Methods of DNA extraction, Miseq sequencing & sequencing data analysis

DNA extraction

DNA was extracted from 200 mg samples using a DNA extration kit (QIAamp DNA Stool Mini Kit;QIAGEN, Hilden, Germany), following the manufacturer's instructions after preparation. DNA concentration and purity was checked by running the samples on 1.2% agarose gels.

PCR amplification of 16SrDNA genes and Miseq sequencing

Polymerase chain reaction (PCR) amplification of target gene was performed using general primers 16S V4-V5: 515F 5'-GTGCCAGCMGCCGCGGTAA-3' ; 926R 5'-CCGTCAATTCMTTTGAGTTT-3'. The primers also contained the Illumina 5'overhang adapter sequences for two-step amplicon library building, following manufacturer's instructions for the overhang sequences. The initial PCR reactions were carried out in 50 μL reaction volumes with 1-2 μL DNA template, 200 uM dNTPs, 0.2 uM of each primer, 5X reaction buffer 10uL and 1U Phusion DNA Polymerase (New England Biolabs, USA). PCR conditions consisted of initial denaturation at 94 ℃ for 2 min, followed by 25 cycles of denaturation at 94 ℃ for 30 s, annealing at 56 C for 30 s and extension at 72 ℃ for 30 s, with a final extension of 72 ℃ for 5 min. The second step PCR with dual 8-base barcodes was used for multiplexing. Eight cycle PCR reactions were used to incorporate two unique barcodes to either end of the 16S amplicons. Cycling conditions consisted of one cycle of 94 ℃ for 3 min, followed by eight cycles of 94 ℃ for 30 s, 56 ℃ for 30 s and 72 ℃ for 30 s, followed by a final extension cycle of 72 ℃ for 5 min. Prior to library pooling, the barcoded PCR products were purified using a DNA gel extraction kit (Axygen, USA) and quantified using the FTC -3000 TM real-time PCR(Funglyn Shanghai). The PCR products from different samples were indexed and mixed at equal ratios for sequencing by 2*300bp paired-end sequencing on the MiSeq platform

using MiSeq v3 Reagent Kit (Illumina) at TinyGen Bio-Tech Co., Ltd. (Shanghai,China)

Bioinformatic analysis

The raw fastq files were demultiplexed based on the barcode. PE reads for all samples were run through Trimmomatic (version 0.35) to remove low quality base pairs using these parameters (SLIDINGWINDOW: 50:20 MINLEN: 50). Trimmed reads were then further merged using FLASH program (version 1.2.11) with default parameters. The low quality contigs were removed based on screen.seqs command using the following filtering parameters, maxambig=0, minlength = 200, maxlength =580, maxhomop= 8.

The 16S sequences were analyzed using a combination of software mothur (version 1.33.3), UPARSE (version v8.1.1756), and R (version 3.6.0).

The demultiplexed reads were clustered at 97% sequence identity into operational taxonomic units (OTUs) and the singleton OTUs were deleted using the UPARSE pipeline.

The OTU representative sequences were assignment for taxonomy against Silva 128 database with confidence score ≥ 0.6 by the classify.seqs command in mothur. OTU taxonomies (from Phylum to Species) were determined based on NCBI.

For the alpha-diversity analysis, Shannon, simpson, Chao1, ACE index and rarefaction curves were calculated were using mothur and plotted by R.

For the beta-diversity metrics, the weighted and unweighted UniFrac distance matrix were calculated using mothur and visualized with Principal Coordinate Analysis (PCoA) and tree by R. The bray curits metrics were calculated by R and visualized also by R.