CHAPTER VIII

Metagenomics Study of *Barilius bendelisis*

VIII.1. Introduction

Metagenomics is the study of the microbial diversity i.e. microbiome in an ecosystem. A microbiome is defined as the collection of microbiota such as bacteria, viruses, archaea and eukaryotes (Backhed et al., 2005; Ellermann et al., 2017). They are ubiquitous in nature and found in the entire environment such as soil, air, water as well as in association with organisms body parts like skin, lung, gastrointestinal, etc. The microbial diversity study with the functional abilities in the specific habitat plays a significant role in understanding more about microbial evolution and ecology (Porchas and Albores, 2015). The studies of microbial diversity are carried out by identifying the different microflora from various environmental sources. However, such studies showed some scientific and technological limitations such as culture-dependent approaches followed by bacterial biochemical identification, which have a much rigorous labor process because of the incapability to culture. Isolation of the microflora and biochemical identification of a huge number of microbes from a sample is a laborious process with many other major limitations. As identifications of all bacterial colonies are not possible, so, the population demonstration pattern of the microbes in the sample vanishes (Tyagi and Singh, 2017). Moreover, the viable but large numbers of non-culturable nature of bacteria are major obstacles for a culture-dependent method (Handelsman, 2004; Streit and Schmitz, 2004). Because of these problems, many researchers have tried to develop new methods for the study of microbial diversity. With that aim about 30 years ago, Pace and co-workers in 1986, for the first time developed a new method by taking the revolutionary idea of DNA cloning from the environmental samples directly to analyze the natural microbial populations complexity (Alves et al., 2018; Pace et al., 1986). After that in 1998, Handelsman established the term called metagenome (Handelsman et al., 1998). It refers to the collection of all genomes and genes of the microbiota from environmental samples like soil, water, or the gut of eukaryotic host organisms and it was transformed into a suitable vector and can identify the whole microbial genome by

some molecular techniques with the help of computer software (Handelsman et al., 1998). The term metagenome is segmented into two approaches mainly with different local microbial community characteristics. The first one is structural metagenomic method indicating study of uncultivated microbial population structure. The community structure study allows to have a deep knowledge about the dealings between the individual components and this is important for interpreting biological or ecological functions among the members (Tringe et al., 2005; Vieites et al., 2009). The second one is functional metagenomic method helping in gene identification and involving the generation of libraries expression with huge numbers of metagenomic clones and that is followed by the activity-based screenings (Schmeisser et al., 2007; Guazzaroni et al., 2015). The 16S rRNA gene investigations are often mentioned as metagenomic studies, where single gene study is focused and can find a link between the analyses of 16S rRNA to metabolic pathways, and is found to be a suitable technique in establishing the functional potential of a microbiome (Li et al., 2008; Thomas et al., 2012; Hunt et al., 2013; Langille et al., 2013). Sanger sequencing technology provided important progress in the analysis of microbial diversity at the initial stage metagenomic studies (Sanger et al., 1977; Gillespie et al., 2002; Breitbart et al., 2003; Uchiyama et al., 2005). A low cost advanced sequencing technology is used for simultaneous sequencing capability of millions of DNA fragments. (Turnbaugh et al., 2009; Klindworth et al., 2013; Oulas et al., 2015; Sunagawa et al. 2015). Aquaculture environment based microbial communities may contain a good relationship with inhabitant microbiota of fishes. Therefore, microbial diversity study of microbiota gut may play a crucial role in fish physiology and health (Ghanbari et al., 2015). 16S rRNA gene sequencing has shown a new light on the composition and diversity of microbial communities within several animal gut systems with the initiation of Sanger sequencing and PCR amplification which is followed by cloning and sequencing, and this has become an important part of environmental microbial phylogenetic and taxonomic characterization (Ley et al., 2008; Lamendella et al., 2011). The microbial gut flora of fish has been studied by several workers including a description of microbial spoilage (Joseph et al., 1988), the relation between environment and fish microflora (Horsley, 1977), monitoring change in fish form (Allen et al., 1983), microbial flora as the food of fish (Kamjunke et al., 2002), microbial flora in production of enzymes (Bairagi et al., 2002) and antibiotic resistance profile of indigenous flora (Spanggaard et al., 2000). The microflora of reared fish has also been studied as a source of protection against diseases (Sissons, 1989). For all these reasons, the study

of microbial flora is important. As the function and structure of the gut microbial community have received significant attention for decades, the present study focuses on the metagenomic study of *Barilius bendelisis* species based on 16S rRNA gene from the gut microbial community. In this study, gut microbiota was collected from the *Barilius bendelisis* species, which is a freshwater fish commonly encountered in the Indian subcontinent and is very common throughout the Western Ghats of India (Raagam and Devi, 2005). It is also a common fish in the *Hel* river of Kokrajhar, Assam, and has a good population density. It is commonly sold as a food fish in the local market of Serfanguri Kokrajhar and is locally known as *Elang*. It has a moderate commercial value. It attains a length of 10–20 cm and weighs about 25–40 g. In the present study, *Barilius bendelisis* fish species belonging to the family of *Cyprinidae* was found to be the highest (15.31 %) in the *Hel* river based on its relative abundance study, and hence, it was selected for the metagenomics study.

VIII.2. Materials and Methods

VIII.2.1. Sample preparation

The sample after the collection was dissected with fine sterile scissor and was submerged immediately with PBS (phosphate-buffered saline) buffer solution and then it was used for further experiment.

VIII.2.2.Method for genomic DNA (gDNA) isolation

Genomic DNA was isolated by using “Xcelgen blood and tissue kit” (Chakraborty et al., 2016). According to the kit, the following steps were followed.

Step 1. 20 µL of proteinase K with 10 µL of RNase was pipetted in a centrifuge tube of 1.5 mL.

Step 2. 10–50 mg of gut fish sample was taken and it was ground in liquid nitrogen for complete lysis.

Step 3. In the sample, 180 µL of lysis buffer was added and mixed to get homogeneous solution by inverting the tubes 8–10 times.

Step 4. It was then incubated at 65°C for 40 min to yield maximum DNA. After incubation, it was then centrifuged for removal of drops from inside of the lid.

Step 5. In the sample, ethanol (200 µL) was added and again mixed by inverting 8–10 times, and centrifuged again.

Step 6. The spin column (2 mL collection tube) was applied without wetting the rim. It was centrifuged for 1 min at 8000 rpm closing the cap.

Step 7. Carefully the spin column was opened and 500 µL of WB1 (kit chemical) was added. It was centrifuged again for 1 min at 8000 rpm by closing the cap. Then the spin column was placed in a collection tube (2 mL) and the tube containing the filtrate was discarded.

Step 8. After that, the spin column was carefully opened and 500 µL of WB2 (kit chemical) was added without wetting the rim. The cap was closed and centrifuged at 8000 rpm for 1 min.

Step 9. Then the spin column was placed in a new collection tube (2 mL) and the filtrate in the old collection tube was discarded, and centrifuged again for 3 min.

Step 10. The spin column was placed in a micro-centrifuge tube (1.5 mL) and the filtrate in the collection tube was discarded. Carefully the spin column was opened and 50 µL elution buffer was added, and incubated for 1 min at room temperature (15–25°C) and then centrifuged for 1 min with full speed.

VIII.2.3. DNA confirmation and quantification

DNA extraction was confirmed by 1.2 % agarose gel electrophoresis with 1 Kb (Kilo base pair) marker and the quality of DNA was checked by using NanoDrop machine.

VIII.2.4. PCR amplification of DNA

The isolated gDNA was amplified using the specific primer V3-V4 hyper-variable region (**Table VIII.1**) for 16S rRNA gene. PCR reaction was carried out in a final reaction volume of 25 µL PCR tube in a Thermal Cycler (Veriti®96 well Thermal Cycler, Model No. 9902). The PCR reaction mixture of 25 µL was prepared containing 2.5–3.0 µL of PCR buffering (10×), 1.0 µL of F primer (10 pmol/µL), 1.0 µL of R primer (10 pmol/µL), 2.0 µL of 20 mM $MgCl_2$, 2.0 µL of 2.5 mM dNTPs (Deoxynucleotide triphosphates), 2.0 µL of Tag DNA polymerase (1 unit/µL), 5.0 µL of template DNA and remaining volume was made up to 25 µL using nucleus free water. The conditions of PCR were initial denaturation for 4 min at 94°C followed by 35 cycles at 94°C for 40 s, annealing at 50°C for 45 s, 72°C for 45 min and the final extension at 72°C for 20 min. The PCR tubes containing the mixture were gently tapped and spin. Then the PCR tube with the component was transferred to a thermal cycler.

Table VIII.1. **Primers used for 16S rRNA in the present study**

Oligo name	Oligo sequence (5′ to 3′)	Length of primer	Product size (Approx.)
Prokaryote V3-Forward	CCTACGGGNBGCASCAG	17	~ 460 bps
Prokaryote V4-Reverse	GACTACNVGGGTATCTAATCC	21	

VIII.2.5. Qualitative and quantitative analysis of PCR products

The PCR product quality was checked on 1.5 % agarose gel (loaded 5 µL) for the presence of the single intact band. The gel was run at 110 V (volt) for 30 min in electrophoresis. Quantification of DNA samples was done following Qubit dsDNA HS Assay kit (Life Tech), and each sample (1 µL) was used for concentration determination using Qubit® 2.0 Fluorometer.

VIII.2.6. Preparation of libraries for 2 × 250 bp (base pair) Run chemistry

Preparation of amplicon library was performed using Nextera XT Index Kit (Illumina Inc.) as per the protocol of 16S metagenomic sequencing library (Part # 15044223 Rev. B). Primers for amplification of V3-V4 hyper-variable region (**Table VIII.1**) of 16S rRNA gene of bacteria and archaea were in-house designed by Xcelris Labs Limited Xcelris NGS Bioinformatics Lab. These primers were synthesized in Xcelris PrimeX facility. As per standard Illumina protocol, amplification of the amplicon with Illumina adaptors was done using i5 and i7 primers that add multiplexing index sequences and common adapters required for the cluster generation (P5 and P7).

VIII.2.7. Cluster generation and sequencing

After getting Qubit concentration for library and mean peak size from the profile of Bioanalyzer, loading of library into Illumina platform was done at the concentration (10–20 PM, particulate matter) for generation and sequencing of cluster. Paired-end sequencing allows the template fragments to be sequenced in both forward and reverse directions on Illumina platform. The kit reagents were used in the binding of samples to complementary adapter oligos on the paired-end

flow cell. Designing of adapters were done to cleave the forward strands selectively after re-synthesis of the reverse strand during sequencing. Then the copied reverse strand was used to sequence from the opposite end of the fragment.

VIII.2.8. Wet-Lab Inference

The library was prepared from the samples after amplifying V3-V4 region 16S segment. The 16S library mean size is 649 bp. The library was sequenced using Illumina 2 × 250 bp to generate ~150 Mb (Mega base pair) of data per library.

VIII.2.9. Bioinformatics analysis (Reads statistics)

The next-generation sequencing was performed using 2 × 250 PE (Paired-end) chemistry on the Illumina platform.

VIII.2.10. QIIME overview and steps for 16S analysis

QIIME (Quantitative insight into microbial ecology) is a suitable open-source tool for the analysis of 16S metagenome data from the platforms of NGS (Next Generation Sequencing) and was implemented in python language. The steps involved for 16S analysis through QIIME were mentioned below.

Step 1. The chimeras DNA sequences were first filtered using the usearch61 algorithm (abundance-based, de novo) from flashed/stitched data. Then, the non-chimeric sequences identified were used for OTU (Operational Taxonomic Unit) pick.

Step 2. The next step involved the cluster of similar sequences and represented together in a form of OTU. All the sequences were clustered into OTUs on the basis of similarity of their sequence and then frequently intended to represent the degree of taxonomic relatedness. The resulting cluster typically represents the same species.

Step 3. The next step consists of picking a representative sequence for each of these OUT's picked. Here, the OTUs consisted of only one sequence was removed, thus retaining OTUs having at least 2 sequences.

Step 4. In the next step, taxonomic names to these representative sequences at 90% sequence similarity were assigned. This was done using UCLUST algorithm, where the query is our representative sequences and subjects are the curated sequences at the green genes database.

Step 5. At last, the diversity calculation for each sample and comparison of the types of communities using the taxonomic assignments was done.

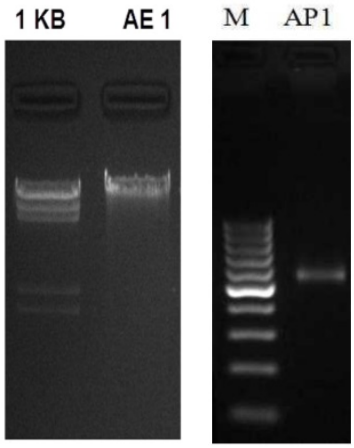


Fig. VIII.1. 1.2% Agarose gel showing extracted DNA of the sample with 1 Kb (Kilo base pair) marker (AE1) and PCR product on 1.5 % Agarose gel (AP1).

Table VIII.2. gDNA quality and quantification NanoDrop machine

Sample	Absorbance at 260/280 nm	Concentration (ng/µL)
Barilius bendelisis	1.91	14.15

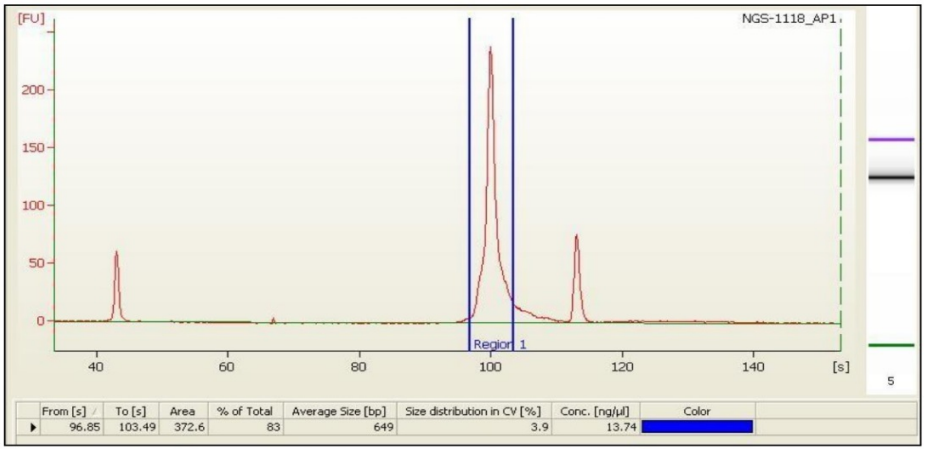


Fig. VIII.2. Bioanalyzer 2100 profile of final libraries of 16S metagenome using DNA 1000 chip.

VIII.3. Results and Discussion

In this study, the result of isolated genomic DNA is shown in **Fig. VIII.1** and **Table VIII.2**. The high quality of DNA was extracted and the quantity was found to be 14.15 ng/µL (**Table VIII.2**). The PCR amplification using the primer V3-V4 hyper-variable region (**Table VIII.1**) showed the successful amplification of DNA (**Fig. VIII.1**) and was quantified by Qubit Fluorometer 2.0 and the concentration was found as 11.4 ng/µL. The purification of amplicon libraries were done by 1X AMpureXP beads, checked on Agilent DNA 1000 chip (2100 Bioanalyzer) (**Fig. VIII.2)**. The reads statistics data (**Table VIII.3**) showed the PE reads as 249,250, total reads (R1+R2) as 498,500, total bases (R1+R2) as 124,625,000 and the flash reads found as 233,473. The result of OTUs in a summary form is shown in **Table VIII.4**.

Table VIII.3. Reads statistics

Sample	PE (Paired-end) reads	Total reads (R1+R2)	Total bases (R1+R2)	Flash reads
Barilius bendelisis	249,250	498,500	124,625,000	233,473

Table VIII.4. OTUs (Operational taxonomic units) summary

Sample	Flash/stitch reads	Non-chimeric sequences	OTUs	OTUs with no singletons
Barilius bendelisis	233,473	227,212	12,549	1948

The results of the taxonomy assignment reported the phylum to species level of taxonomic distributions representing the relative abundance of each class within each microbial community. The results of the taxonomic distribution of *Barilius bendelisis* gut flora at the phylum level indicated that *Proteobacteria* and *Firmicutes* were the most abundantly found in the species (**Fig. VIII.3**; Pie diagram). The top 5 phyla represented were *Firmicutes* (74.8 %), *Proteobacteria* (20.2 %), *Bacteroidetes* (3.9 %), *Cyanobacteria* (0.2 %) and others (0.7 %). The result also showed *Actinobacteria*, *Chloroflexi*, *Planctomycetes*, TM6, TM7, *Tenericutes*, and *Verrucomicrobia* along with some unassigned phyla. Therefore, a total of 12 phyla were found as shown in **Fig. VIII.3** (Pie diagram and % abundant chart list).

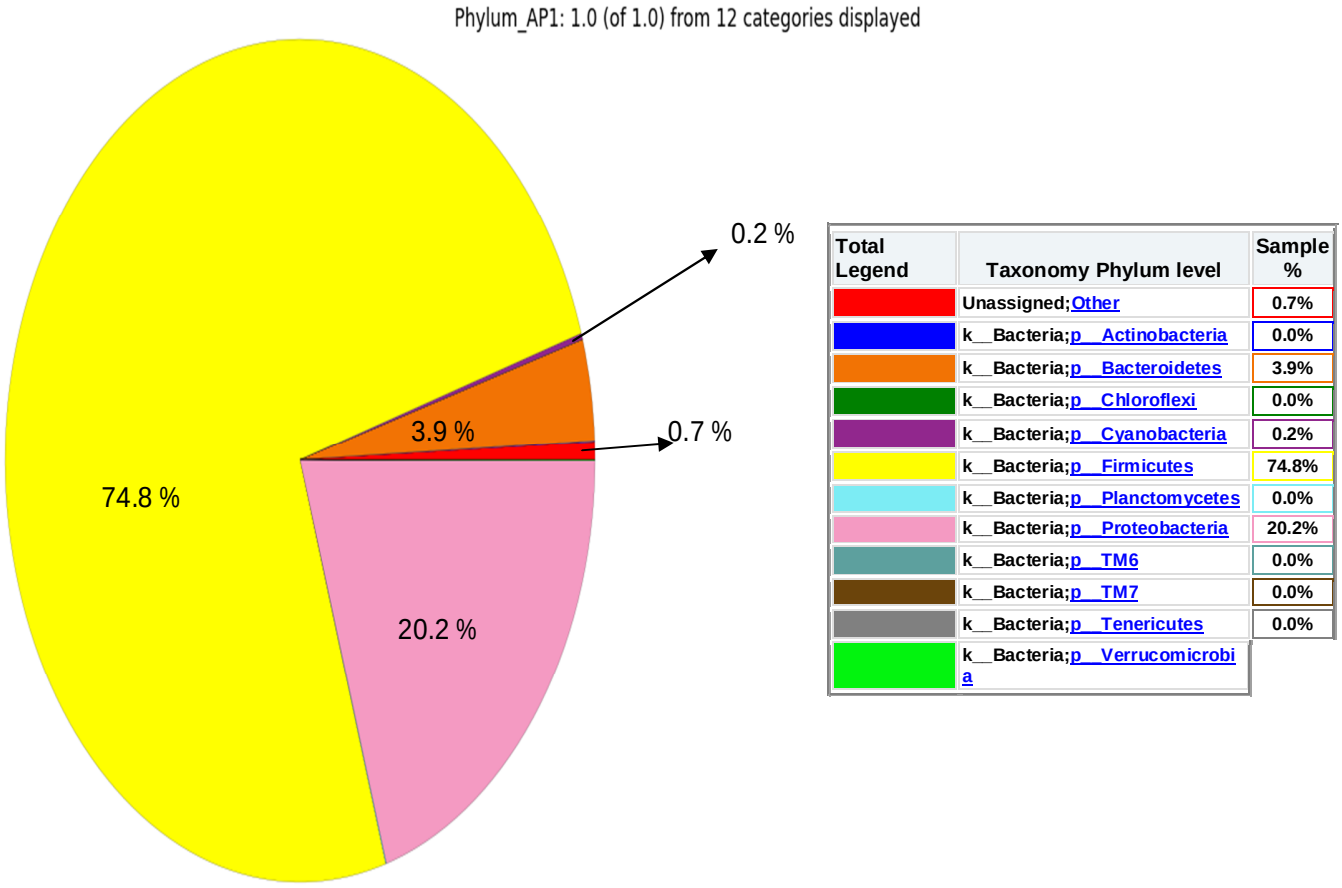
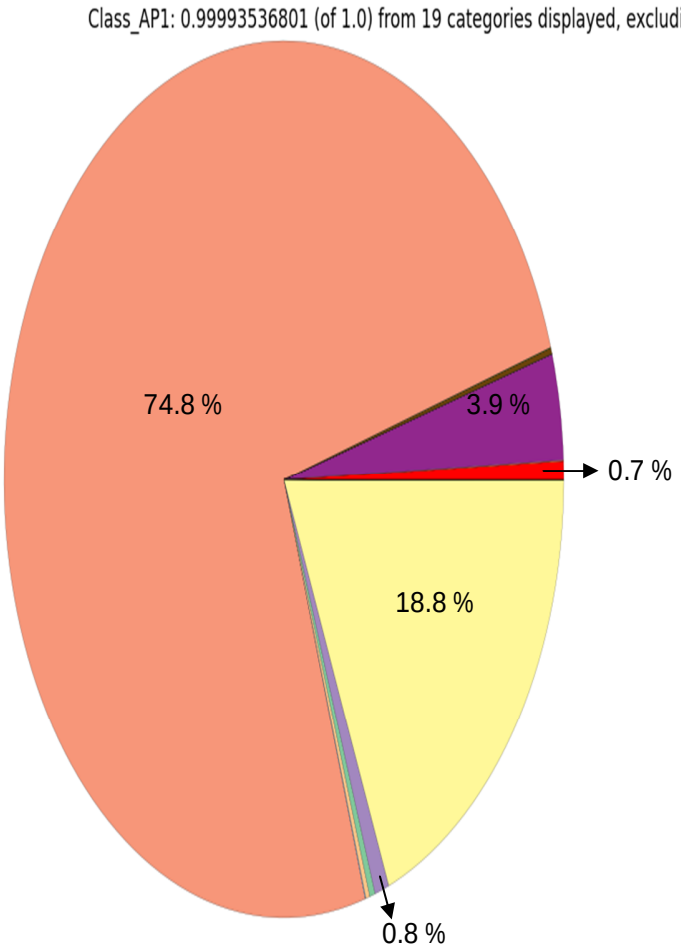


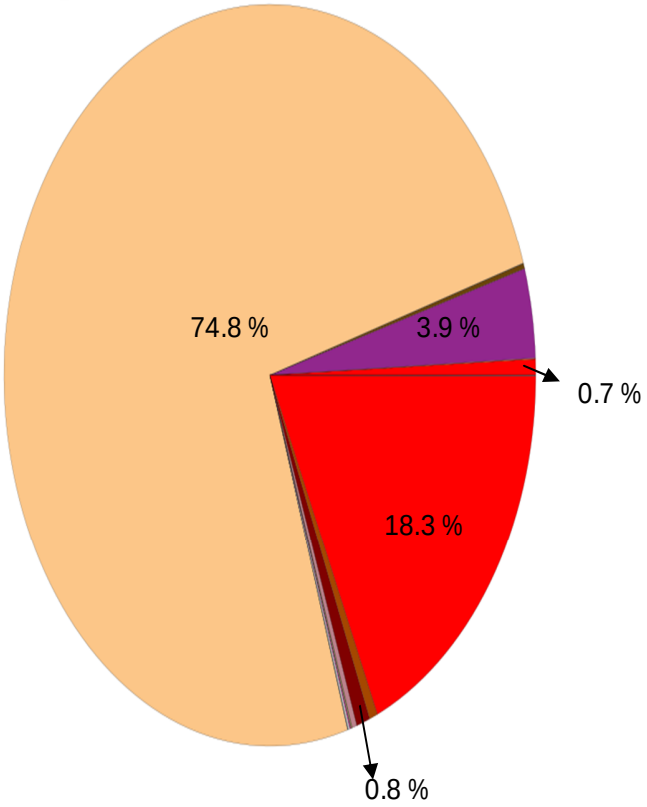
Fig. VIII.3. Taxonomy at phylum level (Top 5 phyla shown in Pie diagram).



Total Legend	Taxonomy Class level	Sample %
	Unassigned;Other; Other	0.7%
	k__Bacteria;p__Actinobacteria;c__Acidimicrobiia	0.0%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria	0.0%
	k__Bacteria;p__Actinobacteria;c__Thermoleophilia	0.0%
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia	3.9%
	k__Bacteria;p__Bacteroidetes;c__Sphingobacteriia	0.0%
	k__Bacteria;p__Chloroflexi;c__Anaerolineae	0.0%
	k__Bacteria;p__Chloroflexi;c__Chloroflexi	0.0%
	k__Bacteria;p__Cyanobacteria;c__4C0d-2	0.0%
	k__Bacteria;p__Cyanobacteria;c__Chloroplast	0.2%
	k__Bacteria;p__Cyanobacteria;c__Synechococcophycideae	0.0%
	k__Bacteria;p__Firmicutes;c__Bacilli	0.0%
	k__Bacteria;p__Firmicutes;c__Clostridia	74.8%
	k__Bacteria;p__Planctomycetes;c__Planctomycetia	0.0%
	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria	0.2%
	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria	0.3%
	k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria	0.8%
	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria	18.8%
	k__Bacteria;p__TM6;c__SJA-4	0.0%
	k__Bacteria;p__TM7;c__TM7-1	0.0%
	k__Bacteria;p__Tenericutes;c__CK-1C4-19	0.0%
	k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae	0.0%
	k__Bacteria;p__Verrucomicrobia;c__[Spartobacteria]	0.0%

Fig. VIII.4. Taxonomy at class level (Top 5 classes shown in Pie diagram).

Order_AP1: 0.999399845806 (of 0.999999999999) from 19 categories displayed, excluding 0.000600154193462 from 19 categories ('All Other Categories')



Total Legend	Taxonomy Order level	Sample %
	Unassigned;Other;Other;Other	0.7%
	k__Bacteria;p__Actinobacteria;c__Acidimicrobiia;o__Acidimicrobiales	0.0%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales	0.0%
	k__Bacteria;p__Actinobacteria;c__Thermoleophilia;o__Solirubrobacterales	0.0%
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	3.9%
	k__Bacteria;p__Bacteroidetes;c__Sphingobacteriia;o__Sphingobacteriales	0.0%
	k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__Caldilineales	0.0%
	k__Bacteria;p__Chloroflexi;c__Chloroflexi;o__[Roseiflexales]	0.0%
	k__Bacteria;p__Cyanobacteria;c__4C0d-2;o__YS2	0.0%
	k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Chlorophyta	0.2%
	k__Bacteria;p__Cyanobacteria;c__Synechococcophycideae;o__Pseudanabaenales	0.0%
	k__Bacteria;p__Cyanobacteria;c__Synechococcophycideae;o__Synechococcales	0.0%
	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales	0.0%
	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales	0.0%
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales	74.8%
	k__Bacteria;p__Planctomycetes;c__Planctomycetia;o__Gemmatales	0.0%
	k__Bacteria;p__Planctomycetes;c__Planctomycetia;o__Pirellulales	0.0%
	k__Bacteria;p__Planctomycetes;c__Planctomycetia;o__Planctomycetales	0.0%
	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales	0.1%
	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales	0.0%
	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales	0.0%
	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales	0.1%
	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales	0.0%
	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Methylophilales	0.0%
	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales	0.3%
	k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfarculales	0.0%
	k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfovibrionales	0.8%
	k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Spirobacillales	0.0%
	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales	0.5%
	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales	18.3%
	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales	0.0%
	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales	0.0%
	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales	0.0%
	k__Bacteria;p__TM6;c__SJA-4;o__	0.0%
	k__Bacteria;p__TM7;c__TM7-1;o__	0.0%
	k__Bacteria;p__Tenericutes;c__CK-1C4-19;o__	0.0%
	k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales	0.0%
	k__Bacteria;p__Verrucomicrobia;c__[Spartobacteria];o__[Chthoniobacterales]	0.0%

Fig. VIII.5. Taxonomy at order level (Top 5 orders shown in Pie diagram).

The results of the taxonomic distribution of *Barilius bendelisis* gut flora at the class level indicated that *Clostridia* and *Gammaproteobacteria* were the most abundantly present in the species (**Fig. VIII.4**; Pie diagram). The top 5 classes represented were *Clostridia* (74.8 %; p-*Firmicutes*), *Gammaproteobacteria* (18.8 %; p-*Proteobacteria*), *Bacteroidia* (3.9 %; p-*Bacteroidetes*), *Deltaproteobacteria* (0.8 %; p-*Proteobacteria*) and others (0.7 %). However, *Acidimicrobiia, Actinobacteria, Thermoleophilia,* (p-*Actinobacteria*), *Sphingobacteriia* (p-*Bacteroidetes*), *Anaerolineae, Chloroflexi* (p-*Chloroflexi*), 4Cod-2, *Chloroplast, Synechococcophycideae* (p-*Cyanobacteria*), *Bacilli* (p-*Firmicutes*), *Planctomycetia* (p-*Planctomycetes*), *Alphaproteobacteria, Betaproteobacteria,* (p-*Proteobacteria*), SJA-4 (p-TM6), TM7-1 (p-TM7), CK-1C4-19 (p-*Tenericutes*), *Verrucomicrobiae, Spartobacteria* (p-*Verrucomicrobia*) were also detected along with some unassigned classes. Therefore, a total of 23 classes belonging to 11 phyla were found as shown in **Fig. VIII.4** (Pie diagram and % abundant chart list).

The results of taxonomic distribution at order level showed that *Clostridiales* and *Enterobacteriales* were the most abundantly present in the species (**Fig. VIII.5**; Pie diagram), and the top 5 orders represented were *Clostridiales* (74.8 %; c-*Clostridia*), *Enterobacteriales* (18.3 %; c-*Gammaproteobacteria*), *Bacteroidales* (3.9 %; c-*Bacteroidia*), *Desulfovibrionales* (0.8 %; c-*Deltaproteobacteria*) and others (0.7 %). However, *Aeromonadales* (c-*Gammaproteobacteria*), *Neisseriales,* (c-*Betaproteobacteria*), *Rhizobiales, Sphingomonadales,* (c-*Alphaproteobacteria*), *Acidimicrobiales, Actinomycetales* (c-*Actinobacteria*), *Solirubrobacterales* (*c-Thermoleophilia*), *Caldilineales* (*c-Anaerolineae*), *Roseiflexales* (*c-Chloroflexi*), *YS2* (*c-4C0d-2*), *Chlorophyta* (c-*Chloroplast*), *Pseudanabaenales, Synechococcales* (*c-Synechococcophycideae*), *Bacillales, Lactobacillales* (c-*Bacilli*), *Gemmatales, Pirellulales, Planctomycetales* (c-*Planctomycetia*), *Rhizobiales, Rhodobacterales, Rhodospirillales* (*c-Alphaproteobacteria*), *Burkholderiales, Methylophilales* (c-*Betaproteobacteria*), *Desulfarculales, Spirobacillales (c-Deltaproteobacteria), Legionellales, Pseudomonadales, Xanthomonadales* (c-*Gammaproteobacteria*), *Verrucomicrobiales* (c-*Verrucomicrobiae*), *Chthoniobacterales* (c- *Spartobacteria*) were also detected along with some unknown and unassigned orders. Therefore, a total of 38 orders belonging to 21 classes were found as shown in **Fig. VIII.5** (Pie diagram and % abundant chart list).

In this study, the top 5 families (**Fig. VIII.A.1**) represented were *Veillonellaceae* (74.6 %; o-*Clostridiales*), *Enterobacteriaceae* (18.3 %; o-*Enterobacteriales*), *Bacteroidaceae* (3.9 %; o-*Bacteroidales*), *Desulfovibrionaceae* (0.8 %; o-*Desulfovibrionales*) and others (0.7 %). In the study, *Aeromonadaceae* (o-*Aeromonadales*), *Neisseriaceae* (*o-Neisseriales*), *Trebouxiophyceae, Chlamydomonadaceae* (o-*Chlorophyta*), *Clostridiaceae, Peptostreptococcaceae, Ruminococcaceae* (o-*Clostridiales*), *Sphingomonadaceae* (o-*Sphingomonadales*) were also detected along with some unknown and unassigned families. Therefore, a total of 61 families belonging to 35 orders were found as shown in **Fig. VIII.A.1**.

The top 5 genera shown in the taxonomic distribution at the genus level (**Fig. VIII.A.2.**) were *Citrobacter* (7.7 %; f-*Enterobacteriaceae*), *Bacteroides* (3.9 %; f-*Bacteroidaceae*), *Proteus* (3.4 %; f-*Enterobacteriaceae*), *Morganella* (2.4 %; f-*Enterobacteriaceae*) and *Enterobacter* (1.1 %; f-*Enterobacteriaceae*). However, *Serratia* (f-*Enterobacteriaceae), Desulfovibrio (f-Desulfovibrionaceae), Providencia (*f-*Enterobacteriaceae), Anaeromusa* (f-*Veillonellaceae*), *Escherichia, Klebsiella* (f-*Enterobacteriaceae*), *Microvirgula* (f-*Neisseriaceae*), *vadinHB04* (f-*Veillonellaceae*), *Clostridium* (f-*Clostridiaceae*), *Sphingomonas* (f- *Sphingomonadaceae*) and *Bilophila* (f-*Desulfovibrionaceae*) were also detected along with some unknown and unassigned genera. Therefore, a total of 113 genera belonging to 50 families were found as shown in **Fig. VIII.A.2**.

The taxonomic distribution of *Barilius bendelisis* gut flora at species level (**Fig. VIII.A.3.**) study showed the top 5 species and these were *morganii* (2.4 %; g-*Morganella*), *marcescens* (0.7 %; g-*Serratia*), coli (0.4 %; g-*Escherichia*), *aerodenitrificans* (0.3 %; g-*Microvirgula*), and *oxytoca* (0.2 %; g-*Klebsiella*) along with the presence of some unknown and unassigned species. Therefore, a total of 129 species belonging to 65 genera were found (**Fig. VIII.A.3.**).

The present study of the gut microbial community of the fish species using 16S rRNA gene Illumina platform showed to be a successful tool for the study of microbial diversity. However, the microbial composition varies among the species and locations, which may be due to age, environmental conditions, individuals, nutritional status, developmental stage, and the complexity of the fish digestive system (Wu et al., 2013). Tyagi et al. (2017) investigated the microbial diversity in association with the NGS of 16S rRNA gene and reported that *Proteobacteria* and *Verrucomicrobia* were the most predominant phyla in the samples of fish gut. They also reported the presence of known cellulose degrader organisms (*Clostridium spp*.,

Eubacterium spp., *Ruminococcus spp.* and *Bacteroides spp.*) and short-chain fatty acid producer organisms (*Veillonella spp., Faecalibacterium prausnitzii* and *Megasphaera spp.*). Garcia et al. (2018) reported that *Gammaproteobacteria, Bacteroidetes, Firmicutes* and *Betaproteobacteria* were the most abundant phyla of the stomach microbiome. Rimoldi et al. (2018) studied the profile of the gut microbial community using High-throughput 16S rRNA gene amplicon sequencing (MiSeq platform, Illumina) and reported the most predominant phyla which were *Firmicutes, Proteobacteria, Bacteroidetes* and *Actinobacteria*. Hsu et al. (2018) reported the most predominant intestinal microbiome of eel species and these were *Clostridium, Cetobacterium, Shewanella, Bacteroides* and *Acinetobacter* that appeared to be the great potential as probiotics. Bledsoe et al. (2016) reported the bacterial phyla present in the gut of catfish throughout ontogeny and these were *Bacteroidetes, Firmicutes, Fusobacteria*, and *Proteobacteria*. Larsen et al. (2014) reported that the gut communities of warm water fish species were dominated by the phylum *Fusobacteria*, specifically the species *Cetobacterium somerae*. Kashinskaya et al. (2015) reported that *Proteobacteria, Bacteroidetes, Firmicutes, Cyanobacteria* and *Actinobacteria* were the most abundant gut microbiome both in the intestine and habitat environments. Li et al. (2015) investigated the gut microbial communities of carps which were dominated by *Firmicutes, Fusobacteria, Bacteroidetes* and *Proteobacteria*. Wu et al. (2012) reported that the grass carp intestine holds a core microbiota composed of *Firmicutes, Proteobacteria,* and *Actinobacteria*. Baldo et al. (2015) reported that *Fusobacteria, Firmicutes,* and *Proteobacteri* were the dominant gut microbiota of Cichlid Fishes. Kessel et al. (2011) also reported that the major phyla were *Fusobacteria, Bacteroidetes, Planctomycetes*, and *Gammaproteobacteria* in *Cyprinus carpio L*. The comparative analyses of our study with the reported results reveal that some of the bacterial community in the intestinal gut is almost similar to those mentioned in the reported research work. However, various types of bacterial microbiome were detected in the gut sample and this variation may be due to fish habit, fish feed, and conditions of the environment. Based on the rank abundance plot (**Fig. VIII.6**), species richness can be viewed as the number of different species on the chart and species evenness is derived from the slope of the line that fits the graph. The steep gradient indicated the low evenness as the high-ranking species have much higher abundance than the low-ranking species. In this study, the refraction plot (**Fig. VIII.7**) is shown based on the constructions for the calculation of species richness for a given number of individual samples. The steep slope

indicated that a large fraction of the species diversity remains to be discovered. The vertical axis demonstrated the diversity of the community, while the horizontal axis showed the number of sequences considered in the diversity calculation.

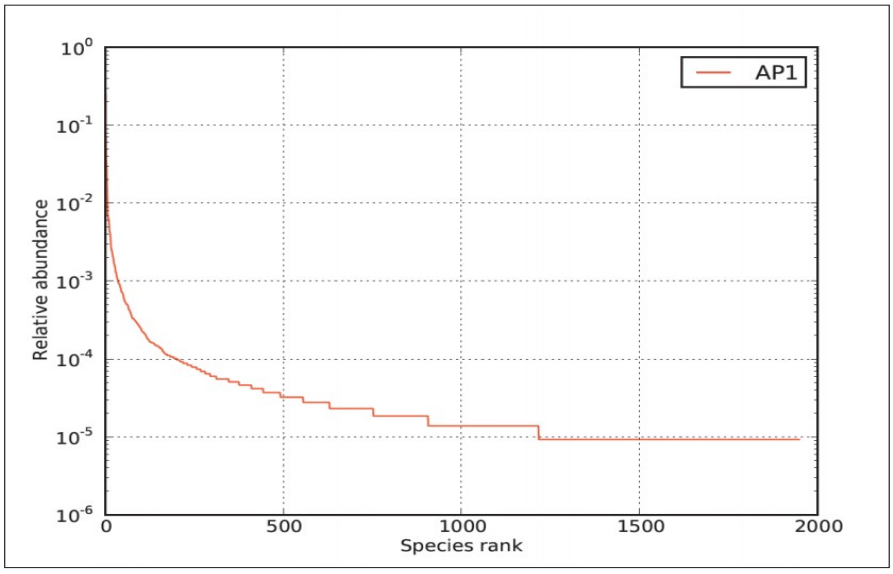


Fig. VIII.6. Rank abundance plot of *Barilius bendelisis* sample.

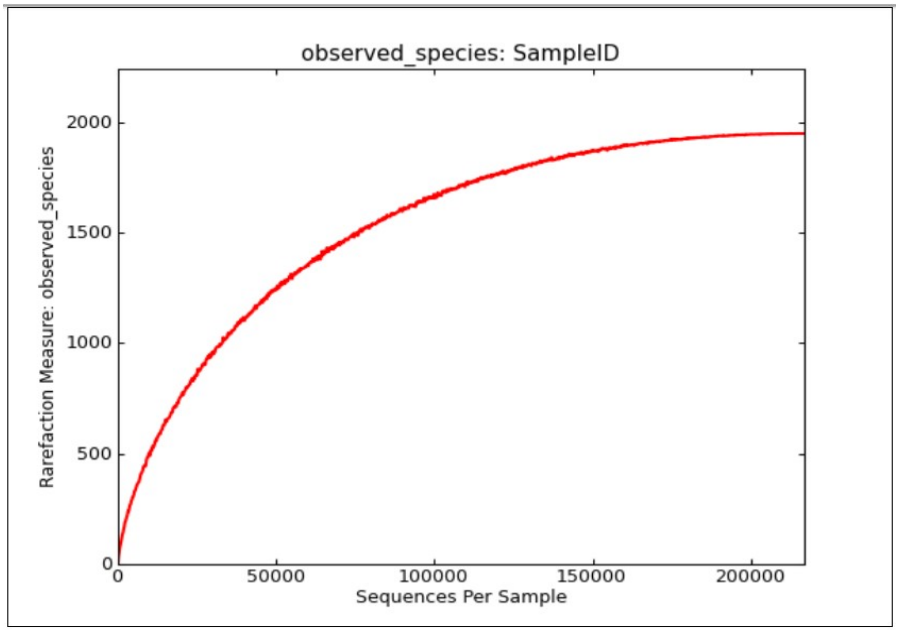


Fig. VIII.7. Rarefaction plot of *Barilius bendelisis* sample.

Table VIII.5. The individual samples summary

Sample	No. of PE reads	Flash/stitch reads	Shannon	Organisms Abundance at phylum level	Organisms Abundance at species level
Barilius bendelisis	249,250	233,473	2.92932425526	*Firmicutes*	*Morganii*

In this study, the OTU table in a form of a heat map representing based on the contribution of percentage is shown in **Fig. VIII.A.4**. This indicated that the high percentage found as k–*Bacteria*; p–*Firmicutes*; c–*Clostridia*; o–*Clostridiales*; f–*Veillonellaceae*; g–; s– was 144654 with denovo3853 and the low percentage found as k–*Bacteria*; p–*Proteobacteria*; c–*Gammaproteobacteria*; o–*Enterobacteriales*; f–*Enterobacteriaceae*; g–*Citrobacter*; s– was 15417 with denovo7933. The Krona graph representing the abundance and hierarchy simultaneously using the summary of taxonomy given by the QIIME is shown in **Fig. VIII.A.5**. This study can be summarized that the highest number of organism's abundance at phylum level was found as *Firmicutes* and the lowest number of organism's abundance at species level was found as *morganii* and Shannon was observed as 2.92932425526 (**Table VIII.5**).

Conclusion

The present study is the report showing the taxonomic distribution of gut microbiome of fish species *Barilius bendelisis* from the *Hel* river, Serfanguri of Kokrajhar, Assam for the first time. The high abundant level of the taxonomic distribution of the phylum was *Firmicutes* (74.8 %), the class was *Clostridia* (74.8 %), the order was *Clostridiales* (74.8 %), the family was *Veillonellaceae* (74.6 %), the genus was *Citrobacter* (7.7 %) and the species was *morganii* (2.4 %). The result of the present study reveals that some of the bacterial community in the intestinal gut is similar to some of the reported research work. However, various types of bacterial microbiome were also detected in the gut sample and this variation may be due to fish habit, fish feed, and environmental conditions. This study gives an idea about the microbial diversity of the selected river, which is a freshwater river source of Kokrajhar that originated from the Indo-Bhutan river.