### 样品测序数据处理方法：

### 1. 数据过滤

对原始的测序数据进行如下处理，获得Clean Data，具体步骤如下：  
1） 采取按窗口去低质量的方法，具体操作如下：设置25 bp的窗口，如果窗口平均质量值低于20，从窗口开始截去后端碱基；截短后read长度低于原始read长度75%的，去掉整条序列。  
2） 去除接头污染reads（默认adapter序列与read序列有15 bp的overlap，设置为15 bp，允许错配数为3）；  
3） 去除含N的reads；  
4） 去除低复杂度reads（默认reads中连续碱基长度≥10的为低复杂度reads ）。  
根据barcode和引物区分样本，barcode序列与测序reads比对允许的错配个数为0 bp

### 2. Tags连接

序列拼接使用软件FLASH（Fast Length Adjustment of Short reads，v1.2.11），利用重叠关系将双末端测序得到的成对reads组装成一条序列，得到高变区的Tags。拼接条件如下：  
1）最小匹配长度15 bp；  
2）重叠区域允许错配率为0.1。  
使用软件：FLASH（Fast Length Adjustment of Short reads，v1.2.11）

### 3. OTU聚类结果统计

OTU聚类有两种方法。Usearch:按照97%序列相似性聚类生成OTU；DADA2：通过去噪的序列以100%的相似度聚类来生成ASV序列，这里统称为OTU。  
Usearch:  
1．利用软件USEARCH（v7 .0.1090）将拼接好的 Tags聚类为OTU。其主要过程如下：  
1) 利用UPARSE在97 %相似度下进行聚类，得到OTU的代表序列；  
2) 利用UCHIME（v4.2.40）将PCR扩增产生的嵌合体从OTU代表序列中去除；  
(16S和ITS采取和已有的嵌合体数据库进行比对的方法去除嵌合体。18S采取De novo的方法去除嵌合体  
16S嵌合体数据库：gold database（v20110519）  
ITS嵌合体数据库：UNITE（v201407 03），分为ITS全长，ITS1和ITS2，按测序区域进行选择)  
3) 使用usearch\_global方法将所有Tags比对回OTU代表序列，得到每个样品的OTU的丰度统计表。  
2． DADA2：利用软件QIIME2中的DADA2（Divisive Amplicon Denoising Algorithm）方法去噪，获得Amplicon Sequence Variants (ASVs)，ASV为100%相似的序列。进而得到特征表（Feature，对ASV/ASV等的统称）。其主要过程如下：  
1) 利用qiime tools import导入过滤后的双端序列；  
2) 利用qiime dada2 denoise-paired命令将导入后的双端序列基于DADA2的方法构建特征表；  
3) 利用qiime tools export将特征表转换成可以直接查看的格式；

### ****Sample Sequencing Data Processing Methods**‌**

### **‌**1. Data Filtering**‌**

#### The raw sequencing data undergoes the following steps to obtain Clean Data:

‌****Sliding-window quality trimming****‌:

Set a 25 bp window; truncate trailing bases if the average quality score within the window falls below 20. Discard entire reads if post-trimming length is <75% of the original read.

‌****Adapter removal****‌:Default settings: 15 bp overlap between adapter and read sequences, allowing ≤3 mismatches.

‌****N-containing read removal****‌: Reads with ambiguous bases (N) are excluded.

‌****Low-complexity read removal****‌:Default threshold: Reads with ≥10 consecutive identical bases are classified as low-complexity.

‌****Sample demultiplexing****‌:Assign reads to samples based on barcode and primer matches (zero mismatches allowed for barcodes).

**‌**2. Tag Assembly**‌**

#### ‌****Software****‌: FLASH (Fast Length Adjustment of Short reads, v1.2.11)

‌****Parameters****‌:

Minimum overlap length: 15 bp.

Maximum mismatch rate in overlapping regions: 0.1.

**‌**3. OTU Clustering & Statistics**‌**

‌****Method 1: Usearch (97% similarity threshold)****‌

‌****Clustering****‌:Use UPARSE to cluster tags into OTUs at 97% similarity, generating representative sequences.

‌****Chimera removal****‌:UCHIME (v4.2.40) filters PCR chimeras from OTU representatives:

16S/ITS: Compare against gold (v20110519) and UNITE (v20140703) databases.

18S: De novo removal.

‌****Abundance mapping****‌:Map all tags back to OTU representatives using usearch\_global to generate OTU abundance tables.

‌****Method 2: DADA2 (100% similarity threshold)****‌

‌****Denoising****‌:Use QIIME2's DADA2 to generate Amplicon Sequence Variants (ASVs).

‌****Feature table construction****‌:Commands:

textCopy Code

qiime tools import --input-path filtered\_fastq

qiime dada2 denoise-paired

qiime tools export --output-path feature\_table

Output: Feature table (ASV/OTU counts).