[Review]

Molecular Approaches to Studying Microbial Communities: Targeting the 16S Ribosomal RNA Gene

Kazumasa Fukuda*, Midori Ogawa, Hatsumi Taniguchi and Mitsumasa Saito

Department of Microbiology, School of Medicine, University of Occupational and Environmental Health, Japan. Yahatanishi-ku, Kitakyushu 807-8555, Japan

Abstract **:** Culture-independent methods to detect microorganisms have been developed in parallel with traditional culture-based methods ever since the classification of bacteria based on 16S rRNA gene sequences was advocated in the 1970s. The development and the prevalence of culture-independent molecular technologies have provided revolutionary progress in microbial studies. The development of these technologies contributes significantly to the research of microorganisms that cannot be detected by traditional methods such as culture-dependent methods. Many molecular methods targeting the 16S rRNA gene, such as fluorescence in situ hybridization (FISH), quantitative PCR, terminal restriction fragment length polymorphism (T-RFLP), denaturing-gradient gel electrophoresis (DGGE), clone library analysis, and next-generation DNA sequencing (NGS) technologies, have been applied to various microbial studies. Notably, the advent of NGS technologies enabled a large-scale research of the bacterial community. Many recent studies using the NGS technologies have revealed that a larger number of bacteria and taxa than previously thought inhabit various parts of the human body and various places on the earth. The principles and characteristics of each molecular method are different, and each method possesses individual advantages; for example target specificity, comprehensiveness, rapidness, and cost efficiency. Therefore it is important that the methods used in studies are suitable for the objective and materials. Herein, we highlights molecular approaches targeting the 16S rRNA gene in bacterial community analysis, and focuses on the advantages and limitations of each technology.

Keywords **:** 16S rRNA gene, bacterial community analysis, DNA sequencing.

(Received February 25, 2016, accepted May 31, 2016)

Introduction

It is important to detect and identify the causative agent(s) in the diagnosis of infectious diseases. Gram-staining and cultivation methods are usually used as basic strategies to detect pathogens, but clinical cases in which a causative agent could not be detected by the culture methods are currently increasing. Moreover, antibiotic treatments according to the results of conventional methods are not effective occasionally. The present situation in the field of infectious disease diagnosis suggests that causative agents which are difficult to detect by traditional methods or complicated pathogens lurk in many patients. The remarkable advance of medical technology has led to the improvement of diagnoses and treatment of patients with infectious disease, although patients taking certain immunosuppressive drugs (such as cancer and transplant patients)

*Corresponding Author: Kazumasa Fukuda, Department of Microbiology, School of Medicine, University of Occupational and Environmental Health, Japan. Yahatanishi-ku, Kitakyushu 807-8555, Japan. Tel: +81-93-691-7242, Fax: +81-93-602-4799, E-mail: kfukuda@med.uoeh-u.ac.jp

are relatively increasing.

Mixed infections consisting of commensal microorganisms -- usually deemed "non-pathogenic microorganism" in healthy subjects -- become the target of antibiotic treatments in immunocompromised patients. In cases in which only commensal microorganisms are detected from clinical specimens, it is difficult to judge whether the detected microorganisms are causative agents or not. In these circumstances, in addition to the conventional methods, exhaustive methods to analyze the "real causative agents" are necessary for the accurate diagnosis of infectious diseases.

The development and prevalence of polymerase chain reaction (PCR) and DNA sequencing technologies [1] have expanded microbiological studies in various environments harboring vast microbial diversity that cannot be covered by culture-methods [2]. Molecular methods using PCR usually detect the "house keeping gene(s)" of the bacteria. The 16S ribosomal RNA (16S rRNA) gene is the most popular target gene in many molecular methods because the nucleotide sequences of the gene are used for the classification of organisms and are required for registering in public databases such as the DNA Data Bank of Japan (DDBJ) (http://www.ddbj.nig.ac.jp), GenBank (http://www.ncbi.nlm.nih.gov), the European Molecular Biology Laboratory (EMBL) (http://www.embl.org), *etc*.

A number of approaches have been developed to improve the detection and resolution methods targeting the 16S rRNA gene, including fluorescence in situ hybridization (FISH) [3], terminal restriction fragment length polymorphism (T-RFLP) [4] and denaturing gradient gel electrophoresis (DGGE)[5]. Though these methods, not a required sequencing process, can clarify the differences in bacterial community composition among samples simultaneously, the results are insufficient for discussing phylogenetic diversity in detail. When more specific phylogenetic information is desired, researchers employ the more laborious strategy of constructing a clone library of 16S rRNA gene amplicons [6]. This approach has made it possible to determine the fine-scale taxonomic assignment of dominant community members. Currently, next-generation sequencing (NGS) technologies, more cost effective tools, are available for studying microbial communities.

In this review, we outline the bacterial community analysis technologies targeting the 16S rRNA gene and describe the present situation of their practical usage. We also discuss their limitations and suitability for analyzing the bacterial community.

About the 16S rRNA gene

The schema of the ribosome structure and the 16S rRNA gene is shown in Fig. 1. Ribosome is a complex of plural proteins and RNA subunits found within all living cells, that plays an important role in biological protein synthesis (translation). Ribosome is composed of two major components: the small ribosomal subunit (30S ribosomal subunit in prokaryotic cells), and the large subunit (50S ribosomal subunit in prokaryotic cells). Each subunit contains one or more ribosomal RNA (rRNA) molecules and a variety of ribosomal proteins. The 16S rRNA gene encodes a ribosomal RNA molecule of 30S ribosomal subunit present in all prokaryotic cells, including bacteria and archaea. The 23S rRNA and 5S rRNA are the rRNA subunits contained in the 50S ribosomal subunit. The genes that encode for the components of ribosome have been mostly conserved, meaning that their structures have changed very little over time due to their important function, translating mRNA into proteins. Therefore, the genes are used as "house keeping genes". The classification of three domains (Eukarya, Bacteria and Archaea) was proposed in accordance with the phylogenetic tree based on rRNA gene sequences [7].

The 16S rRNA gene is a tool commonly used for identifying bacteria for several reasons. First, the gene is relatively short (approximately 1,500 bp). Second, there are ten regions in the 16S rRNA gene sequence that are common among most bacteria (conserved region) and are separated into nine diverse regions (hypervariable regions)(Fig. 1). Therefore, some universal primers are established in the conserved regions [8]. Third, the gene sequences registered in public databases are increasing substantially, because the gene sequence is important information for identification and classification in bacterial taxonomic studies.

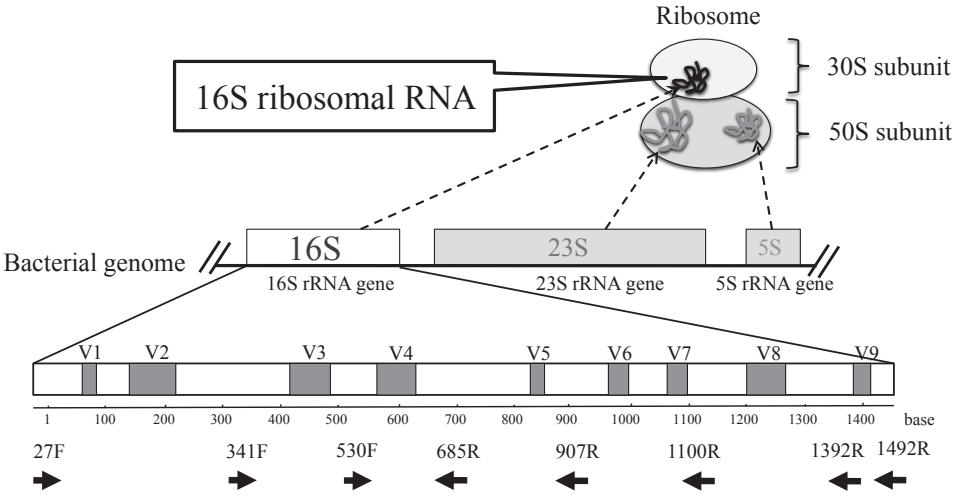

Fig. 1. The schema of ribosome complex and 16S rRNA gene. The white and grey boxes indicate conserved regions and hypervariable regions respectively. The bold arrows are shown approximate positions of universal primers on 16S rRNA gene sequence of *Escherichia coli*. □: conserved regions, ■: hypervariable regions (V1-V9).

Characteristic of the molecular methods targeting the 16S rRNA gene

The characteristics of the major molecular approaches targeting the 16S rRNA gene (Fig. 2) are described as follows.

Fluorescence in situ hybridization (FISH)

Fluorescence in situ hybridization (FISH) enables in situ phylogenetic identification and specification of individual microbial cells by hybridization of fluorescence-labeled oligonucleotide probes [3]. This method doesn't require DNA extraction or PCR processes. Moreover, it can show the shape and localization of target bacteria in the samples within a relatively short time. However, a large number of molecular probes labeled with different fluorescent materials are needed when various kinds of bacteria in the sample are targeted.

Quantitative PCR (Q-PCR)

Quantitative PCR (Q-PCR) has been used in microbial investigations to measure the abundance of the 16S rRNA genes in target bacteria [9]. Q-PCR uses either intercalating fluorescent dyes such as SYBR Green or fluorescent probes (TaqMan) to measure the accumulation of amplicons in real time during each cycle of the PCR. This method can detect only the bacteria whose 16S rRNA gene sequences have been determined.

Terminal restriction fragment length polymorphism (T-RFLP)

Terminal restriction fragment length polymorphism (T-RFLP) is based on DNA sequence variations present in PCR-amplified 16S rRNA genes [4]. PCR is performed with a fluorescently labeled primer. After PCR products are digested with restriction enzyme(s), the fluorescently labeled terminal fragments are separated and detected by high performance liquid chromatography (HPLC) or a DNA sequencer. The bacterial community diversity in a sample is estimated by analyzing the size, numbers, and peak heights of resulting fragment patterns.

Denaturing-gradient gel electrophoresis (DGGE)

The PCR products are amplified from extracted DNA using 16S rRNA gene universal primers possessing an additional Guanine-Cytosine (GC) -rich sequence (GC-clump), and then electrophoresed on a polyacrylamide gel containing a linear gradient of DNA denaturant such as a mixture of urea and formamide [5]. The sequence variation among different

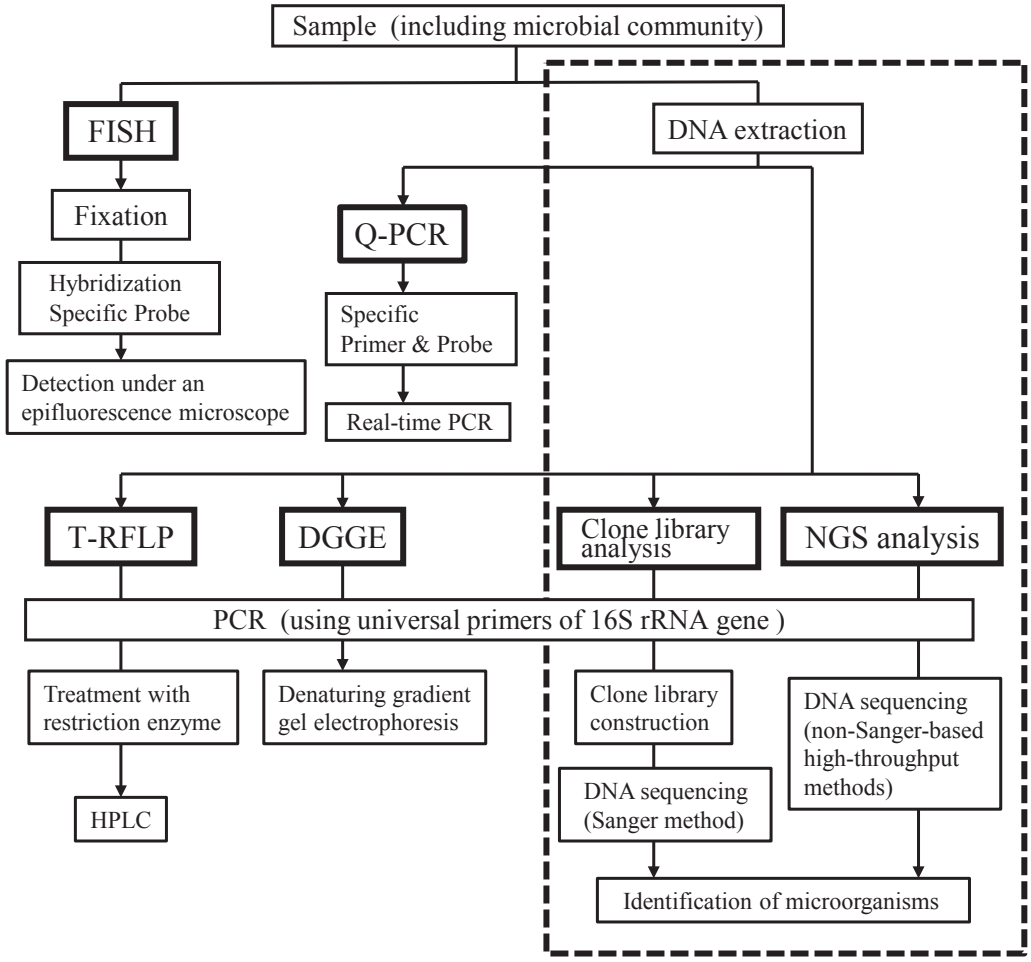

Fig. 2. Work flow of molecular methods targeting 16S rRNA gene. Each molecular method is indicated in a bold box. The broken line encircles the methods based on sequencing technologies. FISH: fluorescence *in situ* hybridization, Q-PCR: quantitative polymerase chain reaction, T-RFLP: terminal restriction fragment length polymorphism, DGGE: denaturing gradient gel electrophoresis, NGS analysis: next-generation sequencing analysis, HPLC: high performance liquid chromatography.

PCR amplicons determines the melting behavior, and therefore different amplicons stop migrating at different positions in the gel. This analysis can rapidly detect the bacterial community diversity and difference of community compositions of each sample.

Clone library analysis

After DNA extraction from the sample, 16S rRNA gene fragments are amplified with PCR, and then the amplicons are cloned into *Escherichia coli* using plasmid vector. The transformed *Escherichia* clones are randomly chosen, and each insert is sequenced by the Sanger method. Sanger sequencing, developed in 1975 by Frederick Sanger [10], is a useful technology that is still being used for various studies. The obtained high-quality sequences are compared to 16S rRNA gene sequences in databases using the Basic Local Alignment Search Tools (BLAST) algorithm. Clone libraries of 16S rRNA genes are suitable for a survey of diversity and identification of novel taxonomic linage in the samples. A capillary type sequencer using the Sanger-method gives the highest accuracy sequences among the current sequencing technologies. Therefore, the resolution of the results covers a wide range of taxonomic hierarchy. Typically, sequences are assigned to phylum, class, order, family, genus, or species at sequence similarity cut-off values of 80, 85, 90, 92, 94, or 97%, respectively [11].

Next-generation sequencing (NGS) technologies

Next-generation sequencing (NGS) represents a number of different current sequencing technologies following first-generation sequencing, known as Sanger sequencing. In the past decade, various NGS platforms have provided low-cost, high-throughput sequencing [12]. NGS platforms, including 454 GS FLX (Roche), HiSeq/MiSeq (Illumina), SOLiD (Applied Biosystems), and Ion PGM (Ion Torrent), possess the ability to sequence millions of DNA fragments in a few days. The process of amplifying the 16S rRNA gene using universal primers is necessary for both the NGS and the clone library method. NGS analyses, however, are strikingly different from the clone library method in the following two points. First, NGS methods do not require the construction of a clone library using *E. coli*. Second, the number of sequences read in one analysis using the platform is considerably larger than the usual Sanger-sequence method.

The NGS technologies, especially 454 GS FLX and Illumina platforms, have been commonly used for bacterial community studies. In the 454 GS FLX platform procedure, each PCR amplicon (16S rRNA genes) that has specific adapters on either end is fixed to a microbead individually, then the DNA fragments are amplified with an emulsion PCR. After that, the resulting beads, each of which contains many cloned copies of the same DNA fragment, are placed into a microwell (~29 μm diameter). The wells are also filled with a sequencing reaction mixture. This platform employs pyrosequencing chemistry, which utilizes pyrophosphate released during a polymerase reaction [13].

Illumina HiSeq and MiSeq platforms are most widely used for microbial community studies. In the sequencing technology of Illumina platforms, DNA fragments with specific adapters added on to either end hybridize to oligonucleotides-probes attached to the flow cell. Each fragment is then amplified to make a cluster of identical fragments (Bridge amplification). The sequencing chemistry of Illumina is similar to Sanger sequencing chemistry, but it is different from the Sanger method in that the dye terminators are reversible. After each reaction cycle, the dye at the 3' end of the nucleotides of the extending nucleotide chains are removed for the next reversible dye-terminated nucleotide [14].

Application of bacterial community analyses using molecular methods

The molecular methods without a sequencing process, such as FISH, Q-PCR, T-RFLP, and DGGE, are continuously used to analyze microbial communities in various samples, mainly from environments [15−19]. In order to evaluate the correlation between the diseases and the microbiota, and also to discover new pathogens, bacterial communities in various clinical specimens have been analyzed using the clone library method with Sanger sequencing [20−22]. The bacterial compositions in the samples were detailed and quantitatively evaluated by this method. In a study of bacterial pleurisy specimens, the clinical significance of the anaerobe, which is difficult to detect by routine culture-methods, was displayed [23]. In addition, it was strongly suggested that anaerobes and oral bacteria play more important roles in community-acquired pneumonia than previously believed [24]. In a study of bacterial vaginosis specimens, the significance of the relative ratio of *Lactobacillus* spp., *Atopobium vaginae*, and anaerobes in the bacterial community of vaginosis was clarified [25]. Moreover, in bacterial conjunctivitis, it has been revealed that indigenous bacteria are often causative-agents [26]. These results revealed that there are many clinical cases with high heterogeneous infections, and it is difficult to find them out only by clinical routine methods, such as culture methods and urinary antigen tests.

NGS platforms run much faster and give larger scale information than the Sanger method [12]. Therefore, these NGS technologies are suitable for analyzing highly diverse communities containing a large number of microbial cells. For example, it is estimated that within the human body there is a large number of microbial cells, more than 10-fold that of human cells [27]. The microbial cells assume the metabolic reactions that are necessary for human health. Therefore the microbial communities harboring in the human body have been compared to "another organ".

The Roche 454 platforms were used in the Human Microbiome Project (HMP), and produced about 7,000 sequences (16S rRNA gene) per specimen from more than 10,000 specimens from healthy adults [28]. The Illumina platform was used in a large cohort of respiratory bacterial communities in patients with adult

cystic fibrosis, to determine the individual bacterial communities in the airways of patients, to assess their relation to host factors, and to determine their dynamics in individual patients [29]. Because of the larger number of reads and the cost performance, the Roche platforms and Illumina platforms are becoming more widely used for 16S rRNA gene-sequence profiling in the microbiome-analysis trend of deeper sampling. The NGS platforms have been increasingly used in microbial community studies in various fields [30−32].

Limitations and problems

DNA extraction is the first process in microbial community analyses (Fig. 2). Analyses of microbial communities by nucleic acids-based methods (except FISH) are based on the assumption that the DNA is extracted equally from all the microbial cells in a sample; nevertheless, there are few descriptions of the cell lysis efficiency in the DNA extraction methods used thus far. The ease of bacterial destruction depends on the bacterial species and the type of sample. In fact, it has been reported that cell lysis efficiency differs according to the soil sample type, even when the same method was used [33]. However, researchers rarely try to judge whether their results correctly reflect the bacterial populations in the communities examined.

Using the clone library method for bacterial community analyses, we concurrently evaluate the cell lysis efficiency resulting from the DNA extraction process. We count bacterial cells stained with ethidium bromide under an epifluorescence microscope before and after the DNA extraction. Then cell lysis efficiency is calculated as the ratio of the number of bacteria remaining after the DNA extraction treatment to the total number before the treatment [23, 24, 26]. We endeavor to choose the effective DNA extraction method for the samples.

The PCR process has to be strictly considered, especially in the selection of primers, because the influences of primer mismatch on bacterial community analysis have been reported previously [34, 35]. The primers well known as “universal primers” were assessed by using the Probe Match program under at the Ribosomal Database Project (RDP II) website (http://rdp.cme.msu.edu). The results are shown in Table 1. The ratios of the bacteria possessing a 16S rRNA gene sequence completely identical to primer sequence vary among universal primers. The 341F, 530F, 907R, and 1392R primers were completely conserved in over 90% of 9,752 16S rDNA sequences of bacterial type strains (being over 1,200 bp with good quality) in the RDP II data set. On the other hand, only less than 40% of the 16S rRNA gene sequences of bacterial type strains completely match with 27F and 1492R primer sequences. A nucleotide sequence of approximately the full length of 16S rDNA is required for the classification of bacterial species. Nearly the full length of the gene is usually amplified by PCR using 27F and 1492R primers. Therefore, the data set contains many sequences lacking in the 27F and 1492R primer regions. Unfortunately, a genuine universal sequence that completely covers all bacteria does not exist on the 16S rRNA gene. Furthermore, it has been suggested that the degree of sequence diversity is different among nine hypervariable regions on the 16S rRNA gene [36]. As described above, researchers should consider the following 2 points when choosing primers for a bacterial community analysis: First, the primer sequences should have a high coverage ratio; second, the resulting PCR amplicon should include enough information for identification.

Until recently, the Sanger sequencing method has been used in molecular microbial surveys. The significant advantage of this technology is its sequence accuracy. The capillary sequencer using the Sanger method can read comparatively long sequences (700 to 1000 bp) with high accuracy (99.999%) (Table 2) [37]. However, using the Sanger technique, the process of clone library construction is required before sequencing. This method is more complicated and needs more time than the procedures of NGS technologies. Consequently, most of the studies using the Sanger method provide only a few hundred sequencing reads from a clone library. Sequencing of a small number of clones determines only the dominants of microbial communities. On the other hand, the Roche platform can produce approximately 1 million reads (read lengths 400 to 600 bp) in 10 hours with an accuracy of over 99% (Table 2) [38]. As a limitation of the Roche platform, relatively error-prone raw sequence data related to homopolymers, especially associated with insertion-

Table 1. Conservation ratio of universal primers in bacterial type strains

Name	Sequence (5' to 3')	Probe match result	Coverage (%)	reference
27F	AGAGTTTGATCMTGGCTCAG	1846/9752	18.9	[8]
109R	ACGYGTTACKCACCCGT	6741/9752	69.1	〃
341F	CCTACGGGAGGCAGCAG	9261/9752	94.9	〃
530F	GTGCCAGCMGCCGCGG	9576/9752	98.2	〃
685R	TCTRCGCATTYCACCGCTAC	5976/9752	61.3	〃
907R	CCGTCAATTCMTTTRAGTTT	9047/9752	92.8	〃
1100R	GGGTTGCGCTCGTTG	8424/9752	86.4	〃
1392R	ACGGGCGGTGTGTRC	9237/9752	94.7	〃
1492R	TACGGYTACCTTGTTACGACTT	3687/9752	37.8	〃

Numbered primers are named for the approximate position on the *E. coli* 16S rRNA molecule. The locations are shown in Fig.1. Probe match program is available at Ribosomal Database Project (RDP-II) website. Sequences completely matched with primer sequences are searched in 9,752 sequences of bacterial type strains (>1,200 bp with high quality).

Table 2. Comparison of Sanger sequencing and major NGS platforms

Sequencer	First generation	NGS technology	
	ABI 3730xl	454 GS FLX	HiSeq 2000
Clonal amplification	Cloning using *E.coli*	Emulsion PCR	Bridge amplification
Sequensing chemistry	Sanger sequencing	Pyrosequencing	Reversible dye terminator
Read length	700 to 1000 bp	400 to 600 bp	100 to 150 bp
Read number/run	96	1 million	1000 million
Accuracy	99.999%	99%	99.5%

NGS: next-generation sequencing

deletions, was reported [38].

The Illumina HiSeq 2000 Genome Analyzer (one of the Illumina platforms) produces approximately 1,000 million reads (100 to 150 bp) with an accuracy of more than 99.5% (Table 2) [38]. The NGS technologies can demonstrate significantly high-throughput ability but are inferior to the Sanger method in the read length and accuracy. Previous studies using these NGS platforms showed that various environmental samples, including soil and water, and specimens from the human body are composed of highly diverse microbial communities. Most of these studies analyzed the microbial communities at a higher taxonomic level, such as phylum, class, order or family. This low resolution might depend on the sequence length and accuracy of the NGS technologies, which were originally developed to determine long DNA sequences (for example "genomic DNA") at once or in a few operations.

The sequence accuracy of the NGS technologies depends on the alignment depth with enormous sequence reads (comparatively short and low quality) rather than on the individual sequence reads. Therefore each sequence read of the NGS technologies is more irregular than the Sanger method in length and accuracy. Though the problems have already been pointed out and improved methods to exclude low accurate reads have been contrived [39–41], the NGS approaches seem to be suitable for studies to elucidate the whole picture and alteration of heavily complicated communities. On the contrary, the Sanger method should be applied in studies to clarify only the predominant bacteria in a community, even in detail (at the genus or species level). The clone library method should be used in studies to identify causative agent(s) in clinical specimens usually containing a small number of microbes.

Conclusion

As a result of the exceptional increases in the number of reads and the lower cost, NGS technologies are becoming more widely used for 16S rRNA gene sequencing and continue to be used in microbial community analysis. The NGS technologies allow for cost-effective large cohorts, which are needed to reach statistically significant conclusions. However, when using 16S rRNA gene sequencing to determine causative agents in patients, more accurate identification is indispensable. In clinical situations that required speed and accuracy, the sequence reads produced by NGS platforms might be excessive and their accuracy might be insufficient. Researchers have to select a method suitable for their study objective and the analyzed sample, understanding both the advantages and limitations of technologies.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Mardis ER (2008): The impact of next-generation sequencing technology on genetics. Trends Genet 24: 133−141
2. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N & Knight R (2011): Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci USA 108: 4516−4522
3. Amann RI, Ludwig W & Schleifer KH (1995): Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59: 143−169
4. Kitts CL (2001): Terminal restriction fragment patterns: a tool for comparing microbial communities and assessing community dynamics. Curr Issues Intest Microbiol 2: 17−25
5. 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
6. Stackebrandt E, Liesack W & Goebel BM (1993): Bacterial diversity in a soil sample from a subtropical Australian environment as determined by 16S rDNA analysis. FASEB J 7: 232−236
7. 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
8. Lane DJ (1991): 16S/23S rRNA sequencing. *In*: Nucleic Acid Techniques in Bacterial Systematics. (Stackebraundt E & Goodfellow M, *ed*). John Wiley & Sons Ltd, New York pp 115−175
9. Bustin SA, Benes V, Nolan T & Pfaffl MW (2005): Quantitative real-time RT-PCR – a perspective. J Mol Endocrinol 34: 597−601
10. Sanger F, Nicklen S & Coulson AR (1977): DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463−5467
11. DeSantis TZ, Brodie EL, Moberg JP, Zubieta IX, Piceno YM & Andersen GL (2007): High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. Microb Ecol 53: 371−383
12. Metzker ML (2010): Sequencing technologies – the next generation. Nat Rev Genet 11: 31−46
13. Ronaghi M, Uhlén M & Nyrén P (1998): A sequencing method based on real-time pyrophosphate. Science 281: 363−365
14. Bentley DR, Balasubramanian S, Swerdlow HP *et al* (2008): Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456: 53−59
15. Caracciolo AB, Bottoni P & Grenni P (2010): Fluorescence *in situ* hybridization in soil and water ecosystems: a useful method for studying the effect of xenobiotics on bacterial community structure. Toxicol Environ Chem 92: 567−579
16. Smith CJ & Osborn AM (2009): Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. FEMS Microbiol Ecol 67: 6−20
17. Fierer N & Jackson RB (2006): The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci USA 103: 626−631
18. Fierer N, Jackson JA, Vilgalys R & Jackson RB (2005): Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. Appl

Environ Microbiol 71: 4117−4120

19. Nakatsu CH, Torsvik V & Øvreås L (2000): Soil community analysis using DGGE of 16S rDNA polymerase chain reaction products. Soil Sci Soc Am J 64: 1382−1388
20. Woo PC, Lau SK, Teng JL, Tse H & Yuen KY (2008): Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. Clin Microbiol Infect 14: 908−934
21. Harris JK, De Groote MA, Sagel SD, Zemanick ET, Kapsner R, Penvari C, Kaess H, Deterding RR, Accurso FJ & Pace NR (2007): Molecular identification of bacteria in bronchoalveolar lavage fluid from children with cystic fibrosis. Proc Natl Acad Sci USA 104: 20529−20533
22. Fredricks DN, Fiedler TL & Marrazzo JM (2005): Molecular identification of bacteria associated with bacterial vaginosis. N Engl J Med 353: 1899−1911
23. Kawanami T, Fukuda K, Yatera K, Kido M, Mukae H & Taniguchi H (2011): A higher significance of anaerobes: the clone library analysis of bacterial pleurisy. Chest 139: 600−608
24. Yamasaki K, Kawanami T, Yatera K *et al* (2013): Significance of anaerobes and oral bacteria in community-acquired pneumonia. PLoS One 8: e63103
25. Yoshimura K, Morotomi N, Fukuda K, Nakano M, Kashimura M, Hachisuga T & Taniguchi H (2011): Intravaginal microbial flora by the 16S rRNA gene sequencing. Am J Obstet Gynecol 205: 235. e1− e9
26. Aoki R, Fukuda K, Ogawa M, Ikeno T, Kondo H, Tawara A & Taniguchi H (2013): Identification of causative pathogens in eyes with bacterial conjunctivitis by bacterial cell count and microbiota analysis. Ophthalmology 120: 668−676
27. The NIH HMP Working Group, Peterson J, Garges S *et al* (2009): The NIH Human Microbiome Project. Genome Res 19: 2317−2323
28. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R & Gordon JI (2007): The human microbiome project. Nature 449: 804−810
29. Kramer R, Sauer-Heilborn A, Welte T, Jauregui R, Brettar I, Guzman CA & Höfle MG (2015): High individuality of respiratory bacterial communities in a large cohort of adult cystic fibrosis patients under continuous antibiotic treatment. PLoS One 10: e0117436
30. Sze MA, Dimitriu PA, Suzuki M *et al* (2015): Host response to the lung microbiome in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 192: 438−445
31. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, Wilkins MJ, Wrighton KC, Williams KH & Banfield JF (2015): Unusual biology across a group comprising more than 15% of domain Bacteria. Nature 523: 208−211
32. Lauber CL, Hamady M, Knight R & Fierer N (2009): Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl Environ Microbiol 75: 5111−5120
33. Zhou J, Bruns MA & Tiedje JM (1996): DNA recovery from soils of diverse composition. Appl Environ Microbiol 62: 316−322
34. Ishii K & Fukui M (2001): Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. Appl Environ Microbiol 67: 3753−3755
35. Sipos R1, Székely AJ, Palatinszky M, Révész S, Márialigeti K & Nikolausz M (2007): Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. FEMS Microbiol Ecol 60: 341−350
36. Chakravorty S, Helb D, Burday M, Connell N & Alland D (2007): A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J Microbiol Methods 69: 330−339
37. Shendure J & Ji H (2008): Next-generation DNA sequencing. Nat Biotechnol 26: 1135−1145
38. Zhang J, Chiodini R, Badr A & Zhang G (2011): The impact of next-generation sequencing on genomics. J Genet Genomics 38: 95−109
39. Tremblay J, Singh K, Fern A, Kirton ES, He S, Woyke T, Lee J, Chen F, Dangl JL & Tringe SG (2015): Primer and platform effects on 16S rRNA tag sequencing. Front Microbiol 6: 771
40. Schloss PD, Gevers D & Westcott SL (2011): Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS One 6: e27310
41. Huse SM, Huber JA, Morrison HG, Sogin ML & Welch DM (2007): Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol 8: R143

16S rRNA遺伝子を標的とした細菌叢解析手法

福田　和正，小川　みどり，谷口　初美，齋藤　光正

産業医科大学　医学部　微生物学

要　　旨：1970年代に16S rRNA遺伝子の塩基配列に基づく細菌の系統的な生物種分類法が提唱されて以来，微生物を検出するため，従来の培養法に加え，培養法に依存しない分子生物学的手法が考案されてきた．その開発と普及は微生物研究に革命的な進歩をもたらし，従来の培養法では検出できない細菌の研究に大きく寄与している．蛍光 *in situ* ハイブリダイゼーション法（FISH），定量PCR（Q-PCR），末端標識制限酵素断片多型分析（T-RFLP），変性剤濃度勾配ゲル電気泳動法（DGGE），クローンライブラリー解析や次世代型DNAシークエンス解析などの16S rRNA遺伝子を標的とした分子生物学的手法は，さまざまな微生物研究に応用されている．特に，次世代型DNAシークエンサーを用いた多くの研究は，大規模な細菌叢解析を可能にしており，最近の多くの研究は人体のさまざまな部位や，地球上のさまざまな場所に予想以上の数と種類の細菌が生息していることを明らかにしている．これらの分子生物学的手法は，それぞれの方法で原理や特徴は異なり，標的特異性，網羅性，迅速性や経済性などにおいて，それぞれ独自の利点を有している．それゆえ，研究の目的や対象に応じて，適した手法を選択することは重要である．本稿では，細菌叢解析に用いられる16S rRNA遺伝子を標的とした手法について概説し，それぞれの手法の利点や限界について考察する．

キーワード：16S rRNA遺伝子，細菌叢解析，DNAシーケンス．

J UOEH（産業医大誌）38（3）：223 − 232（2016）