academicJournals

Vol. 7(49), pp. 5533-5540, 11 December, 2013
DOI: 10.5897/AJMR2013.5966
ISSN 1996-0808 ©2013 Academic Journals
http://www.academicjournals.org/AJMR

African Journal of Microbiology Research

Review

Identifying bacteria and studying bacterial diversity using the 16S ribosomal RNA gene-based sequencing techniques: A review

Khayalethu Ntushelo

Department of Agriculture and Animal Health, University of South Africa, Private Bag X6, Florida, 1710, South Africa.

Accepted 6 November, 2013

This study reflects the usefulness of sequence analysis of the 16S ribosomal RNA gene in identifying bacteria and determining bacterial diversity. Various techniques that are based on utilizing the 16S rRNA gene are discussed. Of critical importance is the use of massively parallel sequencing to study bacterial diversity. Through massively parallel sequencing which is replacing traditional methods of bacterial identification; various bacterial habitats are surveyed to compile their species compositions.

Key words: Bacterial identification, 16S ribosomal RNA gene, DNA.

INTRODUCTION

Human imagination, curiosity and science have all defined the existence of knowledge. Human beings delve into the expanse of our universe hoping for break-throughs in the most complex subjects of life. Life baffles our minds as we continue in the un-ending search for answers. The origin of life is one question that never evades in its perplexity and intricacy. Evolution is the other, so is species diversity. Charles Darwin formulated a theory that life arose in a “warm little pond” (Darwin, 1863). Similarly, the Oparin Ocean Scenario states that life arose spontaneously from an “oceanic soup” of chemical elements (Oparin, 1924). These two accounts and any scientific theory are not sufficient to explain creation and therefore scientists often end up in a maze of false imagination. Regardless of the spark to life and specifically microbial life, human beings continue to seek answers to questions of creation. To seek consolation, biologists study evolution to trace the present microbial diversity to the most primitive form of life and we arrive at single-celled prokaryotes and single-celled eukaryotes as the simplest and precedent functional biological units.

Based on the fact that prokaryotes are simpler and more primitive it is believed that they were first to arise, followed closely by eukaryotes, and subsequently by creatures in the taxa that fall conveniently under the well accepted prokaryote-eukaryote dichotomy. Darwin also made various observations from plants to bees and theorized that plants and animals were adapted in the environment in which they live. Darwin believed that these passed their diversity to their offspring which would then evolve over time. Due to its richness in morphological traits which traditionally formed the basis for classification of organisms, macroscopic life was obviously easier to study than microbial life. In addition to the tedious processes of cultivating microorganisms, microbial morphologies were too simply to provide sufficient information for classification. This difficulty resulted in lack of enthusiasm by students to study microbiology also causing doubt about whether it was worth exploring these unknown “riches” of the microbial world.

However, some early microbiologists steadfastly

E-mail: ntushk@unisa.ac.za.

studied microorganisms just as botanists and zoologists studied animal and plants, respectively. Microorganisms are probably the most diverse group of organisms but their diversity remains relatively poorly understood. Advent of tools like microscopes helped increased our knowledge of microbial diversity and microbes could then be classified as, among other groups, fungi, bacteria and viruses which are themselves very diverse. Among earlier milestones was the classification of the bacterial genus *Pseudomonas* and the compilation of the Bergey's Manual of Systematic Bacteriology which remained the blueprint of bacterial classification for decades after its adoption. While chemists had the periodic table of elements, botanists and zoologists, the Linnaean taxonomy, microbiology had up until the Bergey's manual been devoid of a reference system for bacterial classification.

Although the Bergey's manual seemed to organize microbiology, there still remained a desire to represent bacterial relationships on a single illustration. This would be the phylogenetic tree of life which was based on sequence analysis of ribosomal RNA genes (Woese et al., 1990). The 16S ribosomal RNA gene became a single criterion on which order would be created to reflect bacterial diversity. Further study of diversity within microbial groups was also aided by other molecular-based techniques which traversed beyond the limitations of studying morphology to classify microorganisms. However, the 16S rRNA gene remained the primary reference for bacterial classification. Various laboratories in the world have generated and deposited 16S rRNA gene sequences in web-based databases. So rapid was this process that DNA sequence databases have over a short period been flooded with 16S rRNA gene sequences to afford researchers the opportunity to search these databases to classify bacteria. Sequence databases currently contain over a million full-length 16S rRNA gene sequences representing a broad phylogenetic spectra that are a useful benchmark for identifying bacterial taxa from diverse samples (Cole et al., 2009).

The purpose of this review was to rearrange all information linked to bacterial classification using sequence analysis of the 16S rRNA gene. There is not a recent collation of the information herein presented to reflect an evolutionary process of classifying bacteria. The author reflected on the use of the 16S rRNA gene as a tool for studying bacterial diversity. The look at the 16S rRNA gene cuts across various techniques that are used to study bacterial diversity.

THE 16S rRNA GENE

Cytochrome c and ribulose bisphosphate carboxylase are just two examples that have been used for phylogeny for organisms in general but fell away as favourites because not all organisms have these macromolecules and therefore comparison of distant taxa was not possible. Ribosomal RNA genes are a convenient measure for taxonomy (Woese et al., 1990). For the purposes of classification up to species level, although not sufficient, the 16S rRNA gene is the primary reference for bacteria (Fox et al., 1992). For strain differentiation within a species other measures beyond the scope of this review may be necessary. Among the three types of RNA that are contained by bacterial ribosomes, namely, 16S rRNA, 5S rRNA and 23S rRNA, the 16S rRNA is regarded as the most convenient and reliable measure.

Molecular approaches for detecting and classifying bacteria rely on the PCR amplification and sequence analysis of the 16S rRNA gene. The 16S rRNA gene is a suitable parameter for bacterial classification because the 16S rRNA gene is universal among bacteria and is conserved but has sufficient variation to distinguish between taxa (Gutell et al., 1985; Noller, 1984). One PCR primer pair can target the 16S rRNA gene from a wide range of bacterial species. The supreme advantage of the 16S rRNA gene-based analysis is that it may bypass culturing of bacteria as PCR detection is done on DNA extracted from crude samples. Due to this direct amplification of the 16S rRNA gene from DNA samples, it is possible to detect unculturable bacteria which are estimated to exceed 99% of microorganisms observable in nature (Amann et al., 1995).

Due to these culture-independent surveys, the number of identifiable bacteria has increased drastically in recent times. In these studies, the 16S rRNA gene is PCR-amplified from a DNA sample, the PCR fragment is sequenced, the sequence is queried on a database like the NCBI, sequence hits are pooled from the database, the sequences are used for phylogenetic analysis. The gene databases have facilitated these studies and as a result bacterial diversity surveyors are able to share information eliminating the need for repeated surveys. Various regions of the 16S rRNA gene can be explored to study variation among bacterial phylogenies. Targeting specific regions of the 16S rRNA gene approximately 100 bp instead of longer reads can provide sufficient information for classification (Monstein et al., 2001). Various PCR primers are available to target these regions (Sundquist et al., 2007; Claesson et al., 2010). When analyzing meta-gonomic samples, only a portion (about 500 nucleotides) of the 16S rRNA gene is sufficient for phylogenetic allocation of unknown bacteria into respective taxa (Lane et al., 1985; Claesson et al., 2009). However, the shorter the reads the more it becomes difficult to classify bacteria beyond level above the species level (Claesson et al., 2010). However, use of short reads makes massive parallel sequencing convenient for gross analysis of metagenomic samples because of the short read lengths produced by the high-

throughput sequencing instruments. However, a longer sequence that covers almost the entire 16S rRNA may be preferable.

16S rRNA GENE PCR UNIVERSAL PRIMERS

Because the 16S rRNA gene is common among all known bacteria, it serves as a primary reference for classification. In addition, the 16S rRNA gene is conservative and therefore allows design of universal primers (Greisen et al., 1994; Radstrom et al., 1994). Used in a PCR, a single pair of the 16S rRNA gene universal primers is capable of amplifying the 16S rRNA gene from diverse bacterial taxa (Bottger, 1989). Bottger (1989) demonstrated that a portion of the 16S rRNA gene could be amplified from *Legionella pneumophila*, *Escherichia coli* or *Mycobacterium tubertulosis* using only one pair of primers. This discovery provided a convenient and rapid method of studying bacteria as one set of primers could be used across bacterial taxa. This became more useful in metagenomic studies of assessing bacteria in a sample. The most suitable set of universal primers is one that is designed to flank the entire 16S rRNA gene or the variable regions of the 16S rRNA gene which are on their own sufficient for classifying bacterial taxa. Selecting universal primers to amplify the 16S rRNA gene from a broad range of bacteria remains a daunting task. It appears impossible to design a primer pair that can amplify the 16S rRNA gene from all known bacteria. However, Klindworth et al. (2013) made a selection of best available primers for *Bacteria*, *Archea* and *Eukaryota* and found primers whose amplification of various regions of the 16S rRNA gene to span all three domains.

16S rRNA GENE DATABASES

Sequence information of the 16S rRNA gene sequence is being kept to allow bacterial investigators the opportunity to undertake comparative studies for classifying bacteria. The Ribosomal Database Project (RDP) was established to create a credible bank for 16S rRNA gene information (Cole et al., 2011). The RDP is both a collection of sequence information of the 16S rRNA gene as well as a software suit with tools to organise raw sequences into alignments, annotate sequences and organise information to provide a phylogenetic analysis of the data for understanding bacterial diversity. The RDP has regular updates of its collection as well as a user account system for users to integrate their data with the existing data. Quality of deposited 16S rRNA gene sequences is assured using the Pintail chimera detection program (Ashelford et al., 2005). Analysis tools of the RDP include the Hierarchy Browser which allows users to browse the dataset by taxonomy, publication or sequenced genome. The RDP also has a SeqCart into which a user can add selected sequences so that analysis can be performed only on the sequences in the cart. The RDP uses the naive Bayesian classifier to rapidly and precisely classify bacterial 16S rDNA sequences into phylogenies and it is suitable for both single rDNA reads and libraries of thousands of reads generated by high-throughput genome sequencing platforms (Wang et al., 2007). This classifier is trained on known and continuously updated sequences of the 16S rRNA gene and a few other sequences of value to bacterial diversity. From each query, a base word is selected and probed on the database to determine joint probability of assignment to a genus based on the naive Bayesian assumption using a multiple bootstrap repetition. The RDP Classifier also checks the orientation of the sequences being matched to ensure that reporting of only those sequences that have the same orientation for both full-length and partial 16S rRNA gene fragments. The RDP Library Compare compares differences between taxa from two sample libraries. Sequence Match of the RDP finds sequences in the database with the closest identity to the query sequence. The query sequence can be matched with sequences either in SeqCart, or with sequences on its page or on sequences on the web through the RDP web service interface. The RDP Probe Match allows rapid searching of the database to check primer and probe coverage and specificity of probes up to over 60 bases in length. For near-matching strings Probe Match includes ambiguity codons in the search output. The RDP Tree Builder allows users to draw phylogenetic trees with bootstrap confidence values. Another important tool of the RDP is RDP Taxomatic with which users can create interactive heatmaps based on pairwise distances between 16S rRNA gene sequences. RDP Web Services Interfaces allows programmers to link RDP tools with other web interfaces.

Another useful database is the SILVA which serves as a central repository of aligned rRNA sequences from from *Bacteria*, *Archea* and *Eukarya*. SILVA maintains strict quality control procedures for sequence deposits. Short sequences are rejected and it has a low filtering tolerance for ambiguities. The SILVA online interface suit can be used to query rRNA gene sequences, align sequences for phylogeny-based studies (Pruesse et al., 2007).

The Basic Local Alignment Search Tool (BLAST) of the NCBI (http://blast.ncbi.nlm.nih.gov/) is a widely used bioinformatics program. BLAST has wide applications from nucleotide to protein sequences. A query 16S rRNA gene sequence can be compared with all entries in the database to match it with sequences that resemble it above a defined threshold value and emphasize speed in the search for optimal alignment from a huge dataset.

Hits of 16S rRNA gene sequences come as the BLAST output and they can be arranged according to coverage or similarity. Hits can be retrieved in FASTA formats which can be used to draw a phylogenetic tree.

The traditional 16S-cloning-and-sequencing approach, provides an in-depth analysis of the richness of bacterial species within a sample, however, it remained laborious, costly and would mostly only detect just about a hundred sequences per sample. This warranted alternative assaying techniques, namely terminal restriction fragment length polymorphisms (t-RFLP) (Muyzer et al., 1993; Liu et al., 1997; Fisher and Triplett, 1999) and the denaturing gradient gel electrophoresis (DGGE). These techniques remained relevant and provided useful information for classifying bacteria.

Terminal-restriction fragment length polymorphism

The terminal-restriction fragment length polymorphism technique is used to study bacterial diversity based on the variation of the 16S rRNA gene. The 16S rRNA gene is PCR-amplified with either one or both of the PCR primers fluorescently labelled. The PCR products are digested with a restriction enzyme. The fragments are separated by electrophoresis and the fragment lengths indicate the diversity that is in the analyzed sample.

Temperature gradient gel electrophoresis/denaturing gradient gel electrophoresis

These methods exploit the different denaturing properties of double-stranded 16S rRNA gene copies with sequence differences. Small sequence differences between 16S rRNA gene strands can be detected by these methods. Base differences in the 16S rRNA gene determine the stage of electrophoresis at which the double-stranded 16S rRNA gene melts in a gel with either a temperature or a denaturing chemical gradient. Differences in these denaturing differences can correspond to 16S rRNA gene variation.

METAGENOMIC APPROACHES

Metagenomics began in the 1980s as the idea of gross extraction of DNA from a sample with a mixture of nucleic acid (Olsen et al., 1986; Pace et al., 1986). PCR was used to selectively amplify the target 16S rRNA gene which would then be cloned and sequenced to reveal the identity of the bacterial species from which the 16S rRNA gene came. Sequences representing 16S rRNA gene fragments could be aligned and used to draw a phylogenetic tree. Such a system allowed even relationships between cultured and uncultured bacteria to be determined phylogenetically. Although the idea of metagenomics began in the 1980s, it was not until 1998 that Handelsman referred to this gross genomic sample as a “metagenome” (Handelsman, 1998).

Metagenomic studies inundate data storage bins with massive amounts of data which can be hard to manipulate. This is especially true if the metagenomic analysis is not targeted to a specific gene. For the purposes of bacterial classification, the 16S rRNA gene is a suitable target for analysis. Targeting only this gene helps to scale down data generation in metagenomic studies using high-throughput sequencing platforms. The 16S rRNA gene is PCR-amplified from the metagenome and barcoding allows a mixture with hundreds of samples to be combined in one sequencing run.

High-throughput genomic DNA sequencing from gross DNA samples is able to provide a rapid bacterial survey with a broader range of captured species. Using high-throughput genomic sequencing platforms, metagenomics has become popular in surveys of plant-associated bacteria, animal-associated bacteria, human bacteria and environmental bacteria (Handelsman, 2004). Uses of this approach have gained wide applications in the search for novel bacterial species and novel enzymes (Handelsman, 2004).

Data analysis

Metagenomic data may be analyzed using the phylotyping approach. In phylotyping, sequences are grouped into bins based on their similarity with reference entries in the database. All the 16S rRNA gene sequences are given identities based on previously cultured and classified bacteria. Another approach, the operational taxonomic unit (OTU) approach, does not restrict classification based on already created bins but groups the 16S rRNA gene sequences based on their similarity with each other. The problem with the OTU approach is that cutoffs for assigning a group to the same taxonomic rank e.g. species is not easy to set and therefore placement of 16S rRNA gene sequences remains problematic and the computer algorithms for clustering sequences are slow and demand high amount of computer memory (Schloss and Handelsman, 2005; Schloss et al., 2009; Sun et al., 2009). To evade this problem other workers use Cd-hit, a program which can cluster large datasets with high speed and accuracy (Li and Godzik, 2006). It remains crucial to run 16S rRNA gene datasets on multiple software in order to benefit from the various strengths provided by the different programs.

HIGH-THROUGHPUT GENOMIC SEQUENCING PLATFORMS

Sequencing DNA is an attempt to decipher the permutation of its nucleotides. Early attempts to establish a sequencing system included, the Maxam-Gilbert sequencing and chain termination methods (Maxam and Gilbert, 1980; Sanger et al., 1977). The Maxam-Gilbert method or the chemical sequencing method was developed by Walter-Gilbert and Alan Maxam in 1977. Purified DNA could be used directly in the sequencing reaction. End-labelled DNA is cleaved at specific bases using specific reagents. The labelled fragments cleave using dimethyl sulphate which selectively attacks purines, and hydrazine which selectively attacks pyrimidines. The fragments are electrophoresed in a high resolution polyacrylamide gel to deduce the sequence of the DNA molecule. The chain termination method requires a single strand of the DNA molecule whose sequence must be determined, a DNA primer, a DNA polymerase, labelled nucleotides, nucleotides that serve as terminators because of their lack of 3'-OH group which are required to form phosphodiester bonds between two nucleotides. Based on the sequences of the template DNA, DNA strands of various lengths are synthesized. The newly synthesized strands are denatured by heat and run in a polyacrylamide gel. From the various size fragments on the gel, sequence of the DNA molecule can be deciphered.

The advent of the Sanger sequencing technology (Sanger et al., 1977) afforded scientists the ability to obtain genetic information from any biological sample. Sanger sequencing became widely adopted as knowledge of the genome constitution of various organisms. However, the low throughput of Sanger sequencing could not match the desire and demand for genomic information. This necessitated the advent of the groundbreaking high-throughput technologies by Illumina, Roche (454), Applied Biosystems (SOLiD) and Life Technologies (Ion torrent). With their massively parallel sequencing, these technologies accelerated the survey of habitats to assess the compositions of their bacterial populations by generating multitudes of sequence reads of the 16S rRNA gene. Analysis of the sequence of the 16S rRNA gene of environmental samples has uncovered bacterial phylogenies with little or no representation among bacteria already studied. Some of these novel organisms inhabit fringe ecosystems in which some global chemical cycles are determined. This discovery would have been delayed had it not been the power of massive parallel sequencing that has been adopted by several laboratories worldwide.

In 454 sequencing, adaptors are attached to sequence fragments and the fragments are attached to beads by means of these adaptors. The samples (beads/adaptors/DNA fragments) are loaded into picotiter plates where the beads enter into individual wells. Packing beads are added into the wells to aid the spectrometer in reading the sample.

Illumina relies on chain termination to determine sequences of DNA fragments. Adaptors are attached to DNA fragments, the adaptors hybridise with a flow cell which has a 'forest' of adaptors. The fragments bend over to attach to adaptors and bridge amplification takes place to generate multiple copies of the fragments. This technology uses chemical modification of the nucleotides with fluorescent molecules of different colours to read the bases. Laser detection is incorporated to detect reads.

SOLiD uses emulsion PCR amplification similar to 454. It distributes its fragment library into microbeads that can vary in size and richness of slides they are on. Fluorescence is then emitted when each fragment is ligated into a single strand sequence. Exact call chemistry is employed to capture data. Four different coloured primers are used to map out possible combinations with the sequence.

The Ion torrent sequencing technology does not use any light-based medium to transmit information. It detects pH changes inside the semiconductor chip and those chemical changes are translated into digital bases that appear as output.

Studies to assess plant-associated bacteria

Analysis of bacterial communities associated with the plant root zone benefited tremendously from utilizing sequence analysis of the 16S rRNA gene. No longer are soil microbiologists limited to culturing of soil bacteria and identifying the cultured microorganisms. Their scope of bacteria to be studied also extends to unculturable bacteria which were otherwise not detected by traditional means. By PCR using universal primers, the 16S rRNA gene can be amplified from a variety of bacterial taxa in the targeted sample.

Due to this large-scale generation of sequencing data, soil biologists are able to establish associations between the bacteria, the soil and the rhizosphere rapidly. Such knowledge can speed up efforts to manipulate the ecology of the soil for the benefit of the plant. As means to assess differences between two phytoplasma groups, namely, the lethal yellowing phytoplasma that infects *Cocos nucifera* L. and the Texas Phoenix palm decline group that infects *Phoenix* spp. and *Sabal palmetto*, Ntushelo et al. (2012) used sequence analysis of the 16S rRNA gene. Sequence analysis of the 16S rRNA gene is widely used to classify phytoplasmas. Phytoplasma were generally classified using traditional methods but have now been reclassified using sequence analysis of the 16S rRNA gene (Lee et al., 1998).

Studies of animal-associated bacteria

In animals, sequence analysis of the 16S rRNA gene has been helpful in understanding the make-up of the bacterial communities in the rumen fluid. Studies on the discovery of cellulose-degrading bacteria in the rumen have uncovered novel bacteria that can be used in the fermentation reactions in biofuel plants (Hess et al., 2011). Massively-parallel sequencing of the 16S rRNA gene generated catalogues of cellulose degrading bacteria (Hess et al., 2011). Probing of extracted RNA with oligonucleotides probes would also bypass the need for cultivating bacteria from the rumen. However, the requirement to develop panels of probes proves difficult.

Amplification of the 16S rRNA genes requires the use of universal primers, few of which can amplify this gene from multitudes of species.

Studies of human-associated bacteria

It is estimated that microbes in our bodies are ten-fold the number of human cells. Most of these microbes reside in the gut and influence human physiology and nutrition (Backhed et al., 2005; Hooper et al., 2002). Gut bacteria have been studied extensively using massively-parallel sequencing of the 16S rRNA gene. These studies reveal the dominant bacterial taxa in the human gut and this also creates the necessity to understand the specialized association between humans and the dominant taxa of bacteria in the human gut. The more than 400 bacterial species in the human gut are grouped into a few taxonomic divisions (Eckburg et al., 2005): *Bacteroides, Eubacterium, Clostridium, Ruminococcus* and *Faecalibacterium.* However, each individual appears to have a unique microbial community (Ley et al., 2006; Eckburg et al., 2005; Zoetendal et al., 1998).

Vaginal bacteria have also been surveyed using sequence analysis of the 16S rRNA gene. The vaginal ecosystem is unique and can be more complex in women with bacterial vaginosis (Hillier et al., 1993). Because the composition of the vaginal microflora may determine the state of healthiness of women it is therefore important to survey this microflora in order to understand the underlying causes of vaginal disorders. These studies have been made possible by sequence analysis of the 16S rRNA gene to determine the diversity of the bacteria in the vaginal ecosystem.

Studies of environmental bacteria

Microbial survey of marine ecosystems has also been undertaken. Multiple samples of ocean water were collected on a long transect and millions of sequencing reads of the 16S rRNA gene were analyzed to determine the diversity of the ocean bacteria (Venter et al., 2004). Conclusions drawn from these studies will increase our understanding of the earth as the marine ecosystem is bigger than the terrestrial body. Similar studies have been extended to include remote and extreme environments such as the poles, hypersaline environments such as the Dead Sea and the Great Salt Lake, ancient deposits and various extremophiles have been classified and have been included in the catalogues of bacterial species from various ecosystems. For human hygienic and water management, sewage systems have been surveyed to determine the diversity of bacterial communities found in sewages.

LIMITATION OF THE USE OF SEQUENCE ANALYSIS OF THE 16S rRNA GENE IN IDENTIFYING AND CLASSIFYING BACTERIA

Bacterial species *Deinococcus geothermalis* and *Bacillus subtilis* have no apparent clustering of substitutions in 16S rRNA genes. Some bacterial species were found to have unusual regional diversity among paralogous 16S rRNA genes. Other bacterial species have truncated 16S rRNA genes (Pei et al., 2010). These defects pose limitations on the use of the 16S rRNA gene. Based on the fact that the sequence analysis of 16S rRNA gene requires the use of molecular techniques like sequencing, a single error in the identification process may lead to incorrect identification which may lead to misrepresenttation of the taxa concerned.

The number of rRNA gene regions in prokaryotic chromosomes differs widely (Krawiec and Riley, 1990). The variation in the number of 16S rRNA gene copies in any one cell may bring uncertainty on the number of copies being compared in any comparison. Furthermore, diversities among species which have recently diverged may not be easily recognisable by sequence analysis of their 16S rRNA genes contracting associations established by DNA-DNA hybridisation (Fox et al., 1992).

CONCLUSION

Approaches to identifying and studying bacterial diversity often relied on the traditional methods of plating bacteria on agar. These approaches are still relevant for culturable bacteria but fall short of detecting fastidious and unculturable bacteria. Molecular-based techniques like targeted sequencing of the 16S rRNA gene from gross DNA samples have facilitated surveys of bacterial diversity. The sequencing and cloning of individual sequences is however tedious and cannot provide a comprehensive survey of a bacterial community. The 16S

rRNA gene can be amplified from pure bacterial colonies or can be amplified directly from a crude sample. Amplified from a crude sample, the 16S rRNA gene can be massively sequenced using high-throughput sequen=-cing instruments. Direct amplification of the 16S rRNA gene and its massive sequencing has corrected the underrepresentation of bacteria in many bacterial com-munities. Analysis of bacterial communities is now made easier by the ample data generated from various bacterial communities survey projects.

REFERENCES

Amann RI, Ludwig W, Schleifer KH (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiol. Rev. 59:143-169.

Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl. Environ. Microbiol. 71:7724-7736.

Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI (2005). Host bacterial mutualism in the human intestine. Science 307:1915-1920.

Bottger EC (1989). Rapid determination of bacterial ribosomal RNA sequences by direct sequencing of enzymatically amplified DNA. FEMS Microbiol. Lett. 65:171-176.

Claesson MJ, O'Sullivan O, Wang Q, Nikkila J, Marchesi JR, Smidt H, de Vos WM, Ross RP, O'Toole PW (2009). Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. PLoS One 4:e6669.

Claesson M, Wang Q, O'Sullivan O, Greene-Diniz R, Cole JR, Ross RP, O'Toole PW (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic Acids Res. 38:1-13.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 37:141-145.

Cole JR, Wang Q, Chai B, Tiedje J (2011). The Ribosomal Database Project: Sequences and software for high-throughput rRNA analysis. Handbook of Molecular Microbial Ecology, Volume I: Metagenomics and Complementary Approaches. First Edition. Frans J. De Bruijn (ed) Wiley Blackwell, John Wiley & Sons, Inc.

Darwin C (1863). The doctrine of heterogeny and modification of species. Athenaeum no. 1852, 25 April 1863:554-555. [Reprinted in: van Wyhe J 2009:334-337].

Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA (2005). Diversity of the human intestinal microbial flora. Science 308:1635-1638.

Fisher MM, Triplett EW (1999). Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. Appl. Environ. Microbiol. 65:4630-4636.

Fox GE, Wisotzkey JD, Jurtshuk, JRP (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int. J. Syst. Bacteriol. 42(1):166-170.

Greisen K, Loeffelholz M, Purohit A, Leong D (1994). PCR primers and probes for the 16S rRNA gene of most species of pathogenic bacteria, including bacteria found in cerebrospinal fluid. J. Clin. Microbiol. 32:335-351.

Gutell RR, Weiser B, Woese CR, Noller HF (1985). Comparative anatomy of 16S-like ribosomal RNA. Prog. Nucleic Acid Res. Mol. Biol. 32:155-216.

Handelsman J (2004). Metagenomics: Application of genomics to uncultured microorganisms. Microbiol. Mol. Biol. Rev. 68:669-685.

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998). Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. Chem. Biol. 5:245-249.

Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, Rubin EM (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331(6016):463-467.

Hillier SL, Krohn MA, Rabe LK, Klebanoff SJ, Eschenbach DA (1993). The normal vaginal flora, H_2O_2-producing lactobacilli, and bacterial vaginosis in pregnant women. Clin. Infect. Dis. 16(4):S273-S281.

Hooper LV, Midtvedt T, Gordon JI (2002). How host-microbial interactions shape the nutrient environment of the mammalian intestine. Annu. Rev. Nutr. 22:283-307.

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 41(1):e1.

Krawiec S, Riley M (1990). Organization of the bacterial chromosome. Microbiol. Rev. 54:502-539.

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. 82:6955-6959.

Lee I-M, Gundersen-Rindall DE, Davis RE, Bartoszky IM (1998). Revised classification scheme of phytoplasmas based on RFLP analysis of 16S rRNA and ribosomal protein gene sequences. Int. J. Syst. Bacteriol. 48:1153-1169.

Ley RE, Turnbaugh PJ, Klein S, Gordon JI (2006). Microbial ecology: Human gut microbes associated with obesity. Nature 444:1022-1023.

Li W, Godzik A (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22(13):1658-1659.

Liu W, Marsh T, Cheng H, Forney L (1997). Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Appl. Environ. Microbiol. 63:4516-4522.

Maxam AM, Gilbert W (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. Meth. Enzymol. 65:499-559.

Monstein H, Nikpour-Badi S, Jonasson J (2001). Rapid molecular identification and subtyping of *Helicobacter pylori* by pyrosequencing of the 16S rDNA variable V1 and V3 regions. FEMS Microbiol. Lett. 199:103-107.

Muyzer G, de Waal EC, Uitterlinden AG (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microbiol. 59:695-700.

Noller HF (1984). Structure of ribosomal RNA. Annu. Rev. Biochem. 53:119-162.

Ntushelo K, Harrison NA, Elliott ML (2012). Comparison of the ribosomal RNA operon from Texas Phoenix decline and lethal yellowing phytoplasmas. Eur. J. Plant Pathol. 33(4):779-782.

Olsen GJ, Lane DL, Giovannoni SJ, Pace NR (1986). Microbial ecology and evolution: A ribosomal RNA approach. Ann. Rev. Microbiol. 40:337-365.

Oparin AI (1924). Proiskhozhdenic Zhizny. Moscow: Izd. Moskovski Rabochii.

Pace NR, Stahl DA, Lane DL, Olsen GJ (1986). The analysis of natural microbial populations by rRNA sequences. Adv. Microbiol. Ecol. 9:1-55.

Pei AY, Oberdorf WE, Nossa CW, Agarwal A, Chokshi P, Gerz EA, Jin Z, Lee P, Yang L, Poles M, Brown SM, Sotero S, DeSantis T, Brodie E, Nelson K, Pei Z (2010). Diversity of 16S rRNA genes within individual prokaryotic genomes. Appl. Environ. Microbiol. 76(12):3886-3897.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J,

Glöckner FO (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35(21):7188-7196.
Radstrom P, Backman A, Qian N, Kragssbjerg P, Pahlson C, Olcen P (1994). Detection of bacterial DNA in cerebrospinal fluid by an assay for simultaneous detection of *Neisseria meningitidis, Haemophilus influenzae*, and streptococci using a seminested PCR strategy. J Clin. Microbiol. 32:2738-2744.
Sanger F, Nicklen S, Coulson AR (1977). DNA sequencing with chain-termination inhibitors. Proc. Natl. Acad. Sci. USA. 74(12):5463-5467.
Schloss PD, Handelsman J (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl. Environ. Microbiol. 71:1501-1506.
Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75:7537-7541.
Sun Y, Cai Y, Liu L, Yu F, Farrell ML, McKendree W, Farmerie W (2009). ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences. Nucleic Acids Res. 37:e76.
Sundquist A, Bigdeli S, Jalili R, Druzin ML, Waller S, Pullen KM, El-Sayed YY, Taslimi MM, Batzoglou S, Ronaghi M (2007). Bacterial flora-typing with targeted, chip-based pyrosequencing. BMC Microbiol. 7:108.
Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers Y-H, Smith HO (2004). Environmental genome shotgun sequencing of the Sargasso sea. Science 304:66-74.
Wang Q, Garrity GM, Tiedje JM, Cole JR (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73(16):5261-5267.
Woese CR, Kandler O, Wheelis ML (1990). Towards a natural system of organisms: Proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. Proc. Natl. Acad. Sci. USA 87:4576-4579.
Zoetendal EG, Akkermans AD, De Vos WM (1998). Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bac teria. Appl. Environ. Microbiol. 64:3854-3859.