Methods

Sequencing

1. Genome DNA Extraction

The DNA of the total genome of the samples was extracted by the method CTAB/SDS. DNA concentration and purity were monitored on agarose gels. at 1%. According to the concentration, the DNA was diluted to 1ng/μL using sterile water.

2. Amplicon Generation

The genes 16S rRNA/18SrRNA/ITS of different regions (16SV4/16SV3/16SV3-V4/16SV4-V5, 18S V4/18S V9, ITS1/ITS2, Arc V4) were amplified using a specific primer (for example, 16S V4: 515F-806R, 18S V4: 528F-706R, 18S V9: 1380F-1510R, et al) with the barcode. All PCR reactions were carried out with Phusion® High-Fidelity PCR Master Mix (New England Biolabs).

3. Quantification and Qualification of PCR Products

Mix the same volume of 1X loading buffer (containing SYB green) with PCR products and run 2% agarose gel electrophoresis for detection. Samples with a bright main band between 400 and 450 bp were chosen for other experiments.

4. Mixing and Purification of PCR Products.

The PCR products were mixed in equidensity proportions. The PCR product mixture was then purified with the Qiagen gel extraction kit (Qiagen, Germany).

5. Preparation and Sequencing Library.

Sequencing libraries were generated using NEBNext Ultra DNA Library Pre® Kit for Illumina, following the manufacturer's recommendations and added index codes. Library quality was assessed on Qubit@ 2.0 Fluorometer (Thermo Scientific) and Agilent Bioanalyzer 2100 system. Finally, the library was sequenced on an Illumina platform and 250 bp end-pairs were generated.

Data Analysis.

1. Assembly of Paired Reads and Quality Control.

a. Data Division.

Samples were assigned paired reads based on their code single bars and were truncated by cutting out the barcode and primer sequence.

b. Sequence Montage.

Paired reads were merged using FLASH (V1.2.7, http://ccb.jhu.edu/software/FLASH/)[1], a very fast analysis tool and accurate, which was designed to merge paired reads when at least some of the readings overlap the reading generated from the opposite end of the same DNA fragment, and the splicing sequences are called tags without process.

c. Data Filtering.

The quality filtering on the raw labels was performed under conditions of specific filtering to obtain high quality clean labels[2] according to QIIME (V1.7.0, http://qiime.org/index.html)[3] controlled quality process.

d. Elimination of Chimeras.

The labels were compared with the reference database (database Gold, http://drive5.com/uchime/uchime_download.html) using the algorithm UCHIME UCHIME, (Algorithm http://www.drive5.com/usearch/manua/uchime_algo.html)[4] for chimeric sequences, and then the chimeric sequences[5] were removed. So finally the Effective Tags were obtained.

2. OTU Cluster and Species Annotation.

a. OTU Production.

Sequence analysis was performed with the Uparse software (Uparse v7.0.1001，http://drive5.com/uparse/)[6]. Sequences with ≥97% similarity were assigned to the same OTU. The representative sequence of each OTU was evaluated for get more information.

b. Species Annotation.

For each representative sequence, the GreenGene database (http://greengenes.lbl.gov/cgi-bin/nph-index.cgi)[7] based on the RDP 3 classifier (Version 2.2, http://sourceforge.net/projects/rdp-classifier/)[8] algorithm for annotating taxonomic information.

c. Construction of the Phylogenetics Relationship.

To study the phylogenetic relationship of different OTUs and the difference of the dominant species in different samples (groups), a multiple sequence alignment was performed using the MUSCLE software (Version 3.8.31, http://www.drive5.com/muscle/)[9].

d. Data Normalization.

OTU abundance information is normalized using a sequence number standard corresponding to the sample with fewer sequences. All subsequent analyzes of alpha diversity and beta diversity were performed based on this normalized output data.

3. Alpha Diversity.

Alpha diversity is applied in the analysis of the complexity of species diversity for a sample through 6 indices, including observed species, Chao1, Shannon, Simpson, ACE, Good coverage. All these indices in our samples were calculated with QIIME (Version 1.7.0) and displayed with R software (Version 2.15.3).

Two indices were selected to identify community wealth:

Chao - the Chao1 estimator (http://www.mothur.org/wiki/Chao)

ACE - the ACE estimator (http://www.mothur.org/wiki/Ace)

Two indices were selected to identify community diversity:

Shannon - the Shannon index (http://www.mothur.org/wiki/Shannon)

Simpson - the Simpson index (http://www.mothur.org/wiki/Simpson)

An index to characterize sequencing depth:

Coverage - the Good's coverage (http://www.mothur.org/wiki/Coverage).

4. Beta Diversity.

Beta diversity analysis was used to assess sample differences in species complexity. Beta diversity in weighted and unweighted unifrac was calculated using QIIME software (Version 1.7.0).

Cluster analysis was preceded by principal component analysis (PCA), which was applied to reduce the dimension of the original variables using the FactoMineR package and ggplot2 park in R software (Version 2.15.3).

Principal coordinate analysis (PCoA) was performed to obtain principal coordinates and visualize complex multidimensional data. An unweighted unifrac or weighted distance matrix between the samples obtained above was transformed into a new set of orthogonal axes, whereby the maximum variation factor is shown by the first principal coordinate, and the second maximum by the principal coordinate, and so on. successively. PCoA analysis was shown using the WGCNA package, stat packages, and ggplot2 package in R software (Version 2.15.3).

Unweighted Pair Group Method with Arithmetic Means (UPGMA). The clustering was performed as a type of hierarchical clustering method to interpret the distance matrix using link average and was performed by QIIME software (Version 1.7.0).

Reference

[1] Magoč T, Salzberg S L. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27.21 (2011): 2957-2963.

[2] Bokulich, Nicholas A., et al. Quality-filtering vastly improves diversity estimates from Illuminaamplicon sequencing. Nature methods 10.1 (2013): 57-59.

[3] Caporaso, J. Gregory, et al. QIIME allows analysis of high-throughput community sequencing data. Nature methods 7.5 (2010): 335-336.

[4] Edgar, Robert C., et al. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27.16 (2011): 2194-2200.

[5] Haas, Brian J., et al. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons.Genome research 21.3 (2011): 494-504.

[6] Edgar, Robert C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature methods 10.10 (2013): 996-998.

[7] DeSantis, Todd Z., et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and environmental microbiology 72.7 (2006): 5069-5072.

[8] Wang, Qiong, et al. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and environmental microbiology 73.16 (2007): 5261-5267.

[9] Edgar R C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research32.5(2004): 1792-1797.