sratoolkit安装及使用：https://blog.csdn.net/zx\_zhang01/article/details/120222212?spm=1001.2101.3001.6661.1&utm\_medium=distribute.pc\_relevant\_t0.none-task-blog-2%7Edefault%7ECTRLIST%7ERate-1.pc\_relevant\_default&depth\_1-utm\_source=distribute.pc\_relevant\_t0.none-task-blog-2%7Edefault%7ECTRLIST%7ERate-1.pc\_relevant\_default&utm\_relevant\_index=1

win+r：

cd e：

cd sratoolkit.3.0.0-win64

单独下载

prefetch 序列号

批量下载

bin\prefetch.exe --option-file C:\Users\11146\Desktop\SRR\_Acc\_List.txt 序列号文件的存放路径

.sra文件下载完成后，则可以运行cmd, 进入到相应的SRA文件，我这里以SRR000199为例。

执行：

fastq-dump SRR000199

即可将.sra转化为fastq格式

放excel

="fastq-dump --split-3 --gzip "&A1&".sra -O ./x"

补充一下安装qiime2的步骤，2022-08版本之后的qiime2不支持virtualbox虚拟机安装，那就装旧版本2022-02

下载地址：<https://data.qiime2.org/distro/core/2022.2>

首先装virtualbox：<https://www.virtualbox.org/>

解压core然后打开ovf文件导入虚拟机，cpu跟内存看自己电脑配置分配

安装完成后进入qiime2，密码qiime2

点击外部窗口上的“设备”，选择“安装增强功能”，虚拟内弹出的窗口内，点击“Run”，提示输入密码，输入“qiime2”即可开始安装，完成后敲击回车关闭窗口。restart重启

配置share文件夹，先在自己电脑的Windows系统里创建共享文件夹qiimeshare，在虚拟机中挂载共享文件夹，点击左侧Terminal打开终端窗口。

输入命令：mkdir share ；

然后再输入命令：sudo mount –t vboxsf qiimeshare share即可实现挂载。

设置虚拟机的双向剪贴板同步，这样可以在QIIME虚拟机系统与Windows系统之间执行复制黏贴快捷命令。

制作txt清单导入qiime2

每行的第一个字段是qiime应该使用的样本名sample-id （注意：必须加-，不然报错，这是双端33格式的要求），第二个字段是绝对文件路径absolute-filepath（注意：必须加-，不然报错，这是双端33格式的要求），第三个字段是读取方向direction。以#开头的行和空行将被忽略。sample-id里是你自己的样品名，absolute-filepath里$PWD是目前绝对路径，后面加上你的样品名（双端的话，会有1,2两组数据，记住要在压缩状态，即\*.fq.gz）。direction主要包括：forward和reverse，\*\_1.fq.gz为forward。

sample-id,absolute-filepath,direction

SRR10972199,$PWD/9/SRR10972199\_1.fastq.gz,forward

SRR10972199,$PWD/9/SRR10972199\_2.fastq.gz,reverse

Linux下安装aspera

1. 安装ascp（在conda下）

conda install -c hcc aspera-cli –y

是否安装成功

ascp -h

查找密匙

which ascp（会输出密匙路径，下面代码会用到）

验证是否存在（路径要更改）

ls ~/software/miniconda3/envs/test/etc/asperaweb\_id\_dsa.openssh

2.把ascp添加到.bashrc中

设置环境变量

echo """export PATH=\"/home/USERNAME/.aspera/connect/bin:\$PATH\" """ >> ~/.bashrc

在这里需要注意是/home/USERNAME这里要修改成你自己的用户路径。然后source一下.bashrc文件使新加的环境变量生效。

source ~/.bashrc

3.检查ascp是否安装成功

在命令行窗口输入ascp,出现下面提示信息，说明你已经安装成功了。

Usage: ascp [OPTION] SRC... DEST

SRC to DEST, or multiple SRC to DEST dir

SRC, DEST format: [[user@]host:]PATH

Display full usage: -h,--help

**ascp基本语法**

**[OPTION] 是参数设置，SRC 是远程资源路径，DEST是文件保存路径。若SRC为多个文件，则保存在DEST目录下**

**ascp [OPTION] SRC... DEST**

ascp -l 100M -P 33001 -QT -k 2 -i ~/.aspera/connect/etc/asperaweb\_id\_dsa.openssh era-fasp@fasp.sra.ebi.ac.uk:/vol1/fastq/SRR576/004/SRR5760814/SRR5760814.fastq.gz ./test.fq.gz

[https://www.ebi.ac.uk/ega/about/ftp-aspera](https://www.ebi.ac.uk/ega/about/ftp-aspera%EF%BC%8C)

<https://download.asperasoft.com/download/docs/ascp/2.7/html/index.html>

-v verbose mode 实时知道程序在干啥

-T 取消加密，否则有时候数据下载不了

-i 提供私钥文件的地址

-l 设置最大传输速度，一般200m到500m，如果不设置，反而速度会比较低，可能有个较低的默认值

-k 断点续传，一般设置为值1

-Q 一般加上它

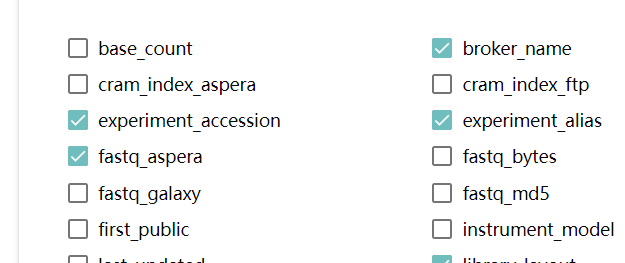
-P 提供SSH port，端口一般是33001

4.用ascp下载数据

ascp -v -k 1 -T -l 200m -i ~/.aspera/connect/etc/asperaweb\_id\_dsa.openssh

命令：

/home/qiime2/miniconda/envs/qiime2-2021.8/bin/ascp -QT -l 300m -P33001 -i /home/qiime2/miniconda/envs/qiime2-2021.8/etc/asperaweb\_id\_dsa.openssh [era-fasp@fasp.sra.ebi.ac.uk:/vol1/fastq/SRR109/051/SRR10918251/SRR10918251.fastq.gz](mailto:era-fasp@fasp.sra.ebi.ac.uk:/vol1/fastq/SRR109/051/SRR10918251/SRR10918251.fastq.gz) .（在ena下的时候注意gz后面空格+.）

勾选fastq-aspera，然后下载tsv/json文件，提取fastq文件下载地址

批量下载SRA文件，ascp命令提供参数--file-list，用于批量下载SRA文件。

建立SRA文件路径列表文件sra\_list.txt

/vol1/srr/SRR157/001/SRR1577021

/vol1/srr/SRR157/002/SRR1577022

ascp -T -i ~/.aspera/connect/etc/asperaweb\_id\_dsa.openssh -k 1 -l 200m --mode recv --host ftp-private.ncbi.nlm.nih.gov --user anonftp --file-list ~/sra/sra\_list.txt ~/sra/

cat run.txt |while read id   
do  
/home/qiime2/miniconda/envs/qiime2-2021.8/bin/ascp -QT -l 300m -P33001 -i  /home/qiime2/miniconda/envs/qiime2-2021.8/etc/asperaweb\_id\_dsa.openssh   era-fasp@$id  .  
done

/home/qiime2/miniconda/envs/qiime2-2021.8/bin/ascp -QT -l 300m -P33001 -i /home/qiime2/miniconda/envs/qiime2-2021.8/etc/asperaweb\_id\_dsa.openssh era-fasp@fasp.sra.ebi.ac.uk:/vol1/fastq/SRR109/041/SRR10972241/SRR10972241\_1.fastq.gz

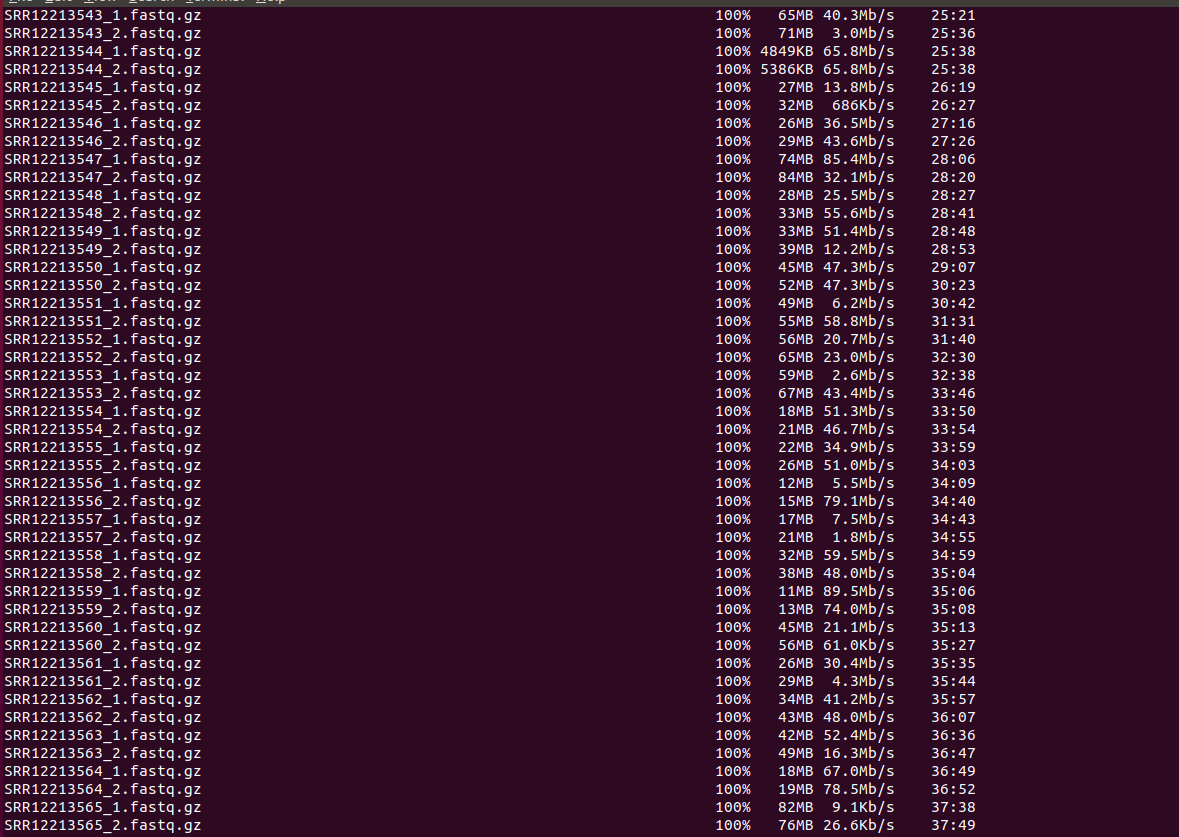
/home/qiime2/miniconda/envs/qiime2-2021.8/bin/ascp -QT -l 300m -P33001 -i /home/qiime2/miniconda/envs/qiime2-2021.8/etc/asperaweb\_id\_dsa.openssh --user era-fasp --mode recv --host fasp.sra.ebi.ac.uk --file-list /home/qiime2/Desktop/list.txt .

*#目的文件后要有指定下载路径，.代表下载到当前目录下*

list.txt模板

/vol1/fastq/SRR109/051/SRR10918251/SRR10918251.fastq.

/home/qiime2/miniconda/envs/qiime2-2021.8/bin/ascp -QT -l 300m -P33001 -i /home/qiime2/miniconda/envs/qiime2-2021.8/etc/asperaweb\_id\_dsa.openssh --user era-fasp --mode recv --host fasp.sra.ebi.ac.uk --file-list /home/qiime2/Desktop/list.txt \ /media/sf\_qiimeshare/fastq



**libstdc++.so.6: version `GLIBCXX\_3.4.26‘ not found**

**qiime2报错：**

**An unexpected error has occurred:**

**/home/qiime2/miniconda/envs/qiime2-2021.8/lib/python3.8/site-packages/pandas/\_libs/window/../../../../../libstdc++.so.6: version `GLIBCXX\_3.4.26' not found (required by /home/qiime2/miniconda/envs/qiime2-2021.8/lib/python3.8/site-packages/scipy/optimize/\_group\_columns.cpython-38-x86\_64-linux-gnu.so)**

**See above for debug info.**

su root进入权限

strings+路径+**libstdc++.so.6 |** | grep GLIBC 查看版本 **GLIBCXX\_**

sudo add-apt-repository ppa:ubuntu-toolchain-r/test

sudo apt update

sudo apt install gcc-9

sudo apt install libstdc++6

strings /usr/lib/x86\_64-linux-gnu/libstdc++.so.6 | grep GLIBCXX

**GLIBCXX\_3.4.26包下载地址:http://www.vuln.cn/9154**

**unzip +包解压**

**Linux命令**

**cd .. 返回上一级**

**返回上次所在的目录；cd –**

**Ctrl + C 终止任务**

设置root

sudo passwd root

su -进入root模式

ctrl+D退出root

**https://forum.qiime2.org/t/qiime2-chinese-manual/838/4**

1. **序列导入qiime2**

**一、sra格式转换为fastq.gz格式**

**二、fastq导入qiime2**

**1. 创建清单文件：txt、csv、tsv均可**

**sample-id,absolute-filepath,direction**

**SRR15182822,$PWD/SRR15182822.fastq.gz,forward（单端）**

**SRR4041257,$PWD/SRR4041257\_1.fastq.gz,forward（双端）**

**SRR4041257,$PWD/SRR4041257\_2.fastq.gz,reverse**

**2. 清单文件导入qiime2生成qza文件**

①单端序列导入：time qiime tools import --type 'SampleData[SequencesWithQuality]' --input-format SingleEndFastqManifestPhred33 --input-path PRJEB32568.txt --output-path PRJEB32568.qza

②双端序列导入：time qiime tools import --type 'SampleData[PairedEndSequencesWithQuality]' --input-path PRJNA382861.txt --output-path PRJNA382861.qza --input-format PairedEndFastqManifestPhred33

FNA文件：<http://qiime.org/tutorials/tutorial.html>

FNA去噪：<http://qiime.org/tutorials/denoising_454_data.html>

fna未对齐序列（未对齐的序列数据包含未对齐的DNA序列（即不包含-或.）的fasta格式文件））

qiime tools import \

--input-path sequences.fna \

--output-path sequences.qza \

--type 'FeatureData[Sequence]'

fna对齐序列（对齐的fasta格式，如下：有-字符，且等长）

qiime tools import \

--input-path aligned-sequences.fna \

--output-path aligned-sequences.qza \

--type 'FeatureData[AlignedSequence]'

**3. qza文件可视化**

time qiime demux summarize --i-data PRJEB32568.qza --o-visualization PRJEB32568.qzv

**4. 去噪/聚类生成特征表和代表性序列**

对应每个仪器看一下最大读取bp值，后面的截掉去

①单端序列去噪：time qiime dada2 denoise-single --i-demultiplexed-seqs manifest17.qza --p-trim-left 0 --p-trunc-len 388 --o-representative-sequences 17rep-seqs.qza --o-table 17table.qza --o-denoising-stats 17stats.qza --p-n-threads 0 --verbose

②双端序列去噪：time qiime dada2 denoise-paired --i-demultiplexed-seqs manifest9.qza --p-trim-left-f 0 --p-trim-left-r 0 --p-trunc-len-f 250 --p-trunc-len-r 241 --o-representative-sequences 9rep-seqs.qza --o-table 9table.qza --o-denoising-stats 9stats.qza --p-n-threads 24 --verbose

--p-n-threads 0(以零做为参数，能调动计算机所有核来进行运行程序，加快运算速度)

--verbose 显示分析进度

**figaro获得dada2截断位点**

wget http://john-quensen.com/wp-content/uploads/2020/03/figaro.yml  
conda env create -n figaro --file figaro.yml

wget https://github.com/Zymo-Research/figaro/archive/master.zip  
unzip master.zip   
mv figaro-master figaro  
cd figaro  
chmod 777 figaro.py

**运行figaro**

conda activate figaro   
python $PWD/figaro/figaro.py -i trim -o trim -f 16 -r 20 -a 300 -F zymo

输出文件示例

trimParameters.json  
forwardExpectedError.png  
reverseExpectedError.png  
要获取推荐的截断参数，请查看trimParameters.json:如下所示  
[  
    {  
        "trimPosition": [  
            141,  
            181  
        ],  
        "maxExpectedError": [  
            2,  
            3  
        ],  
        "readRetentionPercent": 87.24,  
        "score": 82.24249607692565  
    },  
  ]  
推荐的正向截断位置为141，  
推荐的反向截断位置为181。  
修剪和截断后，正向读取的预期错误数为2，  
反向读取的预期错误数为3,  
qiime2 dada2插件合并87.24％的读数

**vsearch对特征表进行聚类97%**

qiime vsearch cluster-features-de-novo \  
  --i-table table.qza \  
  --i-sequences rep-seqs.qza \  
  --p-perc-identity 0.97 \  
  --o-clustered-table table-dn-97.qza \  
  --o-clustered-sequences rep-seqs-dn-97.qza

给定一个特征表和相关的特征序列，根据用户指定的序列的百分比同一性阈值对特征进行聚类。 这不是一种通用的从头聚类方法，而是旨在用于对质量过滤/去重复方法（例如 DADA2）的结果进行聚类，或者用于以比原来更低的百分比身份重新聚类 FeatureTable 最初聚集在。 当输入表中的一组特征被聚类为单个特征时，该单个特征在给定样本中的频率是该样本中被聚类的特征的频率之和。

[技术贴 | 16S专题 | 基于QIIME2 dada2插件的16S扩增子测序数据的分析流程详解（中） (qq.com)](https://mp.weixin.qq.com/s?__biz=Mzg3NjcxNjU4Mw==&mid=2247539474&idx=1&sn=6627f60194ba430992a3436d73ae3dfb&source=41#wechat_redirect)http://www.360doc.com/content/21/0413/18/74684385\_972151461.shtml

**联合R语言进行可视化**

**一、导出特征表**

1. **导出特征表（qiime2）**

mkdir 23a

qiime tools export --input-path 23table.qza --output-path 23a

biom convert -i 23a/feature-table.biom -o 23a/otu\_table.tsv --to-tsv

cd 23a

sed -i '1d' otu\_table.tsv

sed -i 's/#OTU ID/ASV/' otu\_table.tsv

[微生物多样性qiime2分析流程(2）qiime2和dada2处理16s序列 - 哔哩哔哩 (bilibili.com)](https://www.bilibili.com/read/cv8073458) https://www.bilibili.com/read/cv8073458

1. **R更改特征表名称（在qiime2中输入R弹出R的程序）**

install.packages("pacman")

R：

library(pacman)

pacman::p\_load(tidyverse,magrittr,stringr)

otu <- "otu\_table.tsv" %>%

read.delim(check.names = FALSE,header = T,sep="\t")

rown <- paste0("ASV",seq\_len(nrow(otu)))

otu[,1] <- rown

colnames(otu)[1] <- paste0("ASV",colnames(data)[1])

write.table (otu,file ="otu\_table.tsv",sep ="\t",row.names = T)

（生成tsv文件后，用excel删除第一列，第一行后移一位）

这一步output为feature table，代表序列及每个sample的summary。这里feature table类似于之前我们用uparse算法聚类（把有97%相似性的序列聚集在一起成一个OTU，达到去重和聚类的目的）出OTU table类似，但是不论是DADA2还是Deblur， 都可以理解为用100%相似程度来聚类“OTU”，又称为sequence variants，或者sOTU，在qiime2的文档中，用ASV进行统一这些各种的叫法。OTU（Operational Taxonomic Unit ）即分类操作单元，是在系统发生学研究或群体遗传学研究中，为了便于进行分析，人为给某一个分类单元（品系，种，属，分组等）设置的同一标志。在微生物多样性分析中，根据不同的相似度水平，对所有序列进行OTU划分，一般情况下，如果序列之间的相似性高于97% （种水平）就可以把它定义为一个OTU，每个OTU代表一个物种。

为什么用ASV？

1、它比用97%相似程度聚类的“OTU”更加精准，在后续的alpha diversity及物种注释中会更加准确，因为传统意义上的OTU的97%的相似程度是在genus水平上大家认可的，但如果需要到species甚至strain水平的话，使用传统的OTU是很不靠谱的。

2、ASV相比于OTU最大的好处就是可以随时合并不同的数据跑出来的代表序列。因为使用的是100%的identidy，所以一样的序列就是一样的物种，不会受到不同数据、测序平台、文库建立、处理方法等等的误差影响。只要是qiime2上跑出来的数据，即使是不同时间不同人的结果，也可以直接合并，甚至不需要接触到原始的测序数据，也大大方便了大数据之间的整合。所以现在也越来越多的人倡导使用ASV以避免数据不可合并和难以合并的缺点。

**otu\_table.qza --> feature-table.biom ---> otu\_table.tsv**

**现在我们更改了otu\_table.tsv内容后再按照如下步骤将其还原回去：**

**otu\_table.tsv --> feature-table.biom ---> otu\_table.qza**

**3、qiime2将格式转化为qza**

cd 23a

biom convert -i otu\_table.tsv -o feature-table.biom --to-hdf5 --table-type="OTU table"

qiime tools import --input-path feature-table.biom --type 'FeatureTable[Frequency]' --input-format BIOMV210Format --output-path otu\_table-rename.qza

1. **导出代表序列**
2. **导出代表序列**

cd ..

qiime tools export --input-path 23rep-seqs.qza --output-path 23a

1. **更改代表序列名称**

qiime2：

cd 23a

less dna-sequences.fasta |paste - -|sed '1i ASVID,seq' > rep.fa

R：

library(pacman)

pacman::p\_load(tidyverse,magrittr,stringr)

rep <- "rep.fa" %>%

read.delim(check.names = FALSE, row.names = 1) %>%

set\_rownames(paste0(">ASV", seq\_len(nrow(.))))

write.table (rep,file ="rep.xls", sep ="\t", row.names = T)

Qiime2：

less rep.xls|sed '1d'|sed 's/"//g'|sed 's/\r//g'|tr "\t" "\n" > rep-seqs.fasta

1. **将代表序列转换成qza格式**

qiime tools import --type 'FeatureData[Sequence]' --input-path rep-seqs.fasta --output-path rep-seqs-rename.qza

**4、特征表统计**

qiime feature-table summarize --i-table otu\_table-rename.qza --o-visualization otu\_table-rename.qzv

**5、代表序列统计**

qiime feature-table tabulate-seqs --i-data rep-seqs-rename.qza --o-visualization raw.fq.list

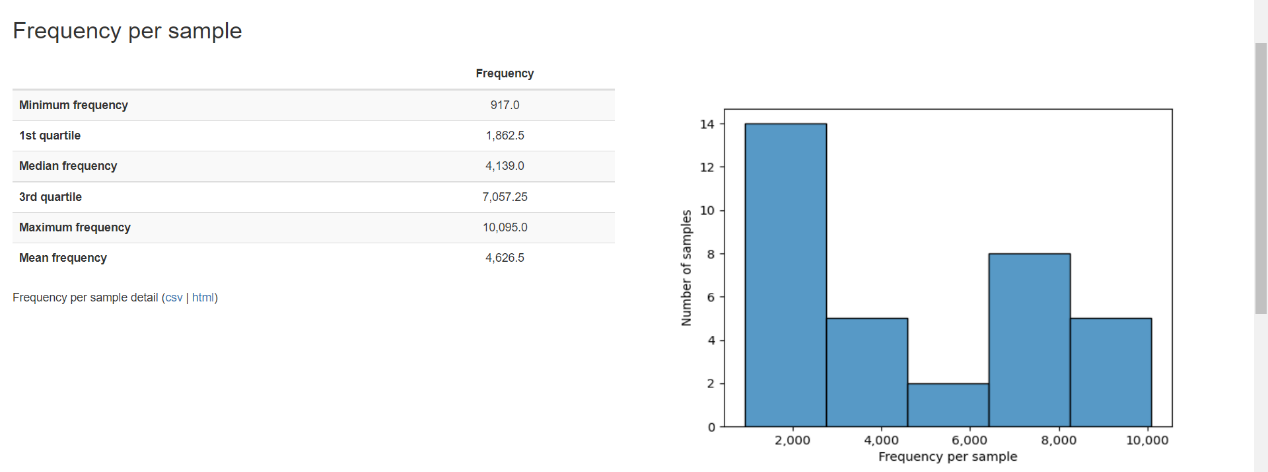
**6、质量过滤，方法参考上文**

**1、按数据量过滤table.qza**

基于总频率的过滤用于根据样本或特征在特征表中出现的频率进行过滤。

例如，在过滤样本时，可以使用此功能来过滤样本，其总频率在样本频率分布中是一个异常值。在许多16S研究中，一些样本只会获得少数（可能个位数至几百）条序列，这可能是由于样本生物量低导致DNA提取率低。在这种情况下，用户可能希望根据其最小总频率（即样本测序的序列总数）删除样本。可以通过以下方式实现（在本例中，总频率小于1500的样本将被过滤掉）。

**qiime feature-table filter-samples --i-table otu\_table-rename.qza --p-min-frequency 1500 --o-filtered-table otu\_table-rename-filtered.qza**

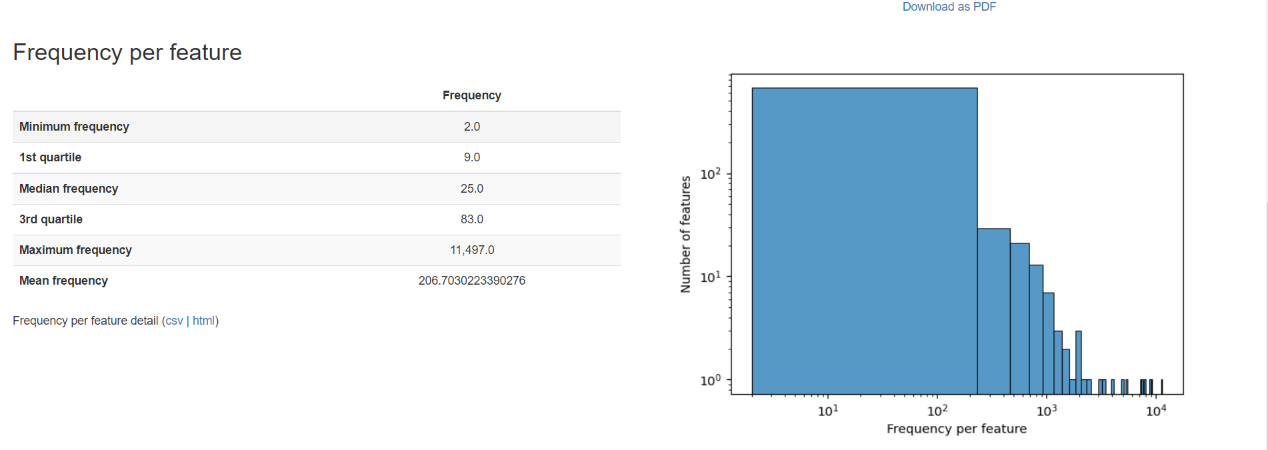


**（通常过滤前面1/4的otu，看1st quartile划分，一般范围在1000-4000左右）**

**2、按特征表的数据量过滤table.qza，只有特征序列总测序量大于10条以上的才保留**

详者注：(实验中会有大量低丰度的特征/OTU，它们会增加计算工作量和影响高丰度结果差异比较时FDR校正pvale，导致假阴性，通常需要选择一定的阈值进行筛选，常用的有相对丰度千分之五、千分之一、万分之五、万分之一；也可根据测序总量，reads频率的筛选阈值常用2、5、10等，大项目有时甚至超过100)

**qiime feature-table filter-features --i-table otu\_table-rename-filtered.qza --p-min-frequency 10 --o-filtered-table otu\_table-rename-filtered2.qza**



**（与上述同理，看1st quartile过滤前1/4，过滤完sample再过滤feature）**

**（差距不大一般取10）**

**3、代表序列过滤**

qiime feature-table filter-seqs --i-table otu\_table-rename-filtered2.qza --i-data rep-seqs-rename.qza --o-filtered-data rep-seqs-filtered.qza

**三、比对代表性序列，并构建系统发育树**

**#现在推荐用sepp构建系统发育树，做meta必须用sepp，不懂sepp可以看看这篇文献：**[https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5904434/](http://links.email.frontiersin.org/ls/click?upn=AAaFa03elZRFPXQ6ShiKwAAnDTaFgVINDYTrMcMBtBHOuqmYhbbm-2BohUh3KGuxQDYbNxChkM5eA31qDqjbLdQg-3D-3DSOgD_VMrn5ToY3axR-2BTypJZF71-2FXMMeJPUK8aGYoIRMsbv2h-2FwcI8dhf4hXxJPdUHmExIcBZHzfUYgputRgP3fMzBwaqG6pDw3r2M7CqXpY1yWGJtc6-2BZhbQkHa1w1lDx-2Fecma0aTP-2FmCBTooMATnGBqXlkqFF81s8FXqVxVMC72S-2B-2BCjSXqPFUM95ykC91gw3YJkuM9S13xlTb9P5tDh2w6Qv7X5Mzif-2B6x09UiQiqiIqYugOaqRJoRajFKvyf4RWklzlyRgtOqeToajqXB86daxV4yx4BmAv0dXiHVanjWl0N1Tc2DYELpzdlKaSoKKEz9l)

qiime phylogeny align-to-tree-mafft-fasttree --i-sequences rep-seqs-filtered.qza --o-alignment aligned-rep-seqs.qza --o-masked-alignment masked-aligned-rep-seqs.qza --o-tree unrooted-tree.qza --o-rooted-tree rooted-tree.qza

**多序列比对结果：aligned-rep-seqs.qza**

**过滤去除高变区后的多序列比对结果：masked-aligned-rep-seqs.qza**

**有根树，用于多样性分析：rooted-tree.qza**

**无根树：unrooted-tree.qza**

**###用sepp法构建系统发育树**

time qiime fragment-insertion sepp --i-representative-sequences rep-seqs-filtered.qza --i-reference-database /media/sf\_qiimeshare/sepp-refs-silva-128.qza --o-tree tree.qza --o-placements tree\_placements.qza --p-threads 6

tree.qza #此种方法中生成的树相当于有根树，可直接用于多样性分析。

tree\_placements.qza #插值法的树文件

**四、物种分类**

qiime feature-classifier classify-sklearn --i-classifier silva-138-99-nb-classifier.qza --i-reads rep-seqs-filtered.qza --o-classification taxonomy.qza

**（silva-138-99-nb-classifier.qza为数据库，**[**https://data.qiime2.org/2022.2/common/silva-138-99-nb-classifier.qza**](https://data.qiime2.org/2022.2/common/silva-138-99-nb-classifier.qza)**）**

**物种可视化：**qiime metadata tabulate  --m-input-file taxonomy.qza  --o-visualization taxonomy.qzv

**ps：用kraken2进行注释，又快又准**

**kraken2-build --db $DBNAME --special silva --threads 20 #silva建库**

**#把qza文件转回成fastq**

**qiime tools export \**

**--input-path xx.qza**

**--output-path clean\_data**

**cd clean\_data**

**mkdir fastq\_files**

**#双端**

**qiime demux emp-paired \**

**--i-seqs demux.qza \**

**--o-per-sample-sequences fastq\_files/ \**

**#单端换成 emp-single**

**kraken2 --db silva --threads 20 \**

**--paried /xx.R1.fq.gz /xx.R2.fq.gz --output xx.kraken2.result --report xx.kraken2.report**

**#bracken校正丰度**

**bracken --db silva --threads 20\#同样用silva数据库**

**-i xx.kraken2.report -o xx.bracken -w xx.bracken.report -r 250**

**#** **-r reads长度，查看下载的silva数据库知，这里可选50，75，100，150，200，250**

**#后面用krakentools脚本输出丰度表，详情看宏基因组kraken部分**

**五：文件转换导出**

**导出OTU表**

mkdir R分析

qiime tools export --input-path otu\_table-rename-filtered2.qza --output-path R分析

biom convert -i ./R分析/feature-table.biom -o ./R分析/otu\_table.tsv --to-tsv

（删掉第一行constuct。。，并将tsv文件转为txt格式）

**导出分类文件**

qiime tools export --input-path taxonomy.qza --output-path R分析

**导出有根进化树文件：**

qiime tools export --input-path rooted-tree.qza --output-path R分析

cd R分析

mv tree.nwk rooted\_tree.tre

mkdir R分析

qiime tools export --input-path otu\_table-rename-filtered2.qza --output-path b

biom convert -i ./b/feature-table.biom -o ./b/otu\_table.tsv --to-tsv

qiime tools export --input-path taxonomy.qza --output-path b

qiime tools export --input-path rooted-tree.qza --output-path b

cd b

mv tree.nwk rooted\_tree.tre

**https://blog.sciencenet.cn/blog-3334560-1229017.html**

**QIIME 2的多样性分析使用q2-diversity插件，该插件支持计算α和β多样性指数、并应用相关的统计检验以及生成交互式可视化图表。我们将首先应用core-metrics-phylogenetic方法，该方法将FeatureTableFrequency抽平到用户指定的测序深度，然后计算几种常用的α和β多样性指数，并使用Emperor为每个β多样性指数生成主坐标分析（PCoA）图。默认情况下计算的方法有：**

**一、α多样性 ：**

**①香农(Shannon’s)多样性指数：群落丰富度的定量度量，即包括丰富度richness和均匀度evenness两个层面）**

**②可观测的OTU (Observed OTUs，群落丰富度的定性度量，只包括丰富度）**

**③Faith’s系统发育多样性（包含特征之间的系统发育关系的群落丰富度的定性度量）**

**④均匀度Evenness（或 Pielou’s均匀度；群落均匀度的度量）**

**二、β多样性**

**①Jaccard距离（群落差异的定性度量，即只考虑种类，不考虑丰度）**

**②Bray-Curtis距离（群落差异的定量度量，较常用）**

**③非加权UniFrac距离（包含特征之间的系统发育关系的群落差异定性度量）**

**④加权UniFrac距离（包含特征之间的系统发育关系的群落差异定量度量）**

**需要提供给这个脚本的一个重要参数是–p-sampling-depth，它是指定重采样（即稀疏/稀疏rarefaction）深度。因为大多数多样指数对不同样本的不同测序深度敏感，所以这个脚本将随机地将每个样本的测序量重新采样至该参数值。例如，提供–p-sampling-depth 500，则此步骤将对每个样本中的计数进行无放回抽样，从而使得结果表中的每个样本的总计数为500。如果任何样本的总计数小于该值，那么这些样本将从多样性分析中删除。选择这个值很棘手。我们建议你通过查看上面创建的表table.qzv文件中呈现的信息并选择一个尽可能高的值（因此每个样本保留更多的序列）同时尽可能少地排除样本来进行选择。**

**查看QIIME 2的table.qzv 对象，尤其是交互式可视化表格。对于采样深度--p-sampling-depth，应该选择什么值呢？根据这个选择，分析中多少个样本将被排除？在core-metrics-phylogenetic命令中，你将分析的总序列是多少条呢？**

**译者注：下面多样性分析，需要基于重采样/抽平(rarefaction)标准化的特征表，标准化采用无放回重抽样至序列一致，如何设计样品重采样深度参数--p-sampling-depth呢？**

**如是数据量都很大，选最小的即可。如果有个别数据量非常小，去除最小值再选最小值。比如此分析最小值为917，我们选择1109深度重采样，即保留了大部分样品用于分析，又去除了数据量过低的异常值。本示例为近10年前测序技术的通量水平，454测序时代抽平至1000条即可，现在看来数据量很小。目录一般采用HiSeq2500或NovaSeq6000的 PE250模式测序，数据量都非常大，通常可以采用3万或5万的标准抽平，仍可保留90%以上样本。过低或过高一般结果也会波动较大，不建议放在一起分析。**

**https://blog.csdn.net/JiaqianLi/article/details/106832075?spm=1001.2014.3001.5502**

**###Alpha & Beta diversity analyse**

**#Calculate core-metric-phylogenetic 计算核心多样性**

qiime diversity core-metrics-phylogenetic \

--i-phylogeny rooted-tree.qza \

--i-table table.qza \

--p-sampling-depth 11562 \

--m-metadata-file sample-metadata.tsv \

--output-dir core-metrics-results

**#Alpha多样性组间显著性分析和可视化**

qiime diversity alpha-group-significance \

--i-alpha-diversity core-metrics-results/faith\_pd\_vector.qza \

--m-metadata-file sample-metadata.tsv \

--o-visualization core-metrics-results/faith-pd-group-significance.qzv

qiime diversity alpha-group-significance \

--i-alpha-diversity core-metrics-results/evenness\_vector.qza \

--m-metadata-file sample-metadata.tsv \

--o-visualization core-metrics-results/evenness-group-significance.qzv

**我们将使用PERMANOVA方法（在Anderson 2001年的文章中首次描述）beta-group-significance分析分类型元数据的样本组间差异**

time qiime diversity beta-group-significance \

 --i-distance-matrix core-metrics-results/unweighted\_unifrac\_distance\_matrix.qza \

 --m-metadata-file sample-metadata.tsv \

 --m-metadata-column body-site \

 --o-visualization core-metrics-results/unweighted-unifrac-body-site-significance.qzv \

 --p-pairwise

# 7s，多种或多样本时计算量指数增长

time qiime diversity beta-group-significance \

 --i-distance-matrix core-metrics-results/unweighted\_unifrac\_distance\_matrix.qza \

 --m-metadata-file sample-metadata.tsv \

 --m-metadata-column subject \

 --o-visualization core-metrics-results/unweighted-unifrac-subject-group-significance.qzv \

 --p-pairwise

# 6s，多种或多样本时计算量指数增长

**不同部分组内和组间差异显著性分析，采用箱线图+统计表呈现**

qiime emperor plot \

--i-pcoa core-metrics-results/unweighted\_unifrac\_pcoa\_results.qza \

--m-metadata-file sample-metadata.tsv \

--p-custom-axes days-since-experiment-start \

--o-visualization core-metrics-results/unweighted-unifrac-emperor-days-since-experiment-start.qzv

qiime emperor plot \

--i-pcoa core-metrics-results/bray\_curtis\_pcoa\_results.qza \

--m-metadata-file sample-metadata.tsv \

--p-custom-axes days-since-experiment-start \

--o-visualization core-metrics-results/bray-curtis-emperor-days-since-experiment-start.qzv

**#稀释曲线Alpha rarefaction plotting (rarefaction curves)**

time qiime diversity alpha-rarefaction --i-table otu\_table-rename-filtered2.qza --m-metadata-file HBV1group.txt --p-metrics 'shannon' 'simpson' 'observed\_features' 'chao1' 'goods\_coverage' --p-max-depth 25000 --p-steps 25 --o-visualization alpha-rarefaction-curves.qzv

[Qiime2+Origin绘制稀释曲线\_环微分析的博客-CSDN博客\_稀疏曲线](https://blog.csdn.net/HUANWEIFENXI/article/details/124127629?spm=1001.2101.3001.6650.1&utm_medium=distribute.pc_relevant.none-task-blog-2%7Edefault%7ECTRLIST%7ERate-1-124127629-blog-119565852.pc_relevant_default&depth_1-utm_source=distribute.pc_relevant.none-task-blog-2%7Edefault%7ECTRLIST%7ERate-1-124127629-blog-119565852.pc_relevant_default&utm_relevant_index=2)

**#Faith系统发育多样性可视化**

qiime diversity alpha-group-significance \

--i-alpha-diversity core-metrics-results/faith\_pd\_vector.qza \

--m-metadata-file sample-metadata.tsv \

--o-visualization core-metrics-results/faith-pd-group-significance.qzv

**#Evenness 均匀度可视化（每个ASV丰度的均匀度）**

qiime diversity alpha-group-significance \

--i-alpha-diversity core-metrics-results/evenness\_vector.qza \

--m-metadata-file sample-metadata.tsv \

--o-visualization core-metrics-results/evenness-group-significance.qzv

**#Shannon diversity可视化（综合Evenness和Richness的指标）**

qiime diversity alpha-group-significance \

--i-alpha-diversity core-metrics-results/shannon\_vector.qza \

--m-metadata-file sample-metadata.tsv \

--o-visualization core-metrics-results/shannon\_vector.qzv

**#Richness可视化**

qiime diversity alpha-group-significance \

--i-alpha-diversity core-metrics-results/observed\_features\_vector.qza \

--m-metadata-file sample-metadata.tsv \

--o-visualization core-metrics-results/observed\_features\_vector.qzv

**https://blog.csdn.net/JiaqianLi/article/details/119775217**