First report on Bacterial Diversity of Potable Spring water of Indian Himalayan Region

Ashish Kumar Singh[1], Saurav Das[2], Samer Singh[3], Varsha Rani Gajamer[1], Nilu Pradhan[1] Yangchen D. Lepcha[4] and Hare Krishna Tiwari[1]*

1. Department of Microbiology, School of Life Sciences, Sikkim University, Gangtok, Sikkim
2. DBT- Advanced Institutional Biotech Hub, Bholanath College, Dhubri, Assam
3. Institute of Microbial Technology, Chandigarh University, Chandigarh, Punjab
4. State Institute of Rural Development, Government of Sikkim, Gangtok, Sikkim, India.

Corresponding Author: Dr. Hare Krishna Tiwari, Department of Microbiology, School of Life Sciences, Sikkim University, Gangtok, Sikkim. Email ID: hktiwari_2005@rediffmail.com

Mob: 8250334595

Abstract:

Water quality of a region directly corroborates with the health index of people. People in the Himalayan hills mainly depend upon the spring water for potability. To determine the microbial ecology of the spring waters of Sikkim, the variable region of 16S rRNA has been sequenced using Illumina MiSeq. Phylum wise annotation showed the East and North district are mostly dominated by *Proteobacteria (*41% and 35.80%), whereas West and South district is dominated by *Planctomycetes* (38.46%) and *Verrucomicrobia* (33%). The consistent dominance phyla in the all the four districts were *Bacteriodetes* (34-24%) which was highest dominancy in North district and lowest in wets district. Genus wise distribution showed the abundance of *Brevifolis*,

Flavobacterium, Verrucomicrobia subdivision3, Emticica, Cytophaga, Prosthecobacter, Planctomycetes, Varivorax, Arcicella, Isosphera, Sedimunibacterium etc. The East district showed highest dominancy of genus *Emticicia* whereas *Planctomycetes* in the West district. The North district was mainly dominated by genus *Arcicella* and *Brevifollis* in the South district. North on the antonymous showed totally different sets of microbial diversity. North district showed an abundance of *Arcicella, Planctomycetes, Schlerensia* and *Azohydromonas.* The heat map produced by Bray Curtis distance method produced three clusters which showed the close relationship between West and East district microbiome that further related to South district. The sample of North district formed out group that showed different community structure from other three districts. The principle component analysis was showed that the east and South district samples are closely related and distantly correlated to the west Sikkim, but the North district showed completely different microbial community. The canonical correspondence analysis showed correlation between bacterial diversity and hydrochemistry and it was found that the bacterial diversity was influenced by the concentration of different metallic ions like sodium, calcium, barium and iron. This is a first report from the Eastern Himalayan region of India and it largely enhances our knowledge about the microbial structure of potable spring water of Eastern Himalayan. This study is useful for Government of India as well as the state government to adopt the different strategic treatment procedures to improve the quality of water that is supplied to the community resides in the Himalayan regions and solely dependent on this untreated spring water.

Keywords: Sikkim, Spring Water, Metagenomics, Bacteria, Proteobacteria, Heat Map

1.0 Introduction:

Microbial communities are the key members of many ecosystems on the earth (Lagkouvardos *et al.*, 2016). The word microbial communities can be described in the terms of richness or evenness as well as composition i.e. abundance of taxa and genes in the sample (Rothschild, 1991). They play a significant role in crucial biological functions ranging from nitrogen and carbon cycles in the environment to regulation of metabolic and immune responses in inhabited animal and human hosts (Rothschild, 1991; Hörmannsperger *et al.*, 2012; Lagkouvardos *et al.*, 2016). Hence, the study of microbial ecology is important for proper understanding of these microbial communities and their functional impact (Lagkouvardos *et al.*, 2016).

Around one-third of global freshwater reserves are placed in subsurface streams and aquifers which represent as a freshwater source for human consumption as well as irrigation (Hemme *et al.*, 2015). As good quality freshwater is an important natural source for domestic as well as the industrial purpose which is gradually exposed by a human (Van Rossum *et al.*, 2015; Vörösmarty *et al.*, 2010) as well as anthropogenic activity that affect the quality of water (Vörösmarty *et al.*, 2010). Throughout the world contamination of water bodies by waterborne pathogens and disease caused by them are a major water quality concern (Pandey *et al.*, 2014). The potential host-specific nature of the microbes they have long been used as an indicator of poor water quality (Staley *et al.*, 2013). Although, these tests for detection of pathogens are based on culture-dependent methods which carry both false positive as well as a false negative result (McLain *et al.*, 2011; Staley *et al.*, 2013). The exploration of microbial communities through culture-dependent methods were used from ancient time, but on the standard laboratory media less than 1%, bacterial species of the environmental communities

were culturable (Staley and Sadowsky, 2016). To overcome the traditional culture methods, now metagenomics delivers a suitable method for monitoring the environmental communities via high-throughput manner. Such techniques have revealed unprecedented taxonomic and functional diversity in aquatic and terrestrial habitats (Sogin *et al.*, 2006). The term metagenomics describes as the sequencing of total DNA isolated directly from the environmental samples i.e. 'metagenomes' which simultaneously provides the access of genetic information of microbial communities from the mixed environment (Dinsdale *et al.*, 2008). By applying this approach, functional and taxonomic microbial diversity can be described and then changes in communities can be monitored over time and space in response to anthropogenic or environmental impacts related to human health (Port *et al.*, 2012). The growing accessibility of next-generation DNA sequencing (NGS) methods has greatly advanced our understanding of microbial diversity in medical and environmental science (Lee *et al.*, 2017). The Next-generation sequencing (NGS) methods have prominently increased sequencing throughput by using massively parallel sequencing (Sogin *et al.*, 2006). The amplification of small but variable regions of the 16s rDNA i.e. V3, V4, V5 or V6 hypervariable regions has provided very profound sequencing of bacterial ecology. This method has used for the identification of very rare populations which is present in very low abundance that may necessary for the functional diversity and ecosystem stability (Sogin *et al.*, 2006; Staley *et al.*, 2013). Several bacterial communities from different environmental samples like marine waters (Brown *et al.*, 2009), soils (Jones *et al.*, 2009), wastewater (Sanapareddy *et al.*, 2009), contaminated water (Das *et al.*, 2017) and several human microbiomes were successfully characterized (Staley *et al.*, 2013; Sanapareddy *et al.*, 2009).

The range of Himalaya is being the source of innumerable perennial rivers, streams as well as springs and the mountain peoples, are largely dependent on this spring water for their sustenance. In mountain regions, there are several natural discharges of groundwater from several aquifers in the form of Springs. The mountain Springs are locally known as Dharas which is in most cases unconfined. In Sikkim, most of the springs are denoted as 'Devithans' which are considered as sacred and protected as well as prevented from any living interferences to maintain the holiness of the respected springs. In Sikkim, about 80% of the rural households population depend on spring water for their drinking as well as household purpose (Tambe *et al.*, 2012). In the current study, we tried to explore the overall bacterial ecology of spring water of rural as well as urban areas of Sikkim, North East India on the basis of a culture-independent study of conserved region V3-V4 of 16S rRNA gene. Along with microbial ecology physicochemical analysis was also done to correlate with the diversity study.

2.0 Materials and Methods:

2.1 Site Description

The northeastern region of India is one of the biodiversity hotspots in the country. It consists of eight states which are often referred to as seven sisters (Assam, Arunachal Pradesh, Manipur, Nagaland, Tripura, Meghalaya, Mizoram) and one brother state (Sikkim). Sikkim is the second smallest State of India lying in ecological hotspot region of the lower eastern Himalayan belt. It has both alpine and subtropical climate with a high mountain range of widespread altitude variation (300-8598 m). It is the host to Kanchenjunga (8598m), the highest peak in India and third highest on earth. The state is subdivided into four districts – East Sikkim, West Sikkim, South Sikkim and North Sikkim. This mountain state is a wonder when it comes to water. It is a

house of waterfalls, springs, river, and lakes, which are always an attractant to tourists and nature lovers. Rainfall is the primary mode of recharge for most of the surface water sources. The precipitation of the water in the mountain range causes surface run-off which often takes the shape of streams, springs, and kholas (local name for springs). These springs remain as an important source of potable water for the local residents **(Fig. 1)**.

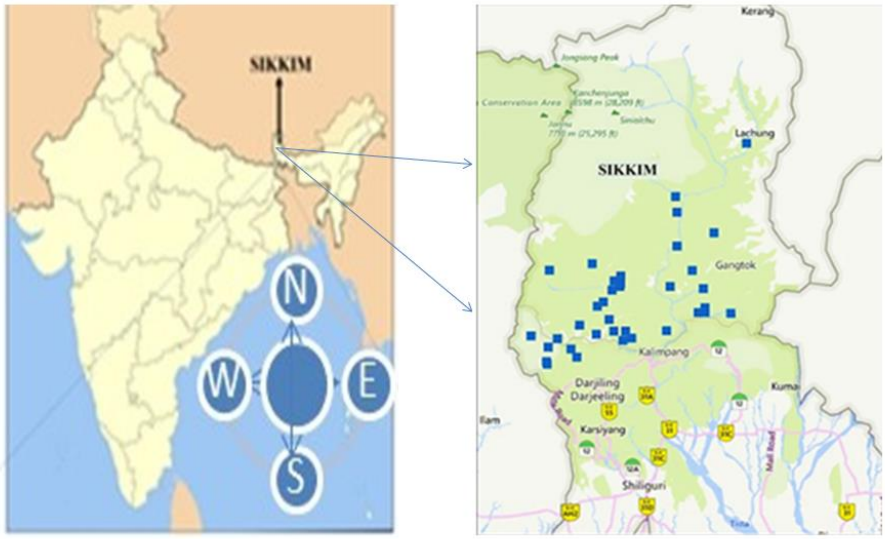


Fig. 1: Map of Study are with regional reference with respect to Map of India. The location is marked with respect to their GPS location [lattitude and longitude] (Supplementary File – S1: Name of the place with their GPS location).

2.2 Sampling Procedures and Processing

A total of 40 samples, 10 from each district was collected in nalgene wide moth sterile sample bottle (ThermoFisher Scientific, USA) for microbiome analysis. Primary physical parameter including pH and temperature were tested using portable multimeter probe(Hi-media, Mumbai). The physicochemical parameters were analyzed with Induced Coupled Plasma- Mass

Spectrometry (ICP-MS) (**Table. 1**). For total microbial diversity analysis, collected samples were thoroughly mixed with the equal ratio (1:1) based on the distribution like East, West, South and North district separately on different vials to prepare four composite samples of 100ml, representative of the four districts.

2.3 DNA Extraction

DNA from water sample was extracted using DNeasy PowerWater Kit (MO BIO Laboratories, Carlsbad, CA, USA) in accordance with the manufacturer's instruction. Quality of the DNA was checked on 0.8% agarose gel and DNA was quantified using Qubit Fluorometer (Vr. 4.0; Thermofisher Scientific, USA), which has a detection limit of 10 -100 ng/ µl (**Table. 1S**).

2.4 Metagenomic Sequencing

2.4.1 Preparation of 2x300 MiSeq library

The amplicon libraries were prepared using Nextera XT Index Kit (Illumina inc.) according to 16S metagenomic sequencing library preparation protocol (Part # 15044223 Rev. B). For amplification of eubacterial and archaeal V3-V4 region, specific primer pair V3 Forward: GCCTACGGNGGCWGCAG and V4 Reverse: ACTACHVGGGTATCTAATCC were synthesized and used. The amplified product was confirmed on 1.2% agarose gel. The adaptor-ligated amplicons were amplified using i5 and i7 primer for the addition of multiplexing index sequence required for cluster generation (P5 and P7 standard of Illumina cycle sequencing). The amplicon library was purified with AMpureXP beads (Beckman Coulter Genomics, Danvers, MA, USA). The amplified library was validated by Bioanalyzer 2100 (Agilent Technologies) using High Sensitivity (HS) DNA chips and concentration was quantified by Qubit fluorometer **(Table. S2)**.

2.4.2 Next-generation Sequencing and Sequence Analysis

Based on the data as obtained from the Qubit fluorometer and the bioanalyzer, 500µl of the 10 pM library was loaded into MiSeq cartridge for cluster generation and sequencing using Paired-end sequencing method. The sequence data were analyzed using the QIIME (Quantitative Insights Into Microbial Ecology) version 1.8.0 software program (Wang *et al.*, 2018). The adapter trimmed sequence was subjected to pre-processing for De-replication, Singleton removal, Chimera filtering with SolexaQA. Sequences with Phred score lower than 20, ambiguous bases having primer mismatch and low read length less than 100bp were removed. Sequences were grouped into Operational Taxonomic Units (OTUs) at 97% sequence similarity level using UCLUST (Edgar, 2010) and a consensus taxonomic classification was assigned to each representative OTU using the UCLUST classifier with a Greengenes 13.8 reference database (DeSantis et al., 2006).

2.5 Statistical Analysis

Normalization of the OTUs relative abundance data was performed by log transformation log10 (xi+1). Diversity analysis of the microbial ecology of spring water was authenticated with alpha, Shannon and Simpson diversity indices and test of significant difference were done by one-way using PAST software (ver. 3.1). Principle component analysis was carried out using R statistics (package – facto extra). Heatmap was constructed using R – statistics (package – ggplot) using Bray Curtis distance method to visualize the comparative differences in the microbiome community by excluding the taxa with less than <0.1% of abundance. Canonical Correspondence Analysis (CCA) was used to determine the correlative relationship between the microbial community (Phylum level) and the physicochemical parameters (R-Statistics, Package – Vegan).

2.6 Data Availability

The sequence obtained through high throughput sequencing method was submitted to NCBI Server and available under the accession number of East Sikkim Fresh water is SRX4016322, West Sikkim is SRX4016323, South Sikkim SRX4016320 and North Sikkim is SRX4016321.

3.0 Results

3.1 Physicochemical Properties of the Samples

Different physicochemical parameters were determined by multiprobe parameter (on site) and IC-PMS as mentioned in **Table.1**. All the water samples collected were normal to alkaline in nature. The temperature was 22 - 25^{o}C in most of the samples except those collected from the North Sikkim, which showed a little drop in the overall temperature with a range of 17 – 22^{o} C.

Physicochemical Parameter	East (E)	South (S)	West (W)	North (N)
pH	6.5 -7.5	6.5–7.5	6.5–7.5	6.5–7.5
Temp (^{o}C)	22 - 25	22 - 25	22 - 25	17 – 22
Ba (Barium) (ng/l)	14.18	5.313	9.761	10.118
Pb (Lead) (ng/l)	0.04	0.023	0.036	0.051
Ag (Silver) (ng/l)	0.018	0.005	0.017	0.018
Al (Aluminium) (ng/l)	1.011	1.315	1.011	1.315
Ca (Calcium) (ng/l)	35.205	8.525	25.71	13.412
Cd(Cadmium) (ng/l)	0.026	0.011	0.012	0.024
Cr(Chromium) (ng/l)	1.517	0.249	1.16	1.069
Cs(Cesium) (ng/l)	0.026	0.017	0.025	0.012
Cu(Copper) (ng/l)	3.025	3.206	1.941	2.106
Fe(Iron) (ng/l)	10.869	1.615	10.823	4.763
Ga(Gallium) (ng/l)	0.709	0.227	0.419	0.436

Hg(Mercury) (ng/l)	0.007	0.007	0.004	0.001
Na(Sodium) (ng/l)	71.618	71.618	52.327	42.317
Total Dissolved Solid (ppm)	41.5	41.5	13.5	39.5
Electric conductivity (µS/cm)	184.3	184.3	96	177.2

Table 1: Physicochemical Parameter of the composite samples. Elemental analysis was done with ICP-MS and pH, the temperature was recorded at the site with a multiprobe parameter (Hanna Instruments).

3.2 Refraction Curve and Diversity Index

Rarefaction curve and diversity index showed **(Fig. 2)** most of the microbial species were sequenced and identified, and significant variations in diversity among the water samples of four districts were observed. Samples from North and East districts showed high diversity as compared to the south and west districts, with Simpson index of N: 0.6988 & E: 0.6928 and S: 0.7499 & W: 0.7356 respectively (Supplementary Table. S3).

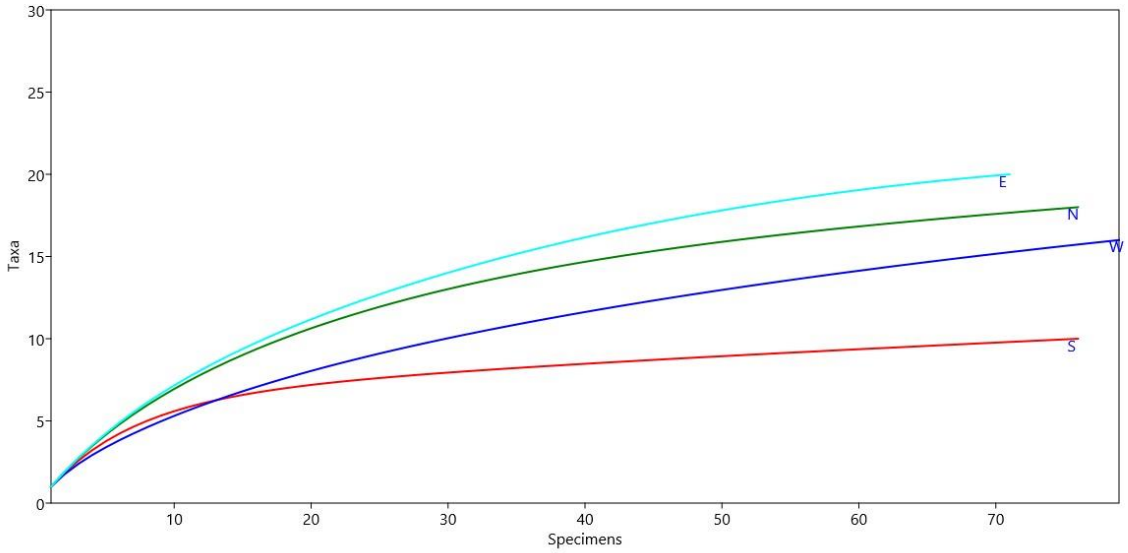
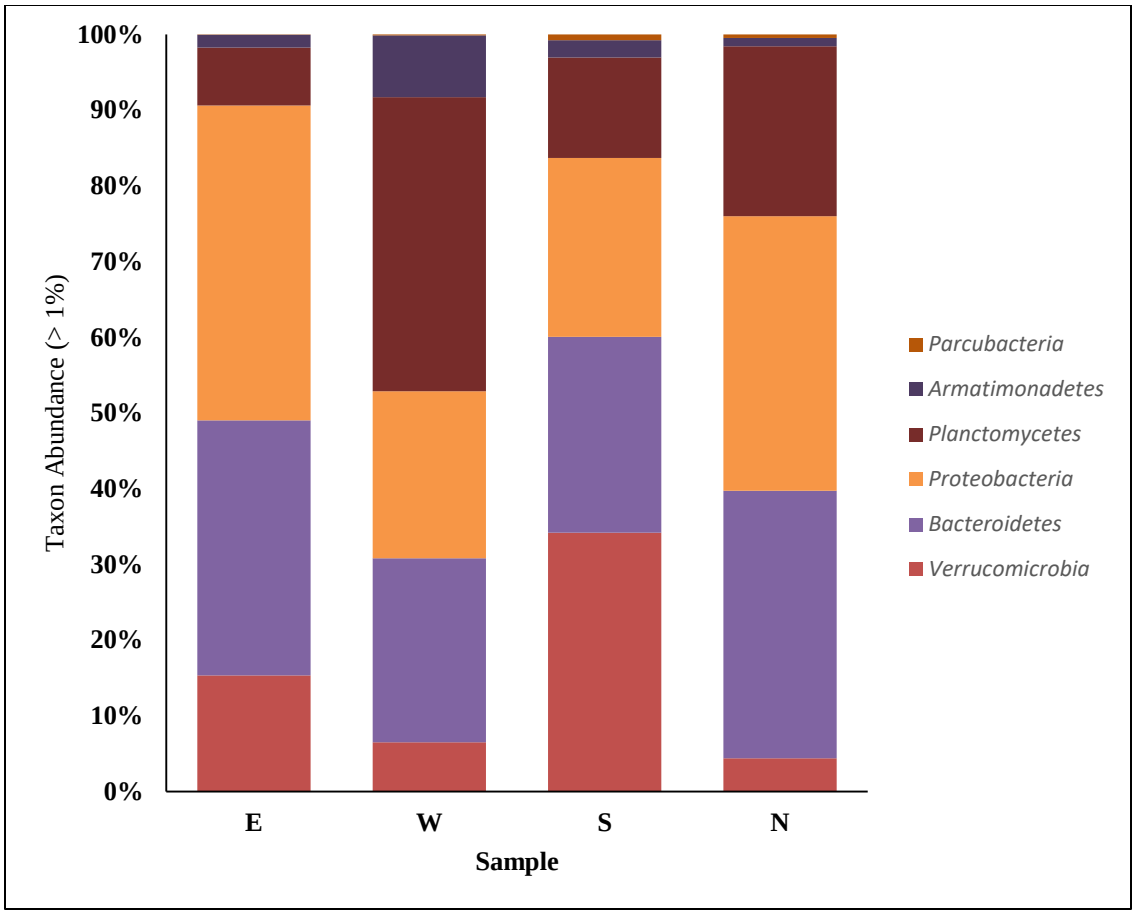


Fig. 2: Rarefaction curve analysis of four districts of Sikkim, East (E), West (W), South (S) and North (N).

3.3 Microbial Diversity

The phylum wise distribution revealed the dominance of *Bacteroidetes, Proteobacteria, Verrucomicrobia, Planctomycetes, Armatimonadetes* and to lesser extent *Parcubacteria, Actinobacteria* with other minor groups like *Aquificae, Firmicutes, Hydrogenedentes, Acidobacteria, Nitrospirae, Deinoococcus-Thermus,* and *Chloroflexi* **(Fig. 3a)**. Diversity ratio was significantly different within the samples of each district. Spring water from east district showed the dominance of *Proteobacteria* (41%), Bacteroidetes (33.21%) followed by *Verrucomicrobia* (15.12%) and *Planctomycetes* (7.54%). Major phylum from the spring water of west district was *Planctomycetes* (38.46%), *Bacteroidetes* (24.04%), *Proteobacteria* (21%) and with least abundance of *Armatimonadetes* (8.08%) and *Verrucomicrobia* (6.47%). Potable spring water of south district was predominated by *Verrucomicrobia* with 33% of relative abundance followed by *Bacteroidetes* (25.46%), *Proteobacteria* (23.25%) and *Planctomycetes* (13.05%). Sample from north district showed the dominance of *Proteobacteria* (35.80%) followed by *Bacteroidetes* (34.90%) and *Planctomycetes* (22.16%). Distant from the validated phyla few *Candidatus* phyla were also recorded though at lower abundance throughout the samples *viz. TM6, WWE3, SR1, SBR1093, FBP, CPR2*.



217

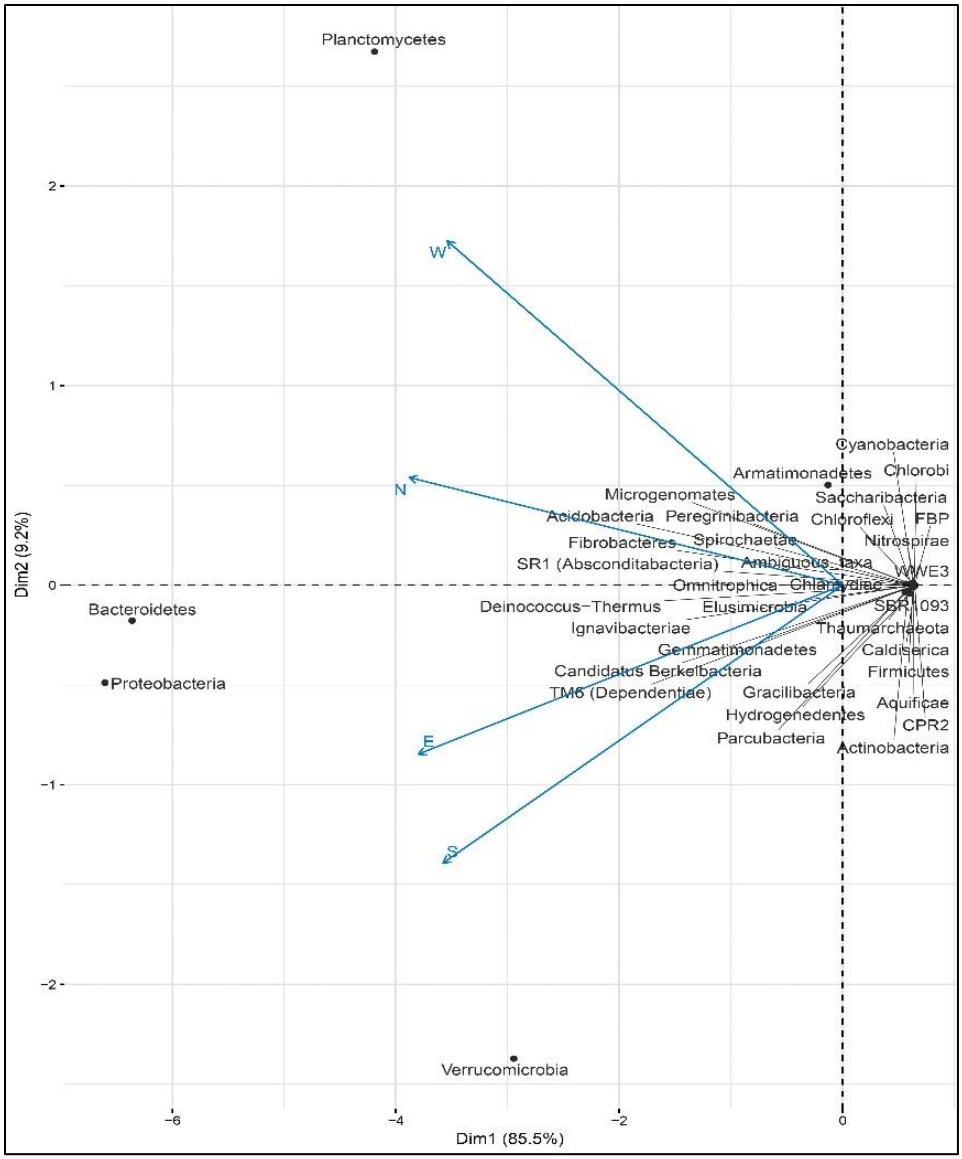
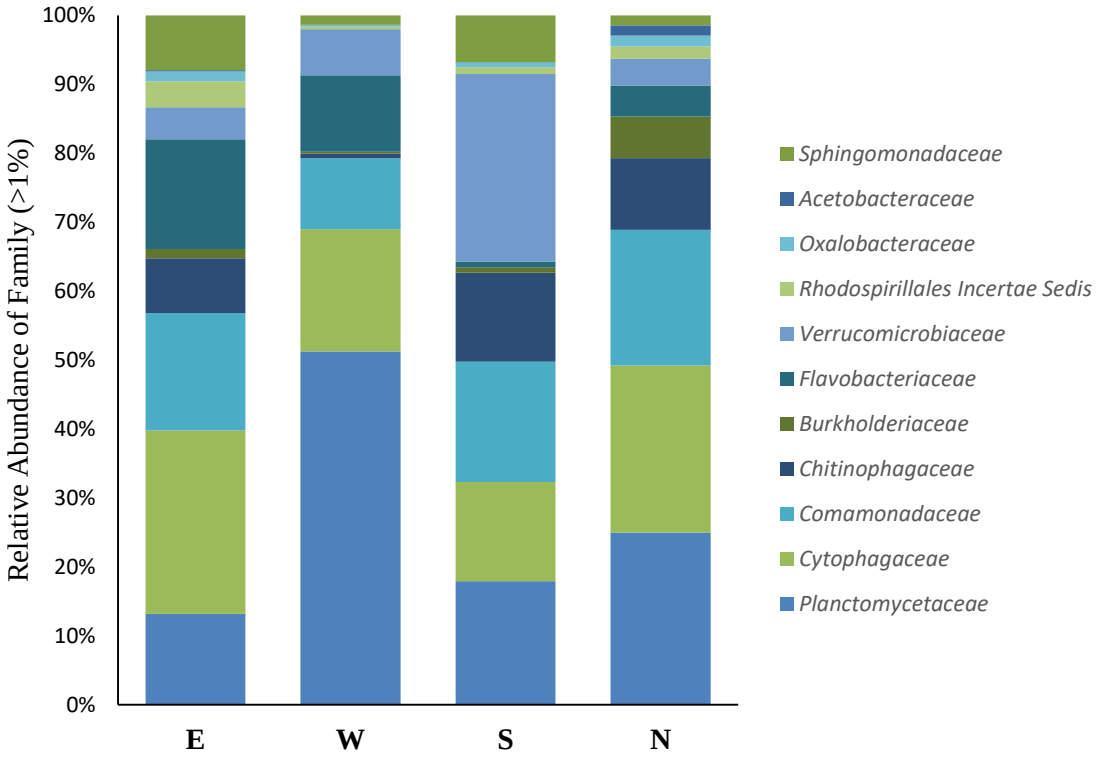


Fig. 3: Relative abundance of phylum as identified from the spring water samples of East (E), West (W), South (S) and North (N) districts. Major phylum as found were *Bacteroidetes, Proteobacteria, Verrucomicrobia,* and *Planctomycetes* and as represented in Bar-plot **(a)** and PCA-Biplot: The principal component analysis using unweighted unifrac distance method showed the microbial distribution in the sample of west and north, east and south are correlated **(b)**.

Family wise distribution **(Fig. 4)** showed the major dominance of *Verrucomicrobiaceae*, *Verrucomicrobia subdivision*3, *Planctomycetaceae*, *Comamonadaceae*, *Cytophagacease*, *Rhodospirillaceae*, *Sphingomonadaceae*, *Chitinophagaceae*, *Glycomycetaceae*, *and Flavobacteriaceae*. Springs from the east district were mainly dominated by *Cytophagacease (15.22%)* while the west district was dominated by *Planctomycetaceae* (38.34%). South district showed the dominance of family *Verrucomicrobiaceae* (19.84%) and the north district was dominated by *Planctomycetaceae* (22.12%). Some of the major genus found in the spring water was *Brevifolis*, *Flavobacterium*, *Verrucomicrobioa subdivision*3, *Emticicia*, *Cytophaga*, *Prosthecobacter*, *Planctomycetes*, *Varivorax*, *Arcicella Isosphera*, *Sediminibacterium*, *Acinetobacter*, *Chitinophaga*, *Rhodopirellula*, *Tenacibaculum*, *Flexibacter*, *Ustilago*, *Lactobactria*, *Flectobacillus*, *Pandoraea* and *Geobacillus*. East district was dominated by genus *Emticicia* (14.83%), west district by *Planctomyces* (36.89%), south district by *Brevifollis* (19.02%) and north district by *Arcicella* (18.29%) **(Fig. 5)**.



238

239 **Fig. 4:** Family wise distribution of microbial flora across four districts, East (E), West (W),

240 South(S) and North (N).

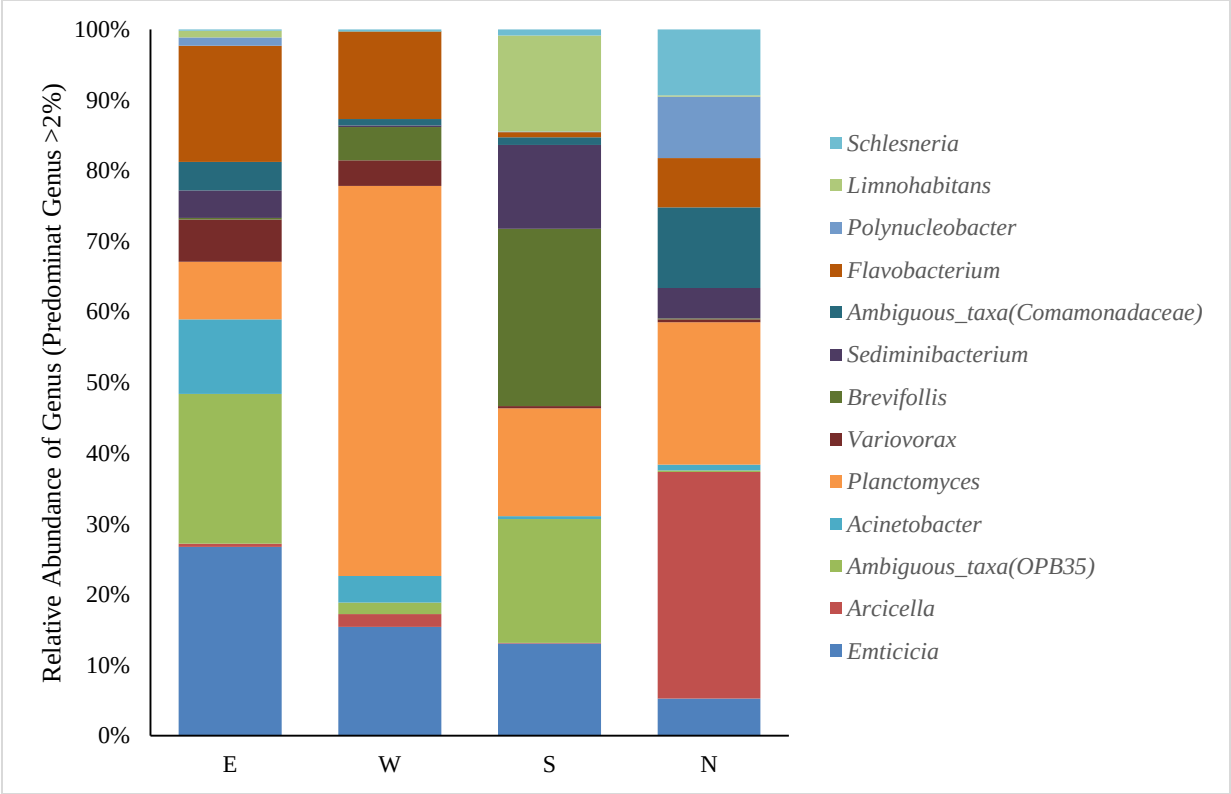
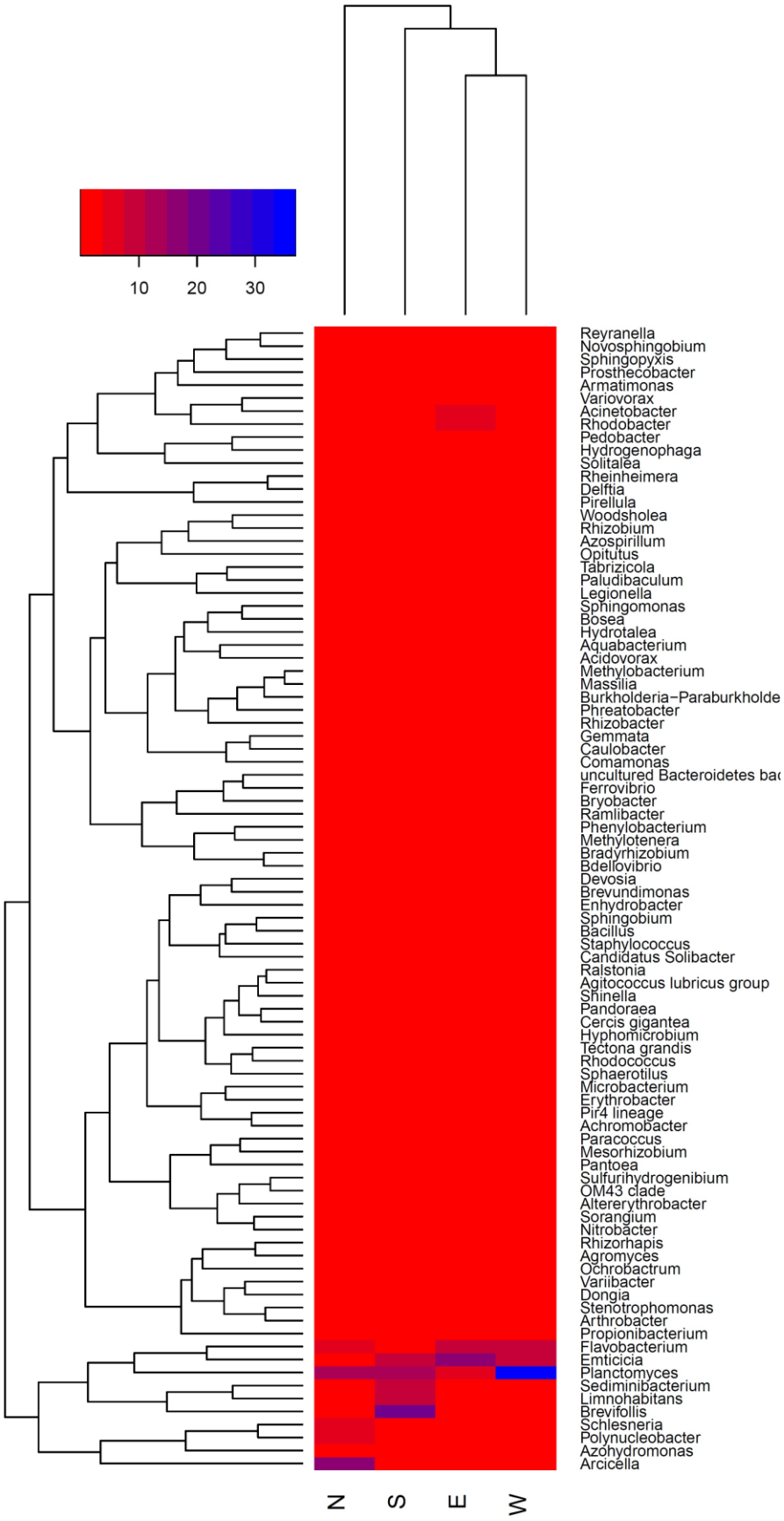


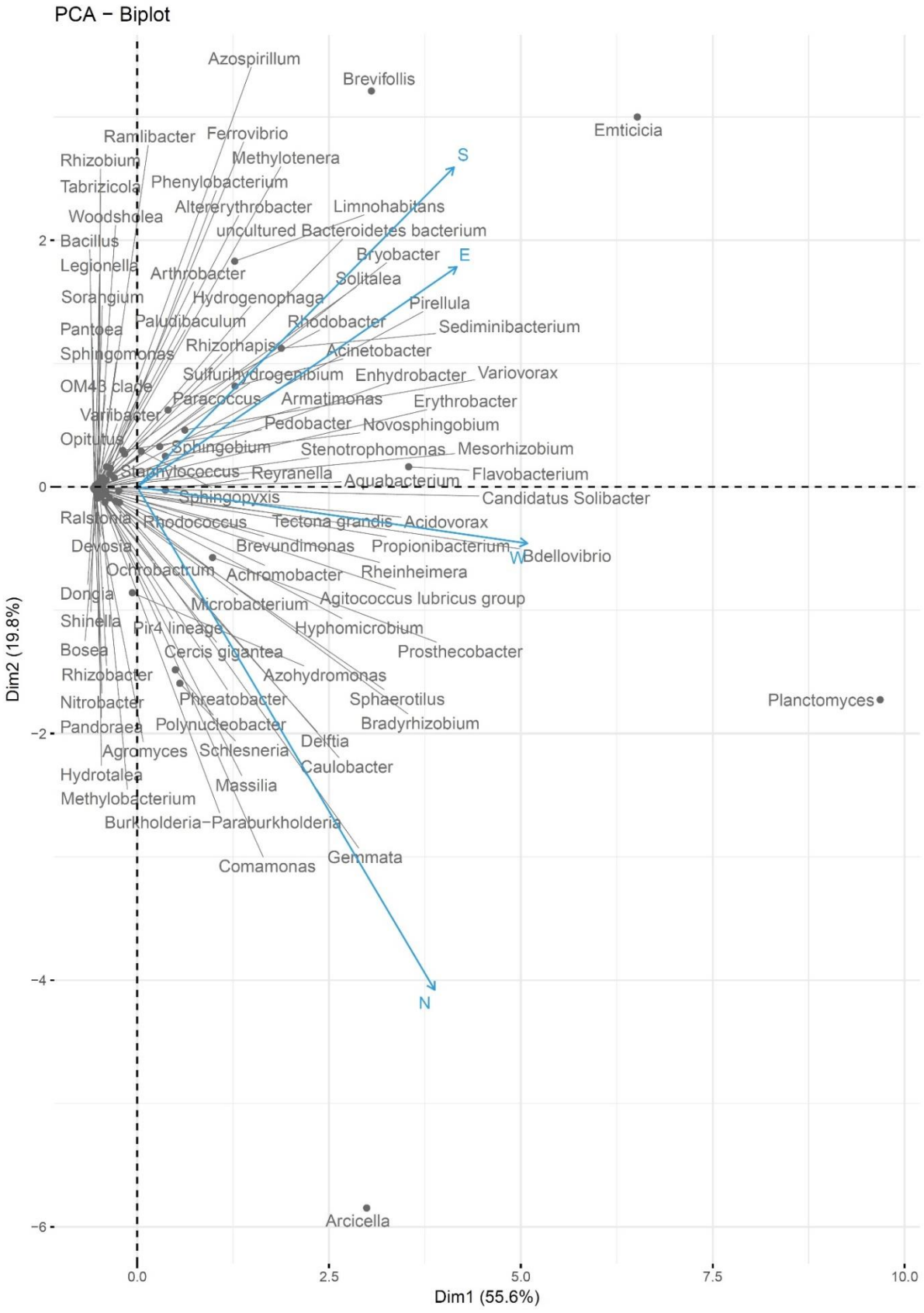
Fig. 5: Relative Abundance of Genus in the spring water of four districts of Sikkim. The chart was prepared excluding the genus with less than 50 reads correspondingly in all the samples.

3.4 Comparative Community Analysis

Comparative community analysis of the Microbiome of spring water from the four districts showed a significant correlation pattern. Genus *Arciella, Planctomycetes, Polynucleobacter, Schlesneria* were mainly found in the spring water of north district with comparably higher relative abundance than the other districts. Genus *Brevifolis, Limnohabitans, Sediminbacterium* were abundant in the waters of south district. Except in south district, genus *Flavobacterium* was found to be predominant in the spring water of north, east and west districts. Although genus *Planctomycetes* were found in all the districts, its dominance in the West district was found to be high. *Emticicia* was a dominant genus of the East district with relatively low abundance in west and south district and in North district presence of genus *Emiticicia* was not recorded. Heatmap

254 produced with the Bray Curtis distance method produced three clusters showing the close
255 relationship between the Microbiome of west district and east district which is again related to
256 the South district. North produces the out-group showing the different community structure from
257 the rest of three districts **(Fig. 6a)**. This can also be observed from the principal component
258 analysis where the Microbiome of east and south are showing close correlation and which is
259 again distantly related to the Microbiome of the west but north have totally different microbial
260 community structure (**Fig. 6b**).



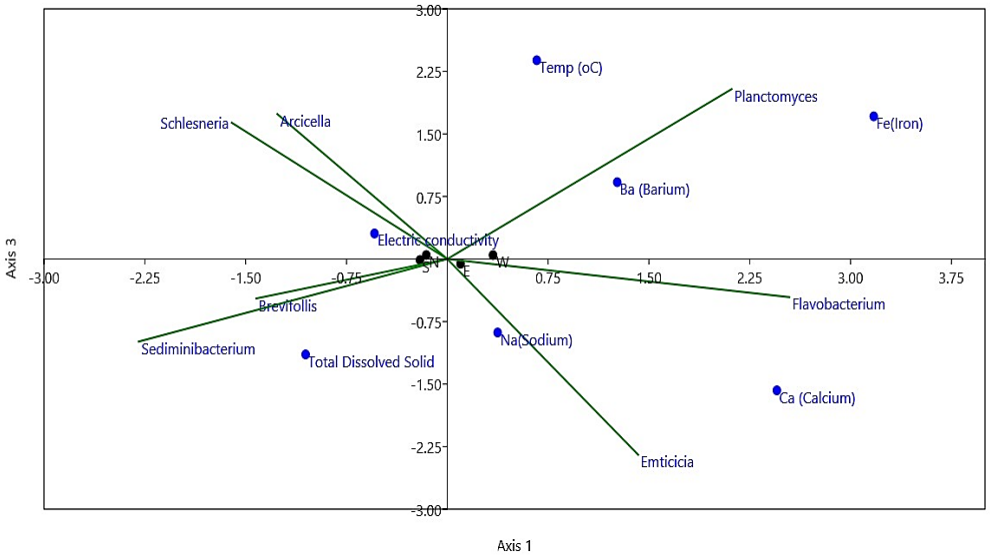


262

Fig. 6: Heatmap drew with Bray Curtis distance method to analyze the comparative difference in microbial diversity among the metagenomic library of spring waters from east, west, south and north district (color key: red = lowest, blue = highest. **(b)** Principale component analysis of the microbial community of four districts (at genus level), the analysis showed a close relationship between the microbial community of east and south, while they are distantly related to the west, but north has totally different community structure forming an outgroup in the plot.

3.5 The relationship between physicochemical characteristics and microbial diversity

Multivariate canonical correspondence analysis of seven physicochemical parameter and seven dominant microorganisms of four districts showed a significant relationship. The distribution of dominant genus i.e. *Emticicia* and *Flavobacterium* in the east district was found to be closely dependent on the concentration of sodium (41.5ng/l) and calcium (35.2ng/l). Correspondingly, the relative abundance of the dominant genus of west district i.e. *Planctomyces* was found to be dependent on the concentration of Barium (9.761), Iron (10.823) and it was also influenced by the temperature. The dominance of *Arcicella* and *Schlesneria* in the North district was correlated with electroconductivity of water. Genus *Brevifolis* and *Sediminibactrium* in South district were correlated with the total dissolved solids present in the springs of South district **(Fig. 7).**



279

280 **Fig. 7:** CCA plot showing the relationship between physicochemical parameters and microbial
281 diversity (genus).

282 ## 4 Discussion

283 Sikkim is a northeastern state lying in the lower region of Eastern Himalaya and neighbors to the
284 countries China, Bhutan and Nepal. It is among the three hot spot ecoregions of India (The
285 Western Ghat, Eastern Himalaya and Indo-Burma) with diverse fauna and flora. In topology, it
286 has mountain terrain with wide altitude variation from 800 -8000 meters. With a population of 6,
287 10, 577 it is the second smallest state of India. Majority of the population living in different
288 altitudes are dependent on the natural spring waters for drinking and household purposes. Quality
289 of life increases when people have access to safe drinking water with adequate sanitation. Better
290 management of water resources to reduce different water-borne diseases and to make water safe
291 for both potable and recreational purpose can save many lives. Water safety and quality are
292 fundamental to human development and well-being (WHO, 2018). However, till date, no such

studies have been conducted in Sikkim to determine the water structure of natural springs from microbiological or chemical aspects. Microbiological analysis is one the main facet of water quality measurement. Taxonomic profiling of dominant microflora can be an environmental indicator of water quality and can forecast future health threats in the surroundings. Knowing the ecology could be helpful in determining future water treatment protocols. But, limitation of culture-dependent methods has always been a barrier for comprehensive microbial ecology analysis. Development of metagenomics approaches and next-generation sequencing has allowed the scientist to overcome the barrier to determine the total microbial biodiversity of an environment. In this study, microbial ecology of the spring waters of Sikkim was determined by next-generation sequencing using variable regions of 16S rRNA gene (V3 -V4). To our knowledge, this is the first report on the microbial ecology of spring waters of Eastern and Western Himalayan Region of India.

The study showed a major dominance of *Bacteroidetes*, *Proteobacteria*, *Verrucomicrobia* and *Planctomycetes* in all the springs of East (E), West (W), South (S) and North (N) districts **(Fig. 3a)**. Springs of East district was dominated by *Proteobacteria (41%)* while springs of West district was dominated by *Planctomycetes* (38.46%). *Planctomycetes* (7.54%) was least dominated in the East as compared to the West district (38.46%). *Planctomycetes* are a phylum of aquatic bacteria, with its habitat in fresh, marine and brackish water (Fieseler *et al.*, 2004). They possess unusual characteristics such as intracellular compartmentalization and they lack peptidoglycan in their cell walls. Remarkably, few genera of this group like *Gemmata* even contain membrane-bound nucleoid similar to the eukaryotic nucleus (Boedeker *et al.*, 2017). Although a recent study reported its close similarity to Gram-negative bacteria however in-depth genetic studies is still lacking (Boedeker *et al.*, 2017). A unique bacterial phylum

316 *Armatimonadetes* (8.08%) was recorded from the samples of the west districts which are not
317 found in springs of other districts. Prior to official classification of *Armatimonadetes* by Tamaki
318 *et al*., (2011), it was classified as candidate division of OP10, first identified by Hugenholtz *et*
319 *al*., (1998) from Obsidian Pool, Yellowstone National Park (Hugenholtz *et al.*, 1998; Tamaki *et*
320 *al.*, 2011; Lee *et al.*, 2014). *Armatimonadetes* as a gram-negative oligotrophic aerobic bacterium
321 was described by Lee *et al.*, 2014. Spring water from South district was dominated by
322 *Verrucomicrobia* (25.46%). *Verrucomicrobia* have only a few described species so far, most of
323 the genus of the phylum is non-culturable and they are ubiquitous in freshwater and soil (Gupta,
324 2016; Griffiths and Gupta, 2007). The spring water of both North and East district was
325 dominated by *Proteobacteria* (35.80 %) and *Bacteroidetes* (34.90%). The principal component
326 analysis of relative phyla diversity also showed a close correlation between samples of the East
327 and North districts while samples of South and West districts were distantly related (**Fig. 3b**).

328 The spring waters from four districts of Sikkim showed considerable differences among
329 their dominant genus **(Fig. 5)**. The spring water of North was most diverse out of the three
330 districts having major dominance of *Arcicella, Planctomycetes, Polynucleobacter, Schlesneria*
331 and *Azohydromonas.* This difference in diversity can be due to the variance in the
332 physicochemical parameter. North district has a comparatively lower temperature (17 -22 ^{0}C) in
333 comparison to other districts. *Arcicella* is aerobic gram-negative bacteria was first proposed by
334 Nikitin *et al*., (2004) (Nikitin *et al.*, 2004). Some of the novel species of *Arcicella* are isolated
335 and identified from aquatic environment like *Arcicella aurantica* from stream water in Southern
336 Taiwan (Sheu *et al.*, 2018), *Arcicella rosea* from tap water (Kampfer *et al*., 2009) and *Arcicella*
337 *rigui* from Niao-Song Wetland Park in Taiwan (Chen *et al.*, 2013). Though the presence of
338 *Planctomycetes* was found in all the four districts with lower relative abundance, its dominance

was maximum at West (38.46%) and least at East district. The springs of East district was mainly dominated by the bacteria *Emticicia* (14.83%). *Emticicia* is a gram-negative bacterial genus belongs to family *Cytophagaceae* and they are ubiquitous in the aquatic environment(Schultz *et al.*, 2013; Seo *et al.*, 2015; Nam *et al.*, 2016). A number of significant members from the genera were identified in aquatic systems *viz. Emticicia aquatica* (Seo *et al.*, 2015), *Emticicia aquatilis* (Ngo *et al.*, 2017), and *Emticicia fontis* (Nam *et al.*, 2016). The bacterial genera *Flavobacterium* was found in all the other three districts except in the south. *Flavobacterium* a gram-negative bacterium belongs to phylum *Bacteroidetes* and are widely distributed in a freshwater ecosystem (Fernández-Gómez *et al.*, 2013). *Flavobacterium* is responsible for bacterial cold water and bacterial gill diseases in different fish species (Strepparava *et al.*, 2014). They are also opportunistic human pathogens, there are several reports on their association with pneumonia and bloodstream infection (Manfredi *et al.*, 1999; Holmes *et al.*, 1984). *Rhodobacter* and *Acinetobacter* were found in the south district which was absent in spring waters of other districts. The diversity of *Rhodobacter* in an aquatic system is ubiquitous; they are photosynthetic bacteria belonging to phylum *Proteobacteria* (IMHOFF *et al.*, 1984). Some of the important *Rhodobacter sp.* identified from aquatic environment are *Rhodobacter adriaticus* isolated from Adriatic Sea (IMHOFF *et al.*, 1984) ; *Rhodovulum aestuarii* isolated from brackish water collected from an estuary (Ramana *et al.*, 2016); *Rhodobacter azollae* and *Rhodobacter lacus* isolated from different pond samples of Kukatpally, India (Suresh *et al.*, 2017); *Rhodobacter vinaykumarii* a marine phototrophic *alpha-proteobacterium* from tidal waters in Visakhapatnam, on the east coast of India (Srinivas *et al.*, 2007). There are also reports from Himalayan regions *viz. Rhodobacter changlensis* which was isolated from snow sample of Changlapass in the Indian Himalaya (Anil Kumar *et al.*, 2007). Presence of *Acinetobacter* in

362 rural drinking water systems dates back to 1989 (Bifulco *et al.*, 1989). Since then, there are
363 several reports of its presence in drinking water sources (Towner, 2006; Krizova *et al.*, 2015;
364 Radolfova-Krizova *et al.*, 2016) and transmission of drug-resistant *Acientobacter* through oral
365 route (Umezawa *et al.*, 2015). Principle component analysis **(Fig. 6b)** of the microbial diversity
366 of four districts showed East and South is correlated in genus wised distribution and they are
367 distantly related with the Wast district but North district produced an out-grouped showing the
368 divergence in diversity.

369 Canonical Correspondence Analysis **(Fig. 7)** confirmed the correlation between hydrochemistry
370 and diversity. The diversity in the springs of the four districts is influenced by the concentration
371 of different metallic compounds like sodium, calcium, barium, and iron. The diversity of
372 *Emticicia* and *Flavobacterium* are influenced by the concentration of sodium and calcium while
373 the diversity of *Planctomyces* was found to be dependent on the concentration of Barium and
374 Iron. Along with the different chemical parameters, physical parameters like temperature, pH
375 and electro-conductivity also had an influence on the diversity of microorganisms. The
376 dominance of *Arcicella*, *Schlesneria*, *Brevifolis* and *Sediminibactrium* are influenced by the
377 electroconductivity and total dissolved solids respectively. The results of this study significantly
378 expand the current understanding of microbiology of the spring water of Sikkim and it also
379 reports the comprehensive knowledge on microbial community structure which will help in near
380 future to determine or to design any water treatment protocols or policies. This study also
381 provides a brief insight of the physicochemical parameters of the spring water of Sikkim and
382 their cross association with the indigenous microbial diversity.

383

Acknowledgement

The authors wish to thank the State Institute of Rural Development Department, Government of Sikkim for their helping hand in providing information about springs locations and water sample collection. The authors also like to thank all the faculty member, non-teaching staff of the Department of Microbiology, Sikkim University for their continuous support and help throughout the study.

Reference

Anil Kumar P, Srinivas TNR, Sasikala C, Ramana C V. (2007). Rhodobacter changlensis sp. nov., a psychrotolerant, phototrophic alphaproteobacterium from the Himalayas of India. *Int J Syst Evol Microbiol* **57**: 2568–2571.

Bifulco JM, Shirey JJ, Bissonnette GK. (1989). Detection of Acinetobacter spp. in rural drinking water supplies. *Appl Environ Microbiol* **55**: 2214–9.

Boedeker C, Schüler M, Reintjes G, Jeske O, van Teeseling MCF, Jogler M, *et al.* (2017). Determining the bacterial cell biology of Planctomycetes. *Nat Commun* **8**: 14853.

Brown M V., Philip GK, Bunge JA, Smith MC, Bissett A, Lauro FM, *et al.* (2009). Microbial community structure in the North Pacific ocean. *ISME J* **3**: 1374–1386.

Chen W-M, Yang S-H, Young C-C, Sheu S-Y. (2013). Arcicella rigui sp. nov., isolated from water of a wetland, and emended descriptions of the genus Arcicella, Arcicella aquatica, Arcicella rosea and Arcicella aurantiaca. *Int J Syst Evol Microbiol* **63**: 134–140.

Das S, Bora SS, Yadav RNS, Barooah M. (2017). A metagenomic approach to decipher the indigenous microbial communities of arsenic contaminated groundwater of Assam. *Genomics*

405 *Data* **12**. e-pub ahead of print, doi: 10.1016/j.gdata.2017.03.013.

406 Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, *et al.* (2008). Functional
407 metagenomic profiling of nine biomes. *Nature* **452**: 629–632.

408 Fernández-Gómez B, Richter M, Schüler M, Pinhassi J, Acinas SG, González JM, *et al.* (2013).
409 Ecology of marine Bacteroidetes: a comparative genomics approach. *ISME J* **7**: 1026–37.

410 Fieseler L, Horn M, Wagner M, Hentschel U. (2004). Discovery of the novel candidate phylum
411 "Poribacteria" in marine sponges. *Appl Environ Microbiol* **70**: 3724–32.

412 Griffiths E, Gupta RS. (2007). Phylogeny and shared conserved inserts in proteins provide
413 evidence that Verrucomicrobia are the closest known free-living relatives of chlamydiae.
414 *Microbiology* **153**: 2648–2654.

415 Gupta RS. (2016). Impact of genomics on the understanding of microbial evolution and
416 classification: the importance of Darwin's views on classification Danchin A (ed). *FEMS*
417 *Microbiol Rev* **40**: 520–553.

418 Hemme CL, Tu Q, Shi Z, Qin Y, Gao W, Deng Y, *et al.* (2015). Comparative metagenomics
419 reveals impact of contaminants on groundwater microbiomes. *Front Microbiol* **6**: 1–12.

420 Holmes B, Owen RJ, Steigerwalt AG, Brenner ' DJ. (1984). Flavobacterium gleum, a New
421 Species Found in Human Clinical Specimens. *Int J Syst Bacteriol* **34**: 21–25.

422 Hörmannsperger G, Clavel T, Haller D. (2012). Gut matters: Microbe-host interactions in
423 allergic diseases. *J Allergy Clin Immunol* **129**: 1452–1459.

424 Hugenholtz P, Pitulle C, Hershberger KL, Pace NR. (1998). Novel division level bacterial
425 diversity in a Yellowstone hot spring. *J Bacteriol* **180**: 366–76.

IMHOFF JF, TRuPER HG, PFENNIG N. (1984). Rearrangement of the Species and Genera of the Phototrophic "Purple Nonsulfur Bacteria" *Int J Syst Bacteriol* **34**: 340–343.

Jones RT, Robeson MS, Lauber CL, Hamady M, Knight R, Fierer N. (2009). A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J* **3**: 442–453.

Kampfer P, Lodders N, Busse H-J. (2009). Arcicella rosea sp. nov., isolated from tap water. *Int J Syst Evol Microbiol* **59**: 341–344.

Krizova L, Maixnerova M, Sedo O, Nemec A. (2015). Acinetobacter albensis sp. nov., isolated from natural soil and water ecosystems. *Int J Syst Evol Microbiol* **65**: 3905–3912.

Lagkouvardos I, Joseph D, Kapfhammer M, Giritli S, Horn M, Haller D, *et al.* (2016). IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Sci Rep* **6**: 1–9.

Lee E, Khurana MS, Whiteley AS, Monis PT, Bath A, Gordon C, *et al.* (2017). Novel primer sets for next generation sequencing-based analyses of water quality. *PLoS One* **12**: 1–17.

Lee KCY, Dunfield PF, Stott MB. (2014). The Phylum Armatimonadetes. In: *The Prokaryotes*. Springer Berlin Heidelberg: Berlin, Heidelberg, pp 447–458.

Manfredi R, Nanetti A, Ferri M, Mastroianni A, Coronado O V, Chiodo F. (1999). Flavobacterium spp. organisms as opportunistic bacterial pathogens during advanced HIV disease. *J Infect* **39**: 146–52.

McLain JET, Rock CM, Lohse K, Walworth J. (2011). False-positive identification of *Escherichia coli* in treated municipal wastewater and wastewater-irrigated soils. *Can J Microbiol*

57: 775–784.

Nam GG, Joung Y, Song J, Lim Y, Cho J-C. (2016). Emticicia fontis sp. nov., isolated from a freshwater pond. *Int J Syst Evol Microbiol* **66**: 5161–5166.

Ngo HTT, Trinh H, Yang J-E, Won K-H, Chu D-H, Kook M, *et al.* (2017). Emticicia aquatilis sp. nov., isolated from a freshwater sample. *Int J Syst Evol Microbiol* **67**: 1703–1708.

Nikitin DI, Strömpl C, Oranskaya MS, Abraham W-R. (2004). Phylogeny of the ring-forming bacterium Arcicella aquatica gen. nov., sp. nov. (ex Nikitin et al. 1994), from a freshwater neuston biofilm. *Int J Syst Evol Microbiol* **54**: 681–684.

Pandey PK, Kass PH, Soupir ML, Biswas S, Singh VP. (2014). Contamination of water resources by pathogenic bacteria. *AMB Express* **4**: 51.

Port JA, Wallace JC, Griffith WC, Faustman EM. (2012). Metagenomic Profiling of Microbial Composition and Antibiotic Resistance Determinants in Puget Sound. *PLoS One* **7**: e48000.

Radolfova-Krizova L, Maixnerova M, Nemec A. (2016). Acinetobacter pragensis sp. nov., found in soil and water ecosystems. *Int J Syst Evol Microbiol* **66**: 3897–3903.

Ramana C V., Sasikala C, Lakshmi KVNS, Bharti D, Divyasree B. (2016). Rhodovulum aestuarii sp. nov., isolated from a brackish water body. *Int J Syst Evol Microbiol* **66**: 165–171.

Van Rossum T V., Peabody MA, Uyaguari-Diaz MI, Cronin KI, Chan M, Slobodan JR, *et al.* (2015). Year-long metagenomic study of river microbiomes across land use and water quality. *Front Microbiol* **6**: 1–15.

Rothschild LJ. (1991). A model for diurnal patterns of carbon fixation in a Precambrian microbial mat based on a modern analog. *BioSystems* **25**: 13–23.

468 Sanapareddy N, Hamp TJ, Gonzalez LC, Hilger HA, Fodor AA, Clinton SM. (2009). Molecular
469 diversity of a north carolina wastewater treatment plant as revealed by pyrosequencing. *Appl*
470 *Environ Microbiol* **75**: 1688–1696.

471 Schultz GE, Kovatch JJ, Anneken EM. (2013). Bacterial diversity in a large, temperate, heavily
472 modified river, as determined by pyrosequencing. *Aquat Microb Ecol* **70**: 169–179.

473 Seo M, Joung Y, Kang H, Kim H, Cho J-C, Ahn T, *et al.* (2015). Emticicia aquatica sp. nov., a
474 species of the family Cytophagaceae isolated from fresh water. *Int J Syst Evol Microbiol* **65**:
475 4358–4362.

476 Sheu S-Y, Yang C-S, Chen M-H, Arun AB, Young C-C, Chen W-M, *et al.* (2018). Arcicella
477 aurantiaca sp. nov., isolated from stream water. e-pub ahead of print, doi: 10.1099/ijs.0.021527-
478 0.

479 Sogin ML, Morrison HG, Huber HA, Welch DM, Huse SM, Neal PR, *et al.* (2006). Microbial
480 Diversity in the Deep Sea and the Underexplored 'Rare Biosphere'. *PNAS* **103**: 12115–12120.

481 Srinivas TNR, Anil Kumar P, Sasikala C, Ramana C V., Imhoff JF. (2007). Rhodobacter vinay
482 kumarii sp. nov., a marine phototrophic alphaproteobacterium from tidal waters, and emended
483 description of the genus Rhodobacter. *Int J Syst Evol Microbiol* **57**: 1984–1987.

484 Staley C, Sadowsky MJ. (2016). Application of metagenomics to assess microbial communities
485 in water and other environmental matrices. *J Mar Biol Assoc United Kingdom* **96**: 121–129.

486 Staley C, Unno T, Gould TJ, Jarvis B, Phillips J, Cotner JB, *et al.* (2013). Application of
487 Illumina next-generation sequencing to characterize the bacterial community of the Upper
488 Mississippi River. *J Appl Microbiol* **115**: 1147–1158.

Strepparava N, Wahli T, Segner H, Petrini O. (2014). Detection and quantification of Flavobacterium psychrophilum in water and fish tissue samples by quantitative real time PCR. *BMC Microbiol* **14**: 105.

Suresh G, Sailaja B, Ashif A, Dave BP, Sasikala C, Ramana CV. (2017). Description of Rhodobacter azollae sp. nov. and Rhodobacter lacus sp. nov. *Int J Syst Evol Microbiol* **67**: 3289–3295.

Tamaki H, Tanaka Y, Matsuzawa H, Muramatsu M, Meng X-Y, Hanada S, *et al.* (2011). Armatimonas rosea gen. nov., sp. nov., of a novel bacterial phylum, Armatimonadetes phyl. nov., formally called the candidate phylum OP10. *Int J Syst Evol Microbiol* **61**: 1442–1447.

Tambe S, Kharel G, Subba S, Arrawatia ML. (2012). Rural Water Security in the Sikkim Himalaya : Status , Initiatives and Future Strategy.

Towner K. (2006). The Genus Acinetobacter. In: *prokaryotes*. pp 746–758.

Umezawa K, Asai S, Ohshima T, Iwashita H, Ohashi M, Sasaki M, *et al.* (2015). Outbreak of drug-resistant Acinetobacter baumannii ST219 caused by oral care using tap water from contaminated hand hygiene sinks as a reservoir. *Am J Infect Control* **43**: 1249–1251.

Vörösmarty CJ, McIntyre PB, Gessner MO, Dudgeon D, Prusevich A, Green P, *et al.* (2010). Global threats to human water security and river biodiversity. *Nature* **467**: 555–561.

Wang X, Wang Z, Jiang P, He Y, Mu Y, Lv X, *et al.* (2018). Bacterial diversity and community structure in the rhizosphere of four Ferula species. *Sci Rep* **8**: 5345.

WHO. (2018). WHO | Water safety and quality. *WHO*. http://www.who.int/water_sanitation_health/water-quality/en/ (Accessed February 5, 2018).

510

511 **Supplementary Files**

512 **Table S1.** The quantity of gDNA as determined by Qubit Fluorometer and mean library

513 fragment size.

514

Sl. No.	Sample ID	The quantity of gDNA ng/ µl	Mean Library Fragment Size (bp)
1.	E	31.7	605
2.	W	15.0	539
3.	S	19.8	598
4.	N	18.5	605

515

516 **Table S2:** FastQ read statistics.

Sample	Number of Reads	Total Bases	Observed OTUs
E	248, 903	116,834,615	687
W	316,594	151,955,051	747
S	198,909	94,297,897	534
N	287,074	198,132,358	729

517

518 **Table S3: Diversity Index**

	S	W	N	E
Simpson_1-D	0.7499	0.7356	0.6988	0.6928

Shannon_H	1.534	1.49	1.36	1.368
Fisher_alpha	8.719	10.7	13.71	8.107
Chao-1	22	25	29	21

519