Microbiol. Biotechnol. Lett. (2020), 48(4), 574–581
http://dx.doi.org/10.48022/mbl.2008.08003
pISSN 1598-642X eISSN 2234-7305

MBL
Microbiology and Biotechnology Letters

Comparison of the Performance of MiSeq and HiSeq 2500 in a Microbiome Study

Hee Sam Na[1,2†], Yeuni Yu[3†], Si Yeong Kim[1,2], Jae-Hyung Lee[4,5,6], and Jin Chung[1,2*]

[1]*Department of Oral Microbiology, School of Dentistry,* [2]*Oral Genomics Research Center,* [3]*Interdisplinary Program of Genomic Science, Pusan National University, Busan 46241, Republic of Korea*
[4]*Department of Oral Microbiology, School of Dentistry,* [5]*Department of Life and Nanopharmaceutical Sciences,* [6]*Kyung Hee Medical Science Research Institute, Kyung Hee University, Seoul 02447, Republic of Korea*

Received: August 6, 2020 / Revised: November 11, 2020 / Accepted: November 13, 2020

Next generation sequencing is commonly used to characterize the microbiome structure. MiSeq is commonly used to analyze the microbiome due to its relatively long read length. However, recently, Illumina introduced the 250x2 chip for HiSeq 2500. The purpose of this study was to compare the performance of MiSeq and HiSeq in the context of oral microbiome samples. The MiSeq Reagent Kit V3 and the HiSeq Rapid SBS Kit V2 were used for MiSeq and HiSeq 2500 analyses, respectively. Total read count, read quality score, relative bacterial abundance, community diversity, and relative abundance correlation were analyzed. HiSeq produced significantly more read sequences and assigned taxa compared to MiSeq. Conversely, community diversity was similar in the context of MiSeq and HiSeq. However, depending on the relative abundance, the correlation between the two platforms differed. The correlation between HiSeq and MiSeq sequencing data for highly abundant taxa (> 2%), low abundant taxa (2–0.2%), and rare taxa (0.2% >) was 0.994, 0.860, and 0.416, respectively. Therefore, HiSeq 2500 may also be compatible for microbiome studies. Importantly, the HiSeq platform may allow a high-resolution massive parallel sequencing for the detection of rare taxa.

Keywords: **Microbiome, HiSeq, MiSeq, next generation sequencing**

Introduction

A collection of microbial taxa in a given environment is defined as microbiota [1]. There has been numerous studies reporting changes of the human microbiome linked to various chronic diseases including periodontitis [2], obesity [3], inflammatory bowel disease [4], cancer [5] and Alzheimer's disease [6]. Many parameters have been reported to influence microbiota composition, ranging from host genotype [7], nutrition [8], inflammation [9], antibiotic usage [10] and other unknown factors such as medication, geological area and host age. Thus, upscaling microbiome studies are required to disentangle the multiple, confounding effects of various factors in the microbiome.

The human oral cavity has been a model for advanced microbiome analysis to gain an understanding of microbial ecology [11]. A wide variety of bacterial types have been reported with more than 700 species present in the oral cavity with distinctive pattern of microbial species on both hard and soft tissue oral surfaces [12]. Although a significant portion (≈30%) of oral microbial species remains uncultivable (http://www.homd.org), oral microbiome is well characterized compared to other microbiome research areas. In diseases such as periodontitis or peri-implantitis, alteration of microbiome in subgingival plaques are also noted [13, 14]. Oral samples can be col-

***Corresponding author**
Tel: +82-51-510-8245, Fax: +82-51-510-8246
E-mail: jchung@pusan.ac.kr
[††]These authors contributed equally to this work.

lected repetitively, easily, and in most cases noninvasively. Thus, oral cavity is very interesting site for microbiome study.

Research into microbial ecology has expanded enormously due to advances in DNA sequencing, which now enables researchers to probe microbial community composition and function in a high-resolution and culture-independent manner. For microbiome study, there are several DNA sequencing platforms, including 454 pyrosequencing [15], MiSeq and HiSeq from Illumina [16], Ion Torrent [17], and so on. Especially, interest in 16S rRNA gene amplicon sequencing on Illumina is growing, largely due to lower cost per sequence and lower sequencing error rate than other platforms, enabling high throughput microbial ecology at the greatest coverage [18]. HiSeq and MiSeq platforms are among the most widely used platform to study microbial communities. But the two platforms differ in the length and amount of reads. MiSeq can run 600 cycles to produce 200 million 300-bp reads, on the other hand, HiSeq 2500 can run 500 cycles to produce 120 million 250 bp.

In this study, we used oral samples from patients with periodontitis simultaneously sequenced by MiSeq and HiSeq platforms to determine the similarity and difference between two platforms.

Material and Methods

Study population and clinical examination

Plaque samples were obtained from patients with periodontitis who were scheduled to undergo periodontal treatment at the Department of Periodontics of Pusan National Dental School, Yangsan, Korea. The samples were collected from a total of 8 patients comprising 2 male patients and 6 female patients. The average age of the patients was 51.4 ± 11.4 years. Buccal, supragingival and subgingival plaque samples were collected with the full-mouth periodontal examination. All participants were requested to refrain from food and oral hygiene (brushing or flossing the teeth) for 2 h before sampling. Samples were collected after isolating the selected sampling site with microbrush. Buccal samples were obtained from the mucosa of both the cheeks. Subgingival and supragingival plaque samples were collected from the molars of each participant using a sterile Gracey curette. The experimental protocol was approved by the Institutional Review Board of Pusan National University (PNUDH-2017-023). Written informed consent was obtained from all participants before the study. The plaque samples were stored at -80°C until analyzed.

Extraction of total genomic DNA

Total DNA was extracted using a Gram positive DNA purification kit (Lucigen, Biosearch Technology, USA) following the manufacturer's instructions. The final concentration was measured with a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, USA) and stored at -80°C until use.

PCR amplification and sequencing analysis

Library construction and sequencing were performed by Macrogen (Korea). Each sequenced sample was prepared according to the Illumina 16S Metagenomic Sequencing Library protocols to amplify the V3 and V4 region (314F-806R). The barcoded fusion primer sequences used for amplifications were as follows: 314F: 5'-CCT ACG GGN GGC WGC AG-3', 806R: 5'-GAC TAC HVG GGT ATC TAA TCC-3'. The DNA quality was measured by PicoGreen and Nanodrop. Input gDNA (10 ng) was PCR amplified. The final purified product was then quantified using qPCR according to the qPCR Quantification Protocol Guide (KAPA Library Quantificatoin kits for Illumina Sequecing platforms) and qualified using the LabChip GX HT DNA High Sensitivity Kit (PerkinElmer, USA). Paired-end sequencing was performed using MiSeq Ragent kit V3 for the MiSeq (2 × 300 bp) or HiSeq Rapid SBS kit V2 for HiSeq 2500 (2 × 250 bp) (Illumina, USA). Raw sequencing data was filtered and trimmed by using QIIME package version 2 (Caporaso *et al.*, 2010). Both 8 bases were trimmed from the start in forward and reverse reads. Same trimming size was applied to both platforms. Forward and backward reads were joined with join_paired_ends.py command. Chimeras were identified and filtered using *usearch* method (Rognes *et al.*, 2016). Finally, the tool was used to pick closed-reference OTUs from the Human oral microbiome database (HOMD) [12].

Bioinformatic analysis, statistical analysis, and visualization

Alpha diversity was used to describe the microbial richness, evenness and diversity within samples using

the Chao1 and Shannon index. Principal coordinate analysis (PCoA) of the Bray-Curtis distance was performed to determine the change in the community structure using the vegan package v2.3-0 in R software v3.2.1. Relative abundance correlation between HiSeq and MiSeq was determined by Pearson's correlation. The taxonomy compositions and abundances of different samples were visualized by R and GraphPad PRISM software (version 4.0).

Data availability

The raw sequencing data have been deposited at NCBI GenBank under BioProject ID PRJNA649363 (BioSample SAMN15664667 - SAMN15664690). Please check the data using the below private reviewer link, https://dataview.ncbi.nlm.nih.gov/object/PRJNA649363?reviewer=fccf7o3gmp30ceb71p23uf4sp0.

Result

Proportions for good quality read sequences of total read

From HiSeq sequencing, we have obtained a total of 6,689,663 raw reads, corresponding to 175,143 to 403,487 reads per sample (average 278,736 ± 50,716). The final data set after removing low-quality reads and checking for chimeras contained 695,643 reads, corresponding to 6,471 to 89,062 reads per sample (average 28,985 ± 22,067). From MiSeq sequencing, we have obtained a total of 3,937,859 raw reads, corresponding to 126,378 to 238,406 reads per sample (average 164,077 ± 25,697). The final data set after removing low-quality and chimera reads, number of reads per sample ranged from 2,877 to 47,380 (average 15,543 ± 11,347) (Table S1).

Read sequence quality

To evaluate and control sequencing data quality, a plot of a random sample was generated using a random sampling of 10000 out of total input sequences without replacement. Each forward and reverse plot represents the parametric seven-number summary of the quality scores at the corresponding position (Fig. 1). The read length produced was 301 bases for MiSeq and 251 bases for HiSeq. The forward reads produced high quality in both MiSeq and HiSeq sequencing. The reverse reads showed lower sequencing quality. Sequences average

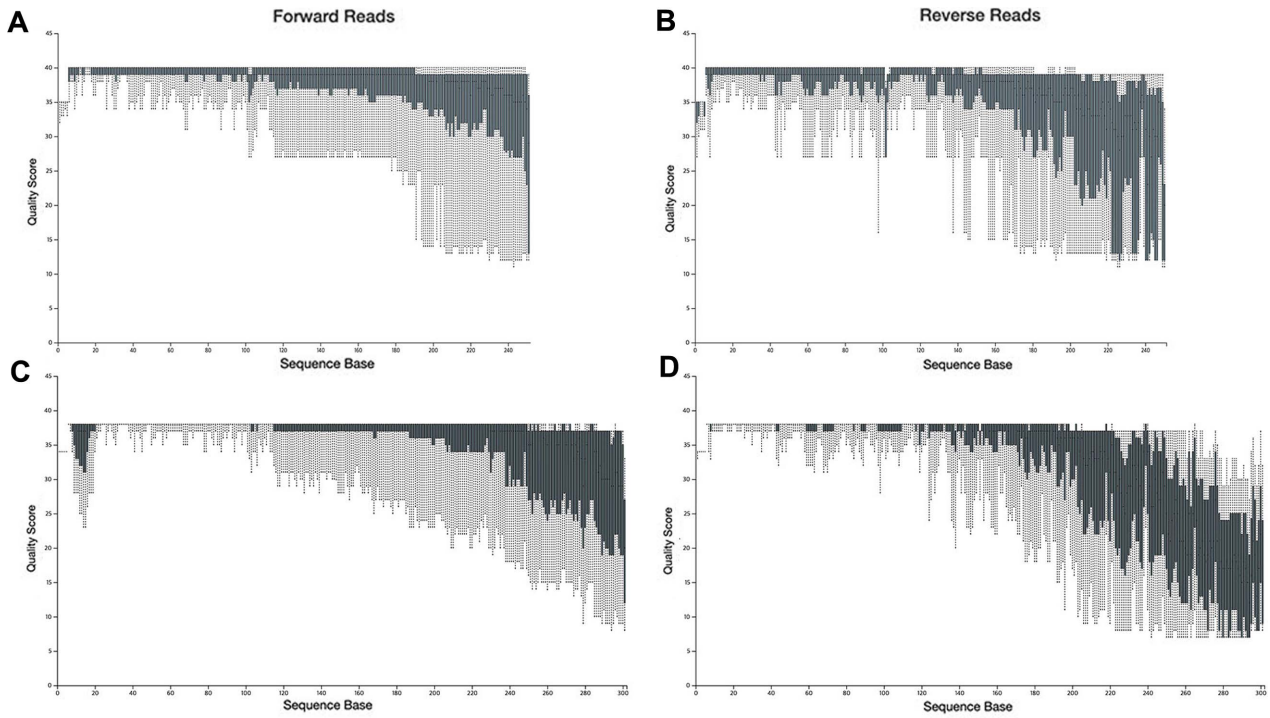

Fig. 1. Representative figures of read sequence quality comparisons between MiSeq and Hiseq. Forward sequence quality score from MiSeq (A) and HiSeq (C). Reverse sequence quality score from MiSeq (B) and HiSeq (D).

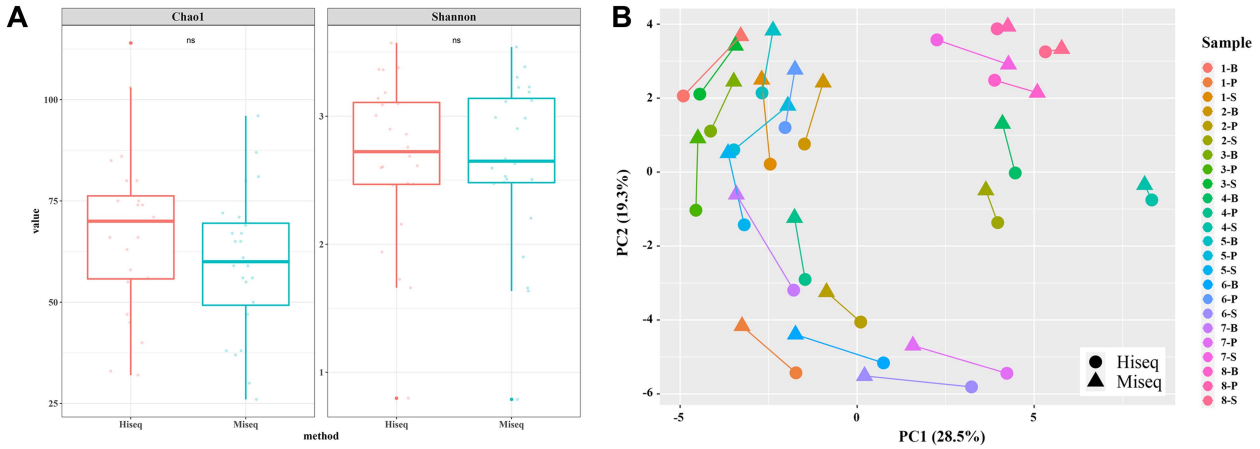

Fig. 2. Bacterial community diversity comparison between MiSeq and HiSeq. (A) Alpha diversity. Alpha diversity was used to describe the microbial richness, evenness and diversity within samples using the Chao1 and Shannon index. (B) Beta diversity of each sample connected with line. Principal coordinate analysis (PCoA) of the Bray-Curtis distance was performed to determine the microbial community structure.

quality score below 30 on a window of 20 bases were considered for trimming site. To compare between platforms, same trimming site was applied with optimum trimming size. Both 8 bases were trimmed from the start in forward and reverse reads. Forward reads were trimmed at 248 and reverse reads were trimmed at 240 bases.

Taxonomical assignment and diversity

Each refined sequencing read was taxonomically assigned by aligning it to sequences in the HOMD database. A total of 8 phyla, 43 genera, and 298 species were detected. The number of species-level operational taxonomic units (OTUs) observed in the HiSeq and MiSeq samples were 279 and 245, respectively. The average number of species in HiSeq and MiSeq samples was 67.42 ± 19.94 and 59.67 ± 17.74, respectively (Table S2). To compare overall outcome, alpha and beta diversity was analyzed. In case of alpha diversity, the Chao1 and Shannon index were similar between MiSeq and HiSeq samples (Fig. 2A). To define bacterial community patterns between MiSeq and HiSeq samples, beta-diversity of the corresponding samples were performed. In the Bray Curtis-based principal coordinates analysis (PCoA), most of the samples were closely positioned suggesting that each community share similar composition (Fig. 2B).

Distribution of species counts depending on relative abundance

Since there could be difference in taxonomy assignment depending on microbial abundance, assigned taxa were clustered into three groups; highly abundant taxa (taxa with more than 2% of each sample), moderately abundant taxa (taxa between 2% and 0.2%) and rare taxa (taxa with less than 0.2% of each sample). When cumulative abundance was calculated, highly abundant taxa consisted 79.2% of the total population while moderately abundant taxa consisted 18.7% and rare taxa consisted only 2.0% of total population in both HiSeq and MiSeq platforms (Fig. 3A). Total taxa count of highly abundant taxa was 244 and all the taxa were assigned in both platforms. In moderately abundant taxa, total taxa count was 622 and 38 taxa was only assigned in HiSeq analysis. In rare taxa, total taxa count was 944 and 413 taxa was assigned by both platforms (43.7%). Interestingly, 192 taxa was uniquely assigned by MiSeq while 340 taxa was assigned by HiSeq analysis (Figs. 3B and 3C). Taken together, in highly abundant taxa, all the assigned taxa showed similar assignment while in rare taxa, there were many taxa that was only assigned by either HiSeq or MiSeq.

Species abundance comparison depending on relative abundance

To evaluate the abundance ratio correlation between MiSeq and HiSeq, scatter plot and bar chart was plotted at the species level depending on their relative abundance. In highly abundant taxa, correlation coefficient was 0.994 ($p < 0.001$) suggesting high correlation

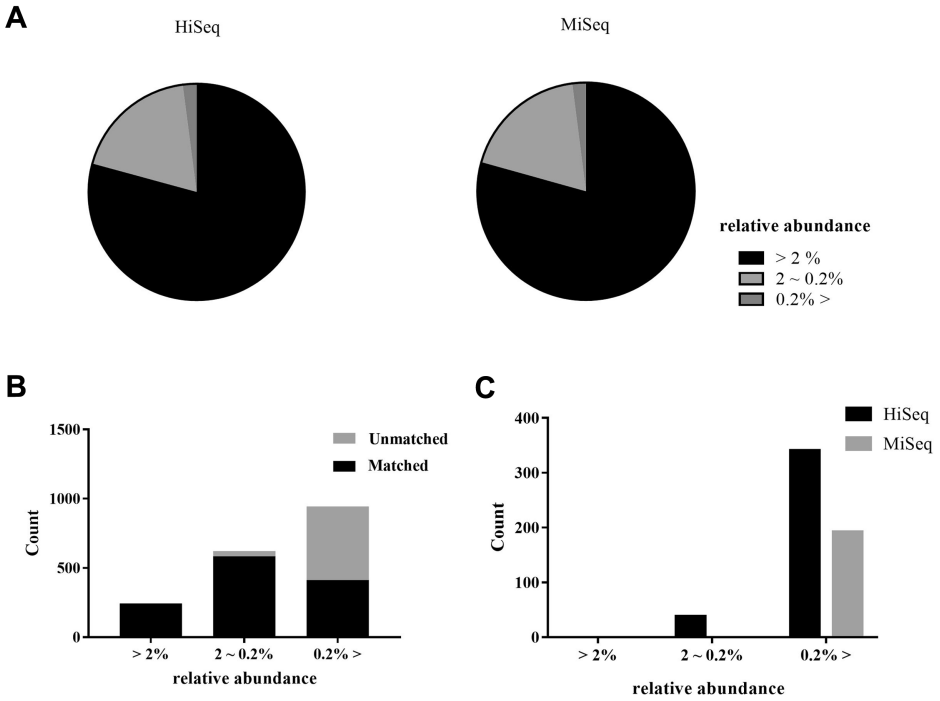

Fig. 3. Overview of species count distribution depending on relative abundance. (A) Proportion of read count included in each group. (B) Number of species counts with matched species samples or with unmatched ones. (C) Number of unmatched species counts in HiSeq and MiSeq.

between HiSeq and MiSeq plat form (Fig. 4A). Also, when 20 top most abundant species were plotted, the relative abundance also showed similar abundance (Fig. 4B). In consistent with distribution overview, when correlation among taxa with low relative abundance were plotted, correlation coefficient was lower as the abundance decreased (0.860 and 0.416 for moderate and low abundant taxa, respectively) (Figs. 5A and 5C). However, when unmatched taxa were removed, correlation coefficient was over 0.6 which suggest there was high correlation between the platforms in the matched taxa (Figs. 5B and 5D). Taken together, in abundant taxa, there was no difference in their relative abundance between platforms while in low abundant taxa, the cor-

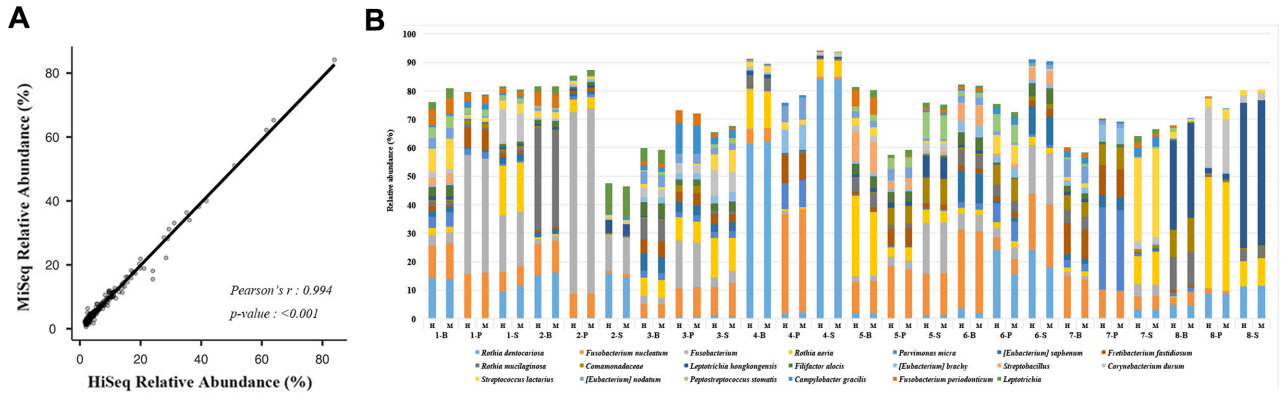

Fig. 4. Correlation between HiSeq and MiSeq in highly abundant species. (A) Correlation plot of species abundance more than 2%. (B) Comparison of top 20 abundant species between MiSeq and HiSeq in each sample.

http://dx.doi.org/10.48022/mbl.2008.08003

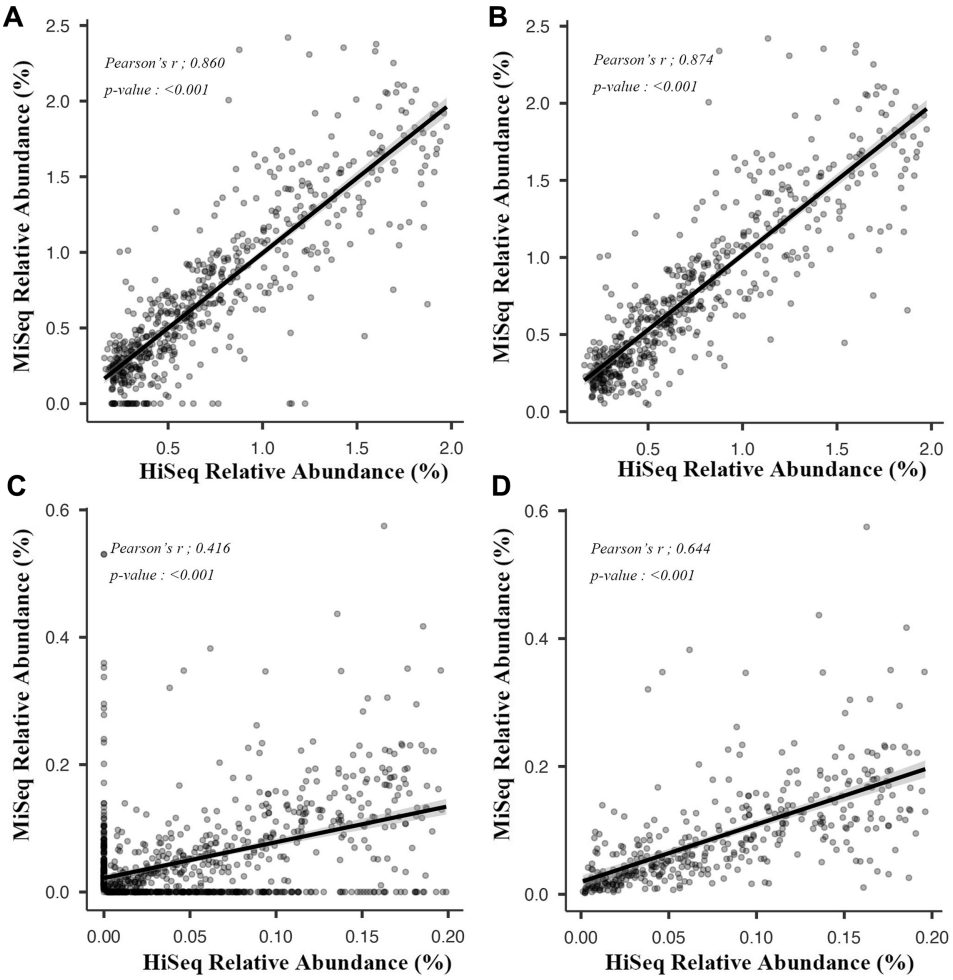

Fig. 5. Correlation between HiSeq and MiSeq in low abundant to rare species. (A) Total correlation plot of species abundance between 2% and 0.2%. (B) Correlation plot of species abundance between 2% and 0.2% after removing unmatched samples. (C) Total correlation plot of species abundance less than 0.2%. (D) Correlation plot of species abundance less than 0.2% after removing unmatched samples.

relation was not as high as the abundant taxa.

Discussion

There have been numerous studies reporting changes of the human microbiome linked to various human diseases [11, 19]. For microbiome study, MiSeq and HiSeq from Illumina are among the most widely used to study microbial community. The HiSeq and MiSeq platforms differ markedly in scale. In this study, we compared the oral microbiome in periodontitis patients using MiSeq and HiSeq platform to determine their suitability for large-scale surveys of oral microbial communities.

Since the number of samples loaded in each platform was not equal in every experiment, direct comparison for read count between HiSeq and MiSeq may not be appropriate. However, practical read count was compared as a prospective to overview the outcomes of each platform. In practice, number of samples recommended to be loaded for HiSeq and MiSeq is around 90 and 380. Considering that the HiSeq2500 produces 62 Gb produces 120 million 250-base paired-end reads and the MiSeq generates 12 Gb from 20 million 300-base paired-end reads, the theoretical average read count from each run for a sample is around 300,000 and 200,000 reads for HiSeq and MiSeq, respectively. In this study, we loaded around 260 and 80 samples for HiSeq and MiSeq analysis, respectively. After quality filter, the overall read pro-

duced by HiSeq and MiSeq was 278,736 ± 50,716 and 164,077 ± 25,697, respectively. Thus, HiSeq was loaded with much more samples simultaneously and produced significantly more read counts compared to MiSeq.

To evaluate and control sequencing data quality, a plot of a random sample was generated. Each forward and reverse plot represents the parametric seven-number summary of the quality scores at the corresponding position. A quality score is determined from the probability that a given sequenced base is wrong. QIIME uses the Phred score and user-defined parameters to remove sequence reads that do not meet the desired quality [20]. The forward reads produced high quality in both MiSeq and HiSeq sequencing. However, the reverse reads showed relatively lower sequencing quality at the distal end. The presence of low-quality bases towards the right end of the sequence adversely affects the joining step, leading to the failure of the joining, and consecutively to the loss of the reads in the middle of the analysis [21]. To reduce the consequences of this problem, it is recommended to trim the reads distal to a point where phred quality score drops below a specific threshold (quality trimming) [22]. With a majority of high-quality, sequences average quality score below 30 on a window of 20 bases were considered optimum for trimming size to retrieve only full-length sequences with low error rates, potentially increasing the discovery rate of rare OTUs. Thus, we trimmed at 248 base for forward reads and 240 base for reverse reads.

Each refined sequencing read was taxonomically assigned by aligning it to sequences in the HOMD database [12]. A total of 8 phyla, 43 genera, and 298 species were detected. The average number of species determined and number of unique taxa assigned was higher in HiSeq compared to MiSeq, suggesting HiSeq may have advantage over MiSeq to identify taxa that are rare in abundance.

To compare microbial complexity, alpha and beta diversity was analyzed. Alpha and beta diversity was similar between MiSeq and HiSeq samples. In accordance with previous study [23], this suggests that microbial diversity is not likely to provide additional insight by increasing the sequencing depth.

Next, correlation of relative abundance assigned by each platform was analyzed. The preprocessed sequences are clustered into Operational Taxonomic Units (OTUs), which in traditional taxonomy represent groups of organisms defined by intrinsic phenotypic similarity that constitute candidate taxa [24]. The threshold level is traditionally set at 97% of sequence similarity [25] and each representative sequences were assigned to specific taxonomy. In the process of OTU clustering and taxonomic assignment, rare taxa likely to be more vulnerable to clustering and representative sequence picking. In highly abundant taxa, all the assigned taxa were assigned in both platforms. In moderately abundant taxa, 38 taxa was uniquely assigned by HiSeq analysis. In rare taxa, only 43.7% of the assigned taxa were assigned by both platforms and 192 taxa was only assigned by MiSeq while 340 taxa was only assigned by HiSeq analysis. This result are in accordance with previous report that deeper sequencing are suggested to be advantageous to identify taxa that are rare in microbial communities [26]. When correlation between HiSeq and MiSeq was compared depending on relative abundance, the correlation coefficient of highly abundant taxa was nearly 1 suggesting a perfect correlation between HiSeq and MiSeq analysis while correlation was weaker in moderately abundant and rare taxa. Since unmatched taxa influence the correlation outcome, we analyzed only matched taxa in rarely abundant taxa and found that correlation coefficient was over 0.6 which suggest that relative abundance was conserved between platforms within the matched rare taxa. Taken together, in highly abundant taxa, all the assigned taxa showed similar assignment between platforms suggesting sequencing depth or platform does not influence the result. On the other hand, in rare taxa, there were many taxa uniquely assigned by either platforms and HiSeq produced more unique taxa suggesting deeper sequencing may provide advantage in detecting rare taxa.

Conclusion

As more microbiome studies are designed, a relevant question arises whether to obtain deeper coverage of samples or to increase the number of samples that are sequenced. In this study, the data generated by HiSeq and MiSeq was compatible but differed mostly in scale. Although MiSeq is most commonly used for microbiome study, HiSeq 2500 was also compatible for microbiome study. Moreover, HiSeq platform may allow massively

parallel sequencing with improved resolution for detecting rare taxa.

Acknowledgment

This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Science & ICT (NRF-2017M3A9B6062027).

Conflict of Interest

The authors have no financial conflicts of interest to declare.

References

1. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007. The human microbiome project. *Nature* **449**: 804-810.
2. Jorth P, Turner KH, Gumus P, Nizam N, Buduneli N, Whiteley M. 2014. Metatranscriptomics of the human oral microbiome during health and disease. *mBio* **5**: e01012-01014.
3. Dong TS, Gupta A. 2019. Influence of early life, diet, and the environment on the microbiome. *Clin. Gastroenterol. Hepatol.* **17**: 231-242.
4. Somineni HK, Kugathasan S. 2019. The microbiome in patients with inflammatory diseases. *Clin. Gastroenterol. Hepatol.* **17**: 243-255.
5. Zackular JP, Baxter NT, Iverson KD, Sadler WD, Petrosino JF, Chen GY, *et al.* 2013. The gut microbiome modulates colon tumorigenesis. *mBio* **4**: e00692-00613.
6. Liu XX, Jiao B, Liao XX, Guo LN, Yuan ZH, Wang X, *et al.* 2019. Analysis of salivary microbiome in patients with Alzheimer's disease. *J. Alzheimers Dis.* **72**: 633-640.
7. Frank DN, Robertson CE, Hamm CM, Kpadeh Z, Zhang T, Chen H, *et al.* 2011. Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflamm. Bowel Dis.* **17**: 179-184.
8. Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JI. 2011. Human nutrition, the gut microbiome and the immune system. *Nature* **474**: 327-336.
9. Cani PD, Possemiers S, Van de Wiele T, Guiot Y, Everard A, Rottier O, *et al.* 2009. Changes in gut microbiota control inflammation in obese mice through a mechanism involving GLP-2-driven improvement of gut permeability. *Gut* **58**: 1091-1103.
10. Langdon A, Crook N, Dantas G. 2016. The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation. *Genome Med.* **8**: 39.
11. Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* **486**: 207-214.
12. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, *et al.* 2010. The human oral microbiome. *J. Bacteriol.* **192**: 5002-5017.
13. Abusleme L, Dupuy AK, Dutzan N, Silva N, Burleson JA, Strausbaugh LD, *et al.* 2013. The subgingival microbiome in health and periodontitis and its relationship with community biomass and inflammation. *ISME J.* **7**: 1016-1025.
14. Apatzidou D, Lappin DF, Hamilton G, Papadopoulos CA, Konstantinidis A, Riggio MP. 2017. Microbiome of peri-implantitis affected and healthy dental sites in patients with a history of chronic periodontitis. *Arch. Oral Biol.* **83**: 145-152.
15. Park OJ, Yi H, Jeon JH, Kang SS, Koo KT, Kum KY, *et al.* 2015. Pyrosequencing analysis of subgingival microbiota in distinct periodontal conditions. *J. Dent. Res.* **94**: 921-927.
16. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, *et al.* 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**: 1621-1624.
17. Junemann S, Prior K, Szczepanowski R, Harks I, Ehmke B, Goesmann A, *et al.* 2012. Bacterial community shift in treated periodontitis patients revealed by ion torrent 16S rRNA gene amplicon sequencing. *PLoS One* **7**: e41606.
18. Junemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, *et al.* 2013. Updating benchtop sequencing performance comparison. *Nat. Biotechnol.* **31**: 294-296.
19. NIH Human Microbiome Portfolio Analysis Team. 2019. A review of 10 years of human microbiome research activities at the US National Institutes of Health, Fiscal Years 2007-2016. *Microbiome* **7**: 31.
20. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, *et al.* 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* **10**: 57-59.
21. Navas-Molina JA, Peraltz-Sanchez JM, Gonzalez A, McMurdie PJ, Vazquez-Baeza Y, Xu Z, *et al.* 2013. Advancing our understanding of the human microbiome using QIIME. *Methods Enzymol.* **531**: 371-444.
22. Kwon S, Park S, Lee B, Yoon S. 2013. In-depth analysis of interrelation between quality scores and real errors in illumina reads. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference.* 2013.
23. Kuczynski J, Liu Z, Lozupone C, McDonald D, Fierer N, Knight R. 2010. Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. *Nat. Methods* **7**: 813-819.
24. Sneath PH, Sokal RR. 1962. Numerical Taxonomy. *Nature* **193**: 855-860.
25. Drancourt M, Collet C, Carlioz C, Martelin R, Gayral JP, Raoult D. 2000. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. *J. Clin. Microbiol.* **38**: 3623-3630.
26. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, *et al.* 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**: 1621-1624.