

# Metagenomic analysis of Dam Reservoirs in Pune city for bacterial fingerprints through BLAST and Kaiju tool

Pritee Chunarkar Patil\*, Darshit Ashokkumar Patel and Vidya Sunil Tale

Department of Bioinformatics, Rajiv Gandhi Institute of IT and Biotechnology,  
Bharati Vidyapeeth (Deemed to be University), Pune, Maharashtra, India -411046.

<http://dx.doi.org/10.13005/bbra/2889>

(Received: 25 October 2020; accepted: 12 January 2020)

Advancement in sequencing techniques and transformative progress in metagenomics provides an unprecedented platform for functional and taxonomic characterization of the enormous microbial diversity inhabiting and governing various biochemical processes of the freshwater sources. Metagenomic analysis of freshwater resources has led to the discovery and identification of novel microbial genes and an understanding of how microorganisms mediate energy and carbon. In this study, we report the taxonomical classification of bacterial sequences obtained from 6 dam reservoir sites in Pune city, Maharashtra, India. The analysis was performed using two different alignment tools: BLAST and Kaiju. The bacterial diversity was dominated by the presence of *Vogecella indigofera*, uncultured Proteobacterium, *Wolinella Succinogenes*, *Chromobacterium violaceum*, and *Heliobacter billis*. It was further observed that, despite an identical bacterial composition over various reservoir sites, there were nominal differences in the relative abundance of the inhabitant species. Almost all reservoirs were dominated by *Vogecella indigofera* (~29%) and uncultured Proteobacterium (~15%). A seasonal analysis performed using BLAST resulted in a number of species exclusive to the season and the site of their growth. A high proportion of unidentified sequences were also reported which demands sequential identification. The results obtained through BLAST and Kaiju were significantly different, suggesting inconsistencies and inaccuracies in existing metagenomic reads comparison.

**Keywords:** Metagenomics, Freshwater Metagenomics, BLAST, Kaiju Tool, Bacterial community, Environmental Microbiology.

Until the late 20<sup>th</sup> century, the outcomes of microbiology were restricted by the meagre potential of pure culturing techniques in terms of throughput data and accuracy. This is largely because of the dependency of culturing techniques on the ability of a microbe to grow in the appropriate medium. Although culturing techniques have led to the discovery of a plethora of microorganisms, a century worth of research has only been able to identify 1% of the speculated microbial population. Acknowledging the limitations of pure culturing techniques, Pace and colleagues utilized the

available sequencing techniques to sequence the 16s rRNA and 5s rRNA in order to describe the microbial diversity in an environmental soil sample<sup>1</sup>. This was one of the pioneer projects that led to the development of a new field of environmental ecology, called metagenomics. First introduced by Jo Handelsman, metagenomics is an innovative approach of analyzing the total genetic material from an environmental source in order to understand the taxonomic composition and the role of each member in the community<sup>2</sup>. Simultaneously, advancement in sequencing technologies leading

\*Corresponding author E-mail: [preeti.chunarkar@bharatividyaapeeth.edu](mailto:preeti.chunarkar@bharatividyaapeeth.edu)



to Next Generation Sequencing (NGS), had a remarkable effect on the frequency and accuracy of the studies performed under the domain of metagenomics. Since the start of the 21<sup>st</sup> century, efforts towards metagenomic analysis have increased several folds with multiple benchmark projects aimed at understanding the complex taxonomic structure of the microbial communities inhabiting various environmental sources and their role in maintaining the biogeochemical cycles responsible for governing the flow of carbon and energy in the ecosystem. Some of the most remarkable studies have been published covering a wide range of environmental samples that are directly or indirectly involved in the health of our species. One of the most interesting of these studies started in 2003 was “The Sorcerer II Global Ocean sample expedition” (popularly known as GOS) that followed the renowned route of Darwin’s HMS Beagle, answering one of the most complex questions regarding the diversity of microbes in global marine environments (data sets of which are available on NCBI and CAMERA databases)<sup>3</sup>. The uncultured viral community associated with the human fecal matter was another vital study carried out by 4 under the paradigms of metagenomics revealing important data regarding ~60 strains of phages and other viruses colonizing the lower areas of the gut microbiome. Another set of studies deeply investigates local environmental samples, examples of which include acid mine drainage biofilms, Sargasso Sea, Whale falls, Eel river sediments, Minnesota farm soil, Pleistocene bear caves fossils, Soudan Mine, Hawaii Ocean, Mammoth fossils, Bras del Port Saltern, Coral holobiont, Neanderthal microbiome, Mediterranean Sea, Coral Reefs in Australia, Tasmanian Tiger Genome and many more. The collection of data from these studies has enhanced our understanding of various factors of human evolution, gut microbiome, and most importantly the role of microorganisms in the environmental samples that are directly involved in the general human lifestyle of that particular locality<sup>5</sup>. The data collected from the various samples can be categorized as either sequence driven data or function-driven data. While the former has a great potential to discover novel microbial species and their evolutionary relationship with existing taxa, the latter reveals important features about

the genetic and proteomic factors responsible for carrying out the physiochemical processes within a particular microbial community. Considering the wide success of various large-scale environmental projects, Metagenomics has inevitably permeated in the research of freshwater bodies and related sciences like limnology to address the lacunae in our understanding of the microbial factors responsible for 90% of global organic degradation and nutrient recycling<sup>6</sup>.

Different kingdoms of microorganisms are known to inhabit all kinds of extreme and non-extreme environments. Similarly, each freshwater body has its indigenous microbiome or microbial community that might not occur in other systems like marine, soil, or terrestrial ecosystems<sup>7</sup>. Microorganisms from all kingdoms like protozoa, fungi, bacteria, and archaea present in these freshwater ecosystems dominate a large number of biogeochemical processes that occur in the respective habitat. Most of the members of this community are also directly involved in maintaining the quality of water and pollutants. Bacterial species (Gram-positive and negative) are one of the most frequent inhabitants of freshwater bodies (more than 90%) like rivers, lakes, and other inland waters<sup>8</sup>. Studies by<sup>9</sup> further confirmed that bacterial species found in a particular freshwater system is closely related to communities from other freshwater sources but differed significantly when compared to marine ecosystems. They also established that certain bacterial colonies are typical to the freshwater systems and were universally observed in different freshwater systems, examples of such families of bacteria include Actinobacteria (ACK M1), Betaproteobacteria (GKS98, polynucleobacter, and R-BT065), and Bacteroides (SOL). It is also observed in many species that despite the cosmopolitan presence of certain families of bacteria in different climatic zones, they did show adaptation to a range of temperature but could only acclimatize to the particular temperature they are subjected to<sup>10</sup>. Furthermore, freshwater habitats also provide the essential conditions to favor the growth of a range of bacterial infections like Botulism, Campylobacteriosis, Cholera, E. coli Infection, M. marinum infection, Dysentery, Legionellosis (two distinct forms: Legionnaires’ disease and Pontiac fever), Leptospirosis, Otitis

Externa (swimmer's ear), Salmonellosis, Typhoid fever and Vibrio Illness<sup>11</sup>. These infections are directly responsible for an estimated 1.8 million death (4.1% of the global burden of diseases) globally. Consequently, the uses of antibiotics by a large population contribute to the development of antibiotic-resistant genes in the bacterial colonies through fecal matter and industrial contamination.

Advancement in sequencing techniques and transformative progress in metagenomics provides an unprecedented platform for functional and taxonomic characterization of the enormous microbial diversity inhabiting and governing various biochemical processes of the freshwater sources. Metagenomic analysis of freshwater resources has led to the discovery and identification of novel microbial genes and an understanding of how microorganisms mediate energy and carbon. A combination of techniques developed under the paradigms of functional metagenomics and sequence-driven metagenomics has been utilized in identifying novel microbial species and their peculiar biogeochemical processes, enzymatic or catalytic properties, or their pathogenic properties that are primary markers in deciding the quality of water. Various projects recently carried out in the concerned field have led to the discovery of antibiotic-resistant bacterial strains, novel viral families, gene families, and pathogenic strains. The freshwater metagenomics approach further provides the potential to compare bacterioplankton of different water bodies, identifying generic microbial families of specific water body systems (lake, river, pond) and phylogenetic classification of resident microbial members. Most importantly, Metagenomics has extensively helped in categorizing genes from various bacterial strains that are directly involved in antibiotic resistance under a new term, called resistome<sup>12</sup>.

Realizing the potential and success of metagenomics in decoding microbial communities from local freshwater resources, we focused this study on the analysis of 65 samples collected over 6 dams (Khadakwasla, Panshet, Varasgaon, Bhatgar, Mulshi, and Temghar) in Pune city (Maharashtra, India) to elucidate the bacterial composition and establish their taxonomic relationship. The building of river dams has a significant effect on the riverine environment causing a myriad of physical, biological, and chemical variations. These effects

are key factors in deciding the composition of the microbial communities inhabiting within them as discussed before<sup>13</sup>. We further analyzed, samples from the same sources over various seasons (Summer, Winter, and Monsoon), and from the surface, middle, bottom layers. The seasonal analysis was carried out with an objective to understand the effect of the annual hydrological cycle on the bacterial community.

## MATERIAL AND METHODS

### Chemicals

All the bulk chemicals, solvents used in the study were of analytical reagent (AR) grade and procured from suppliers including SD fine-chemicals, Loba Chemie, Sisco Research Laboratories and Qualigenes, India.

Fine chemicals like sodium acetate, STE buffer, lysozyme, proteinase K, dithiothrietol, HEXfluorescent dye were obtained from Sigma-Aldrich, Germany, and geneOmbio Technologies, India. Membrane filters of 45µm were procured from Himedia, India. DNA ladder mix, RNAase was procured from Promega, UK; dNTP, HF PCR buffer, Taq polymerase, was PCR primers were provided by IDT, USA and PCR purification kit was purchased from geneOmbio Technologies, India.

### Water sample Collection

Total 72 water samples were collected from six dams, Temghar, Bhatgar, Khadakwasla, Varasgaon, Panshet, and Mulshi of Pune city (18.5204° N, 73.8567° E). This study had been taken three samples (Bottom layer, Middle layer, and Upper layer) at a time for every season (rainy, winter, and summer). Total of nine samples were taken from one place. All water samples were collected in a sterile Labifite borosilicate glass reagent bottle with a screw cap. While taking the sample, the bottle was used to rinse three times with the respective sample (river/lake/dam water). After rinsing, the bottle was submerged below the water level till it will fill with the sample below 2.5 cm below the lid to maintain homogeneity. After the collection of the samples, till analysis, it is used to store at 400C in the fridge.

### DNA isolation from water samples

15ml samples were collected and stored in sterilized bottles followed by the addition of 1.5ml

3M Sodium acetate and 33ml absolute ethanol. The mixture was stored at 40C until extraction.

For DNA extraction, the mixture was centrifuged at 10000 rpm for 20 mins (60C) and the supernatant was discarded. The pellet was washed and re-suspended in 0.9% saline. Nucleic acid extraction was performed by the addition of the STE buffer, lysozymes, and Proteinase K (10µg/ml). Purification of nucleic acid was performed by the NaCl and SDS lysis followed by phenol: chloroform: isoamyl alcohol-based extraction. Extracted and purified DNA was deproteinized thrice using the Tris-saturated phenol (isoamyl

alcohol: CHCl<sub>3</sub>: Phenol in ratio 2:48:50), followed by CHCl<sub>3</sub>: isoamyl alcohol (24:1). The DNA was precipitated using 2% sodium acetate and absolute alcohol. The DNA was dried and dissolved in 30µL nuclease-free water.

#### Qualitative and Quantitative assessment

The extracted community DNA has been subjected to qualitative and quantitative assessment. By using a Qubit™ Fluorometer (Invitrogen, USA) kit quality assessment of genomic DNA was performed by 1% agarose gel electrophoresis. DNA extraction has been estimated with the help of positive control sample (*E. coli*)

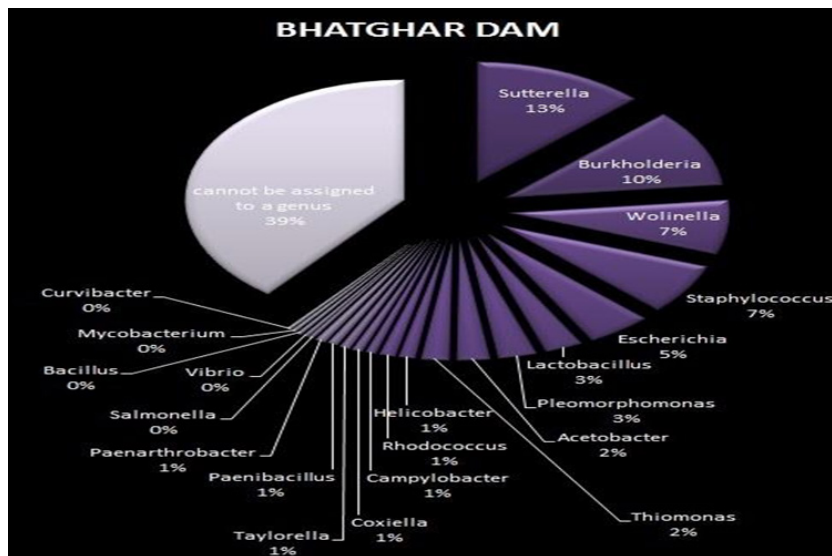


Fig. 1. Bacterial diversity analyzed by Kaiju tool of Bhatghar Dam

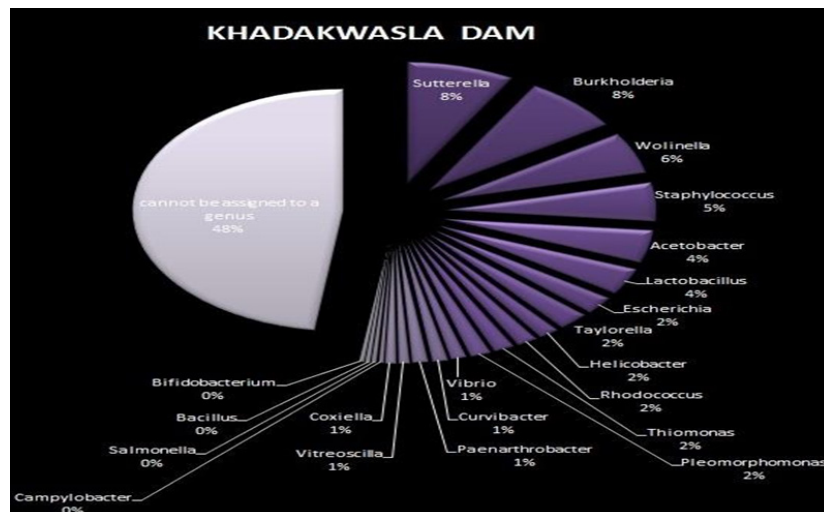


Fig. 2. Bacterial diversity analyzed by Kaiju tool of Khadakwasla Dam

and negative control (plain saline) to rule out the possibility of extraction failure. Quantitative assessment has been carried out using Qubit® DNA BR (Broad-Range) Assay Kit with the Qubit® 2.0 Fluorometer (Life Technologies USA). The final concentration of the community DNA was ranging from 930 (µg/mL) to 950 (µg/mL).

**T-RFLP analysis of community DNA**

The community DNA contains all types of microbial diversity. The current study only focuses on bacterial diversity. The bacterial 16S rRNA gene was amplified by PCR using the primer set 8F: 5'-AGAGTTTGATCATGGCTCAG-3' and

1492R: 5'-GGCTACCTTGCCACGACTTC-3' (Lane, 1991). The 8F primer was labelled at the 5' end with HEXfluorescent dye. The PCR mixture was 0.5 µL of each primer (10 mM), 5 µL of the PCR buffer, 1 µL of dNTP (2.5 mM), 0.5 µL of Taq polymerase (2.5 U/µL) and double-distilled water for final reaction volume of 50 µL. PCR was performed at 95 C for 5 min; 30 cycles of 95 C for 1 min, 55 C for 1 min and 72 C for 1 min; and 72 C for 10 min. PCR products were checked on 1% agarose gels and purified with the geneO-Spin PCR product Purification kit (geneOmbio Technologies, India) according to the manufacturer's protocol.

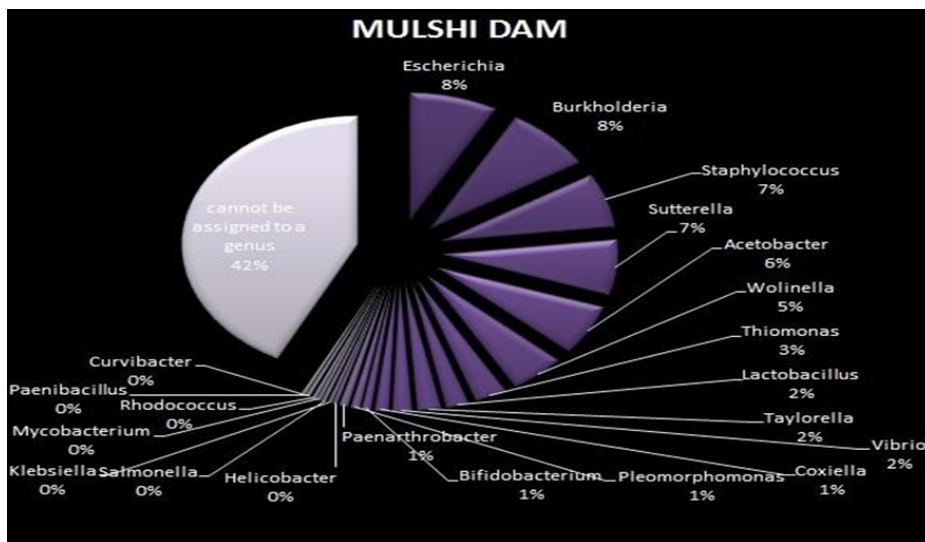


Fig. 3. Bacterial diversity analyzed by Kaiju tool of Mulshi Dam

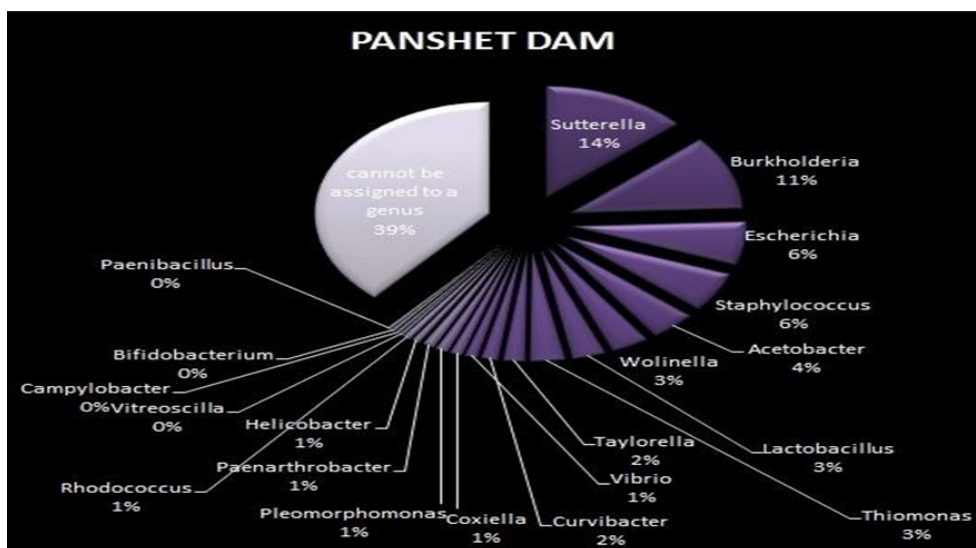


Fig. 4. Bacterial diversity analyzed by Kaiju tool of Panshet Dam

Following PCR, 10 µL of PCR products were digested with 0.5 U of TaqI restriction enzymes (New England Biolabs) for 3 hours at 37°C in 20 µL reaction volumes. Digestions were separated on 3% agarose gels in 1X TBE buffer containing ethidium bromide, and visualized under UV light.

10 µL of the digested DNA was desalted using the following procedure: 2.5 µL of 125mM EDTA and 1/10 volume of 3M Sodium Acetate pH 5.2 was added to 10 µL of digested DNA. Further

2.5 volume of ice cold ethanol was added into the tubes and mixed well. Tubes were then centrifuged at 12000 RPM for 20 minutes at 18°C. Supernatant was removed being careful not to dislodge the pellet. Pellet was then washed with 60 µL of 70% Ethanol twice with centrifugation at 12000 RPM for 20 minutes at 18°C temperature. Pellet was then dried at 37 °C for 30 minutes.

**Sample Preparation and loading**

Hi-Di Formamide (9.7 µL) from Applied Biosystems was added to the dried pellet. Each

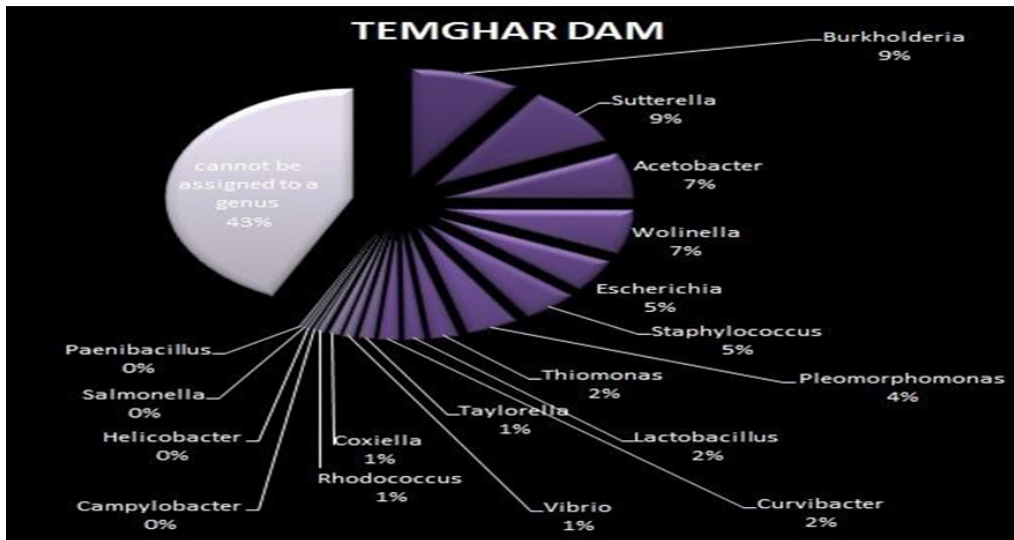


Fig. 5. Bacterial diversity analyzed by Kaiju tool of Panshet Dam

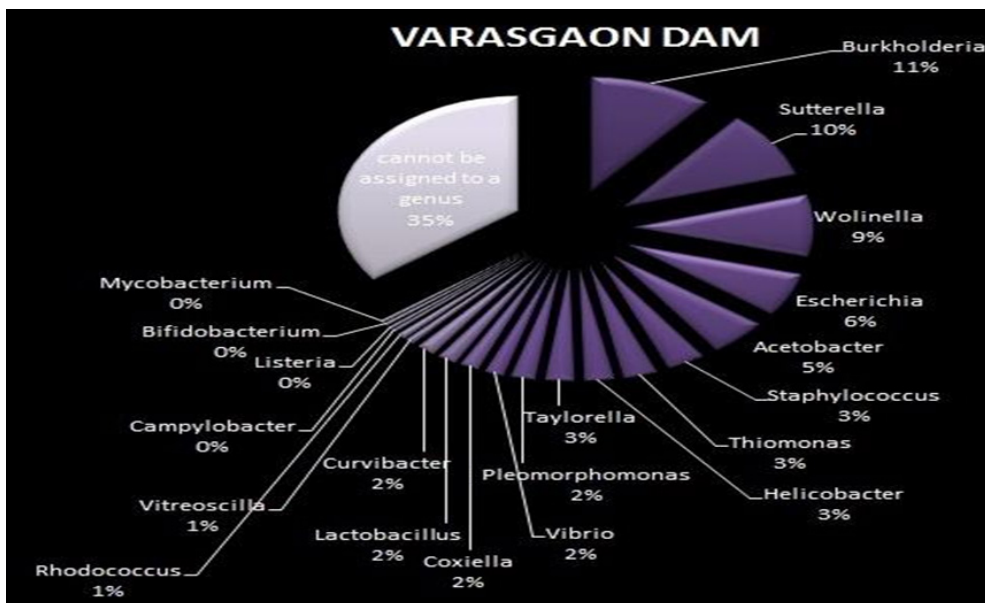


Fig. 6. Bacterial diversity analyzed by Kaiju tool of Varasgaon Dam

sample was also added with 0.3 µL GenScan 500 LIZ Internal Size Standard. This mixture was denatured at 95°C for 3 minutes and immediately chilled on ice before loading. The samples were

then subjected to electrophoresis on the 3130 Genetic Analyzer using the FA\_36\_POP-7™ run module and G5 dye set.

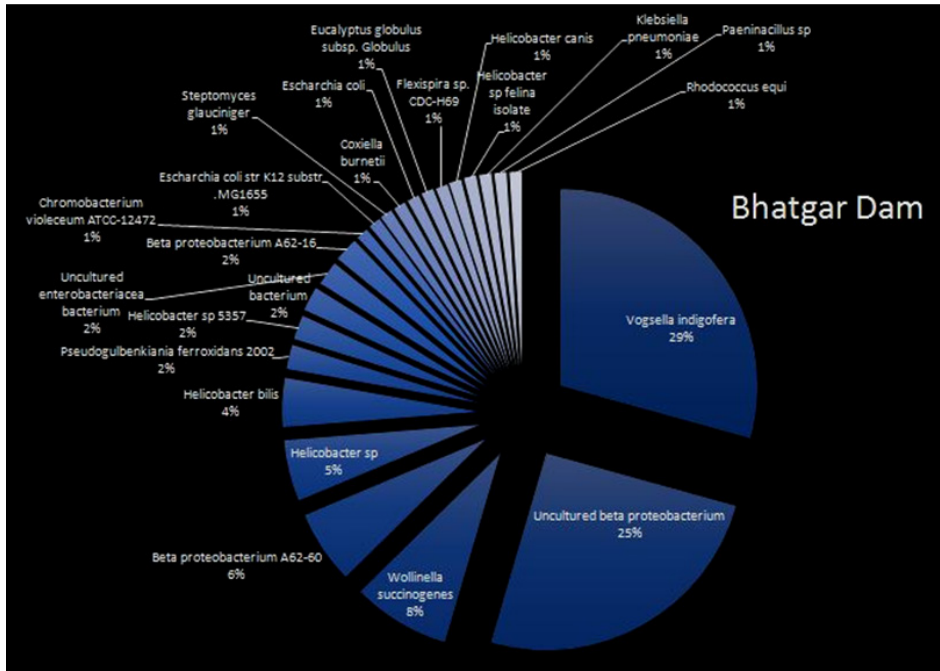


Fig. 7. Bacterial diversity analyzed by BLAST tool of Bhatgar Dam

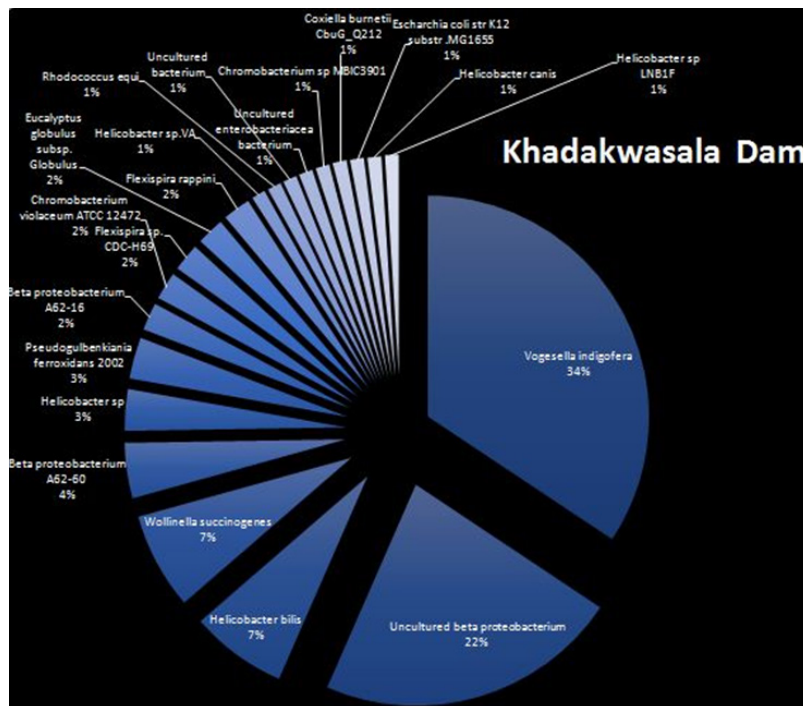
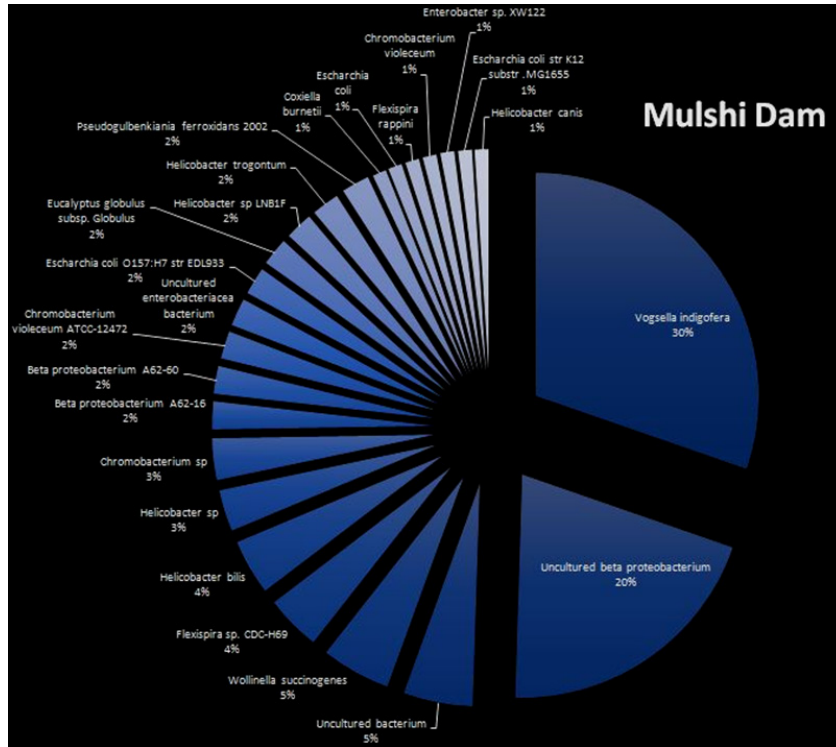


Fig. 8. Bacterial diversity analyzed by BLAST tool of Khadakwasala Dam

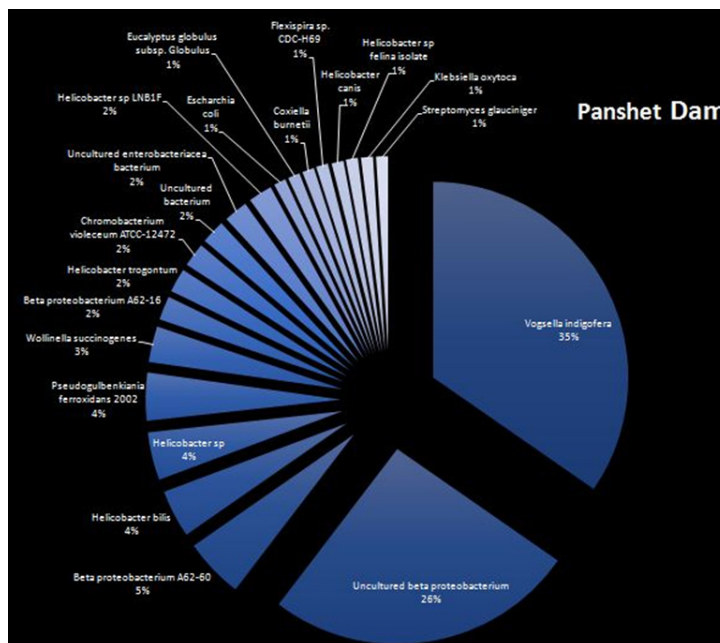
**Gene Mapper data analysis**

GeneMapper software based analysis was performed for fragment analysis after completion of the capillary electrophoresis.

Output from automated sequencers is in the form of an electropherogram, with peaks representing fluorescently labeled T-RFs detected over time in relation to the size standard. The duration and



**Fig. 9.** Bacterial diversity analyzed by BLAST tool of Mulshi Dam



**Fig. 10.** Bacterial diversity analyzed by BLAST tool of Panshet Dam



intensity of the fluorescent signal from T-RFs is reflected in the area and height of each peak detected, respectively. Software specific to each

sequencing unit collects data from each run. The ABI 3730 capillary sequencer operates GeneMapper v3.5 (AppliedBiosystems), which

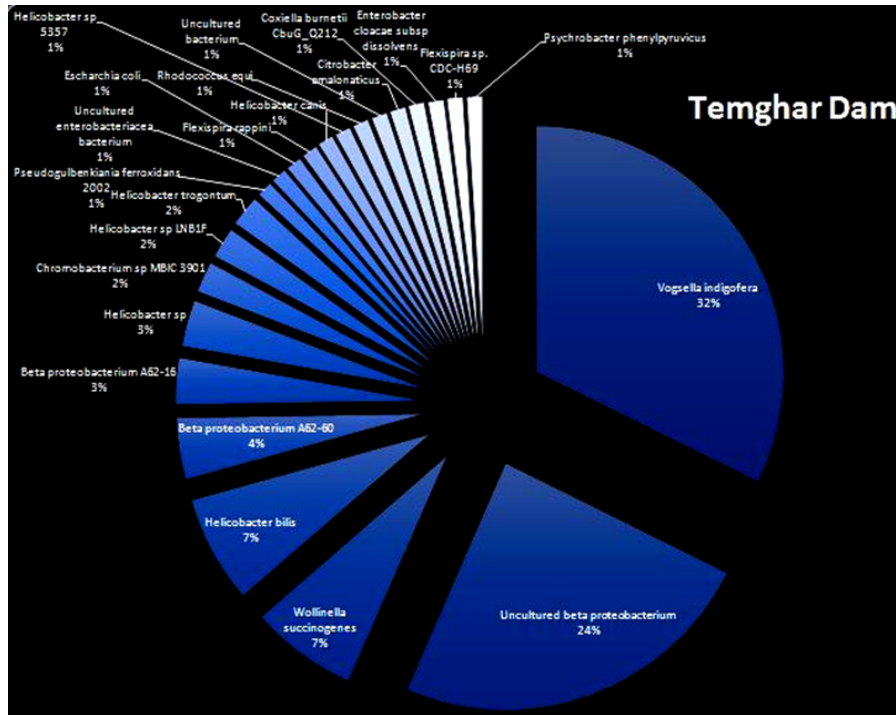


Fig. 11. Bacterial diversity analyzed by BLAST tool of Temghar Dam

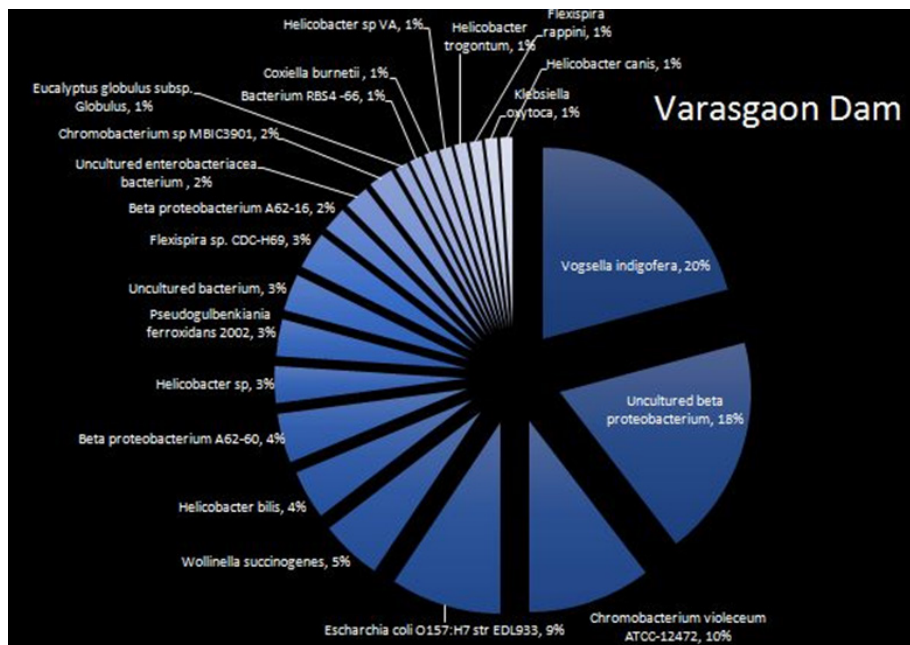


Fig. 12. Bacterial diversity analyzed by BLAST tool of Varasgaon Dam

performs the functions of both GeneScan and Genotyper. Either data collection program provides researchers with several algorithms for sizing sample fragments by comparing their mobility with that of the size standard. Once data are processed and fragment lengths assigned, the dataset is typically imported into a spreadsheet program, such as Microsoft Excel (Microsoft Corp., Redmond, WA). In the spreadsheet, sample identifiers can be added and presence/absence (1, 0) matrices developed. Other manipulations, such as matrix inversions, can also be performed. The T-RFs for each sample run should be closely examined and the entire run evaluated for the average number of T-RFs detected per sample and the number of T-RFs contained in the various size classes.

The result of a T-RFLP profiling is a graph called Electropherograms, which is an intensity plot representation of an electrophoresis experiment (gel or capillary). In an Electropherograms the X-axis marks the sizes of the fragments (base pairs) while the Y-axis marks the fluorescence intensity in fluorescent unit of each fragment. In a T-RFLP profile each peak assumingly corresponds to one genetic variant in the original sample while its height corresponds to its relative abundance in the specific community. T-RFs having size less than 40 bases were eliminated from the analysis as they might result from primer-dimers. Fluorescent signal threshold was set to 10 fluorescent units as per the standards to minimize the background signal and signals arising from ssDNA non-specific amplicons/fragments. Bacterial richness ranged from 7 to 114 with an average of 33.66 with T-RFLP technique. Sample T-RFLP has been taken further for cloning and sequencing for identification of bacterial population.

#### **Cloning of 16S rDNA amplicons**

Cloning of 16S rDNA amplicons was performed using Promega TA cloning kit consisting of pGEMT vector and *E. coli* XL1 blue Chemical competent cells. PCR products obtained after colony PCR were checked on 1% agarose gel by agarose gel electrophoresis.

#### **Purification and DNA sequencing**

PCR products were purified using Purelink PCR purification kit (Invitrogen, USA). The sequencing of purified DNA was performed using the BigDye Terminator V 3.1 Cycle sequencing kit (Applied Biosystems, USA). The DNA sequencing

reaction is set using 2µl Sequencing primer used for sequencing T7 promoter (1.6pMol concentration), 2µl Purified PCR product (template DNA ~ 50 -100 ng/µL concentration), 1.8µl sequencing buffer, 0.5µl ready reaction mix and 3.7µl Nuclease free water.

Applied Biosystems 3130 Genetic Analysis (Automated Sequencing Analysis machine) was used for the analysis of the sequencing products obtained from the cycling kit. The sequences were generated in .abl format which was then exported to FASTA format for analysis on BLAST and Kaiju tools.

## **RESULTS AND DISCUSSION**

72 samples collected from 6 different dam sites were compared and analyzed to the existing databases using BLAST and Kaiju tools for analysis of bacterial composition. It was observed that sequences from Burkholderia, Sutterella, Wolinella, Escherichia, Acetobacter, and Staphylococcus were relatively abundant compared to the others, while an average of 40% nucleotide sequences obtained from various samples remained unidentified following the comparison with the mentioned database. It is a clear indication of uncultured bacterial communities. A consistent bacterial composition was observed over various dam reservoirs at different locations in the city.

#### **Analysis through Kaiju tool**

Figure 1 to 6, shows the relative abundance of different bacterial groups within each of the reservoir systems. Among the 6 dam reservoir systems, Khadakwasla Dam has the highest proportion of unidentified sequences at 49% while Varasgaon Dam has the lowest percentage of unidentified sequences at 35%. While all the systems resulted in a relatively higher majority of Burkholderia (8-11%), Sutterella (8-14%), Wolinella (5-9%), Staphylococcus (3-7%), Acetobacter (4-7%), Lactobacillus (2-4%), and Escherichia (5-8%), a few cases were observed to have a significantly lower population of the mentioned species, for example, Mulshi dam comprising 7% Burkholderia, Panshet comprising 3% Wolinella and Khadakwasla comprising only 2% Escherichia sequences. Albeit, in lower concentrations, another set of sequences observed throughout the 6 systems were from Pleomorphonas

(1-4%), *Helicobacter* (0-3%), *Rhodococcus* (0-2%), *Coxiella* (1-2%), *Taylorella* (1-3%), *Vibrio* (0-2%), and *Thiomonas* (2-3%). With an exception of Mulshi Dam, *Campylobacter* was observed in all the other systems at relatively low concentrations ranging from 0-1%. *Paenobacillus* was completely absent in Khadakwasla and Varasgaon dam while ranging at 0-1% in other systems. Similar trends were observed for other sequences from *Paenathrobacter* (0-1%), *Salmonella* (0-1%), *Curvibacter* (0-2%), and *Bifidobacterium* (0-1%), where each of these species was absent from 2 reservoir systems: *Paenathrobacter* from Temghar and Varasgaon dam, *Salmonella* from Panshet and Varasgaon Dam, *Curvibacter* from Panshet and Varasgaon Dam, and *Bifidobacterium* from Bhatgar and Temghar Dam. Sequences from the *Bacillus* group were only obtained from Bhatgar and Khadakwasla Dam in ultra-trace quantities ranging from 0-1%. *Mycobacterium* sequences were further observed to be absent from Khadakwasla, Temghar, and Panshet dam, while *Vitrioscella* sequences were missing from Bhatgar, Mulshi, and Temghar Dams. Furthermore, Varasgaon Dam and Mulshi Dam reported the presence of sequences from exclusive groups, *Listeria* and *Klebsiella*, respectively.

The presence of the *Staphylococcus aureus* and *Burkholderiales cepacia* complexes, suggests an important link to the human respiratory tract as the former is a commensal in the microbiome while the latter is an opportunistic pathogen causing respiratory infections<sup>14, 15</sup>. *Burkholderia* genus comprises a range of 30 different species that are involved in various bioremediation, ecological, and pathogenic processes. Some of the species from *Burkholderia* have been known pathogens of plants (carnation, onions, blight) and have also found early applications as bioweapons (Germany, during World War I) for its ability to cause infectious diseases in humans, horses, mules, and donkeys like melioidosis, septicemia, and pneumonia<sup>16, 17</sup>. Furthermore, *Burkholderia* species have been actively utilized in industrial bioremediation processes, nutrition delivery in plants, and disintegration of biohazardous waste<sup>18</sup>. A high percentage presence (~18%) of *Suterella* sp. also indicates fecal contamination since the bacterial species has been most commonly isolated from human feces<sup>19</sup>. Similarly, a considerable

composition of reservoir freshwater bacterial colonies was identified to be commensals of the human GI tract microbiome like various strains of *Escherichia coli*, while sequences of various pathogenic strains from the family *Campylobacter* like *Wolinella succinogenes* and *Helicobacter cinaedi* were recovered in trace quantities<sup>20, 21, 22</sup>. *Escherichia* also shows a peculiar ability to grow under low nutrient and other physiological stress, therefore, showing a high species richness. It also contributes to the microbial resistome of the habitat by expressing various genes for multidrug antibiotic resistance<sup>23</sup>. Various pathogenic species of the *Helicobacter* genus including *H. bilis*, *H. canis*, *H. trogontum*, and *H. hepaticus* responsible for causing a range of infections in mammals were also persistently recovered throughout the year, although in trace quantities<sup>24, 25, 26</sup>. *Coxiella burnetii*, a Rickettsia looking bacteria, further adds to the list of pathogens recovered from the river samples. It is a characteristic mild pathogen, resistant to extreme environmental changes and is the sole etiological factor responsible for the Q fever in cattle and humans<sup>27</sup>.

#### Analysis through BLAST

A seasonal analysis was performed for each reservoir site using BLAST to understand the effect of the annual hydrological cycle on the bacterial composition. BLAST analysis resulted in a much more sensitive taxonomic analysis compared to that of Kaiju with significant differences in the bacterial composition. The results obtained through BLAST alignment reported various species from the bacterial groups identified in the Kaiju analysis, for example, *Wolinella succinogenes*, *Helicobacter bilis*, *Helicobacter pylori*, and *Bacillus pumilus* among others. Additionally, sequences from bacterial taxa, *Vogisella*, *Chromobacterium*, *Psuedogulbekiana*, *Citrobacter*, *Flexispira*, *Streptomyces*, *Neisseria*, *Pantoea*, *Enterobacterium*, *Glyxomyces*, and *Iodobacter* were also identified from various samples suggesting the presence of a wide spectrum of ecologically important bacteria in the reservoir sites. Figure 7-12, shows the relative abundance of various species obtained from different sites in Summer, Winter, and Rainy seasons.

According to the BLAST analysis, *Vogisella indigofera* was the most dominating bacterial species found among all the dams.

*Vogisella indigofera* is a known mesophilic freshwater proteobacterium that shows a perennial presence in relatively high concentrations in rivers<sup>28</sup>. In comparison to other seasons, winter can be attributed to a rise in the population of *Wolinella Succiongenes* in both the rivers that might possibly affect the anaerobic consumption of sulfites, polysulfide, nitrites, fumarates, and nitrates<sup>29</sup>. One of the rare bacteria, *Chromobacterium violaceum*, was slightly more abundant during the rainy season compared to the summer or winter. *C. violaceum* is common to soil or stagnant waters, therefore, its presence can be directly correlated with the interplay of rivers with the environmental soil, ponds, or lakes. While ingestion of water contaminated by *C. violaceum* might not be harmful, it can cause localized skin infections with a potential to progress into septicemia, hence increasing the chances for multiorgan failure<sup>30</sup>. A high percentage of bacterial sequences obtained from unknown species of Proteobacteria phylum or Enterobacteriaceae family might have a potential role in maintaining ecological cycles involving nitrogen, oxygen, sulfur, and carbon. Another agent of contamination, *Bacillus pumilus* was reported from Bhatgar and Khadakwasla dam. *Bacillus pumilus* is a Gram-positive aerobic bacteria that is commonly an inhabitant of soil microbial communities and is actively involved in rhizosphere formation and fungal resistance in various plants like red peppers and wheat<sup>31</sup>. Furthermore, it serves the ecology of the surrounding environment by taking the responsibility for fixing atmospheric nitrogen to ammonia<sup>32</sup>. It is also rarely pathogenic to humans causing toxic effects on epithelial cells through a complex of lipopeptide and pumilacidins. *Streptomyces glaucinger* represents the cosmopolitan mesophilic freshwater actinomycete that actively degrades starch, xanthin, casein, and hypoxanthine. It has applications in anti-fungal therapies for its pathogenic activity towards clinically important *Candida albicans*<sup>33</sup>. Industrially important strains of bacteria were also found to be active members of the Pune dam microbiome community, like, *Pantoea agglomerans*, a species that can be exploited as a rich source for various antibiotics like phenazine and pantocins<sup>19</sup>. Furthermore, *P. agglomerans* have been recently modified to cause antimalarial effects towards *Plasmodium* in mosquito guts<sup>34</sup>.

There were certain bacterial species exclusive to certain dams in specific seasons. An unusual presence of *Psychrobacter pacificensis* was reported in Khadakwasla dam during the rainy season, as this is a non-motile, deep water bacteria adapted to psychrophilic conditions that show oxidative and catalytic activities<sup>35</sup>. The summer season further marks the growth of *Zooglea ramigera* in Khadakwasla dam, suggesting deposition of minerals, since *Z. ramigera* growth indicates the presence of an enriched aqueous environment<sup>36</sup>. Panshet dam accounts for one of the highest species richness among all the environments studied. In addition to species obtained in other resources, Panshet dam sustains the growth of *Acidiferrobacter thiooxydans*, *Achromatium*, and *Neisseria elongata*. *A. thiooxydans* are a highly resistant species towards acidic and thermal stress belonging to the family, Ectothiorhodospiraceae. It might have a potential role in the anaerobic oxidation of Iron and Sulfur radicals, hence contributing to the overall ecology of the environment<sup>37</sup>. Another sulfur oxidising class of bacterial genus called, *Achromatium* was also isolated from Panshet Dam. *Achromatium* has an interesting genome that is highly similar to many environmental genomes and often contributes to the bacterial population at the interface of river sediments<sup>38</sup>. An ancestor of the pathogenic Neisseriaceae family, *N. elongata* was also isolated from Panshet Dam. Usually a commensal in the pharynx of humans, these organisms can otherwise cause systematic diseases like endocarditis, osteomyelitis, and septicemia<sup>39</sup>. A biologically and industrially important genus, *Vitrioscella*, was isolated from Mulshi dam during the rainy season. It is one of the earliest bacteria that has the capability to biosynthesize bacterial hemoglobin and find many applications in industrial biotechnology for its ability to promote cell growth, fermentation, biodegradation of toxic substances, and expulsion of toxic xenobiotics from cells, enhance protein synthesis, and metabolic productivity<sup>40</sup>. A chemoorganoheterotrophic bacterium, *Leptothrix fluviatis* was isolated from Temghar dam, indicating neutral to slightly acidic environmental conditions. *Leptothrix* further suggests the presence of rust contamination in the dam as the organism is highly ferruginous<sup>41</sup>.

## CONCLUSION

A deep metagenomic investigation of bacterial composition at various dam reservoir sites in Pune resulted in the identification of various etiologically, industrially, and ecologically important bacterial strains. However, an analysis performed using tools, Kaiju, and BLAST, reported significant differences in the taxonomic classification obtained. This suggests a lot of inaccuracy and inconsistency in results obtained through the existing tools for the alignment of metagenomic reads. To develop an efficient metagenomic analysis technique is the need of the hour.

## ACKNOWLEDGMENT

We would like to acknowledge our institute for providing us the infrastructure facilities. We would like to thank Ms. Sanober Khatoon for helping us in technical issues.

### Funding Source

Bharati Vidyapeeth (Deemed to be University) has provided the funding for the current project.

### Conflict of Interest

Authors declare no conflict of interests.

## REFERENCES

1. Wooley J, Godzik A, and Friedberg I. A primer on Metagenomics. *PLoS Computational Biology*. 2010. **6**(2) : e100067
2. Handlesman J, Rondon M, Brady S, Clardy J, and Goodman R. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry and Biology*. 1998. **5**(10) :245.
3. Rusch D, Halpern A, Sutton G, Heidelberg K, Williamson S, *et al.* The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biology* 2007. **5**(3) :77.
4. Breitbart M, Hewson I, Felts B, Mahaffy J, Nulton J, *et al.* Metagenomic Analyses of an Uncultured Viral Community from Human Feces. *Journal of Bacteriology*. 2003. **185**(20) :6220.
5. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. 2010. **20**(9) : 1297.
6. Wetzel R. Freshwater ecology: changes, requirements, and future demands. *Limnology*. 2000. **1**: 3.
7. Shreeve J. The blueprint of life. *US News World Reports*. 2005. **139**(16) :70.
8. Hiorns W, Methe' B, Nierzwickbauer S, and Zehr J. Bacterial diversity in Adirondack mountain lakes as revealed by 16S rRNA gene sequences. *Applied Environmental Microbiology*. 1997. **63**: 2957.
9. Zwart G, Crump B, Kamst-van A, Hagen F, and Han S. Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat Microb Ecol*. 2002. **28**:141-155
10. Hahn M, and Pockl M. Ecotypes of planktonic Actinobacteria with identical 16S rRNA genes adapted to thermal niches in temperate, subtropical, and tropical freshwater habitats. *Applied Environmental Microbiology*. 2005. **71**:766.
11. Sharma S, Sachdeva P, and Viridi J. Emerging water-borne pathogens. *Applied Microbiol Biotechnolgy*. 2003. **61**: 424.
12. Nnadozie C, and Odume O. Freshwater environments as reservoirs of antibiotic resistant bacteria and their role in the dissemination of antibiotic resistance genes. *Environmental pollution*. 2009. **254**(part B) :113067.
13. Yan Q, Bi Y, Deng Y, He Z, Wou L, *et al.* Impacts of the Three Gorges Dam on microbial structure and potential function. *Scientific Reports*. 2015. **5**: 8605.
14. Masalha M, Borovok I, Schreiber R, Aharonowitz Y, and Cohen G . Analysis of transcription of the *Staphylococcus aureus* aerobic class Ib and anaerobic class III ribonucleotide reductase genes in response to oxygen. *Journal of Bacteriology*. 2001. **183**(24): 7260.
15. Chandler L. Challenges in Clinical Microbiology Testing. *Accurate results in ththe clinical laboratory*. 2013. **315**.
16. Godoy D, Gaynor R, Simpson A, Aanensen DM, Pitt T, *et al.* Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, *Burkholderia pseudomallei* and *Burkholderia mallei*. *Journal of Clinical Microbiology*. 2003. **41**(5) :2068.
17. Vazquez N, Cesar V, Rocio A, Yendi N, Marco L, *et al.* Modifications of bacterial populations in anthracene contaminated soils. *Applied soil ecology*. 2012. **61** :113.
18. Coenye T. and Vandamme P. Diversity and significance of *Burkholderia* species occupying diverse ecological niches. *Environmental*

- Microbiology*. 2003. **5(9)** :719.
19. Wang L, Christophersen C, Sorich M, Gerber J, Angley M, *et al.* Increased abundance of *Sutterella* spp. and *Ruminococcus torques* in feces of children with autism spectrum disorder. *Molecular Autism*. 2013. **4(1)** :42.
  20. Baar C, Eppinger M, Raddatz G, Simon J, Lanz C, *et al.* Complete genome sequence and analysis of *Wolinella succinogenes*. *Proceedings of the National Academy of Sciences of the United States of America*. 2003. **100(20)**: 11690.
  21. Batt C. *Alcaligenes*. *Encyclopedia of Food Microbiology Second edition*. 2014 :**38**.
  22. Kawamura Y, Tomida J, Morita Y, Fujii S, Okamoto T, *et al.*. Clinical and bacteriological characteristics of *Helicobacter cinaedi* infection. *Journal of Infection and Chemotherapy*. 2014. **20(9)** : 517.
  23. Kaper J, Nataro J and Mobley H. Pathogenic *Escherichia coli*. *Nat Rev Microbiol*. 2004. **2** :123.
  24. Dutasta F, Samaha E, Carayol N, Masse J, Bourillon C, *et al.*. Acute Colitis Caused by *Helicobacter trogontum* in Immunocompetent Patient. *Emerging Infectious Diseases*. 2016. **22(2)** :335.
  25. Abidi M, Wilhelm M, Neff J, Hughes J, Scott C, *et al.*. *Helicobacter canis* Bacteremia in a Patient with Fever of Unknown Origin. *Journal of clinical microbiology*. 2013. **51(3)** : 1046.
  26. Falsafi T, and Mahboub M. *Helicobacter hepaticus*, a new pathogenic species of the *Helicobacter* genus: Similarities and differences with *H. pylori*. *Iranian Journal of Microbiology*. 2013. **5(3)**. 185.
  27. Walker D. Rickestti. *International Encyclopedia of Public Health (Second Edition)* .2017.
  28. Gleim D, Kracht M, and Weiss N.. Prokaryotic Nomenclature Up-to-date. *Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures*. 2020.
  29. Simon J, and Kroneck M. Microbial Sulfite Respiration. *Advances in Microbial Physiology*. 2013. **62**: 45.
  30. Kumar M. *Chromobacterium violaceum*: A rare bacterium isolated from a wound over the scalp. *International journal of applied & basic medical research*. 2012. **2(1)**: 70.
  31. Sari, E, Etebarian, R, and Aminian H. The effects of *Bacillus pumilus*, isolated from wheat rhizosphere, on resistance in wheat seedling roots against the Take-all fungus, *Gaeumannomyces graminis* var. *tritici*. *Journal of Phytopathology*. 2007. **155**: 720.
  32. Hernandez J, Bashan, L, Rodriguez D, Rodriguez Y, and Bashan Y. Growth promotion of freshwater microalga *Chlorella vulgaris* by the nitrogen-fixing, plant growth-promoting bacterium *Bacillus pumilus* from arid zone soils. *European Journal of Soil Biology*. 2009. **45**: 88.
  33. Huang Y, Li W, Wang L, Lanoot B, Vancanneyt M, Rodriguez C *et al.* *Streptomyces glauciniger* sp. nov., a novel mesophilic streptomycete isolated from soil in south China. *International Journal of Systematic and Evolutionary Microbiology*. 2004. **54(6)**: 2085.
  34. Lim J, Hwan L; Byoung K, and Sunggi H. Draft genome sequence of *Pantoea agglomerans* R190, a producer of antibiotics against phytopathogens and foodborne pathogens. *Journal of Biotechnology*. 2014. **188**: 7.
  35. Maruyama A, Honda D, Yamamoto H, Kitamura K, and Higashihara T. Phylogenetic analysis of psychrophilic bacteria isolated from the Japan Trench, including a description of the deep-sea species *Psychrobacter pacificensis* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*. 2000. **50(2)**: 835.
  36. Roinestad F, and Yall I. Volutin granules in *Zoogloea ramigera*. *Appl Microbiol*. 1970 **19(6)**: 973.
  37. Hallberg K, Hedrich S, and Johnson DB. *Acidiferrobacter thiooxydans*, gen. nov. sp. nov.; an acidophilic, thermo-tolerant, facultatively anaerobic iron- and sulfur-oxidizer of the family *Ectothiorhodospiraceae*. *Extremophiles*. 2011. **15(2)** :271.
  38. Ionescu D, Bizic M, and Maio N. Community-like genome in single cells of the sulfur bacterium *Achromatium oxaliferum* . *Nat Commun*. 2017. **8**: 455
  39. Grant P, Brenner D, Steigerwalt A, Hollis D, and Weaver R. N. *elongata* subsp. *nitroreducens* subsp. nov., Formerly CDC Group M-6, a Gram-Negative Bacterium Associated with Endocarditis. *Journal of Clinical Microbiology*. 1990: 2591.
  40. Stark B, Dikshit K, and Pagilla K. Recent advances in understanding the structure, function, and biotechnological usefulness of the hemoglobin from the bacterium *Vitreoscilla*. *Biotechnol Lett*. 2011. **33(9)** : 1705.
  41. Emerson D, Reitner J, and Thiel V. *Leptothrix*. *Encyclopedia of Geobiology Springer Dordrecht*. 2013.