

# Methods

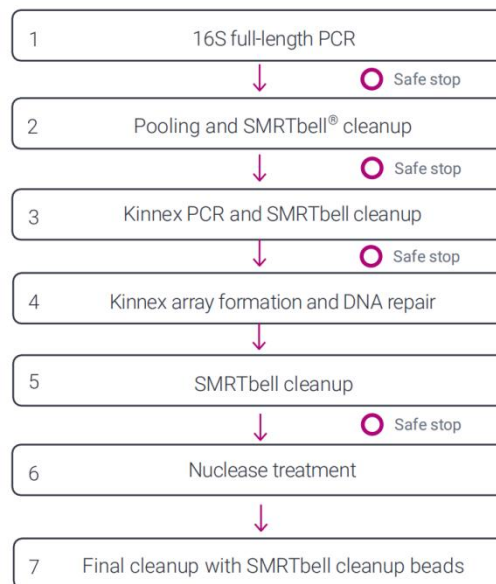
## Sequencing

### 1. Sample Quality Control

Methods of sample quality control refer to QC report.

### 2. Library Construction, Quality Control and Sequencing

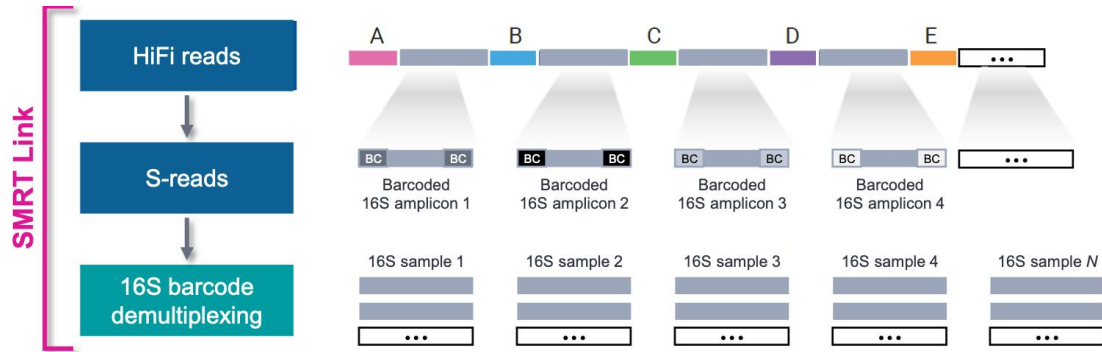
To prepare Kinnex libraries from 16S rRNA amplicons, the process starts with PCR amplification using barcoded primers. After pooling the barcoded 16S PCR amplicons, Kinnex adapters are added to facilitate the concatenation of these amplicons into longer sequences. Kinnex array formation, involves treating the Kinnex PCR products with Kinnex enzyme, ligase, and barcoded Kinnex terminal adapters to assemble the 16S segments into a linear array. Lastly, a final cleanup with SMRTbell cleanup beads is performed to ensure the library is clean and pure. The experimental procedures of DNA library preparation are shown:



**Workflow of library construction**

# Data analysis

## 1. Data Quality Control



Read segmentation: Use SMRT Link Read Segmentation data utility to split arrayed Kinnex library HiFi reads into individual segmented reads (S-reads) that represent the original barcoded 16S amplicon sequences.

Demultiplex barcodes: Use SMRT Link Demultiplex barcodes data utility to separate sequence reads by barcode.

## 2. ASVs Denoise and Species annotation

### 2.1 ASVs Denoise

For the clean data obtained previously, denoise is performed with DADA2 or deblur module in the QIIME2 software to obtain initial ASVs (Amplicon Sequence Variants) (default: DADA2), and then ASVs with abundance less than 5 were filtered out.

### 2.2 Species Annotation

Species annotation is performed using QIIME2 software. For 16S/18S, the annotation database is Silva Database, while for ITS, it is Unite Database.

### 2.3 Phylogenetic Relationship Construction

In order to study phylogenetic relationship of each ASV and the differences of the dominant species among different samples (groups), multiple sequence alignment is performed using QIIME2 software.

### 2.4 Data Normalization

The absolute abundance of ASVs is normalized using a standard of sequence number corresponding to the sample with the least sequences. Subsequent analysis of alpha diversity and beta diversity were all performed based on the output normalized data.

### 3. AlphaDiversity

In order to analyze the diversity, richness and uniformity of the communities in the sample, alpha diversity is calculated from 7 indices in QIIME2, including Observed\_otus, Chao1, Shannon, Simpson, Dominance, Good's coverage and Pielou\_e.

Three indices were selected to identify community richness:

Observed\_otus - the number of observed species ([http://scikit-bio.org/docs/latest/generated/skbio.diversity.alpha.observed\\_otus.html](http://scikit-bio.org/docs/latest/generated/skbio.diversity.alpha.observed_otus.html));

Chao - the Chao1 estimator (<http://scikit-bio.org/docs/latest/generated/skbio.diversity.alpha.chao1.html#skbio.diversity.alpha.chao1>);

Dominance - the Dominance index (<http://scikit-bio.org/docs/latest/generated/skbio.diversity.alpha.dominance.html#skbio.diversity.alpha.dominance>);

Two indices were used to identify community diversity:

Shannon - the Shannon index (<http://scikit-bio.org/docs/latest/generated/skbio.diversity.alpha.shannon.html#skbio.diversity.alpha.shannon>);

Simpson - the Simpson index (<http://scikit-bio.org/docs/latest/generated/skbio.diversity.alpha.simpson.html#skbio.diversity.alpha.simpson>);

One indice is used to calculate sequencing depth:

Coverage - the Good's coverage ([http://scikit-bio.org/docs/latest/generated/skbio.diversity.alpha.goods\\_coverage.html#skbio.diversity.alpha.goods\\_coverage](http://scikit-bio.org/docs/latest/generated/skbio.diversity.alpha.goods_coverage.html#skbio.diversity.alpha.goods_coverage));

One indice is used to calculate species evenness:

Pielou\_e - Pielou's evenness index ([http://scikit-bio.org/docs/latest/generated/skbio.diversity.alpha.pielou\\_e.html#skbio.diversity.alpha.pielou\\_e](http://scikit-bio.org/docs/latest/generated/skbio.diversity.alpha.pielou_e.html#skbio.diversity.alpha.pielou_e)).

## 4. Beta Diversity

In order to evaluate the complexity of the community composition and compare the differences between samples(groups), beta diversity is calculated based on weighted and unweighted unifrac distances in QIIME2.

Cluster analysis is performed with principal component analysis (PCA), which is applied to reduce the dimension of the original variables using the ade4 package and ggplot2 package in R software. Principal Coordinate Analysis (PCoA) is performed to obtain principal coordinates and visualize differences of samples in complex multi-dimensional data. A matrix of weighted or unweighted unifrac distances among samples obtained previously is transformed into a new set of orthogonal axes, where the maximum variation factor is demonstrated by the first principal coordinate, and the second maximum variation factor is demonstrated by the second principal coordinate, and so on. The three-dimensional PCoA results were displayed using QIIME2 package, while the two-dimensional PCoA results were displayed using ade4 package and ggplot2package in R software.

To study the significance of the differences in community structure between groups, the adonis and anosim functions in the QIIME2 software were used to do analysis. To find out the significantly different species at each taxonomic level (Phylum, Class, Order, Family, Genus, Species), the R software is used to do MetaStat and T-test analysis. The LEfSe software is used to do LEfSe analysis (LDA score threshold: 4) so as to find out the biomarkers.

Further, to study the functions of the communities in the samples and find out the different functions of the communities in the different groups, the PICRUST2 software is used for function annotation analysis.

## 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. 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.
3. LiMinjuan, Shao Dantong, Zhou Jiachen et al. Signatures within esophageal microbiota with progression of esophageal squamous cell carcinoma.[J] .*Chin J Cancer Res*, 2020, 32: 755-767.