



# Amplicon sequencing and imputed metagenomic analysis of waste soil and sediment microbiome reveals unique bacterial communities and their functional attributes

Surajit De Mandal · Vabeiryureilai Mathipi ·  
Rajendra Bose Muthukumaran ·  
Guruswami Gurusubramanian ·  
Esther Lalnunmawii · Nachimuthu Senthil Kumar

Received: 29 April 2019 / Accepted: 10 October 2019 / Published online: 29 November 2019  
© Springer Nature Switzerland AG 2019

**Abstract** The discharge of solid and liquid waste from domestic, municipal, and hospital premises pollutes the soil and river ecosystems. However, the diversity and functions of the microbial communities present in these polluted environments are not well understood and may contain harmful microbial communities with specialized metabolic potential. In this present study, we adapted the Illumina sequencing technology to analyze microbial communities and their metabolic capabilities in polluted environments. A total of 1113884 sequences of v3–v4 hypervariable region of the 16S rRNA were obtained using Illumina sequencing and assigned to the corresponding taxonomical ranks using Greengenes

databases. Proteobacteria and Bacteroidetes were dominantly present in all the four studied sites (solid waste dumping site (SWD); Chite river site (CHR), Turial river site (TUR), and Tuikual river site (TUKR)). It was found that the SWD was dominated by *Firmicutes*, *Actinobacteria*; CHR by *Acidobacteria*, *Verrucomicrobia*, *Planctomycetes*; TUR by *Verrucomicrobia*, *Acidobacteria*; and TUKR by *Verrucomicrobia* and *Firmicutes*, respectively. The dominant bacterial genus present in all samples was *Acinetobacter*, *Flavobacterium*, *Prevotella*, *Corynebacterium*, *Comamonas*, *Bacteroides*, *Wautersiella*, *Cloacibacterium*, *Stenotrophomonas*, *Sphingobacterium*, and *Pseudomonas*. Twenty-seven putative bacterial pathogens were identified from the contaminated sites belonging to *Salmonella enterica*, *Pseudomonas aeruginosa*, *Escherichia coli*, and *Staphylococcus aureus*. Functional analysis showed a high representation of genes in the KEGG pathway involved in the metabolism of amino acids and carbohydrates and identified several genes associated with antibiotic resistance and xenobiotic degradation in these environments, which can be a serious problem for human health and environment. The results from this research will provide a new understanding of the possible management practices to minimize the spread of pathogenic microorganisms in the environment.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10661-019-7879-0>) contains supplementary material, which is available to authorized users.

S. De Mandal · V. Mathipi · E. Lalnunmawii ·  
N. S. Kumar (✉)  
Department of Biotechnology, Mizoram University, Aizawl,  
Mizoram 796004, India  
e-mail: nskmzu@gmail.com

S. De Mandal  
College of Agriculture, South China Agricultural University,  
Guangzhou, China

R. Muthukumaran  
Department of Chemistry, Mizoram University, Aizawl, Mizoram  
796004, India

G. Gurusubramanian  
Department of Zoology, Mizoram University, Aizawl, Mizoram  
796004, India

**Keywords** Waste discharge · Bacterial community ·  
Pathogen · Antibiotic-resistant gene · Xenobiotic

## Introduction

With a rapidly growing population, solid and liquid wastes are increasing at an alarming rate in global ecosystems (Guerrero et al. 2013). Limited resources and the lack of proper management systems create dumping sites and pollute river ecosystems, which leads to the spread of human pathogenic bacteria and antibiotic resistance (AR) genes in both terrestrial and aquatic ecosystems (Blomström et al. 2016). However, knowledge of pathogenic bacteria in the natural environment is limited. Thus, an in-depth microbiological investigation is important for understanding the risks associated with such environments (Mwaikono et al. 2016). Most of the bacterial species are not suitable for laboratory cultivation, and therefore information on pathogenic species present in contaminated ecosystems is limited.

High-throughput sequencing (HTS) allows for in-depth metagenome research and can address the entire bacterial communities in any ecosystem. It has been successfully applied to monitor bacteria communities and their metabolic potentials in diverse environments (Panda et al. 2015; De Mandal et al. 2017). Deep amplicon sequencing of 16S rRNA has been used to detect rare species (Welch and Huse 2011). An analysis of the contaminated environmental metagenome may be useful for assessing pathogenic strains present in the environment. Sequencing of such metagenomes would also lead to the discovery of novel bacterial taxa, virulence genes, and other mobile genetic elements.

Mizoram, a state in Northeast India, is known as a “biodiversity hotspot” and has a unique ecology and geography with mountainous landscape. The Mizo-tribal community has a municipal waste management system, but due to the hilly terrain, the wastes are often mixed with streams and rivers. Discharges from hospitals, dumping sites, and domestic waste cause significant water pollution, which leads to an increase in infectious diseases to the local population and animals. To date, no studies have been conducted to evaluate the bacterial populations present in these contaminated sites. Therefore, it is important to determine the dominant bacterial communities, their diversity, and distribution in these environments.

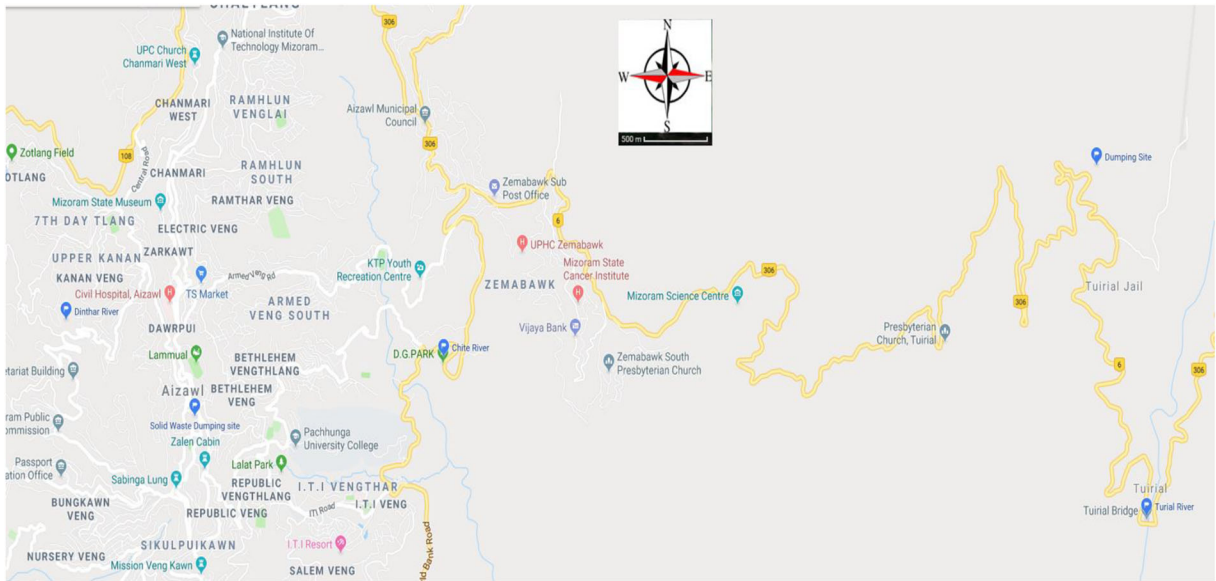
In the present study, we examined the bacterial community structure and diversity using paired-end Illumina sequencing of V3–V4 regions of 16S rRNA in the domestic and hospital wastes in polluted soil (dumping land) and sediment (river) of Mizoram, North East India. The aim of this study was to analyze the bacterial communities and their imputed metabolic potentials in the contaminated sites. Questions of interest include:

1. What are the bacterial communities and potential human bacterial putative pathogens occurring in contaminated sites?
2. What are the functional roles, especially the antibiotic resistance and xenobiotic degrading genes prevailing in these environments?

## Materials and methods

### Site selection and soil sampling

Soil and sediment samples were collected from three rivers and one dumping site contaminated by hospital and domestic waste in the Aizawl city of Northeast India. Aizawl city is the state capital of Mizoram, Northeast India, located at an average altitude of about 1132 miles. The population according to the 2011 census was 293,416. Based on the data on the drainage system and sewage management of the Public Health Department and Aizawl Municipal Council, four sampling sites were selected where maximum pollution could occur. Chite river site (CHR) is contaminated by discharges from Hospital 1, Hospital 2, Hospital 3 and Hospital 4. Turial river site (TUR) is largely polluted by Hospital 5 and Hospital 6. Tuikual river site (TUKR) is contaminated by the discharges from the Hospital 7, and most of the solid wastes from the Aizawl city are disposed at the solid waste dumping site (SWD) (Figs. 1 and 2). Soils and sediments (about 50 g) were sampled from three contaminated rivers and one dumping site from 0 to 10 cm depth. Ten replicate samples were collected at different locations for four individual sites and mixed uniformly to make a composite sample. All the collected samples were transferred to ethanol-disinfected core tubes inside the ice box and transferred into the laboratory, and kept at  $-80^{\circ}\text{C}$  until processed.

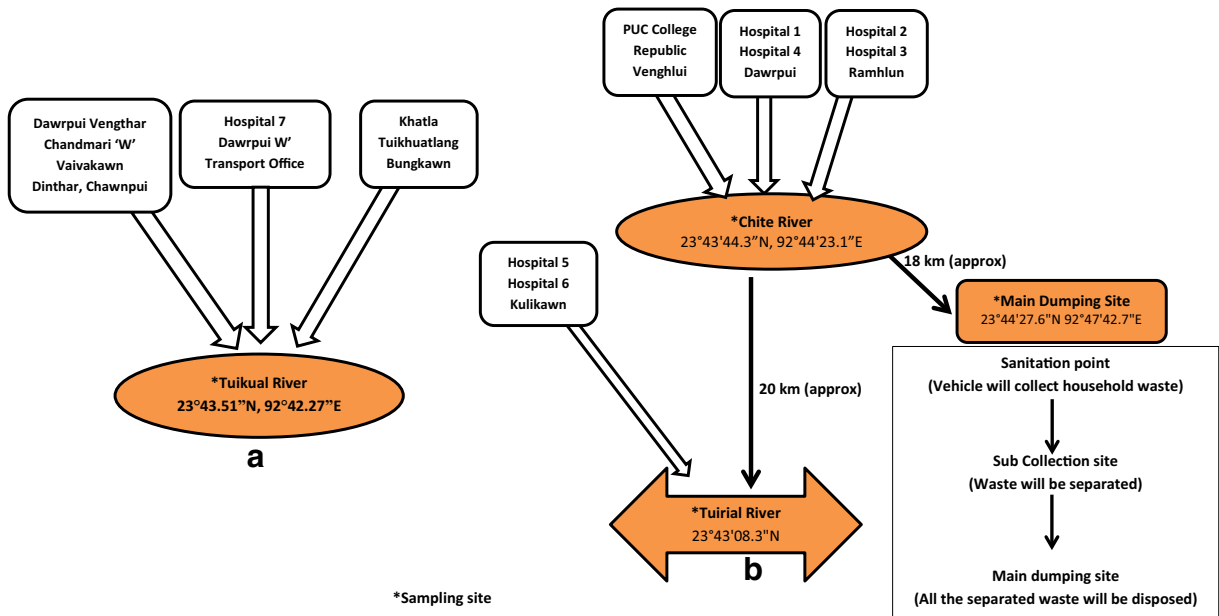


**Fig. 1** Geographical location of sampling sites: Tuikual river site (TUKR), Chite river site (CHR), Tuirial river site (TUR), and solid waste dumping site (SWD).

DNA isolation and Illumina sequencing

Bacterial DNA was isolated from approximately 500 mg of soil and sediment samples using the Fast DNA spin kit (MP Biomedical, Solon, OH, USA) according to the manufacturer’s protocol. The DNA concentration was quantified using a microplate reader (Spectra Max 2E, Molecular

Devices, CA, USA). Each sequenced sample was prepared according to the Illumina 16S Metagenomic Sequencing Library protocols. The quantification of DNA and the DNA quality was measured by Pico Green and Nanodrop. Amplification was performed using the forward primer 341F (TCGTCGGCAGCGTCAGATGTGTATAA GAGACAG) with locus-specific V3 sequence



**Fig. 2** Schematic representation of major households effluent and domestic waste flow. a Eastern site of Aizawl City. b Western site of Aizawl City (Public Health Dept., Govt of Mizoram, 2017).

(CCTACGGGNGGCWGCAG) and reverse primer 805R (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG) with locus-specific V4 sequence (GACTACHVGGGTATCTAATCC) (Klindworth et al. 2013). Genomic DNA (2 ng) was amplified using 5 pmol forward-tailed target-specific primer, 5 pmol reverse-tailed target-specific primer, and Hercules II polymerase (Agilent). PCR was carried out in triplicate for each variable region in a 25- $\mu$ l reaction as follows: 95 °C for 3 min, followed by 25 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s, and a final elongation step at 72 °C for 5 min. The amplicon libraries were removed of excess nucleotides, salts, and enzymes using 20  $\mu$ l of Agencourt AMPure XP system (Beckman Coulter Genomics) and eluted in 25  $\mu$ l TE buffer. Ten microliters of this first step reaction was subjected to a second amplification step following Nextera XT Index Primer (N7xx), Nextera XT Index Primer (S5xx), and Hercules II polymerase (Agilent). PCR for each variable region was carried out in triplicate in a 25- $\mu$ l reaction as follows: 95 °C for 3 min, followed by 10 cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s, and a final elongation step at 72 °C for 5 min. The amplicon libraries were cleaned to avoid excess nucleotides, salts, and enzymes using 20  $\mu$ l of the Agencourt AMPure XP system (Beckman Coulter Genomics) and eluted in 25  $\mu$ l of TE buffer. The final purified product was quantified using qPCR (KAPA Library Quantification kits for Illumina Sequencing platforms) and qualified using the TapeStation DNA screentape D1000 (Agilent). Paired-end Illumina sequencing of the V3–V4 hypervariable region of the 16S rRNA gene was performed (Macrogen Inc., Korea), and raw reads were deposited into the NCBI (SRA accession number SRR7079871–SRR7079874).

### Analysis of Illumina sequences

The adapter sequences were removed from the raw fastq sequences using Scythe, and low-quality bases were trimmed using Sickle v1.33 (Joshi and Fass 2011). The adapter trimmed sequences were analyzed using QIIME software package v.1.8.0 (Caporaso et al. 2010). Sequences with poor quality were filtered using the `split_libraries` command. USEARCH was used to remove chimeric sequences followed by deletion of singletons (Edgar et al. 2011). The consensus sequences were clustered into operational taxonomic units (OTUs) based on their sequence similarity by the Uclust program (similarity cutoff = 0.97) (Edgar 2010). Finally,

the representative sequence of each OTU was aligned to the Greengenes core set reference database using the PyNAST program and classified using Greengenes database (DeSantis et al. 2006). Shannon diversity and observed species indices were calculated from the Illumina dataset by QIIME software.

### Human pathogen identification from Illumina sequencing data

The bacterial pathogen database (HBPD) was used for the search for human bacterial pathogens from Illumina sequencing reads (Cai and Zhang 2013). All the pre-processed V3–V4 16S-rRNA Illumina reads obtained in this study were compared against HBPD database for the sequence similarity using the NCBI-BLAST algorithm. Only sequences with a blast hit with  $\geq 97\%$  identity were considered for the analysis (Banskar et al. 2016).

### Predicted metagenome analysis

Phylogenetic investigation of communities by reconstruction of unobserved states (PICRUSt) (Langille et al. 2013) was used to predict the community function in the contaminated sites based on a constructed phylogenetic tree of 16S rRNA marker gene sequences. The OTU table obtained by `pick_closed_reference_otus.py` script against the Greengenes database was employed for predicted 16S rRNA gene copy number normalization, followed by metagenome functional analysis. The resulting table was collapsed at KO level 3 within the pathway hierarchy of KEGG using the `categorize_by_function.py` script.

### Statistical analysis

The Shannon index represents OTU abundance and estimates both richness and evenness, while the observed species metric indicates only the unique OTUs present in the samples. In this study, beta diversity between the bacterial communities was determined by weighted and unweighted UniFrac (Lozupone and Knight 2005). Weighted and unweighted UPGMA trees were constructed by performing Jackknife test. Heatmap was used to determine the relative abundance of the top 50 bacterial genus using the CIMminer tool (<http://discover.nci.nih.gov/>). Differences in the metabolic potential of soil and sediment samples were examined using Welch's *t* test in the STAMP software.

**Result and discussion**

Environmental monitoring is an important aspect for understanding the processes and activities of an environment and is also used to assess the human activities which carry a risk of harmful effects on the natural environment. Next-generation sequencing (NGS) technologies have been applied in various complex microbial environments to answer diverse ecological questions and have enabled scientists to routinely use this technology for environmental monitoring (Kwon et al. 2011). In the present study, we employed high-throughput Illumina sequencing method to characterize the taxonomical and imputed functional composition in four different sites contaminated with domestic and or hospital wastes. A total of 1,113,884 16S rRNA sequence tags were obtained by Illumina sequencing with an average length of 200 bp (Table 1). The pre-processed high-quality Illumina sequences ranged from 37,999 (SWD) to 42,387 sequences (CHR) with a mean of 39,985 and SD 1791. A detailed description of the sequence statistics is provided in Table 1.

**Bacterial diversity**

NGS method identified the abundance and diversity of bacteria in soil and sediment samples of Mizoram, Northeast India. Rarefactions were performed to a depth of 37,999 reads. Rarefaction curves for richness in all four samples plateau at the maximum depth, indicating an adequate sampling procedure (Fig. 3). Shannon index suggests that the microbial diversity is different between the study sites, with the lowest diversity associated with SWD (8.58) and highest in CHR (11.18). Chao richness estimates suggest that the greatest number of the least number of sequences were captured in TUR sediment

**Table 1** Raw read statistics obtained from Illumina and nanopore sequencing

Sample ID	Illumina dataset	
	Total no. of reads	Total size (in bp)
CHR	283,120	85,219,120
TUKR	264,644	79,657,844
SWD	268,060	80,686,060
TUR	298,060	89,701,612

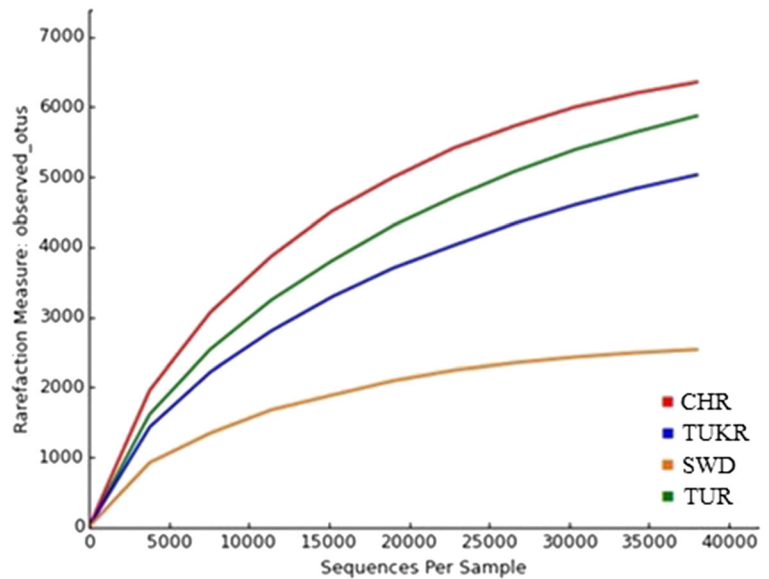
TUKR Tuikual river site, CHR Chitae river site, TUR Turial river site, SWD solid waste dumping site

and lowest in SWD (Table S1). A total of 20108 OTU was found ranging from 2537 (SWD) to 6492 (CHR). Both soil and sediment samples contain high bacterial diversity. A large number of untreated domestic waste are discarded into these studied sites, which may serve as an inoculum to increase the overall bacterial diversity in these environments. Differences between the bacterial communities were carried out using principal coordinate’s analysis (PCoA) and UPGMA hierarchical clustering analysis using QIIME. PCoA (Fig. 4a, b) revealed that the microbial communities were different in the studied samples. Bacterial communities present in TUR and TUKR were more closely related than the CHR and SWD samples. Hierarchical clustering analysis (Fig. 4) using both soil and sediment samples showed that the microbial communities are significantly different from each other. The differences in the alpha diversity, as well as bacterial communities between the samples, may be due to the variation in physiochemical parameters (Ibekwe et al. 2013).

**Composition of the bacterial communities**

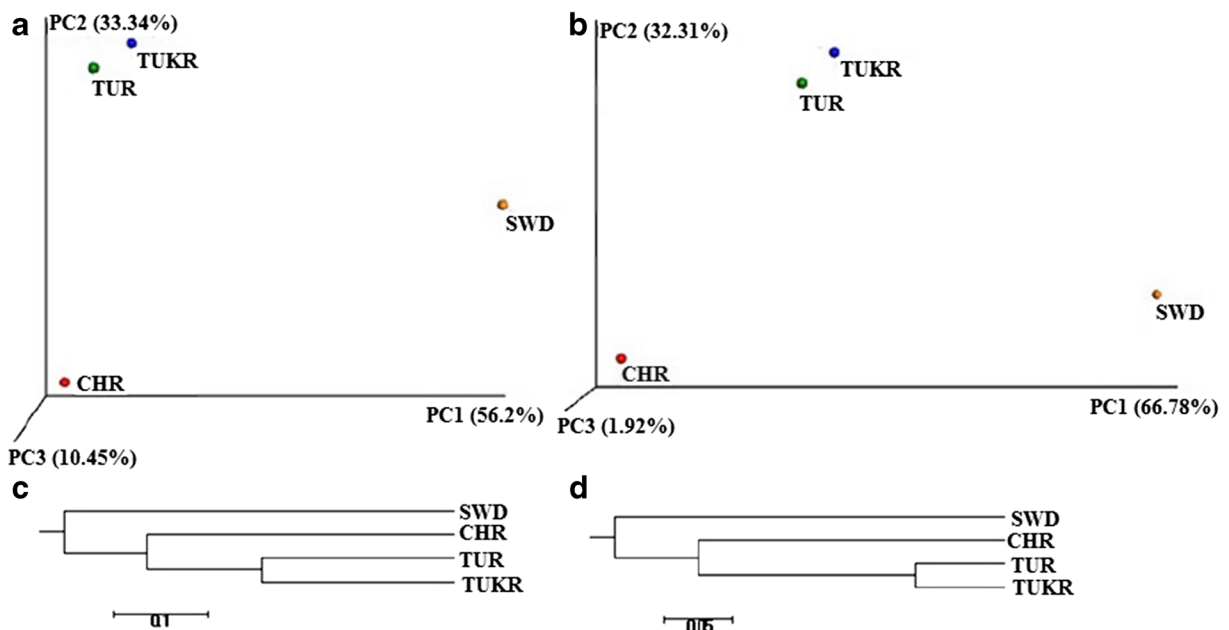
Domestic or hospital waste has a major impact on microbial populations (Liu et al. 2015, 2018). In the present study, 50 bacterial phyla were identified in the soil and sediment samples using Illumina sequencing (Fig. 5). Bacterial members (phyla, genus, species etc.) with a higher relative abundance in the complete dataset are referred to as dominant bacteria. The most dominant bacterial phyla present in these samples were *Proteobacteria* (48.1%), *Bacteroidetes* (23.6%), *Acidobacteria* (4.9%), *Verrucomicrobia* (4.20 %), *Firmicutes* (3.1 %), *Actinobacteria* (2.6 %), *Chloroflexi* (1.90 %), *Planctomycetes* (1.9%), and OD1 (1.3%). Other bacterial phyla were present in < 1% of the total bacterial community. Abundance of *Gammaproteobacteria* and *Bacteroidetes* were previously identified in municipal waste landfill sites (Song et al. 2015). Presence of these phyla was commonly found in anaerobic ecosystem and plays a major role in organic matter degradation and carbon cycling. Phyla such as *Firmicutes* (8.8%) and *Actinobacteria* (6.7%) were dominated in the dumping site; *Acidobacteria* (14.1%), *Verrucomicrobia* (7.6%), and *Planctomycetes* (5.3%) in Chitae river sediment; *Verrucomicrobia* (4.5%) and *Acidobacteria* (3.8%) in Turial river sediment, and *Verrucomicrobia* (4.6%) and *Firmicutes* (1.8%) in Tuikual river sediments. The presence of the

**Fig. 3** Rarefaction curve of the samples (cutoff of 3%): alpha-rarefaction plots were generated in QIIME using the observed species metric for estimating alpha diversity.



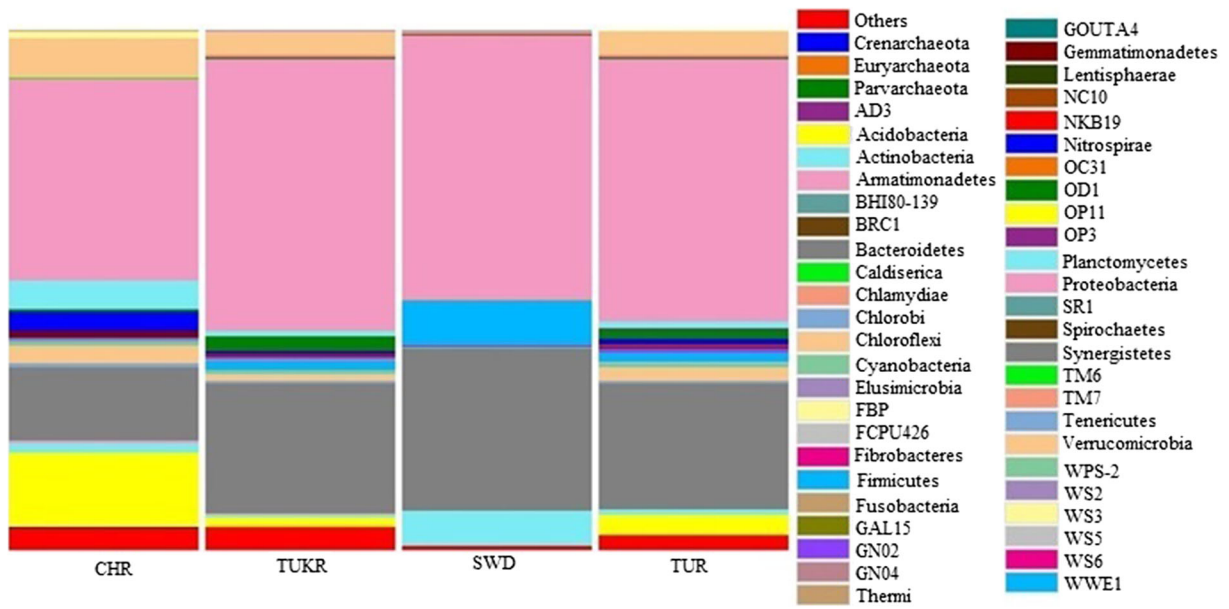
bacterial phyla *Firmicutes* indicates that possible anaerobic and methanogenic decomposition processes are likely to occur in dumping sites (Song et al. 2015). Comparison of soil and sediment samples revealed that soil samples were dominated with the bacterial phyla *Proteobacteria* (50.90%), *Bacteroidetes* (31.10%),

*Firmicutes* (8.80%), and *Actinobacteria* (6.70%) whereas sediment samples were dominated by *Proteobacteria* (47.00%), *Bacteroidetes* (20.90%), *Acidobacteria* (6.70%), *Verrucomicrobia* (5.60%), *Chloroflexi* (2.50%), *Planctomycetes* (2.50%), OD1 (1.80%), *Nitrospirae* (1.70%), *Firmicutes* (1.20%), and



**Fig. 4** Principal coordinate analysis (PCoA) and UPGMA cluster obtained using the UniFrac distance matrix. **a** The PCoA showed the unweighted Unifrac distance. **b** The PCoA showed the weighted Unifrac distance. **c** UPGMA cluster based on the unweighted Unifrac measurements. **d** UPGMA cluster based on the weighted

Unifrac measurements. The variance explained by each principal coordinate axis is shown in parentheses. Datasets were subsample to equal depth prior to the UniFrac distance computation. Legend: TUKR Tuikual river site, CHR Chitae river site, TUR Turial river site, SWD solid waste dumping site.



**Fig. 5** Relative abundance of the bacterial community at Phylum level. The relative abundances at the bacterial community at the phylum level were calculated using QIIME. Each bacterial family

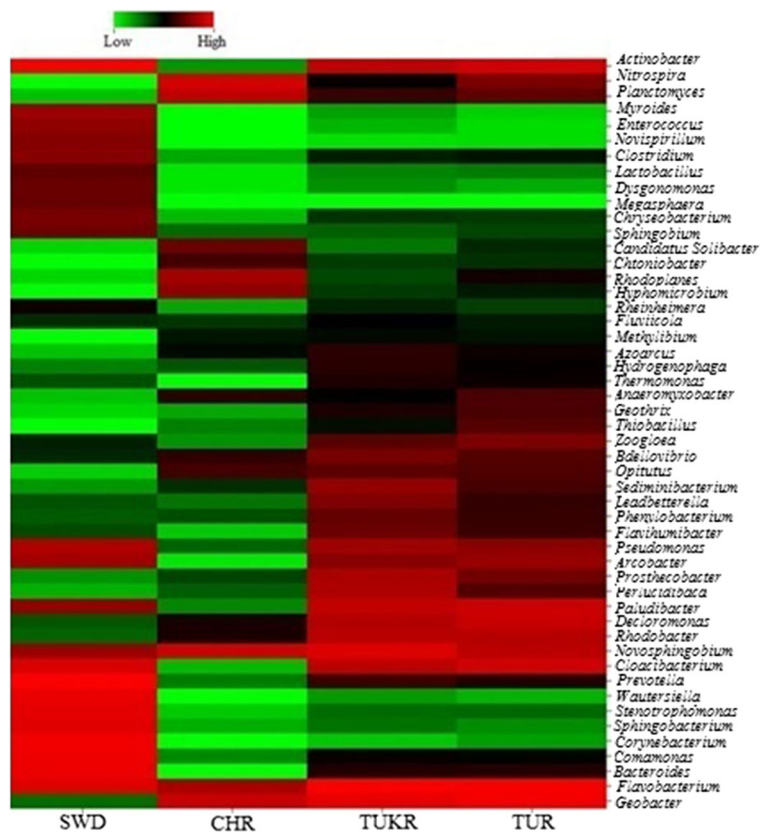
is represented as a different color in the bar graphs below. Legend: TUKR Tuikual river site, CHR Chitae river site, TUR Tural river site, SWD solid waste dumping site.

*Actinobacteria* (1.20%) (Fig. S1). Presence of *Verrucomicrobia* in all the sediment samples is consistent with the previous literatures, but little is known about their physiology (Shen et al. 2017) (Fig. 2).

The relative richness of the 95 bacterial genera identified at these four sites showed that the most abundant genus was *Flavobacterium* (4.1%), *Acinetobacter* (3.7%), *Geobacter* (2.5%), *Prevotella* (1.8%), *Novosphingobium* (1.70%), *Cloacibacterium* (1.4%), *Corynebacterium* (1.1%), *Bacteroides* (1.1%), *Paludibacter* (1.1%), and *Comamonas* (1.1%). A heat map of the top 50 bacterial genera is shown in Fig. 6. Dominant bacterial genera present in > 1% of the total bacterial community in both soil and sediment samples were *Acinetobacter*, *Flavobacterium*, *Prevotella*, *Corynebacterium*, *Comamonas*, *Bacteroides*, *Wautersiella*, *Cloacibacterium*, *Stenotrophomonas*, *Sphingobacterium*, and *Pseudomonas*, whereas soil samples were specifically dominated by the bacterial genera *Geobacter*, *Novosphingobium*, and *Paludibacter* and sediment samples with *Arcobacter* and *Myroides*. Members of the genus *Geobacter* have activity for the reduction of metal and sulfur (Yan et al. 2012). Further long-read nanopore sequencing identified two species under this genus: *G. metallireducens* and *G. sulfurreducens*. Both the species were involved in metal reduction (Lovley et al. 1993; Caccavo et al.

1994). The second most dominant genus in the river sediments was *Novosphingobium*, a gram-negative bacterium involved in the degradation of polycyclic aromatic hydrocarbons (Sohn et al. 2004). Whereas *Paludibacter* is a gram-negative, strictly anaerobic, chemoorganotrophic and non-motile genus, its possible function in the river sediment is not yet known (Qiu et al. 2013). Identification of the genus *Dechloromonas* indicates the oxidation process of aromatic compounds in river sediments (Coates et al. 2001). *Nitrospira* is found in different ecosystems and involved in the nitrogen cycle (Lucker et al. 2010; Van Kessel et al. 2015). The presence of *Nitrospira* in the sediment samples suggests that nitrification may occur in these environments. Members under the genus *Cloacibacterium* are facultative anaerobic bacteria and have been identified in wastewater (Allen et al. 2006). The present study also identifies the genus *Acinetobacter*. Members under this genus were previously reported from healthcare-associated infections. They also have the ability to acquire AR much faster than other gram-negative organisms (Manchanda et al. 2010). *Prevotella* is typically found in human gut microbiome, and the presence of this genus in the soil ecosystem indicates the possible fecal contamination through domestic waste (Tanaka et al. 2008). The Illumina dataset also detected a large number of *Stenotrophomonas maltophilia*. Members

**Fig. 6** Heatmap showing the top 50 major bacterial genera: TUKR Tuikual river site, CHR Chitae river site, TUR Turial river site, SWD solid waste dumping site



under this species are opportunistic pathogen with multidrug-resistant properties against multiple antibiotics (Brooke 2012).

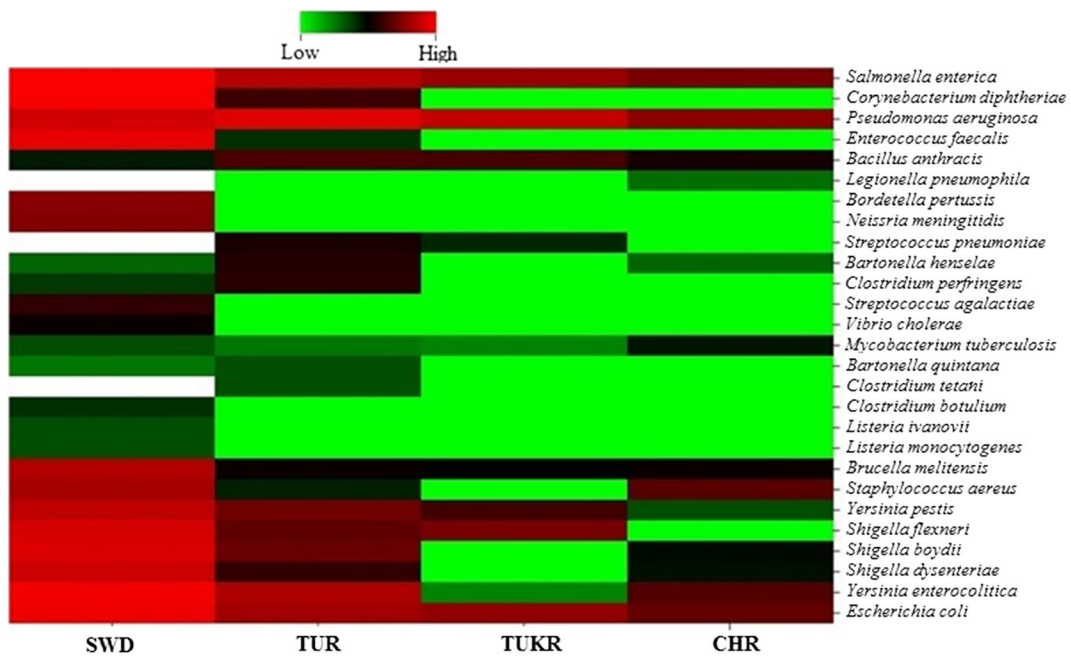
#### Human pathogen in different samples

Improper management of domestic and hospital wastes is one of the main reason for the occurrence of pathogens in different ecosystems (Watson et al. 2016). Understanding pathogens in the natural environment provides direct evidence of possible health risks and helps to develop preventive action strategies (Yang et al. 2008; Ivanek et al. 2009; Ibekwe et al. 2013). In the present study, the presence of putative bacterial pathogens in domestic- and hospital waste-contaminated soil and sediment samples was evaluated using the HBDP database. This database contains the 16s rRNA gene sequences derived from the well-characterized bacterial pathogens (Cai and Zhang 2013). A similar approach was used to identify pathogens from bat intestine using Ion Torrent sequencing of the V3 region of 16s rRNA

(Banskar et al. 2016). Twenty-seven different species of well-known bacterial pathogens were identified with 97% sequence similarity cutoff and  $\geq 99\%$  query coverage. The main pathogenic strains dominantly present in all the sites were *Salmonella enterica*, *Pseudomonas aeruginosa*, *Escherichia coli*, and *Staphylococcus aureus*. Other dominant species found in SWD were *Corynebacterium diphtheriae*, *Enterococcus faecalis*, *Brucella melitensis*, *Yersinia enterocolitica*, *Y. pestis*, *Shigella boydii*, and *S. flexneri*. *Yersinia enterocolitica* and *S. flexneri* were also abundantly found in TUR and TUKR, respectively (Fig. 7).

In the present study, major putative bacterial pathogen was *Salmonella enteric*, which causes a spectrum of diseases, including typhoid and gastroenteritis (Giannella 1996). Two species of *Yersinia*, *Y. enterocolitica* and *Y. pestis*, have also been identified. While *Y. enterocolitica* can cause yersiniosis and diarrhea in human, *Y. pestis* causes plague (Kenneth and Ray 2004; Fàbrega and Vila 2012). The presence of these species indicates the possible transmission of





**Fig. 7** Heatmap showing the major bacterial pathogens present in the soil and sediment samples. Legend: TUKR Tuikual river site, CHR Chitae river site, TUR Turial river site, SWD solid waste dumping site

pathogenic bacterial species through animal activities in the dumping land. Another identified pathogen, *P. aeruginosa*, has been reported to cause several nosocomial and opportunistic infections (De Bentzmann and Plésiat 2011). Two important bacterial pathogenic species, *Corynebacterium diphtheria* and *Mycobacterium tuberculosis*, have also been observed in the Illumina dataset (Kenneth and Ray 2004). Both the genera *Corynebacterium* and *Mycobacterium* can spread through the droplet nuclei and have been reported from the sediments affected by urban runoff (Kenneth and Ray 2004). An abundant number of the bacterial pathogen, *Enterococcus faecalis*, was identified mainly in the dumping soil. Members under this species can acquire AR genes from a medical waste-contaminated environment that plays a major role in pathogenesis (Kenneth and Ray 2004). Three species under the genus *Shigella* were identified (*S. flexneri*, *S. boydii*, and *S. dysenteriae*) in abundant number at dumping sites compared to sediment samples. Members under this genus were involved in several human and animal diseases and also have the ability to acquire multidrug resistance (García et al. 2010; Doumith et al. 2012; Fritah et al. 2014; Mwaikono et al. 2016). The association of this genus with dumping sites was previously studied by Mwaikono et al. (2016), and they postulated

that interaction between animals, humans, and microbes through the dumping sites could be a public health risk factor (Mwaikono et al. 2016).

Predictive functional role of the bacterial community

The PICRUSt tool used in this study may provide useful insights about the imputed functions of the bacterial community (Langille et al. 2013). Based on the prediction, KEGG database categorized the identified genes under the putative pathways involved in metabolism (50.02 %), genetic information processing (16.23 %), environmental information processing (12.90 %), cellular processes (4.37 %), organismal systems (0.79 %), and human diseases (1.10 %). A high representation of amino acid metabolism (20.780 %), carbohydrate metabolism (19.97%), and energy metabolism (11.87%) was observed. In addition, high proportions of gene families involved in lipid metabolism (7.46 %), metabolism of cofactors and vitamins (8.456 %), nucleotide metabolism (6.42 %), and xenobiotics biodegradation and metabolism (6.75%) were also predicted (Fig. S2). These dominant putative pathways present in bacterial communities were vital for the survival of bacterial communities (Erickson et al. 2012; Lamendella et al. 2011).

AR in pathogenic bacteria has become a worldwide problem caused primarily by the abuse and misuse of antibiotics, and a high prevalence of AR genes has been commonly found in contaminated ecosystems including river sediments (Goossens et al. 2005). In the present study, selected sites were contaminated with domestic and hospital waste. Therefore, it is important to study the AR genes in these environments (Allen et al. 2009). The presence of abundant putative AR genes involved in beta-lactam antibiotic resistance (K01207, K01467, K03585, K03587, K05366, K05515, K08218, and K12340) was found primarily 51 taxonomically diverse families (Table S2). In the sediment samples, these genes had a higher predicted prevalence among the bacterial family *Bacillaceae*, *Bacteroidaceae*, *Bdellovibrionaceae*, *Chitinophagaceae*, *Comamonadaceae*, *Cytophagaceae*, *Corynebacteriaceae*, *Enterococcaceae*, *Erysipelotrichaceae*, *Flavobacteriaceae*, *Nocardiaceae*, *Hydrogenophilaceae*, *Hyphomicrobiaceae*, *Mycobacteriaceae*, *Sphingobacteriaceae*, *Staphylococcaceae*, *Streptococcaceae*, *Micrococcaceae*, *Streptomycetaceae*, and *Micromonosporaceae*. In the soil sample, this KO is predicted to occur mainly in *Bacteroidaceae*, *Enterobacteriaceae*, *Corynebacteriaceae*, *Hyphomicrobiaceae*, *Nocardiaceae*, and *Corynebacteriaceae* (Fig. 5). Identified AR genes are involved in the degradation of  $\beta$ -lactam (ampicillin) antibiotic and commonly used as antimicrobial agent to treat bacterial infection (Elander 2003). Thus, the presence of such genes leads to speculation about the misuse or overuse of antibiotics in this region.

Due to the improper disposal system, the household and hospital wastes containing xenobiotic compounds were deposited in the environment. The presence of recalcitrant xenobiotic compounds exerts selective pressure on the microbial community leading to the elimination of susceptible species (Nojiri et al. 2004; Devpura et al. 2017). These communities adopt a number of metabolic processes to survive in these environments (Nojiri et al. 2004; Devpura et al. 2017). They acquire novel enzymatic mechanisms as well as degradative metabolic pathways through natural selection for the decomposition of recalcitrant compounds (Galvão et al. 2005). Thus, an analysis of the metabolic activity of these communities will help to understand the possible impact of these domestic wastes on the ecosystems. In the present study, we analyzed various genes

involved in such metabolic process in samples of contaminated soil and sediments in Northeastern India and several genes encoding various enzymes involved in the degradation of xenobiotic compounds. A total of 8, 4, 12, 7, 12, 8, 12, and 8 putative xenobiotic degradation genes involved in the degradation of benzoate, aminobenzene, naphthalene, fluorobenzoate, chloroalkane and chloroalkene, chlorocyclohexane and chlorobenzene, and nitrotoluene and toluene, respectively, were found in the soil and sediment samples. Major bacterial OTU's associated with benzoate degradation were *Acinetobacter*, *Arthrobacter*, *Comamonas*, *Diaphorobacter*, *Geobacter*, *Novispirillum*, *Phenylobacterium*, *Pseudoxanthomonas*, *Pseudoxanthomonas*, and *Rhodoplanes*. The predicted dominant genera associated with the degradation of aminobenzene were *Sphingomonas*, *Novosphingobium*, *Sulfuricurvum*, and *Telmatospirillum*. Similarly, the degradation of naphthalene was carried out by the major bacterial OTUs assigned under the genera *Arthrobacter*, *Comamonas*, *Diaphorobacter*, *Geobacter*, *Klebsiella*, *Leptothrix*, and *Novosphingobium*. OTU under the bacterial genera *Acinetobacter*, *Agrobacterium*, *Citrobacter*, *Devosia*, *Diaphorobacter*, *Erwinia*, *Hyphomicrobium*, *Klebsiella*, *Leptothrix*, *Novosphingobium*, *Perluclidibaca*, *Pseudomonas*, *Rhodoplanes*, and *Variovorax* mainly contributed to the fluorobenzene degradation. A complete list of bacterial OTUs associated with the xenobiotic degradation is shown in Table S3. The abundance of these genes in the metagenomics datasets indicates that the hospital and domestic wastes may contain the high concentration of benzoate, aminobenzene, naphthalene, fluorobenzoate, chloroalkane and chloroalkene, chlorocyclohexane and chlorobenzene, and nitrotoluene and toluene.

## Conclusion

The present study examined the taxonomical and metabolic diversity of the bacterial communities present in contaminated soil and sediment samples. The bacterial community at polluted sites was dominated by Actinobacteria and Proteobacteria. All the sites were associated with pathogenic bacteria, and the imputed functional analysis identified putative genes associated with antibiotic resistance as well as xenobiotic degradation in these environments. This study improves our

understanding of bacterial community and its metabolic activity in the contaminated ecosystem, which will help us to mitigate the health hazards and develop environmental remediation plans.

**Funding information** This research was supported by the Bioinformatics Infrastructure Facility (BTISNeT) sponsored by Department of Biotechnology, Govt. of India, New Delhi. The grant has enabled us to establish computational facility to work on Metagenomics pipeline, and the Advanced State Biotech Hub grant has helped in sampling and generate NGS data. The DeLCON facility has enabled for study design, analysis, and interpretation of data.

#### Compliance with ethical standards

**Conflict of interest** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

#### References

- Allen, T. D., Lawson, P. A., Collins, M. D., Falsen, E., & Tanner, R. S. (2006). *Cloacibacterium normanense* gen. nov., sp. nov., a novel bacterium in the family Flavobacteriaceae isolated from municipal wastewater. *International Journal of Systematic and Evolutionary Microbiology*, 56(6), 1311–1316. <https://doi.org/10.1099/ijs.0.64218-0>.
- Allen, H. K., Moe, L. A., Rodbumrer, J., Gaarder, A., & Handelsman, J. (2009). Functional metagenomics reveals diverse B-lactamases in a remote Alaskan soil. *ISME Journal*, 3(2), 243–251. <https://doi.org/10.1038/ismej.2008.86>.
- Banskar, S., Bhute, S. S., Suryavanshi, M. V., Punekar, S., & Shouche, Y. S. (2016). Microbiome analysis reveals the abundance of bacterial pathogens in Roussetus leschenaultii guano. *Scientific Reports*, 6. <https://doi.org/10.1038/srep36948>.
- Blomström, A.-L., Lalander, C., Komakech, A. J., Vinnerås, B., & Boqvist, S. (2016). A metagenomic analysis displays the diverse microbial community of a vermicomposting system in Uganda. *Infection ecology & epidemiology*, 6, 32453. <https://doi.org/10.3402/IEE.V6.32453>.
- Brooke, J. S. (2012). *Stenotrophomonas maltophilia*: an emerging global opportunistic pathogen. *Clinical Microbiology Reviews*. <https://doi.org/10.1128/CMR.00019-11>.
- Caccavo, F., Lonergan, D. J., Lovley, D. R., Davis, M., Stolz, J. F., & McInerney, M. J. (1994). *Geobacter sulfurreducens* sp. nov., a hydrogen- and acetate-oxidizing dissimilatory metal-reducing microorganism. *Applied and Environmental Microbiology*, 60(10), 3752–3759.
- Cai, L., & Zhang, T. (2013). Detecting human bacterial pathogens in wastewater treatment plants by a high-throughput shotgun sequencing technique. *Environmental Science and Technology*, 47(10), 5433–5441. <https://doi.org/10.1021/es400275r>.
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*. <https://doi.org/10.1038/nmeth.f.303>.
- Coates, J. B., Chakraborty, R., Lack, J. G., O'Connor, S. M., Cole, K. A., Bender, K. S., & Achenbach, L. A. (2001). Anaerobic benzene oxidation coupled to nitrate reduction in pure culture by two strains of *Dechloromonas*. *Nature*, 411(6841), 1039–1043. <https://doi.org/10.1038/35082545>.
- De Bentzmann, S., & Plésiat, P. (2011). The *Pseudomonas aeruginosa* opportunistic pathogen and human infections. *Environmental Microbiology*, 13(7), 1655–1665.
- De Mandal, S., Chatterjee, R., & Kumar, N. S. (2017). Dominant bacterial phyla in caves and their predicted functional roles in C and N cycle. *BMC Microbiology*, 17(1), 90.
- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, 72(7), 5069–5072. <https://doi.org/10.1128/AEM.03006-05>.
- Devpura, N., Jain, K., Patel, A., Joshi, C. G., & Madamwar, D. (2017). Metabolic potential and taxonomic assessment of bacterial community of an environment to chronic industrial discharge. *International Biodeterioration and Biodegradation*, 123, 216–227. <https://doi.org/10.1016/j.ibiod.2017.06.011>.
- Doumith, M., Dhanji, H., Ellington, M. J., Hawkey, P., & Woodford, N. (2012). Characterization of plasmids encoding extended-spectrum  $\beta$ -lactamases and their addition systems circulating among *Escherichia coli* clinical isolates in the UK. *Journal of Antimicrobial Chemotherapy*, 67(4), 878–885. <https://doi.org/10.1093/jac/dkr553>.
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27(16), 2194–2200. <https://doi.org/10.1093/bioinformatics/btr381>.
- Elander, R. P. (2003). Industrial production of beta-lactam antibiotics. *Applied Microbiology and Biotechnology*, 61(5–6), 385–392. <https://doi.org/10.1007/s00253-003-1274-y>.
- Erickson, A. R., Cantarel, B. L., Lamendella, R., Darzi, Y., Mongodin, E. F., Pan, C., Shah, M., Halfvarson, J., Tysk, C., Henrissat, B., Raes, J., Verberkmoes, N. C., Fraser, C. M., Hettich, R. L., & Jansson, J. K. (2012). Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. *PLoS One*, 7(11). <https://doi.org/10.1371/journal.pone.0049138>.
- Fàbrega, A., & Vila, J. (2012). *Yersinia enterocolitica*: Pathogenesis, virulence and antimicrobial resistance. *Enfermedades Infecciosas y Microbiología Clínica*, 30(1), 24–32. <https://doi.org/10.1016/j.eimc.2011.07.017>.
- Fritah, S., Lhocine, N., Golebiowski, F., Mounier, J., Andrieux, A., Jouvion, G., et al. (2014). Sumoylation controls host antibacterial response to the gut invasive pathogen *Shigella flexneri*. *EMBO Reports*, 15(9), 965–972. <https://doi.org/10.15252/embr.201338386>.

- Galvão, T. C., Mohn, W. W., & De Lorenzo, V. (2005). Exploring the microbial biodegradation and biotransformation gene pool. *Trends in Biotechnology*. <https://doi.org/10.1016/j.tibtech.2005.08.002>.
- García, A., Fox, J. G., & Besser, T. E. (2010). Zoonotic enterohemorrhagic *Escherichia coli*: A One Health perspective. *ILAR Journal / National Research Council, Institute of Laboratory Animal Resources*, 51(3), 221–232. <http://www.ncbi.nlm.nih.gov/pubmed/21131723>.
- Giannella, R. (1996). *Salmonella - Medical Microbiology - NCBI Bookshelf. Medical Microbiology*.
- Goossens, H., Ferech, M., Vander Stichele, R., & Elseviers, M. (2005). Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. *Lancet*, 365(9459), 579–587. [https://doi.org/10.1016/S0140-6736\(05\)17907-0](https://doi.org/10.1016/S0140-6736(05)17907-0).
- Guerrero, L. A., Maas, G., & Hogland, W. (2013). Solid waste management challenges for cities in developing countries. *Waste Management*, 33(1), 220–232. <https://doi.org/10.1016/j.wasman.2012.09.008>.
- Ibekwe, A. M., Leddy, M., & Murinda, S. E. (2013). Potential human pathogenic bacteria in a mixed urban watershed as revealed by pyrosequencing. *PLoS One*, 8(11). <https://doi.org/10.1371/journal.pone.0079490>.
- Ivanek, R., Gröhn, Y. T., Wells, M. T., Lembo, A. J., Sauders, B. D., & Wiedmann, M. (2009). Modeling of spatially referenced environmental and meteorological factors influencing the probability of *Listeria* species isolation from natural environments. *Applied and Environmental Microbiology*, 75(18), 5893–5909. <https://doi.org/10.1128/AEM.02757-08>.
- Joshi, N., & Fass, J. (2011). Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at <https://github.com/najoshi/sickle>.
- Kenneth, J. R., & Ray, C. G. (2004). *Sherris Medical Microbiology. Vasa*. <https://doi.org/10.1036/0838585299>
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, 41(1), e1–e1.
- Kwon, S., Moon, E., Kim, T.-S. S., Hong, S., & Park, H.-D. D. (2011). Pyrosequencing demonstrated complex microbial communities in a membrane filtration system for a drinking water treatment plant. *Microbes Environ*, 26(2), 149–155. <https://doi.org/10.1264/jsme2.ME10205>.
- Lamendella, R., Santo Domingo, J. W., Ghosh, S., Martinson, J., & Oerther, D. B. (2011). Comparative fecal metagenomics unveils unique functional capacity of the swine gut. *BMC Microbiology*. <https://doi.org/10.1186/1471-2180-11-103>.
- Langille, M., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, 31(9), 814–821. <https://doi.org/10.1038/nbt.2676>.
- Liu, J., He, X., Lin, X., Chen, W., Zhou, Q., Shu, W., & Huang, L. (2015). Ecological effects of combined pollution associated with E-waste recycling on the composition and diversity of soil microbial communities. *Environmental Science & Technology*, 49(11), 6438–6447. <https://doi.org/10.1021/es5049804>.
- Liu, J., Chen, X., Shu, H. Y., Lin, X. R., Zhou, Q. X., Bramryd, T., et al. (2018). Microbial community structure and function in sediments from e-waste contaminated rivers at Guiyu area of China. *Environmental Pollution*, 235, 171–179. <https://doi.org/10.1016/j.envpol.2017.12.008>.
- Lovley, D. R., Giovannoni, S. J., White, D. C., Champine, J. E., Phillips, E. J. P., Gorby, Y. A., & Goodwin, S. (1993). *Geobacter metallireducens* gen. nov. sp. nov., a microorganism capable of coupling the complete oxidation of organic compounds to the reduction of iron and other metals. *Archives of Microbiology*, 159(4), 336–344. <https://doi.org/10.1007/BF00290916>.
- Lozupone, C., & Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, 71(12), 8228–8235. <https://doi.org/10.1128/AEM.71.12.8228-8235.2005>.
- Lucker, S., Wagner, M., Maixner, F., Pelletier, E., Koch, H., Vacherie, B., et al. (2010). A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proceedings of the National Academy of Sciences*, 107(30), 13479–13484. <https://doi.org/10.1073/pnas.1003860107>.
- Manchanda, V., Sanchaita, S., & Singh, N. (2010). Multidrug resistant acinetobacter. *Journal of Global Infectious Diseases*, 2(3), 291–304. <https://doi.org/10.4103/0974-777X.68538>.
- Mwaikono, K. S., Maina, S., Sebastian, A., Schilling, M., Kapur, V., & Gwakisa, P. (2016). High-throughput sequencing of 16S rRNA gene reveals substantial bacterial diversity on the municipal dumpsite. *BMC Microbiology*, 16(1). <https://doi.org/10.1186/s12866-016-0758-8>.
- Nojiri, H., Shintani, M., & Omori, T. (2004). Divergence of mobile genetic elements involved in the distribution of xenobiotic-catabolic capacity. *Applied Microbiology and Biotechnology*. <https://doi.org/10.1007/s00253-003-1509-y>.
- Panda, A. K., Bisht, S. S., Kumar, N. S., & De Mandal, S. (2015). Investigations on microbial diversity of Jakrem hot spring, Meghalaya, India using cultivation-independent approach. *Genomics Data*, 41, 156–157.
- Qiu, G., Song, Y. H., Zeng, P., Duan, L., & Xiao, S. (2013). Characterization of bacterial communities in hybrid upflow anaerobic sludge blanket (UASB)-membrane bioreactor (MBR) process for berberine antibiotic wastewater treatment. *Bioresour Technol*, 142, 52–62. <https://doi.org/10.1016/j.biortech.2013.04.077>.
- Shen, C., Ge, Y., Yang, T., & Chu, H. (2017). Verrucomicrobial elevational distribution was strongly influenced by soil pH and carbon/nitrogen ratio. *Journal of Soils and Sediments*, 17(10), 2449–2456. <https://doi.org/10.1007/s11368-017-1680-x>.
- Sohn, J. H., Kwon, K. K., Kang, J. H., Jung, H. B., & Kim, S. J. (2004). *Novosphingobium pentaromativorans* sp. nov., a high-molecular-mass polycyclic aromatic hydrocarbon-degrading bacterium isolated from estuarine sediment. *International Journal of Systematic and Evolutionary Microbiology*, 54(5), 1483–1487. <https://doi.org/10.1099/ijs.0.02945-0>.
- Song, L., Wang, Y., Tang, W., & Lei, Y. (2015). Bacterial community diversity in municipal waste landfill sites. *Applied Microbiology and Biotechnology*, 99(18), 7745–7756. <https://doi.org/10.1007/s00253-015-6633-y>.

- Tanaka, S., Yoshida, M., Murakami, Y., Ogiwara, T., Shoji, M., Kobayashi, S., et al. (2008). The relationship of *Prevotella intermedia*, *Prevotella nigrescens* and *Prevotella melaninogenica* in the supragingival plaque of children, caries and oral malodor. *Journal of Clinical Pediatric Dentistry*, 32(3), 195–200. <https://doi.org/10.17796/jcpd.32.3.vp65717781561811>.
- Van Kessel, M. A. H. J., Speth, D. R., Albertsen, M., Nielsen, P. H., Op Den Camp, H. J. M., Kartal, B., et al. (2015). Complete nitrification by a single microorganism. *Nature*, 528(7583), 555–559. <https://doi.org/10.1038/nature16459>.
- Watson, S. C. L., Paterson, D. M., Queirós, A. M., Rees, A. P., Stephens, N., Widdicombe, S., & Beaumont, N. J. (2016). A conceptual framework for assessing the ecosystem service of waste remediation: In the marine environment. *Ecosystem Services*. <https://doi.org/10.1016/j.ecoser.2016.06.011>.
- Welch, D. B. M., & Huse, S. M. (2011). Microbial diversity in the deep sea and the underexplored “rare biosphere”. In *Handbook of Molecular Microbial Ecology II: Metagenomics in Different Habitats* (pp. 243–252). <https://doi.org/10.1002/9781118010549.ch24>.
- Yan, J., Ritalahti, K. M., Wagner, D. D., & Löffler, F. E. (2012). Unexpected specificity of interspecies cobamide transfer from *Geobacter* spp. to organohalide-respiring *Dehalococcoides mccartyi* strains. *Applied and Environmental Microbiology*, 78(18), 6630–6636. <https://doi.org/10.1128/AEM.01535-12>.
- Yang, W., Chen, P., Villegas, E. N., Landy, R. B., Kanetsky, C., Cama, V., Dearen, T., Schultz, C. L., Orndorff, K. G., Prelewicz, G. J., Brown, M. H., Young, K. R., & Xiao, L. (2008). *Cryptosporidium* source tracking in the Potomac river watershed. *Applied and Environmental Microbiology*, 74(21), 6495–6504. <https://doi.org/10.1128/AEM.01345-08>.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.