eukaryota (13,851), 0.31% from archaea (7,471), 0.18% from viruses (4,395), and 0.04% from unclassified sequences (1001) [Figure 3].

The total numbers of different species annotations, as a function of the number of sequences sampled, were determined by plotting the rarefaction curve of the annotated species richness. The curve initially exhibited an exponential growth rate when the most common species were found, but as more reads were processed and analyzed, much rarer species in the sample came to light, causing the curve to take the shape of a plateau [Figure 4].

Within viruses, 50% of reads belonged to potyviridae, 41% to lambdaviruses, Human endogenous retrovirus K and some unculturable phages comprise 3% of the total share of viral reads each, while human papilomavirus and alphapolyomavirus, together with some other viruses, share a 2% stack each. Among fungi, basidiomycota made up the biggest portion (75%), followed by ascomycota (21%), and uncultured fungi (approximately 4%). In the sediment samples from Kanpur, the predominant phylum was ascomycota and not basidiomycota, with a diverse range of fungal species. The major organisms found were *Gibberellazeae*, *Schizosaccharomyces pombe, Neosartorya fumigata*, and *Cryptococcus neoformans* [39]. In our analysis of the archaeal reads, we found that *Euryarchaeota* made up 35% of all archaeal reads, uncultured archaeons made up 47%, and other archaeal reads formed 18% of the total archaeal reads. Comparing this with another study in Kanpur, the most prevalent phylum was *Euryarchaeota*, with *Methanosarcinaceae*, *Halobacteriaceae*, and *Thermococcaceae* being the three most abundant families [Figure 5] [39].

The stacked bar graph shown in [Figure 6] represents the total number of hits across all the previously mentioned databases. These consisted of ribosomal RNA databases, protein databases that provide information on the functional hierarchy, and protein databases. The e-value range is used to color the bars, indicating annotated readings.

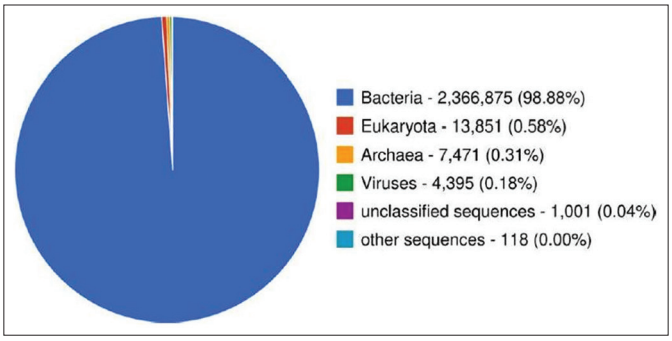

Figure 3: Taxonomic hits distribution at the domain level for sampled water.

3.4. Relative Abundance

A stacked bar plot was plotted using Phyloseq R packages, showing the average relative abundance for class, order, and family separately [Figure 7]. The stacked bar plots exhibit the microbial diversity within each of these categories. Among the classes, *Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria* are more prevalent. In order, Sphingomonadales, *Burkholderiales*, and *Pseudomonadales* share the largest diversity in terms of average relative abundance. In the case of the families *Sphingomonadaceae*, *Erythrobacteraceae*, *Moraxellaceae*, and *Comamonadaceae,* they are the most prominent contributors to average relative abundance.

When we plotted the extracted high-quality reads using Ampvis2 R packages, a heatmap was generated [Figure 8], visualizing the taxonomic percentage of reads at the phyla level. Proteobacteria with 85.75% share the largest portion of percentage reads, followed by Bacteroidetes (4%), and Actinobacteria with 3.12% stack. *Firmicutes* and *Deinococcus-Thermus* also share a major part, with 1.51% and 1.27%, respectively. *Verrucomicrobia* and *Chloroflexi* comprise 0.47% each, while *Cyanobacteria* share 0.6% of the total percentage. Some of the other phyla include *Euryarchaeota* (0.38%), *Chordata* (0.13%), *Ascomycota* (0.07%), *Basidiomycota* (0.05%), *Nitrospirae* (0.07%), Spirochaetes (0.12%), *Arthropoda* (0.03%), Synergistetes (0.04%), *Chlorophyta* (0.02%), *Cnidaria* (0.08%), *Aquificae* (0.06%), Chlamydiae (0.03%), and *Nematoda* (0.01%).

3.5. Analysis Based on Phylogeny

After subjecting the processed or trimmed reads to multiple sequence alignment and running them through RefDB of MG-RAST, we generated a phylogenetic tree [Figure 9] at the species level of the top 50 species detected in the sample; some of these are *Novosphingobium aromaticivorans*, *Erythrobacter litoralis*, *Erythrobacter* spp. NAP1, *Erythrobacter* spp. SD-21, *Sphingobium japonicum*, *Acinetobacter baumannii*, *Citromicrobiumbathyo marinum*, *Sphingobium chlorophenolicum*, *Polaromonas* spp. JS666, *Acidovorax* spp. JS42, *Rhodopseudomonas palustris*, and *Acinetobacter* spp. ADP1.

The vast majority of the species on the top 50 list that we extracted were found in two distinct orders, Sphingomonadales and *Burkholderiales*, which together contain 11 and 13 species, respectively. *Sphingopyxis alaskensis*, *Erythrobacter* spp. SD-21, *Sphingomonas wittichii*,

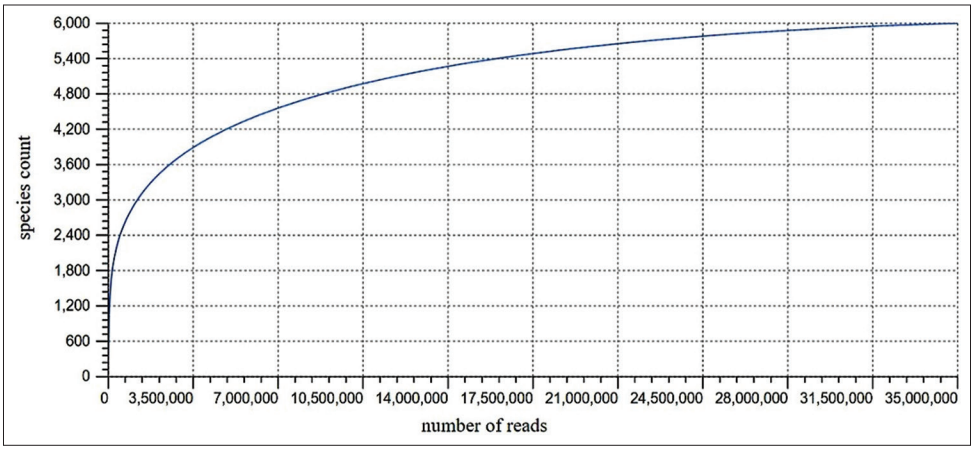

Figure 4: Rarefaction curve for the sampled water.

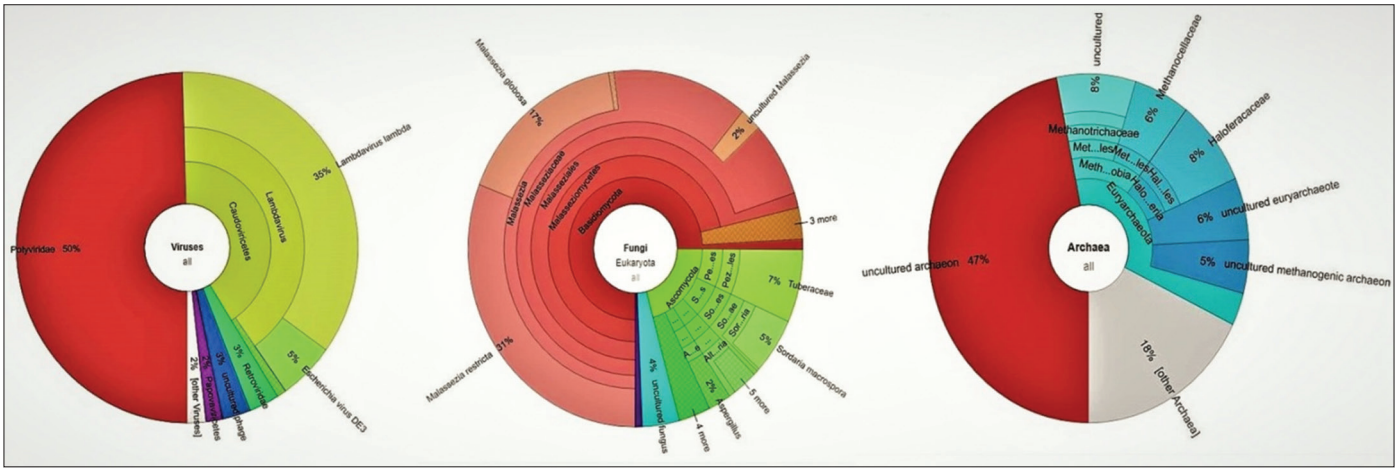

Figure 5: Taxonomic classification of viruses, fungi and archaea based on reads.

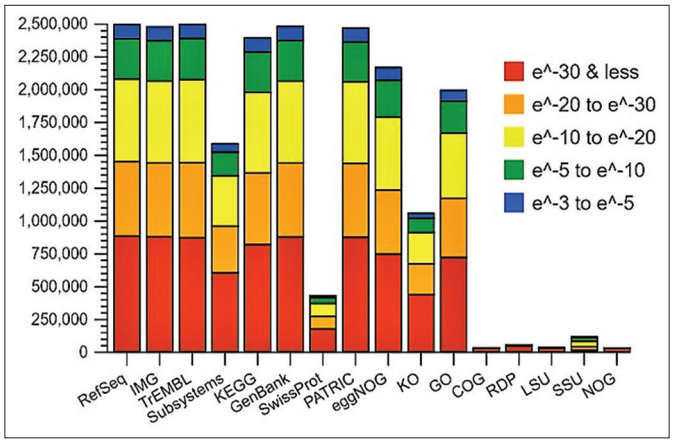

Figure 6: The graph displays the number of hits in the different databases listed. These include protein databases, protein databases with functional hierarchy information, and ribosomal RNA databases. The bars representing annotated reads are colored by an e-value range.

Sphingomonas spp. SKA58, *Citromicrobiumbathyo marinum*, and *Sphingobium indicum* belong to the order Sphingomonadales. *Burkholderiales* is represented by *Albidiferax ferrireducens*, *Acidovorax citrulli*, *Polaromonas naphthalenivorans*, *Variovorax paradoxus*, *Methylibium petroleiphilum*, and *Alicycliphilus denitrificans*.

4. CONCLUSION

The metagenome sample from the Ganga River shows an enriched distribution of numerous organisms from various domains, according to our research. The majority of the discovered creatures are bacteria, while eukaryotes, viruses, and archaea have demonstrated a noteworthy presence. It may be inferred from the phylogenetic analysis of the top 50 species that microbial species have undergone significant evolutionary diversity. A considerable difference in the relative abundance of particular microorganisms is also highlighted in the text. High anthropogenic activities at the sampling locations are also one of the reasons for such a high microbial diversity. Surveillance studies are vital for future enhancement and management strategies. They support explorations of various communities, environmental monitoring, and metagenomic resource utilization for their functional characterization. It would be very valuable for industrial uses where these organisms are needed, as well as for government agencies concerned with the ecological health of the river. In the future, new methods must be developed for the isolation of these organisms for commercial use and to protect and conserve the riverine ecology of the river.

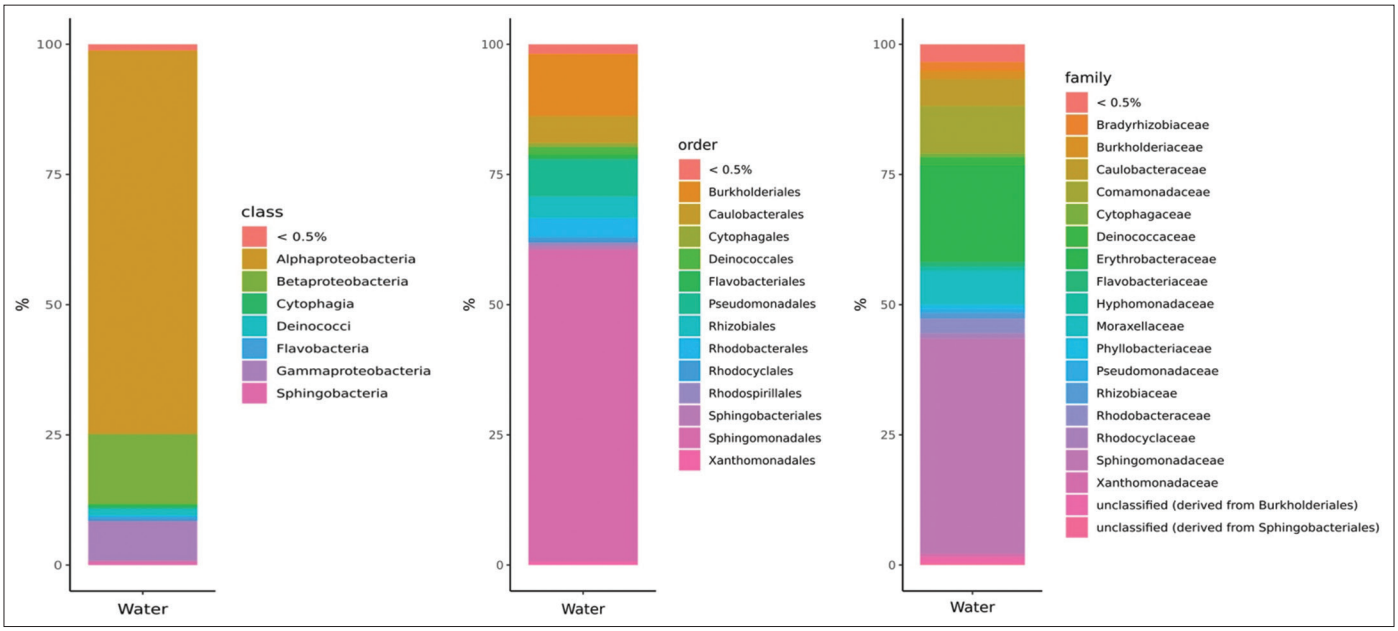

Figure 7: Stacked bar plots showing the average relative abundance of each taxa at the levels of class, order and families.

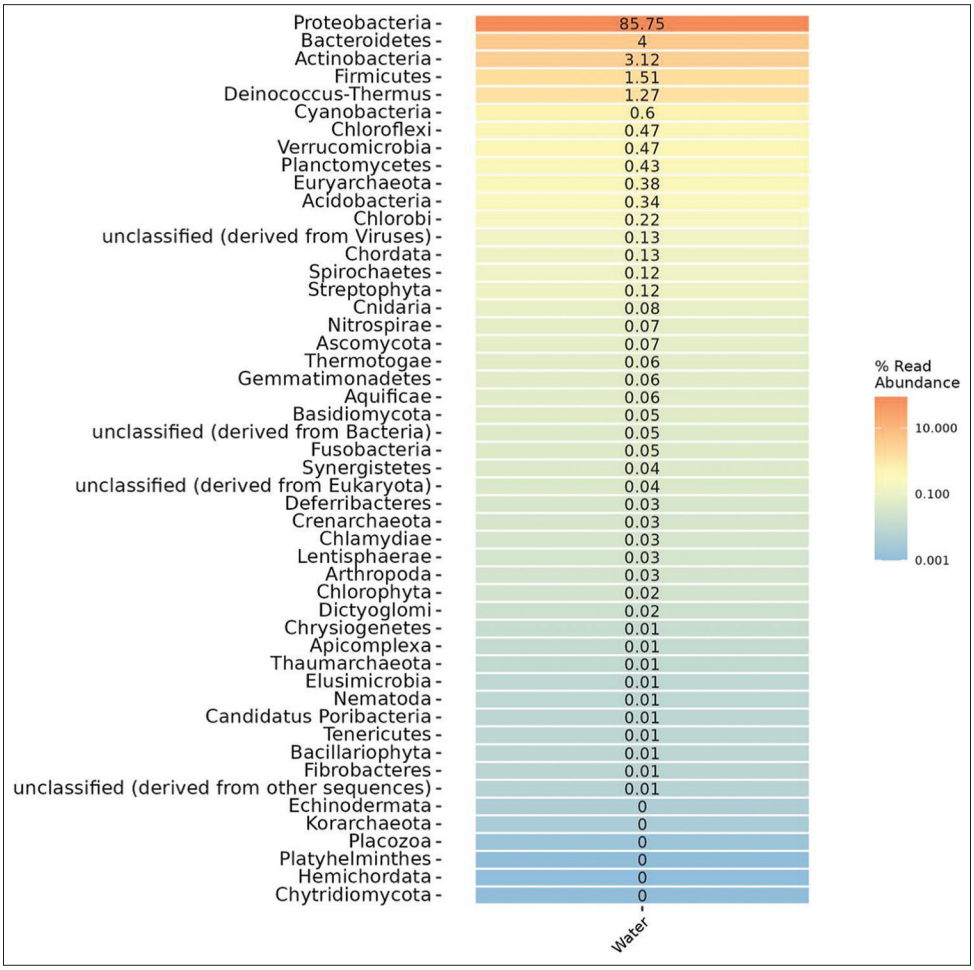

Figure 8: Heatmap at the phyla level showing the percentage read abundance at the phyla level.

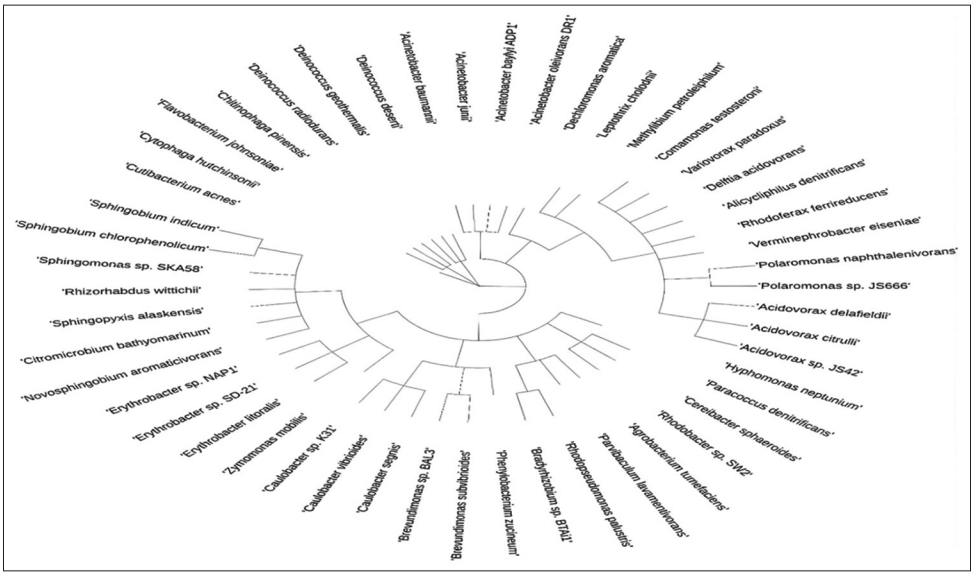

Figure 9: Phylogenetic tree at the species level based on the top 50 species from the sampled water using MG-RAST reference database.

5. ACKNOWLEDGEMENT

The authors wish to thank the Head, Department of Botany and Microbiology, Gurukula Kangri (deemed to be a university), Haridwar, for providing the necessary laboratory facilities.

6. AUTHOR'S CONTRIBUTIONS

AK and SC outlined and wrote the main manuscript text, analyzed the data, and prepared figures and tables. HVC, HC, and RCD reviewed the manuscript and made the necessary corrections. All authors read and approved the final manuscript.

7. FUNDING

There is no funding to report.

8. CONFLICTS OF INTERESTS

The authors declare that there is no conflict of interest.

9. ETHICAL APPROVALS

This study does not involve experiments on animals or human subjects.

10. DATA AVAILABILITY

All data supporting the findings of this study are available within the paper.

11. PUBLISHER'S NOTE

This journal remains neutral with regard to jurisdictional claims in published institutional affiliations.

REFERENCES

1. Singh M, Singh AK. Bibliography of environmental studies in natural characteristics and anthropogenic influences on the Ganga River. Environ Monit Assess 2007;129:421-32.
2. Singh IB. Geological evolution of ganga plain—An overview. J Palaeontol Society Ind 1996;41:99-137.
3. Joshi DM, Kumar A, Agrawal N. Assessment of the irrigation water quality of river Ganga in Haridwar district. Rasayan J Chem 2009;2:285-92.
4. Rao RJ. Biological resources of the Ganga River, India. Hydrobiologia 2001;458:159-68.
5. Rai B. Pollution and conservation of ganga river in Modern India. Int J Sci Res Publ 2013;3:1-4.
6. Kumar D. River Ganges-historical, cultural and socioeconomic attributes. Aquat Ecosyst Health Manag 2017;20:8-20.
7. Dwivedi S, Chauhan PS, Mishra S, Kumar A, Singh PK, Kamthan M, *et al*. Self-cleansing properties of Ganga during mass ritualistic bathing on Maha-Kumbh. Environ Monit Assess 2020;192:221.
8. NMCG-NEERI. Assessment of Water Quality and Sediment to Understand the Special Properties of River Ganga; 2017. Available from: https://nmcg.nic.in/writereaddata/fileupload/nmcgneeri%20ganga%20Report.pdf [Last accessed on 2023 Jul 27].
9. Nautiyal CS. Scientific validation of incorruptible the self-purificatory characteristic of Ganga water. Asian Agrihist 2009;13:53-6.
10. Nautiyal CS. Self-purificatory ganga water facilitates death of pathogenic *Escherichia coli* O157:H7. Curr Microbiol 2009;58:25-9.
11. Kumar N, Gupta AK, Sudan SK, Pal D, Randhawa V, Sahni G, *et al*. Abundance and diversity of phages, microbial taxa, and antibiotic resistance genes in the sediments of the river ganges through metagenomic approach. Microb Drug Resist 2021;27:1336-54.
12. Pankratova EM. Functioning of *Cyanobacteria* in soil ecosystems. Eurasian Soil Sci 2006;39:S118-27.
13. Zorriehzahra MJ, Delshad ST, Adel M, Tiwari R, Karthik K, Dhama K, *et al*. Probiotics as beneficial microbes in aquaculture: An update on their multiple modes of action: A review. Vet Q 2016;36:228-41.
14. Behera BK, Patra B, Chakraborty HJ, Sahu P, Rout AK, Sarkar DJ, *et al*. Metagenome analysis from the sediment of river Ganga and Yamuna: In search of beneficial microbiome. PLoS One 2020;15:e0239594.
15. Ahn JH, Grant SB. Characteristics of storm runoff and sediment dispersal in the San Pedro Channel, southern California. Water Sci Technol 2007;55:519-26.
16. Li K, Hu J, Li T, Liu F, Tao J, Liu J, *et al*. Microbial abundance and diversity investigations along rivers: Current knowledge and future directions. Wiley Interdiscip Rev Water 2021;8:e1547.
17. Staley C, Sadowsky MJ. Application of metagenomics to assess

microbial communities in water and other environmental matrices. J Mar Biol Assoc UK 2016;96:121-9.
18. Regar RK, Kamthan M, Gaur VK, Singh SP, Mishra S, Dwivedi S, *et al*. Microbiome Divergence Across Four Major Indian Riverine Water Ecosystems Impacted by Anthropogenic Contamination : A Comparative Metagenomic Analysis. Durham: Research Square; 2020. p. 1-24.
19. Ghosh A, Yash, Kumar C, Bhadury P. Metagenomic insights into the functional microbial diversity of the lower stretch of the River Ganga: Mapping antibiotic and metal resistance genes. Environ Res Commun 2023;5:095012.
20. Hong PY, Mantilla-Calderon D, Wang C. Metagenomics as a tool to monitor reclaimed-water quality. Appl Environ Microbiol 2020;86:e00724-20.
21. Choudhury N, Sahu TK, Rao A, Rout AR, Behera BK. An improved machine learning-based approach to assess the microbial diversity in major North Indian River Ecosystems. Genes (Basel) 2023;14:1082.
22. Matta G. Assessment of pollution on water quality and phytoplankton diversity in canal system of River.World J Pharm Res 2015;4:889-908.
23. Mohanty TR, Tiwari NK, Das BK, Kumari S. Plankton diversity and ecology of River Ganga. Sci Cult 2022;88:379-84.
24. Matta G, Uniyal DP. Assessment of species diversity and impact of pollution on limnological conditions of River Ganga. Int J Water 2017;11:87-102.
25. Bilgrami KS, Kumar S. Bacterial contamination in water of the River Ganga and its risk to human health. Int J Environ Health Res 1998;8:5-13.
26. Chaturvedi P, Chaurasia D, Pandey A, Gupta P. Co-oocurence of multidrug resistance β-lactamase and plasmid mediated AmpC genes in bacteria isolated from river Ganga, northern. Environ Pollut 2020;267:115502.
27. Behera BK, Chakraborty HJ, Patra B, Rout AK, Dehury B, Das BK, *et al*. Metagenomic analysis reveals bacterial and fungal diversity and their bioremediation potential from sediments of River Ganga and Yamuna in India. Front Microbiol 2020;11:556136.
28. Shekhawat SS, Kulshreshtha NM, Saini P, Upadhyay A, Gupta AB, Subramanian V, *et al*. Antibiotic resistance genes and bacterial diversity: A comparative molecular study of treated sewage from different origins and their impact on irrigated soils. Chemosphere 2022;307:136175.
29. Simon A. FastQC: A Quality Control Tool for High Throughput; 2010. Available from: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=FastQC%3A+a+quality+control+tool+for+high+throughput+sequence+data.+2010&btnG [Last accessed on 2023 Jul 30].
30. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884-90.
31. Glass E, Wilkening J, Wilke A, Antonopoulos D, Meyer F. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. Cold Spring Harb Protoc 2010;2010:pdb.prot5368.
32. Marcelino VR, Clausen PT, Buchmann JP, Wille M, IredellJR, Meyer W, *et al*. CCMetagen: Comprehensive and accurate identification of eukaryotes and prokaryotes in metagenomic data. Genome Biol 2020;21:103.
33. Federhen S. The NCBI taxonomy database. Nucleic Acids Res 2012;40:D136-43.
34. Letunic I, Bork P. Interactive tree of life (iTOL) v4: Recent updates and new developments. Nucleic Acids Res 2019;47:W256-9.
35. McMurdie PJ, Holmes S. Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 2013;8:e61217.
36. Andersen K, Kirkegaard R, Karst S. ampvis2: An R Package to Analyse and Visualise 16S rRNA Amplicon Data. BioRxiv; 2018; p.299537.
37. Reddy B, Dubey SK. Exploring the allochthonous pollution influence on bacterial community and co-occurrence dynamics of River Ganga water through 16S rRNA-tagged amplicon metagenome. Environ Sci Pollut Res 2021;28:26990-7005.
38. Behera BK, Sahu P, Rout AK, Parida PK, Sarkar DJ, Kaushik NK, *et al*. Exploring microbiome from sediments of River Ganga using a metagenomic approach. Aquat Ecosyst Health Manag 2021;24:12-22.
39. Rout AK, Dehury B, Parida PK, Sarkar DJ, Behera B, Das BK, *et al*. Taxonomic profiling and functional gene annotation of microbial communities in sediment of river Ganga at Kanpur, India: Insights from whole-genome metagenomics study. Environ Sci Pollut Res 2022;29:82309-23.

How to cite this article:
Katara A, Chand S, Chaudhary H, Chandra H, Dubey RC. Microbial diversity of river Ganga at Haridwar (Uttarakhand) through metagenomic approaches. JApp Biol Biotech. 2024;12(2):295-303.