Microbiota Analysis Using Two-step PCR and Next-generation 16S rRNA Gene Sequencing

Shailesh K. Shahi[1], **Kasra Zarei**[2], **Natalya V. Guseva**[1], **Ashutosh K. Mangalam**[1,2,3,4]

[1]Department of Pathology, University of Iowa

[2]Medical Scientist Training Program, University of Iowa

[3]Graduate Program in Immunology, University of Iowa

[4]Graduate Program in Molecular Medicine, University of Iowa

Abstract

The human gut is colonized by trillions of bacteria that support physiologic functions such as food metabolism, energy harvesting, and regulation of the immune system. Perturbation of the healthy gut microbiome has been suggested to play a role in the development of inflammatory diseases, including multiple sclerosis (MS). Environmental and genetic factors can influence the composition of the microbiome; therefore, identification of microbial communities linked with a disease phenotype has become the first step towards defining the microbiome's role in health and disease. Use of 16S rRNA metagenomic sequencing for profiling bacterial community has helped in advancing microbiome research. Despite its wide use, there is no uniform protocol for 16S rRNA-based taxonomic profiling analysis. Another limitation is the low resolution of taxonomic assignment due to technical difficulties such as smaller sequencing reads, as well as use of only forward (R1) reads in the final analysis due to low quality of reverse (R2) reads. There is need for a simplified method with high resolution to characterize bacterial diversity in a given biospecimen. Advancements in sequencing technology with the ability to sequence longer reads at high resolution have helped to overcome some of these challenges. Present sequencing technology combined with a publicly available metagenomic analysis pipeline such as R-based Divisive Amplicon Denoising Algorithm-2 (DADA2) has helped advance microbial profiling at high resolution, as DADA2 can assign sequence at the genus and species levels. Described here is a guide for performing bacterial profiling using two-step amplification of the V3–V4 region of the 16S rRNA gene, followed by analysis using freely available analysis tools (i.e., DADA2, Phyloseq, and METAGENassist). It is believed that this simple and complete workflow will serve as an excellent tool for researchers interested in performing microbiome profiling studies.

Keywords

Biology; Issue 152; Fecal DNA extraction; library preparation; 16S variable region; 16S rRNA next generation sequencing; gut microbiota

Introduction

Microbiota refers to a collection of microorganisms (bacteria, viruses, archaea, bacteriophages, and fungi) living in a particular environment, and the microbiome refers to the collective genome of resident microorganisms. As bacteria are one of the most abundant microbes in humans and mice, this study is focused only on bacterial profiling. The human gut is colonized by trillions of bacteria and hundreds of bacterial strains[1]. The normal gut microbiota plays a vital role in maintaining a healthy state in the host by regulating functions (i.e., maintenance of an intact intestinal barrier, food metabolism, energy homeostasis, inhibition of colonization by pathogenic organisms, and regulation of immune responses)[2,3,4,5]. Compositional perturbations of the gut microbiota (gut dysbiosis) have been linked to a number of human diseases, including gastrointestinal disorders[6], obesity[7,8], stroke[9], [10]cancer, diabetes[8,11], rheumatoid arthritis[12], allergies[13], and central nervous system-related diseases such as multiple sclerosis (MS)[14,15] and Alzheimer's disease (AD)[8,16]. Therefore, in recent years, there has been growing interest in tools for identifying bacterial composition at different body sites. A reliable method should have characteristics such as being high-throughput and easy-to-use, having the ability to classify bacterial microbiota with high resolution, and being low-cost.

Culture-based microbiological techniques are not sensitive enough to identify and characterize the complex gut microbiome due to the failure of several gut bacteria to grow in culture. The advent of the sequencing-based technology, especially 16S rRNA-based metagenomic sequencing, has overcome some of these challenges and transformed microbiome research[17]. Advanced 16S rRNA-based sequencing technology has helped in establishing a critical role for the gut microbiome in human health. The Human Microbiome Project, a National Institutes of Health initiative[18], and the MetaHIT project (a European initiative)[19] have both helped in establishing a basic framework for microbiome analysis. These initiatives helped kick-start multiple studies to determine the role of the gut microbiome in human health and disease.

A number of groups have shown gut dysbiosis in patients with inflammatory diseases[12,14,15,20,21,22]. Despite being widely used for taxonomic profiling due to the ability to multiplex and low costs, there are no uniform protocols for 16S rRNA-based taxonomic profiling. Another limitation is the low resolution of taxonomic assignment owing to smaller sequencing reads (150 bp or 250 bp) and use of only forward sequencing read (R1) due to low quality reverse sequencing reads (R2). However, advances in sequencing technology have helped to overcome some of these challenges, such as the ability to sequence longer reads using paired-end reads (e.g., Illumina MiSeq 2×300bp).

The present sequencing technology can sequence 600 bp good quality reads, which allows merging of R1 and R2 reads. These merged longer R1 and R2 reads allow better taxonomic

assignments, especially with open-access R-based Divisive Amplicon Denoising Algorithm-2 (DADA2) platform. DADA2 utilizes amplicon sequence variant (ASV)-based assignments instead of operational taxonomic unit (OTU) assignments based on 97% similarity utilized by QIIME[23]. ASV matches result in an exact sequence match in the database within 1–2 nucleotides, which leads to assignment at genus and species levels. Thus, the combination of longer, good quality paired-end reads and better taxonomic assignment tools (such as DADA2) have transformed microbiome studies.

Provided here is a step-by-step guide for performing bacterial profiling using two-step amplification of the V3–V4 region of 16S rRNA and data analysis using DADA2, Phyloseq, and METAGENassist pipelines. For this study, human leukocyte antigen (HLA) class II transgenic mice are used, as certain HLA class II alleles are linked with a predisposition to autoimmune diseases such as MS[20,24,25]. However, the importance of HLA class II genes in regulating the composition of gut microbiota is unknown. It is hypothesized that the HLA class II molecule will influence gut microbial community by selecting for specific bacteria. Major histocompatibility complex (MHC) class II knockout mice (AE.KO) or mice expressing human HLA-DQ8 molecules (HLA-DQ8)[24,25,26] were used in order to understand the importance of HLA class II molecules in shaping the gut microbial community. It is believed that this complete and simplified workflow with R-based data analysis will serve as an excellent tool for researchers interested in performing microbiome profiling studies.

The generation of mice lacking endogenous murine MHC class II genes (AE.KO) and AE $^{-/-}$. HLA-DQA1*0103, DQB1*0302 (HLA-DQ8) transgenic mice with a C57BL/6J background has been described previously[26]. Fecal samples are collected from mice of both sexes (8–12 weeks of age). Mice were previously bred and maintained in the University of Iowa animal facility as per the NIH and institutional guidelines. Contamination control strategies such as weaning of the mice inside a laminar flow cabinet, changing of gloves between different strains of mice, and proper maintenance of mice are critical steps for profiling of gut microbiome.

Proper personal protective equipment (PPE) are highly recommended during the entire procedure. Appropriate negative controls should be included when performing DNA isolation, PCR1 and PCR2 amplification, and sequencing steps. Use of sterile, DNase-free, RNase-free, and pyrogen-free supplies is recommended. Designated pipettor for microbiome work and filtered pipette tips should be used throughout the protocol. Microbiota analysis consists of seven steps: 1) fecal sample collection and processing; 2) extraction of DNA; 3) 16S rRNA gene amplification; 4) DNA library construction using indexed PCR; 5) clean-up and quantification of indexed PCR (library); 6) MiSeq sequencing; and 7) data processing and sequence analysis. A schematic diagram of all protocol steps is shown in Figure 1.

Protocol

The protocol was approved by the Institutional Animal Care and Use Committee of the University of Iowa.

1. **Fecal Sample Collection and Handling**

1. Sterilize the divider boxes (see Table of Materials, Supplementary Figure 1) with 70% ethanol.

2. Pre-label microcentrifuge tubes (one per mouse) with the sample ID and treatment group (if applicable).

3. Place the mice in sterilized divider boxes and allow them to defecate normally for up to 1 h.

4. Collect the fecal pellets in an empty, pre-labeled 1.5 mL microcentrifuge tube using sterile forceps and close the tube securely. Sterilize the forceps after collecting from each mouse.

5. Place the microcentrifuge tube containing fecal pellets in a −80 °C freezer until further processing.

   NOTE: The divider boxes are advantageous because they allow simultaneous collection of fecal samples from multiple mice (up to 12) at one time.

2. **Extraction of DNA**

1. Remove the fecal samples (mouse or human) from the freezer and thaw at room temperature (RT).

   NOTE: It is advisable to thaw human stool samples overnight at 4 °C as needed to collect 200 mg or the required amount from stock samples.

2. Use 200 mg of starting materials and elute DNA to a final volume of 50 μL.

3. Include a DNA isolation kit blank in which no fecal sample is added but is processed through all DNA extraction steps.

   NOTE: A specific DNA isolation kit (see Table of Materials) was used, as it contains specific reagents to remove inhibiting materials such as biosolids, undigested plant material, and heme compounds from lysed red blood cells present in human and mouse stool samples.

4. Place the bead tube into a homogenizer (see Table of Materials) and homogenize the samples for 45 s at RT and a speed of 4.5 m/s.

   NOTE: Bead-beating homogenizer from any manufacturer can be used. However, it is recommended to standardize the method, specifically the speed and duration of homogenization, when using bead-beating homogenizer from another vendor.

5. Isolate DNA from individual mouse fecal samples using a bacterial DNA isolation kit following the manufacturer's protocol (see Table of Materials) with minor modifications. Quantify the isolated DNA by loading 1 μL of the DNA on a fluorimeter or on the high sensitivity electrophoresis chip (see Table of Materials).

NOTE: The expected yield of DNA can range from 500–2,500 ng when starting with 200 mg of the fecal sample.

6. Adjust the concentration of DNA to 4–20 ng/μL using elution buffer. Requantify the DNA (as done in step 2.5) before proceeding to 16S rRNA gene amplification (PCR1), if PCR1 is not performed on the same day.

3. 16S rRNA Gene Amplification (PCR1)

1. Set up 16S rRNA gene amplification (PCR1) in a 96 well PCR plate using a 25 μL reaction volume.

2. Using a multichannel pipette, add 12.5 μL of 2x high-fidelity polymerase enzyme mix including buffer, in addition to dNTPs (see Table of Materials): 40 ng of DNA in up to 10.5 μL total volume (adjust the total volume with PCR grade water): 1 μL (each) of forward and reverse primers at 1 μM concentration.

    NOTE: Sequences of the primers are as follows: forward primer = 5'-TCGTCGGCAGCGTCA GATGTGTATAAGAGA CAGCCTACGGGNGGCWGCAG-3' reverse primer = 5'-GTCTCGTGGGCTCGGAGATGTGTA TAAGAGACAGGACTACHVGGGTATCTAATC C-3'. Include a kit blank from step 2.3 (kit reagent control) in the PCR plate.

3. Seal the PCR plate and centrifuge at $1,000 \times g$ at 20 °C for 1 min in a tabletop plate centrifuge (see Table of Materials) and perform PCR in a thermal cycler programmed for: 95 °C for 3 min; 25 cycles of 1) 95 °C for 30 s, 2) 55 °C for 30 s, 3) 72 °C for 30 s; final extension at 72 °C for 5 min; and hold at 4 °C.

    NOTE: Although this 16S rRNA gene amplification method should work with different types of biospecimens, it is advised to standardize the number of amplification cycles when starting a new project.

4. Confirm the size of PCR1 product by loading 1 μL of the DNA on a high sensitivity electrophoresis chip. Alternatively, run 5 μL of 16S rRNA-amplified product on a 1.5% agarose gel to confirm 550 bp of the PCR1 product.

    NOTE: Clean-up of 16S rRNA amplified product is optional and depends on the DNA isolation kit/method being used. If using an in-house DNA isolation method, PCR1 product can be cleaned utilizing a microbiome DNA purification kit as per the manufacturer's protocol (see Table of Materials). The DNA isolation kit used here yields an ultrapure DNA and does not require clean-up of the PCR1 product.

4. DNA Library Construction Using Indexed PCR (PCR2)

1. Place the Index 1 and Index 2 barcoded primers in a special rack (see Table of Materials) for 96 libraries.

    1. Arrange Index 1 primer in columns 1–12 and Index 2 primer in rows A–H of the special rack.

2. Add 2.5 μL of Index 1 in columns 1–12 and Index 2 primer in rows A–H using multichannel pipettes. Place the new cap (see Table of Materials) on Index 1 and Index 2 adopter primers and store it in a −20 °C freezer.

2. Using a multichannel pipette, add 12.5 μL of 2x high fidelity polymerase enzyme mix containing a buffer, in addition to dNTPs (**see** Table of Materials); 5 μL of PCR grade water and 2.5 μL of 16S rRNA-amplified product.

    NOTE: Add unique indices to each sample for multiplexing of more than 96 libraries in a single run, as described in Kiernan et al.[27]. The present protocol uses adapters from a commercial kit (e.g., Nextera XT Index Kit) as per the manufacturer's instruction provided in the 16S metagenomics sequencing method (e.g., Illumina).

3. Seal and centrifuge the Indexed PCR plate at $1,000 \times g$ at 20 °C for 1 min and perform PCR in a thermal cycler programmed for: 95 °C for 3 min; 8 cycles of 1) 95 °C for 30 s, 2) 55 °C for 30 s, 3) 72 °C for 30 s; and final extension at 72 °C for 5 min.

4. Confirm the 630 base size of the indexed PCR product by loading 1 μL of DNA on a high sensitivity electrophoresis chip. Alternatively, run 5 μL of indexed PCR product on a 1.5% agarose gel to confirm the size and intensity of the product.

5. Clean-up of Indexed PCR (PCR2) and Quantification

1. Pool 5 μL of PCR2 amplicon using multichannel pipettes from each well into a multichannel reservoir tray free of detectable DNase, RNase, human DNA, and Pyrogenic bacteria (see Table of Materials).

2. Transfer the pooled product from the multichannel reservoir tray into an empty, pre-labeled 1.5 mL microcentrifuge tube and vortex to mix.

3. Purify PCR2 product using standard magnetic beads kit (**see** Table of Materials) as per the manufacturer's instruction. Seal and store the remaining 20 xL of PCR2 in the same plate at −80 °C for further use, if needed.

4. Prepare fresh 80% ethanol by adding 4 mL of 100% ethanol to 1 mL of PCR grade water.

5. Equilibrate the magnetic bead to RT and vortex for 30 s to disperse the beads evenly.

6. Briefly vortex and spin down the pooled PCR2 amplicon samples.

7. Add 80 μL of magnetic beads into a pre-labeled, sterile 1.5 mL microcentrifuge tube with 80 μL of pooled PCR products, then vortex and spin down briefly to evenly resuspend the magnetic beads.

8. Incubate the contents at RT without disturbing the tubes for 15 min.

9. Place the tube with the DNA and magnetic beads on a magnetic stand (**see** Table of Materials) for 5 min.

10. Carefully remove and discard 150 μL of the supernatant.

11. Add 200 μL of freshly prepared 80% ethanol without disturbing the beads and incubate for 30 s.

12. Carefully remove, and discard all the supernatant.

13. Repeat steps 5.11–5.12. Remove the remaining volume with a P10 pipette. Allow the beads to air-dry for 15 min, with the index PCR tube remaining on the magnetic stand.

14. Remove from the magnetic stand. Add 33 μL of elution buffer (elution buffer of DNA kit is acceptable). Vortex well, and perform a quick spin to remove any remaining liquid on the side. Incubate for 2 min and place on the magnetic stand for 5 min.

15. Transfer 30 μL of the supernatant (clean PCR products) to a pre-labeled 1.5 mL microcentrifuge tube.

16. Quantify the purified pool by loading 1 μL of the purified pool on a fluorimeter or high sensitivity electrophoresis chip, as this will be required during sequencing. Perform MiSeq as detailed below.

6. MiSeq Sequencing

1. Create a sample sheet containing sample-specific barcode information for metagenomics workflow and demultiplexing on the MiSeq instrument (**see** Table of Materials). Upload this sample sheet to the software (e.g., Illumina Experiment Manager).

2. Dilute the pooled libraries from step 5.15 to 4 nM.

3. Denature pooled libraries by combining 5 μL of the 4 nM library pool with 5 μL of freshly prepared 0.2 M NaOH in a 1.5 mL microcentrifuge tube. Vortex briefly to mix, centrifuge briefly and incubate for 5 min at ambient RT.

4. Add 990 μL of ice-cold hybridization buffer (HB buffer) and pipette gently to mix. This will yield a 20 pM library.

5. Combine 2 μL of the 10 nM control library with 3 μL of EBT buffer (10 mM Tris-HCl, pH = 8.5, with 0.1% Tween 20) to yield a 4 nM control library. Add 5 μL of freshly prepared 0.2 N NaOH and vortex briefly to mix. Incubate for 5 min at RT.

6. Add 990 μL ice-cold hybridization buffer (HB buffer) and pipette gently to mix. This will yield 20 pM control libraries.

   NOTE: Denatured 20 pM control libraries can be stored at −20 °C up to for 4 weeks. After 4 weeks, cluster numbers tend to decrease.

   1. Combine 210 μL of the 20 pM library with 40 μL of the 20 pM control library (final concentration = ~18%), and add 350 μL of HB buffer. Load the library at a final concentration of 7 pM.

NOTE: Input details should be adjusted as per run performance.

7. Incubate samples for 2 min at 96 °C. Put on ice for 5 min. Load 600 μL of the final pool into the appropriate well of the MiSeq cartridges.

    NOTE: Section 6 above can be performed at a genomic/DNA core facility.

7. Data Processing and Sequence Analysis

1. Use R software (version 3.5) for DADA2 data processing and analyses. For steps 7.1–7.4, use the open-access software as outlined in the previously developed DADA2 online tutorial found at <https://benjjneb.github.io/dada2/tutorial.html>.

    NOTE: A readily usable R script has been attached as a Supplemental File 2, and users must change the name and source of sequencing files (e.g., SAMPLENAME_R1_001.fastq and SAMPLENAME_R2_001.fastq).

2. Visualize the quality profiles of the forward and reverse reads using the **plotQualityProfile** command.

3. Trim nucleotides from forward and reverse reads based on the quality plot. These parameters are specified by the truncLen parameter in DADA2.

    NOTE: Here, 280 is used as the length threshold for which the forward reads would be discarded, and 260 is used as the length threshold for which the reverse reads would be discarded.

4. Process the raw 16S data as fastq files by the DADA2 pipeline as outlined in the online tutorial (found at <https://benjjneb.github.io/dada2/tutorial.html>) to merge R1 and R2 reads and form amplicon sequence variants (ASVs), which are then used to assign taxa with the Silva reference database[23].

    NOTE: A sample amplicon sequence variant table generated from the DADA2 pipeline is included as Supplemental File 3. Either a Greengene or Silva reference database can be used, as no differences in bacterial classification were found using either of these databases.

5. Generate a user-defined mapping file that contains the metadata (i.e., genotype, gender, treatment, etc.) for each sample. A sample metadata file has been included as Supplemental File 4.

6. Calculate alpha diversity (Shannon index) and beta diversity using principal coordinates analysis (PCA) based on the rarefied OTU counts using Phyloseq[28] as outlined in the online tutorial, found at <https://benjjneb.github.io/dada2/tutorial.html>.

7. Perform the following analysis in METAGENassist[29].

    NOTE: Perform the differential abundance analysis using the Wilcoxon rank-sum test at the genus level. Heat maps and differentially abundant taxa are highlighted using METAGENassist[29], a publicly available and web-based analysis pipeline.

1. Upload the taxonomic abundance table (CSV format) and select **Samples in the column**.

2. Upload the mapping file (CSV format) and select **Samples in a row**.

3. Select **Options to remove variables** with over 50% zeroes, and exclude unassigned and unmapped reads.

4. Select **Options to normalize rows** by sum and log normalize columns.

5. Make a **Volcano, PCA**, or **PLSDA** plot by clicking the same in the left-hand column and click **Remove samples name** to make the graph.

6. Perform a *t*-test (if only two groups) or ANOVA (if greater than two groups) to visualize the features (bacteria) that differ among groups. Click **Selected features** to visualize specific bacteria that differ between groups.

7. Click **Dendrogram** or **Heat map** to create respective plots. Additional analysis, such as sample visualization by groups or *t*-test/ANOVA-based top 25 features, can be performed.

8. Click **RandomForest** to create graphs showing features that can be used for classification. Click the **Variance** tab on top to create graphs for the top features that differ between/among groups. Click **Feature details** to see a list of bacteria that differ among groups and click each bacterium to create a graphical summary of the same.

9. Click the **Outlier** tab on top to visualize samples that are outliers.

10. Finally, click **Download** and select either 1) a zip file containing all the analysis performed or 2) the desired features to download. This file should be saved as a unique name and will need to be unzipped before use.

NOTE: For detailed statistical tests performed during microbiome analysis, refer to the works of Chen et al. and Hugerth et al.[14,30].

Representative Results

As MHC class II molecules (HLA in humans) are central players in the adaptive immune response and show strong associations with a predisposition to MS[24,25,26], it was hypothesized that the HLA class II molecule would influence gut microbial composition. Therefore, mice lacking the MHC class II gene (AE.KO) or expressing human HLA-DQ8 gene (HLA-DQ8) were utilized to understand the importance of HLA class II molecules in shaping the gut microbial community.

Fecal samples were collected from AE.KO (n = 16) and HLA-DQ8 (n = 12) transgenic mice, bacterial DNA was extracted, and the V3–V4 region of the 16S rRNA gene was amplified. The amplicon size (550 bp) was confirmed by running the samples on a 1.5% agarose gel (Figure 2A, **lanes 1**–**6**). Further confirmation of 16S rRNA amplicon size (550 bp) was

performed by loading 1 μL of the PCR1 product on a high sensitivity electrophoresis chip (Figure 2B, **lanes 1**–**7**).

An electropherogram was generated from 16S rRNA PCR product, which showed peak regions with fragments sized ~550 bp (Figure 2C). Dual indices and sequencing adapters were attached using indexed PCR (PCR2) that assigned a unique identity to each sample and allowed multiplexing of many samples in a single MiSeq sequencing run. Confirmation of indexed PCR was performed by agarose gel electrophoresis (Figure 2A, **lanes 7**–**12**) and a high sensitivity electrophoresis chip (Figure 2B, **lanes 8–12**), Figure 2D]. All the samples from PCR2 were pooled, purified, and loaded onto a next-generation sequencer that yielded forward R1 and reverse R2 reads of good quality (Figure 3). The median obtained reads after quality filtering and trimming were 88,125 (range of 9,597–111,848).

Community ecology analysis was performed using the DADA2 analysis pipeline and visualized with Phyloseq and METAGENassist to demonstrate differences in alpha diversity (Figure 4) and beta diversity (Figure 5), as well as differences at the genus and species levels between groups. DADA2 analysis generated an abundance table with comma-separated-values in csv format, which was used for further downstream analysis using a web-based platform (i.e., Phyloseq and/or METAGENassist). The alpha and beta diversity analyses were performed based on user-defined categories listed in a mapping file.

Shannon diversity analysis revealed an overall lower alpha diversity for AE.KO mice compared to HLA-DQ8 transgenic mice (Figure 3). Ordination with principal coordinates analysis showed a distinct spatial clustering between AE.KO mice and HLA-DQ8 transgenic mice (Figure 5). An abundance table from DADA2 was also used to perform a comprehensive metagenomic analysis using open-access software METAGENassist[29]. Heat map-based clustering of bacterial abundance (genus level) (Figure 6A) and a box plot for specific bacteria showing the differences between two groups (Figure 6B) were generated utilizing a METAGENassist pipeline.

Heat map analysis showed that certain bacteria such as *Allobacullum, Desufovibrio*, and *Rikenella* were more abundant in HLA-DQ8 transgenic mice. In contrast, *Biolophila* was more abundant in AE.KO mice (Figure 6B). Relative abundances of individual bacteria (*Bilophila* and *Rikenella)* are shown in a representative box plot (Figure 6B). Altogether, the data demonstrate that AE.KO mice possess a distinct microbial community compared to that of HLA-DQ8 transgenic mice, with an absence of specific bacteria in AE.KO mice. The data also suggest that MHC class II molecules play an important role in the abundance of certain bacteria. In summary, this simple and detailed protocol will help researchers who are new to the microbiome field as well as those who need updates on the methods for achieving higher taxonomic resolution.

Discussion

The described protocol is simple, with easy-to-follow steps to perform microbiome profiling using 16S rRNA metagenomic sequencing from a large number of biospecimens of interest. Next-generation sequencing has transformed microbial ecology studies, especially in human

and pre-clinical disease models[31,32]. The main advantage of this technique is its ability to successfully analyze complex microbial compositions (culturable and non-culturable microbes) in a given biospecimen at a high throughput level and at a low cost[32]. However, several factors (i.e., batch effects, selection of primers for 16S rRNA gene, and sequence data analysis) remain a major obstacle in the widespread use of this technology.

Advanced 16S rRNA-based MiSeq sequencing technologies (2×300bp)[33] allow sequencing of ~600 nucleotides out of the 1,500 nucleotide-long 16S rRNA gene of bacteria, which overcame the earlier bottleneck of short sequencing reads. Primers specific for a different region of rRNA such as V1–V2, V3–V5, or V6–V9 can be used with each region-specific primers showing some bias for over- or under-detection of particular taxa[34,35]. Some groups prefer the V1–V2 region[21], which shows an increased bias for *Clostridium* but underdetection for certain *Bacteroidetes* species. In contrast, other groups prefer the V4, V3–V4, and V3–V5 regions, which demonstrate the least biased classification of bacterial taxa[15,20,36,37].

The present protocol uses primers specific for V3–V4 regions of the 16S rRNA gene, as it covers the longer region of 16S rRNA with two hypervariable regions (V3 and V4) compared to V4 alone. Additionally, V3–V4 specific amplicon allows merging of both forward (R1) and reverse (R2) reads, leading to better resolution of bacterial classification. Although the V3–V5 region provides longer reads and covers more hypervariable regions (V3, V4, and V5), it is challenging to merge R1 and R2 reads due to no/little overlapping regions between R1 and R2 reads. Therefore, a number of studies have used data only from R1 reads for bacterial classification when performing 16S rRNA metagenomic sequencing using V3–V5 region[14].

Proper biospecimen storage and handling are critical for microbiome analysis to prevent degradation of bacterial DNA or environmental contamination[38]. Long-term storage of biospecimen at RT or 4 °C can lead to overgrowth of certain bacteria or fungi. Samples can be either frozen immediately or transported at 4 °C (for 1–3 days), then frozen. Biospecimen can also be stored directly in preservatives such as nucleic acid stabilization solution (e.g., RNA-later), 95% ethanol, or a commercial storage kit. The general consensus is that either of these storage methods does not cause significant differences in bacterial community profiles[39,40]. Although a solution with preservatives allows for storage and transportation at RT, these samples cannot be used for RNA-based assays, metabolite analyses, or fecal transplant experiments. These issues have been discussed in detail elsewhere[39,40].

One of the critical steps in this protocol is use of a bead-beating-based method for the mechanical disruption of gram-positive bacteria and *Archaea*. An earlier study showed the highest bacterial diversity using the bead-beating-based method compared to other methods of cell lysis[41]. An incomplete bacterial lysis or contamination during DNA extraction can introduce bias in gut microbiome data[42]. Another important point to consider is contamination due to laboratory reagents included in kits called the kitome[43,44,45]. Samples with large biomass and rich bacterial diversity such as soil or feces show less of these problems compared to samples with lower biomass such as skin. Therefore, a water

extraction control should be included with each extraction and processed with other samples to identify the introduction of potential contamination due to DNA extraction[43,44,45].

In microbiome analysis, diversity within a sample can be measured by alpha diversity, a measure of the richness of species, or beta diversity, which estimates dissimilarities between samples/groups. Popular methods for measuring α-diversity include UniFrac (weighted and unweighted) coupled with multivariate statistical techniques such as principal coordinate analysis (PCoA) or Brey-Curtis dissimilarity. While UniFrac is based on phylogenetic distances, Brey-Curtis dissimilarity analysis utilizes bacterial abundance for generating plots. In depth descriptions about α- and β-diversity iare discussed elsewhere[30,46,47]. A number of statistical methods can be utilized to compare differences in microbial communities between groups[14,48]. It is advised to use adjusted p-values instead of raw p-values to correct for multiple testing[14].

There is a variety of bioinformatics software to analyze targeted sequencing data independently[49]. The proposed protocol uses R-based opensource software packages, which allows user-friendly and fast profiling of bacterial taxa through R-based DADA2 pipelines. The abundance table generated from DADA2 can be used for downstream analysis using phyloseq and METAGENassist. DADA2 pipeline is advantageous over QIIME because it does not require special features (i.e., installation of virtual machines or Docker containers), which need relatively large computational resources and special technical expertise. Especially for beginners to the 16S analysis, R is appealing, as it is free and allows users to take advantage of accessible online tutorials and analysis scripts that are easy to execute. Importantly, these analysis tools require relatively small computational resources and can be run on a PC, Macintosh, or Linux-based platform. Additionally, METAGENassist can use abundance tables generated from DADA2 pipelines as well as biological observation matrix (BIOM) files generated from QIIME/MG-RAST to perform analysis such as PCA, partial least squares discriminant analysis (PLS-DA), volcano plots, *t*-tests (comparing two groups), ANOVA (comparing three or more groups), heat map plots, random forest analysis, etc. METAGENassist was found be very user-friendly.

In summary, this protocol describes a simple 16S rRNA metagenomic profiling pipeline, with a detailed guide on sample collection, DNA extraction, metagenomic library preparation, sequencing on Illumina MiSeq, and user-friendly data analysis using freely available platforms (i.e., DADA2, phyloseq, and METAGENassist). Although 16S rRNA metagenomic-based taxonomic profiling is reliable for characterization of bacteria present in particular biospecimens, shotgun metagenomic sequencing may be a better approach for detailed metabolic pathway analysis and strain-specific bacterial identification.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Lynch SV, Pedersen O The Human Intestinal Microbiome in Health and Disease. New England Journal of Medicine. 375 (24), 2369–2379 (2016). [PubMed: 27974040]

2. Blacher E, Levy M, Tatirovsky E, Elinav E Microbiome-Modulated Metabolites at the Interface of Host Immunity. Journal of Immunology. 198 (2), 572–580 (2017).

3. Jarchum I, Pamer EG Regulation of innate and adaptive immunity by the commensal microbiota. Current Opinion in Immunology. 23 (3), 353–360 (2011). [PubMed: 21466955]

4. Rescigno M The intestinal epithelial barrier in the control of homeostasis and immunity. Trends in Immunology. 32 (6), 256–264 (2011). [PubMed: 21565554]

5. Zhang K, Hornef MW, Dupont A The intestinal epithelium as guardian of gut barrier integrity. Cell Microbiology. 17 (11), 1561–1569 (2015).

6. Ghoshal UC, Park H, Gwee KA Bugs and irritable bowel syndrome: The good, the bad and the ugly. Journal of Gastroenterology and Hepatology. 25 (2), 244–251 (2010). [PubMed: 20074148]

7. Ley RE, Turnbaugh PJ, Klein S, Gordon JI Microbial ecology: human gut microbes associated with obesity. Nature. 444 (7122), 1022–1023 (2006). [PubMed: 17183309]

8. Naseer MI et al. Role of gut microbiota in obesity, type 2 diabetes and Alzheimer's disease. CNS & Neurological Disorders - Drug Targets. 13 (2), 305–311 (2014). [PubMed: 24059313]

9. Yin J et al. Dysbiosis of Gut Microbiota With Reduced Trimethylamine-N-Oxide Level in Patients With Large-Artery Atherosclerotic Stroke or Transient Ischemic Attack. Journal of American Heart Association. 4 (11), (2015).

10. Azcarate-Peril MA, Sikes M, Bruno-Barcena JM The intestinal microbiota, gastrointestinal environment and colorectal cancer: a putative role for probiotics in prevention of colorectal cancer? American Journal of Physiology-Gastrointestinal and Liver Physiology. 301 (3), G401–424 (2011). [PubMed: 21700901]

11. Su M et al. Diversified gut microbiota in newborns of mothers with gestational diabetes mellitus. PLoS ONE. 13 (10), e0205695 (2018). [PubMed: 30332459]

12. Horta-Baas G et al. Intestinal Dysbiosis and Rheumatoid Arthritis: A Link between Gut Microbiota and the Pathogenesis of Rheumatoid Arthritis. Journal of Immunological Research. 4835189 (2017).

13. Panzer AR, Lynch SV Influence and effect of the human microbiome in allergy and asthma. Current Opinion in Rheumatology. 27 (4), 373–380 (2015). [PubMed: 26002029]

14. Chen J et al. Multiple sclerosis patients have a distinct gut microbiota compared to healthy controls. Science Reports. 6, 28484 (2016).

15. Tremlett H et al. Gut microbiota composition and relapse risk in pediatric MS: A pilot study. Journal of Neurological Science. 363, 153–157 (2016).

16. Pistollato F et al. Role of gut microbiota and nutrients in amyloid formation and pathogenesis of Alzheimer disease. Nutritional Reviews. 74 (10), 624–634 (2016).

17. Caporaso JG et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Science U. S. A 108 Suppl 1, 4516–4522 (2011).

18. Turnbaugh PJ et al. The human microbiome project. Nature. 449 (7164), 804–810 (2007). [PubMed: 17943116]

19. Dusko Ehrlich S, Meta H. I. T. c. [Metagenomics of the intestinal microbiota: potential applications]. Gastroenterologie Clinique et Biologique. 34 Suppl 1, S23–28 (2010). [PubMed: 20889001]

20. Jangi S et al. Alterations of the human gut microbiome in multiple sclerosis. Nature Communications. 7, 12015 (2016).

21. Miyake S et al. Dysbiosis in the Gut Microbiota of Patients with Multiple Sclerosis, with a Striking Depletion of Species Belonging to Clostridia XIVa and IV Clusters. PLoS ONE. 10 (9), e0137429 (2015). [PubMed: 26367776]

22. Shahi SK, Freedman SN, Mangalam AK Gut microbiome in multiple sclerosis: The players involved and the roles they play. Gut Microbes. 8 (6), 607–615 (2017). [PubMed: 28696139]

23. Callahan BJ et al. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods. 13 (7), 581–583 (2016). [PubMed: 27214047]

24. Das P et al. Complementation between specific HLA-DR and HLA-DQ genes in transgenic mice determines susceptibility to experimental autoimmune encephalomyelitis. Human Immunology. 61 (3), 279–289 (2000). [PubMed: 10689117]

25. Mangalam AK, Rajagopalan G, Taneja V, David CS HLA class II transgenic mice mimic human inflammatory diseases. Advances in Immunology. 97, 65–147 (2008). [PubMed: 18501769]

26. Mangalam A et al. HLA-DQ8 (DQB1*0302)-restricted Th17 cells exacerbate experimental autoimmune encephalomyelitis in HLA-DR3-transgenic mice. Journal of Immunology. 182 (8), 5131–5139 (2009).

27. Kiernan MG et al. The Human Mesenteric Lymph Node Microbiome Differentiates Between Crohn's Disease and Ulcerative Colitis. Journal of Crohn's & Colitis. 13 (1), 58–66 (2019).

28. McMurdie PJ, Holmes S phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE. 8 (4), e61217 (2013). [PubMed: 23630581]

29. Arndt D et al. METAGENassist: a comprehensive web server for comparative metagenomics. Nucleic Acids Research. 40, W88–95 (2012). [PubMed: 22645318]

30. Hugerth LW, Andersson AF Analysing Microbial Community Composition through Amplicon Sequencing: From Sampling to Hypothesis Testing. Frontiers in Microbiology. 8 1561 (2017). [PubMed: 28928718]

31. Srinivasan R et al. Use of 16S rRNA gene for identification of a broad range of clinically relevant bacterial pathogens. PLoS ONE. 10 (2), e0117617 (2015). [PubMed: 25658760]

32. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW Transforming clinical microbiology with bacterial genome sequencing. Nature Reviews Genetics. 13 (9), 601–612 (2012).

33. Buermans HP, den Dunnen JT Next generation sequencing technology: Advances and applications. Biochimica et Biophysica Acta. 1842 (10), 1932–1941 (2014). [PubMed: 24995601]

34. Cai L, Ye L, Tong AH, Lok S, Zhang T Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. PLoS ONE. 8 (1), e53649 (2013). [PubMed: 23341963]

35. Kuczynski J et al. Experimental and analytical tools for studying the human microbiome. Nature Reviews Genetics. 13 (1), 47–58 (2011).

36. Mangalam A et al. Human Gut-Derived Commensal Bacteria Suppress CNS Inflammatory and Demyelinating Disease. Cell Reports. 20 (6), 1269–1277 (2017). [PubMed: 28793252]

37. Shahi SK et al. Prevotella histicola, A Human Gut Commensal, Is as Potent as COPAXONE(R) in an Animal Model of Multiple Sclerosis. Frontiers in Immunology. 10, 462 (2019). [PubMed: 30984162]

38. Bliss DZ et al. Comparison of subjective classification of stool consistency and stool water content. Journal of Wound Ostomy & Continence Nursing. 26 (3), 137–141 (1999).

39. Kim D et al. Optimizing methods and dodging pitfalls in microbiome research. Microbiome. 5 (1), 52 (2017). [PubMed: 28476139]

40. Sinha R et al. Collecting Fecal Samples for Microbiome Analyses in Epidemiology Studies. Cancer Epidemiology, Biomarkers & Prevention. 25 (2), 407–416 (2016).

41. Salonen A et al. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. Journal of Microbiological Methods. 81 (2), 127–134 (2010). [PubMed: 20171997]

42. Santiago A et al. Processing faecal samples: a step forward for standards in microbial community analysis. BMC Microbiology. 14, 112 (2014). [PubMed: 24884524]

43. Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. Gut Pathogens. 8, 24 (2016). [PubMed: 27239228]

44. Salter SJ et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology. 12, 87 (2014). [PubMed: 25387460]

45. Weiss S et al. Tracking down the sources of experimental contamination in microbiome studies. Genome Biology. 15 (12), 564 (2014). [PubMed: 25608874]

46. Lozupone C, Lladser ME, Knights D, Stombaugh J,Knight R UniFrac: an effective distance metric for microbial community comparison. ISME Journal. 5 (2), 169–172 (2011). [PubMed: 20827291]

47. Schloss PD Evaluating different approaches that test whether microbial communities have the same structure. ISME Journal. 2 (3), 265–275 (2008). [PubMed: 18239608]

48. Xia Y, Sun J Hypothesis Testing and Statistical Analysis of Microbiome. Genes & Diseases. 4 (3), 138–148 (2017). [PubMed: 30197908]

49. Malla MA et al. Exploring the Human Microbiome: The Potential Future Role of Next-Generation Sequencing in Disease Diagnosis and Treatment. Frontiers in Immunology. 9, 2868 (2018). [PubMed: 30666248]
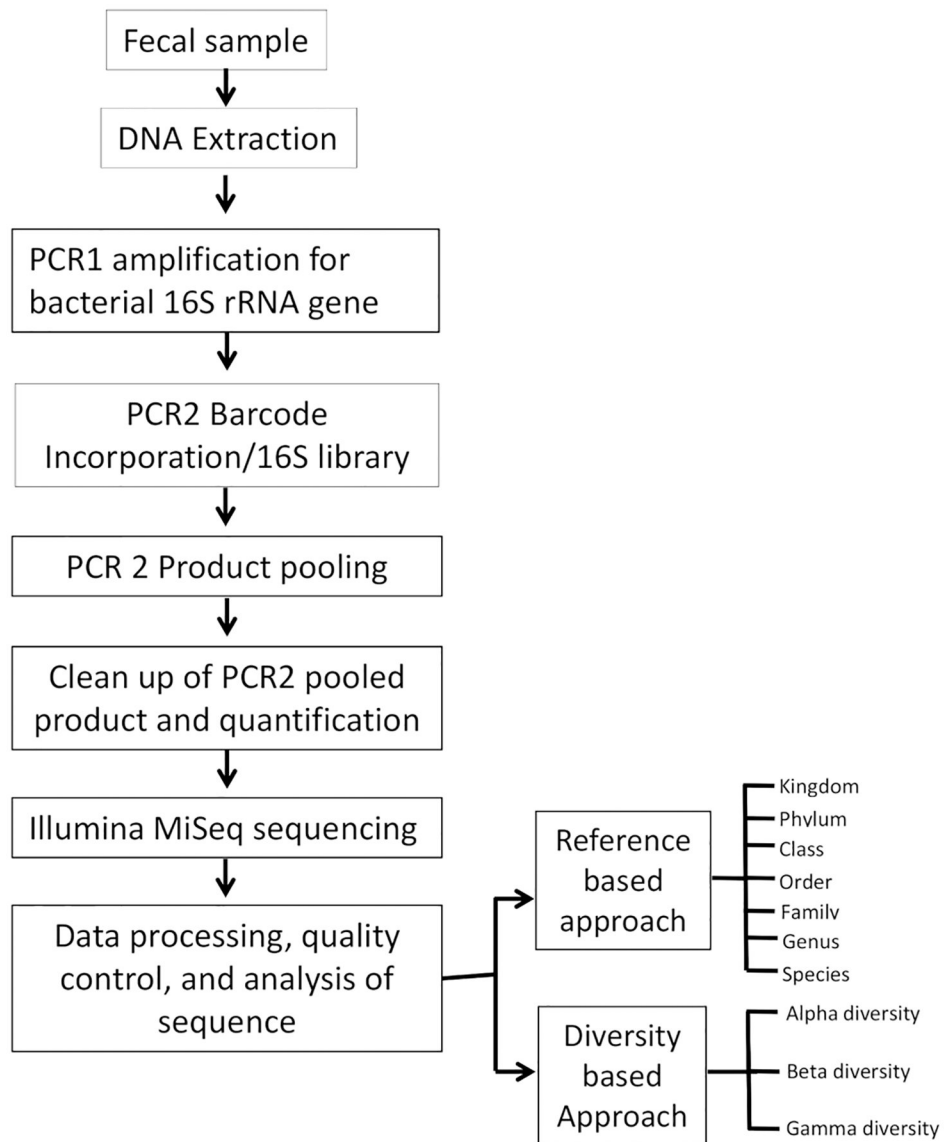
Figure 1: Flow diagram of gut microbiome sequencing.
All steps of the microbiome sequencing (sample collection to microbiome data analysis) are displayed.
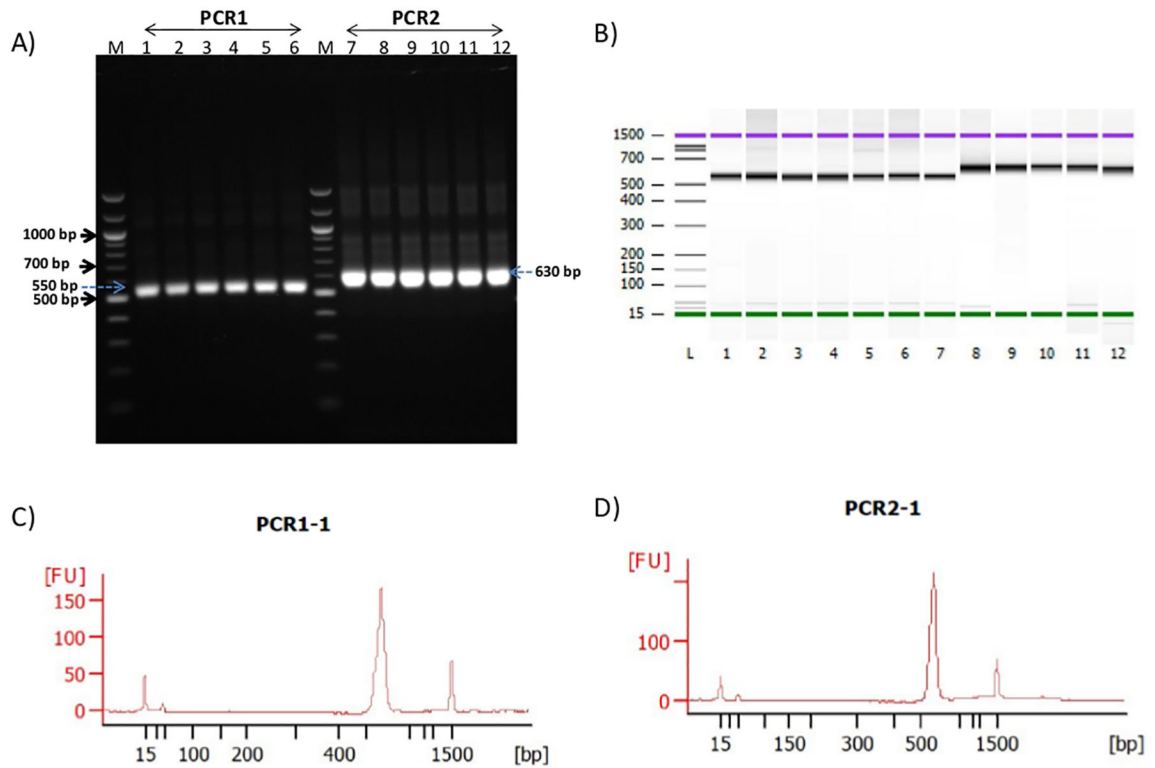
Figure 2: 16S rRNA gene amplification and quality control analysis of V3–V4 region.

(**A**) Representative agarose gel electrophoresis image of the 16S amplicon (PCR1, size 550 bp) and indexed PCR (PCR2, size = 630 bp) from AE.KO mice (lanes 1–3 and 7–9) and HLA-DQ8 transgenic mice (lanes 4–6 and 10–12). (**B**) Representative gel image of the 16S amplicon (PCR1, size = 550 bp) and indexed PCR (PCR2, size = 630 bp) of AE.KO mice (lanes 1–3 and 8–10) and HLA-DQ8 transgenic mice (lanes 4–7 and 11–12) resolved by electrophoresis. (**C**) Representative electropherogram of the 16S amplicon (PCR1) showed a peak region containing fragments that were sized ~550 bp. (**D**) Representative electropherogram of indexed PCR (PCR2) showed a peak region comprising of fragments sized ~630 bp.
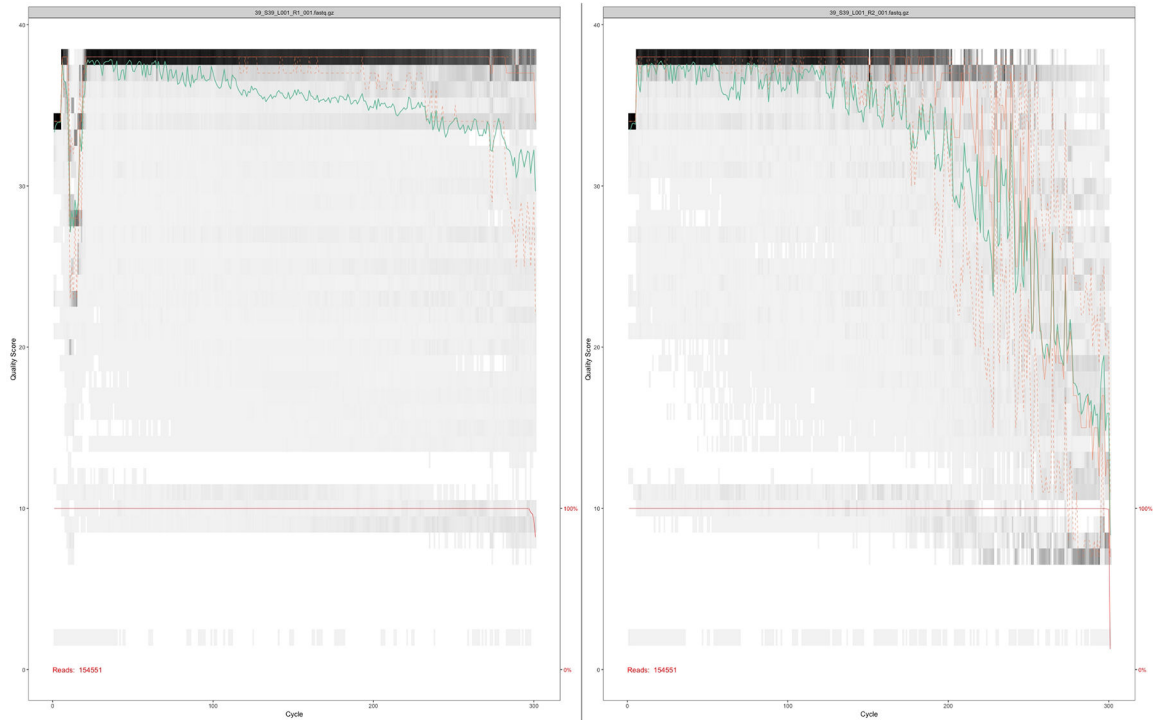
Figure 3: Quality profile of forward reads (R1, top) for two representative samples and corresponding reverse reads (R2, bottom) for the same samples.

The analysis was performed using a DADA2 pipeline, in which the x-axis shows read length (0–300 bases) and y-axis shows quality of the reads. Green line represents the median quality score, whereas the orange line represents quartiles of the quality score distribution at each base position. Forward reads (R1) always showed better quality than reverse reads (R2).
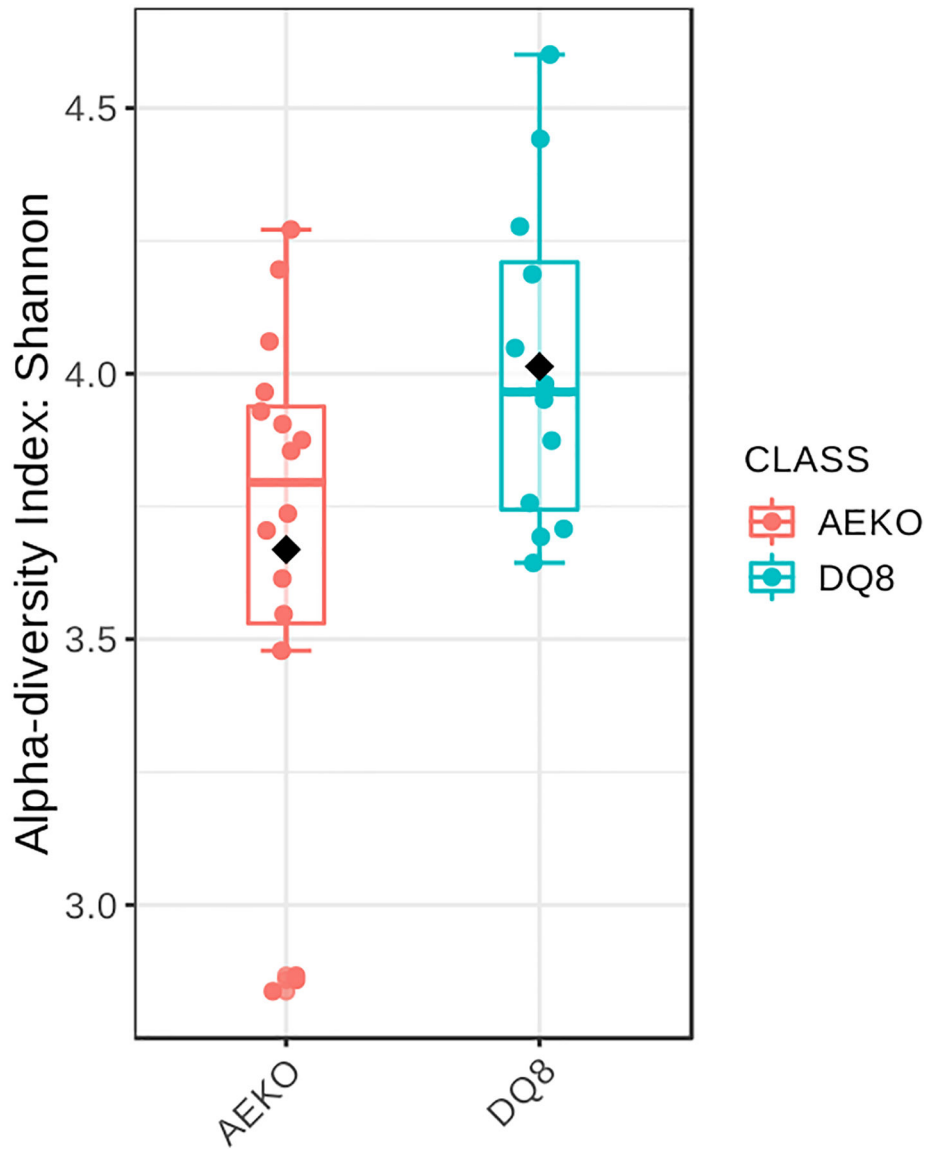
Figure 4: Alpha diversity measures (Shannon diversity) of AE.KO mice and HLA-DQ8 transgenic mice.

Each dot represents α-diversity (Shannon diversity) in a sample from a single mouse.

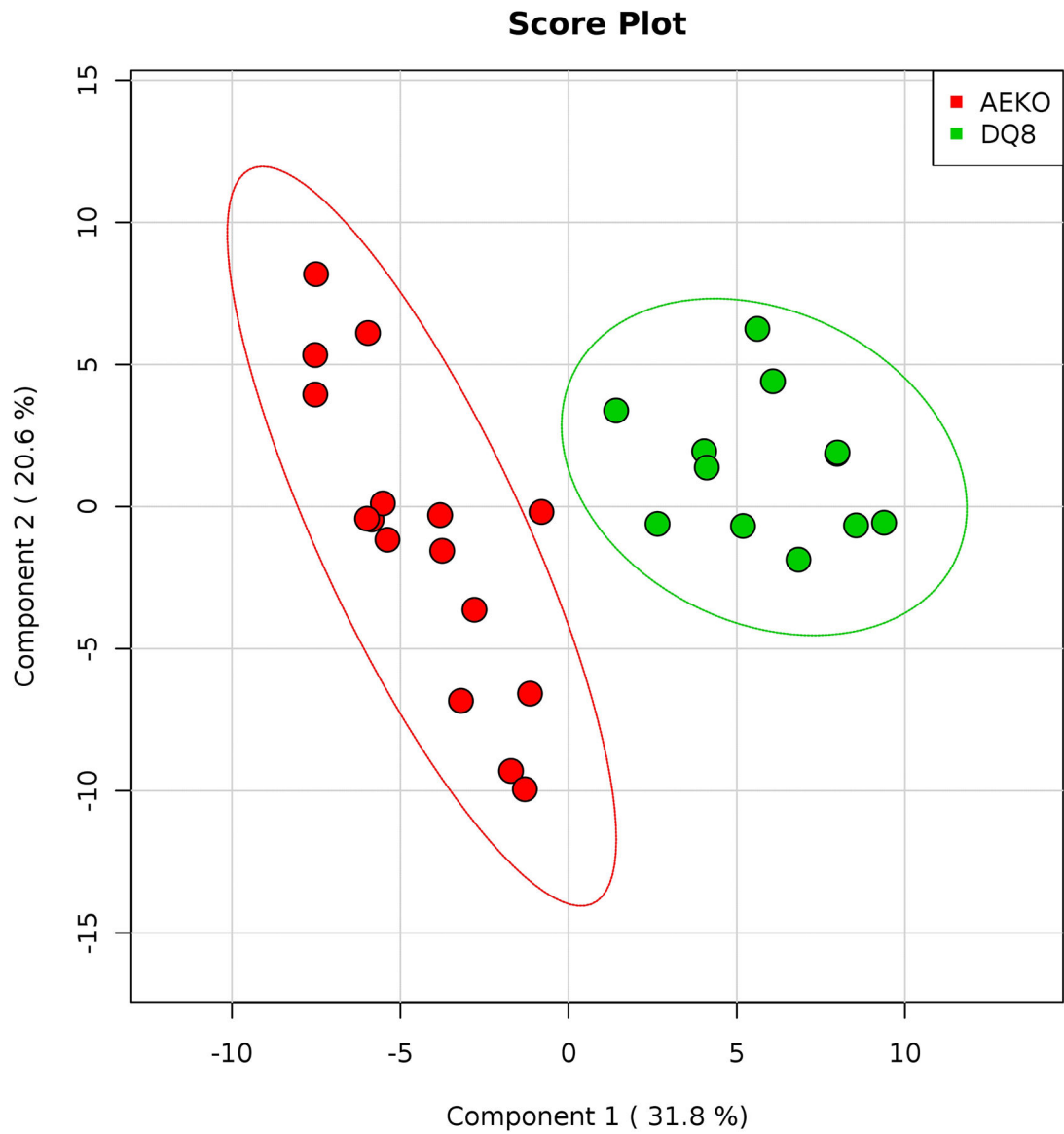Shannon diversity was overall lower for AE.KO mice compared to HLA-DQ8 transgenic mice.

Score Plot

Figure 5: Ordination with partial least squares-based dimension analysis plot.
The plot shows a clear separation between AE.KO mice and HLA-DQ8 transgenic mice.
Each dot represents bacterial composition within a sample, and dotted eclipses indicate 80%
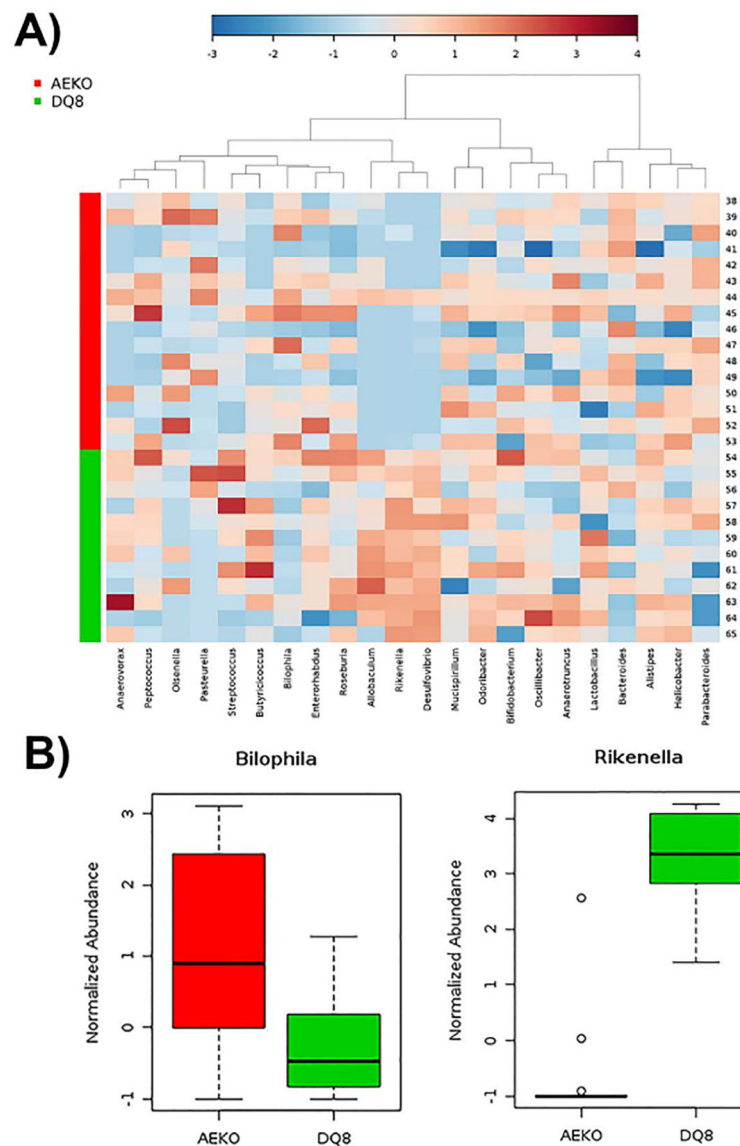confidence intervals. The PLS-DA plots were generated using METAGENassist.

Figure 6: AE.KO mice showing distinct microbial community compared to HLA-DQ8 transgenic mice, with an absence of specific bacteria in AE.KO mice.

(**A**) Heat map combined with agglomerative hierarchical clustering showing the relative abundance of bacteria (genus level). (**B**) Box plot showing a normalized relative abundance of two representative bacteria (*Bilophila* and *Rikenella*) in AE.KO and HLA-DQ8 transgenic mice. Both plots were generated using METAGENassist.