Methods

Sequencing

1. Extraction of genome DNA

Total genome DNA from samples was extracted using CTAB/SDS method. DNA concentration and purity was monitored on 1% agarose gels. According to the concentration, DNA was diluted to l ug/μL using sterile water.

2. Amplicon Generation

16S rRNA/18S rRNA/ITS genes of distinct regions(16S V4/16S V3/16S V3-V4/16S V4-V5,18S V4/18S V9,ITS1/ITS2, Arc V4) were amplified used specific primer(e.g.16S V4:515F-806R,18S V4: 528F-706R, 18S V9:1380F-1510R,et.al) with the barcode. All PCR reactions were carried out with 15 μL of Phusion® High-Fidelity PCR Master Mix (New England Biolabs); 0.2 μM of forward and reverse primers, and about 10 ng template DNA. Thermal cycling consisted of initial denaturation at *98°C* for 1 min, followed by 30 cycles of denaturation at *98°C* for 10 s, annealing at *50°C* for 30 s, and elongation at 72°C for 30 s. Finally 72℃ for 5 min.

3. PCR Products quantification and qualification

Mix same volume of IX loading buffer (contained SYB green) with PCR products and operate electrophoresis on 2% agarose gel for detection. PCR products was mixed in equidensity ratios. Then, mixture PCR products was purified with Qiagen Gel Extraction Kit(Qiagen, Germany).

4. Library preparation and sequencing

Sequencing libraries were generated using TruSeq® DNA PCR-Free Sample Preparation Kit (Illumina, USA) following manufacturer's recommendations and index codes were added. The library quality was assessed on the Qubit@2.0 Fluorometer (Thermo Scientific) and Agilent Bioanalyzer 2100 system. At last, the library was sequenced on an Illumina NovaSeq platform and 250 bp paired-end reads were generated.

Data analysis

1.Paired-end reads assembly and quality control

1.1 Data split

Paired-end reads was assigned to samples based on their unique barcode and truncated by cutting off the barcode and primer sequence.

1.2 Sequence assembly

Paired-end reads were merged using FLASH (VI.2.7,http://ccb.jhu.edu/software/FLASH/)[1],a very fast and accurate analysis tool, which was designed to merge paired-end reads when at least some of the reads overlap the read generated from the opposite end of the same DNA fragment, and the splicing sequences were called raw tags.

1.3 Data Filtration

Quality filtering on the raw tags were performed under specific filtering conditions to obtain the high-quality clean tag[2]according to the QIIME(V1.9.1, http://qiime.org/scripts/split libraries fastq.html[3]quality controlled process.

1.4 Chimera removal

The tags were compared with the reference database(Silva database, https://www.arb-silva.de/) using UCHIME algorithm (UCHIME Algorithm,http://www.drive5.com/usearch/manual/uchime_algo.html)[4]to detect chimera sequences, and then the chimera sequences were removed[5].Then the Effective Tags finally obtained.

Reference

[1] Magoc 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.