Pharmacognosy Communications, 2023; 13(1):24-33.
https://www.phcogcommn.org

Original Article

Analyzing Bacterial Species from Different Environments Using Direct 16S rRNA Gene Sequencing Methods

Olubunmi Faniyan[1], Victor Akpe[2], Ian E Cock[2,3,*]

[1]Department of Science Laboratory Technology, Ekiti State University, Ado-Ekiti, NIGERIA.
[2]School of Environment and Science, Nathan Campus, Griffith University, Nathan, Brisbane, Queensland, AUSTRALIA.
[3]Centre for Planetary Health and Food Security, Nathan Campus, Griffith University, Nathan, Brisbane, Queensland, AUSTRALIA.

ABSTRACT

Multiple bacterial species including Escherichia coli, Bacillus spp., Streptococcus spp., Proteus spp., *Pseudomonas aeruginosa*, *Serratia* spp., *Staphylococcus aureus*, *Enteroccus faecalis* and *Klebsiella* spp. were collected and isolated from multiple geographical niches of Ado-Ekiti, Ekiti state in Nigeria. 16S ribosomal RNA (rRNA) gene sequencing analysis was applied using two universal primers. The bacterial isolates had similarity scores ≥ 96% when compared with the sequence-based bacterial type isolates in the GenBank of the National Center for Biotechnology Information (NCBI). The similarity score values were subsequently used to identify the bacterial isolates to genus and species levels. Thus, the inter-specific polymorphisms of the universal primer of the16S rRNA gene among different genera were found to be suitable for this study. Comparison of bacterial 16S rRNA gene sequences has emerged as a preferred genetic technique compared to the phenotypic identification of bacteria based on cell morphology, gram's reaction, colony morphology, catalase test, sporulation tests, gas production from glucose, e.t.c, and its validity in identifying bacterial species from different environments is verified in this study.

Keywords: Bacterial species identification, 16S rRNA, Universal primers, Similarity score values, Polymorphisms, Gene sequences.

Correspondence:
Dr. Ian E Cock[1,2]
[1]School of Environment and Science, Nathan Campus, Griffith University, 170 Kessels Rd, Nathan, Brisbane, Queensland 4111, AUSTRALIA.
[2]Centre for Planetary Health and Food Security, Nathan Campus, Griffith University, 170 Kessels Rd, Nathan, Brisbane, Queensland 4111, AUSTRALIA.
Email id: i.cock@griffith.edu.au

Received: 17-09-2022;
Revised: 06-10-2022;
Accepted: 22-11-2022.

INTRODUCTION

The bacteria identification and classification threshold standards used for microbiome analysis have been replaced and updated with the widespread adoption of 16S rRNA sequencing techniques since the 1970s.[1,2] These techniques allow for the robust identification of prokaryotes to the species level much more expediently than via the usage of bacterial genome sequencing methods. In addition, 16S rRNA sequencing methods have revolutionized bacterial taxonomy studies, allowing bacteriologists to classify prokaryotes based on their phylogenetic similarities. Previously, polymerase chain reaction (PCR) analysis of genomic DNA was used to detect and classify pathogens, with multiple individual primers developed to detect species-specific genes.[3] However, real time PCR (RT-PCR) of genomic sequences suffers from problems that include the unexpected absence of expected amplification product(s), non-specificity of amplified products, damage to the amplified products hampering identification, and mixtures of primer-dimer bands in the agarose gel electrophoresis which can hamper the visualization of product bands.[4]

The 16S rRNA gene sequencing method, commonly used for bacterial identification, classification and quantization, involves hybridization of small ribosomal subunit genetic marker sequences universally across all bacteria. These sequences also have hyper-variable regions that have region-specific sites in the variable domains or conserved regions in the sequencing units which allow for differentiation between bacterial species. These incremental differences between 16S rRNA sequences have enhanced our understanding of the bacterial evolution. Thus, 16S rRNA gene sequencing are a versatile tool for bacterial phylogeny, ecology, and taxonomy studies.[4,5] This molecular approach has been used to identify and classify environmental and clinical bacterial isolates.[6] *Cronobacter* spp. are highly diverse and share many phenotypic traits with other *Enterobacteriaceae* members, highlighting the need to use multiple methods to confirm the identity of pathogens of this genus. None of the biochemical, chromogenic or PCR primers are reliable for confirmation of the identity of *Cronobacter* spp. isolates as they either give false positives or false negative results. Therefore, 16S rRNA sequencing is a valuable tool used to confirm the identity of bacterial isolates.[7]

It is now common practice to use 16S rRNA gene molecular testing as a means of rapidly identifying specific pathogens

DOI: 10.5530/pc.2023.1.5

present in clinical isolates and clinical specimens. The 16S rRNA gene sequencing technique also provides a means for rapidly identifying pathogens directly from clinical samples without to the need to employ time-consuming culturing methods.[8-10] Thus, these approaches have been increasingly used for the identification of pathogenic bacteria species, including methicillin-resistant *Staphylococcus aureus* (MRSA) and *Clostridium difficile,* for which rapid identification is paramount for improving patient health outcomes.[11]

The development of 16S rRNA gene molecular testing techniques has resulted in several molecular markers that are useful for identifying specific microbial taxa and their phylogenetic classifications.[6,12-16] The 16SrRNA gene is the most common housekeeping genetic marker because the small ribosomal subunit is present universally amongst bacteria. Whilst it contains some highly conserved regions, it also contains hypervariable regions containing species-specific variability with sequences that have diverged during evolution.[4,5] The use of 16S rRNA sequencing is now widely recognized as the 'gold standard' for bacterial identification. Indeed, 16S rRNA gene sequencing has been used in several studies for universal PCR primers and partial 16S gene amplicons, which have been amplified and then used to confer taxonomic identifications based upon bioinformatic alignments against sequence databases.[17-20] Moreover, the use of PCR techniques has enabled the analysis of small amounts of genetic material to analyze genetic variations in specific genomic regions. PCR is a relatively rapid, highly sensitive and highly specific method to examine bacteria genetic coding sequences.[21,22] In contrast, genomic sequencing technologies provide the most detailed, unbiased information of all nucleic acid-based methods and can accurately differentiate and identify unknown organisms. However, depending on the application, genomic sequencing can be prohibitively cost-intensive and time-consuming. Multiplex sequencing reduces the costs but decreases the sensitivity of the analytes, which may be problematic when pathogens are in low abundances in clinical samples.[22]

Several studies have also demonstrated the application of 16S rRNA gene sequencing in microbiology.[23-25] One study compared phenotypic and 16S rRNA based identifications for a collection of 177 isolates, of which 81 isolates originated from clinical samples. The approach was found to be efficient in most cases, with 88.7% of isolates identified to the genus level and 76.3% identified to the species level. Even for isolates that cannot be identified to the genus level, a phylogenetic position may be assigned. In contrast, phenotypic identification may often be compromised with errors and variable character expressions. The overall performance of 16S rRNA sequence analysis is excellent for identifying bacterial strains and can resolve 90% of identifications, it allows the identification of microbes at the genus or the species level and in addition, it offers the possibility of recognizing as yet undescribed taxa, because ribosomal DNA (rDNA) similarity reflects phylogenetic relationships.[25] Another study evaluated 66 cystic fibrosis sputum isolates for the concordance of 16S rRNA and phenotypic identification of *Pseudomonas* spp. where 16S rRNA sequence data were used to design PCR assays to provide genus or species-level identification.[26] The results indicated that preliminary phenotypic testing had misidentified several isolates and they were able to design a 16S rRNA-based PCR assay that provided rapid, simple, and reliable identification of *P. aeruginosa* and its differentiation from other phylogenetically closely related *Pseudomonas* species.[27] Despite these previous studies, a systematic and broad evaluation study of 16S rRNA genes for identification of clinically relevant organisms is lacking. In addition, the existing studies examined limited panels of organisms and the identifications in those studies was based on sequence alignment and similarity against databases with very limited diversity (i.e., MicroSeq, SmartGene, RIDOM and 16SpathDB, etc.), which often results in poor classification, depth, and ambiguous species level identities.[28-33]

MATERIALS AND METHODS

Study area

Ado-Ekiti, the capital of Ekiti state in Nigeria is located within the north-western part of Benin-Owena River basin development area (Figure 1). The city lies between latitude 7^0 3' and 7^0 44' north of equator and longitude 5^0 11' and 5^0 18' east of the Greenwich meridian. It has several satellite towns surrounding it. As seen in Figure 1, Iworoko is to the north, 16km away from the city. Approximately 20km to the east are Igbemo and Ijan-Ekiti, and to the west are Iyin and Igede. Ado-Ekiti has a relatively low relief, with hills that are isolated and dome-shaped inselbergs. The change in economic and political status of Ado-Ekiti over recent years has resulted in a corresponding increase in population. Indeed, the population of Ado-Ekiti increased from 274,205 in 1995 to more than 1.1million in a 2013 census, of which 82% lived in the urban centre.[34] Ado-Ekiti is largely an urban setting containing associated facilities, including multiple higher education institutions. This rapid urban development has impacted on the quality of different water resources in the city, necessitating the development of rapid and reproducible methods of monitoring potable water sources for the identification and quantification of pathogens.

Sample collection

A total of 20 samples were collected from four different sources: dumpsite, sewage, stream, and soil (agricultural farmland). These samples were randomly collected from different geographical niches of Ado-Ekiti (including Fajuyi, Doctors Quarters Basiri South/Owolabi Junction, G.R.A, Adebayo/Adamolekun, Ado/Agric-Road, NTA Road, and Omolayo areas) for bacteriological analysis. Four samples each from the dumpsite, stream, sewage, and agricultural farmland were collected using clean sterile glass containers. These samples were aseptically collected and stored in

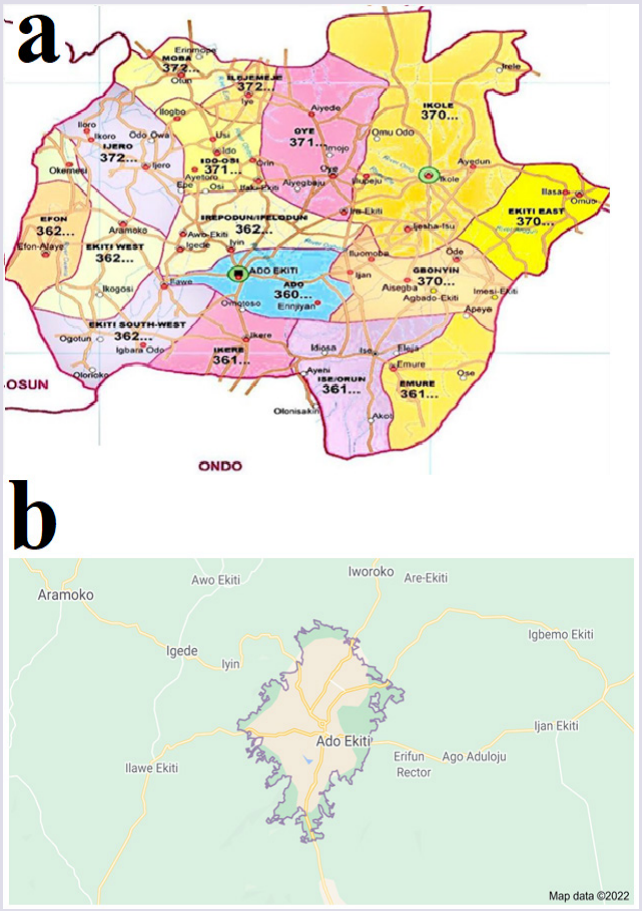

Figure 1: A map of Ekiti State, Nigeria showing (a) local governments and notable areas and (b) the location of the Ado-Ekiti study area. The maps were adapted from.[34]

a sample collection ice pack to ensure that the sample organisms remained viable whilst awaiting study. Samples were transferred to the laboratory and processed within 6 hr of collection.

Total bacterial count

A 1 ml volume of $10^7$and 10^8 serial diluted in 0.9% normal saline solution of the bacterial cultures was inoculated onto nutrient agar (Scantrik Medical Supplies) and incubated for 24 hr at 37°C before counting the number of colonies on each plate. A 2ml volume of the diluted samples was inoculated into each of the 5 tubes and the tubes were incubated at 35°C ± 0.5°C. The reactions were examined at 24 ± 2 hr for gas, (i.e. displacement of medium in fermentation vial or effervescence when tubes are gently agitated). The gas-negative tubes were re-incubated for an additional 24 hr and examined and the reactions were recorded again at 48 ± 3 hr.[35]

Media for isolation and identification experiments

The bacteria that were isolated from different sources and geographical niches of Ado-Ekitiwere cultivated using several selective media, including SSA (*Salmonella-Shigella* agar), EMB (eosin methylene blue medium), and CLED (cysteine-lactose electrolyte deficient agar) which were obtained from KOEMAN group of companies, Lagos.

Isolation and identification of Enterobacteriaceae species

Membrane filtration with selective media was used for bacterial isolation. DNA from soil (CLED) S_1A, dumpsite (SSA) S_4A, sewage (EMB) S_2A, stream (CLED) S_1B and dumpsite (EMB) S_4B (which were collected from Fajuyi, G.R.A, Adebayo/Adamolekun, Ado/Agric-Road and NTA Road respectively) were sequenced for molecular identification and relatedness and filtered using 0.45µm sterile membrane filters (Sterlitech). The membrane filters were subsequently placed onto MacConkey agar plates.[12]

Extraction of total DNA and PCR amplification

Total DNA extracts of bacterial isolates were obtained using a Genomic DNA Purification Kit (Promega, A1120). From the total DNA, the 16S rRNA gene was amplified using the universal primers 27F (5' AGA GTT TGA TCM TGG CTC AG 3') and 1492R (5' CGG TTA CCT TGT TAC GAC TT 3').[13] The PCR mixtures comprised of 1µl of 10X buffer, 0.4µl of 50mM $MgCl_2$, 0.5µl of 2.5mM, 0.5µl 5mM forward and reverse primers, 0.05µl of 5U/µl Taq with 2µl of template DNA and 5.05µl of distilled water to yield total 10µl reaction mix volumes. The PCR method used an initial denaturation temperature of 94°C for 3min, followed by 35 cycles of 94°C for 30 sec, 60°C for 40 sec, 72°C for 60 sec and the final extension temperature of 72°C for 5 min. PCR products were resolved on a 1% agarose gel. The gel was visualized using a UV trans-illuminator (Major Science, Taiwan).[14]

Purification of PCR products

The amplicon was further purified before the sequencing using 2M sodium acetate wash techniques. To 10µl of the PCR product, 1µl of 2M sodium acetate buffer (NaAct) pH 5.2 was added, followed by 20µl absolute ethanol and the mixture was maintained at -20°C for 1hr. It was then centrifuged at 10,000rpm for 10 min, washed with 70% ethanol and air dried. The precipitate was suspended in 5µl sterile distilled water and stored at 4°C whilst awaiting sequencing.

Sequencing analysis

DNA capillary sequencing was performed at Macrogen Inc., using an ABI 3730xl genetic analyzer (Applied Biosystems, USA). The results obtained were checked for quality and trimmed using Codon Code Aligner software (Codon Code Corporation) before subjecting them to BLAST searches using GenBank database. Alignment of the obtained and downloaded sequences was established using the ClustalX2 software.[15] A phylogenetic tree was constructed using MEGA 6 software[16] employing the neighborhood-joining method, based on the Poisson model with Nearest-Neighbor Interchange and a Bootstrap test of phylogeny.

The Bootstrap was set to test 100 replicates to increase the reliability of the tree.

RESULTS

Molecular identification of bacterial isolates using 16S rRNA sequencing

This study revealed amplified sequences of approximately 1500 bp in length of 16S rRNA (Figure 2). The 16S rRNA sequence for each isolate compared with NCBI and showed relationship ranges from 96 to 100 percent to standard isolates of *Escherichia coli, Enterobacter aerogenes, Klebsiella pneumoniae, Klebsiella oxytoca, Raoultella planticola* and *Serratia marcescens* (Figure 3). This study also adopted the Vitek 2 diagnostic biochemical system, which analyzes environmental isolates using universal primers in comparison with the 16S rRNA sequencing results. Notably, these results align closely when compared with the 16S rRNA gene sequencing for the identification of *Streptococcus* spp. of dairy importance.[36] A computerized integrated VITEK 2 system (bioMérieux, Marcy l'Etoile, France) was used to identify the bacterial isolates in the microwells of thin plastic cards using a fluorescent-based method. The system combined with ID-GNB and ID-GPC cards contain at least 43 different tests: 18 tests for sugar assimilation; 18 tests for sugar fermentation; 2 decarboxylase tests (ornithine and lysine); and tests for urease, utilization of malonate and tryptophan deaminase.[37,38] The VITEK2 system was found to identify correctly 15 out of 24 strains (equivalent accuracy of 62.5%) and incorrectly identified 9 strains on the genus level (37.5%) (Figure 4). This revealed low identification potential for such system for streptococci implicated in dairy sector but the amplification of 16S rRNA gene proved to be reliable method for the identification of *Streptococcus* spp. in the diary sector (Figure 5). These results also confirm previous studies which used 16S rRNA sequences to determine the relationship among *Enterobacteriaceae* species in environmental samples, or to identify species.[39]

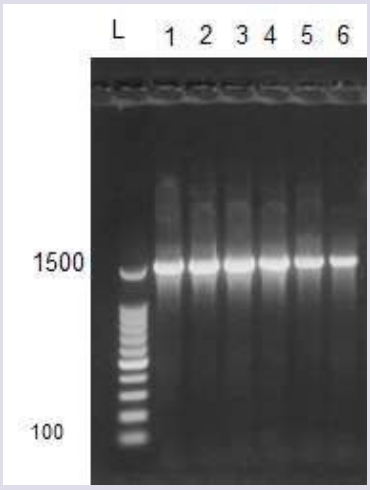

Figure 2: Agarose gel electrophoresis of 16S rRNA gene (1500bp), lane L, ladder (100 - bp DNA ladder); lanes 1-6 positive results of the 16S rRNA amplicons, at 100V for 90 min.

The microorganisms isolated from diverse settings were cultured using different selective media using CLED (cysteine-lactose electrolyte deficient agar), EMB (eosin methylene blue medium) and SSA (*Salmonella-Shigella* agar). The sample sources were gathered from streams, sewage, dumpsites, and soil, from which five samples were selected from the various environmental niches. Bacteria were determined to be different organisms on the basis of colony morphology and colour, as well as micoscopic characteristics (Table 1).

For the purposes of this study, to increase species level sensitivity and specificity of 16S based identification; a model-based methodology was used for species level classification of organisms in different environments. We demonstrated very high confidence genus level identification and good species level identification with microorganisms isolated from different sources, which included *Escherichia coli, Bacillus* spp., *Streptococcus* spp., *Proteus* spp., *Pseudomonas aeruginosa, Shigella* spp., *Staphylococcus aureus, Enterococcus faecalis,* and *Klebsiella* spp. (Table 2). Figure 6 shows the colour mapping representation of the bacterial species isolated from stream water, dumpsite, sewage, farmland soil and the percentage distribution of the bacterial species isolated from the four different sources:

The bacteria count of the samples from stream water ranged from 7.64 $\log_{10}$cfu/ml in dilution 10^{-7} to 7.10 $\log_{10}$cfu/ml in the 10^{-8} dilution. Sewage samples ranged from 8.16 $\log_{10}$cfu/ml in the 10^{-7} dilution to 7.43 $\log_{10}$ cfu/ml in the 10^{-8} dilution. Samples from the dumpsite ranged from 8.13 $\log_{10}$cfu/ml in the 10^{-7} dilution to 7.31 $\log_{10}$cfu/ml in the 10^{-8} dilution, and the farmland soil samples ranged from 7.98 $\log_{10}$ cfu/ml in the 10^{-7} dilution to 7.09 $\log_{10}$ cfu/ml in the 10^{-8} dilution (Table 3). In Figure 3, a neighbor-joining phylogenic tree was developed and shows the relationship between *S. marcescens* strain NBRC and representative members of all other *Enterobacteriaceae* isolates as an example of the analyses. The bars represent 0.005 substitutions per site. Values

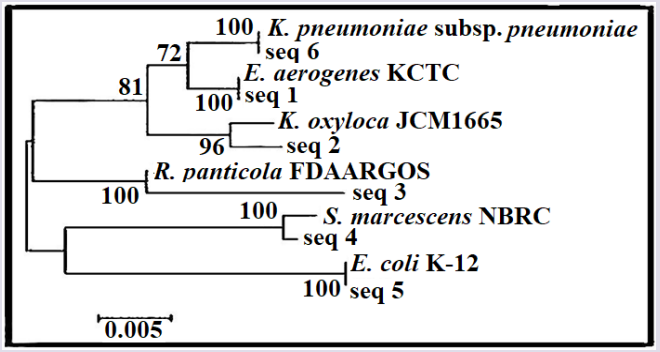

Figure 3: Neighbor - joining phylogenetic tree of deduced 16S rRNA gene sequences compared to *Serratia marcescens*. The diagram was adapted from.[50]

No	Strain	Identification based on VITEK2 system	Identification based on 16S rRNA gene sequencing	Gene accession No.	Correct identification
1	USC 1	*S. agalactiae*	*S. agalactiae*	--------	Yes
2	USC 3	*S. agalactiae*	*S. agalactiae*	KC510215	Yes
3	USC 45	*S. agalactiae*	*S. agalactiae*	--------	Yes
4	USC 47	*S. agalactiae*	*S. agalactiae*	KC510221	Yes
5	LHICA142	*S. agalactiae*	*E. faecium*	--------	No
6	USC 13	*S. dysgalactiae*	*S. dysgalactiae* subsp. *dysgalactiae*	KC510218	Yes
7	USC 17	*S. dysgalactiae*	*S. dysgalactiae* subsp. *dysgalactiae*	KC510219	Yes
8	LHICA 143	*S. dysgalactiae*	*E. faecalis*	--------	No
9	LHICA 144	*S. dysgalactiae*	*E. faecalis*	--------	No
10	LHICA 145	*S. dysgalactiae*	*E. faecium*	--------	No
11	LHICA 146	*S. dysgalactiae*	*E. faecium*	--------	No
12	USC 5	*S. uberis*	*S. uberis*	KC510216	Yes
13	USC 7	*S. uberis*	*S. uberis*	-------	Yes
14	USC 11	*S. uberis*	*S. uberis*	KC510217	Yes
15	USC 55	*S. uberis*	*S. uberis*	KC510223	Yes
16	USC 69	*S. uberis*	*S. uberis*	KC510224	Yes
17	USC 71	*S. uberis*	*S. uberis*	KC510225	Yes
18	USC 80	*S. uberis*	*S. uberis*	KC510226	Yes
19	LHICA 994	*S. uberis*	*E. faecium*	-------	No
20	USC 83	*S. gallolyticus*	*S. gallolyticus* subsp. *gallolyticus*	KC510227	Yes
21	USC 84	*S. gallolyticus*	*S. gallolyticus* subsp. *gallolyticus*	KC510228	Yes
22	USC 52	*S. canis*	*S. canis*	KC510222	No
23	USC 87	*S. canis*	*E. faecalis*	-------	No
24	LHICA 150	*S. canis*	*Lactococcus garvieae*	-------	No
		Total no. of correctly identified strains=**15**		Total no. of incorrectly identified strains=**9**	

Figure 4: Identification of *Streptococcus* spp. based on VITEK 2 system and 16S rRNA gene sequencing. The diagram was adapted from.[36]

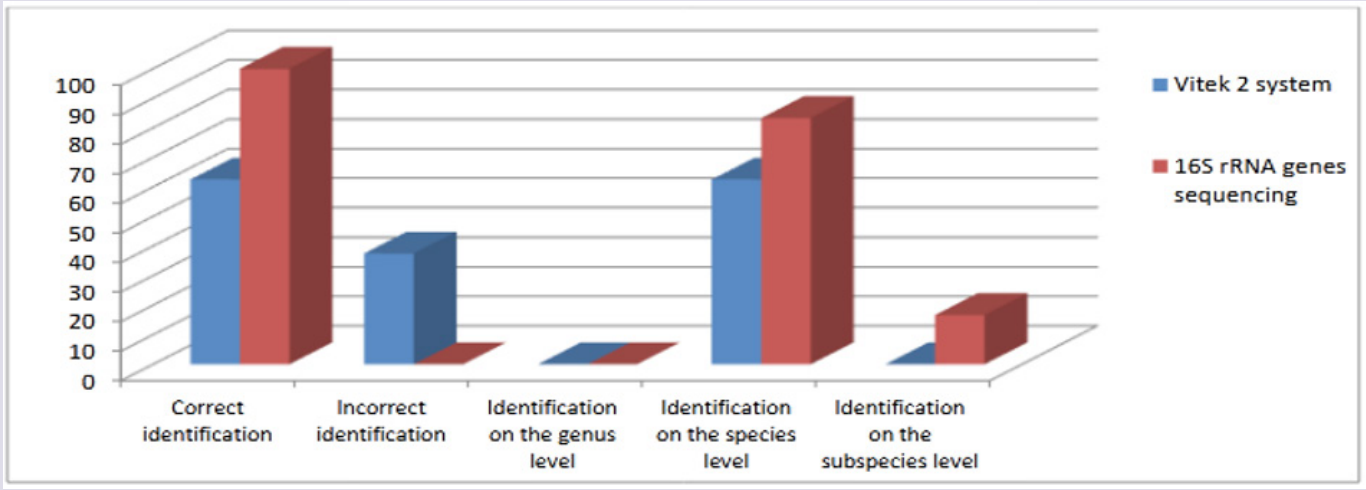

Figure 5: Graphical representation showing the reliability of the 16s rRNA gene sequencing for the identification of *Streptococcus* spp. over Vitek 2.[36]

displayed at the nodes (n = 100) indicate bootstrap values (type strain of *S. marcescens* NBRC 30121T). The numbers next to each node in the tree represent a measure of support for the node and are given as percentages, where 100 represents maximal support. The 'bootstrapping' produced high values, which provides strong evidence of phylogenetic similarity. The sequences to the right of the node cluster together to the exclusion of any other sequences. Bootstrapping is a common method for assessing confidence in phylogenetic analyses. Although bootstrapping was first applied in phylogenetics to assess the repeatability of a given result, bootstrap results are commonly interpreted as a measure of the probability that a phylogenetic estimate represents the true phylogeny.[40]

Circular consensus sequencing of the full 16S rRNA gene

Circular consensus sequencing (CCS),[41] combined with sophisticated denoising algorithms[42] were used to remove PCR and sequencing errors. Thus, it is possible to discriminate between millions of sequences reads that differ by as little as one nucleotide across the entire gene. Together, these technological

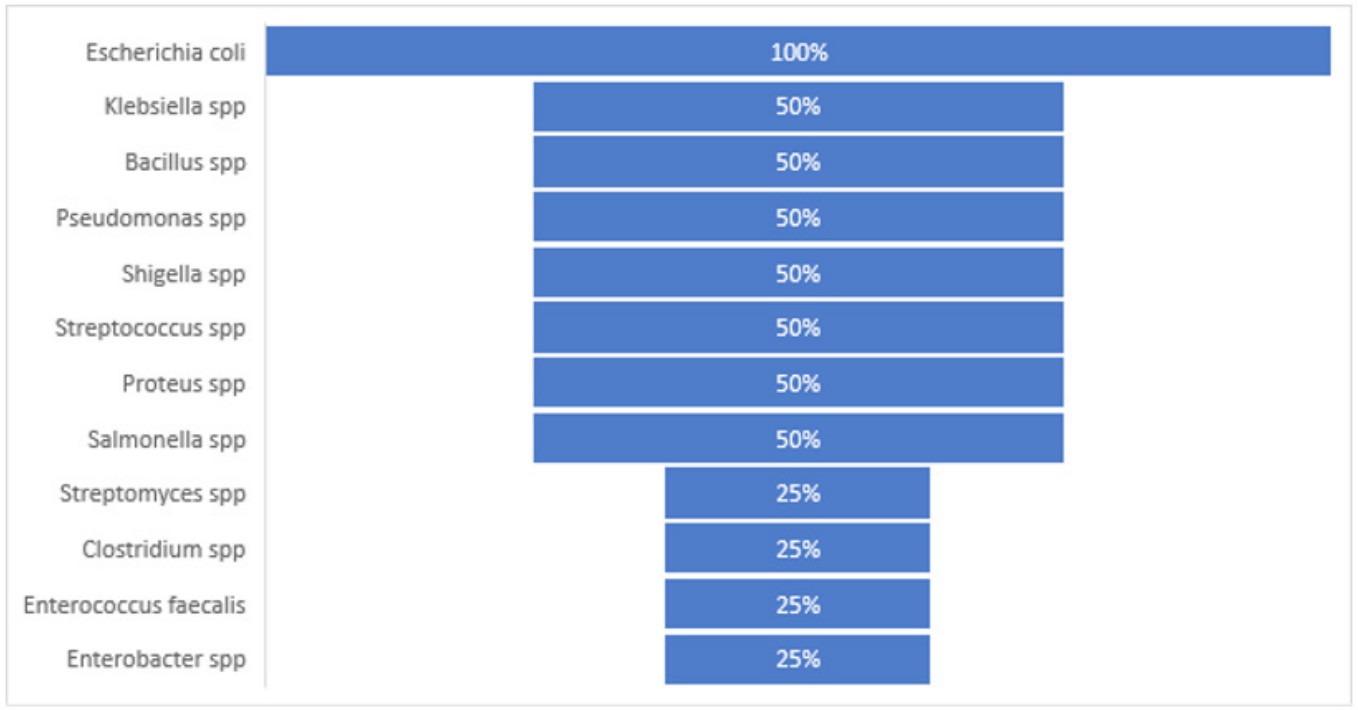

Figure 6: Graphical representation showing the bacterial species based on their presence in the different sources.

Table 1: Isolates recovered from different sources on different media (CLED, EMB and SSA media).

Sources	Samples (n)	Isolates recovered on different media (CFU)			
		(CLED)	(EMB)	(SSA)	Total
Stream	5.0	6.0	7.0	4.0	17.0
Sewage	5.0	4.0	5.0	4.0	13.0
Dumpsite	5.0	3.0	4.0	5.0	12.0
Soil	5.0	4.0	8.0	4.0	16.0

Table 2: Bacterial species isolated from different sources.

Stream water	Dumpsite	Sewage	Farmland Soil
Klebsiella spp.	*Bacillus* spp.	*Klebsiella* spp.	*Bacillus* spp.
Escherichia coli	*Escherichia coli*	*Escherichia coli*	*Escherichia coli*
Salmonella spp.	*Pseudomonas* spp.	*Pseudomonas* spp.	Streptomyces spp.
Shigella spp.	*Streptococcus* spp.	*Shigella* spp.	*Streptococcus* spp.
Clostridium spp.	*Staphylococcus* spp.	*Enterococcus faecalis*	*Staphylococcus* spp.
Enterobacter spp.	*Proteus* spp.	*Salmonella* spp.	*Proteus* spp.

Table 3: Mean total bacteria count (TBC) of environmental samples taken from different sources in Ado-Ekiti.

S/N	Mean sample	Source	10^{-7} CFU/ml	10^{-8} CFU/ml
1	5.0	Stream water	7.64	7.10
2	5.0	Sewage	8.16	7.43
3	5.0	Dumpsite	8.13	7.31
4	5.0	Farmland soil	7.98	7.09

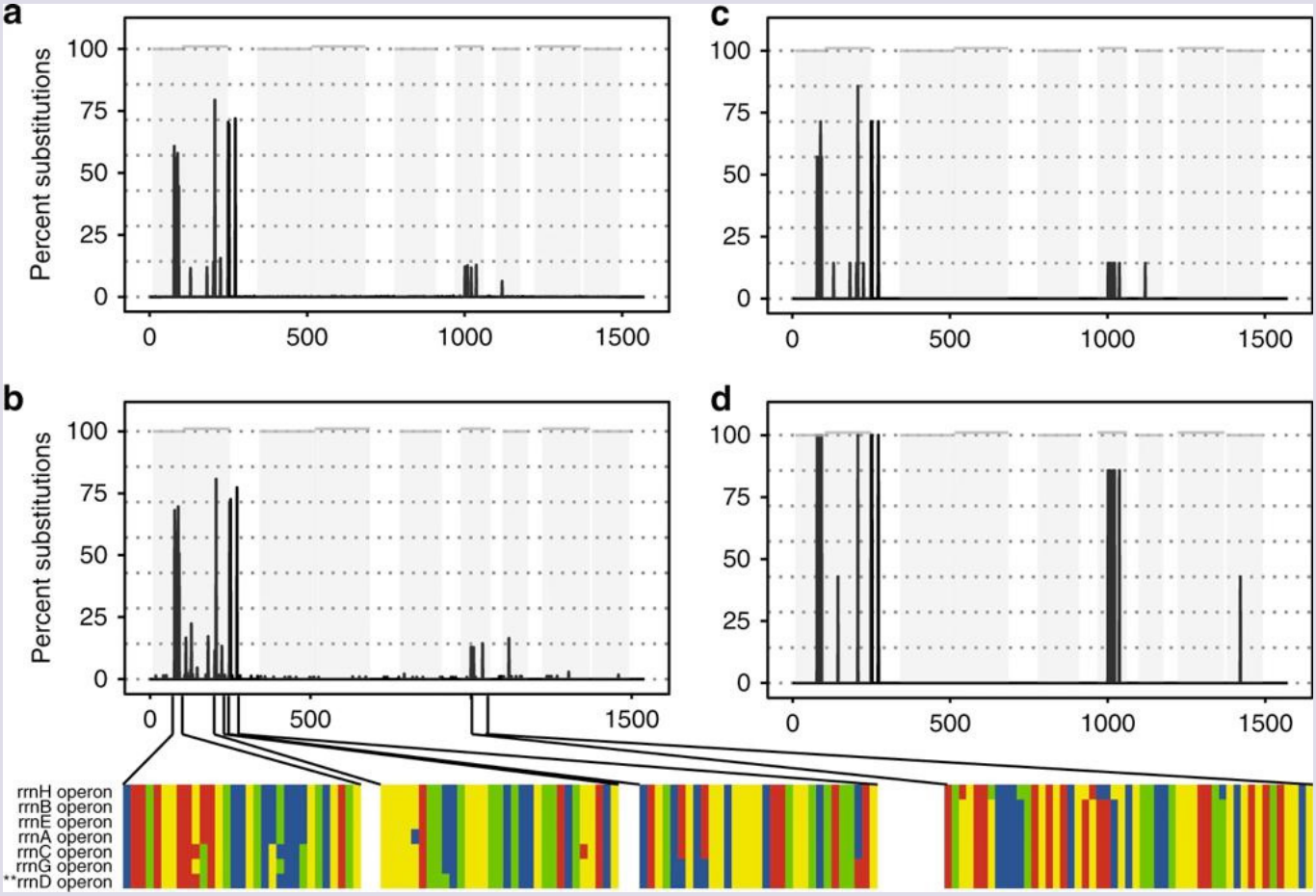

Figure 7: Polymorphisms in *E. coli* 16S rRNA gene sequences including: (a) The position and frequency of substitutions appearing in *E. coli* strain K-12 MG1655 V1–V9 amplicons sequenced on the PacBio RS II platform. (b) The position and frequency of substitutions in reads generated from genomic sequencing of the isolated *E. coli* strain K-12 MG1655 on the Illumina MiSeq platform. Magnified regions show respective positions in the alignment of all seven 16S rRNA genes present in the *E. coli* K-12 MG1655 reference genome. The 16S rRNA sequence from the rrnD operon (**) is used as the reference for all SNP phasing. (c) The predicted nucleotide substitution profile of *E. coli* K-12 MG1655 based on aligning the 16S rRNA gene sequences present in the reference genome. (d) The predicted substitution profile of *E. coli* O157 Sakai based on aligning the 16S rRNA gene sequences present in the reference genome.[48]

and methodological advances mean that it is now possible to exploit the full discriminatory potential of 16S rRNA in a high-throughput manner. Assuming that CCS sequencing can produce 16S rRNA reads with a low frequency of substitution errors, it is likely that a proportion of the substitution errors within accurately aligned reads should reflect variation attributable to 16S rRNA polymorphisms within a species genome.[43] For example, reads aligned to the *E. coli* strain K-12 subs. MG1655 showed a substitution profile which mirrored exactly that predicted by aligning all the 16S rRNA sequences known to be present in a genome[43] (Figures 7a and 7c). The stoichiometry of these nucleotide substitutions was further validated by quantifying variation in comparably aligned Illumina WGS reads (Figure 7b), which demonstrated that a similar substitution profile was reproducible across multiple sequencing runs. Alignments to other reference sequences showed a similar trend of abundant substitutions localized to specific base positions along the 16S rRNA gene, although it was noted that the signal-to-noise ratio increased significantly when the 16S rRNA gene had fewer than 100 aligned reads.

DISCUSSION

Accurate identification of bacteria is important to understand the pathobiology of infectious clinical syndromes and better use of specific antibiotic and infection control strategies for patients and populations. Given that many different species might have similar biochemical profiles, it is doubtful whether culture-based methods can accurately differentiate distinct isolates. Indeed, some species not considered "difficult" to identify based on routine culture-based testing were misidentified upon subsequent sequence-based molecular identification.[44]

In this study, we used high quality sequences generated from 16S rRNA gene amplicons to analyze environmental isolates

using universal primers. Our study revealed amplified sequences of approximately 1500 bp in length of 16S rRNA (Figure 2). Subsequent 16S rRNA analysis using the suggested universal primers may then be applied to select more specific bacterial DNA sequences, which may be more appropriate for molecular analysis of bacterial compositions of environmental and clinical isolates. Additionally, sequence analysis provides an opportunity to compare the local isolates with global isolates in the NCBI library.

The sources of the fecal coliforms including *Escherichia coli, Enterococcus faecalis* and *Klebsiella* spp. includes treatment plant effluent, leaking septic systems, storm water runoff, sewage discharged or dumping from recreational boats, domestic animal and wildlife waste, improper land application of manure or sewage, and runoff from manure storage areas, pastures, rangelands, and feedlots. It is perhaps not surprising that they were recovered from those sources (i.e., stream, sewage, and dumpsite).[45] *Escherichia coli* are the most prevalent member of the feacal coliform group and their presence in the stream, soil, sewage, and dumpsite is an indicator that other disease-causing agents such as viruses, bacteria, and protozoa may also be present.[47] These bacteria produce enzymes including DNase, hyluronidase, staphylokinase, staphylolysin, streptokinsase etc. that help degrade waste materials at dumpsites.[44]

Perhaps not surprisingly, the sewage sample had the highest mean bacterial growth count of 8.16 $\log_{10}$cfu/ml, whilst the lowest counts were for farmland soil, with a mean TBC of 7.98 $\log_{10}$ cfu/ml. For dilution 10^{-8}, sewage also had the highest mean TBC of 7.43 $\log_{10}$cfu/ml, whilst farmland soil had the lowest mean TBC (7.09 $\log_{10}$ cfu/ml). It is likely that the sewage may contain many growth factors that could be utilized by the bacteria which are not available in the farmland soil.

It has been demonstrated that identifying bacteria isolated by sequence rather than phenotype may enhance clinical microbiology by better detecting strains that are poorly documented, seldom isolated, or biochemically abnormal. This can be achieved by sequencing the DNA of the bacteria. The identification of bacteria based on their 16S rRNA gene sequences is more reliable, reproducible, and accurate than the identification achieved via phenotypic testing. The findings of the exam are less open to interpretation. Analysis of the sequence of the 16S rRNA gene may lead to the identification of previously unknown infections. Researchers could use the 16S rRNA gene sequence to find bacteria that have never been grown in a lab. This would free them from having to control the conditions for growth. Additionally, it is essential to accurately identify bacteria as this can assist in disease diagnosis and treatment.

The sequencing of 16S rRNA genes has traditionally only had a limited use in the process of identifying microorganisms. This was primarily due to the high costs involved, the high level of technical expertise required, and the absence of user-friendly comparative sequencing analysis software and validated databases. However, the availability of enhanced DNA sequencing methods, greatly enlarged databases, and kits and software that are more readily accessible makes this technology a viable alternative to regular microbial identification procedures for some categories of species, such as *Mycobacterium* spp. It is possible that the prices will also be similar to those of conventional identification techniques for other slow-growing and species that are difficult to identify. This is especially true if a DNA sequencer is already available in the laboratory. The sequencing of the 16S rRNA gene may also play an important and vital role in classifying microorganisms into correct groups for future research.

The analysis of 16S rRNA gene sequences, even though it is accurate, is currently not widely used outside of major labs and reference facilities due to the complexities and expenses involved. Therefore, one of the future challenges for the larger microbiological, reference, and research laboratories will be to translate information from 16S rRNA gene sequencing into practical biochemical testing schemes. This will make the accuracy of genotypic identification available to smaller, less well-equipped microbiology laboratories.

CONCLUSION

Analysis of 16S rRNA gene sequences permit rapid, robust, reproducible and accurate bacterial identification. Additionally, the test results are less subjective than other methods. 16S rRNA gene sequence analysis can identify non-cultured bacteria, allowing independence from the constraint of growth conditions. Identification of the correct taxonomy or name assignment can make a difference in clinical outcomes. Therefore, there is a demand for the accurate identifications that 16S rRNA gene sequence analysis can provide. The limitations of using molecular identification by 16S rRNA have been described by many researchers. The primary issues include the lack of available high-quality sequences, reliance on phylogenetic clustering methods and/or sequence alignment scores, databases comprising of miss-annotated user-submitted sequences, and/or databases with sparse sequence coverage for some organisms. Obtaining full-length or near full-length 16S rRNA sequences is crucial for making confident genus and species level taxonomic placements. While next-generation sequencing technologies have gained popularity for microbiology applications,[46] the current method of sequencing read lengths falls short of reaching the length of the 16S rRNA gene, limiting their usefulness for high resolution taxonomic placement. However, despite these limitations, 16S rRNA gene sequencing remains the fastest, easiest, and most precise method for the accurate identification of environmental microorganisms.

ABBREVIATIONS

GenBank: The National Institute of Health genetic sequence database; **NCBI:** The National Center for Biotechnology Information; **rRNA:** The ribosomal RNA; **PCR:** Polymerase Chain Reaction; **VITEK 2:** Automated system used to analyse bacterial identification in this project; **MicroSeq:** Rapid microbial Identification System; **RIDOM:** Ribosomal Differentiation of Microorganisms; **16SpathDB:** 16SpathDB is a Web-based 16S rRNA gene sequence database was used analyse the bacterial gram-negative organisms in this project; **CLED:** Cystine–lactose–electrolyte-deficient agar or medium; **SSA:** Salmonella Shigella (SS) Agar; **EMB:** Eosin methylene blue (EMB, also known as "Levine's formulation"); **TBC:** The total bacteria count.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTIONS

Olubunmi Faniyan conceived the project. All authors contributed to the study design. Material preparation, data collection and analysis were performed by Olubunmi Faniyan. The first draft of the manuscript was written by Olubunmi Faniyan and Victor Akpe and Ian Cock edited the draft version. All authors contributed to revision of the manuscript and all authors read and approved the final version of the manuscript.

REFERENCES

1. Ragan MA, Bernard G, Chan CX. Molecular phylogenetics before sequences: Oligonucleotide catalogs as k-mer spectra. RNA Biol. 2014;11(3):176-85. doi: 10.4161/rna.27505, PMID 24572375.
2. Fitzgerald DM, Rosenberg SM. What is mutation? A chapter in the series: How microbes "jeopardize" the modern synthesis. PLOS Genet. 2019;15(4):e1007995. doi: 10.1371/journal.pgen.1007995, PMID 30933985.
3. Skow A, Mangold KA, Tajuddin M, Huntington A, Fritz B, Thomson Jr RB, *et al*. Species-level identification of staphylococcal isolates by real-time PCR and melt curve analysis. J Clin Microbiol. 2005;43(6):2876-80. doi: 10.1128/JCM.43.6.2876-2880.2005, PMID 15956411.
4. Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. J Clin Microbiol. 2007;45(9):2761-4. doi: 10.1128/JCM.01228-07, PMID 17626177.
5. Větrovský T, Baldrian P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. PLOS ONE. 2013;8(2):e57923. doi: 10.1371/journal.pone.0057923, PMID 23460914.
6. Clarridge JE. IIIImpact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev. 2004;17(4):840-62. doi: 10.1128/CMR.17.4.840-862.2004, PMID 15489351.
7. Iversen C, Waddington M, Farmer JJ, Forsythe SJ. The biochemical differentiation of *Enterobacter sakazakii* genotypes. BMC Microbiol. 2006;6(1):94. doi: 10.1186/1471-2180-6-94, PMID 17067387.
8. Barbut F, Monot M, Rousseau A, Cavelot S, Simon T, Burghoffer B, *et al*. Rapid diagnosis of *Clostridium difficile* infection by multiplex real-time PCR. Eur J Clin Microbiol Infect Dis. 2011;30(10):1279-85. doi: 10.1007/s10096-011-1224-z, PMID 21487764.
9. Hallin M, Maes N, Byl B, Jacobs F, De Gheldre Y, Struelens MJ. Clinical impact of a PCR assay for identification of *Staphylococcus aureus* and determination of methicillin resistance directly from blood cultures. J Clin Microbiol. 2003;41(8):3942-4. doi: 10.1128/JCM.41.8.3942-3944.2003, PMID 12904425.
10. Lehmann LE, Hunfeld KP, Emrich T, Haberhausen G, Wissing H, Hoeft A, *et al*. A multiplex real-time PCR assay for rapid detection and differentiation of 25 bacterial and fungal pathogens from whole blood samples. Med Microbiol Immunol. 2008;197(3):313-24. doi: 10.1007/s00430-007-0063-0, PMID 18008085.
11. Brown J, Paladino JA. Impact of rapid methicillin-resistant *Staphylococcus aureus* polymerase chain reaction testing on mortality and cost effectiveness in hospitalized patients with bacteraemia: A decision model. Pharmacoeconomics. 2010;28(7):567-75. doi: 10.2165/11533020-000000000-00000, PMID 20550222.
12. Adékambi T, Drancourt M, Raoult D. The rpoB gene as a tool for clinical microbiologists. Trends Microbiol. 2009;17(1):37-45. doi: 10.1016/j.tim.2008.09.008, PMID 19081723.
13. Bartual SG, Seifert H, Hippler C, Luzon MA, Wisplinghoff H, Rodríguez-Valera F. Development of a multilocus sequence typing scheme for characterization of clinical isolates of *Acinetobacter baumannii*. J Clin Microbiol. 2005;43(9):4382-90. doi: 10.1128/JCM.43.9.4382-4390.2005, PMID 16145081.
14. Enright MC, Knox K, Griffiths D, Crook DW, Spratt BG. Molecular typing of bacteria directly from cerebrospinal fluid. Eur J Clin Microbiol Infect Dis. 2000;19(8):627-30. doi: 10.1007/s100960000321, PMID 11014627.
15. Mathema B, Kurepina NE, Bifani PJ, Kreiswirth BN. Molecular epidemiology of tuberculosis: Current insights. Clin Microbiol Rev. 2006;19(4):658-85. doi: 10.1128/CMR.00061-05, PMID 17041139.
16. Slack AT, Symonds ML, Dohnt MF, Smythe LD. Identification of pathogenic *Leptospira species* by conventional or real-time PCR and sequencing of the DNA gyrase subunit B encoding gene. BMC Microbiol. 2006;6(1):95. doi: 10.1186/1471-2180-6-95, PMID 17067399.
17. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, *et al*. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A. 2011;108(Suppl 1);Suppl 1:4516-22. doi: 10.1073/pnas.1000080107, PMID 20534432.
18. Bosshard PP, Abels S, Zbinden R, Böttger EC, Altwegg M. Ribosomal DNA sequencing for identification of aerobic gram-positive rods in the clinical laboratory (an 18-month evaluation). J Clin Microbiol. 2003;41(9):4134-40. doi: 10.1128/JCM.41.9.4134-4140.2003, PMID 12958237.
19. Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. Short Pyrosequencing reads suffice for accurate microbial community analysis. Nucleic Acids Res. 2007;35(18):e120. doi: 10.1093/nar/gkm541, PMID 17881377.
20. Bartram AK, Lynch MD, Stearns JC, Moreno-Hagelsieb G, Neufeld JD. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. Appl Environ Microbiol. 2011;77(11):3846-52. doi: 10.1128/AEM.02772-10, PMID 21460107.
21. Euzeby JP. List of prokaryotic names with standing in nomen-clature: Genus Legionella; 2013. Fr/l/legionella. Html. Available from: http://www.bacterio.cict [cited May 30 2022]. Available from: https://lpsn.dsmz.de/archive/-typessup.
22. Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, *et al*. Complete genome sequence of DSM 30083(T), the type strain (U5/41(T)) of *Escherichia coli,* and a proposal for delineating subspecies in microbial taxonomy. Stand Genomic Sci. 2014;9:2. doi: 10.1186/1944-3277-9-2. PMID 25780495.
23. Dela Rosa CJ, Rivera W. Identification of Bacteroides spp. From ducks using 16S rRNA Gene Pcr Assay: Prelude to its Application in Microbial Source Tracking. J microb biotech food sci;11(3). doi: 10.15414/jmbfs.4101.
24. Starke R, Pylro VS, Morais DK. 16S rRNA gene copy number normalization does not provide more reliable conclusions in metataxonomic surveys. Microb Ecol. 2021;81(2):535-9. doi: 10.1007/s00248-020-01586-7, PMID 32862246.
25. Drancourt M, Bollet C, Carlioz A, Martelin R, Gayral JP, Raoult D. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. J Clin Microbiol. 2000;38(10):3623-30. doi: 10.1128/JCM.38.10.3623-3630.2000, PMID 11015374.
26. Bosshard PP, Abels S, Altwegg M, Böttger EC, Zbinden R. Comparison of conventional and molecular methods for identification of aerobic catalase-negative gram-positive cocci in the clinical laboratory. J Clin Microbiol. 2004;42(5):2065-73. doi: 10.1128/JCM.42.5.2065-2073.2004, PMID 15131171.
27. Spilker T, Coenye T, Vandamme P, LiPuma JJ. PCR-based assay for differentiation of *Pseudomonas aeruginosa* from other Pseudomonas species recovered from cystic fibrosis patients. J Clin Microbiol. 2004;42(5):2074-9. doi: 10.1128/JCM.42.5.2074-2079.2004, PMID 15131172.
28. Fontana C, Favaro M, Pelliccioni M, Pistoia ES, Favalli C. Use of the MicroSeq 500 16S rRNA gene-based sequencing for identification of bacterial isolates that commercial automated systems failed to identify correctly. J Clin Microbiol. 2005;43(2):615-9. doi: 10.1128/JCM.43.2.615-619.2005, PMID 15695654.
29. Woo PC, Ng KH, Lau SK, Yip KT, Fung AM, Leung KW, *et al*. Usefulness of the MicroSeq 500 16S ribosomal DNA-based bacterial identification system for identification of clinically significant bacterial isolates with ambiguous biochemical profiles. J Clin Microbiol. 2003;41(5):1996-2001. doi: 10.1128/JCM.41.5.1996-2001.2003, PMID 12734240.
30. Simmon KE, Croft AC, Petti CA. Application of SmartGene IDNS software to partial 16S rRNA gene sequences for a diverse group of bacteria in a clinical laboratory. J Clin Microbiol. 2006;44(12):4400-6. doi: 10.1128/JCM.01364-06, PMID 17050811.
31. Harmsen D, Rothgänger J, Frosch M, Albert J. RIDOM: Ribosomal differentiation of medical micro-organisms database. Nucleic Acids Res. 2002;30(1):416-7. doi: 10.1093/nar/30.1.416, PMID 11752353.
32. Woo PC, Teng JL, Yeung JM, Tse H, Lau SK, Yuen KY. Automated identification of medically important bacteria by 16S rRNA gene sequencing using a novel comprehensive database, 16SpathDB. J Clin Microbiol. 2011;49(5):1799-809. doi: 10.1128/JCM.02350-10, PMID 21389154.
33. Teng JL, Ho TC, Yeung RS, Wong AY, Wang H, Chen C, *et al*. Evaluation of 16SpathDB 2.0, an automated 16S rRNA gene sequence database, using 689 complete

bacterial genomes. Diagn Microbiol Infect Dis. 2014;78(2):105-15. doi: 10.1016/j.diagmicrobio.2013.10.019, PMID 24295571.

34. Oriye O. Urban expansion and urban land use in Ado Ekiti, Nigeria. Am J Res Commun. 2013;1(2). doi: 10.5897/jgrp2015.0482.
35. Feng P, Weagant SD, Grant MA, Burkhardt W. Shellfish M, water B. BAM. Enumeration of *Escherichia coli and the Coliform Bacteria.* Bacteriological analytical manual. 2002;13(9):1-3. doi: 10.3403/bseniso9308.
36. Alnakip MEA, Rhouma NR, Abd-Elfatah EN, Quintela-Baluja M, Böhme K, Fernández-No I, *et al*. Discrimination of major and minor streptococci incriminated in bovine mastitis by MALDI-TOF MS fingerprinting and 16S rRNA gene sequencing. Res Vet Sci. 2020;132:426-38. doi: 10.1016/j.rvsc.2020.07.027. PMID 32777539.
37. Haanperä MJ, Jalava J, Huovinen P, Meurman O, Rantakokko-Jalava K. Identification of alpha-hemolytic streptococci by Pyrosequencing the 16S rRNA gene and by use of VITEK 2. J Clin Microbiol. 2007;45(3):762-70. doi: 10.1128/JCM.01342-06, PMID 17215341.
38. Nyberg SD, Meurman O, Jalava J, Rantakokko-Jalava K. Evaluation of detection of extended-spectrum beta-lactamases among *Escherichia coli* and *Klebsiella* spp. isolates by VITEK 2 AST-N029 compared to the agar dilution and disk diffusion methods. Scand J Infect Dis. 2008;40(5):355-62. doi: 10.1080/00365540701704706, PMID 17934979.
39. Hoyer MV, Donze JL, Schulz EJ, Willis DJ, Canfield Jr DE. Total coliform and *Escherichia coli* counts in 99 Florida lakes with relations to some common limnological factors. Lake ReservManag. 2006;22(2):141-50. doi: 10.1080/07438140609353891.
40. Fan H, Ives AR, Surget-Groba Y, Cannon CH. An assembly and alignment-free method of phylogeny reconstruction from next-generation sequencing data. BMC Genom. 2015;16(1):522. doi: 10.1186/s12864-015-1647-5, PMID 26169061.
41. Garg N, Bhattacherjee AK, Shukla PK, Singh B. Influence of Imidacloprid on bacterial community diversity of mango orchard soil assessed through 16S rRNA sequencing-based metagenomic analysis. Environ Monit Assess. 2021;193(2):102. doi: 10.1007/s10661-021-08885-7, PMID 33515343.
42. Jiao X, Zheng X, Ma L, Kutty G, Gogineni E, Sun Q, *et al*. A Benchmark Study on Error Assessment and Quality Control of CCS Reads Derived from the PacBio RS. J Data Mining Genomics Proteomics. 2013;4(3). doi: 10.4172/2153-0602.1000136. PMID 24179701.
43. Acinas SG, Marcelino LA, Klepac-Ceraj V, Polz MF. Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. J Bacteriol. 2004;186(9):2629-35. doi: 10.1128/JB.186.9.2629-2635.2004, PMID 15090503.
44. Duru C, Anyadoh S. Seed rot fungal pathogen of post harvest *Irvingia gabonensis* in the Niger Delta region of Nigeria. International Journal of Tropical Agriculture and Food Systems. 2010;3(2). doi: 10.4314/ijotafs.v3i2.50043.
45. Silva M, McLellan S. Environmental and Social Impact of Stormwater Outfalls at Lake Michigan Beaches. International Journal of Social Ecology and Sustainable Development. 2010;1(3):34-48. doi: 10.4018/jsesd.2010070104.
46. Fournier PE, Raoult D. Prospects for the future using genomics and proteomics in clinical microbiology. Annu Rev Microbiol. 2011;65:169-88. doi: 10.1146/annurev-micro-090110-102922, PMID 21639792.
47. Freddolino PL, Amini S, Tavazoie S. Newly identified genetic variations in common *Escherichia coli* MG1655 stock cultures. J Bacteriol. 2012;194(2):303-6. doi: 10.1128/JB.06087-11, PMID 22081388.
48. Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, *et al*. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat Commun. 2019;10(1):5029. doi: 10.1038/s41467-019-13036-1. PMID 31695033.
49. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30(12):2725-9. doi: 10.1093/molbev/mst197, PMID 24132122.

Cite this article: Faniyan O, Akpe V, Cock IE. Analyzing Bacterial Species from Different Environments using Direct 16S rRNA Gene Sequencing Methods. Pharmacognosy Communications. 2023;13(1):24-33.