

# Microbiome divergence across four major Indian riverine water ecosystems impacted by anthropogenic contamination: A comparative metagenomic analysis

**Raj Kumar Regar**

CSIR-Indian Institute of Toxicology Research

**Mohan Kamthan**

Jamia Hamdard, New Delhi

**Vivek Kumar gaur**

CSIR-Indian Institute of Toxicology Research

**Satyendra Pratap Singh**

NBRI: National Botanical Research Institute CSIR

**Seema Mishra**

Deen Dayal Upadhyaya Gorakhpur University

**Sanjay Dwivedi**

NBRI: National Botanical Research Institute CSIR

**Aradhana Mishra**

National Botanical Research Institute CSIR

**Natesan Manickam**

CSIR-Indian Institute of Toxicology Research

**Chandra Shekhar Nautiyal** (✉ [nautiyalnri@lycos.com](mailto:nautiyalnri@lycos.com))

National Botanical Research Institute CSIR <https://orcid.org/0000-0002-3581-9273>

---

## Research

**Keywords:** Bacteriophage, Microbial diversity, Metal resistance genes, Ganga river, Cauvery river, Narmada river

**Posted Date:** November 18th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-107257/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

# Background

Indian rivers are a major source of livelihood as river water is used for drinking, agriculture, and religious purposes to a large population. In this study, we report comparative microbial structures and functional potential of four major rivers of India, namely Ganga, Narmada, Cauvery, and Gomti. Comparative microbiome study of these geographically distinct rivers was performed using the samples collected from the source to the downstream sites of each river. We employed metagenomic approach to comprehensively determine the taxonomic and functional potential of river microbiome.

## Results

In this study, we report the pollution influences on microbial composition and functional potential of four distantly located rivers. Results revealed significant microbial diversity in contaminated locations as compared to the upstream samples. A total number of 37 bacterial phyla were detected out of which *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, *Planctomycetes*, and *Verrucomicrobia* were abundant. Microbial diversity in respect to anthropogenic activities revealed the prevalence of *Acidobacteria*, *Actinobacteria*, *Verrucomicrobia*, *Firmicutes*, and *Nitrospirae* phyla, whereas a decline in *Proteobacteria* and *Bacteroides*. Virulent and temperate bacteriophages were found high in Ganga when compared to others. Interestingly, the abundance of bacteriophage decreased with increasing pollution load in the river Ganga, unlike in other rivers. The carbon utilization studies indicated a correlation with functional genes occurred in metal contaminated sites. Ganga water has relatively higher trace elements at pristine-upstream than in the Narmada and Cauvery, indicating its origin from Himalayan rocky mountains and also both Ganga and Cauvery rivers found to harbour a large number of metal resistance genes.

## Conclusion

Our findings indicate a correlation between pollution and the microbiome composition. The insights obtained suggest the role of high abundance of microbial communities with implications for human health and demonstrate the functional capabilities contributed by the microbial communities. Among the four rivers studied, the distinctiveness of Ganga in comparison to others, particularly upstream of Ganga revealed a highly dynamic microbial structure. Bhagirathi and Alaknanda confluence to form Ganga, the microbiome revealed that Alaknanda has the foremost contribution to Ganga with respect to microbial community, bacteriophages, and the type of trace elements and heavy metals detected.

## 1. Background

Rivers are not only the primary source of freshwater crucial to the welfare of humankind, but they also have an important place in culture, tradition, and as a natural mode of transportation [1-5]. Ganga, the largest river of the Indian subcontinent originates from Himalayan glaciers at an elevation of about 13,200 ft. It discharges into the Bay of Bengal after travelling for over 2525 km. The Narmada originates from the north-eastern end of Satpura range in Amarkantak and discharges into the Arabian Sea after passing 1312 km through the valley of Satpura and Vindhya ranges. The Cauvery originates from Talacauvery pond at an elevation of 4,400 ft on the Brahmagiri range in the Western Ghats in the state of Karnataka. It flows for about 800 km before amalgamating in the Bay of Bengal. River Gomti is a tributary of river Ganga originating from Gomat Taal from the Tarai region of Himalayan foothills in Pilibhit district of Uttar Pradesh, India. The main focus of the current study was to evaluate the occurrence of specific microbial communities of each river, particularly of Ganga, which has special significance in Hindu culture. Ganga is believed to be soul purifier and believed to have antimicrobial properties [2]. A recent study has shown that the Ganga river contains higher bacteriophage density than the river Yamuna [1].

The increasing population and continuous developmental activities of our society have enhanced the discharge of municipal and industrial wastes into the rivers. This discharge mainly consists of organic and inorganic pollutants which are major stressors for the river water ecosystem and severely affected the water quality. Moreover, these pollutants make way into the food chain and adversely affect both animals and humans [6-9]. The ever increasing municipal, hospital, and agricultural waste along with industrial wastewater from pharmaceuticals, pesticides, dyes, paper, pulp, and tanneries, burden these rivers with complex mixtures of pollutants. Usually, these wastes contain metals, solvents, organic chemicals that are highly persistent in the river ecosystem [3, 8, 10-13]. These contaminants, directly and indirectly, found to be responsible for the development of antimicrobial resistance (AMR) and other biocide resistance in microorganisms [13, 14]. Therefore, many river water ecosystems are becoming a reservoir of pathogens, contributing to AMR dissemination and as a breeding sink for spreading AMR-genes through horizontal gene transfer in the ecosystem [15-18].

Since microbial communities are key players in the river ecosystem to maintain the water quality by degrading the pollutants. Microorganisms have a large number of metabolic genes that help in maintaining biogeochemical cycling in environments [19, 20]. Therefore, a comprehensive metagenomic study was conducted to understand the microbial community structure and their role from pristine-upstream to downstream sites of the rivers. In this study, we also report the extensive physicochemical analysis of four rivers Ganga, Narmada, Cauvery, and Gomti having different geographical path was conducted. To the best of our knowledge, this is the first comparative microbiome study of four Indian rivers which correlates the river microbial diversity to their function, physicochemical parameters, metals, and bacteriophage. The study further aimed to elucidate the effect of anthropogenic interventions on the microbial community and overall water quality of the rivers.

## 2. Results

### 2.1. Overview of metagenome sequencing

A comparative microbiome study of four rivers, namely Ganga, Gomti, Narmada, and Cauvery situated in different geographical locations of India, was comprehensively performed to understand the specificity as well as similarity. Only 1% of the total genome can be identified by employing culture-based approaches [21]. However, metagenomics offers an attractive tool to study the structure of microbial communities present in the aquatic ecosystems [17, 22]. The sequence reads generated using high-throughput Illumina HiSeq 2500, were filtered to remove low-quality reads, followed by dereplication, which generated more than 29 million high-quality reads for each sample (Table 1). A total of 560,321,208 high-quality paired-end reads were produced from 19 samples representing ~90% of the total data. The dereplicated fragments ranging from 19.51-33.14 million was obtained for all the four river samples. The rarefaction curve for the microbial taxonomy reached saturation plateau exhibiting high coverage of data, indicating that the maximum richness of microbes was covered (Additional file 1: Fig. S1). Dendrogram hierarchical clustering at species level showed closeness among the downstream samples, whereas the upstream samples were more closeness to each other. Further, close association is seen in the microbial diversity of upstream and middle stretch of the Ganga (G1-G5) and upstream of Cauvery river (C1) (Additional file 1: Fig. S2). Whereas anthropogenically influenced downstream of Ganga (G6 & G7), Cauvery (C3) and Narmada (N4) form a distinct group.

## 2.2. Comparative GC content characteristics of river metagenomic datasets

GC content of a microbiome was reported to exhibit a dependency on the physicochemical characteristics of respective habitat [23, 24]. It was found that upstream locations of river samples G1, C1, N1, and C3 sample showed low GC content (~53% GC) and similarity was observed in GC distribution pattern of the sequences. In the diversity analysis, it was observed that these four samples showed similar bacterial diversity by forming a separate clade in operational taxonomic unit (OTU) cluster analysis (Fig 7). Whereas downstream samples G6, G7, C4 and N4 showed similar GC content distribution patterns and also formed a separate clade in OTU cluster analysis. Ganga river samples G6 and G7 showed similar GC distribution patterns, whereas G1, G3, G4 and G5 showed similar GC distribution patterns, interestingly, G2 sample showed an entirely different GC pattern with respect to other samples of Ganga. Cauvery river sample C1 was closer to C3, and C2 was found to be closer to C4 in terms of GC content distribution pattern. In the Narmada river, N1 exhibited very different GC content (~53%) distribution as compared to the other four samples N2, N3, N4 and N5, which had 58 - 62%. All three samples of Gomti river showed high GC content of 58 - 62%. It was also observed that the GC content distribution pattern was highly similar to the occurrence of bacterial diversity clustered at OTUs level. Thus, GC distribution pattern correlates very well with the OTUs cluster analysis, as shown in additional file 1: Fig. S3 and Fig 7.

## 2.3. Microbial diversity structure

Assessment of alpha diversity indices revealed significant differences between the pristine and polluted sites of all the rivers. The estimates of the observed OTUs richness and evenness were found to increase at the polluted site in comparison to the pristine sites in all the rivers. Furthermore, the statistical significance of the grouping based on experimental factors was also estimated using either parametric or nonparametric tests. Alpha diversity measurement across all the samples for the given diversity index is provided in Fig.2. The boxplot analysis of alpha diversity richness and evenness results indicated that Gomti river has the highest bacterial diversity, followed by Narmada, Ganga, and Cauvery (Fig. 1) in that order. Chao-1 index showed that Lko2\_Bhatpur have the highest richness (chao1-2229) followed by N4\_Hoshangabad (chao1-2222), G6\_Bithoor (chao1-2175) and C4\_Erode (chao1-2162) representing the Gomti, Narmada, Ganga, and Cauvery rivers respectively. Shannon index showed that Lko3\_Gaughat have highest evenness (Shannon -23.24) followed by C4\_Erode (Shannon-22.76), N4\_Hoshangabad (Shannon-16.71), and G6\_Bithoor (Shannon- 11.65) representing the Gomti, Cauvery, Narmada, and Ganga, rivers respectively (Table 1). When multiple samples were compared, a clear separate bracket in PCoA and NMDS plot was formed based on the pollution load using the phyloseq package3 in MicrobiomeAnalyst (Fig. 3). This bracket has downstream samples of all the rivers that include G6 and G7 of Ganga, C4 of Cauvery, N4 and N5 of Narmada and all three samples of Gomti.

## 2.4. Bacterial community composition

The bacterial community structure showed a significant increase in bacterial diversity with increasing anthropogenic activities at genus, order, class, and other higher levels of taxonomy. A total of 37 bacterial phyla were detected out of which the *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, *Planctomycetes*, and *Verrucomicrobia* were found to be prevalent in all samples (Fig. 4A). The increasing anthropogenic activity showed the dominance of *gram*-positive phyla, namely *Acidobacteria*, *Actinobacteria*, *Verrucomicrobia*, *Firmicutes* and *Nitrospirae* and *gram*-negative phyla such as *Proteobacteria* and *Bacteroides*. At class level *Alphaproteobacteria* and *Betaproteobacteria* were low and *Deltaproteobacteria*, *Gammaproteobacteria*, and *Planctomycetia* were found to be more abundant (Fig. 4B). For easy presentation and understanding, a graphical representation using the krona chart was prepared. These krona charts show bacterial diversity from the phylum level to species level with their relative abundance (Additional file 3-6).

## 2.5. Core microbiome analysis

Core bacterial orders identified by using MicrobiomeAnalyst tool indicated the unchanged microbiome composition across the samples from different river stretches. Core microbiome analysis at order level showed *Burkholderiales*, *Rhizobiales*, *Planctomycetales*, *Rhodospirillales*, *Sphingomonadales*, *Verrucomicrobiales*, *Rhodobacterales*, *Cytophagales*, *Acidimicrobiales*, and *Nevskiales* prevailed in all the rivers (Additional file 1: Fig. S4) (Additional file 2: Table S1). Core microbiome analysis is adopted from the core function in R package microbiome [25]. The result showing comparative core microbiome occurrence is shown in a heatmap (Additional file 1: Fig. S4).

## 2.6. Bacteriophages profiling

In the present study, both virulent and temperate phages were analyzed in all the samples. We studied the virulent phages of the six most commonly found pathogenic bacteria in river water systems. These bacteria are *Staphylococcus aureus*, *Enterococcus faecalis*, *Pseudomonas aeruginosa*, *Salmonella typhimurium*, *Escherichia coli* and *Vibrio cholera*. For this, dsDNA of phage that is incorporated into the bacteria (prophage) were analyzed from the total DNA. Both virulent and temperate phages are key drivers of bacterial community composition in natural aquatic systems. However, prophages were found to play a vital role in horizontal gene transfer [26-29]. Prophage is believed to provide several advantages to the host bacteria such as stress tolerance, biofilm formation, antibiotic resistance etc.

Analysis of virulent phage against six bacteria showed that river Ganga exhibits high diversity as compared to other rivers (Table 2). Interestingly, it was observed that the presence of *S. aureus* phage in G1\_Bhagirathi and G3\_Deprayag, the upstream samples of river Ganga. However, it was not observed in downstream samples which were more influenced by anthropogenic activities. Intriguingly, river Ganga also showed an abundance of phage in pristine samples. Whereas, Cauvery and Narmada showed an abundance of phages in samples where anthropogenic interferences were more. In the case of Gomti river, as it receives a large volume of wastewater, it showed the highest concentration of phages for the hosts *E. coli* and *E. faecalis* (Table 2).

High throughput metagenomic sequencing revealed that the prophage related genes observed in this study are associated with 218 bacterial species. Alaknanda sample of river Ganga showed the highest abundance of phage specific genes, followed by Bhatpur sample of river Gomti. In comparison to other rivers, Ganga showed a much higher abundance of prophage specific genes, namely phage prohead/capsid assembly, phase terminase large subunit, phase portal protein, endolysin activity. In Ganga river, prophages are mainly contributed by G2\_Alaknanda river which exhibits the highest number of prophage genes among all samples. These results suggested that like virulent phage, there is also an abundance of temperate phage presence in river Ganga. The temperate phage abundance pattern in various rivers indicates that increasing anthropogenic activities has led to a decrease in the population of prophage (Fig. 5A). Analysis of the host bacteria harbouring the prophage showed a preponderance of phylum *Proteobacteria* and *Terrabacteria*. Among the top 15 bacteria that were harbouring maximum phage genes, 13 belonged to *Proteobacteria* and two others were found to be associated with *Terrabacteria*. At the class level, the dominant bacteria harbouring bacteriophages were from *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, *Deltaproteobacteria*, *Firmicutes* and *Chloroflexi* (Fig. 5B).

The prophage specific genes were analysed using the ACLAME database. Our analysis showed that for most of the identified genes, ACLAME functional annotation was not available. However, the functionally annotated genes were found to code for prohead/capsid assembly followed by phage terminase large subunit, phage portal protein and others were present (Fig S5). River Ganga showed the highest abundance of prophage specific genes. Interestingly, the samples from industrially polluted sites of rivers Ganga, Gomti and Cauvery showed a significant dip in the abundance of prophage specific genes.

## 2.7. Microbial community carbon utilization pattern analysis using Biolog

Functional metabolic diversity of microbial communities at different sites of rivers is reflected in terms of Average Well Colour Development (AWCD, Additional file 1: Fig. S6 A-D). Among the four rivers, the maximum AWCD range was observed in Ganga (0.701-1.946; Additional file 1: Fig. S6 B), whereas, the lowest AWCD range was found in Narmada (0.523-0.98; Additional file 1: Fig. S6 D). Moreover, higher AWCD was documented in the downstream sites of all the rivers in comparison to the upstream sites. The maximum AWCD was found at G6\_Bithoor (1.94) and G7\_Jajmau (1.75) sites of Ganga and lowest in Narmada N1\_Udgam (0.522; upstream sampling site). The diversity richness and evenness indices show significant differences between upstream and downstream sampling sites (Additional file 2: Table S2). The microbial communities based on carbon substrate utilization pattern showed two distinct clusters, for instance, river Ganga and Narmada clustered together, whereas Gomti and Cauvery formed a distinct cluster (Fig. 6 A). Notably, an analysis of carbon utilization pattern of samples individually showed an interesting trend, as the water samples collected from downstream sites clustered separately, except Gomti water samples (Additional file 1: Fig. S7 A). Cauvery showed the presence of similar carbon utilization pattern at C1\_Talacauvery and C3\_KRS Dam regions. However, the microbial communities of C2\_Bhagamandla and C4\_Erode showed a distinct pattern of substrate usage (Additional file 1: Fig. S7 B). In Ganga, G3\_Deprayag, G2\_Alaknanda and G4\_Rishikesh, and G5\_Haridwar were clustered close to each other. Also, the microbial communities from G7\_Jajmau showed a distant pattern of carbon substrate usage and cluster in an entirely distinct way (Additional file 1: Fig. S7 C). A similar pattern was also noted for Narmada, which showed a close clustering within the microbial communities of N1\_Udgam and N2\_Kapildhara followed by N3\_Bhedaghat and N4\_Hoshangabad. The microbial communities of N5\_Omkereshwar utilized different carbons as compared to the other four samples (N1 to N4) and hence clustered separately (Additional file 1: Fig. S7 D). The higher utilization pattern of carbon substrates in MT plates was recorded by Ganga followed by Narmada, Gomti and Cauvery (Fig. 6 B). In amino acids, higher utilization of Serine and L-Proline was noticed in Gomti, which was followed by Serine in Cauvery. Ganga and Narmada showed a maximum preference for Glutamine (Fig. 6 B). In carbohydrates and miscellaneous, casein was utilized maximum in Ganga. Furthermore, chitin was the most preferred polymer in all rivers, and Gomti and Ganga utilized dextrin (Fig. 6 B). In carboxylic acids group, microbial communities of Gomti showed a preference for succinic acid, whereas Narmada, Ganga and Cauvery favoured the usage of citric acid as primary carbon source (Fig. 6 B). Persistent pesticides were also found to be utilized by different bacterial communities present in the rivers, for examples the Gomti river sample showed a preference for anthracene, Cauvery for endosulfan, Ganga and Narmada for most of the pollutants studied (Fig. 6 B). Moreover, site-specific carbon substrate utilization pattern revealed that G6\_Bithoor and G7\_Jajmau utilize a maximum site-specific carbon source (Additional file 1: Fig. S8). Microbial community of G4\_Rishikesh sample utilized maximum carbohydrates. Furthermore, N5\_Omkareshwar also showed higher use of different substrates (Additional file 1: Fig. S8).

## 2.8. Cluster analysis of microbial community in respect to carbon utilization pattern

Microbial community metabolic potential assay was performed using Biolog® Eco and MT plates. The AWCD data revealed the metabolic activity of microbial community which reflect the functional potential of actively available microbes. We used the AWCD data generated by growing bacterial communities on different carbon sources and correlated the above data with the microbial communities at species level. Data analysis was performed by using the

Microbiome Analyst tool. It was observed that sample G1, C1, N1 and C3 (mostly upstream sites); sample G2, G3, G4, and G5 (Ganga middle stretch); samples N2, N3, N5, Lko1, Lko2, Lko3 and C2 (mostly the middle stretch of Narmada, Cauvery and all three samples of Gomti), and samples G6, G7, N4 and C4 (downstream of rivers) formed separate clusters (Fig.6). Based on these carbon sources utilization pattern and their correlation with microbial diversity at the species level it may be concluded that anthropogenic activities influence the structure, occurrence and adaptation of the microbial communities in these perturbed ecosystems.

## 2.9. Impact of physicochemical parameters on microbial diversity

In an ecosystem, physicochemical parameters play a crucial role for the growth of organisms and therefore, the microbial diversity results were correlated with the physicochemical parameters obtained in the study (Additional file 2: Table S3). A significant correlation between microbiome and major physicochemical parameters of the samples analyzed through PCoA plot is provided as an Additional file 1: Fig. S9. Identification of statistically significant and enriched OTUs with respect to various parameters was performed through ANOVA and Parametric Student T-test. The number of key OTUs that are enriched within each metagenome given in Additional file 2: Table S4. We were able to identify OTUs that are uniquely enriched in one or more parameters using Gene Matrix algorithm. Nitrate and pH have the highest number of OTUs that are uniquely enriched in the respective comparisons. With respect to high pH (>8) genus *Bordetella*, *Sphingomonas*, *Achromobacter*, *Phenylobacterium*, *Phenylobacterium*, *Sphingobium*, *Erythrobacter* and *Erythrobacter* was found enriched whereas nitrate has led to the enrichment of *Haliscomenobacter*, *Malonomonas*, *Hansschlegelia*, *Hyphomicrobium*, *Chelativorans*, *Peptostreptococcus*, *Psychrobacter*, *Psychrobacter*, *Psychrobacter* and *Psychrobacter*.

## 2.10. Metal contamination and its correlation with microbial community

High levels of metals were detected in large number of samples that may be generated and introduced owing to the industrial discharges. High concentrations 70 and 100 µg/L of trace elements such as Mn, Fe, Co, Cu were recorded in Bhagirathi and Alaknanda (the upstream of Ganga) respectively as compared to 60 µg/L at Narmada Udgam. Their concentration further gradually increased in the downstream samples of Ganga (over 2500µg/L at Jajmau), however there was no significant change in Narmada. Cauvery has a relatively lower concentration of trace elements, except Cu (with <15µg/L total trace elements). In the river Gomti, the level of both trace and toxic metals were relatively high at all the sampling sites, i.e. the total level of trace and toxic elements up to 208 µg/L and 150µg/L respectively. The samples of Ganga from Bithoor (G6) and Jajmau (G7) showed the highest concentration of all the elements except Cd, which was highest in Gomti river (Fig 8). In Narmada and Cauvery, chromium (Cr) and arsenic (As) were mostly present at downstream while trace of lead (Pb) and cadmium (Cd) were present at all sites. Interestingly, the level of As was higher at all sites of Ganga in comparison to other rivers (Fig 8). Consequently, metal and antibiotic resistant bacteria identified as *Acinetobacter baumannii* species from the Bhagirathi river and *Alcaligenes faecalis* species from Cauvery river were isolated (data not shown) and characterized.

It is significant to note that high levels of metal resistant genes (MRGs) were detected in pristine water samples as compared to downstream polluted samples. The highest MRGs were detected in C1 sample of Cauvery, followed by G2 in Ganga and N2 of Narmada river (Fig. 9A), which are upstream locations of the rivers. Maximum MRGs was detected for copper followed by chromium (Cr), arsenic, zinc and iron. Most of the MRGs were located on bacterial chromosome (91.5%), whereas rest (8.5%) were found to be located in plasmids (Fig S10). The highest number of plasmids located MRGs were recorded in Cauvery C1 (22.32%) and Ganga G2 (14.32%) samples, whereas all other samples have low percentage, ranging from 4 to 10% (Fig. 9D). Besides MRGs, biocides and other chemical resistance genes for ethidium bromide, rhodamine 6G, acriflavine and triclosan were also found in abundance in the pristine sample as compared to their respective downstream samples (Fig. 9B). The PCoA plot of bacterial data at species level with respect to the metal concentration using the phyloseq package3 in MicrobiomeAnalyst depicted a significant relation of trace metals with respective microbial communities (Fig. 10) revealing their direct correlation with the presence of trace metals.

## 3. Discussion

Rivers have been lifeline as water source however in many parts of the world massive human settlements surrounding the rivers have become a critical issue along with huge volumes of untreated effluents and waste sewage discharged into the rivers. As per the 2018 report of Central Pollution Control Board (CPCB), New Delhi, India, an estimated 6.07 billion litres per day of industrial and sewage wastewater was discharged in the Ganga river [30]. Severe contamination in these waters is expected to pose alterations in the native microbiota of the river water ecosystem. Worldwide, around 4.0% of deaths and 5.7% of the global disease burden was caused by waterborne diseases [31]. Whereas in India, an estimated 37.7 million people were affected by the waterborne diseases annually, and an estimated 1.5 million children die every year due to diarrhoea alone [32]. The rivers chosen for the study are geographically very distantly located, having varying climatic conditions, almost covering North, Middle and Southern part of India. The analysis of the metagenome based microbial community data of these four river samples has revealed certain common and yet unique patterns. Data obtained also provides a complete picture on the alterations on microbial community structure and acquisition of pathogenesis because of anthropogenic particularly industrial influences causing a greater risk to public health.

All the four rivers studied showed the prevalence of phylum *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, *Planctomycetes*, and *Verrucomicrobia* suggesting that these are the most flexible phyla which may get enriched even under constant variation of climate, human interference, soil types and physicochemical water quality parameters commonly observed in any river ecosystem. We have also observed different community structure with increasing anthropogenic activities for example, a decline in phyla *Proteobacteria*, and *Bacteroides* whereas phyla *Acidobacteria*, *Actinobacteria*, *Verrucomicrobia*, *Firmicutes* and *Nitrospirae* were specifically found enriched. The phyla *Acidobacteria*, *Actinobacteria*, and *Firmicutes* were reported to be associated with wastewater and fecal contaminants [22]. Similarly, while moving downstream along the rivers, the classes *Deltaproteobacteria*, *Gammaproteobacteria* and *Planctomycetia*

were found enriched. Interestingly, many pathogenic genera such as *Acinetobacter*, *Aeromonas*, *Escherichia*, and *Pseudomonas* belonged to class *Gammaproteobacteria*. This is an indication of possible human interference, enhancing the load of pathogenic microbial communities.

Furthermore, it was observed by alpha diversity analysis that Gomti River showed the highest richness and evenness thus exhibiting highest bacterial diversity. This can be correlated to the high human influence in the Gomti River as also evident from previous studies [33, 34]. Furthermore, Gomti river owing to its origin from plains and with human interferences exhibits high contamination in origin itself as compared to the upstream samples of other rivers. This leads to increased diversification and richness in the Gomti river sample. Similarly, as evident from Fig. 2 within the river samples, the increase in contamination leads to increased richness and evenness in the samples in all the four rivers. However, C1\_ Talacauvery and C3\_KRS dam sample of Cauvery River showed less richness and evenness, and the reason to this may be attributed to relatively static water conditions at these locations. Previous studies have also suggested the distinct variations in sediment bacterial community in dam-affected sites [35]. The studies indicate that dam construction can cause changes in microbial diversities. More detailed studies are required to study, the effect of dam constructions on *ecosystems* and its overall impact, on the environment. At order level classification in core microbiome analysis, the most abundant five orders were *Burkholderiales*, *Rhizobiales*, *Actinobacteria*, *Bacteroidetes*, and *Planctomycetia*. It was previously reported that these orders were common to freshwater, lakes and rivers [36-38]. They are oligotrophic in nature and are able to adapt to diverse variations in their niches such as metal, pH, pesticides, and other toxicants or low availability of metabolic substrate [39, 40]. Thus, the core bacterial orders such as *Burkholderiales*, *Rhizobiales*, *Planctomycetales*, *Rhodospirillales*, *Sphingomonadales*, *Rhodobacterales* and *Acidimicrobiales* play a crucial role in maintaining the ecosystem [41].

Bacteriophage analysis suggests that out of the four rivers analysed, Ganga has the highest abundance and diversity of both virulent and temperate phages. In contrast to other rivers studied, the pristine samples of river Ganga also showed a prevalence of several phages including *S. aureus* phage. Since infections of *S. aureus* are often associated with wounds causing pain and pus formation, therefore the presence of *S. aureus* phage further strengthens the belief that Ganga water has healing properties. It is expected that the host specific phage would control the population of *S. aureus* within the wounds causing relief from pain when water is applied to the wounds. The uniqueness of Ganga was further emphasized by the highest abundance of prophages and higher abundance of phages in comparison to other rivers. However, the increasing anthropogenic activity has led to a significant decline in the phage population. From these indications it is evident that bacteriophages play an important role in shaping the diversity and abundance of microbial communities in natural river water systems. Therefore, shift in the abundance and types of bacteriophage with anthropogenic activities would also cause a change in bacterial communities.

Metal contamination has been reported to affect the river microflora and leads to the acquisition of metal tolerance in the microorganisms [42]. The higher levels of trace elements in Ganga upstream (G1-G5) samples in comparison to other locations are attributed to its hilly origin. While the increased level of toxic elements in downstream samples, e.g. G6 and G7 is attributed to the increase in human bathing, domestic discharges, industrial wastes and wide cattle grazing activities [8, 43, 44]. Gomti river samples showed high level of elements throughout the sampling sites, suggesting the origin of the river through surface runoff. The occurrence of the high load of *E. coli* phages and *E. faecalis* phages in Gomti river also indicates the high sewage and fecal load presence in the river.

Carbon substrate utilization profiling has great potential as a technique to evaluate the quality of different sources of water bodies [45]. This technique has previously been employed for analyzing the metabolic fingerprints of microbial communities [2, 46-49]. We have observed a different carbon utilization pattern for each water samples. This may be due to the difference in the water quality, which nurtures diverse structure, the composition of microbial communities based on the nutrients [50, 51] and other physico-chemical characteristics. The microbial communities based on carbon substrate utilization showed interesting results for instance, in Ganga, the samples from Alaknanda to Haridwar (G2 to G5) clustered together demonstrating that river Ganga gets most of its attributes from the source stream Alaknanda. In addition, the dissolved solids and mineral load in Alaknanda are significantly higher than Bhagirathi [52]. This difference probably results in enriching different microbial communities in Bhagirathi than Alaknanda as observed by the results of carbon source utilization pattern. Similarly, highly contaminated sites of Bithoor (G6) and Jajmau (G7) also distinctly positioned based on principal component analysis for carbon utilization pattern. Our results suggest that downstream samples showed a higher carbon utilization pattern than upstream sampling sites which were more polluted because of sewage and industrial effluents discharges [53]. Also, the organic carbon processing is a characteristic property of heterotrophic microorganisms whose number may increase because of the increasing contamination [54]. Therefore, it is strictly recommended that the industrial effluents should be fully treated and diluted before its release into the rivers [55]. The physiological profiling of the microbial communities demonstrates that the amino acids, carbohydrates and carboxylic acids were most utilized in the upstream sites compare to downstream locations. It was documented that several heavy metals such as Cr, Cd, Hg, Pb, As and other pollutants are regularly disposed into the river bodies [43, 44, 56, 57]. Therefore, the availability of these substrates becomes reduced to the bacterial communities to assimilate them once high concentrations of pollutants reach rivers.

Moreover, the higher utilization of pollutants found in Ganga may be due to the reason that it covers most of the land of Uttar-Pradesh, Uttarakhand, Bihar and West Bengal (almost 1,000,000 km<sup>2</sup>) which are highly industrialized and under intense agricultural practices. Thus, pollutants like heavy metals, pharmaceuticals and pesticides reach routinely in Ganga river owing to industrial effluents and surface runoffs [8, 53, 58-60]. The present findings highlight that microbial communities differed in all the river water samples in terms of their composition quantitatively from upstream to downstream. The microbial analysis results were significantly correlated with carbon utilization profiling of all the water samples which were in relative proportion to their community members that were subjected to physicochemical changes in the environment (water body) as well as by the physiological and metabolic perturbations of the organisms.

## 4. Conclusions

The metagenomic strategy employed in this study allowed us to identify common microbiome as well as few unique signatures of microbial communities occurring in the four major geographically distantly located rivers namely Ganga, Gomti, Narmada and Cauvery. Understandably, microbial activities shape the biogeochemistry, water quality parameters and efficiency in maintaining the health of river ecosystems. The results showed that the studied rivers have

specific microbial community in upstream samples based on their geography and environment of origin. River Ganga showed several unique characteristics in terms of higher concentration of trace elements, abundant microbial communities, unique community structure, and abundance of bacteriophages. However, increasing anthropogenic activities and the confluence of different waste streams in the rivers alter their microbial community structure which leads to enrichment of specific phyla such as *Acidobacteria*, *Actinobacteria*, *Verrucomicrobia*, *Firmicutes* and *Nitrospirae*. The metabolic variations in river Ganga reflected in the pristine and contaminated stretches vary differently. A similar pattern in the increase in heterotrophic microorganism due to increased human interferences in downstream of the rivers was observed through carbon spectrum utilization studies. The insights obtained through this comprehensive study are in terms of microbial diversity, contaminations, genome plasticity, metal resistance and the overall metabolic diversity that will help in making strategy for river cleanup and protection of human and environmental health. The holy rivers Bhagirathi and Alaknanda confluence at Devprayag (Uttarakhand) thereafter called river Ganga. Comparative microbiome analysis of Ganga revealed that, the Alaknanda river has the foremost contribution in Ganga with respect to bacteriophage, microbial diversity, physiochemical parameters, trace elements, and heavy metals. To the best of our knowledge this is the first study that reveals the rapid and dynamic response of major river water microbiome in response to human activities and pollution load.

## 5. Methods

### 5.1. Sampling sites

The sampling sites of each river was decided on the basis of covering both upper region and downstream stretches which are usually anthropogenically influenced. To cover the aforementioned stretches, seven sites of Ganga, five of Narmada, four of Cauvery and three of Gomti rivers were selected along a gradient of pollution, from the pristine source of the river to the heavily polluted downstream site located in the densely populated urban areas (Fig. 1). The details of sampling sites and their GPS coordinates are provided in Table 1.

### 5.2. Sample collection and processing

The water sample from each location was collected aseptically by using airtight 10 L sterile container during the season of December 2014 to March 2015. The samples were stored at 4°C for metagenomic studies and bacterial analysis, while a portion of the sample was subjected to physiochemical analysis.

### 5.3. Physicochemical characterization of water samples

Physicochemical properties of water sample such as pH, temperature, dissolved oxygen, turbidity, and conductivity were measured on-site using a multi-parameter water quality meters (Horiba U-50, Kyoto, Japan). Other parameters, viz. total dissolved solids, nitrate, phosphate, chloride, fluorides, sulphates, total alkalinity, total hardness, calcium hardness, magnesium hardness, and silica were tested in water testing division of our institute as per APHA, standard protocols [61].

### 5.4. Metagenomic DNA extraction

For metagenomic DNA extraction, water samples were filtered using the Millipore filtration apparatus (Merck Millipore, USA) through a 0.45 µm filter membrane to trap sufficient microorganism for total DNA isolation. DNA was extracted using a Meta-G-Nome DNA isolation kit (Epicenter Biotechnologies, Madison, WI, USA) following the manufacturer's instruction. The DNA quality was monitored on 1% agarose gel on electrophoresis. The DNA purity and concentration was assessed by measuring a ratio of absorbance at 280/230 nm and 260/280 nm using a NanoDrop 1000 (Thermo Scientific Inc. Wilmington, DE, U.S.A.). Using the extracted DNA High-Throughput Sequencing (HTS) was performed on Illumina HiSeq 2500 platform. Rest of the extracted DNA samples was stored at -80 °C until further analysis.

### 5.5. Library preparation and sequencing

TruSeqDNAPCR-Free Library Preparation kit (Illumina Inc., San Diego, California, USA) was used to prepare the Illumina shotgun library as per the manufacturer instructions. In brief, approximately 200 nucleotide DNA fragments were generated from an equal concentration of DNA collected from each sample, followed by end repairing using a T4 DNA polymerase, Klenow fragment, and T4 Polynucleotide Kinase, sequentially. Furthermore, the adapters were ligated, and the DNA library was prepared at the time of PCR. The qualified libraries were further taken for sequencing using the Illumina HiSeq 2500 platform (Illumina Inc., San Diego, California, USA). Minimum 25 million reads were generated to ensuring high depth. The sequencing was outsourced to Bionivid Technology Pvt Ltd, Bangalore, India [62].

### 5.6. Sequencing data quality control

Estimation and elimination of high base call error reads were done using NGS QC Toolkit v2.3.3 with default parameters [63]. During quality control, the reads containing adapter sequences were discarded. Furthermore, the reads were merged to obtain a single fragment, and dereplication was performed using Vsearch [64].

### 5.7. Whole environmental microbiome analysis

Genome-wide alignment of the dereplicated sequence generated after QC was performed using DIAMOND against the NCBI non-redundant protein database [65]. The output was then imported to MEGAN6 [66]. Homology binning (includes quantification) was done in a combination of Megan's inclusive LCA assignments and conservative NCBI blastx assignments. Sequences were assigned to a bin in the case of passing both the above methods at all taxonomic levels (kingdom, phylum, class, order, family, genus and species).

## 5.8. Bacteriophage isolation

The double agar overlay method was employed for the isolation of bacteriophage from water samples. Filtered water (1 mL) was added to top agar (0.7%) which contain 100 µL of overnight grown targeted bacteria. The mixture was then poured on to the agar plates and the plates were analysed for the formation of lytic zones. The plaque forming units (pfu) were counted for enumeration. Samples in which no lytic zones appeared were further analysed using the method of [67] with slight modifications. A hundred millilitres of the water sample was added to 100 mL of double strength TSB. The medium was then inoculated with overnight grown bacteria of interest and incubated at 30°C. After 12 h of incubation, the cell free supernatant was filtered with 0.45µm filters (Merck Millipore, Billerica, MA), followed by spotting on plates containing a lawn of bacteria of interest. The plates were then incubated at 37°C for 24 h for the development of lytic zones.

## 5.9. Biosystem enrichment and depletion analysis of environments

The functional analysis of aligned sequences was performed using MEGAN6. Gene function, subsystem classification, and identification were performed using SEED Genome Databases analysis, KEGG (Kyoto Encyclopedia of Genes and Genomes) [68] was used for pathway elucidation with EC numbers of enriched enzymes. A Classification of Mobile Genetic Elements (ACLAME) [69] and Antibacterial Biocide and Metal Resistance Genes Database (BacMet) were used to identify mobile genetic elements and metal resistance gene, respectively [70].

## 5.10. Analysis of microbial diversity using carbon source utilization pattern

To explore the functional diversity of microbial communities, the carbon utilization pattern was checked by using BIOLOG. The carbon source utilization pattern of microorganisms in all the river water samples was determined by using MT and Biolog Eco plates (Biolog, Inc., Hayward, CA, USA). From the water samples diluted with saline (1:100), 150 µL was inoculated in each well of MT and Biolog Eco plates followed by incubation at 28±2 °C. A redox indicator dye, tetrazolium changes color from colorless to purple, which indicate the substrate utilization. Average Well Colour Development (AWCD) was determined as an indicator of microbial activity by measuring at 590 nm for 7 days [71]. Evenness and diversity indexes were calculated as per [2], and Principal component analysis (PCA) was performed as described by [72].

## 5.11. Quantification of trace and toxic elements

The toxic elements (viz., Cd, Pb, Cr, and As) and trace mineral nutrients (viz., Mn, Zn, Fe, Co, Cu, Se) were analysed using Inductively Coupled Plasma Mass spectrometer (ICP-MS 7500ex, Agilent Technologies, Japan). Internal standardization was maintained as described previously by Dwivedi et al. [73]. The instrument was calibrated using multi-element calibration standard 3 (8500-6948) and 2A (8500-6940) from Agilent Technologies, USA. The analytical accuracy and precision of the ICP-MS was maintained as per the requirements of the National Accreditation Board for Testing and Calibration Laboratories (NABL) accredited lab (certificate no. T-1381). Repeated analysis of spiked river water samples (n=5) were performed for each analytical batch for calibration and quality assurance. More than 98% recovery of As, Mn, Fe, Cr, Co, Zn, Cu, Cd, Pb, and Se was recorded with a detection limit of 1 µg/L for each sample.

## 5.12. Statistical analysis

Statistical analysis of microbiome data was performed using Microbiome Analyst, an online tool for comprehensive statistical, visual, and meta-analysis of microbiome data at different taxonomic level [25, 74]. To determine the effect of a physicochemical factor on the bacterial communities, physicochemical data were statistically correlated with diversity and validated by performing the principal component analysis (PCA) and regression with respect to each environmental factor to the bacterial community through MicrobiomeAnalyst. Alpha diversity (observed, chao1, and Shannon) and Beta diversity analysis by calculating Bray-Curtis distances method show at operational taxonomic unit (OTU) level with nonmetric multidimensional scaling (NMDS) and PCA plots perform to show differences between groups based on the bacterial community composition.

## 6. Additional Files

Additional file 1: Supplementary figure S1 to S10.

Additional file 2: Table S1 to S4.

Additional file 3: - River Cauvery samples microbial diversity Krona chart.

Additional file 4: - River Ganga samples microbial diversity Krona chart.

Additional file 5: - River Gomti samples microbial diversity Krona chart.

## Abbreviations

KRS Dam: Krishna Raja Sagar Dam; ARGs: Antibiotic resistance genes; MRGs: metal resistant genes; OTU: Operational taxonomic unit; RPM: Reads per million mapped reads; PCoA and NMDS; ANOVA; PCA: Principal component analysis; QC: Quality control; NCBI: National Center for Biotechnology Information; LCA: lowest common ancestor; KEGG: Kyoto Encyclopedia of Genes and Genomes; ACLAME: A Classification of Mobile Genetic Elements; BacMet: Antibacterial Biocide and Metal Resistance Genes Database; AWCD: Average Well Colour Development.

## Declarations

### Acknowledgements

The work described in this paper was supported by Institutional project grants from the Council of Scientific & Industrial Research-National Botanical Research Institute (CSIR-NBRI) and CSIR-Indian Institute of Toxicology Research (CSIR-IITR). CSN is grateful to Prof. Murli Manohar Joshi, Former Union Minister of Science and Technology, Government of India for several useful discussions during the conceptualization of this project. CSN is grateful to the Science and Engineering Research Board, Department of Science & Technology, Govt. of India, New Delhi, for the award of J. C. Bose Fellowship. We thank Prof. Saroj Barik, Director, CSIR-NBRI and Prof. Alok Dhawan, Director, CSIR-IITR for extending their valuable support for this project.

### Author contributions

The authors contributions are as follows: **RKR**: Sample collection from Cauvery and downstream of Ganga river, Execution of experiments, Metagenomic study, Data curation, Statistical analysis, First draft preparation. **VKG**: Sample collection from Cauvery, Gomti and downstream of Ganga river and First draft preparation. **MK**: Sample collection from Ganga and Cauvery rivers, bacteriophage study and writing of this part of MS. **SP & AM**: Sample collection from Narmada river, Microbial carbon utilization profiling using Biolog and writing of this part of MS. **SD**: Sample collection from Ganga, Narmada and Gomti rivers, Metal analysis and writing of this part of MS. **SM**: Metal analysis and writing of this part of MS, critical revision of MS and intellectual inputs. **NM**: Sample collection from Ganga and Cauvery rivers, planning of experiments, project implementation, supervision, writing and editing of the manuscript. **CSN**: Conceptualization, funding of project, administration, supervision, editing and final approval of the MS.

### Funding

This work was funded by CSIR-Indian Institute of Toxicology Research and CSIR-National Botanical Research Institute, Lucknow, India.

### Availability of data and materials

All raw sequence datasets related to this study have been made available at the NCBI Sequence Read Archive (SRA) database (SRR12274735- SRR12274750) under the bio-project accession number PRJNA647427

### Declaration of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Ethics approval and consent to participate

Not applicable

### Consent for publication

Not applicable

### Competing interests

The authors declare no competing financial interests.

### Author details

<sup>1</sup>Environmental Biotechnology Laboratory, CSIR-Indian Institute of Toxicology Research, Vishvigyan Bhavan, 31 Mahatma Gandhi Marg, Lucknow, Uttar Pradesh, 226001, India.

<sup>2</sup>CSIR-National Botanical Research Institute, Rana Pratap Marg, Lucknow, Uttar Pradesh 226 001, India.

<sup>3</sup>Department of Biochemistry, School of Dental Sciences, Babu Banarsi Das University, Lucknow, Uttar Pradesh, India.

<sup>4</sup>Swami Rama Himalayan University, Jolly Grant, Dehradun, 248016, India

<sup>5</sup>Present address: Department of Biochemistry, School of Chemical & Life Sciences, Jamia Hamdard, New Delhi 110062, India.

<sup>6</sup>Present address: Deen Dayal Upadhyaya Gorakhpur University, Gorakhpur-273009, India

## References

1. 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. *Environmental Monitoring and Assessment*. 2020;192(4):1-15.
2. Nautiyal CS. Self-purificatory Ganga water facilitates death of pathogenic *Escherichia coli* O157: H7. *Current microbiology*. 2009;58(1):25-9.
3. Jani K, Dhotre D, Bandal J, Shouche Y, Suryavanshi M, Rale V, et al. World's largest mass bathing event influences the bacterial communities of Godavari, a holy river of India. *Microbial ecology*. 2018;76(3):706-18.
4. Satinsky BM, Fortunato CS, Doherty M, Smith CB, Sharma S, Ward ND, et al. Metagenomic and metatranscriptomic inventories of the lower Amazon River, May 2011. *Microbiome*. 2015;3(1):39.
5. Satinsky BM, Zielinski BL, Doherty M, Smith CB, Sharma S, Paul JH, et al. The Amazon continuum dataset: quantitative metagenomic and metatranscriptomic inventories of the Amazon River plume, June 2010. *Microbiome*. 2014;2(1):17.
6. Stuart M, Lapworth D, Crane E, Hart A. Review of risk from potential emerging contaminants in UK groundwater. *Science of the Total Environment*. 2012;416:1-21.
7. Vasquez MI, Lambrianides A, Schneider M, Kümmerer K, Fatta-Kassinos D. Environmental side effects of pharmaceutical cocktails: what we know and what we should know. *Journal of Hazardous Materials*. 2014;279:169-89.
8. Dwivedi S, Mishra S, Tripathi RD. Ganga water pollution: a potential health threat to inhabitants of Ganga basin. *Environment international*. 2018;117:327-38.
9. Radović T, Grujić S, Petković A, Dimkić M, Laušević M. Determination of pharmaceuticals and pesticides in river sediments and corresponding surface and ground water in the Danube River and tributaries in Serbia. *Environmental Monitoring and Assessment*. 2015;187(1):4092.
10. Zhang C, Nie S, Liang J, Zeng G, Wu H, Hua S, et al. Effects of heavy metals and soil physicochemical properties on wetland soil microbial biomass and bacterial community structure. *Science of the Total Environment*. 2016;557:785-90.
11. Xu Y, Guo C, Luo Y, Lv J, Zhang Y, Lin H, et al. Occurrence and distribution of antibiotics, antibiotic resistance genes in the urban rivers in Beijing, China. *Environmental pollution*. 2016;213:833-40.
12. Coutinho FH, Silveira CB, Pinto LH, Salloto GR, Cardoso AM, Martins OB, et al. Antibiotic resistance is widespread in urban aquatic environments of Rio de Janeiro, Brazil. *Microbial ecology*. 2014;68(3):441-52.
13. Marathe NP, Berglund F, Razavi M, Pal C, Dröge J, Samant S, et al. Sewage effluent from an Indian hospital harbors novel carbapenemases and integron-borne antibiotic resistance genes. *Microbiome*. 2019;7(1):97.
14. Zhou Z-C, Zheng J, Wei Y-Y, Chen T, Dahlgren RA, Shang X, et al. Antibiotic resistance genes in an urban river as impacted by bacterial community and physicochemical parameters. *Environmental Science and Pollution Research*. 2017;24(30):23753-62.
15. Li B, Yang Y, Ma L, Ju F, Guo F, Tiedje JM, et al. Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. *The ISME journal*. 2015;9(11):2490-502.
16. Nakayama T, Hoa TTT, Harada K, Warisaya M, Asayama M, Hinenoya A, et al. Water metagenomic analysis reveals low bacterial diversity and the presence of antimicrobial residues and resistance genes in a river containing wastewater from backyard aquacultures in the Mekong Delta, Vietnam. *Environmental Pollution*. 2017;222:294-306.
17. Mittal P, PK VP, Dhakan DB, Kumar S, Sharma VK. Metagenome of a polluted river reveals a reservoir of metabolic and antibiotic resistance genes. *Environmental Microbiome*. 2019;14(1):5.
18. Gupta S, Arango-Argoty G, Zhang L, Pruden A, Vikesland P. Identification of discriminatory antibiotic resistance genes among environmental resistomes using extremely randomized tree algorithm. *Microbiome*. 2019;7(1):123.
19. Liu T, Zhang AN, Wang J, Liu S, Jiang X, Dang C, et al. Integrated biogeography of planktonic and sedimentary bacterial communities in the Yangtze River. *Microbiome*. 2018;6(1):16.
20. Tang X, Xie G, Shao K, Hu Y, Cai J, Bai C, et al. Contrast diversity patterns and processes of microbial community assembly in a river-lake continuum across a catchment scale in northwestern China. *Environmental Microbiome*. 2020;15:1-17.
21. Jiang B, Jin N, Xing Y, Su Y, Zhang D. Unraveling uncultivable pesticide degraders via stable isotope probing (SIP). *Critical reviews in biotechnology*. 2018;38(7):1025-48.
22. Vadde KK, Feng Q, Wang J, McCarthy AJ, Sekar R. Next-generation sequencing reveals fecal contamination and potentially pathogenic bacteria in a major inflow river of Taihu Lake. *Environmental Pollution*. 2019;254:113108.
23. da C Jesus E, Marsh TL, Tiedje JM, de S Moreira FM. Changes in land use alter the structure of bacterial communities in Western Amazon soils. *The ISME journal*. 2009;3(9):1004-11.
24. Navarrete AA, Kuramae EE, de Hollander M, Pijl AS, van Veen JA, Tsai SM. Acidobacterial community responses to agricultural management of soybean in Amazon forest soils. *FEMS Microbiology Ecology*. 2013;83(3):607-21.
25. Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J. *MicrobiomeAnalyst*: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic acids research*. 2017;45(W1):W180-W8.

26. Zhang Y, LeJeune JT. Transduction of bla<sub>CMY-2</sub>, tet (A), and tet (B) from *Salmonella enterica* subspecies *enterica* serovar Heidelberg to *S. Typhimurium*. *Veterinary microbiology*. 2008;129(3-4):418-25.
27. Lood R, Ertürk G, Mattiasson B. Revisiting antibiotic resistance spreading in wastewater treatment plants–bacteriophages as a much neglected potential transmission vehicle. *Frontiers in microbiology*. 2017;8:2298.
28. Chen J, Novick RP. Phage-mediated intergeneric transfer of toxin genes. *science*. 2009;323(5910):139-41.
29. Moon K, Jeon JH, Kang I, Park KS, Lee K, Cha C-J, et al. Freshwater viral metagenome reveals novel and functional phage-borne antibiotic resistance genes. *Microbiome*. 2020;8(1):1-15.
30. Scarr S, Cai W, Kumar V, Pal A. The race to save the river Ganges. <https://graphics.reuters.com/INDIA-GANGES/010081VQ3BT/index.html> (2019). Accessed JANUARY 18 2019.
31. Yang K, LeJeune J, Alsdorf D, Lu B, Shum C, Liang S. Global distribution of outbreaks of water-associated infectious diseases. *PLoS Negl Trop Dis*. 2012;6(2):e1483.
32. Chauhan A, Goyal P, Varma A, Jindal T. Microbiological evaluation of drinking water sold by roadside vendors of Delhi, India. *Applied Water Science*. 2017;7(4):1635-44.
33. Malik A, Singh K, Mohan D, Patel D. Distribution of polycyclic aromatic hydrocarbons in Gomti river system, India. *Bulletin of environmental contamination and toxicology*. 2004;72(6):1211-8.
34. Malik A, Verma P, Singh AK, Singh KP. Distribution of polycyclic aromatic hydrocarbons in water and bed sediments of the Gomti River, India. *Environmental monitoring and assessment*. 2011;172(1-4):529-45.
35. Chen J, Wang P, Wang C, Wang X, Miao L, Liu S, et al. Bacterial communities in riparian sediments: a large-scale longitudinal distribution pattern and response to dam construction. *Frontiers in microbiology*. 2018;9:999.
36. Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. A guide to the natural history of freshwater lake bacteria. *Microbiology and molecular biology reviews*. 2011;75(1):14-49.
37. Staley C, Gould TJ, Wang P, Phillips J, Cotner JB, Sadowsky MJ. Species sorting and seasonal dynamics primarily shape bacterial communities in the Upper Mississippi River. *Science of the Total Environment*. 2015;505:435-45.
38. Kou W, Zhang J, Lu X, Ma Y, Mou X, Wu L. Identification of bacterial communities in sediments of Poyang Lake, the largest freshwater lake in China. *SpringerPlus*. 2016;5(1):401.
39. Pang CM, Liu W-T. Biological filtration limits carbon availability and affects downstream biofilm formation and community structure. *Applied and environmental microbiology*. 2006;72(9):5702-12.
40. Aburto-Medina A, Adetutu EM, Aleer S, Weber J, Patil SS, Sheppard PJ, et al. Comparison of indigenous and exogenous microbial populations during slurry phase biodegradation of long-term hydrocarbon-contaminated soil. *Biodegradation*. 2012;23(6):813-22.
41. Wakelin SA, Colloff MJ, Kookana RS. Effect of wastewater treatment plant effluent on microbial function and community structure in the sediment of a freshwater stream with variable seasonal flow. *Applied and environmental microbiology*. 2008;74(9):2659-68.
42. Valls M, De Lorenzo V. Exploiting the genetic and biochemical capacities of bacteria for the remediation of heavy metal pollution. *FEMS microbiology Reviews*. 2002;26(4):327-38.
43. Aktar MW, Paramasivam M, Ganguly M, Purkait S, Sengupta D. Assessment and occurrence of various heavy metals in surface water of Ganga river around Kolkata: a study for toxicity and ecological impact. *Environmental monitoring and assessment*. 2010;160(1-4):207-13.
44. Purkait S, Ganguly M, Aktar MW, Sengupta D, Chowdhury A. Impact assessment of various parameters polluting Ganga water in Kolkata Region: a study for quality evaluation and environmental implication. *Environmental monitoring and assessment*. 2009;155(1-4):443-54.
45. Kamitani T, Oba H, Kaneko N. Microbial biomass and tolerance of microbial community on an aged heavy metal polluted floodplain in Japan. *Water, air, and soil pollution*. 2006;172(1-4):185-200.
46. Chaudhry V, Rehman A, Mishra A, Chauhan PS, Nautiyal CS. Changes in bacterial community structure of agricultural land due to long-term organic and chemical amendments. *Microbial ecology*. 2012;64(2):450-60.
47. Zhang H-H, Huang T-L, Chen S, Guo L, Yang X, Liu T. Spatial pattern of bacterial community functional diversity in a drinking water reservoir, Shaanxi Province, Northwest China. *J Pure Appl Microb*. 2013;3:1647-54.
48. Gordon-Bradley N, Li N, Williams H. Bacterial community structure in freshwater springs infested with the invasive plant species *Hydrilla verticillata*. *Hydrobiologia*. 2015;742(1):221-32.
49. Zhang H-H, Chen S-N, Huang T-L, Ma W-X, Xu J-L, Sun X. Vertical distribution of bacterial community diversity and water quality during the reservoir thermal stratification. *International Journal of Environmental Research and Public Health*. 2015;12(6):6933-45.
50. Paul D, Sinha SN. Isolation and characterization of a phosphate solubilizing heavy metal tolerant bacterium from River Ganga, West Bengal, India. *Songklanakarinn Journal of Science and Technology*. 2015;37(6):651-7.
51. Oest A, Alsaffar A, Fenner M, Azzopardi D, Tiquia-Arashi SM. Patterns of change in metabolic capabilities of sediment microbial communities in river and lake ecosystems. *International Journal of Microbiology*. 2018;2018.
52. Chakrapani G, Saini R, Yadav S. Chemical weathering rates in the Alaknanda–Bhagirathi river basins in Himalayas, India. *Journal of Asian Earth Sciences*. 2009;34(3):347-62.
53. Paul D, Sinha SN. Assessment of various heavy metals in surface water of polluted sites in the lower stretch of river Ganga, West Bengal: a study for ecological impact. *Discovery Nature*. 2013;6(14):8-13.

54. Rasmussen LD, Sørensen SJ. Effects of mercury contamination on the culturable heterotrophic, functional and genetic diversity of the bacterial community in soil. *FEMS microbiology ecology*. 2001;36(1):1-9.
55. Sa'idi M. Experimental studies on effect of heavy metals presence in industrial wastewater on biological treatment. *International journal of environmental sciences*. 2010;1(4):666-76.
56. Dutta S, Kole R, Ghosh S, Nath D, Vass K. Impact assessment of lead on water quality of river Ganga in West Bengal, India. *Bulletin of environmental contamination and toxicology*. 2005;75(5):1012-9.
57. Sinha SN, Biswas M, Paul D, Rahaman S. Biodegradation potential of bacterial isolates from tannery effluent with special reference to hexavalent chromium. *Biotechnology Bioinformatics and Bioengineering*. 2011;1(3):381-6.
58. Rai P, Mishra A, Tripathi B. Heavy metal and microbial pollution of the River Ganga: A case study of water quality at Varanasi. *Aquatic Ecosystem Health & Management*. 2010;13(4):352-61.
59. Katiyar S. Impact of tannery effluent with special reference to seasonal variation on physico-chemical characteristics of river water at Kanpur (UP), India. *J Environ Anal Toxicol*. 2011;1(117):4.
60. Bhatnagar M, Singh R, Gupta S, Bhatnagar P. Study of tannery effluents and its effects on sediments of river Ganga in special reference to heavy metals at Jajmau, Kanpur, India. *Journal of Environmental Research and Development*. 2013;8(1):56-9.
61. Federation WE, Association APH. Standard methods for the examination of water and wastewater. American Public Health Association (APHA): Washington, DC, USA. 2005.
62. Regar RK, Gaur VK, Bajaj A, Tambat S, Manickam N. Comparative microbiome analysis of two different long-term pesticide contaminated soils revealed the anthropogenic influence on functional potential of microbial communities. *Science of the Total Environment*. 2019;681:413-23.
63. Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PloS one*. 2012;7(2):e30619.
64. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ*. 2016;4:e2584.
65. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nature methods*. 2015;12(1):59-60.
66. Huson DH, Beier S, Flade I, Górská A, El-Hadidi M, Mitra S, et al. MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS computational biology*. 2016;12(6):e1004957.
67. Carvalho C, Susano M, Fernandes E, Santos S, Gannon B, Nicolau A, et al. Method for bacteriophage isolation against target *Campylobacter* strains. *Letters in applied microbiology*. 2010;50(2):192-7.
68. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*. 2000;28(1):27-30.
69. Leplae R, Hebrant A, Wodak SJ, Toussaint A. ACLAME: a CLAssification of Mobile genetic Elements. *Nucleic acids research*. 2004;32(suppl\_1):D45-D9.
70. Pal C, Bengtsson-Palme J, Rensing C, Kristiansson E, Larsson DJ. BacMet: antibacterial biocide and metal resistance genes database. *Nucleic acids research*. 2014;42(D1):D737-D43.
71. Garland JL. Patterns of potential C source utilization by rhizosphere communities. *Soil Biology and Biochemistry*. 1996;28(2):223-30.
72. Garland JL, Mills AL. Classification and characterization of heterotrophic microbial communities on the basis of patterns of community-level sole-carbon-source utilization. *Applied and environmental microbiology*. 1991;57(8):2351-9.
73. Dwivedi S, Tripathi R, Srivastava S, Singh R, Kumar A, Tripathi P, et al. Arsenic affects mineral nutrients in grains of various Indian rice (*Oryza sativa* L.) genotypes grown on arsenic-contaminated soils of West Bengal. *Protoplasma*. 2010;245(1-4):113-24.
74. Chong J, Liu P, Zhou G, Xia J. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. *Nature Protocols*. 2020;15(3):799-821.

## Tables

**Table 1:** Details of total reads of Bacteria, Archaea and Eukaryota and their diversity index

Name of River	Sampling sites	Sample code	Latitude	Longitude	Reads (in million)			Studied domains			Richness	Ever
					Total no. of reads	HQ reads	HQ reads %	Bacteria	Archaea	Eukaryota	Chao 1	Shai inde
<b>Ganga</b>	Bhagirathi	G1	30°08'49.2"N	78°35'51.4"E	35.49	32.02	90.22	7412319	98702	92895	1606.5	1.30
	Alaknanda	G2	30°08'43.4"N	78°35'56.1"E	37.64	33.81	89.83	7515144	10776	113439	1643	6.05
	Devprayag	G3	30°08'43.5"N	78°35'51.6"E	32.81	29.31	89.34	7248761	39706	194360	1843	0.08
	Rishikesh	G4	30°07'35.4"N	78°21'13.5"E	29.83	26.75	89.69	7396816	23377	215193	1659	4.66
	Haridwar	G5	29°57'22.5"N	78°10'15.2"E	29.69	26.28	88.53	7168744	20434	221953	1762	4.66
	Bithoor	G6	26°36'45.9"N	80°16'31.0"E	29.57	26.83	90.72	6631699	29910	120020	2175.6	11.6
	Jajmau	G7	26°25'13.3"N	80°25'21.8"E	31.82	28.24	88.75	6268661	17066	354601	1787.857	1.40
<b>Narmada</b>	Udgam	N1	22°40'28.8"N	81°45'28.9"E	29.38	25.68	87.42	5966344	13639	296001	1894	2.89
	Kapildhara	N2	22°42'03.0"N	81°42'19.0"E	30.54	27.13	88.86	6772850	16801	85776	2004	1.58
	Bhedaghat	N3	23°07'44.0"N	79°49'04.0"E	35.99	32.84	91.25	6669261	94566	101200	1922.75	4.21
	Hoshangabad	N4	22°43'13.0"N	77°48'12.0"E	31.39	28.45	90.63	7092344	18546	146660	2203.75	16.7
	Omkareshwar	N5	22°15'01.0"N	76°08'48.0"E	34.33	31.32	91.25	6614525	35570	144861	2083.6	11.6
<b>Cauvery</b>	Talacauvery	C1	12°23'07.4"N	75°29'27.6"E	33.66	29.61	87.98	6575577	4524	154535	1346.5	2.60
	Bhagamandla	C2	12°23'03.8"N	75°31'59.7"E	35.97	32.06	89.12	7119654	35387	73889	2115.75	11.1
	Krishna Raja Sagara (KRS) Dam	C3	12°25'34.8"N	76°34'28.6"E	33.83	30.22	89.32	7146673	22616	184008	1490	0.00
	Erode	C4	11°21'46.6"N	77°44'31.3"E	31.79	28.78	90.51	6879557	21665	177008	2162.333	22.7
<b>Gomti</b>	Neemsar	LK01	27°21.014'N	80°27.615'E	34.01	30.81	90.59	6669081	44411	120650	2203.429	12.6
	Bhatpur	LK02	27°11.323'N	80°48.129'E	29.73	26.99	90.79	6876931	26450	62761	2229	14.7
	Gaughat	LK03	26°53.105'N	80°54.047'E	36.81	33.17	90.13	6461647	58595	118899	2222	23.2

Table: S2. Presence of virulent phages at selected sites of Ganga, Narmada, Cauvery and Gomti

Name of river	Sampling sites	Sample code	Name of virulent phages			
			<i>V. cholera</i>	<i>E. coli</i>	<i>P. aeruginosa</i>	<i>E. faecalis</i>
<b>Ganga</b>	Bhagirathi	G1	NO	YES*	NO	NO
	Alaknanda	G2	NO	NO	NO	NO
	Devprayag	G3	NO	YES*	NO	NO
	Rishikesh	G4	YES*	NO	NO	NO
	Haridwar	G5	YES*	YES*	NO	NO
	Bithoor	G6	YES*	YES*	NO	NO
	Jajmau	G7	YES*	YES*	NO	YES*
<b>Narmada</b>	Udgam	N1	NO	NO	NO	NO
	Kapildhara	N2	NO	YES*	NO	NO
	Bhedaghat	N3	NO	NO	NO	NO
	Hoshangabad	N4	YES*	YES*	NO	NO
<b>Cauvery</b>	Talacauvery	C1	NO	NO	NO	NO
	Bhagamandla	C2	NO	NO	NO	YES*
	Krishna Raja Sagara (KRS) Dam	C3	YES*	NO	NO	YES*
	Erode	C4	NO	NO	NO	NO
<b>Gomti</b>	Neemsar	LK01	YES*	10 <sup>2</sup> (pfu ml <sup>-1</sup> )	NO	YES*
	Bhatpur	LK02	YES*	10 <sup>2</sup> (pfu ml <sup>-1</sup> )	NO	YES*
	Gaughat	LK03	YES*	YES*	NO	NO

\*Phages detected after enrichment.

## Figures

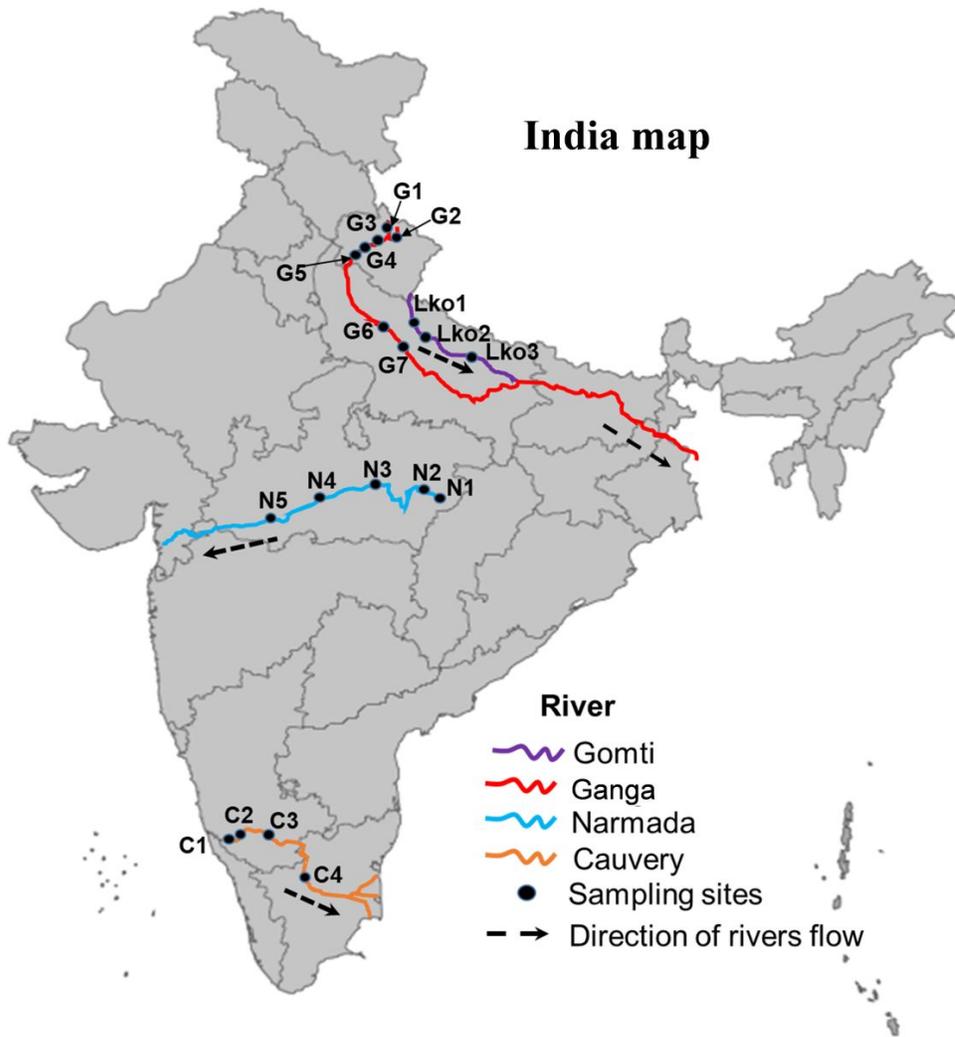
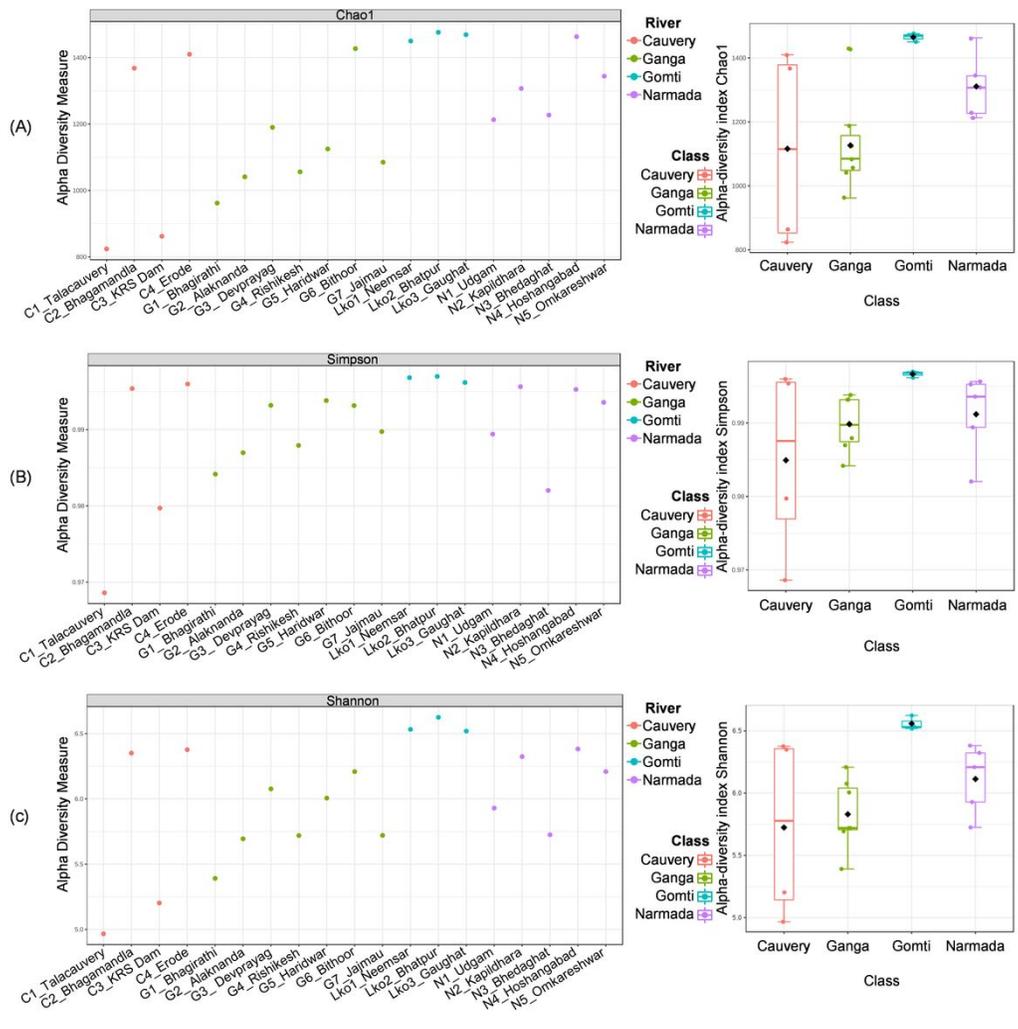
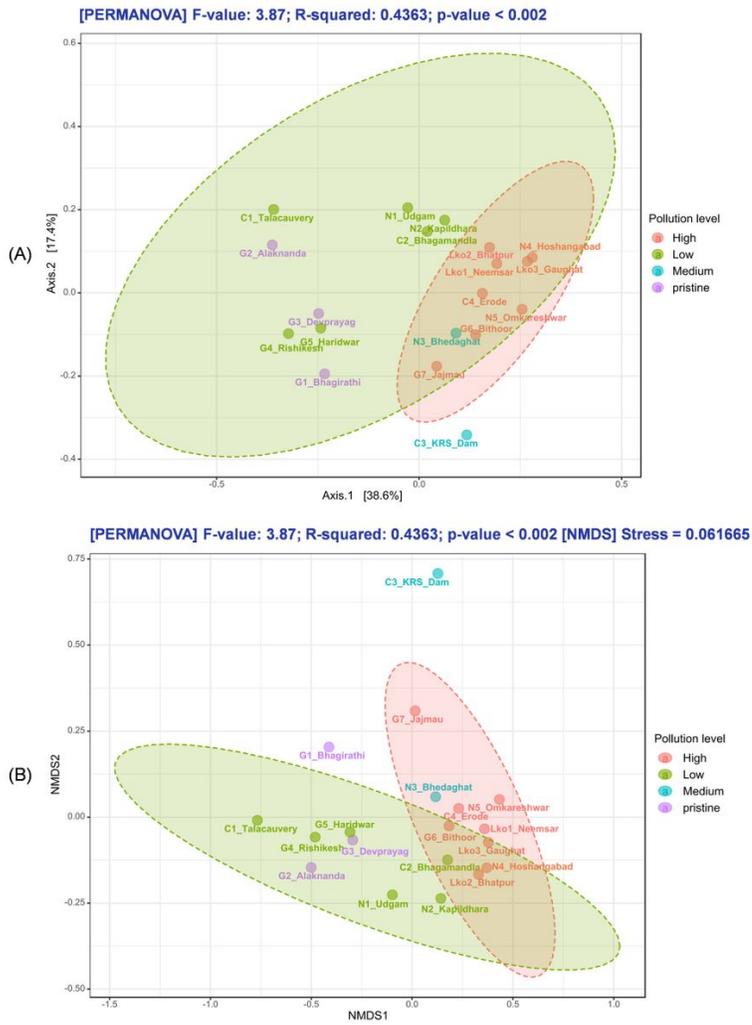


Figure 1

Map of India showing the sample collection locations of four major rivers.

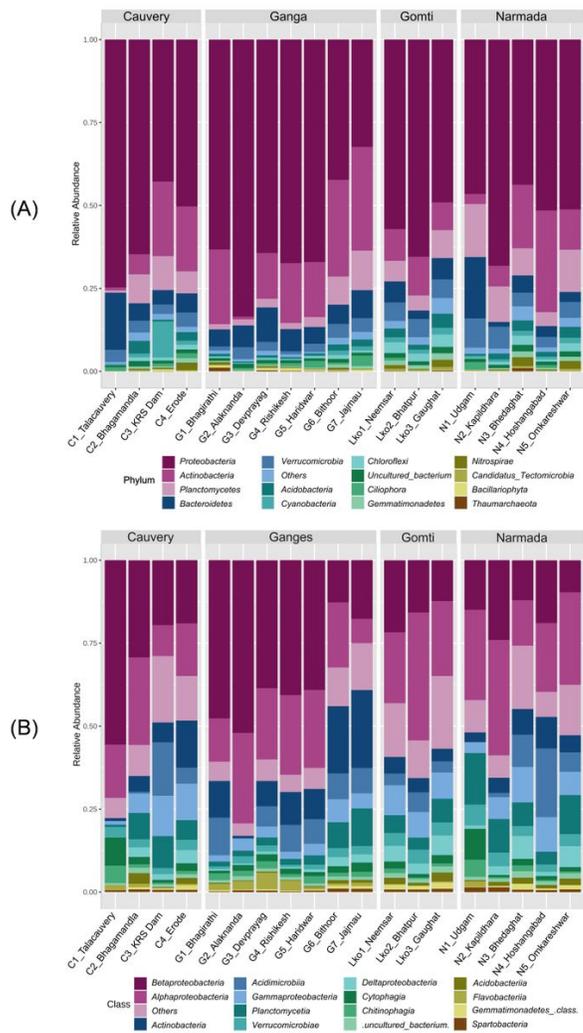


**Figure 2** Alpha diversity measured using (A) Chao1 for richness where (B) Shannon and (C) Simpson index representing evenness at operational taxonomic unit (OTU) level across the samples. Each boxplot represents the diversity distribution present within a river group.

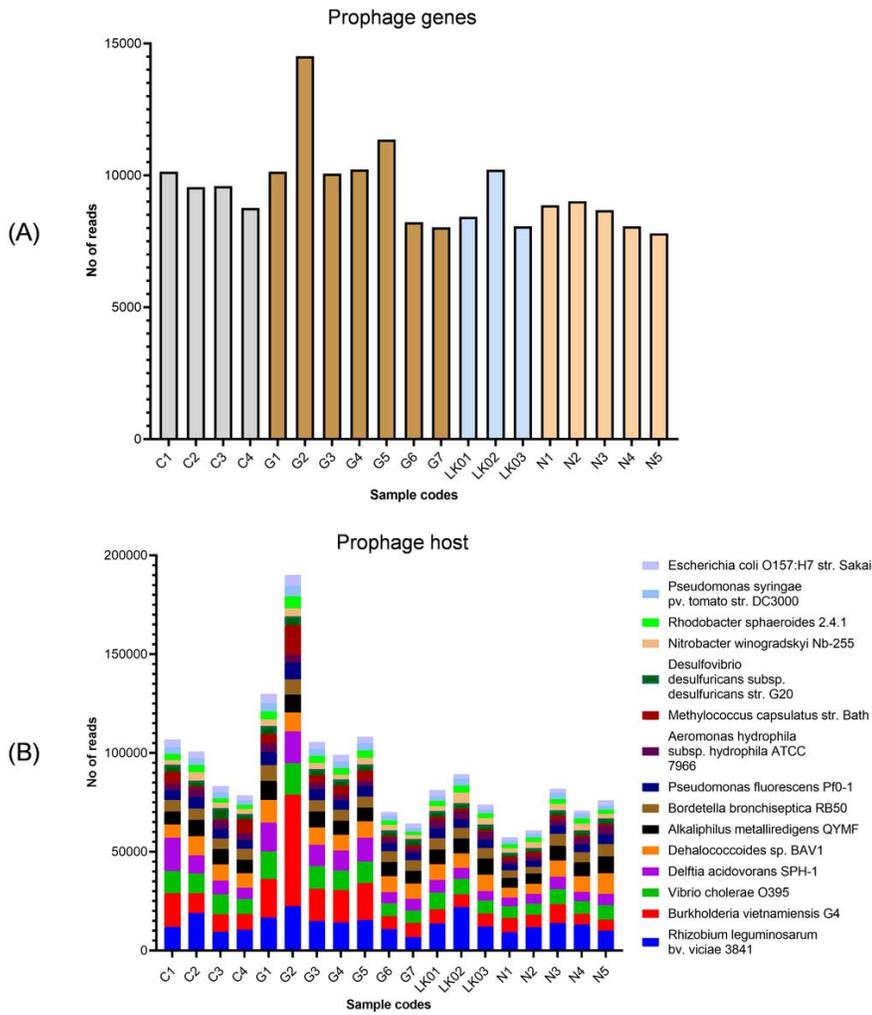


**Figure 3**

Plot showing the Beta diversity analysis and significance testing, (A) PCoA and (B) NMDS plot. Plots depicted the variance among all samples microbial communities (between samples) and grouped based on their similarity.

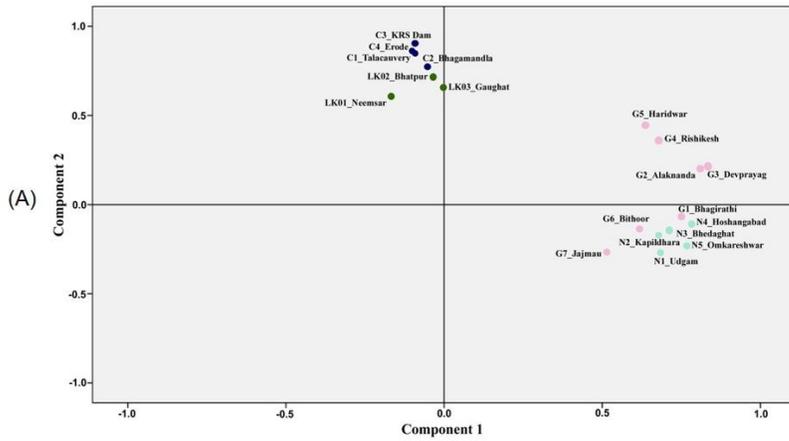


**Figure 4**  
 Bacterial community compositions across the four rivers. (A) Phylum level (B) Class level distribution of top 15 dominant bacterial groups among all four rivers.

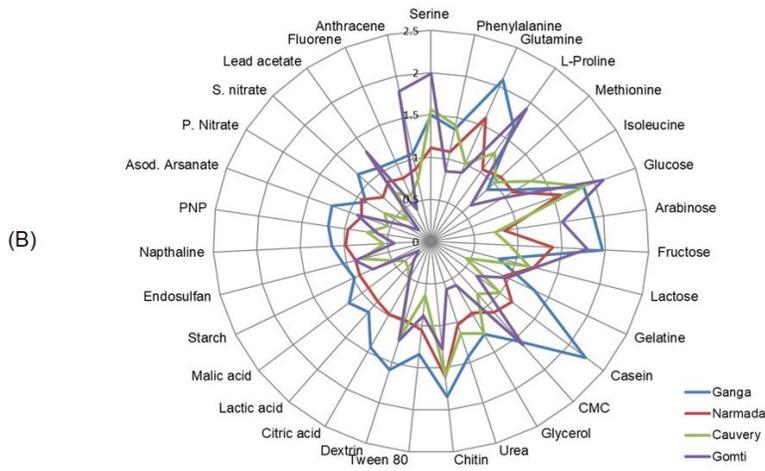


**Figure 5**

Bar plot for (A) Abundance of prophage genes, (B) Bacterial species likely to be associated with prophages.



Substrate specific carbon utilization pattern of all river samples



**Figure 6**  
 (A) Principal component analysis (PCA) of all river water samples based on carbon source utilization pattern. The green dots represent the sample of Cauvery river, blue for Ganga, red for Gomti and indigo for Narmada. (B) Substrate specific carbon utilization pattern of all river samples in MT plate.

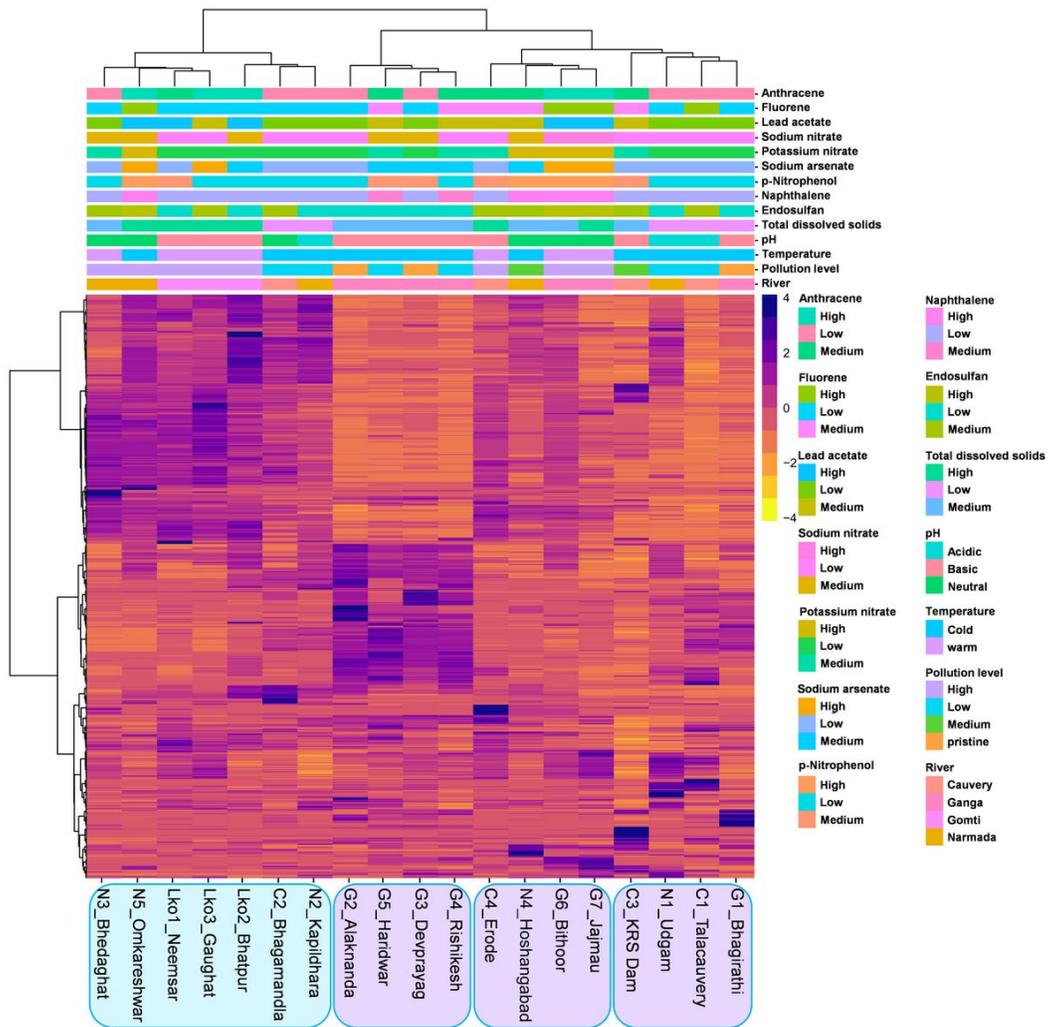


Figure 7

Heatmap representing correlation analysis of carbon utilization pattern with microbial diversity at the species level.

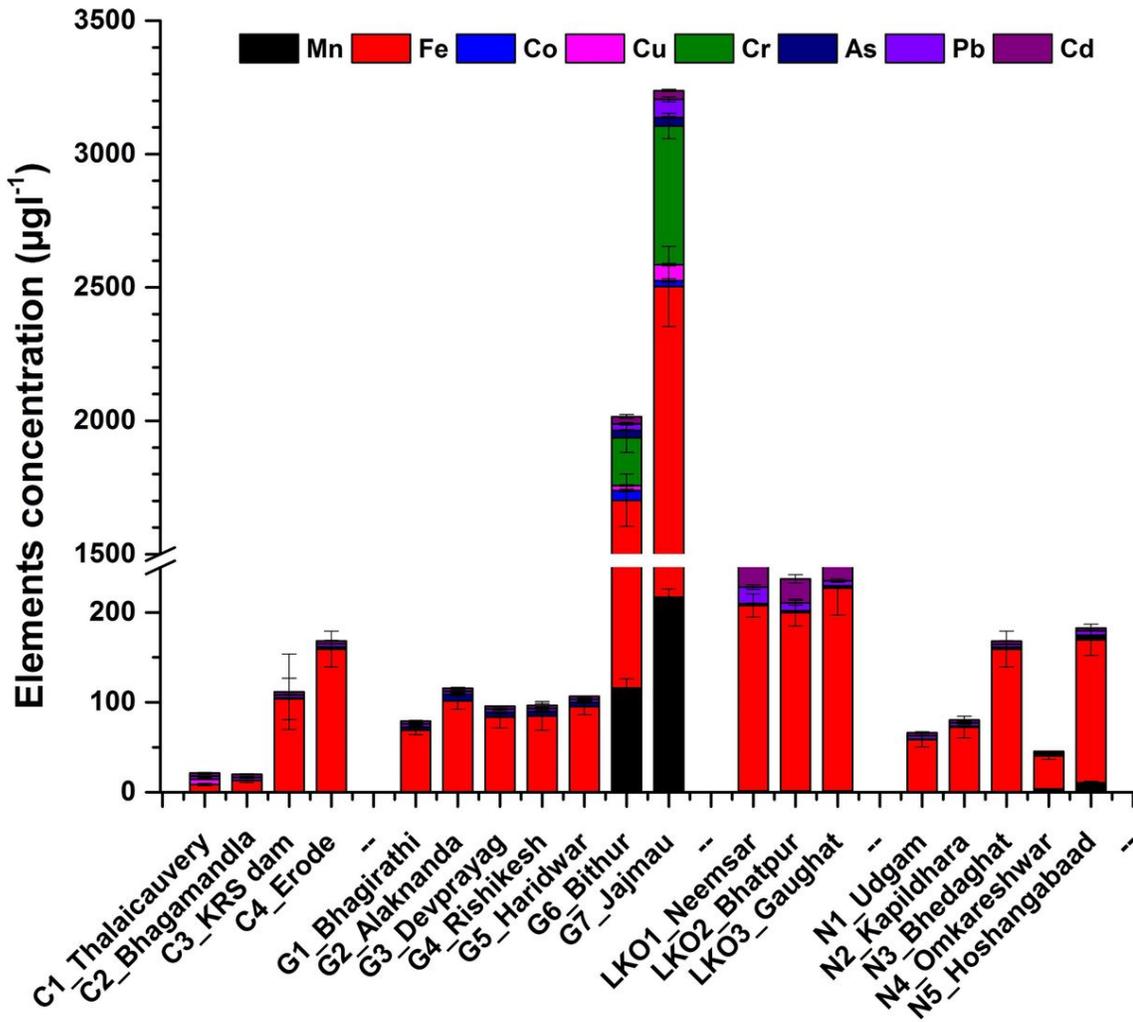
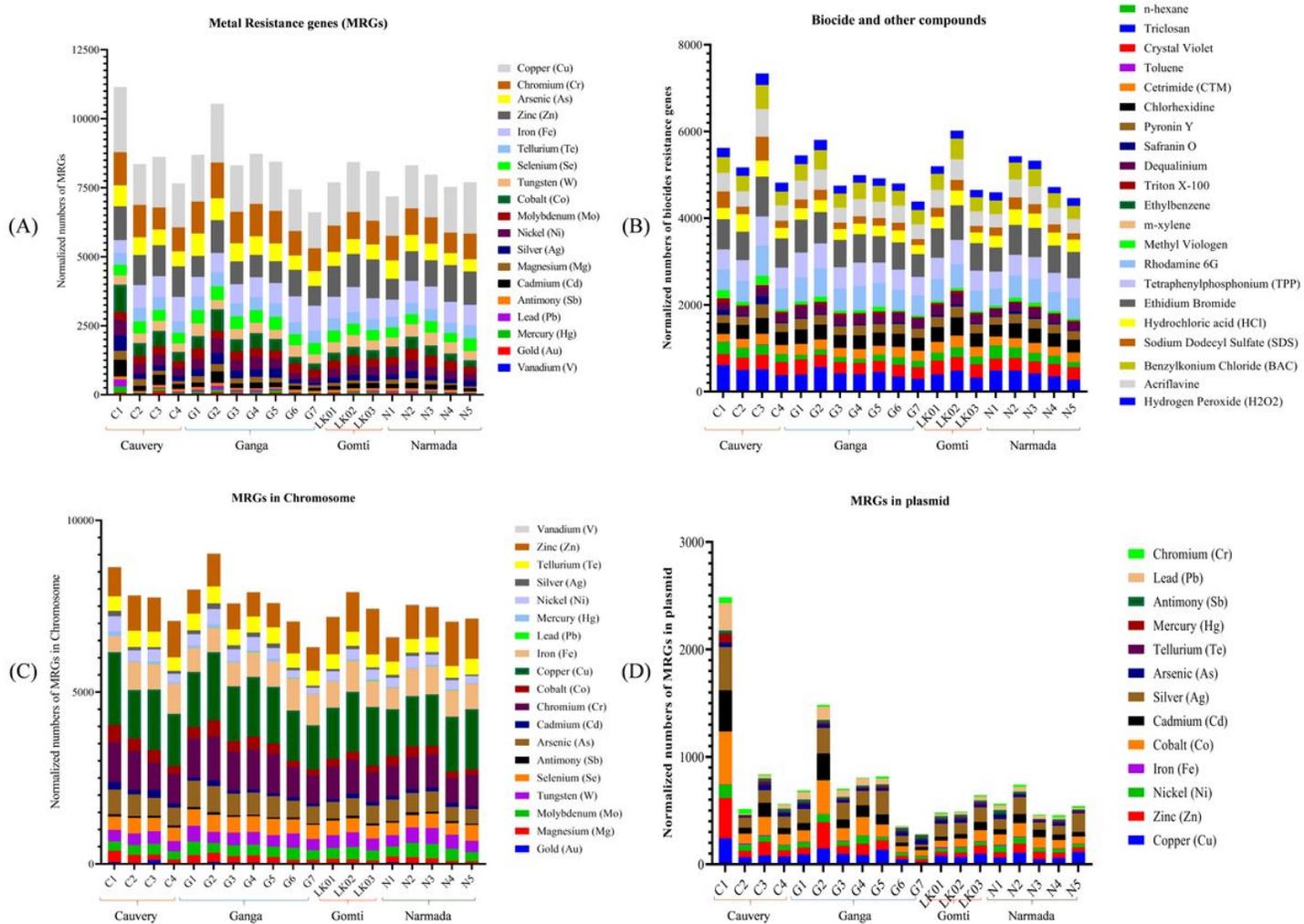


Figure 8

Level of trace and toxic elements at selected sites of Ganga, Cauvery, Narmada and Gomti rivers.



**Figure 9**

Stacked bar plot showing the distribution of metal and biocides resistance genes in all the water samples from four rivers. (A) Metal resistance genes, (B) Biocides resistance genes, (C) Abundance of MRGs located in chromosomes, and (D) Abundance of MRGs located in plasmid. Note, in this representation some of the genes are represented in more than one group due to broad gene specificity.

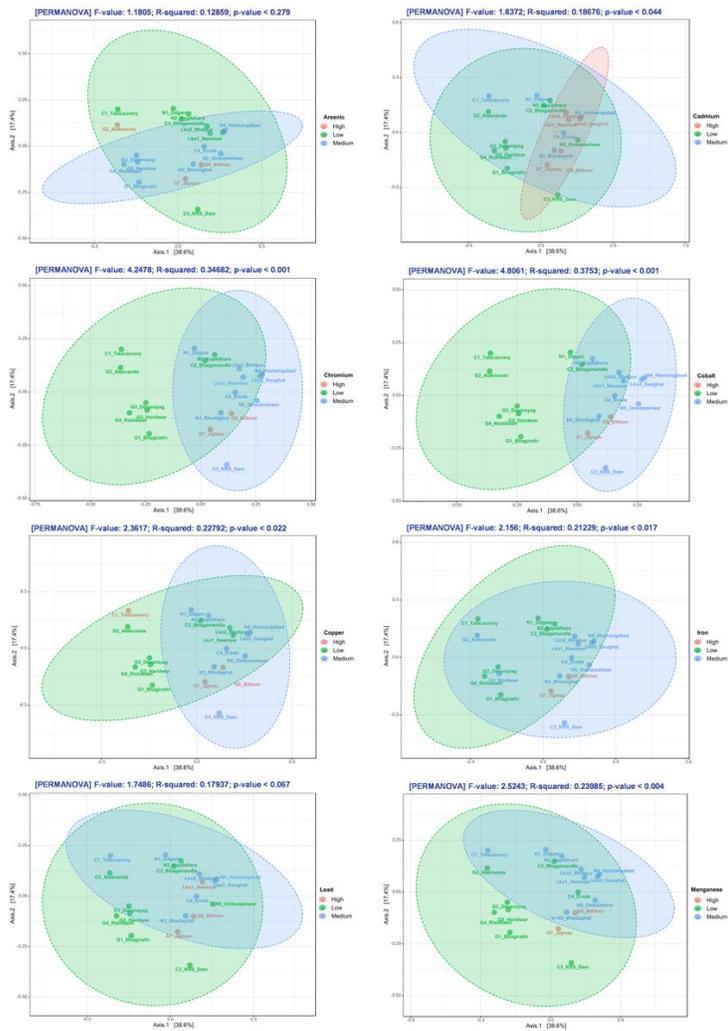


Figure 10

This PCoA plot showed the variation in diversity in reference to detected trace elements and toxic metals concentration. The explained variances are shown in brackets with reference to different metals present in water samples.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1SupplementaryfigureS1toS10.docx](#)
- [Additionalfile2TableS1toS4.docx](#)
- [Additionalfile3CauveryRiverKronachart.html](#)
- [Additionalfile4GangaRiverKronachart.html](#)
- [Additionalfile5GomtiRiverKronachart.html](#)
- [Additionalfile6NarmadaRiverKronachart.html](#)