Metagenomic approach to characterize soil microbial diversity of Phumdi at Loktak Lake
Sampada Puranik, Rajesh Ramavadh Pal, Ravi Prabhakar More and Hemant J. Purohit

ABSTRACT
Loktak, one of the largest freshwater lakes of India, is known for floating islands (Phumdi), being made up of a heterogeneous biomass of vegetation and soil. This ecological site represents an exclusive environmental habitat wherein the rhizospheric microbial community of Phumdi plays a key role in biogeochemical cycling of nutrients. A culture-independent whole genome shotgun sequencing based metagenomic approach was employed to unravel the composition of the microbial community and its corresponding functional potential at this environmental habitat. Proteobacteria (51%) was found to be the most dominant bacterial phylum followed by Acidobacteria (10%), Actinobacteria (9%) and Bacteroidetes (7%). Furthermore, Loktak metagenome data were compared with available metagenomes from four other aquatic habitats, varying from pristine to highly polluted eutrophic habitats. The comparative metagenomics approach aided by statistical analysis revealed that Candidatus Solibacter, Bradyrhizobium, Candidatus Koribacter, Pedosphaera, Methylobacterium, Anaeromyxobacter, Sorangium, Opitutus and Acidobacterium genera are selectively dominant at this habitat. Correspondingly, 12 different functional categories were found to be exclusively prevalent at Phumdi compared to other freshwater habitats. These differential features have been attributed to the unique habitat at Phumdi and correlated to the phenomenon of bioremediation at Loktak Lake.

Key words | Loktak Lake, metagenomics, MG-RAST, microbial diversity, Phumdi, STAMP

INTRODUCTION
Freshwater lake sediment harbors a high diversity of microbes, which ultimately drive biogeochemical cycles at this habitat (Hölker et al. 2013). These ecosystems are undermined by anthropogenic activities, which have altered many lake systems drastically (Hoverman & Johnson 2012). One of the observed effects is eutrophication, which leads to a detrimental imbalance of nutrient cycling in lake ecosystems (Conley et al. 2009; Finlay et al. 2013). Loktak Lake (LL), spanning an area of around 289 square kilometers, is one of the largest freshwater habitats in India and plays an important role in the socio-economic development of Manipur state. However, due to the many hydropower projects, fisheries, and other anthropogenic activities, there is an environmental threat to this system (Garg 2015). It is known for its floating islands (Phumdi), which are a heterogeneous biomass of vegetation, soil, and organic matter at various stages of decomposition. Phumdi also constitute a dense rhizosphere extending to the lake sediment, and thus create a distinct ecological habitat close to a constructed wetland system. Constructed wetlands harbor the phenomenon of bioremediation performed by microbial communities and have been employed effectively in treating wastewater streams (Bell et al. 2014). These microbial communities perform bioremediation by a variety of mechanisms such as organic matter mineralization, biological control against soil-borne pathogens, biological nitrogen fixation, and root growth promotion (Pii et al. 2015).

This study aims at identification and classification of the microbial community residing in the habitat constituted by the Phumdi of LL. The metagenomics approach is widely used to characterize the microbial communities of an environmental habitat. Numerous researchers have used this tool to determine the composition of the microbial community in freshwater lake ecosystems (Oh et al. 2011; Smith...
It has also been used to study microbial diversity in the rhizosphere of individual plants with phytoremediation activity (Chen et al. 2012; Wei et al. 2014). However, Phumdi represent a variety of vegetation, thus harboring a myriad of the microbial community. Hence, it is expected that this unique habitat could house an exclusively enriched collection of microbes in comparison to other freshwater habitats. Therefore, a comparison of the microbial diversity from Phumdi with available diversity reports from freshwater habitats is necessary to reveal the favored microbial residents at this habitat. Furthermore, a close correlation of the microbial community of LL habitat with a freshwater habitat can be extrapolated with its pristine nature.

In this study, a comparative metagenomic approach was implemented to explore the favored microbial community associated with Phumdi at LL. To the best of our knowledge, this is the first study employing a metagenomic approach to describe the microbial diversity for this habitat. In addition, we compared microbial diversity data with those of other freshwater lakes to derive an exclusive association of the microbial community corresponding to Phumdi. Data for other freshwater lakes, ranging from oligotrophic to eutrophic states, were selected for comparative metagenomics. By using an in silico based comparative metagenomics approach we could discriminate LL from eutrophic lake habitat and thus have a comparative measure of health for this habitat.

**MATERIALS AND METHODS**

**Sampling site description and metagenome sequencing**

LL (24°33′N93°47′E) is located near Moirangin Manipur state, India. Five soil samples were collected randomly from the rhizospheric sites of different floating islands (Phumdi), covering a fairly large area at the lake. Jogesh & Dey (2014) have reported that discharge of municipal sewage, domestic wastes, fertilizers and pesticides from agricultural practices have degraded the water quality at LL. The wastewater from fertilizers and pesticides is the main source of the organic load and other aromatic pollutants at this habitat. All soil samples were mixed together, representing a composite sample of Phumdi at LL. The metagenomic DNA was extracted in triplicates from 0.5 g of the composite sample using the FastDNA Soil Kit (MP Biomedicals, CA, USA) and following the manufacturer’s protocol. DNA was pooled before being subjected to high-throughput sequencing on the NextSeq500 platform to generate a paired-end sequencing (2 × 150 bp paired-end chemistry) library using the Illumina TruSeq Nano DNA HT Library Preparation Kit. De novo assembly of high-quality data was accomplished using MetaVelvet (version 1.20.02) on optimized parameters. The assembled data were further curated for submission to NCBI. As per NCBI whole genome shotgun (WGS) sequence submission guidelines, the sequences <199 nt and containing ‘N’ at termini were removed using in-house Perl scripts. Subsequently, the metagenomes were submitted to WGS for deposition at the DDBJ/EMBL/GenBank under the BioProject with ID PRJNA230141. The metadata information can be viewed at NCBI with BioSample accession number SAMN02438464.

**In-silico bioinformatics analysis**

MG-RAST is a widely used online tool, being employed for metagenome sequence analysis (Meyer et al. 2008; Yadav et al. 2015). The metagenome sequence data of a rhizospheric soil sample from Phumdi were submitted to the MG-RAST (version 3.3.6) server for automated annotation and analysis. The data were subjected to quality control (QC) including quality filtering (removing sequences with > ambiguous nucleotide), length filtering (removing sequences more than 2.0 standard deviations from the mean sequences’ length) and dereplication (removing similar sequences that are artifacts of shotgun sequencing). For functional annotation of metagenome data, the hierarchical classification option was used as the data type and SEED-based subsystems as the annotation source (with parameters of identity cut-off – 60%, e-value cut off – 1×10⁻⁵ and alignment length cut-off – 15 for amino acids). Additionally, the taxonomy of oxygenase coding sequences was identified by BLAT against the M5NR protein database. Similarly, taxonomic annotations were assigned by using the parameters described above, against the M5NR annotation source and best classification as a data type. A rarefaction curve was generated by plotting a number of reads versus the number of species. This curve also yields a total number of species being annotated for the Phumdi metagenome data.

**Comparative metagenomics**

For comparative metagenome analysis, metagenome sequence data derived from other freshwater lakes, aquifers and freshwater lakes from the Antarctica region were compared with data derived from freshwater habitat with
eutrophic blooms and the LL metagenome. Overall eight metagenome data (Table 1), available in the public database of the MG-RAST server and corresponding to diverse freshwater habitats, were subjected to comparative metagenome analysis. The data, representing eutrophic blooms from Lake Erie and Grand Lakes, were considered as a reference for habitat experiencing extreme eutrophication. Steffen et al. (2022) have reported the microbial diversity of these lakes, which are especially dominated by Cyanobacteria, through metagenome analysis of cyanobacterial blooms. Aquifers are underground freshwater habitats, thus are less prone to anthropogenic pollution. Similarly, owing to a remote location in Antarctica, the freshwater lakes in this region would be unaffected by anthropogenic activities. Metagenome data for aquifers and lakes in Antarctica provided oligotrophic control in our comparative study, revealing microbial communities that predominate in healthy freshwater habitat with a pristine nature. A group parameter was selected for metagenomes representing a single habitat in order to remove inter-sample variability and make the data more comparable. Thus, four groups were generated representing Freshwater, Aquifers, Eutrophic blooms and lakes from Antarctica. Functional and taxonomic feature abundance profiles for four groups were generated by using similar parameters as described for the Phumdi metagenome data. Subsequently, four groups were compared with these data using a statistical tool to derive differential functional and taxonomical features.

**Statistical analysis**

Statistical analysis of the Metagenomic Profiles (STAMP) (version v2.0.3) package was employed to analyze statistically significant differential abundance of taxonomic and functional features among four groups with respect to Phumdi (Parks et al. 2014). A feature abundance table, created for each group and Phumdi at a functional and taxonomic level using MG-RAST, was imported to the STAMP tool. Principal component analysis (PCA) was carried out at the genus level of each group in order to compare freshwater habitats based on their microbial community composition. LL, being unique in possessing floating islands made up of profuse vegetation, the microbial community must vary in the rhizospheric soil compared to the remaining freshwater habitats. Thus, in order to identify the bacterial genera being predominant in the Phumdi, it was compared with the remaining groups by extended error bar analysis. Error bar plots were generated via two sample analysis, by applying Fischer’s exact test and Storey’s FDR multiple test correction. The results were curated to exhibit

<table>
<thead>
<tr>
<th>No.</th>
<th>Metagenome name</th>
<th>MG-RAST ID</th>
<th>Country, location and lake</th>
<th>Habitat</th>
<th>Total contigs</th>
<th>Post QC contigs</th>
<th>Base pair count</th>
<th>Predicted protein</th>
<th>Size (Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LL - EGD-LL4</td>
<td>4554336.3</td>
<td>India, Manipur</td>
<td>Freshwater</td>
<td>215,988</td>
<td>206,254</td>
<td>106,935,262</td>
<td>238,310</td>
<td>115.9</td>
</tr>
<tr>
<td>2</td>
<td>GS020 Shotgun – freshwater – Panama Canal – Lake Gatun – Panama</td>
<td>4441590.3</td>
<td>Panama, Lake Gatun</td>
<td>Freshwater</td>
<td>296,355</td>
<td>296,355</td>
<td>315,151,399</td>
<td>340,669</td>
<td>315.1</td>
</tr>
<tr>
<td>3</td>
<td>Metagenome name</td>
<td>4618288</td>
<td></td>
<td>Freshwater</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>681.6</td>
</tr>
<tr>
<td>4</td>
<td>Erie Bloom</td>
<td>4467029.3</td>
<td>USA, Erie Harbor, Lake Erie</td>
<td>Eutrophic bloom</td>
<td>504,233</td>
<td>354,384</td>
<td>153,986,130</td>
<td>264,567</td>
<td>204.4</td>
</tr>
<tr>
<td>5</td>
<td>Grand Lake St Marys (GLSM) Bloom</td>
<td>4467059.3</td>
<td>USA, Ohio</td>
<td>Eutrophic bloom</td>
<td>219,716</td>
<td>140,179</td>
<td>60,656,381</td>
<td>110,292</td>
<td>88.2</td>
</tr>
<tr>
<td>6</td>
<td>A5</td>
<td>4453083.3</td>
<td>Australia, South Australia</td>
<td>Aquifer</td>
<td>409,743</td>
<td>288,776</td>
<td>121,615,569</td>
<td>237,754</td>
<td>158.8</td>
</tr>
<tr>
<td>7</td>
<td>A1</td>
<td>4453064.3</td>
<td>Australia, South Australia</td>
<td>Aquifer</td>
<td>64,506</td>
<td>45,902</td>
<td>19,066,746</td>
<td>36,879</td>
<td>24.9</td>
</tr>
<tr>
<td>8</td>
<td>AntarcticaAquatic_5 – Marine Derived Lake</td>
<td>4443682.3</td>
<td>Antarctica, Ace Lake Vestfold Hills</td>
<td>Antarctica</td>
<td>283,663</td>
<td>283,663</td>
<td>284,069,722</td>
<td>314,206</td>
<td>284</td>
</tr>
<tr>
<td>9</td>
<td>AntarcticaAquatic_6 – Ace Lake</td>
<td>4443684.3</td>
<td>Antarctica, Ace Lake</td>
<td>Antarctica</td>
<td>281,490</td>
<td>281,490</td>
<td>281,056,691</td>
<td>306,354</td>
<td>281</td>
</tr>
</tbody>
</table>

Table 1 | Data summary of metagenome data used for comparative metagenomics
only features with a q-value <0.05 and >1% difference in relative abundance of a given genera between two samples. Subsequently, Venn analysis was applied to the lists of statistically significant abundant microbial genera corresponding to Phumdi soil habitat against the remaining groups using the Venny tool (Oliveros & Venny 2007). Similarly, selectively enriched functional features for Phumdi were retrieved in comparison to the remaining freshwater habitats.

### RESULTS AND DISCUSSION

LL has socio-economic importance to the local area and thus requires regular monitoring to assess its health. Chronic anthropogenic pollution in this habitat has posed a threat to its health. The microbial residents are the ones that look after and respond to any environmental abuse in an ecological habitat. Phumdi at this lake represent a unique habitat in which plant-microbe interaction collectively may confer a self-bioremmediation caliber against consistent anthropogenic pollution. This study aims at exploring the composition of the microbial counterpart and corresponding functional potential of this habitat. Metagenome sequence data were generated for a rhizospheric soil sample from Phumdi and subsequently analyzed in-silico. Furthermore, the comparative metagenomic approach was employed in the present study for metagenomes derived from lakes with different ecophysiological characteristics. Analyses led to the identification of bioindicators pertaining to the Phumdi rhizospheric site.

### Metagenome sequencing

The WGS project of Phumdi soil has been deposited at the DDBJ/EMBL/GenBank under the accession JAHE00000000 and consists of sequences JAHE01000001-JAHE01215988 (http://www.ncbi.nlm.nih.gov/nuccore/JAHE0000000). The version described in this paper is JAHE00000001. Table 2 illustrates sequence QC and annotation features for the Phumdi soil metagenome accomplished by the MG-RAST server. Figure 1(g) shows the rarefaction curve for microbial diversity corresponding to the Phumdi rhizospheric soil, indicating that sufficient sequencing depth was achieved in this study. However, it wasn’t exhaustive, as the rarefaction curve had not approached a plateau. As can be seen, about 3,100 microbial species are present in Phumdi soil in different proportions.

### Microbial community composition of Phumdi habitat

Metagenome analysis revealed a plethora of microbial communities in the Phumdi. From the results of the taxonomic classification, three major microbial domains were found: bacteria, archaea, and eukaryote. Bacterial domain dominated with 82% abundance, compared to the Eukaryota and Archaea domains (Figure 1). Additionally, as reported in earlier studies (Debroas et al. 2011; Smith et al. 2012), major bacterial groups were also annotated in the LL Phumdi metagenome. Figure 1(a)–1(f) also depict, at phylum level, classification of bacteria dominating the microbial abundance profile, Proteobacteria (51%) being the most dominant phylum followed by Acidobacteria (10%), Actinobacteria (9%) and Bacteroidetes (7%). With almost half of the total bacterial counts, all five classes of Proteobacteria phylum were detected, the major classes present being Alphaproteobacteria (18%), Betaproteobacteria (14%), Deltaproteobacteria (10%) and Gammaproteobacteria (7%), which ranked among the top five of the total bacterial classes. At the order level, Rhizobiales (14%) were found to be most dominant, indicating the abundance of bacterial activity belonging to the rhizosphere. Rhizobiales were followed by Burkholderiales (10%), Actinomycetales (8%), and Solibacteriales (5%) in total abundance. At the family level, Bradyrhizobiaceae (6%) dominated among all bacterial families. At the genus level, surprisingly Candidatus Solibacter (6%), belonging to the Acidobacteria phylum, was found to be most abundant. This indicates that although Proteobacteria dominated at the phylum level, the Acidobacteria phylum is also a major contributor at genus level in governing the functional potential of this ecosystem.

The results are in contrast to observations made by Eiler et al. (2014) in which the authors found that Actinobacteria dominate in temperate freshwater lake habitat, instead of Proteobacteria, as at Phumdi. In the Eiler et al. (2014) study, Betaproteobacteria was found to be most significant activity.
Figure 1 | Microbial diversity of Phumdi habitat. Pie charts show bacterial community distribution at respective taxonomic levels. (a) Domain, (b) Phylum (c) Class, (d) Order, (e) Family, (f) Genus and (g) the rarefaction curve of microbial diversity of Phumdi. After domain level, the top 10 taxa are illustrated in the pie chart and the remainder were grouped into ‘Others’.
prevailing bacterial class. In contrast, Alphaproteobacteria dominated in the class level, with Rhizobiales being the most prevalent order and Bradyrhizobiaceae the most dominant family in Phumdi. This lineage of bacteria exclusively includes plant-associated bacteria such as Bradyrhizobium, suggesting that the prevalence of a plant-associative bacterial group helps in nutrient cycling in this ecosystem. Candidatus Solibacter belonging to the Acidobacteria phylum dominated among the bacterial genera in relative abundance (5%). It has been hypothesized that Candidatus Solibacter has the ability to degrade complex polymers associated with vegetation, conferring it an advantage in environments with a low carbon concentration (Pearce et al. 2012). Moreover, five strains from the Acidobacteria phylum, isolated from Sphagnum peat bogs, were shown to be capable of hydrolysing polymers of plant origin such as pectin, xylan, laminarin, lichenan and starch (Pankratov & Dedysh 2010). On the other hand, Bradyrhizobium, belonging to the Rhizobiales order, is a well-established plant growth promoting bacteria. This group of bacteria can colonize the rhizosphere at several orders of magnitude greater than the surrounding bulk soil, and are capable of degrading xenobiotic compounds (de-Bashan et al. 2012). Furthermore, Bradyrhizobium is an endophyte, which can interact more intimately with its plant host. Plants benefit from their endophytes during phytoremediation of organic contaminants in soils, as they possess degradation pathways and metabolic capabilities not inherent in the plant. Such plant-microbe associations result in the most efficient degradation of pollutants. The analysis suggests that the Phumdi habitat selectively harbors an active microbial community that can better respond to high organic loads in the form of aromatics and other pollutants.

Extended error bar plot analysis by STAMP enabled comparative analysis of the dominant microbial features among selected freshwater habitats (Figure 2). In order to retrieve the dominant microbial feature of the Phumdi habitat, it was compared with the remaining three freshwater habitats’ metagenome data. This comparative metagenomic approach helped to uncover selective microbial genera and corresponding functional features of the Phumdi habitat. As depicted in Figure 2, error bar plots considered the relative abundance profile of bacterial genera in Phumdi compared to the remaining freshwater habitats. In comparison with pristine freshwater habitat, Candidatus Solibacter, Bradyrhizobium, Candidatus Koribacter, Geobacter, Pedosphaera and Acidobacterium were abundant in Phumdi (Figure 2(ia)). However, when compared to eutrophic habitat, bacterial genera Mycobacterium, Rhodopseudomonas, Anaeromyxobacter, Streptomyces, Methylobacterium, Opitutus, Sorangium, and Burkholderia were more prevalent in Phumdi, in addition to those listed in comparison with pristine habitats (Figure 2(ib)). Figure 2(ic) and Figure 2(id) show a similar comparison of Phumdi with aquifers and freshwater lakes from the Antarctica region respectively. Figure 2(iia)–2(id) illustrate similar analysis considering the relative abundance profile of the metabolic pathways corresponding to the selected datasets. The subsequent Venn analysis of the list of the dominant microbial attributes of the Phumdi habitat with respect to the remaining habitats yielded a list of common features through four comparisons (Figure 3). As can be inferred, Candidatus Solibacter, Bradyrhizobium, Candidatus Koribacter, Pedosphaera, Methylbacterium, Anaeromyxobacter, Sorangium, Opitutus and Acidobacterium are selectively prevalent in Phumdi. Similarly, metabolic pathways corresponding to monosaccharides, fatty acids, fermentation, carbohydrates, di- and oligosaccharides, peripheral and central pathways of aromatic compound catabolism and sulfur metabolism can be ascertained to prevail selectively in the Phumdi habitat.

Among nine bacterial genera, the bacteria from the Acidobacteria phylum such as Candidatus Solibacter, Candidatus Koribacter, and Acidobacterium have been reported to possess genes involved in the breakdown and utilization of diverse structural and storage polysaccharides (Ward et al. 2009; Rawat et al. 2012). Bradyrhizobium has been reported to biodegrade complex compounds such as 5-nitroanthranilic acid, indole and its derivatives, catechin and Ethyl tert-butyl ether (ETBE) (Hopper & Mahadevan 1997; Qu & Spain 2010; Le Digabel et al. 2013; Arora et al. 2013). Similarly, based on genome analysis of Opitutus terrae, Bawn reported many genes required in the biodegradation of lignocellulosic polysaccharides (Bawn 2012). Anaeromyxobacter dehalogenans has been characterized for its ability to dechlorinate chloro phenols under anaerobic conditions (Arora & Bae 2014). Thus, the Phumdi habitat can recycle not only plant originated organic compounds but also other xenobiotics. Consequently, it can be inferred that owing to the unique scenario presented by the Phumdi, some bacterial genera have been selectively enriched at this habitat, which has relevance to efficient bioremediation activity.

**Functional metagenome analysis at Phumdi**

The MG-RAST functional annotation pipeline characterized metagenomic sequences into 28 defined subsystems based on the SEED database, each representing a collection of functionally-related protein families belonging to a particular metabolic domain (Figure 4). As can be observed
Extended error bar plots demonstrate the differential distribution of microbial features in a given pair of habitats. (i) Error bar plots considering abundance of bacterial genera. (ii) Error bar plots considering abundance of functional features at level 2 of subsystem based hierarchical functional classification. (a) Differential representation of microbial features between Phumdi and freshwater habitat. (b) Differential representation of microbial features between Phumdi and eutrophic freshwater habitat (bloom). (c) Differential representation of microbial features between Phumdi and aquifer habitat. (d) Differential representation of microbial features between Phumdi and freshwater habitat from Antarctica region. Each extended error bar plot illustrates only those genera with a difference of at least 1% in their relative abundances at a respective pair of habitats.
all the major housekeeping metabolic subsystems were identified for this habitat. Owing to the abundant vegetation in the form of the Phumdi, the plant-derived alkaloids and other aromatic compounds are expected in abundance in this habitat. Hence, the subsystem of the metabolism of aromatic compounds was selectively analyzed in order to explore the microbial potential to deal with aromatic compounds. As can be observed, a variety of aerobic as well as anaerobic metabolic pathways relating to aromatic compound metabolism is present in the LL habitat (Figure 4, inset). Pathways involving n-phenylalkanoic acid degradation, anaerobic and aerobic benzoate catabolism, N-heterocyclic aromatic compound degradation, salicylate and gentisate catabolism, phenylpropanoid compound degradation, and the protocatechuic and catechol branch of the beta-ketoadipate pathway were found primarily at the Phumdi habitat.

The comparative metagenomics approach clearly differentiated the Phumdi habitat from other freshwater habitats depending on functional characteristics. Figure 5 supports this inference, depicting the prevalence of nutrient cycling at Phumdi related to carbohydrates, amino acids and derivatives, nitrogen and sulfur metabolism, metabolism of aromatic compounds, fatty acids, lipids, and isoprenoids. It indicates that owing to the unique habitat of Phumdi, creating a massive flow of organic matters at various stages of decomposition, microbes have selectively been chosen for these metabolic features. Statistical analysis by the STAMP tool helped the understanding of the distinct metabolic pathways, which are significantly differential in the Phumdi habitat. Pathways related to the peripheral and central pathways for aromatic compound degradation, monosaccharides, fatty acids, and fermentation were found to be significantly enriched by extended error bar plot analysis, as illustrated in Figure 2(ii). These
results corroborate with inferences from Figures 4 and 5, collectively suggesting an efficient potential for bioremediation of the resident microbial community.

MG-RAST BLAT results, containing species information, were used to assign the oxygenases to specific bacteria at the genus level. Figure 6 shows the diversity of oxygenase coding bacteria in the Phumdi habitat. The analysis revealed that Bradyrhizobium and Burkholderia are the two major bacteria involved in aromatic compound degradation of this habitat. This also corroborates with the results illustrated in Figure 5.
Analysis to compare Phumdi at LL with other freshwater habitats

The comparative metagenomics approach helped in identifying bacterial genera selective to Phumdi compared to the remaining freshwater habitats. PCA further supported inferences from the metagenome analysis, illustrating similarity in the microbial community composition of the Phumdi habitat with freshwater habitats that have a comparatively pristine nature (Figure 6). However, based on PCA of functional categories, the Phumdi showed unique characteristics compared to the remaining habitat, attributed to the unique habitat of the Phumdi.

Figure 6 illustrates PCA plots for four selected freshwater habitats and Phumdi metagenome data. The PCA corresponding to the community composition of the selected habitats illustrates that LL is close to aquifer and freshwater habitats with relatively healthy ecology (Figure 6(a)). This suggests that instead of regular anthropogenic activities, LL, owing to its bioremediation potential, has managed to maintain its healthy ecophysiology. Highly eutrophic habitat with characteristics of cyanobacterial bloom appeared distant from the remaining freshwater habitats, indicating that those have a differential composition of the microbial community. The PCA plot based on the functional characteristics of the selected habitats corroborated with that of taxonomy based analysis, although with little difference (Figure 6(b)). Here LL showed unique functional characteristics to the remaining habitats. Figure 7 depicts comparative bar charts for...

![Figure 6](https://iwaponline.com/wst/article-pdf/74/9/2075/457582/wst074092075.pdf)

**Figure 6** | Principal component analysis (PCA). (a) PCA plot illustrating correlation among eight freshwater habitat metagenome samples based on microbial composition at genus level. (b) PCA plot illustrating correlation among eight freshwater habitat metagenome samples based on level three of a subsystem-based hierarchical functional classification.

![Figure 7](https://iwaponline.com/wst/article-pdf/74/9/2075/457582/wst074092075.pdf)

**Figure 7** | BLAT Distribution of oxygenase among bacterial genera. The pie chart illustrates the distribution of oxygenase among bacterial genera in the Loktak Lake habitat corresponding to a subsystem of metabolism of aromatic compounds.
subsystems that are selectively preponderant in the Phumdi compared to the remaining freshwater habitats.

CONCLUSIONS

In conclusion, the metagenome sequence analysis based approach proved to be helpful in identifying microbial diversity and its functional attributes characteristic to Phumdi in LL. The comparison of the Phumdi metagenome with other freshwater habitats with differential ecophysiological states revealed the unique microbial composition of this habitat. Compared to highly eutrophic freshwater lakes, the microbial community at Phumdi was found to be close to habitats with relatively pristine conditions. This was attributed to the presence of unique habitat in Phumdi that, owing to the unique plant-microbe interactions, might be efficiently performing the phenomenon of bioremediation at LL. Thus, in spite of regular anthropogenic activities at this lake, by virtue of the Phumdi and its unique microbial community composition, it has succeeded in maintaining its ecophysiological health. In addition to LL habitat, bioindicators for eutrophic conditions were also identified by a comparative metagenomics approach. Such bioindicators can be exploited in monitoring strategies to follow the eutrophication process.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Dhananjay Raje, MDS Bio-analytics Pvt. Ltd, Nagpur for statistical support. Two of the authors, Sampada Puranik and Rajesh R. Pal, would like to thank CSIR and UGC respectively for a senior research fellowship. This work was supported by a grant from CSIR Project ESC0108.

REFERENCES

sp. IFP 2042 and Bradyrhizobium sp. IFP 2049 isolated from a polluted aquifer. *Applied Microbiology and Biotechnology* 97 (24), 10531–10539.


Pankratov, T. A. & Dedysh, S. N. 2010 *Granulicella paladica* gen. nov., sp. nov., *Granulicella pectinivorans* sp. nov., *Granulicella aggregans* sp. nov. and *Granulicella rosea* sp. nov., acidophilic, polymer-degrading acidobacteria from *Sphagnum* peat bogs. *International Journal of Systematic and Evolutionary Microbiology* 60 (12), 2951–2959.


