water

MDPI

Article

The Nitrogen Removal Performance and Functional Bacteria in Heterotrophic Denitrification and Mixotrophic Denitrification Process

Xinghao Ren, Yan Wang, Jinquan Wan, Zhicheng Yan，Yongwen Ma, Guihua Zhang and Bin Zhu

Supplementary Materials

Table S1. The concentrations of micronutrients in this experiment.

Micronutrients	Concentration(mg/L)
$ZnSO_4$	22
$CaCl_2$	55
$MnCl_2 \cdot 4H_2O$	50.6
$(NH_4)_6Mo_7O_{24} \cdot 4H_2O$	11
$CuSO_4 \cdot 5H_2O$	15.7
$CoCl_2 \cdot 6H_2O$	16.1

Test S1 high-throughput sequencing method

1. Extraction of genome DNA

Total genome DNA from samples was extracted according to manufacturer's protocols. DNA concentration was monitored by Qubit® dsDNA HS Assay Kit. The preparation of next generation sequencing libraries and Illumina sequencing was conducted by GENEWIZ (Suzhou, China).

2. Amplicon generation and library preparation

The sequencing library was constructed using a Meta VX Library Preparation Kit (GENEWIZ, Suzhou, China). Briefly, 20–30 ng of DNA was used to generate amplicons that cover V3 and V4 hypervariable regions of the 16s rRNA gene of bacteria. The forward primer contained the sequence 'CCTACGGRRBGCASCAGKVRVGAAT' and the reverse primers contained the sequence 'GGACTACNVGGGTWTCTAATCC'. The 25 μl PCR mixture was prepared with 2.5 μl of Trans Start buffer, 2 μl of dNTPs, 1 μl of each primer, 0.5 μl of Trans Start Taq DNA polymerase and 20 ng template DNA. The PCR was performed by the following program: 3 min of denaturation at 94℃, 24 cycles of 5s at 95℃, 90s of annealing at 57℃, 10s of elongation at 72℃, and a final extension at 72℃ for 5min. Indexed adapters were added to the ends of the amplicons by limited cycle PCR. Finally, the library was purified with magnetic beads.

3. Illumina sequencing

The concentration was detected by a microplate reader (Tecan, Infinite 200 Pro) and the fragment size was detected by 1.5% agarose gel electrophoresis which was expected at 600 bp. Next generation sequencing was conducted on an Illumina Miseq/Novaseq Platform (Illumina, San Diego, USA). Automated cluster generation and 250/300 paired-end sequencing with dual reads were performed according to the manufacturer's instructions.

4. Data Analysis

The forward and reverse reads obtained from double-end sequencing were first spliced in two, and sequences containing N in the splicing result were filtered, and

sequences with a length greater than 200bp were retained. After quality filtering to remove chimeric sequences, the final obtained sequences were used for OTU clustering, and sequence clustering was performed using VSEARCH (1.9.6) (sequence similarity is set to 97%), and the 16S rRNA reference database for comparison was Silva 138. The representative sequences of OTUs were then analyzed for species taxonomy used the RDP classifier (Ribosomal Database Program), and the community composition of each sample was counted at different species classification levels.

Based on the results of the analysis obtained from OTU, the alpha diversity indices such as Shannon and Chao1 were calculated to reflect the species abundance and diversity of the community by used a random sampling of sample sequences for equalization.

Test S2 Metagenomic sequencing Method

Next generation sequencing library preparations were constructed following the manufacturer's protocol. 200μg genomic DNA was randomly fragmented by Covaris to an average size of 300-350 bp. The fragments were treated with End Prep Enzyme Mix for end repairing (5′ Phosphorylation and 3′ adenylated), to add adaptors to both ends. Size selection of Adaptor-ligated DNA was then performed by DNA Cleanup beads. Each sample was then amplified by PCR for 8 cycles using P5 and P7 primers, with both primers carrying sequences which can anneal with flowcell to perform bridge PCR and P7 primer carrying a six-base index allowing for multiplexing. The PCR products were cleaned up and validated using an Agilent 2100 Bioanalyzer. The qualified libraries were sequenced pair end PE150 on the illumina HiseqXten/Novaseq / MGI2000 System.

Raw shotgun sequencing reads were trimmed using cutadapt(v1.9.1). Low-quality reads, N-rich reads and adapter-polluted reads were removed. Then host contamination reads were removed using BWA(v0.7.12). Samples were each assembled *de novo* to obtain separate assemblies. Whole genome *de novo* assemblies were performed using MEGAHIT(v1.13) with different k-mer. The best assembly result of Scaffold, which has the largest N50, was selected for the gene prediction analysis.

Genes of each sample were predicted using Prodigal (v3.02). MMseq2 was used to cluster genes derived from all samples with default identity of 0.95 and coverage of 0.95. In order to analyze the relative abundance of unigenes in each sample, paied-end clean reads were mapped to unigenes using SOAPAligner (version 2.2.1) to generate read coverage information for unigenes. Gene abundance was calculated based on the number of aligned reads and normalized to gene length.

Diamond (version v0.8.15.77) was used to search the protein sequences of the unigenes with the NR database, CAZy database, eggNOG database, CARD database and KEGG database with E < 1e-5. The matched result with best scores was selected for annotation. In order to explore the microbial composition of the samples, we used Diamond to align the unigene sequences with the NR database, and obtained the species annotation results of each sequence through the taxonomic annotation information corresponding to each sequence in the NR database. The abundance of a specie in one sample equal the sum of the gene abundance annotated for the specie.

To determine the similarity or difference of taxonomic and functional components between different samples, relative Clustering analysis and Principal Component Analysis (PCA) were performed. Meanwhile, there were a series of advanced analysis items available to explore the environmental samples, such as LEfSe, significant difference, CCA/RDA, prediction of secreted protein an annotation of VFDB, MvirDB, PHI, and TCDB.