The bias generated by sequencing platforms in analyzing of maternal-neonate gut microbiota profiles and diversity

Jie Wei
Xi'an Jiaotong University
Yue Cheng
Xi'an Jiaotong University
Xin Cao
Xi'an Jiaotong University
Xiaohong Tuo
Xi'an Jiaotong University
Kaichong Jiang
Xi'an Jiaotong University
Jia Lv
Xi'an Jiaotong University
Adediji Omolade Monisayo
Xi'an Jiaotong University
Bei Han (hanbei@mail.xjtu.edu.cn)
Xi'an Jiaotong University https://orcid.org/0000-0002-7841-5543

Research article

Keywords: 16S rDNA sequencing, high-throughput platforms, microbiota

Posted Date: December 19th, 2019

DOI: https://doi.org/10.21203/rs.2.19254/v1

Abstract

Background: Microbiome is an important internal ecosystem closely related to host health. Most of the bacteria existed in the internal ecosystem cannot be isolated with laboratory bacteriological culture methods, while 16S rDNA sequencing is considered and used for the bacterial identification by through the high-throughput platforms. The aim of this study was to compare the microbiota analysis result using two next-generation sequencing platforms and bioinformatics pipelines.

Results: 56 maternal-neonate fecal samples were sequenced and analyzed by 16S rRNA amplicon sequencing both by Ion Torrent S5-xl and Illumina Hiseq 2500 with standard protocols at same lab. For the richness and diversity of microbiota, index of chao1, observed_specise, PD_whole_tree, simpson and good_converage varied significantly except Shannon index at two platforms ($P<0.05$). The relative abundance of bacteria at different taxonomy levels is checked from phylum to species level, the more species of bacteria sequenced and annotated, the lower the correlation of the relative abundance of the bacteria founded between two platforms. The sequencing results are consistent between two platforms. Principal component analysis (PCA) results showed that more than 87% of samples were concentrated. According to principal coordinate analysis (PCoA), 56 samples of the two platforms were divided into two clusters, and the compliance rate of the two platforms is 71.43%. The differences between microbial community structures generated from two platforms were tested by multi-response permutation procedure (MRPP), which showed significant differences at family and genus levels separately ($A=0.094$, $P=0.001$; $A=0.085$, $P=0.002$). When maternal and neonate samples were considered, at family level, there was no difference in microbiota composition between two platform for maternal group ($A=0.006$, $P=0.149$), while in the neonate group, it showed significant differences ($A=0.035$, $P=0.006$). At the genus level, there existed significant differences in microbiota both in maternal and neonate group ($A=0.0216$, $P=0.004$; $A=0.098$, $P=0.001$).

Conclusion: Although the relative abundance of microbiota sequenced at two different platforms is basically similar, the diversity and correlation coefficient are still quite different. To increase reproducibility and reliability in cohort studies, it is important to use the same sequencing platforms and the corresponding pipeline to reduce the systematic error in microbiome analysis.

Background

Microorganisms are an important part of the earth's biology, and the number of cells exceeds the total number of animal and plant cells by two to three orders of magnitude. Many microorganisms inhabit the human body at least ten times more than the somatic and germline cells they contain. The host gut and its microbiota have a co-evolving complex relationship. The mammalian gastrointestinal tract is colonized by 10^{100} trillion microbes, and the diversity of gut microbiota maintains microbial balance. The stable composition and integrity of the species are essential, and these microorganisms are necessary for the host health, where more and more researches are focused on the human intestinal microbiota [1, 2]. However, only a small part of microorganisms could be cultivated by classical bacteriological culture, and

most of the bacterial forms observed in nature cannot be isolated and cultured with laboratory methods. At present, many scientists have carried out their research by metagenomic analysis technology where the microorganisms are identified by molecular and genomic fingerprint [3, 5, 6].

The 16S rRNA subunit is considered as the universal target for bacterial identification from DNA with the aid of sequencing. Among current research methods, amplicon sequencing is of great importance. For the bacteria species classification, the most typical of 16S rDNA molecules, using universal primers to design PCR primers for high-throughput sequencing and variable region sequences, is the best target molecule for studying bacterial communities [7.8]. 16S rDNA is located on the ribosomal small subunit of prokaryotic cells, including 10 conserved regions and 9 hypervariable regions. The conserved regions are not significantly different among bacteria. The hypervariable regions have genus or species specificity, and there exist differences with kinship, which makes it the most suitable indicator for bacterial phylogeny and taxonomic identification [9]. Usually 1~2 hypervariable regions are selected, the universal primers are used for PCR, and the sequences will be analyzed based on the high-variation regions. Compared to traditional amplicon sequencing, there is no need for isolation, culture, accuracy, rapidity, and sensitivity [10.11].

There are many platforms for second-generation sequencing, and the Roche 454 genome sequencing platform is based on the principle of pyrophosphate release. The Roche 454 was produced and commercialized in 2005 and was advertised as the first available high-throughput sequencing system. But Roche had closed 454 in mid-2016 and stopped supporting this platform [12]. Illumina sequencing platform first appeared in 2006 and was subsequently acquired by Illumina in 2007. Illumina has a range of the most commonly used sequencing systems and has been rapidly adopted by many researchers around the world. Illumina offers the highest throughput and lowest cost per base for all platforms, with read lengths up to 300bp and is compatible with almost any type of application [13]. However, its sample loading is technically challenging [14]. Life Technology introduced the SOLiD platform in 2007, and this system uses a unique chemical method for sequencing by attaching an oligonucleotide adaptor to the DNA fragment and immobilizing the ligation product on the beads, which are then placed in a water-oil emulsion [15]. Ion Torrent launched the Personal Genome Machine (PGM) in 2010 as a cost-effective platform for DNA sequencing. Unlike other sequencing technologies, ion flooding does not utilize light signals, but instead uses an enzymatic cascade to generate signals [16].

Different sequencing platforms had their advantages and disadvantages, and the sequencing platforms are upgrading continually. Thus, the comparison of microbiota sequencing data generated from different platforms in the cohort study is possible, also directly download those kind of biological sequence data from different online database, which is the enhancing of reproducibility for research community. Some recent studies have attempted to identify errors or bias generated by the characteristics of different platforms, and indicated the technical protocols and sequencing platforms have variable impact on output results [11, 21].

In this study, we reported the sequenced samples with known microbiota difference (maternal and neonate), to investigate comparison bias existed in different sequencing platforms of amplicon sequencing of 16S rDNA hypervariable regions represented the microbiota, 56 gut microbiota samples (from 28 pairs of women and neonates) were sequenced on both the Ion S5-xl platform and the Hiseq 2500 platform accordingly, then the raw reads were assembled, quality controlled, OTU cluster and species annotated under the same procedure, finally the microbiota structures of the same samples sequenced at different platforms were analyzed and compared.

Results

Sequencing quality assessment and comparison

By the 16s rDNA V3-V4 region sequencing of gut microbiota from 28 mother-neonate samples (28 late pregnant women and their infants) sequenced at Ion S5-xl platform and Hiseq 2500 platform at the same time, there had a total of 4, 231, 149 and 4, 325, 273 raw reads obtained separately, after passing quality control, a total number of 4, 030, 479 and 2, 894, 467 reads got, finally identified a total of 2532 and 4222 OTUs respectively (Table 1).

The dilution curve can intuitively reflect the depth of coverage, as the depth of sequencing increase, the curve of the sequencing data is more reliable when the curve tends to be flat [11, 12]. As shown in Figure S1, dilution curves of the 56 samples tend to be flat at the two sequencing platforms, which indicated the reasonable sequencing depth and the OTU coverage.

Gut microbiota composition analyzed by data generated from Ion and Hiseq platform

Although there defined 2532 OTUs in data generated from Ion S5-xl platform, which is higher than that of Hiseq 2500 platform (4222 OTUs), there had more species annotated at difference taxonomy level in Ion S5-xl platform than that from Hiseq2500 platform within the samples. There had a total of 38 phylum, 54 classes, 130 orders, 248 families and 587 genus annotated in the OTUs data generated from Ion platform, while from Hiseq 2500 platform, the corresponding taxonomy are 39 phylums, 62 classes, 137 orders, 252 families and 386 genus (Table S1) .

Alpha diversity analysis

Alpha diversity analysis is used to analyze the richness and diversity of microbial communities. For the comparison of 56 samples sequenced at 2 platforms, the value of chao1 (544.74 vs 444.15, P=0.001), observed_specise (498.16 vs 392.52, P<0.001), and PD_whole_tree (95.20 vs 40.49, P<0.001) calculated by sequence results obtained from Hiseq platform are higher, Shannon (4.74 vs 4.92, P=0.726), simpson (0.84 vs 0.88, P=0.003) index are lower than those obtained from the Ion platform (FigureS2). When the sample source of population considered, for the healthy adult women, except Shannon and simpson index, the other alpha diversity index of chao1, observed_specise, PD_whole_tree and good_converage were all significant different between two sequencing platforms (P<0.01); for the neonates, only index of

simpson and PD_whole_tree showed significant difference between two sequencing platforms (P<0.01) (Table 2). If only Shannon index was considered, the gut microbiota of the same samples showed no statistical difference when sequenced at two different platforms.

Microbiota composition revealed by different sequencing platforms

Abundance and diversity of bacterial communities were showed with analyses of Beta diversity. For data generated from Ion platform, the top 4 phylum are Firmicute (40.20%), Bacteroidetes (26.99%), Proteobacteria (25.61%), Actinobacteria (6.15%), compared with Firmicute(39.72%), Proteobacteria(30.23%), Bacteroidetes(16.62%), Actinobacteria(11.10%) in Hiseq platform. There are differences in the relative abundance of the level of the phylum, such as Bacteroidetes (P=0.001) and Actinobacteria (P=0.003). At the class level, there had statistical difference in Bacteroidia (Ion 26.97% vs Hiseq 16.57%, P=0.001), unidentified_Actinobacteria (Ion 5.8% vs Hiseq 9.78%, P=0.008), which belong to the top 4 classes. For the top 5 orders, only the relative abundance of Bacteroidia showed statistical difference between two platforms (Ion 29.39% vs Hiseq 16.32%, P=0.002). For the top 5 families, only the relative abundance of Prevotellaceae showed statistical difference between two platforms (Ion 6.37% vs Hiseq 0.32%, P<0.001) (Fig 1).

The relative abundance of bacteria at different taxonomy levels of the two sequencing platforms is related. The correlation of bacteria with a relative abundance of more than 1% at phylum level were analyzed between two sequencing platforms (Fig 2). The Pearson correlation coefficient (r) of all bacteria at phylum and class level, 82% of bacteria at order level, 61% of bacteria at and family level is more than 0.5, and the correlation trends is good (Table 3). However, as the taxonomy levels is checked from phylum to species, the more species of bacteria sequenced and annotated, the lower the correlation of the relative abundance of the bacteria founded.

Microbiota diversity impacted by sequencing platforms

In order to evaluate differences of species complexity of the same 28 maternal-neonate-pair samples, and verify the microbiota diversity impacted by sequencing platforms, the PCA was performed and analyzed. It is found that the scatter points of two platforms are relatively close. It can be shown that the sequencing results of repeatability and accuracy are consistent at two platforms. However, in explaining the efficiency of variation, microbiota results at the family level (PCA1=14.47%; PCA2=8.77%) (Fig 3A) was better than that of genus level (PCA=9.86%; PCA= 7.07%) (Fig 3B).

To further improve the ability to explain variation, we stratified the gut microbiota of 56 samples into maternal and neonate groups. For the maternal group, the comparison of PCA between family and genus level yielded similar results, where the diversity were less than that of the overall group. The stability of results from the Ion platform is better than that of Hiseq platform. Compared with the neonate group, the PCA of mother group can explain the most variation (PCA1=30.39%; PCA2=11.68%) (Fig3C, 3D). For the gut microbiota hierarchical analysis of PCA results in neonate group, as the taxonomy levels is higher, the

variation of PCA value is higher, which means that intra-group variation of neonate group is larger than that of maternal group (Fig3E, 3F).

MRPP inter-group difference analysis was used to analyze whether the differences in microbial community structure differences between groups were significant. By MRPP analysis, there was a significant difference in microbial community structure between the two platforms both in family level and genus level (A=0.094, *P*=0.001; A=0.085, *P*=0.002). Then the 56 samples were divided into maternal and neonate groups, for the maternal microbiota composition, it showed no difference at family level (A=0.006, *P*=0.149) and significant difference at genus level (A=0.0216, P=0.004); while for the neonate microbiota composition, the differences were all significant both at family level and genus level (A=0.035, *P*=0.006; A=0.098, *P*=0.001).

We chose the primary coordinate analysis based on the Unweighted Unifrace distance and selected the primary coordinate combination with the highest contribution rate to the sample. The PCoA abscissa represents a principal component, the ordinate represents another principal component, and each dot in the figure represents a sample. At the genus level, we identified the enterotypes of the 56 maternal-neonate samples, where the optimal number of clusters was different (Fig4). For PCoA results from Ion platform (Fig4A), there had 25 samples gathered and automatically grouped into cluster 1 and the rest 31 samples in cluster 2; while for the Hiseq platform (Fig4C), there had 33 samples in cluster 1 and the rest 23 samples in cluster 2. When samples in the same cluster were compared, 19 identical samples sequenced in 2 platforms were included in cluster 1(17 are of neonate samples), for cluster 2, there had 21 identical samples included (20 are of maternal samples) (Table S2). There had a coincidence rate of 71.43% (40/56) in the PCoA grouping for the two platforms.

Discussion

The Ion S5-xl platform and the Hiseq 2500 platform are two commercially available sequencing platforms that target both clinical applications and laboratories [17]. There are many 16s sequencing platforms for microbiota analysis. The direct comparison and analysis of the data generated by different platforms may have certain errors. Therefore, the purpose of this study is to explore which aspects of the comparison will be used when comparing data from different platforms.

When the sample collection and treating were controlled to eliminate any potential upstream variations, then using the same computational pipeline and 16S reference database to annotate the sequenced results to control the downstream variations in analysis. And the possible variability will focused on the sequencing platform used.

In our study, for low-quality sequence removal rate, it seems that Ion S5-xl platform showed better than Hiseq 2500 platform. For the microbiota analysis, alpha diversity index is usually used to show the diversity within individual, where Chao1, Shannon, simpson, observed_specise, PD_whole_tree, good_coverage are the general index included [18, 19, 20].There existed significant statistical difference in the alpha diversity index in this study between the two sequencing platforms (*P*<0.01), except the

Shannon index, which is to evaluate the diversity of the bacterial community. As expected, microbial species sequenced by two platforms are basically the same evaluated by Shannon index [21]. While other index difference indicated that the number of species from the same sample sequenced by two platforms was quite different. Therefore, when evaluating the microbiota by alpha diversity, it is not comprehensive and precise to consider only one or two index, which choice is likely to lead to analysis bias. For example, Chao1 is used to assess the richness of bacterial communities [22]. When OTU composition and richness are considered to reflect the community structure of the microbiota, it may not accurately reflect the real microbiota, because of species corresponding to one or more OTUs. As we know, the relationship between species interactions in one community is very complex, such as predation, symbiosis, partial symbiosis, and microbial quorum sensing, so alphy diversity index cannot be accurate and comprehensive in reflecting community diversity. And the large fluctuation of alphy diversity index does not represent the actual results well.

The calculation of beta diversity is an indicator to characterize the similarity of microbiata composition among individuals. The presence and inconsistency of species between individuals usually affects the beta diversity index, and the beta diversity is affected by the alpha diversity. In our data generated by two platforms, the category of bacteria with relative abundance more than 1% in different taxonomy is consistent with high similarity, but the relative abundance of each bacteria was not exactly the same, even showed significant difference. The species with a relative abundance of more than 1% in 56 fecal samples were Firmicute, Bacteroidetes, Proteobacteria and Actinobacteria, but the relative abundance of those four bacteria were not the same. For example, the relative abundance of Bacteroidetes (Ion 26.99% vs Hiseq 16.62%, P=0.001), Actinobacteria (Ion 6.15% vs Hiseq 11.10%, P=0.003) showed significant difference, which may lead to wrong conclusion to some extent, and we speculate this might be due to the depth of different sequencing platforms. Especially in the comprehensive Enterotype analysis based on 16S rRNA gene amplicon data by the relative abundance of those four major phyla [23], the sequence data generated platforms should be pay more attention. With the further comparisons, such as correlation analysis, PCA and PCoA analysis, it was shown that there existed significant statistical differences between the two platforms.

Whether the alphy diversity or the relative abundance of the sample population is not exactly the same, and regardless of the sequencing platform, PCR-based high-throughput amplicon sequencing will bias the results, so using the same sequencing platform to reduce errors in interpretation will be better [24]. Especially in cohort studies, the acquired biological samples were collected at different times, so to maintain the activity of the biological samples, sequencing was performed within a certain period of time, which resulted in different batches of samples not being carried out on the same machine. Furthmore, it is suggested that when multi-factor analysis of microbiota sequencing data, the sequencing platform should be considered as much as possible, or when doing other analysis, the microbial data cannot be simply downloaded from the public database for comparison, which may led to inaccurate conclusion. So it is recommended to use the same platform and the proper internal standard for sequencing to reduce systematic bios and minimize error in interpretation.

Conclusion

By analysis the same fecal samples at two different PCR-based high-throughput amplification regardless of the sequencing platform with the same data process pipeline, there still have some difference which effect the richness and diversity of gut microbiota analysis. It is strongly suggested to use the same sequencing platform to reduce systematic error and bios, draw accurate biological conclusions.

Methods

Fecal samples collection and bacteria DNA extraction

28 mother-infant pairs of fecal samples (28 late pregnant women and their infants) were collected and used from a Chinese birth cohort to study the gut microbiota during infancy and its effect for child health. Fecal samples were collected and transported to the lab in the ice box within 4 hours. The study was approve by Ethics Committee of Health Science Center, Xi'an Jiaotong University (approval ref.2016-114). The approvals and the informed consent from the mothers were obtained prior to collection data and samples.

For the DNA extraction, about 0.2 g of the stool sample was weighed, and the bacterial genome DNA was extracted using the QIAamp Fast DNA Stool Mini Kit (QIAGEN, GER). The extracted DNA concentration and purity was detected on 1.0% agarose gel electrophoresis and Ultra-micro spectrophotometer (NanoDrop 2000, Thermo, US). And the DNA was diluted to 1ng/μl using sterile H_2O.

16S rDNA sequencing of the fecal microbiota

The V3-V4 region of the 16S rDNA sequence (465 bp) was amplified using a specific primer with barcode. The primer sequence were: 515F (3'ACT CCT ACG GGA GGC AGC AG5') and 806R (3'GGA CTA CHV GGG TWT CTA AT5'). PCR amplification was carried out in 50μl reaction mixtures containing 0.5 U of Phusion® High-Fidelity PCR Master Mix (New England Biolabs, UK) and 50 ng of DNA as template. PCR was performed on an automated thermocycler (Bio-Rad MyCycler, US) for 30 cycles consisting of 30 s at 94°C, 30 s at 50°C, and 30 s at 72°C; Amplicons were visualized on 1.5% agarose and checked for length; products with the desired size (approximately 460bp) were purified using a QIAquick gel extraction kit (QIAGEN, GER). DNA quality and amounts were assessed using an Invitrogen Qubit® dsDNA BR kit. The experiments were performed in triplicate. The purified fragments in each sample were normalized and pooled to generate the sequencing libraries. The library was sequenced on an Illumina HiSeq 2500 platform and Ion S5-xl platform separately in lab of Novogene Biotech Co., Ltd., Beijing, China.

Microiota data analysis

The raw reads were subject successively to data split, assembly and merged by FLASH V1.2.7, quality filtered by QIIME V1.7.0, chimera removed by UCHIME algorithm to get qualified reads. Then qualified reads were clustered to generate operational taxonomic units (OTUs) at the 97% similarity level using

USEARCH. The Silva132 was used to annotate taxonomic information. The alpha diversity is applied in analyzing complexity of species diversity for a sample by 6 indices (Observed-species, Chao1, Shaanon, Simpson, ACE and Good-coverage) and calculated by QIIME V1.7.0. Beta diversity analysis was used to evaluate differences of samples in species complexity, the Beta diversity on both weighted and unweighted unifric distance were also calculated by QIIME V1.7.0. Cluster analysis was preceded by principal component analysis (PCA) by using the FactoMineR package and ggplot2 package in R V2.15.3.

For the different sequencing platform, the structure of the OTU annotation cannot be combined. Therefore, we take the two levels of detectable species as the analysis unit, and perform PCA, PCoA analysis on 56 sequencing samples at these two levels. For maternal-neonate stratified reanalysis, ade4 and ggplot2 packages in R were used. Enterotype analysis based on the genus-level abundance profiles were performed using ade4 package in R [25].

The sequenced data were deposited into the Sequence Read Archive (SRA) of NCBI (http://www.ncbi.nlm.gih.gov/sra) and can be accessed via accession number PRJNA544346.

Statistical analysis

Statistical analysis was performed by SPSS 18.0. Paired t-test and Wilcoxon rank sum test were used to compare the alpha index and the relative abundance of the two groups. $P < 0.05$ indicates that the difference is statistically significant. Pearson correlation analysis was used to test the correlation of the relative abundance of the two groups of bacteria.

Declarations

Acknowledgements

The authors would like to acknowledge Prof. Lingxia Zeng of Xi'an Jiaotong University for her assistance in the fecal sample collection.

Funding

This work was financially supported by the National Natural Science Foundation of China (81673199), the Natural Science Foundation of Shaanxi province, China (2018JM7054), Fundamental Research Funds for the Central Universities (No. xjh012019054)

Availability of data and material

The datasets used in the current study are available from the corresponding author upon reasonable request.

Authors' contributions

JW and YC performed bioinformatics and statistical of data, JW, XC and AOM contributed to manuscript writing and editing. XHT and KCJ performed DNA isolation of samples. XC and JL contributed to sample collection. YC and BH designed the project, wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare they have no completing interests.

References

1. Kim YS, Kim J, Park SJ. High-throughput 16S rRNA gene sequencing reveals alterations of mouse intestinal microbiota after radiotherapy. Anaerobe. 2015; 33:1-7.
2. Sommer F, Anderson JM, Bharti R, et al. The resilience of the intestinal microbiota influences health and disease. Nat Rev Microbiol. 2017; 15(10):630-638.
3. Nikolaki S, Tsiamis G. Microbial diversity in the era of omic technologies. Biomed Res Int. 2013; 2013: 958719.
4. Woo PC, Lau SK, Teng JL, et al. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. Clin Microbiol Infect. 2008; 14(10):908-34.
5. Liu F, Yuan Z, Zhang X, et al. Characteristics and diversity of endophytic bacteria in moso bamboo (*Phyllostachys edulis*) based on 16S rDNA sequencing. Arch Microbiol. 2017; 199(9):1259-1266.
6. Loong SK, Khor CS, Jafar FL, et al. Utility of 16S rDNA sequencing for identification of rare pathogenic bacteria. J Clin Lab Anal. 2016; 30(6):1056-1060.
7. Tian Y, Li YH. Comparative analysis of bacteria associated with different mosses by 16S rRNA and 16S rDNA sequencing. J Basic Microbiol. 2017; 57(1):57-67.
8. Davidson RM, Epperson LE. Microbiome sequencing methods for studying human diseases. Methods Mol Biol. 2018; 1706:77-90.
9. Andreas H, Joachim G, Udo R, et al. Analyses of intestinal microbiota: culture versus sequencing. ILAR J. 2015; 56(2):228-40.
10. Christensen JJ, Andresen K, Justesen T, et al. Ribosomal DNA sequencing: experiences from use in the Danish National Reference Laboratory for identification of bacteria. APMIS. 2005; 113(9):621-8.

11. Hahn A, Sanyal A, Perez GF, et al. Different next generation sequencing platforms produce different microbial profiles and diversity in cystic fibrosis sputum. J Microbiol Methods. 2016; 130:95-99.
12. Malla MA, Dubey A, Kumar A, et al. Exploring the human microbiome: The potential future role of next-generation sequencing in disease diagnosis and treatment. Front Immunol. 2019; 9:2868.
13. Liu L, Li Y, Li S, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012; 2012:251364.
14. Van Dijk EL, Auger H, Jaszczyszyn Y, et al. Ten years of next-generation sequencing technology. Trends Genet. 2014; 30(9):418-26.
15. Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005; 437(7057):376-80.
16. Rothberg JM, Hinz W, Rearick TM, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011; 475(7356):348-52.
17. Shin S, Kim Y, Oh SC, et al. Validation and optimization of the Ion Torrent S5 XL sequencer and Oncomine workflow for BRCA1 and BRCA2 genetic testing. Oncotarget. 2017; 8(21):34858-34866.
18. Mao H, Lv Z, Sun H, et al. Improvement of biochar and bacterial powder addition on gaseous emission and acterial community in pig manure compost. Bioresour Technol. 2018; 258:195-202.
19. Ma J, Tang JY, Wang S, et al. Illumina sequencing of bacterial 16S rDNA and 16S rRNA reveals seasonal and species-specific variation in bacterial communities in four moss species. Appl Microbiol Biotechnol. 2017; 101(17):6739-6753.
20. Wang J, Chapman SJ, Ye Q, et al. Limited effect of planting transgenic rice on the soil microbiome studied by continuous $^{13}CO_2$ labeling combined with high-throughput sequencing. Appl Microbiol Biotechnol. 2019; 103(10): 4217-4227.
21. Allali I, Arnold JW, Roach J, et al. A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome. BMC Microbiol. 2017;17(1):194.
22. Chiu CH, Chao A. Estimating and comparing microbial diversity in the presence of sequencing errors. PeerJ. 2016; 4:e1634.
23. Besser J, Carleton HA, Gerner-Smidt P, et al. Next-generation sequencing technologies and their application to the study and control of bacterial infections. Clin Microbiol Infect .2018; 24(4):335-341.
24. Arumugam M, Raes J, Pelletier E, et al. Enterotypes of the human gut microbiome. Nature. 2011; 473(7346): 174-180.

Tables

Table 1. The sequencing description of 56 fecal samples at 2 sequencing platform

platform	OTUs	Raw reads	Clean reads	Average Length(nt)	Q20	GC%
Ion S5-xl	2537	4231149	4030479	419.19	84	52.77
Hiseq2500	4222	4325273	2894467	418.37	97	53.43

Table 2. Alpha diversity index comparison between two different sequencing platforms

Index	Sample group[#]	Ion S5-xl	Hiseq 2500	*P* value
chao1	M-N	444.15	544.74	*P*=0.001
	M	396.49	555.86	*P*=0.001
	N	491.81	532.21	*P*=0.29
shannon	M-N	4.92	4.74	*P*=0.726
	M	5.11	5.22	*P*=0.076
	N	4.73	4.25	*P*=0.072
simpson	M-N	0.88	0.84	*P*=0.003
	M	0.92	0.93	*P*=0.927
	N	0.85	0.76	*P*<0.001
observed_specise	M-N	395.52	498.16	*P*<0.001
	M	342	506.07	*P*<0.001
	N	449.03	490.25	*P*=0.272
PD_whole_tree	M-N	40.49	95.2	*P*<0.001
	M	34.03	78.42	*P*<0.001
	N	46.97	111.98	*P*<0.001
good_coverage	M-N	0.99	0.99	*P*=0.009
	M	0.99	0.99	*P*=0.003
	N	0.99	0.99	*P*=0.231

[#]M-N, Maternal-neonate group (56 samples); M, maternal group (28 samples); N, neonate group (28 samples)

Table 3 Pearson correlation analysis of the relative abundance of microbiota at the same taxonomy level sequenced in different platforms.

taxonomy level	microbiota	Relative abundance（%）		Pearson correlation coefficient	*P* value
		Ion S5-xl	Hiseq 2500		
Phylum	Firmicutes	40.2	39.72	0.83	*P*<0.001
	Proteobacteria	25.61	30.23	0.901	*P*<0.001
	Bacteroidetes	26.99	16.62	0.619	*P*<0.001
	Actinobacteria	6.15	11.1	0.785	*P*<0.001
	Verrucomicrobia	0.15	0.98	0.714	*P*<0.001
Class	Clostridia	29.39	25.99	0.848	*P*<0.001
	Bacteroidia	26.97	16.57	0.618	*P*<0.001
	Gammaproteobacteria	24.65	28.31	0.901	*P*<0.001
	Bacilli	6.95	9.78	0.708	*P*<0.001
	unidentified_Actinobacteria	5.81	9.78	0.762	*P*<0.001
	Negativicutes	2.8	2.6	0.643	*P*<0.001
	Erysipelotrichia	1.06	1.35	0.811	*P*<0.001
	Alphaproteobacteria	0.61	1.54	0.5	*P*<0.001
	Coriobacteriia	0.17	1.23	0.738	*P*<0.001
Order	Clostridiales	29.39	25.99	0.848	*P*<0.001
	Bacteroidales	26.6	16.32	0.608	*P*<0.001
	Enterobacteriales	22.01	24.86	0.879	*P*<0.001
	Lactobacillales	5.85	7.95	0.756	*P*<0.001
	Bifidobacteriales	5.19	7.99	0.74	*P*<0.001
	Selenomonadales	2.8	2.6	0.643	*P*<0.001
	unidentified_Gammaproteobacteria	1.68	1.74	0.678	*P*<0.001
	Bacillales	1.1	1.83	0.469	*P*<0.001
	Erysipelotrichales	1.06	1.35	0.811	*P*<0.001
	Coriobacteriales	0.17	1.23	0.737	*P*<0.001
	Pasteurellales	0.54	0.89	0.378	*P*=0.04
Family	Enterobacteriaceae	22.01	24.86	0.879	*P*<0.001
	Ruminococcaceae	16.14	13.37	0.823	*P*<0.001
	Bacteroidaceae	15.36	12.94	0.696	*P*<0.001
	Lachnospiraceae	9.21	10.22	0.829	*P*<0.001
	Prevotellaceae	6.37	0.35	0.356	*P*<0.001

Bifidobacteriaceae	5.19	7.99	0.74	*P*<0.001
Enterococcaceae	2.3	3.7	0.485	*P*<0.001
Veillonellaceae	2.17	2.51	0.713	*P*<0.001
Streptococcaceae	1.92	3.44	0.586	*P*<0.001
unidentified_Clostridiales	1.85	0.89	0.727	*P*<0.001
Rikenellaceae	1.74	1.27	0.694	*P*<0.001
Burkholderiaceae	1.59	1.35	0.226	*P*=0.093
Lactobacillaceae	1.57	0.68	0.112	*P*=0.410
Tannerellaceae	1.55	1.18	0.536	*P*<0.001
Erysipelotrichaceae	1.06	1.35	0.811	*P*<0.001
Staphylococcaceae	0.98	1.52	0.398	*P*=0.002
Peptostreptococcaceae	0.96	1.02	0.26	*P*=0.053
Coriobacteriaceae	0.12	1.08	-0.234	*P*=0.082

Figures

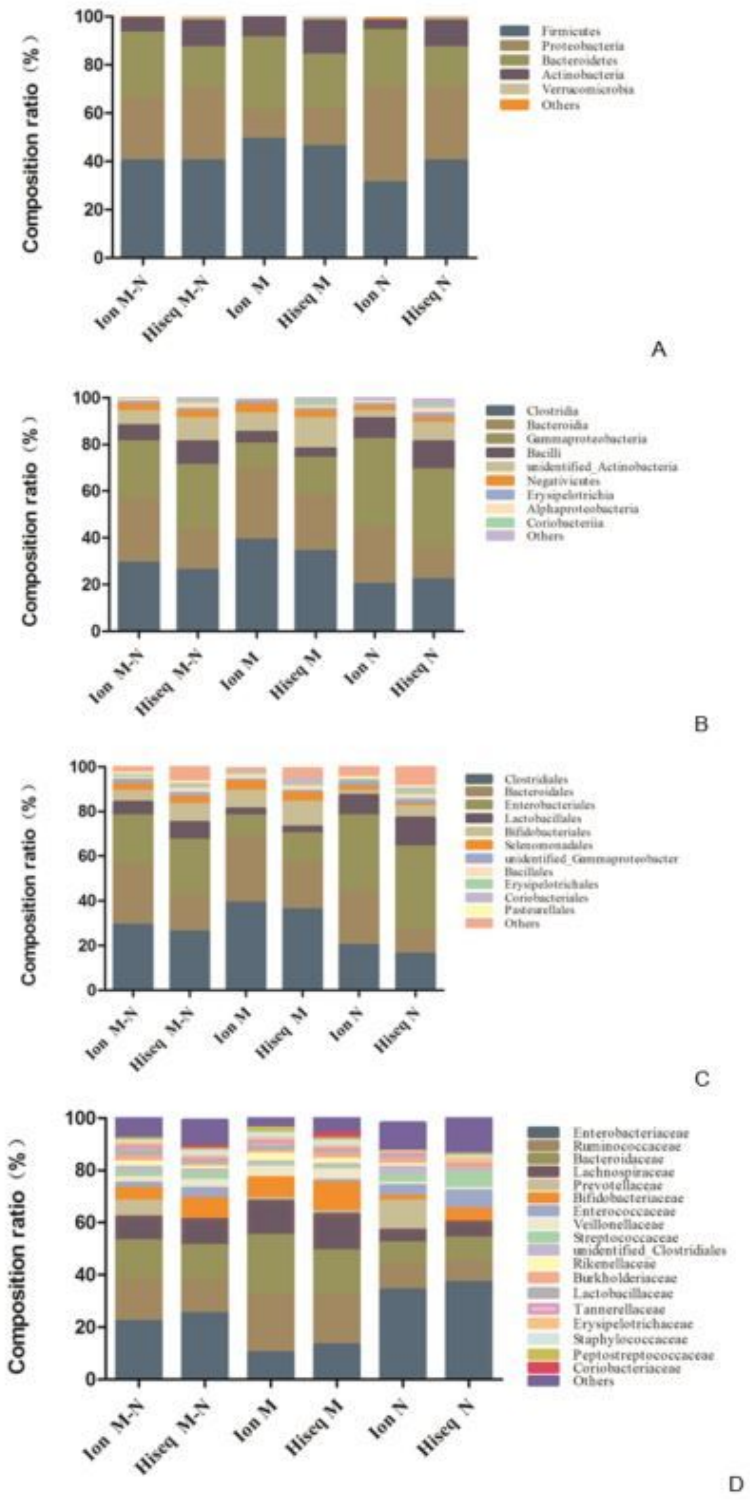

Figure 1

Microbiota distributions at phylum (A), class (B), order (C) and family (D) based on 16S rRNA V3-V4 region sequencing of 28 maternal-neonate pairs samples at 2 sequencing platforms. Ion, Ion S5-xl; Hiseq, Hiseq 2500; M-N, 56 maternal-neonate samples; M, 28 maternal samples; N, 28 neonate samples.

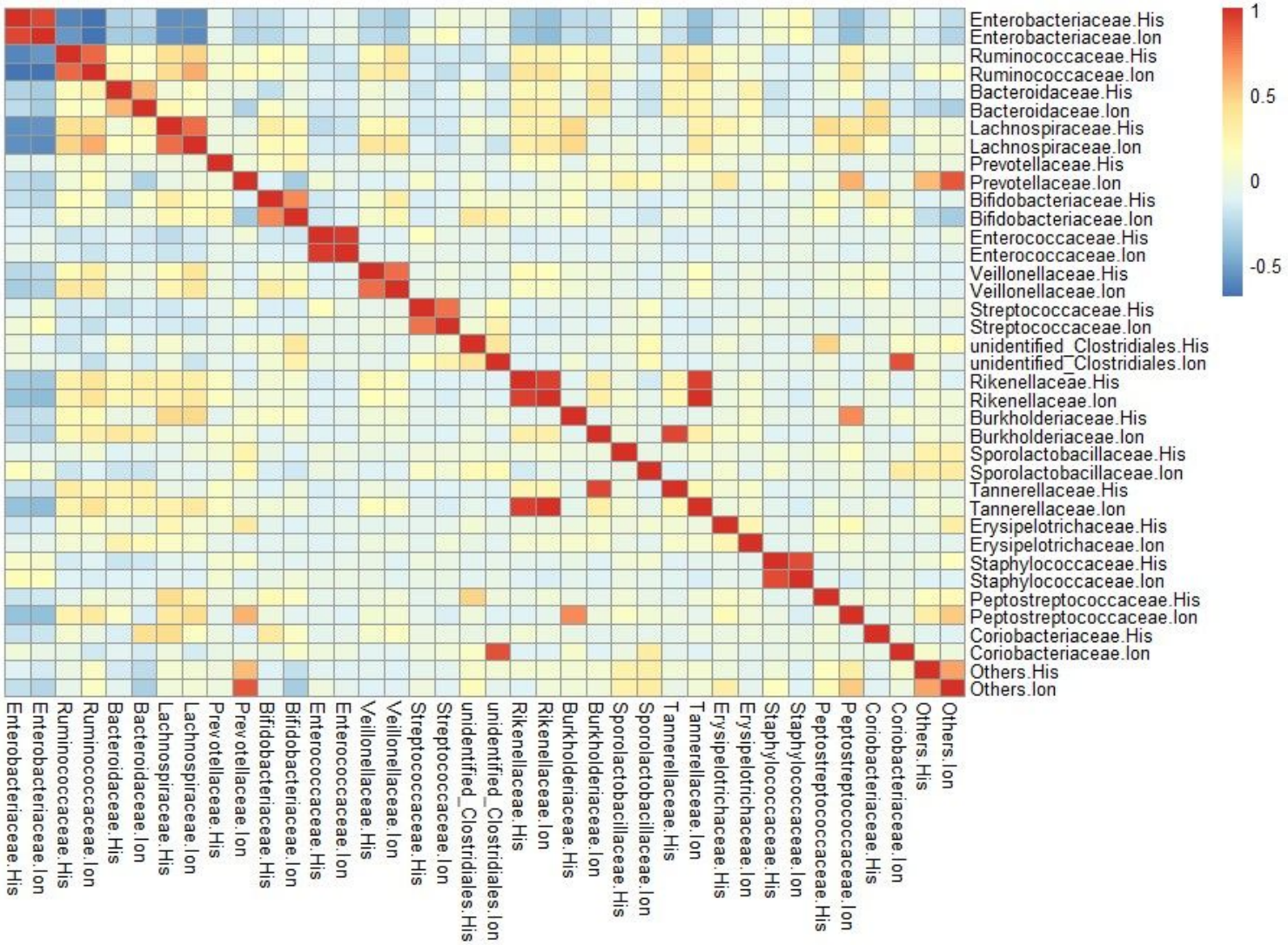

Figure 2

Co-occurrence patterns among the bacteria with a relative abundance of more than 1% at phylum level between sequencing platforms of Ion S5-xl (Ion) and Hiseq 2500(His), as determined by the Spearman's rank correlation analysis.

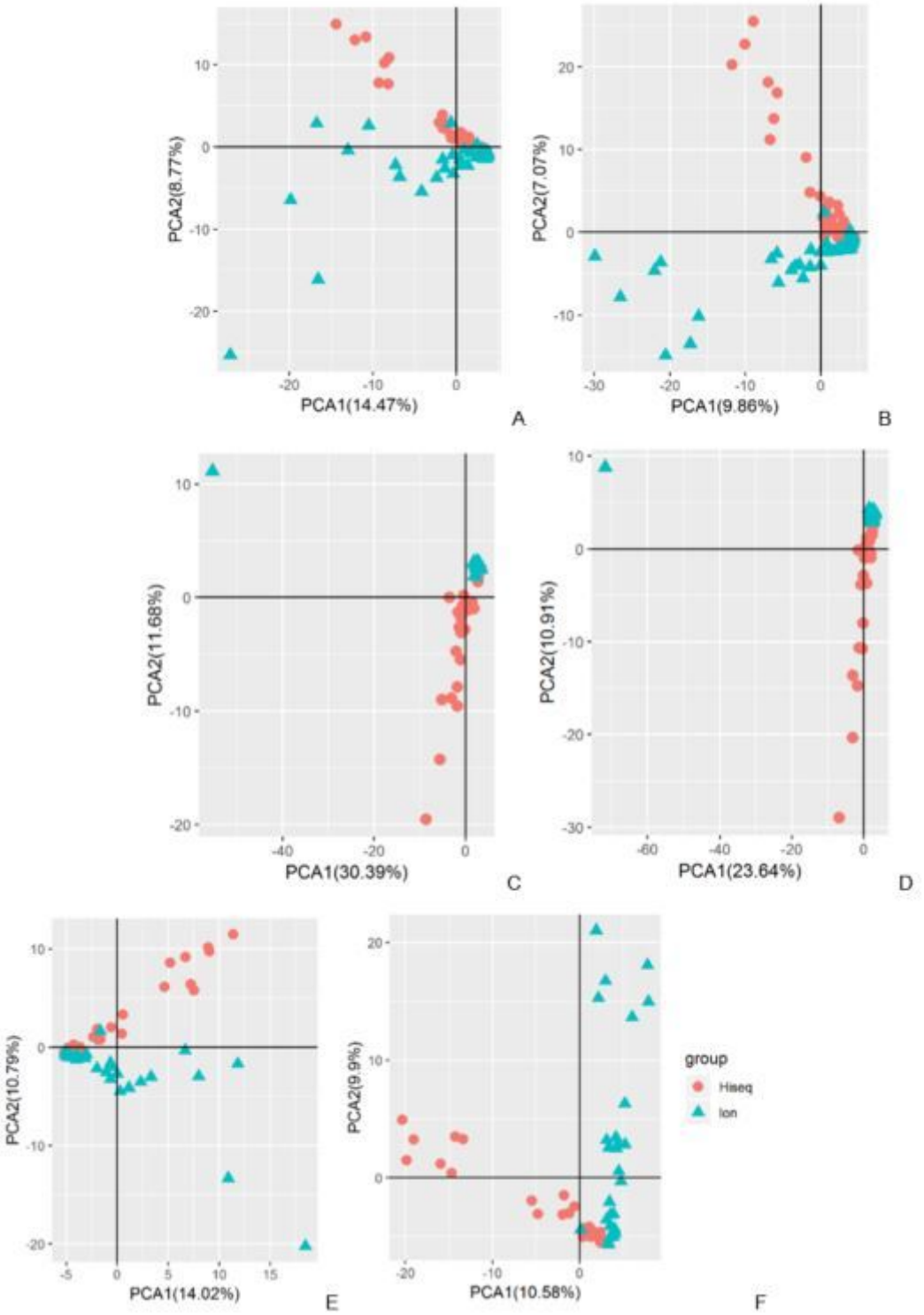

Figure 3

Beta-diversity based Principal Component Analysis (PCA) plots of gut microbiomes of 56 maternal-neonate samples (A, Family level; B, Genus level), 28 maternal samples (C, Family level; D, Genus level), 28 neonate samples (E, Family level; F, Genus level) at Ion(S5-xl) and Hiseq (2500) sequencing platforms, using unweighted unifrac distances.

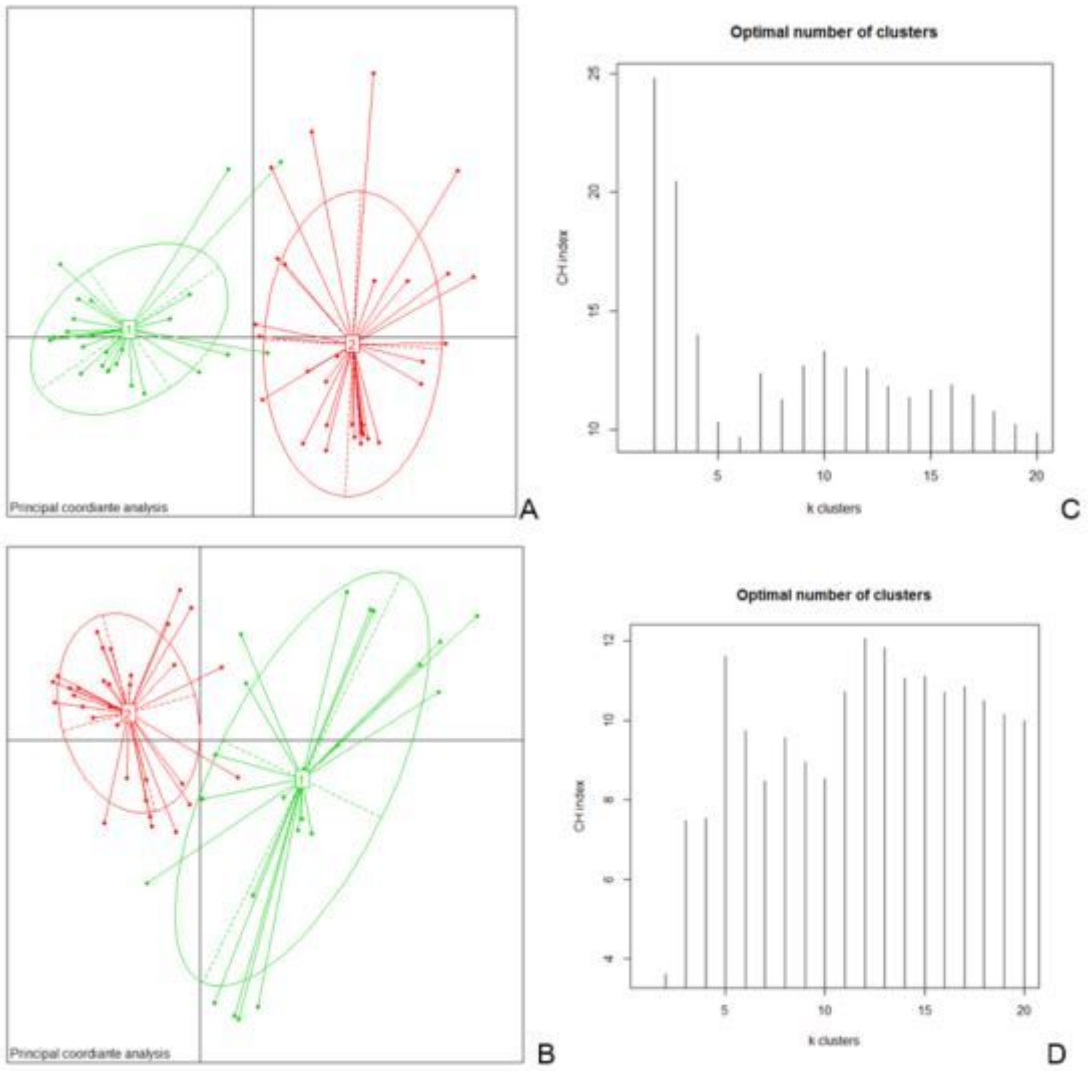

Figure 4

Microbiota clusters obtained from Principal Coordinate Analysis (PCoA) identified in 56 maternal-neonate samples gut microbiomes data sequenced at Ion S5-xl (A) and Hiseq 2500 (B) platforms, and the corresponding clustering assessment of 56 maternal-neonate samples gut microbiomes data sequenced at Ion S5-xl (C) and Hiseq 2500 (D) platforms. There had 19 identical samples sequenced in 2 platforms were included in cluster 1, and 21 identical samples included in cluster 2.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplmentarymaterial.docx