ABSTRACT

**Background:** The microbial ecosystem in the human intestine is complex and it plays a great role in health and nutrition. Cultural techniques have been used over the years to study the gut microbiota but studies suggest that a greater percentage of these bacteria found in the gut cannot be cultivated using the conventional methods of bacteria isolation.

**Aim:** To increase understanding in this area, we characterized the bacterial diversity (both cultivated and non-cultivated bacteria) in the gut of diarrhoeic individuals using 16S rRNA gene (rDNA) sequences.

**Methodology:** PCR amplification, sequencing and phylogenetic analysis of the 16S ribosomal DNA (rDNA) sequences were done on 10 diarrhoeic stool samples.

**Results:** After quality filtering and chimeric sequence removal, 72313 sequences from all 10 diarrhoeic stool samples subjected to clustering generated 2767 Operational Taxonomic Units (OTUs) of which 2073 were new and unassigned. Representative sequences of the bacteria OTUs cluster were used to construct a bacteria...
### 1. INTRODUCTION

Diarrhoea occurs when the intestine absorbs or secretes fluid more fluid than normal [1]. Most diarrhoea is self-limiting mild infections that can be resolved on its own but some can be acute, severe and life-threatening. Though some diarrhoea is caused by chemical irritations, metabolism and organic disturbances, a vast majority is by infectious pathogens like virus, fungi, parasite or bacteria [2] with bacteria diarrhoea being more common in developing countries. The commonly associated enteric pathogens include bacteria composed of *Escherichia coli* species, *Salmonella* species, *Shigella* species, *Campylobacter jejuni* species, *Vibrio* species, *Yersinia* species, *Aeromonas* species, *Clostridium difficile* species., parasites like *Cyclospora*, *Gardia lambia*, *Entamoeba histolitica*, *Cryptosporidium* spp., viruses like *Rotavirus*, *Calicivirus* and other enteric viruses with *Rotavirus* as leading cause in young children [3,4]. All over the world, there are about 1.7 billion diarrhoea cases every year and the major reason for malnutrition in children under 5 years [5].

The most used method for assessing microbial diversity is the culture-based method. It has been based on selective and differential plating of samples on culture media and identifying the pure culture of the bacteria to the species level. A major challenge in studying the gut microbiota is the inability of culturing most of the gut microorganisms [6]. Even with these intensive investigations, however, there is much concern that culture-based methods do not provide a complete picture of the diversity of the predominant organisms of the gut flora.

Molecular genetic tools have indicated that 60 to 80% of the organisms in the total human microflora have not been cultivated [7]. The use of 16S rRNA gene sequences has been by far the most common housekeeping genetic marker used to study bacterial phylogeny and taxonomy [8]. The phylogenetic analysis of bacterial 16S rRNA genes (rDNAs), amplified directly from complex communities, have provided an efficient strategy for exploring the biodiversity of a particular biota. To derive detailed phylogenetic biodiversity of the bacterial community in the gut of diarrhoeic patients, we analysed bacterial 16S rRNAs extracted from 10 diarrhoeic faecal samples. This method has facilitated access to both cultivated and non-cultivated microorganisms. Sequences generated were clustered into OTUs which were used to construct a phylogenetic tree to reveal the wide variety of bacteria and their contribution in the individual samples.

### 2. MATERIALS AND METHODS

#### 2.1 Study Area

Rivers State was selected as the study area because of its varied socio-economic conditions. Based on the census done in 2006, Rivers State is the sixth most heavily populated states in the country with a population of 5,185,400. It is located on longitude 6°50'E and latitude 4°45'N and covers an area of 468 km². 11077 km².

The study was carried out in three different locations in Nigeria. Diagnostic Laboratory in Rivers State University of Science and Technology Nkpolu Port Harcourt Molecular Microbiology Laboratory, Medical Laboratory
This research work was carried out among 10 individuals who presented with signs and symptoms of diarrhoea in the two major tertiary health Facilities University of Port Harcourt Teaching Hospital (UPTH) and Braithwaite Memorial Specialist Hospital (BMSH) Port Harcourt in Rivers State. Cases of diarrhoea were seen as individuals who passed loose stool three or more times within 24 hours. These hospitals were chosen because they are amongst the busiest health facilities in the state that are accessible to people of different ethnic groups that live in and around the state. The randomly selected individuals are supposed to represent a subgroup of individuals in the state.

2.2 The Conventional Method of Cultivation of Microorganisms

An aliquot of each of the 10 faecal samples were inoculated into different media plates (Mac Conkey agar, Salmonella-Shigella agar, Deoxycholate Citrate agar, Nutrient agar and Thioglycolate Citrate Bile Salt agar) incubated at 37°C for 24 hours for the growth of pure single colonies which were then collated for identification using standard microbiological methods.

2.3 DNA Extraction

Total DNA was extracted and purified using a Zymo Research (ZR) fungal/bacterial DNA MiniPrep TM 50 extraction kit model D6005 (Zymo Research, California, USA) supplied by Inqaba Biotec Africa's Genomics Company, South Africa. 16S rRNA gene of 5-10 pure cultures for each of the 10 faecal samples was mixed with 750 µl of lysis solution and 200 µl of isonic buffer in a ZR Bashing Bead lyses tubes. A bead beater built in a 2 ml tube holder was used to hold the tubes and spun at maximum speed for 5 minutes. The ZR bashing bead lyses tubes were spun at 10,000g for 1 minute. Four hundred (400) µl of the liquid lying above the sediments after centrifugation was put in a collection tube containing the Zymo-Spin IV spin Filter (orange top) and spun at 7000xg for 1 minute. One thousand two hundred (1200) µl of fungal/bacterial DNA binding buffer was put into the collection tube containing the filtrate making the final volume to 1600 µl, 800 µl was then moved to another collection tube containing the Zymo-Spin IIC column and spun for 1 minute at 10,000xg, the flow-through was thrown away from the collection tube. The remaining volume was moved to the same Zymo-spin and spun. Two hundred (200) µl of the DNA Pre-Wash buffer and 500 µl of fungal/bacterial DNA Wash Buffer were added to a new collection tube containing the Zymo-spin IIC and spun for 2 minutes at 10,000xg. The Zymo-spin IIC column was moved to a clean 1.5 µl of fungal/bacterial DNA Wash Buffer centrifuge tube, 100 µl of DNA elution buffer was put into the column matrix and spun for 30 seconds at 10,000xg to elute the DNA. The ultra-pure DNA was then stored at -20 degrees for further analyses. The concentration of DNA and size was estimated by 1.5% agarose gel using the DNA of known molecular weight.

2.4 16S rRNA Amplification and Sequencing

The 16s rRNA region of the rRNA genes of the isolates was amplified using the 27F (AGAGTTTGATCMTGGCTCAG) forward primer and 1492R (CGGGTTACCTTGTTACGACTT) reverse primers in an ABI 9700 Applied BIo systems thermal cycler at a final volume of 50 microlitres for 35 cycles. The PCR mix included: The X2 Dream Taq Master mix supplied by Inqaba, South Africa (Taq polymerase, DNTPs, MgCl), the primers at a concentration of 0.4M and the extracted DNA as the template. The PCR conditions were as follows: Initial denaturation, 95°C for 5 minutes; denaturation, 95°C for 30 seconds; annealing, 52°C for 30 seconds; extension, 72°C for 30 seconds for 35 cycles and final extension, 72°C for 5 minutes after which the machine keeps the amplicons cool at 4°C. PCR products were purified and concentrated with a QIA quick spin PCR purification kit (Qiagen, S.A., Courtaboeuf, France) and amplicons detected on an agarose gel by agarose gel electrophoresis. After successful amplification and detection of 16s rRNA gene fragments, reads of all samples were filtered using Illumina sequencing. These reads were generated when trying to determine the relatedness of organisms by subjecting the sequences to BLAST with already known sequenced genes in the gene bank. Taxonomic classification of all samples was done. The data retrieved from sequencing using Illumina Miseq (in fastq – format) was demultiplexed and quality screened with MOTHUR software (v.1.39.0) [9]. Only the sequences with a minimum length of 250 bp and average quality
score 25 were retained. Sequences were aligned to the Silva reference alignment (release 123) [10]. Pre clustering was performed to remove sequences with possible sequencing errors. Chimeric sequences were identified and removed with UCHIME2 using Silva gold alignment as a reference dataset. The unique sequences were classified using GREENGENES (May 2013 release) reference taxonomy and assigned to operational taxonomic units (OTUs) with phylotype command in MOTHUR. UPGMA-dendrogram was visualized using FIGTREE. Newick phylogenetic tree of data which was done with the R program using Vegan, Phylloseq and BiodiversityR packages [11,12,13]. The neighbour-joining phylogenetic tree was also constructed using MEGA7 [14]. The evolutionary distances were computed using the Jukes-Cantor method with the trees bootstrapped 1000 times.

3. RESULTS

Conventional Cultural method of bacteria isolation identified *Escherichia coli*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, Vibrio species, *Bacillus* species, *Salmonella* species and *Clostridium* species. *Escherichia coli* was found to be highest in prevalence amongst the enteric bacteria followed by *Staphylococcus aureus* then *Pseudomonas aeruginosa* while *Vibrio* species was the organism that was least present in the samples. After successful 16SrRNA amplification and sequencing, 72313 sequences were generated. The sequences from all 10 diarrheic stool samples subjected to clustering generated 2767 OTUs of which 2073 were new and unassigned. The reads generated were used to try to determine the relatedness of organisms by subjecting the sequences to BLAST with already known sequenced genes in the gene bank.

3.1 Similarities between Samples Expressed on a Dendrogram

Analysis of the distances between the different samples and sampling parameters using the Jaccard coefficient was used to generate a dendrogram as shown below. The dendrogram generated is a representation of the similarities between all the samples using the OTUs generated from each sample. The result showed that the dendrogram was in 2 groups. Sample V5 on one group and the other Samples V1, V2, V3, V4, V6, V7, V8, V9 and V10 on the other side of the group. Samples V2 and V3, V6 and V8 and V9 and V10 more closely related to each other than the others in the group because they are on the same clade. Sample V5 formed a different phyletic line because its microbial community is not so related to the other samples. The observed clades formed within the tree was tested for statistical significance using tree Parsimony. No significant differences were observed.

![Fig. 1. Similarities between samples expressed on a dendrogram](image)
Tree parsimony was used to determine if the distances observed within the tree was statistically significant. The Pars significance of 1.0000 indicated that there was no significant difference between the various samples.

3.2 Phylogenetic Biodiversity of the Bacterial Community in the Gut

Initial phylogenetic identification was made using BLAST. The BLAST search program was used to check for close relatives and phylogenetic affiliation. The search results were used as a guide for phylogenetic tree construction using MEGA7. The evolutionary relationship of the 80 most abundant organisms and their contribution from each sample is represented in Fig. 2.

3.3 Evolutionary Relationship of the 50 Most Abundant Organisms per Sample

The evolutionary relationship of the 50 most abundant organisms per sample is represented in Fig. 3a-3).

4. DISCUSSION

The gut is analysed mainly by culture-based methods [15]. In this study, examining 10 faecal samples from diarrhoeic patients by culture-based method yielded the growth of Escherichia coli, Staphylococcus aureus, Klebsiella pneumoniae, Pseudomonas aeruginosa, Vibrio species, Bacillus species, Salmonella species and Clostridium species. Studies have reported cases of Clostridium species, Bacillus species, Salmonella species, Pseudomonas aeruginosa and Vibrio species in human faeces[16,17,18,19]. Other studies have also reported the presence of Escherichia coli, Klebsiella pneumonia and Staphylococcus aureus in human faeces [20,21,22]. The result of this work also agrees with a study in Nigeria in 2010, which revealed Escherichia coli, Klebsiella pneumonia, Staphylococcus aureus and Pseudomonas species as highly prevalent bacteria isolated in diarrhoeic stool samples [23]. Other studies, on the other hand, have shown that 60 to 80% of the microorganisms in the total human gut microbiota have not been cultivated [24] making the whole human gut microbiota still not yet established. Therefore, studying the array of microorganisms in the gut is very important because studies have revealed that many pathogenic microorganisms can go into a new physiological state where their cells remain viable and they fail to grow on the standard laboratory media on which they would normally grow and develop into colonies [25]. The study of evolutionary relatedness among various groups of organisms in a community is known as microbial phylogeny [26]. The use of 16S rRNA gene sequences has been by far the most common housekeeping genetic marker used to study bacterial phylogeny and taxonomy [27]. In this study, 72313 sequences from all 10 diarrheic stool samples subjected to clustering generated 2767 OTUs of which 2073 were new and unassigned. To derive detailed phylogenetic biodiversity of the bacterial community in the gut of diarrhoeic patients, we analysed bacterial 16S rDNAs extracted from the 10 diarrheic faecal samples used for the culture method and it yielded a large percentage of both unculturable and unknown microorganisms available in the gut. A total of 9 Kingdoms, 22 Phyla, 30 Classes, 50 Orders, 74 Families and 670 Blast output results were detected in 16S rRNA metagenomics sequencing of all sample reads. Sequences generated were clustered into OTU's which were used to construct a phylogenetic tree to reveal the wide variety of bacteria and their contribution in the individual samples. The result suggests that several unknown species inhabit the human intestinal tract of which cannot be cultivated. This is in agreement with studies by Stackebrandt and Rainey, [28] which shows that there are difficult to culture bacteria in the human gut which cannot be identified using culture methods of bacteria identification. Suau et al., [29] used molecular-biological techniques to overcome the limit of cultivation and reported that 284 clones were classified into 82 species or phylotypes. Of them, 20 (24.4%) were known species. Using Phyloseq R, bacteria phylogenetic tree was created with representative sequences of the bacteria Operational Taxonomic Unit (OTU) and evolutionary distance was computed using the Jukes-cantor method. The results revealed that the OTUs formed different clades when compared to their closest relatives in the Gen Bank. The evolutionary relationship of the 50 most abundant organisms per samplerevealed the phylum Firmicutes to be most abundant in samples V1, V2, V3, V4, V7, V9 and V10. The phylum Proteobacteria is most abundant in samples V6 and V8 while sample V5 had Bacteroidetes as the most abundant phylum. Making Firmicutes the major phyla in all the samples followed by Proteobacteria, Bacteroides and Actinobacteria. The evolutionary relationship of the 80 most abundant organisms and their contribution from each sample shows the phylum
Firmicutes to have contributed most in the samples followed by Bacteroidetes and this corroborates the work by Eckburg et al., [30] and Sester et al., [31]. Fewer contributions were.
made by the phylum *Proteobacteria, Actinobacteria, Tenericutes* and *Cyanobacteria*. Comparison of the top phyla showed that the major phyla in all the samples were *Firmicutes* followed by *Proteobacteria, Bacteroides* and *Actinobacteria*. *Firmicutes* phylum has been identified as the major phyla inhabiting the intestine of humans [30]. *Proteobacteria* constituted about 44.39% of the most abundant phyla compared across the samples and this agrees with the study by Sester et al., [31] who examined the global pattern of bacterial communities from various habitats and found out that the average level occupied by *Proteobacteria* in the bacterial population was as high as 40%. The abundance of *Clostridia* (43.85%) and *Bacilli* (89.59%) in the comparison of the class classification confirms the Phylum *Firmicutes* as the most abundant throughout the samples. This corroborates with the studies by Eckburg et al., [30] which identifies *Firmicutes* phylum and *Clostridia* class as the most abundant in the Human Intestine. In the comparison of top-order classification of all samples, *Lactobacillales* is the most dominant order. The *Lactobacillales* also is known as lactic acid bacteria play a great role in maintaining a healthy microflora of human mucosal surfaces and is said to preserve the immune function during human Immunodeficiency virus infections [32]. Top blast output results of all samples yielded a high relative abundance of uncultured bacteria which includes uncultured *gamma*, uncultured *lachnospiraceae*, uncultured *organism*, uncultured *klebsiella*, uncultured *bacilli*, uncultured *streptococcus*, uncultured *marine*, uncultured *acetivibrio*, uncultured *organism*, uncultured *romboutsia*, uncultured *bacteroidetes* and uncultured *ruminococcaceae*. Others bacteria were *Collinsella aerofaciens*, *Enterococcus faecalis*, *Bifidobacterium breve*, *Lactobacillus fermentum*, No hits, *faecilis*, *torques*, *Enterococcus durans*, *Bacteroides vulgatus*, *Escherichia coli*, *Bacteroides species*, *Bacteroides vulgatus*, *Veillonella parvula*, *Clostridium species*, *Faecalibacterium prausnitzii*, *Bacteroides dorei*, *Parabacteroides distans*, *Streptococcus salivarius*, *Enterobacteriaceae bacteria*,

![Fig. 3b. Evolutionary relationship of the 50 most abundant organisms in sample 2](image)
Fig. 3c. Evolutionary relationship of the 50 most abundant organisms in sample 3

Clostridium innocuum, Parabacteroides species, solanacearum, Unidentified oral, Roseburia
Clostridium saccharobutylicum, Ralstonia species, Lactococcus species, Enterococcus
durans, Lactobacilli paracasei, Lactobacilli species, Enterococcus faecalis, Lachnospiracea bacterium, Bacteroides vulgatus, Escherichia albertii, Enterococcus durans, Victivallis vadensis and Enterococcus faecium. Most of these bacteria have not yet been characterized. We believe that an improvement of the culture methods would result in the cultivation and identification of new intestinal microorganisms.

Fig. 3e. Evolutionary relationship of the 50 most abundant organisms in sample 5

Fig. 3f. Evolutionary relationship of the 50 most abundant organisms in sample 6
Fig. 3g. Evolutionary relationship of the 50 most abundant organisms in sample 7

Fig. 3h. Evolutionary relationship of the 50 most abundant organisms in sample 8
Fig. 3i. Evolutionary relationship of the 50 most abundant organisms in sample 9

Fig. 3j. Evolutionary relationship of the 50 most abundant organisms in sample 10
5. CONCLUSION

In conclusion, this research was able to identify culturable and unculturable bacteria in the gut of diarrhoeic people in Rivers state and also show the biodiversity and interrelatedness of these microorganisms using molecular methods. We were able to characterize several diverse microorganisms in the gut by using 16S rRNA libraries and a culture-based method and it has been confirmed that the number of microorganisms identified in the gut of diarrhoeic people in Rivers State using the molecular method of bacteria identification far exceeds that of the conventional cultural method. 16S rRNA metagenomic sequence analysis yielded difficult to culture microorganisms with the high level of unknown bacteria of which majority may be of public health significance to humans. Therefore, 16S rRNA techniques for detection and identification of predominant bacteria create new opportunities for non-cultivation studies of the human intestinal microflora which will also help in the proper diagnosis of infectious diseases.

CONSENT AND ETHICAL APPROVAL

Following ethical approval from the Ethics Committee of the University of Port Harcourt Teaching Hospital and Rivers State Hospitals Management Board Port Harcourt, Rivers State. The participants were informed about the study, a questionnaire shared and written informed consents were obtained from each of the participants before the stool specimen was collected from them.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

12. McMurdie PJ, Holmes S. Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome


