JRMI
Journal of Rehman Medical Institute

Volume 5, No. 2
April - June 2019
www.jrmi.pk

ORIGINAL ARTICLE

Large scale differences in characterization of gut microbiome between 16S amplicon and shallow shotgun sequencing

Bibi Sabiha, Syed Adnan Haider, Hanifullah Jan, Ome Kalsoom Afridi, Johar Ali

Submitted
March 28, 2019
Accepted
April 15, 2019

Author Information

From: Center for Genomic Sciences (CGS), Rehman Medical Institute (RMI), Peshawar, Khyber Pakhtunkhwa (KP), Pakistan.

Ms. Bibi Sabiha
Scientific Officer

Syed Adnan Haider
Scientific Officer

Hanifullah Jan
Technologist

Ms. Ome Kalsoom Afridi
Research Associate
Department of Biology Education, Kyungpook National University, Daegu 41566, Republic of Korea

Dr. Johar Ali
Head, Center for Genomic Sciences
(Corresponding author)
Email: johar.ali1@rmi.edu.pk

Citation: Sabiha B, Haider SA, Jan H, Afridi OK, Ali J. Large scale differences in characterization of gut microbiome between 16S amplicon and shallow shotgun sequencing. J Rehman Med Inst. 2019 Apr-Jun;5(2):3-6.

ABSTRACT

Introduction: The advent of next generation sequencing (NGS) technology aided exponentially to the understanding of the gut microbiome. The most commonly used approaches of NGS are 16S rRNA and shotgun metagenome sequencing. 16S rRNA sequencing provides microbial resolution only up to the genus level, while deep shotgun metagenome sequencing provides taxonomic profiling down to the species level but the high cost hampers its usage. Alternate method, Shallow shotgun metagenome sequencing (SSMS) emerged as a cost effective high throughput method for microbial profiling.

Objective: To compare the consistency of 16S rRNA and SSMS NGS methods in gut microbial profiling.

Materials & Methods: DNA was extracted from the stool samples of five infants, and the full 16S rRNA region was amplified. Sequencing libraries were prepared for 16S rRNA sequencing and SSMS, loaded onto the flow cell, sequenced using the MiSeq platform (Illumina), and analyzed through custom bioinformatics analysis pipeline.

Results: NGS data analysis revealed that gut microbiome characterized using both 16S rRNA sequencing and SSMS showed large scale differences in terms of reported species. Some of the microbes detected using 16S rRNA sequencing were found to be missing while using SSMS and vice versa. Very few microbes were commonly detected by both sequencing approaches.

Conclusion: The application of both NGS sequencing approaches is warranted for accurate bacterial profiling from different sources.

Keywords: Gut Microbiome; 16S rRNA Sequencing; Shallow Shotgun Metagenomic Sequencing; Microbial Profiling.

The authors declared no conflict of interest. All authors contributed substantially to the planning of research, data collection, data analysis, and write-up of the article, and agreed to be accountable for all aspects of the work.

INTRODUCTION

The 400 B.C. Hippocrates quote “death lies within the bowel, and poor digestion is the root of all diseases” shows the importance of gut microbiome in human health and disease. In the first days of life, the gut microbiota develop and continue to increase and diversify with time.[1] Imbalanced gut microbiota have been associated with various diseases.[2] In addition to playing several essential roles in human health and conditions, microbiota have emerging roles in ecology, making their analysis in different environments, animals, and humans essential.[3] Though it is now clear that human gut microbiome is associated with health and disease, yet complete knowledge of gut bacterial repertoire is undefined and have been unrecognized since long due to unavailability of state of the art techniques.[4] Rapid identification of bacterial pathogens is also critical for classification, proper diagnosis, and treatment.[5]

The emergence of High Throughput Sequencing (HTS) technology revolutionized traditional microbial identification approaches by making identification very authentic, robust, and specific. Next Generation Sequencing (NGS) greatly aided in understanding the role of gut microbiota in healthy and diseased conditions and accelerated diagnostic and therapeutic measures.[6] NGS is the gold standard technique to characterize and identify microbial community structures from different habitats and robustly provide critical insights into various attributes of the human gut microbiome. For this purpose, Deep Shotgun Metagenome Sequencing (DSMS) is most commonly employed to study the gut microbial community. DSMS is considered the best in terms of high taxonomic resolution but is extremely expensive for large scale studies. Therefore the affordable alternative methods such as 16S rRNA and Shallow Shotgun Metagenome Sequencing (SSMS) have been widely accepted for cost effective microbial profiling.

Bacterial 16S rRNA is considered a potential marker for identification and phylogenetic analysis of bacteria.[7] More attention has been given to 16S rRNA sequencing since it can better identify non-cultured bacteria at low cost.[8] The 16S rDNA is a

stretch of DNA (codes RNA component of the 30S subunit of bacterial ribosome) with 1500bp consisting of highly conserved and variable regions.

This highly conserved region provides an opportunity to amplify 16S rRNA regions of different bacteria with the same primer set, whereas the hypervariable region supports the rapid identification and classification of bacteria using the amplified region. These variable regions vary from V1 to V9 which are diverse among different bacteria and species specific. Therefore, they are used for diagnostic and scientific investigations.[9] While searching for effective and economical methods keeping in view that either 16S rRNA sequencing or SSMS technique is best for bacterial identification, this study compares the 16S rRNA sequencing and SSMS approaches.

MATERIALS & METHODS

A. Next Generation Sequencing

Sample collection and DNA extraction

The study was conducted at the Center for Genomic Sciences (CGS) in Rehman Medical Institute (RMI), Peshawar, Khyber Pakhtunkhwa, Pakistan. Fecal samples were collected from five infants residing in city area of Peshawar, Khyber Pakhtunkhwa (KP), Pakistan. After giving their informed consents, parents of sampled children collected fecal samples in sterilized containers and stored them at home refrigerators for a maximum of 6 hours.. Homogenized stool samples of 200 mg were prepared in 2ml sterilized screw capped tubes and stored at -80°C before DNA extraction. Bacterial DNA was extracted from each stool sample of infants, and each sample was subjected to both 16S rRNA and SSMS sequencing methods to identify eubacteria. Microbial DNA was extracted using the QIAamp DNA stool mini kit (Qiagen, Valencia, CA, USA) following protocol. The quality and quantity of extracted DNA was determined using agarose gel electrophoresis and Qubit fluorometer using dsDNA High Sensitivity kit, respectively (Qubit, Cat. #Q32851; Invitrogen). Each quantified DNA was further used for both 16S rRNA and SSMS.

PCR amplification of 16S rDNA

Primers covering 16S rDNA region (1500bp) of most of eubacteria (Table 1)[10] were used to amplify the targeted region. Phusion High Fidelity PCR Master Mix was used for PCR amplification and was performed in a T100™ Thermal Cycler (Bio Rad) with incubation step of 95°C for 10 sec and 20 cycles at 95°C for 03 sec, 66°C for 10 sec and 72°C for 01 min, followed by a final extension step of 72°C for 3 minutes. The amplified products were run on 1.5% agarose gel and quantified through Qubit fluorometer and normalized for library preparation.

Table 1. Primers used for the amplification of 16S rDNA region

Primer name	Primer Sequence	Size (bp)
16S-F	AGAGTTTGATCCTGGCTCAG	1500
16S-R	ACGGCTACCTTGTTACGACTT	

F: Forward, R: Reverse

Library preparation

Libraries of the amplified products and genomic DNA were prepared through Nextera XT DNA Library Preparation Kit (Illumina, San Diego, USA) following the manufacturer instructions. One ng of input DNA was used for library preparation, where in the first step tagmentation was done, followed by PCR amplification of libraries. Then the cleanup and normalization of the libraries were done by bead based method. Normalized libraries were pooled and loaded onto the Illumina MiSeq using V2 sequencing reagent kit, and 150bp paired end sequencing was performed using NGS technology at Rehman Medical Institute (RMI), Hayatabad, Peshawar, Pakistan.

NGS data analysis

FASTQ files generated from NGS data were analyzed using a variety of publicly available bioinformatics softwares. CASAVA v1.8.2 package[11, 12] was used to demultiplex all FASTQ files. All the raw FASTQ files were filtered using Trimmomatic v.0.36 and all technical biases, low quality reads, and adapters were removed. A computational tool KneadData v. 0.6.1[13] was used to decontaminate the microbial reads from the host DNA. Sequence alignment of the microbial reads against the NCBI non-redundant protein reference database (NCBI-NR) was performed using DIAMOND v.0.9.25.[14] The final microbial profiling was carried out using MEtaGenome ANalyzer (MEGAN 6) and Metaphlan 2.0 for SSMS and 16S rRNA sequencing, respectively.[15]

B. Missing samples validation

Validation of missing eubacterial species in 16S rRNA and SSMS

After the comparison of the data, the missing eubacterial species from each technique were further validated. 16S rRNA region of missing bacteria was downloaded from NCBI and aligned with filtered reads of 16S rRNA for validation of missing samples in 16S rRNA sequencing. Similarly, for the validation of missing bacteria in SSMS, the reference genome of the misssing bacteria was dowloaded from NCBI and aligned with filtered reads of shotgun samples.

RESULTS

A. 16S rRNA VERSUS SSMS

In this study eubacteria in the gut of infant stool samples was detected by two different approaches, 16S rRNA and SSMS. A total of five samples were used for both 16S rRNA and SSMS. After library preparation and sequencing the data was analyzed through in house developed pipeline. In all five samples, we found a more significant number of bacteria in SSMS than the 16S rRNA. Some bacteria were found common in both WGS and SSMS.

Some eubacteria were missing in 16S rRNA and present in SSMS while some were present in 16S rRNA and missing in SSMS. Table 2 shows the number of eubacteria identified, missing, and common in both approaches.

Table 2: Details of the number of eubacteria reported in SSMS and 16S rRNA sequencing.

Sample ID	Number of eubacteria reported in SSMS	Number of eubacteria reported in 16S rRNA	Number of eubacteria common in SSMS and 16S rRNA	Number of eubacteria missing in 16S rRNA	Number of eubacteria missing in SSMS
1S	28	19	01	18	27
2S	19	19	01	18	18
3S	15	21	01	14	20
4S	36	24	02	34	22
5S	18	22	02	16	20

B. VALIDATION

Missing eubacteria in each technique were further validated by aligning their respective data with the reference sequence. When the reference sequence of the 16S rRNA region of missing eubacteria in 16S rRNA sequencing were aligned with filtered reads data of 16S rRNA for validation none of the eubacteria were found.

Similarly, none of the eubacteria were found in shotgun data when missing eubacteria in SSMS reference geneome were aligned with filtered reads of shotgun data.

DISCUSSION

The development of NGS has enabled the researchers to study the microorganism in a better way. It not only enables to characterize the genome but also facilitates the more profound taxonomic identification of microbial community.[16] The flexibility of NGS technology allows the identification of uncultured bacteria through a culture independent strategy from different sources. [17] NGS is mainly applied for bacterial identification in different ways:16S rRNA and shotgun metagenome sequencing.[18] Both methods have their advantages and limitations. 16S rRNA sequencing for bacterial identification is widely used because of its low cost, but it cannot accurately identify the microbial population as it faces PCR amplification biases. In contrast, the shotgun metagenome sequencing allows unbiased microbial profiling but demand a high rate of sequencing depth which make it more expensive.[19] Recently the other method, SSMS become more popular for taxonomic profiling by sequencing the whole genome at a shallower depth.[20]

To select the most suitable method for bacterial identification different studies have compared these methods[21]. Some studies have preferred shotgun metagenome sequencing[22] while others studies prefer 16S rRNA sequencing as it is more microbiome specific and eliminate the other organism DNA contamination which will be supportive in identification of low abundant bacteria.[23]

The advantages of 16S rRNA sequencing for bacterial identification is that it is cost effective; its data is analyzed by established pipelines and availability of extensive archived data for reference. In 16S rRNA sequencing, the microorganism is identified based on the hyper variable regions of the 16S rDNA[24] while shotgun metagenomics identification is based on whole genome data and not only identifies bacteria at the species level but also to the strain level along with functional characterization.[25,26]

The other method being used is SSMS, an economical way of sequencing for bacterial profiling at a shallower depth than the deep shotgun sequencing.[20] Here in this study, we compared the two affordable NGS based sequencing approaches, 16S rRNA and SSMS, for the identification of eubacteria in gut of infants. This comparative study showed that the outcome of SSMS is more descriptive and informative than 16S rRNA sequencing.

These two methods are not consistent with bacterial identification. The number of microbes reported in SSMS was higher than the number reported through 16S rRNA sequencing. Some eubacteria were unique in SSMS, and some were unique in 16S rRNA sequencing. The difference in abundance of eubacteria in both sequencing methods may be due to difference in the primer binding region in the conserved region of bacteria and low coverage in the case of SSMS. Previous studies have shown that conserved areas are not truly conserved, and they offer substantial variability, which should be considered when using 16S rRNA as an identification marker.[27] Comparative study of 16S rRNA and shotgun sequencing of DNA from stool samples in different platforms of NGS has reported that both methods have almost similar results.[28] Similarly, water sample DNA analyzed by the V3-V4 region of 16S rRNA and shotgun metagenome sequencing showed the same number of phyla in both methods.[29] Both studies are different from the result of this study, which we assume may be that SSMS missed the eubacteria detected by 16S rRNA sequencing due to low coverage.

CONCLUSION

Comparison of 16S rRNA sequencing and SSMS applied to identify eubacteria in gut of infants provided novel insights. Although 16S rRNA sequencing and SSMS are considered as the cost effective method for taxonomic profiling, but individually they cannot identify all the eubacteria found in samples. The coupling of 16S rRNA and SSMS sequencing can lead to better identification of bacteria, especially when dealing with complex microbiota.

REFERENCES

1. Asnicar F, Manara S, Zolfo M, Truong DT, Scholz M, Armanini F, et al. Studying vertical microbiome transmission from mothers to infants by strain level metagenomic profiling. mSystems 2017;2(1):e00164-00116.
2. Shen T-CD. Diet and gut microbiota in health and disease. In: Isolauri E, Sherman PM, Walker WA (eds): Intestinal Microbiome: Functional Aspects in Health and Disease. Nestlé Nutr Inst Workshop Ser. Nestec Ltd. Vevey/S. Karger AG Basel, © 2017, vol 88, pp 117-26.
3. Laudadio I, Fulci V, Palone F, Stronati L, Cucchiara S, Carissimi C. Quantitative assessment of shotgun metagenomics and 16S rDNA amplicon sequencing in the study of human gut microbiome. OMICS. 2018 Apr;22(4):248-54.
4. Sekirov I, Russell SL, Antunes LCM, Finlay BB. Gut microbiota in health and disease. Physiol Rev. 2010;90(3):859-904.
5. Kai S, Matsuo Y, Nakagawa S, Kryukov K, Matsukawa S, Tanaka H, et al. Rapid bacterial identification by direct PCR amplification of 16S rRNA genes using the MinION™ nanopore sequencer. FEBS Open Bio. 2019 Jan 29;9(3):548-57.
6. Kashyap PC, Chia N, Nelson H, Segal E, Elinav E. Microbiome at the frontier of personalized medicine. Mayo Clin Proc. 2017 Dec;92(12):1855-64.
7. Yang B, Wang Y, Qian P-Y. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. BMC Bioinformatics. 2016;17(1):1-8.
8. Clarridge JE. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev. 2004 Oct;17(4):840-62.
9. Chakravorty S, Helb D, Burday M, Connell N, Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J Microbiol Methods. 2007;69(2):330-9.
10. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol. 1991;173(2):697-703.
11. Mehrad B, Clark NM, Zhanel GG, Lynch III JP. Antimicrobial resistance in hospital-acquired gram-negative bacterial infections. Chest. 2015;147(5):1413-21.
12. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics. 2012;13(1):31.
13. Nelson MT, Pope CE, Marsh RL, Wolter DJ, Weiss EJ, Hager KR, et al. Human and extracellular DNA depletion for metagenomic analysis of complex clinical infection samples yields optimized viable microbiome profiles. Cell Reports. 2019;26(8):2227-40. e2225.
14. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nature Methods. 2015;12(1):59-60.
15. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. Nature Methods. 2012;9(8):811-4.
16. Peker N, Garcia-Croes S, Dijkhuizen B, Wiersma HH, van Zanten E, Wisselink G, et al. A Comparison of three different bioinformatics analyses of the 16S–23S rRNA encoding region for bacterial identification. Frontiers Microbiol. 2019;10:620.
17. Boers SA, Jansen R, Hays JP. Understanding and overcoming the pitfalls and biases of next-generation sequencing (NGS) methods for use in the routine clinical microbiological diagnostic laboratory. Eur J Clin Microbiol Infect Dis. 2019;38(6):1059-70.
18. Tessler M, Neumann JS, Afshinnekoo E, Pineda M, Hersch R, Velho LFM, et al. Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. Scientific Reports. 2017;7(1):1-14.
19. Cottier F, Srinivasan KG, Yurieva M, Liao W, Poidinger M, Zolezzi F, et al. Advantages of meta-total RNA sequencing (MeTRS) over shotgun metagenomics and amplicon-based sequencing in the profiling of complex microbial communities. NPJ Biofilms Microbiomes. 2018;4(1):1-7.
20. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, et al. Evaluating the information content of shallow shotgun metagenomics. mSystems. 2018 Nov 13;3(6):e00069-18.
21. Clooney AG, Fouhy F, Sleator RD, O'Driscoll A, Stanton C, Cotter PD, et al. Comparing apples and oranges?: next generation sequencing and its impact on microbiome analysis. PloS One. 2016;11(2):e0148028.
22. Ranjan R, Rani A, Metwally A, McGee HS, Perkins DL. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. Biochem Biophys Res Comm. 2016;469(4):967-77.
23. Jovel J, Patterson J, Wang W, Hotte N, O'Keefe S, Mitchel T, et al. Characterization of the gut microbiome using 16S or shotgun metagenomics. Frontiers Microbiol. 2016;7:459.
24. Bukin YS, Galachyants YP, Morozov I, Bukin S, Zakharenko A, Zemskaya T. The effect of 16S rRNA region choice on bacterial community metabarcoding results. Scientific Data. 2019;6:190007.
25. Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. Microbial strain-level population structure and genetic diversity from metagenomes. Genome Res. 2017;27:626-38.
26. van der Walt AJ, van Goethem MW, Ramond J-B, Makhalanyane TP, Reva O, Cowan DA. Assembling metagenomes, one community at a time. BMC Genomics. 2017;18:1-13.
27. Martinez-Porchas M, Villalpando-Canchola E, Suarez LEO, Vargas-Albores F. How conserved are the conserved 16S-rRNA regions? Peer J. 2017;5:e3036.
28. Mitra S, Förster-Fromme K, Damms-Machado A, Scheurenbrand T, Biskup S, Huson DH, et al. Analysis of the intestinal microbiota using SOLiD 16S rRNA gene sequencing and SOLiD shotgun sequencing. BMC Genomics. 2013;14(S5):S16.
29. Chan CS, Chan K-G, Tay Y-L, Chua Y-H, Goh KM. Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. Frontiers Microbiol. 2015;6:177.