

使用 QIIME 2 2025.4 分析微生物组 16S

rDNA 基因和 ITS 扩增子测序数据

Using QIIME 2 2025.4 to Analyze Microbiome 16S rDNA Gene and ITS Amplicon Sequencing Data

杨海飞^{1,2}, 曾美尹², 高云云³, 陈同^{4,*}, 刘永鑫^{2,*}

¹生命科学学院, 青岛农业大学, 青岛

²农业农村部基因数据分析重点实验室, 中国农业科学院深圳农业基因组研究所, 深圳

³北京林业大学, 生态与自然保护学院, 北京

⁴中国中医科学院, 中药资源中心, 北京

*通讯作者邮箱: chentong@nrc.ac.cn; liuyongxin@caas.cn

摘要

QIIME 系列是目前微生物组扩增子分析领域使用最广泛的流程, 其论文在发表后十五年被引用超 7 万次。尽管 QIIME 第一版在微生物组数据分析领域具有里程碑意义, 但随着数据分析的快速发展, 已难以满足当前需求。全新的 QIIME 2 采用 Python 3 编写, 结合最新算法, 提供交互式图表和强大的可扩展性插件, 适应大数据和可重复分析的要求。然而, QIIME 2 无法在主流 Windows 系统下直接运行, 且用户使用文档长达十万多字, 对缺少生物信息背景的研究人员来说仍具巨大挑战。本文介绍了 QIIME 2 的安装方法和标准分析流程, 方便同行快速上手; 同时, 文章解读了中间步骤和参数, 指导用户掌握参数优化的方法, 以获得更合理的结果; 此外, 文中总结了软件安装和使用过程中的常见问题和解决方案, 包括数据导入导出、特征表生成、alpha 和 beta 多样性分析、物种组成分析、差异物种分析以及数据可视化等。本文提供配套视频、分析代码、测序数据和预期结果, 方便同行学习和复现分析过程。

关键词: 微生物组, 扩增子, QIIME 2, rDNA, 可视化

引用格式: 杨海飞, 曾美尹, 高云云, 陈同, 刘永鑫 (2025). 使用 QIIME 2 2025.4 分析微生物组 16S rDNA 基因和 ITS 扩增子测序数据. *Bio-101* e1011032. Doi: 10.21769/BioProtoc.1011032.

How to cite: Yang, H. F., Zeng, M. Y., Gao, Y. Y., Chen, T. and Liu, Y. X. (2025). Using QIIME 2 2025.4 to Analyze Microbiome 16S rDNA Gene and ITS Amplicon Sequencing Data. *Bio-101* e1011032. Doi: 10.21769/BioProtoc.1011032. (in Chinese)

Copyright: © 2025 The Authors; exclusive licensee Bio-protocol LLC.

研究背景

微生物无处不在，从自然环境到极端环境，几乎生态系统都能发现它们的身影。微生物组是特定环境或生态系统中所有微生物及其遗传信息的总和[1]，它不仅深刻影响着环境和生态平衡[2]，还与人类健康[3]、农业发展[4]和环境保护[5]等方面密切相关。近年来，微生物组研究已成为生命科学领域的热点，相关研究文献的数量持续激增，反映了这一领域的重要性[6]和广泛关注。在微生物组研究中，扩增子分析作为一种高效且经济的技术，因其特定基因片段（如 16S rDNA）进行分析的特性，在揭示微生物群落结构和多样性方面表现尤为突出。凭借成本低、特异性强、操作简便等优势[7]，扩增子分析已被广泛应用于环境微生物监测、医学微生物学、农业科学等多个领域[8]。

QIIME (Quantitative Insights Into Microbial Ecology) [9]是扩增子数据分析的经典工具，自发表以来就因其易用性和高效性深受研究人员的青睐。然而，随着高通量测序技术的发展，数据体量迅速增加，传统的 QIIME 已难以满足现代研究对计算性能和功能复杂性的要求。为了应对这些挑战，QIIME 2 [10]应运而生。

QIIME 2 在功能上全面升级，不仅保留了传统 QIIME 的优势，还提供了更强大的工具和流程来处理扩增子数据。其交互性分析功能和可视化能力显著提升了用户体验，同时保证了研究的可重复性。然而，由于 QIIME 2 涉及的生物信息学知识较多，对缺乏经验的新手而言，仍存在一定的学习门槛。

本研究旨在提供一套基于 QIIME 2 2025.4 版本的标准化实验方案，针对微生物组 16S rDNA 基因扩增子测序数据的处理与分析。同时，该实验方案也属于 EasyAmplicon 扩增子分析流程[11]的一部分，EasyAmplicon 的分析结果可导入 QIIME 2 2025.4 进行后续分析。该方案的核心原理包括以下几个关键步骤：

1. 数据导入

将原始测序数据以 QIIME 2 支持的格式导入，以确保数据兼容性和分析的顺畅。

2. 去噪与特征生成

使用 DADA2[12]等方法去除测序误差，生成高分辨率的特征 (ASV, Amplicon Sequence Variants)。ASV 是单碱基分辨率的序列单位，比传统的 OTU (Operational Taxonomic Units) 分类方法更准确，能够反映微生物群落的真实组成。

3. 多样性分析

通过计算 α 多样性（如香农指数、观察物种数）和 β 多样性（如 PCoA 和 NMDS 分析）量化微生物群落的多样性和相似性，揭示不同样本之间的生态学差异。

4. 物种注释

利用分类学数据库（如 SILVA 或 Greengenes），对 ASV 进行物种注释，解析群落组成，并挖掘潜在的功能信息。

5. 差异分析

引用格式：杨海飞，曾美尹，高云云，陈同，刘永鑫 (2025). 使用 QIIME 2 2025.4 分析微生物组 16S rDNA 基因和 ITS 扩增子测序数据. *Bio-101* e1011032. Doi: 10.21769/BioProtoc.1011032.

How to cite: Yang, H. F., Zeng, M. Y., Gao, Y. Y., Chen, T. and Liu, Y. X. (2025). Using QIIME 2 2025.4 to Analyze Microbiome 16S rDNA Gene and ITS Amplicon Sequencing Data. *Bio-101* e1011032. Doi: 10.21769/BioProtoc.1011032. (in Chinese)

通过统计学方法（如 ANCOM 或 LEfSe），识别在不同实验组之间具有显著差异的微生物种类，探索其与环境因素或健康状态的关联。

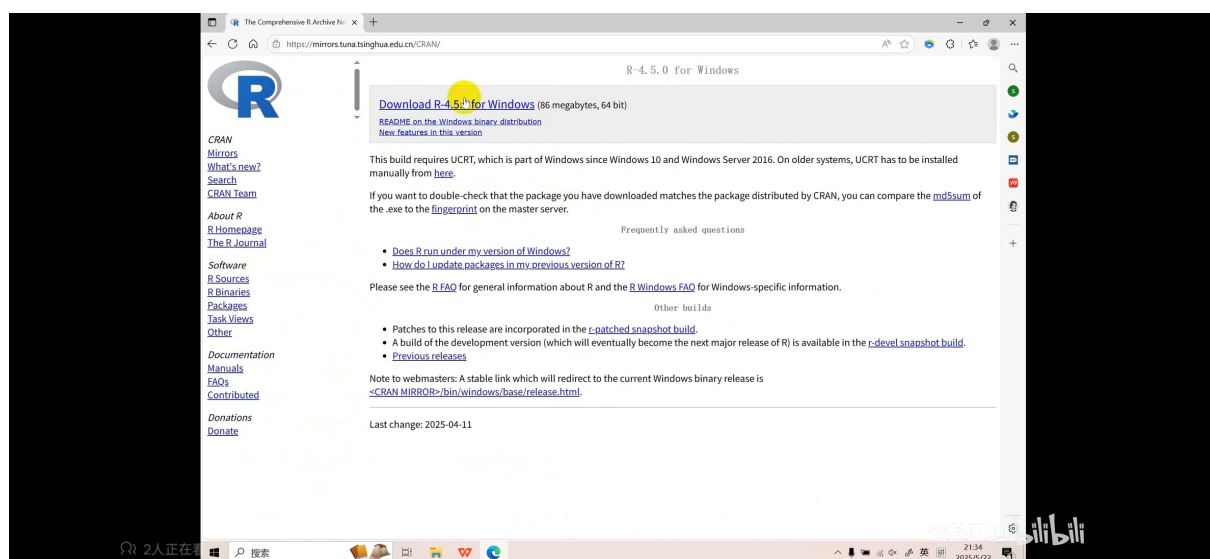
通过以上步骤，本方案不仅能够高效完成扩增子数据的标准化处理，还为研究者提供了直观的结果可视化和交互式探索工具。本文还附有操作视频，对代码实现和关键分析步骤进行详细讲解，以帮助新手快速掌握 QIIME 2 的使用方法，为同行提供便捷的学习和参考资源。

仪器设备

1. （可选）推荐使用计算服务器（操作系统：Linux 主流发行版本，如 CentOS 8+/Ubuntu 22.04+；CPU：4 核+；内存：16G+；硬盘：> 10 GB，且大于原始数据大小 3 倍），网络访问畅通。
2. 个人电脑推荐 Windows 10 系统，内存 8G+。需先在应用商店中安装 Linux 子系统（如 Ubuntu 24.04 LTS），然后安装 QIIME 2 2025.4；也可使用 VirtualBox 虚拟机运行 QIIME 2 镜像，但效率较低不推荐使用；Mac 系统可直接安装 QIIME 2 2025.4。（可选）Windows 用户远程访问服务器需安装 XShell 8.0+ 或 Putty 等终端类软件，Mac 使用系统内置终端即可远程访问计算服务器。

软件和数据库

软件及数据库介绍（视频 1）



视频 1. 软件及数据库介绍 (<https://www.bilibili.com/video/BV1u47nzTEax/>)

引用格式：杨海飞，曾美尹，高云云，陈同，刘永鑫 (2025). 使用 QIIME 2 2025.4 分析微生物组 16S rDNA 基因和 ITS 扩增子测序数据. *Bio-101* e1011032. Doi: 10.21769/BioProtoc.1011032.

How to cite: Yang, H. F., Zeng, M. Y., Gao, Y. Y., Chen, T. and Liu, Y. X. (2025). Using QIIME 2 2025.4 to Analyze Microbiome 16S rDNA Gene and ITS Amplicon Sequencing Data. *Bio-101* e1011032. Doi: 10.21769/BioProtoc.1011032. (in Chinese)

1. QIIME 2 可运行的四种环境任选其一：Linux 服务器（推荐，适合大数据）、Windows 10 子系统 Ubuntu 24.04 LTS（推荐，适合小数据）、Windows 中 VirtualBox 虚拟机中运行 Ubuntu 24.04 LTS（不推荐，小数据集且效率低）、Mac 系统（不推荐，兼容性问题较多）
2. 软件管理器 Miniconda3 Linux 64-bit (Python 3.10+)： <https://docs.anaconda.net.cn/miniconda/miniconda-other-installer-links/>
3. QIIME 2 [10]发行版 2025.4： <https://docs.qiime2.org/2025.4/index.html>
Greengenes 2 [13] 分类器（参考序列和物种分类信息）：
https://ftp.microbio.me/greengenes_release/2022.10/2022.10.backbone.full-length.fna.qza
https://ftp.microbio.me/greengenes_release/2022.10/2022.10.backbone.tax.qza
4. 流程示例参考代码和结果文件详见：
https://github.com/YongxinLiu/EasyAmplicon/blob/master/qiime2/pipeline_qiime2.sh
5. (可选) 远程文件传输工具 WinSCP 客户端 6.5： <https://winscp.net/eng/download.php>
6. (可选) Windows 远程访问服务器终端工具 Xshell 8.0： <https://www.xshell.com/zh/xshell/>
7. (可选) R 语言环境，下载适合自己系统的安装包（版本：4.4.3+）： <https://www.r-project.org/>，本次演示最新版 R4.5.0 下载安装。
8. (可选) R 语言开发环境 RStudio, 用于执行流程，下载适合自己系统的安装包（版本：2024.12.1+563）：
<https://posit.co/download/rstudio-desktop/>

软件安装和数据库部署

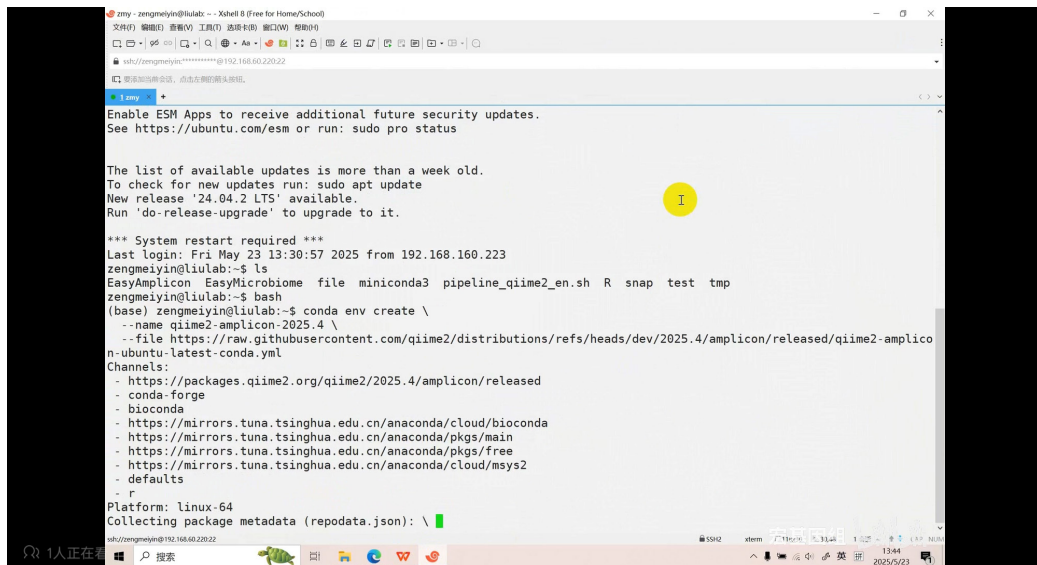
QIIME 2 不支持在 Windows 系统下直接安装。我们主要介绍远程访问 Linux 服务器和 Windows 10 下安装 Linux 子系统并使用 QIIME 2 的两种方法，任选其一即可。

方法 1. 远程访问 Linux 服务器：Windows/Mac 用户安装 WinSCP 客户端，用于上传测序数据至服务器或数据中心，也可用于下载分析结果本地查看。Windows 用户安装 Xshell 用于远程访问服务器并开展分析，Mac 用户可使用系统自带 Terminal 中的 ssh 命令远程访问服务器。

方法 2. Windows 10 的 1609 以后的版本可以安装 Linux 子系统：开始→Microsoft Store→搜索“Ubuntu”→选择“Ubuntu 24.04 LTS”→安装。安装前的系统设置和常见问题请阅读《[Windows10 安装 Linux 子系统 Ubuntu 20.04LTS](#)》。安装成功后可以在开始中启动“Ubuntu 24.04 LTS”的命令行，也可选在 RStudio 中设置默认 Terminal 为“Bash (Windows Subsystem for Linux)”，打开新终端即可使用。

在 Linux 系统下，以 Miniconda3 软件和 Python3 虚拟环境安装 QIIME 2 流程；然后下载 16S rRNA 基因数据库，建立物种分类器（视频 2）。

注：下文代码行添加灰色底纹背景，其中需要根据系统环境修改的部分标为蓝色。



视频 2. 安装 QIIME 2 2025.4 (<https://www.bilibili.com/video/BV1w67nztESG/>)

1. 安装 Miniconda3 Linux 64-bit (已安装请跳过)

```
wget -c https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
bash Miniconda3-latest-Linux-x86_64.sh
~/miniconda3/bin/conda init
```

2. Conda 新建环境安装 QIIME 2 (原始链接)

```
conda env create \
--name qiime2-amplicon-2025.4 \
--file
https://raw.githubusercontent.com/qiime2/distributions/refs/heads/dev/2025.4/amplicon/released/qiime2-amplicon-ubuntu-latest-conda.yml
```

备用链接

```
wget -c http://www.imeta.science/db/qiime2/qiime2-amplicon-ubuntu-latest-conda.yml

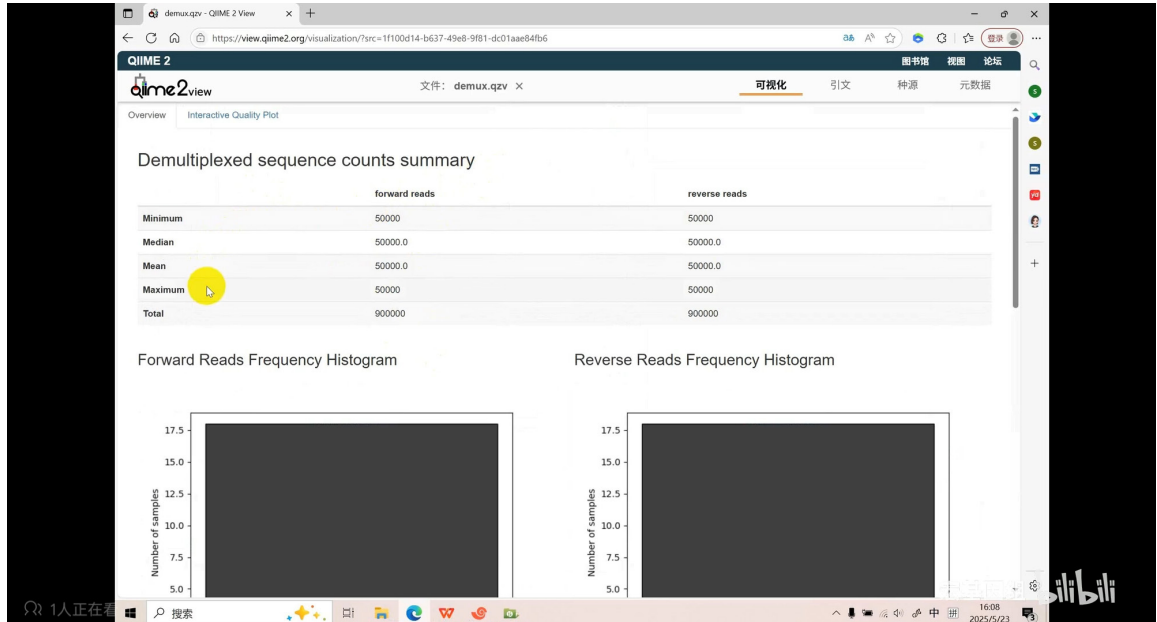
conda env create \
--name qiime2-amplicon-2025.4 \
--file qiime2-amplicon-ubuntu-latest-conda.yml
```

检查是否安装成功

```
conda activate qiime2-amplicon-2025.4
```


实验步骤

QIIME 2 分析流程演示和讲解见视频 3 和视频 4。



视频 3. 数据导入、特征表、Alpha 和 Beta 多样性分析 (<https://www.bilibili.com/video/BV1Cs7WzNEVQ/>)

Feature ID	Taxon	Confidence
005447134fa2808c8d15bd73dc76c0d	d__Bacteria; p__Actinobacteriota; c__Actinomycetia; o__Mycobacteriales; f__Pseudonocardiales; g__Pseudonocardia	0.9997027022154275
0066304c258bb8f503a5007837c289	d__Bacteria; p__Verrucomicrobiota; c__Verrucomicrobia; o__Chthoniobacteriales; f__Chthoniobacteraceae; g__VFJQ01	0.9943227905630342
0091167b8dc904da47b754475194893	d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Burkholderiales; f__Burkholderiaceae; g__Burkholderia; s__Burkholderia	0.9999999999999999
00993b7fa8d72adef0c02c360de94	d__Bacteria; p__Actinobacteriota; c__Actinomycetia; o__Mycobacteriales; f__Micromonosporaceae; g__Virgisporangium; s__Virgisporangium	0.988472114014454
01061ad5b34963006479b1620363c5c	d__Bacteria; p__Planctomycetia; c__Planctomycetia; o__Planctomycetia; f__Planctomycetia; g__Planctomycetia; s__Planctomycetia	0.994251060791022
0152071386f540302d90e3d4d936d	d__Bacteria; p__Actinobacteriota; c__Thermoleptophila; o__Gaiellales; f__Gaiellales; g__Palsa-739; s__Palsa-739	0.9347281949854271
01d1c2b1a2eb16fc26499343847e67e	d__Bacteria; p__Chlamydia; c__Chlamydia; o__Chlamydia; f__Chlamydia; g__Chlamydia; s__Chlamydia	0.999997963663047
01e26c95617448e11eeddd576a118d59	d__Bacteria; p__Verrucomicrobiota; c__Verrucomicrobia; o__Chthoniobacteriales; f__Chthoniobacteraceae; g__VFJQ01; s__VFJQ01	0.993882209393236
021c5b891400b6d9b57112b2890f0c	d__Bacteria; p__Bacteroidia; c__Bacteroidia; o__Flavobacteriales; f__Weeksellaceae; g__Chryseobacterium; s__Chryseobacterium	0.9999999498292921
021cbbf00a9d9d5ae24eab8b5004ef1	d__Bacteria; p__Firmicutes; c__Bacilli; o__Paenibacillales; f__NBRC-103111; g__Paenibacillus; s__Paenibacillus	0.999824352273753
02a2382cb1e49db9b11889a4c450a2ae	d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Burkholderiales; f__Burkholderiaceae; g__Burkholderia; s__Burkholderia	0.7129074692533067
03972f450c97cab485bf121801bb8a9	d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Xanthomonadales; f__Xanthomonadaceae; g__Xanthomonas; s__Xanthomonas	0.9532728544660587
03c9cc02356700beeb8459f61e24569	d__Bacteria; p__Bacteroidia; c__Bacteroidia; o__Chitinophagales; f__Saprospiraceae; g__OLB9; s__OLB9	0.9999191480703169
03d3e7842b0834ab8144cd7c35598bb	d__Bacteria; p__Chloroflexia; c__Chloroflexia; o__Chloroflexales; f__Roseiflexaceae; g__Kousethrix; s__Kousethrix	0.8499999177011435

视频 4. 物种组成、差异比较 (<https://www.bilibili.com/video/BV1Ps7WzKETS/>)

引用格式: 杨海飞, 曾美尹, 高云云, 陈同, 刘永鑫 (2025). 使用 QIIME 2 2025.4 分析微生物组 16S rDNA 基因和 ITS 扩增子测序数据. *Bio-101* e1011032. Doi: 10.21769/BioProtoc.1011032.

How to cite: Yang, H. F., Zeng, M. Y., Gao, Y. Y., Chen, T. and Liu, Y. X. (2025). Using QIIME 2 2025.4 to Analyze Microbiome 16S rDNA Gene and ITS Amplicon Sequencing Data. *Bio-101* e1011032. Doi: 10.21769/BioProtoc.1011032. (in Chinese)

开始分析前，我们应处于项目所在目录（如 windows 下 C 盘的 qiime2 目录为/mnt/c/qiime2） ， 并启动软件所在的 Conda 环境。

```
mkdir -p qiime2-2025.4
cd qiime2-2025.4
conda activate qiime2-amplicon-2025.4
```

1. 准备和导入元数据和测序数据

下载示例元数据供编写自己课题对应的元数据时作为参考

```
wget -c http://www.imeta.science/db/qiime2/metadata.txt
```

通常测序公司会返回原始数据，如 Illumina 双端测序的文件，每个样本有一对文件。本文使用的数据来自发表于 Science 杂志关于拟南芥根系微生物组研究的文章[14]，GSA 项目号为 PRJCA001296。为方便演示流程的使用，我们从中选取三个组（每组包括 6 个生物学重复共 18 个样本），并且随机抽取了 50,000 对序列作为本教程的测试数据，该数据可以从中国科学院基因组研究所的原始数据归档库（Genome Sequence Archive, GSA, <https://bigd.big.ac.cn/gsa/>） [15]中按批次编号 CRA002352 搜索并手动逐个下载至 seq 目录。可选使用 awk 语言配合 wget 命令根据样本元数据中批次和样本编号批量下载至 seq 目录，代码如下。

```
mkdir -p seq
awk '{system("wget -c ftp://download.big.ac.cn/gsa/"$5"/"$6"/"$6"_f1.fq.gz
-O seq/"$1"_1.fq.gz")}' <(tail -n+2 metadata.txt)
awk '{system("wget -c ftp://download.big.ac.cn/gsa/"$5"/"$6"/"$6"_r2.fq.gz
-O seq/"$1"_2.fq.gz")}' <(tail -n+2 metadata.txt)
```

awk 为 Linux 下的一种字符处理语言，可同时使用文本中的多个字段；使用 system 命令调用 wget，实现根据列表批量下载、改名的功能。

检查文件大小，确定是否下载完整或正常。

```
ls -lsh seq
```

接下来根据样本名和测序文件位置手动编写样本文件的索引列表

（manifest），格式参考流程目录中示例 manifest 文件。可选使用 awk 语言根据 metadata 编写自动生成 manifest 文件，实现全流程的可重复计算。

```
awk 'NR==1{print "sample-id\tforward-absolute-filepath\treverse-absolute-
filepath"} \
NR>1{print $1"\t"$PWD/seq/"$1"_1.fq.gz\t"$PWD/seq/"$1"_2.fq.gz"}' \
metadata.txt > manifest
```

数据导入 qiime2

```
qiime tools import \
--type 'SampleData[PairedEndSequencesWithQuality]' \
--input-path manifest \
--output-path demux.qza \
--input-format PairedEndFastqManifestPhred33V2
```

导入 1G 大小的 fq 文件用时需 7 分钟，本文中使用测试数据仅需 34 秒。

注：QIIME 2 的起始文件为每个样本 1 个或 1 对文件。对于混池测序未拆分样本的原始数据，需要使用 QIIME[9] 中的脚本进行拆分，或要求测序服务商提供拆分后的单样本 fastq 格式测序文件。

通过运行 summarize 命令来查看数据的质量（表 1）

```
qiime demux summarize \
--i-data demux.qza \
--o-visualization demux.qzv
```

表 1. 拆分序列汇总计数(Demultiplexed sequence counts summary)

分位数	正向	反向
最小值 (Minimum)	50000	50000
中位数 (Median)	50000.0	50000.0
平均值 (Mean)	50000.0	50000.0
最大值 (Maximum)	50000	50000
总计 (Total)	900000	900000

2. 生成特征表和代表序列

DADA2 是基于 R 语言编写的扩增子分析流程，可以实现扩增子序列去除测序噪音、错误和嵌合体，并挑选扩增序列变体 (amplicon sequence variant, ASV) 和生成特征表 (feature table) 的功能[12]。dada2 denoise-paired 方法需要两个用于质量过滤的参数：--p-trim-left，截取左端低质量序列，用于切除低质量序列、barcode 或引物，本文--p-trim-left 参数根据示例数据引物长度设置；--p-trunc-len，序列截取长度，去除右端低质量序列，本文示例数据质量较好，所以--p-trunc-len 参数设置为 0。具体参数应根据序列引物长度、标签序列和质量高低进行设置。

支持多线程加速，如测试平台 96 线程 (p) 环境下，可以使用--p-n-threads 参数指定。不同线程下的计算时间（以分钟为单位，m）：

0（使用全部）/96 p, 34 m；

24 p, 44 m;

8 p, 77 m;

1 p, 462 m。

```
time qiime dada2 denoise-paired \
--i-demultiplexed-seqs demux.qza \
--p-n-threads 8 \
--p-trim-left-f 29 --p-trim-left-r 18 \
--p-trunc-len-f 0 --p-trunc-len-r 0 \
--o-table dada2-table.qza \
--o-representative-sequences dada2-rep-seqs.qza \
--o-denoising-stats denoising-stats.qza
```

可使用 dada2 结果并导入主流程, 或可选从外部其他流程结果的特征表和代表序列进行导入继续分析, 详见常见问题 1。QIIME 2 结果 qza/qzv 文中结果导出方法见常见问题 2。

```
cp dada2-table.qza table.qza
cp dada2-rep-seqs.qza rep-seqs.qza
```

统计特征表

```
qiime feature-table summarize \
--i-table table.qza \
--o-visualization table.qzv \
--m-sample-metadata-file metadata.txt
```

结果 qzv/qza 文件可上传到 <https://view.qiime2.org/> 网站查看。QIIME 的结果 qza 为数据文件, qzv 为图表文件, 本质上都是 zip 格式的压缩包, 也可使用压缩软件解压查看内容, qzv 解压的目录中包括分析结果的网页报告和相关的图表文件。我们先观察每个样本的测序量的总体概述表 (表 2), 用于确定多样性分析的抽平标准化阈值, 如本示例样本在特征表中使用的测序数据量最小值为 27,059, 即选择最小值; 如果最小值和第一分位数差别特别大, 则需要结合样本测序量和样本量分布图 (图 1) 选择最小值和第一分位数间的合适数值, 尽量保留足够样本量的前提下选择较大的抽平阈值以使用更多的测序数据, 注意低于阈值的样本将不会参与多样性分析。更多抽平阈值的交互式选择详见网页中“交互样本细节 (Interactive Sample Detail)”页面。

注: 抽平阈值最小为 1,000 是基于早期 454 测序时代的标准, 当前 Illumina 测序通量较大, 最小值一般不小于 5,000, 推荐 1 万, 且越大越好。

表 2. 每个样本的测序量 (Frequency per sample)

分位数	测序量
最小值 (Minimum frequency)	27,059
第一分位数 (1st quartile)	28,594
中位数 (Median frequency)	30,867.5
第三分位数 (3rd quartile)	32,861.5
最大值 (Maximum frequency)	34,667
平均值 (Mean frequency)	30,781.7

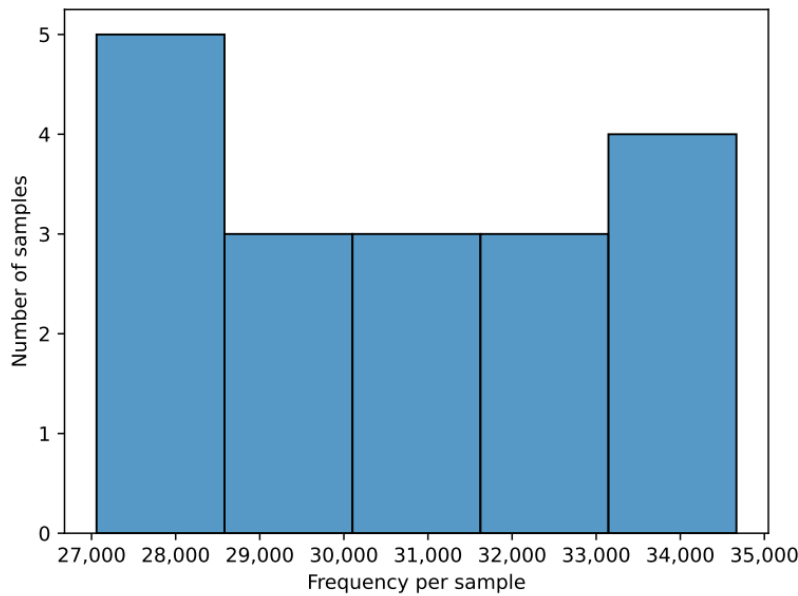


图 1. 样本可用测序量和样本量分布图。X 轴表示样本的可用测序量，轴须线为样本所在位置；Y 轴为样本量分布。本图可指导用户筛选合理的抽平阈值。

统计代表序列

```
qiime feature-table tabulate-seqs \
--i-data rep-seqs.qza \
--o-visualization rep-seqs.qzv
```

结果可上传至 <https://view.qiime2.org/>中查看（以 qzv 结尾的文件，我们接下来分析均默认在 QIIME 2 网页预览工具中查看，以后不再赘述），会显示序列长度统计，还有每个特征序列的 ID、长度和序列，其中序列可以点击跳转 NCBI BLAST 显示比对结果。

3. Alpha 和 Beta 多样性分析

构建进化树用于多样性分析

```
qiime phylogeny align-to-tree-mafft-fasttree \
--i-sequences rep-seqs.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
```

多样性分析，低于重采样深度的样本将会丢弃，通常重采样深度会选择样本测序量最小值以保留较多样本，同时要兼顾保留总体测序量最大化。因此需要根据样本量、数据分布等实际情况选择适合的值尽量使数据利用率最大化，具体见上面 table.qzv 结果中交互式筛选页面的辅助筛选工具可用。本研究数据分布较均匀，最小值即为最优阈值。最后会在 core-metrics-results 目录生成 4 种常用 Alpha 和 Beta 多样性结果。

```
qiime diversity core-metrics-phylogenetic \
--i-phylogeny rooted-tree.qza \
--i-table table.qza \
--p-sampling-depth 27059 \
--m-metadata-file metadata.txt \
--output-dir core-metrics-results
```

4. Alpha 多样性组间显著性分析和可视化

可选的 alpha 多样性指数有 faith_pd、shannon、observed_features 和 evenness。faith_pd 是综合物种间进化树信息的多样性指数[16,17]，shannon 是综合丰度和均匀度的指数，observed_features 是丰富度，evenness 是均匀度，中文简介进一步阅读《[Alpha 多样性箱线图](#)》，详细的介绍、计算方法和参考文献详见 scikit-bio 文档 <http://scikit-bio.org/docs/latest/generated/skbio.diversity.alpha.html>。此处以 observed_features 为例，在之前的版本中被称作 observed_otus。

```
index=observed_features
qiime diversity alpha-group-significance \
--i-alpha-diversity core-metrics-results/${index}_vector.qza \
--m-metadata-file metadata.txt \
--o-visualization core-metrics-results/${index}-group-significance.qzv
```

查看 core-metrics-results 目录下 observed_features-group-significance.qzv 结果文件，包括各种多样性指数分布的箱线图（图 2），和基于 Kruskal-Wallis 两两组较的 p-value 和 q-value（表 2），可以下载 svg 格式的矢量图和 tsv 格式的表格，还可以切换列（Column）探索不同分组方式下的分布和统计结果。

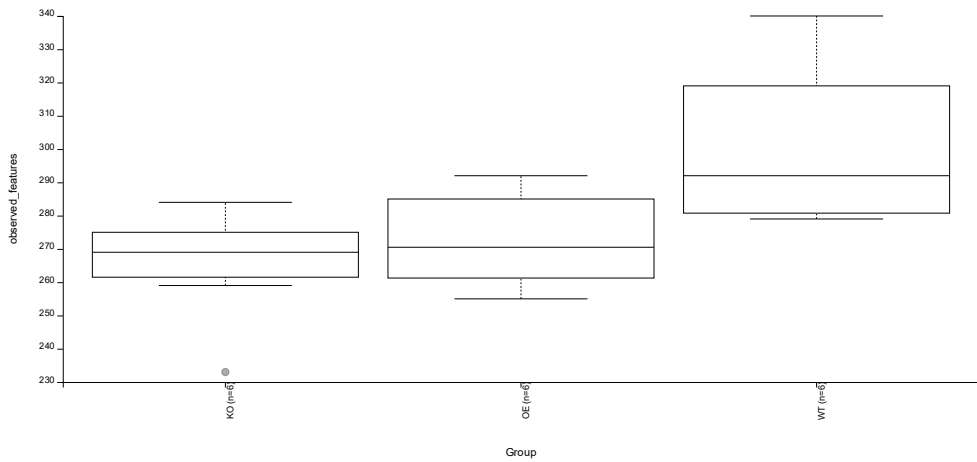


图 2. 各组 Alpha 多样性指数的箱线图分布。X 轴为分组，Y 轴为 observed features 多样性指数值，各组间统计的 *P*-value 值见表 3。

表 3. Alpha 多样性指数各组间比较的统计结果（Kruskal-Wallis pairwise）

Group 1	Group2	H	p-value	q-value
KO (n=6)	OE (n=6)	0.0231579	0.630356	0.630356
KO (n=6)	WT (n=6)	6.610329	0.010139	0.030417
OE(n=6)	WT (n=6)	3.705263	0.054241	0.081362

5. Alpha 多样性稀释曲线

max-depth 参数通常调置为样本测序量最大值，如果最大值为异常值（outlier），可以在第三分位数和最大值间选择合适的值，请参考 table.qzv 结果选择。

```
qiime diversity alpha-rarefaction \
--i-table table.qza \
--i-phylogeny rooted-tree.qza \
--p-max-depth 34667 \
--m-metadata-file metadata.txt \
--o-visualization alpha-rarefaction.qzv
```

结果 alpha-rarefaction.qzv 中有 shannon, faith_pd 和 observed_otus 三种 Alpha 多样性指数可切换，以展示各样本组随测序深度（Sequencing Depth）增加对应多样性指数分布的箱线图，图例可点选实现控制分组显示开/关。

6. Beta 多样性组间显著性分析和可视化

可选的 beta 指数有 unweighted_unifrac、bray_curtis、weighted_unifrac 以及 jaccard。UniFrac 是结合特征间进化关系计算群落间距离的方法[18]，weighted 和 unweighted 分别是指是否考虑特征的丰度权重。Bray-Curtis[19]是一种生态学常用的距离计算方法。Jaccard 类似于非加权的 Bray-Curtis 距离。中文

简介进一步阅读《[Beta 多样性 PCoA 和 NMDS 排序](https://www.osgeo.cn/scikit-bio/diversity.html)》，Unifrac 的详细的介绍详见 scikit-bio 文档（<https://www.osgeo.cn/scikit-bio/diversity.html>）。

指定 beta 多样性指数和分组用于减少计算量，因为置换检验较耗时。

```
distance=weighted_unifrac
column=Group
qiime diversity beta-group-significance \
--i-distance-matrix core-metrics-results/${distance}_distance_matrix.qza \
--m-metadata-file metadata.txt \
--m-metadata-column ${column} \
--o-visualization core-metrics-results/${distance}-${column}-significance.qzv \
--p-pairwise
```

此处可得到三个组别中的样本是否具有显著差异的分析结果，我们先查看 core-metrics-results 目录中的 weighted_unifrac_emperor.qzv，颜色选择分组 Group，此外还可以设置点的透明度、大小、形状等，确定图形样式后，在图右上角设置按钮选择“Save plot”——“SVG+Label”，可以保存主图和图例两个 SVG 格式的矢量图，发表时可使用矢量图编辑软件拼接（图 3）。

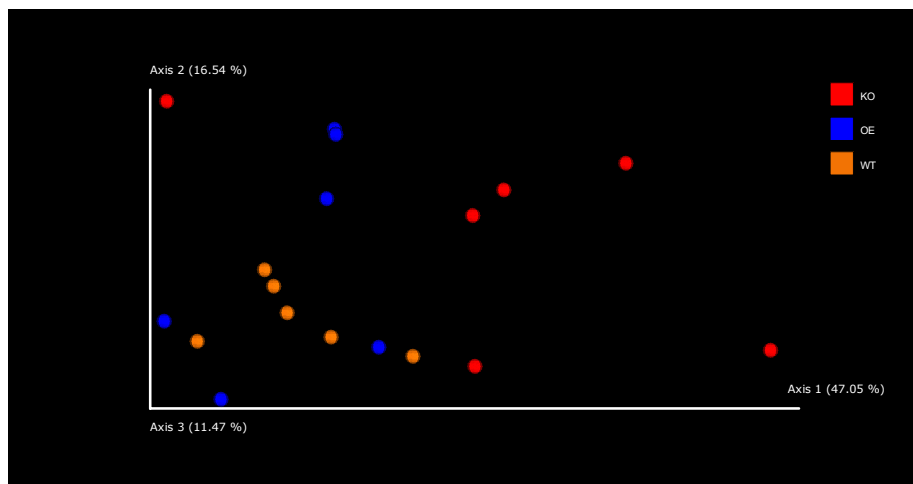


图 3. 基于 Weighted unifrac 距离的主坐标分析图。图中展示前三轴，括号中为解析率。图中点代表样本，颜色对应分组信息，对应关系见右上角图例。

再查看结果 weighted_unifrac-Group-significance.qzv，有组间距离分布图（图 4），还有组间成对 permanova 比较的结果（表 3），可以看到各组间均存在显著差异（ $p\text{-value} < 0.05$, $q\text{-value} < 0.05$ ）。

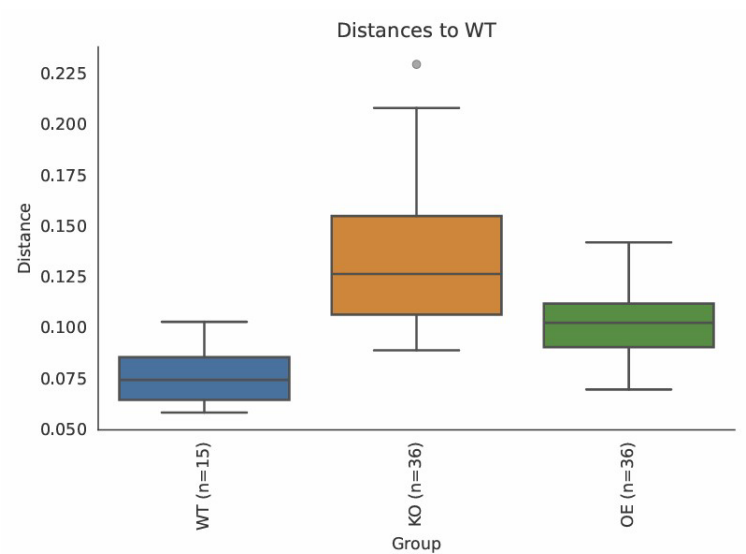


图 4. 相对于 WT 组样本的 Weighted Unifrac 距离分布箱线图。WT 组 6 个样本间非自身距离有 15 种 ($6 \times 5 \div 2$)，而组间为 36 种 (6×6)。组间的显著性检验结果见表 4。

表 4. Beta 多样性指数各组间比较的统计结果 (PERMANOVA/ADONIS)

Grou1	Grou2	Sample size	Permutations	pseudo-F	p-value	q-value
KO	OE	12	999	3.680430	0.010	0.015
KO	WT	12	999	3.982189	0.022	0.022
OE	WT	12	999	2.744504	0.009	0.015

7. 物种组成分析

训练分类器--全长（通用）

下载参考序列和物种分类信息

```
wget -c --no-check-certificate
https://ftp.microbio.me/greengenes_release/2022.10/2022.10.backbone.full-length.fna.qza
```

```
wget -c --no-check-certificate
https://ftp.microbio.me/greengenes_release/2022.10/2022.10.backbone.tax.qza
```

分类器训练，耗时 2 小时左右

```
time qiime feature-classifier fit-classifier-naive-bayes \
--i-reference-reads 2022.10.backbone.full-length.fna.qza \
--i-reference-taxonomy 2022.10.backbone.tax.qza \
--o-classifier classifier-full.qza
```

(可选) 训练分类器—指定 V 区分类器

使用与测试数据对应的 V5（799F） - V7（1193R） 引物为例进行提取序列，耗时约 6 分钟。

```
time qiime feature-classifier extract-reads \
--i-sequences 2022.10.backbone.full-length.fna.qza \
--p-f-primer AACMGGATTAGATACCCCKG \
--p-r-primer ACGTCATCCCCACCTTCC \
--p-trunc-len 350 \
--o-reads ref-seqs.qza
```

基于筛选的指定区段，生成实验特异的分类器，耗时约 30 分钟。

```
time qiime feature-classifier fit-classifier-naive-bayes \
--i-reference-reads ref-seqs.qza \
--i-reference-taxonomy 2022.10.backbone.tax.qza \
--o-classifier classifier_greengenes_V5-V7.qza
```

物种注释使用 classifier_greengenes_V5-V7.qza，也可以使用全长序列的训练集，耗时约 2 小时，具体时间由输入数据和数据库大小决定。如果你使用的研究非此引物，可以使用全长训练集，同时推荐使用实验中采用的引物训练特异的分类器应用于此，详见数据库部署。

```
qiime feature-classifier classify-sklearn \
--i-classifier classifier_greengenes_V5-V7.qza \
--i-reads rep-seqs.qza \
--o-classification taxonomy.qza
```

可视化物种注释

```
qiime metadata tabulate \
--m-input-file taxonomy.qza \
--o-visualization taxonomy.qzv
```

结果文件 taxonomy.qzv 为交互式网页表格（图 5），包括特征 ID、分类注释结果和置信度 3 列，可以实现排序和查找等功能。

Feature ID	Taxon	Confidence
005d4f7134fa2808c6d15bd73dc76c0d	d__Bacteria; p__Actinobacteriota; c__Actinomycetia; o__Mycobacteriales; f__Pseudonocardiales; g__Pseudonocardia	0.9997927922154275
0066304c256bb8fd503a5007f837c289	d__Bacteria; p__Verrucomicrobiota; c__Verrucomicrobiae; o__Chthoniobacteriales; f__Chthoniobacteraceae; g__VFJQ01	0.9043227905630342
0091167b8dc904da47b75f4475194893	d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Burkholderiales_592522; f__Burkholderiaceae_A_592522	0.9999999839821454
00993b7fa8df72adef8f0cf2b360de94	d__Bacteria; p__Actinobacteriota; c__Actinomycetia; o__Mycobacteriales; f__Micromonosporaceae; g__Virgisporangium; s__Virgisporangium aliadipaligenens	0.988472114014454
01061ad5b34963006479b16203fd3c5c	d__Bacteria; p__Planctomycetota; c__Phycisphaerae; o__Phycisphaerales; f__UBA1924; g__s__	0.994251060791022
0152071386c9f540302d99ed3dad936d	d__Bacteria; p__Actinobacteriota; c__Thermoleophilales; o__Gaielliales; f__Gaiellaceae; g__Palsa-739; s__Palsa-739 sp003139545	0.9347281949854271
01d1c2b1ac2eb16fc26499343847e87e	d__Bacteria; p__Chlamydia; c__Chlamydia; o__f__g__s__	0.999997963663047
01e26c95617448e11eeddd76a118d59	d__Bacteria; p__Verrucomicrobiota; c__Verrucomicrobiae; o__Chthoniobacteriales; f__Chthoniobacteraceae; g__VFJQ01; s__VFJQ01 sp009885995	0.993882209393236

图 5. 物种注释信息

堆叠柱状图展示 (图 6)

```
qiime taxa barplot \
--i-table table.qza \
--i-taxonomy taxonomy.qza \
--m-metadata-file metadata.txt \
--o-visualization taxa-bar-plots.qzv
```

结果文件 taxa-bar-plots.qzv 为交互式网页图片 (图 6), 如可以调整柱宽至中等, 切换不同分类级别至 Level2, 修改配色方案为 schecDark2, 按分组 Group 排序, 再按 Proteobacteria 降序排列, 然后保存 SVG 的柱状图和图例。也推荐保存表格的 csv 数据用于个性化绘图。

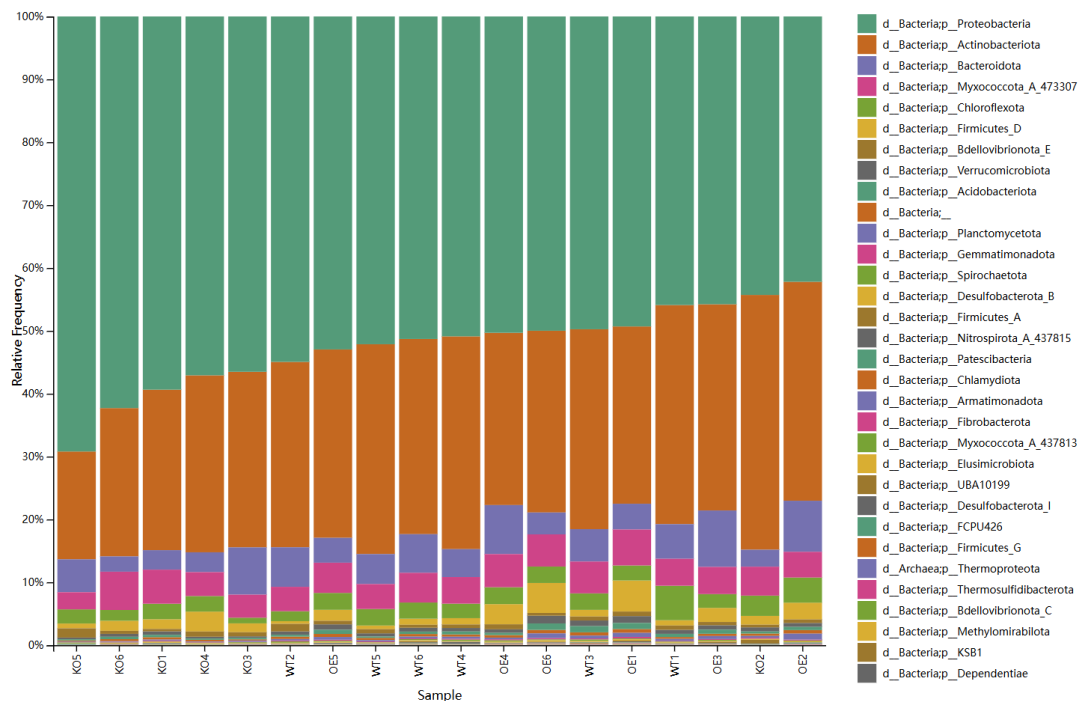


图 6. 堆叠柱状图展示样本门水平组成。图中按分组和变形菌门丰度依次排序。

8. 差异分析 ancom

格式化特征表, 添加伪计数

```
qiime composition add-pseudocount \
--i-table table.qza \
--o-composition-table comp-table.qza
```

ancom[20]计算差异特征, 指定分组类型比较, 耗时约 7 m。

```
column=Group
```

```
time qiime composition ancom \
--i-table comp-table.qza \
--m-metadata-file metadata.txt \
--m-metadata-column ${column} \
--o-visualization ancom-${column}.qzv
```

结果 ancom-Group.qzv 用散点图显示显著差异的 ASV (图 7), 在交互式图形页面中可直接看到显著差异 ASV 统计结果。本研究组间差异较小, 仅有 1 个显著差异扩增序列变体 (amplicon sequence variants ASV) 为 b46f62815f5bc0f8c18c3c374acabe23 (ASV_672), 在下图最右上角以红色标注, 在 taxonomy.qzv 中查找其分类信息为 d__Bacteria; p__Actinobacteriota; c__Actinomycetia; o__Mycobacteriales; f__Pseudonocardia; g__Pseudonocardia。

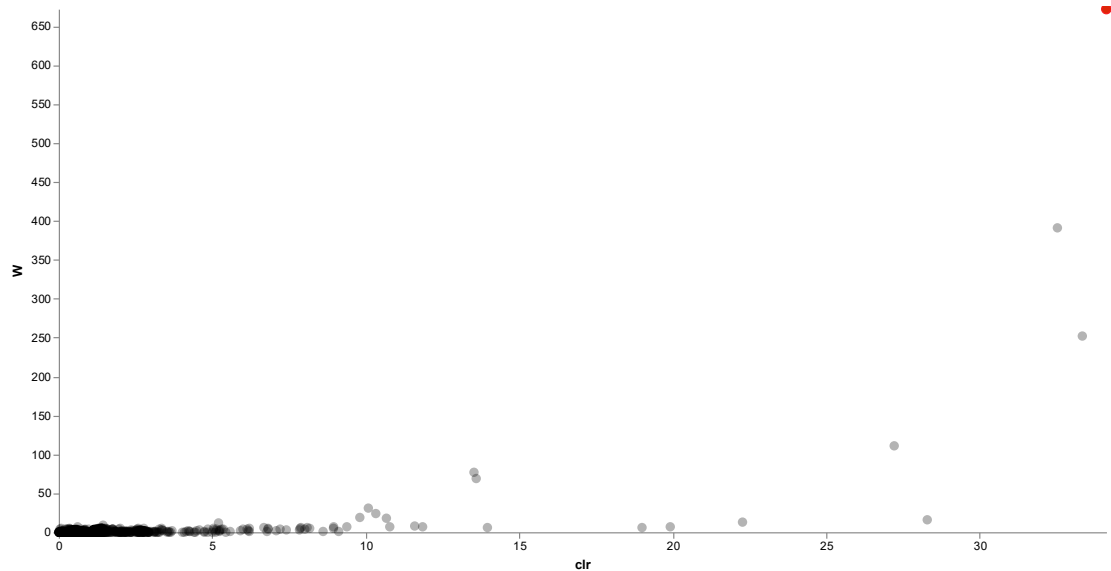


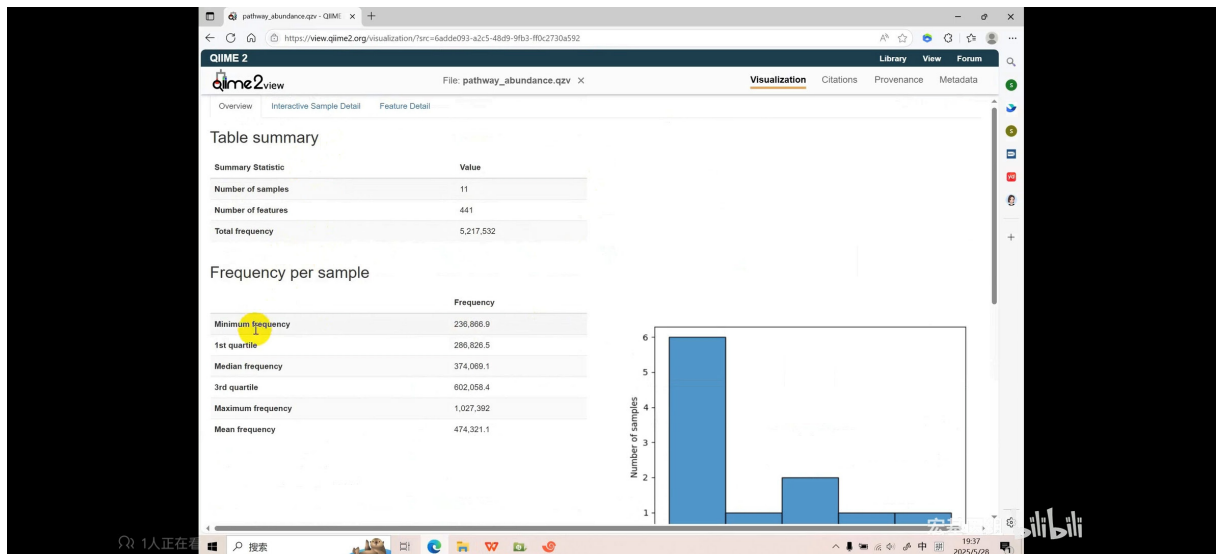
图 7. 属水平 ancom 差异分析散点图

QIIME 2 实用非核心插件

QIIME 2 微生物组分析功能可通过插件提供给用户。除 QIIME 2 2025.4 包含的微生物组分析使用的核心插件, QIIME 2 论坛中还有一些实用的非核心插件, 本文将重点介绍以下 3 个插件(Q2-picrust2、Q2-ITSxpress 和 Q2-FMT) 的使用, 更多相关插件详见表 8。

q2-picrust2 插件 (功能预测)

用于运行 PICRUST2 管道, 以获得基于 16S 数据的 EC、KO 和 MetaCyc 通路预测[21]。EPA-NG 或 SEPP 可用于将序列放入所需的参考系统发育中。本文将重点介绍如何运行 SEPP, 它占用的内存较少。



视频 5. 运行 q2-picrust2 插件 (<https://www.bilibili.com/video/BV1yp7WziEgx/>)

1. 安装

创建目录

```
mkdir -p q2-picrust2-tutorial
cd q2-picrust2-tutorial
```

使用 conda 安装 q2-picrust2 (原始链接)

```
conda env create --name q2-picrust2-amplicon-2024.5 --file
https://raw.githubusercontent.com/picrust/q2-
picrust2/refs/heads/master/environment-files/q2-picrust2-qiime2-amplicon-
2024.5.yml
```

备用链接

```
wget -c http://www.imeta.science/db/qiime2/q2-picrust2-amplicon-
2024.5.yml
```

```
conda env create --name q2-picrust2-amplicon-2024.5 --file q2-picrust2-
amplicon-2024.5.yml
```

激活环境

```
conda activate q2-picrust2-amplicon-2024.5
```

2. 下载 PICRUST2 教程中的测试文件

引用格式: 杨海飞, 曾美尹, 高云云, 陈同, 刘永鑫 (2025). 使用 QIIME 2 2025.4 分析微生物组 16S rDNA 基因和 ITS 扩增子测序数据. *Bio-101* e1011032. Doi: 10.21769/BioProtoc.1011032. 18

How to cite: Yang, H. F., Zeng, M. Y., Gao, Y. Y., Chen, T. and Liu, Y. X. (2025). Using QIIME 2 2025.4 to Analyze Microbiome 16S rDNA Gene and ITS Amplicon Sequencing Data. *Bio-101* e1011032. Doi: 10.21769/BioProtoc.1011032. (in Chinese)

```
wget -c
http://kronos.pharmacology.dal.ca/public_files/picrust/picrust2 tutorial
files/mammal_biom.qza
wget -c
http://kronos.pharmacology.dal.ca/public_files/picrust/picrust2 tutorial
files/mammal_seqs.qza
wget -c
http://kronos.pharmacology.dal.ca/public_files/picrust/picrust2 tutorial
files/mammal_metadata.tsv
```

这些文件对应于示例的 ASV 计数表、ASV 序列和元数据。总共有 11 个样本和 37 个 ASV。这些样本是从北极狼、土狼、海狸和豪猪的哺乳动物粪便中收集的。

3. 功能预测

创建临时目录并设置环境变量

```
mkdir -p ~/tmp/qiime2_picrust2_temp
export TMPDIR=~/tmp/qiime2_picrust2_temp
```

运行 SEPP

```
qiime picrust2 full-pipeline \
--i-table mammal_biom.qza \
--i-seq mammal_seqs.qza \
--output-dir q2-picrust2_output \
--p-placement-tool sepp \
--p-threads 8 \
--p-hsp-method pic \
--p-max-nsti 2 \
--verbose
```

输出结果文件有 ec_metagenome.qza-EC 宏基因组预测、ko_metagenome.qza-KO 宏基因组预测和 pathway_abundance.qza-MetaCyc 通路丰度预测。

4. 使用此命令获取有关通路丰度文件的摘要信息（表 5）

```
qiime feature-table summarize \
--i-table q2-picrust2_output/pathway_abundance.qza \
--o-visualization q2-picrust2_output/pathway_abundance.qzv
```

表 5. 每个样本的频率(Frequency per sample)

分位数	频率
最小值(Minimum frequency)	236,866.9
第一分位数(1st quartile)	286,826.5

中位数(Median frequency)	374,069.1
第三分位数(3rd quartile)	602,058.4
最大值(Maximum frequency)	1,027,392
平均值(Mean frequency)	474,321.1

5. 计算多样性指标

```
qiime diversity core-metrics \
--i-table q2-picrust2_output/pathway_abundance.qza \
--p-sampling-depth 236867 \
--m-metadata-file mammal_metadata.tsv \
--output-dir pathabun_core_metrics_out \
--p-n-jobs 1
```

查看结果 pathabun_core_metrics_out/bray_curtis_emperor.qzv (图 8)，它显示了此数据集在食肉动物和啮齿动物之间的明显差异。

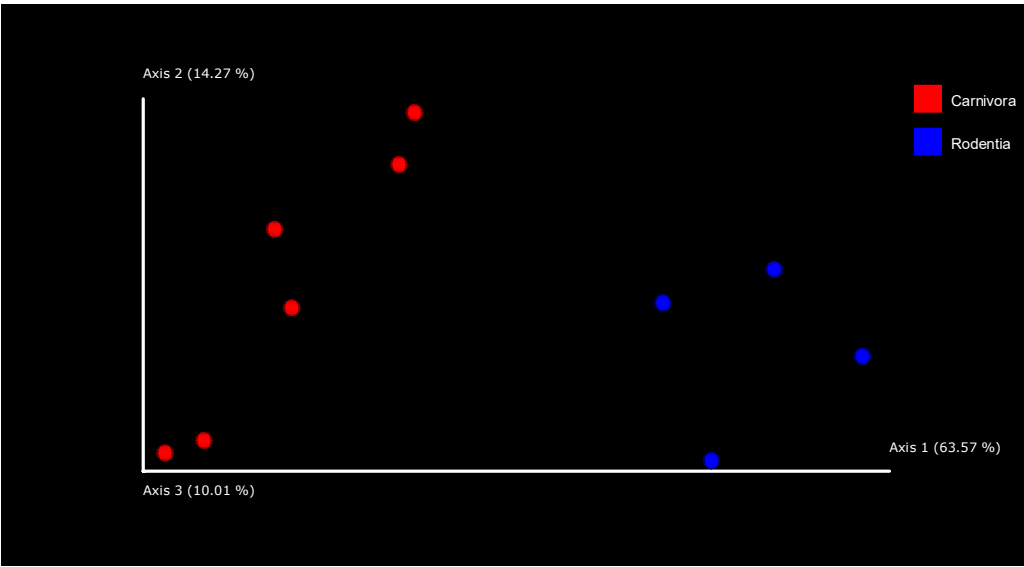


图 8 基于 Weighted unifrac 距离的主坐标分析图

q2-ITSxpress 插件（修剪 ITS 序列的保守侧翼区域）

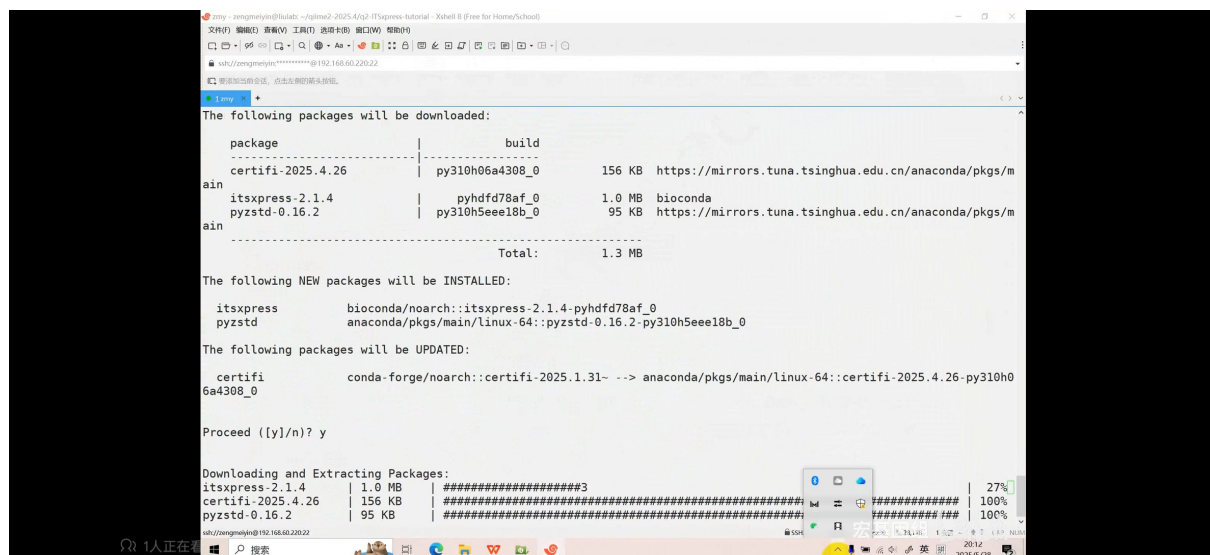
用于快速修剪内部转录的间隔序列的软件，并具有标记基因分析的质量分数。ITSxpress 通过获取 FASTQ 数据、反复制序列，然后使用 HMMSearch 识别起始和终止位点[22]。分析结果并返回修剪后的文件。本文将介绍 ITS 工作流程的第一部分：

- (1) 使用 ITSxpres 修剪 ITS 区域
- (2) 使用 DADA2 或 Deblur 调用序列变体

(3) 训练 QIIME 2 分类器

(4) 按分类学对序列进行分类

在本教程中，我们将从两个双端样本开始，这两个样本已经拆分为正向和反向 FASTQ 文件。



视频 6. 运行 q2-ITSxpress 插件 (<https://www.bilibili.com/video/BV1yp7WziE4L/>)

1. 安装

创建目录

```
mkdir -p q2-ITSxpress-tutorial
cd q2-ITSxpress-tutorial
```

激活 QIIME 2 环境

```
conda activate qiime2-amplicon-2025.4
```

使用 Bioconda 安装 ITSxpress

```
conda install -c bioconda itsxpress
```

刷新插件

```
qiime dev refresh-cache
```

检查是否安装成功

```
qiime itsxpress
```

引用格式: 杨海飞, 曾美尹, 高云云, 陈同, 刘永鑫 (2025). 使用 QIIME 2 2025.4 分析微生物组 16S rDNA 基因和 ITS 扩增子测序数据. *Bio-101* e1011032. Doi: 10.21769/BioProtoc.1011032. 21

How to cite: Yang, H. F., Zeng, M. Y., Gao, Y. Y., Chen, T. and Liu, Y. X. (2025). Using QIIME 2 2025.4 to Analyze Microbiome 16S rDNA Gene and ITS Amplicon Sequencing Data. *Bio-101* e1011032. Doi: 10.21769/BioProtoc.1011032. (in Chinese)

2. 下载示例数据

我们将使用来自两个土壤样品的数据，这些样品的 ITS2 区域已用真菌引物扩增。它们已被子采样到 10,000 个读取对，以便更快地处理。sample1_r1.fq.gz 和 sample1_r2.fq.gz；sample2_r1.fq.gz 和 sample2_r2.fq.gz；清单文件：manifest.txt 和映射文件：mapping.txt。

```
wget -c http://www.imeta.science/db/qiime2/sample1\_r1.fastq.gz
wget -c http://www.imeta.science/db/qiime2/sample1\_r2.fastq.gz
wget -c http://www.imeta.science/db/qiime2/sample2\_r1.fastq.gz
wget -c http://www.imeta.science/db/qiime2/sample2\_r2.fastq.gz
wget -c http://www.imeta.science/db/qiime2/manifest.txt
wget -c http://www.imeta.science/db/qiime2/mapping.txt
```

3. 导入序列数据

```
qiime tools import \
--type SampleData[PairedEndSequencesWithQuality] \
--input-format PairedEndFastqManifestPhred33 \
--input-path manifest.txt \
--output-path sequences.qza
```

通过运行 summarize 命令来查看数据的质量（表 6）

```
qiime demux summarize \
--i-data sequences.qza \
--o-visualization sequences.qzv
```

表 6. 拆分序列汇总计数(Demultiplexed sequence counts summary)

分位数	正向	反向
最小值 (Minimum)	2500	2500
中位数 (Median)	2500.0	2500.0
平均值 (Mean)	2500.0	2500.0
最大值 (Maximum)	2500	2500
总计 (Total)	5000	5000

4. 使用 ITSxpress 修剪 ITS 样本

```
qiime itsxpress trim-pair-output-unmerged \
--i-per-sample-sequences sequences.qza \
--p-region ITS2 \
--p-taxa F \
--p-cluster-id 1.0 \
--p-threads 16 \
```

```
--o-trimmed trimmed_exact.qza
```

使用以下命令聚类 99.5% 的序列相似度。建议将聚类到 100%相似性，因为速度优势可以忽略不计。

```
qiime itsxpress trim-pair-output-unmerged \  
--i-per-sample-sequences sequences.qza \  
--p-region ITS2 \  
--p-taxa F \  
--p-cluster-id 0.995 \  
--p-threads 16 \  
--o-trimmed trimmed.qza
```

5. 使用 DADA2 识别序列变异

```
qiime dada2 denoise-paired \  
--i-demultiplexed-seqs trimmed_exact.qza \  
--p-trunc-len-r 0 \  
--p-trunc-len-f 0 \  
--output-dir dada2out
```

汇总数据以进行目视检查(表 7)

```
qiime feature-table summarize \  
--i-table dada2out/table.qza \  
--o-visualization tableviz.qzv
```

表 7. 每个样本的频率(Frequency per sample)

分位数	频率
最小值 (Minimum frequency)	848
第一分位数 (1st quartile)	862.2
中位数 (Median frequency)	876.5
第三分位数 (3rd quartile)	890.8
最大值 (Maximum frequency)	905
平均值 (Mean frequency)	876.5

6. 从 UNITE 下载用于真菌分类的参考数据

原始下载链接

```
wget -c https://s3.hpc.ut.ee/plutof-public/original/db1d6ddb-a35d-48c5-8b1a-ad9dd3310c6d.tgz
```

备用下载链接

```
wget -c http://www.imeta.science/db/qiime2/db1d6ddb-a35d-48c5-8b1a-ad9dd3310c6d.tgz
```

解压

```
tar -xzf db1d6ddb-a35d-48c5-8b1a-ad9dd3310c6d.tgz
```

将最新的 UNITE 数据导入 QIIME 2

```
qiime tools import \
  --type 'FeatureData[Sequence]' \
  --input-path sh_refs_qiime_ver10_dynamic_04.04.2024.fasta \
  --output-path unite.qza
```

```
qiime tools import \
  --type 'FeatureData[Taxonomy]' \
  --input-format HeaderlessTSVTaxonomyFormat \
  --input-path sh_taxonomy_qiime_ver10_dynamic_04.04.2024.txt \
  --output-path unite-taxonomy.qza
```

7. 训练 QIIME 2 分类器, 耗时 1 小时左右

```
qiime feature-classifier fit-classifier-naive-bayes \
  --i-reference-reads unite.qza \
  --i-reference-taxonomy unite-taxonomy.qza \
  --o-classifier classifier.qza
```

8. 对序列变体进行分类

```
qiime feature-classifier classify-sklearn \
  --i-classifier classifier.qza \
  --i-reads dada2out/representative_sequences.qza \
  --o-classification taxonomy.qza
```

汇总结果

```
qiime metadata tabulate \
  --m-input-file taxonomy.qza \
  --o-visualization taxonomy.qzv
```

堆叠柱状图展示 (图 9)

```
qiime taxa barplot \
  --i-table dada2out/table.qza \
  --i-taxonomy taxonomy.qza \
  --m-metadata-file mapping.txt \
  --o-visualization taxa-bar-plots.qzv
```

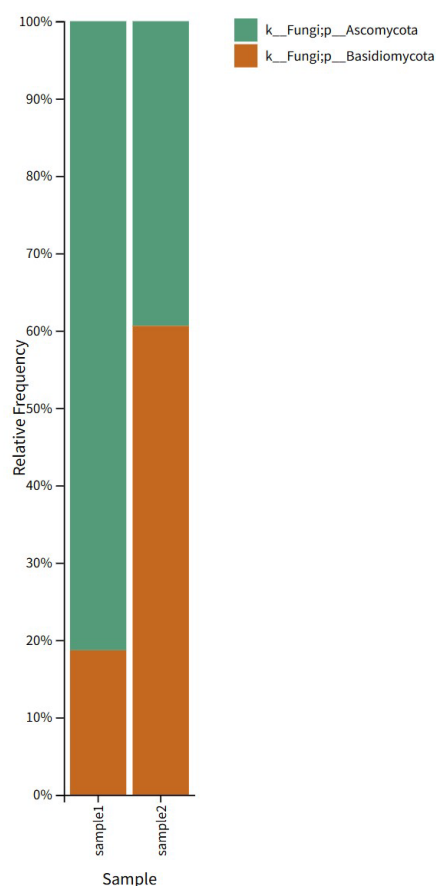


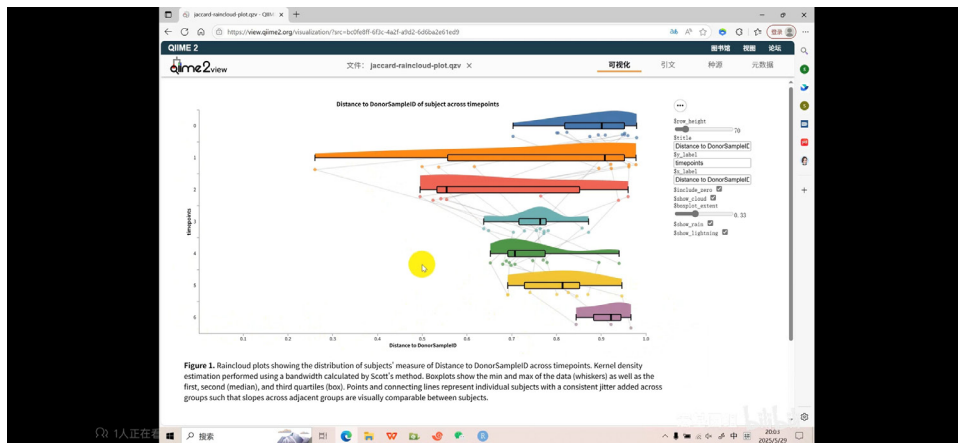
图 9. 堆叠柱状图展示样本门水平组成

q2-FMT 插件（评估粪菌移植后的植入范围）

Q2-FMT 是 QIIME 2 的一个插件，专门用于评估粪菌移植（FMT）后微生物组的定植程度(Herman 2024)，其主要功能包括以下几个方面：

- （1）评估嵌合不对称群落融合：通过分析受体微生物组与供体微生物组的融合情况，判断移植后的微生物组是否成功定植。
- （2）识别供体微生物组指示特征：找出在受体中定植的供体微生物组特征，帮助了解哪些微生物在移植后能够在受体中存活并发挥作用。

（3）分析时间稳定性：研究移植后微生物组随时间的变化情况，评估其在受体中的稳定性和持久性。
本教程介绍该插件的安装及绘制云雨图了解受体与供体样本在粪菌移植（FMT）后如何变化，更多内容请参照 <https://q2-fmt.readthedocs.io/en/latest/tutorial/tutorial.html>



视频 7. 运行 q2-FMT 插件 <https://www.bilibili.com/video/BV1UL7Wz3EA9/>

1. 安装

下载清单文件（原始链接）

```
wget -c https://raw.githubusercontent.com/qiime2/q2-fmt/refs/heads/dev/environment-files/q2-fmt-qiime2-amplicon-2024.10.yml
```

备用链接

```
wget -c http://www.imeta.science/db/qiime2/q2-fmt-qiime2-amplicon-2024.10.yml
```

conda 新建环境安装

```
conda env create --name q2-fmt-amplicon-2024.10 --file q2-fmt-qiime2-amplicon-2024.10.yml
```

安装过程中遇到提示 pip 安装一个依赖项无法从 GitHub 下载 q2-fmt 的压缩包时，手动下载从本地安装

```
wget -c http://www.imeta.science/db/qiime2/q2-fmt-2024.11.1.zip

~/miniconda3/envs/q2-fmt-amplicon-2024.10/bin/python -m pip install ./q2-fmt-2024.11.1.zip
```

激活环境

```
conda activate q2-fmt-amplicon-2024.10
```

2. 下载元数据和特征表

创建目录并进入

```
mkdir q2-fmt-tutorial
cd q2-fmt-tutorial
```

数据下载

```
wget -O 'sample-metadata.tsv' \
'https://q2-fmt.readthedocs.io/en/latest/data/tutorial/sample-
metadata.tsv'
wget -O 'feature-table.qza' \
'https://q2-fmt.readthedocs.io/en/latest/data/tutorial/feature-table.qza'
```

汇总特征表

```
qiime feature-table summarize \
--i-table feature-table.qza \
--m-sample-metadata-file sample-metadata.tsv \
--o-visualization autofmt-table-summ.qzv
```

3. 计算多样性指标

要开始我们的多样性分析，我们首先需要确定我们想要选择什么样的均匀采样深度（或“稀疏深度”）来计算我们的多样性指标。由于大多数多样性指标对不同样本的不同采样深度敏感，因此通常会将每个样本的计数随机子采样为特定值。

```
qiime diversity core-metrics \
--i-table feature-table.qza \
--p-sampling-depth 10000 \
--m-metadata-file sample-metadata.tsv \
--output-dir diversity-core-metrics
```

4. 绘制云雨图了解受体与供体的距离（图 10）

```
qiime fmt cc \
--i-diversity-measure diversity-core-metrics/jaccard distance matrix.qza \
--m-metadata-file sample-metadata.tsv \
--p-distance-to donor \
--p-compare baseline \
--p-time-column timepoints \
--p-reference-column DonorSampleID \
--p-subject-column PatientID \
--p-filter-missing-references \
--p-against-group 0 \
--p-p-val-approx asymptotic \
--o-stats jaccard-raincloud-stats.qza \
--o-raincloud-plot jaccard-raincloud-plot.qzv
```

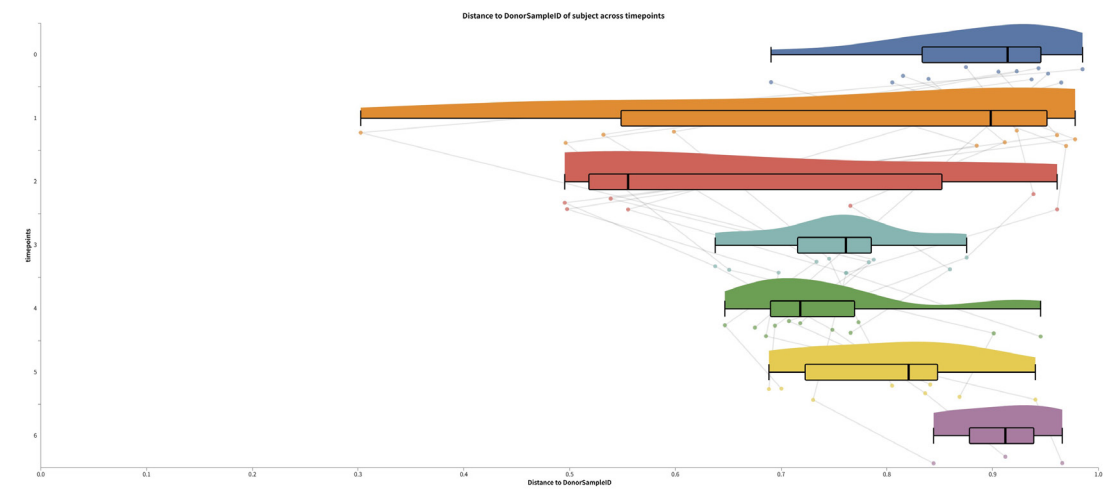


图 10. 受试者在不同时间点的“与供体样本 ID 的距离”测量值的分布情况

在得到的云雨图中，我们看到在癌症治疗后（时间点 0-2），受体与供体的距离相对较远。在 FMT 干预后（时间点 3），受体与供体的距离变近了，两者的微生物组看起来更相似。我们看到这有一些稳定性，但到最后一个时间点，接受者的微生物组看起来大多与其捐赠的微生物组不同。

表 8. QIIME 2 官网论坛实用插件

名称	链接	简介	参考文献
Q2-krona	https://forum.qiime2.org/t/q2-krona-plugin-for-creating-krona-plots/21294	Krona 是一个创建交互式多层饼图的工具	[23]
Q2-SCRuB	https://forum.qiime2.org/t/q2-scrub-release/26609	SCRuB 是一种概率计算机净化方法，它结合了多个样品和对照品之间的共享信息，以精确识别和去除污染物	[24]
Q2-SCNIC	https://forum.qiime2.org/t/q2-scnic-a-tool-for-making-correlation-networks-finding-modules-of-observations-and-summarizing-them/6116	SCNIC 是一种从特征表构建相关网络、查找所述网络中的模块并汇总这些模块的工具	[25]
Q2-fondue	https://forum.qiime2.org/t/q2-fondue-plugin-for-meta-data-retrieval-from-sra/22595	用于从 SRA 检索（元）数据	[26]
Q2-ITSxpress	https://forum.qiime2.org/t/q2-itsxpress-a-tutorial-on-a-qiime-2-plugin-to-trim-its-sequences/5780	用于修剪 ITS 序列，通过 HMM 图谱与侧翼保守基因的末端相匹配来修剪 FASTA 序列	[22]

Q2-srs	https://forum.qiime2.org/t/q2-srs-qiime2-plugin-for-library-size-normalization-by-scaling-with-ranked-subsampling-srs/17661	用于通过使用排名子采样 (SRS) 进行缩放来规范库大小	[27]
Q2-Phylogenize	https://forum.qiime2.org/t/q2-phylogenize-a-pipeline-for-associating-microbial-genes-with-environments-accounting-for-phylogeny/9971	使用系统发育回归将微生物基因与环境中的流行率或该环境相对于其他环境的特异性联系起来	
Q2-picrust2	https://forum.qiime2.org/t/q2-picrust2-updated-pipeline-for-metagenome-inference-with-denoised-sequences/6520	基于 16S 数据的 EC、KO 和 MetaCyc 通路预测	[21]
Q2-ghost-tree	https://forum.qiime2.org/t/q2-ghost-tree-plugin-community-tutorial-for-creating-hybrid-gene-phylogenetic-trees/6139	将来自两个遗传标记数据库的序列数据组合成一个可用于多样性分析的混合系统发育树	[28]
Q2-clawback	https://forum.qiime2.org/t/using-q2-clawback-to-assemble-taxonomic-weights/5859	组装分类权重	[29]
Q2-perc-norm	https://forum.qiime2.org/t/q2-perc-norm-community-plugin/3922	对数据进行百分位归一化以允许跨不同数据集汇集数据的方法	[30]
Q2-corncob	https://forum.qiime2.org/t/q2-corncob-community-tutorial/6138	corncob 是一个使用丰度表和样本数据的单个分类单元回归模型	[31]
Q2-gcn-norm	https://forum.qiime2.org/t/q2-gcn-norm-plugin-for-normalizing-sequences-by-16s-rrna-gene-copy-number/12715	基于 rrnDB 数据库的 QIIME 2 插件，用于通过 16S rRNA 基因拷贝数 (GCN) 对序列进行归一化	[32]
Q2-qurro	https://forum.qiime2.org/t/visualizing-omic-feature-rankings-and-log-ratios-with-q2-qurro/14795	可视化“组学特征排名和对数比”	[33]
Q2-hitac	https://forum.qiime2.org/t/q2-hitac-a-plugin-for-taxonomic-classification-of-fungal-its-sequences/25031	用于真菌 ITS 序列分类	[34]
Q2-sourmash	https://forum.qiime2.org/t/q2-sourmash-a-plugin-for-generating-and-comparing-minhash-sketches/6276	用于生成和比较 MinHash 草图的插件	[35]
Q2-repeat-rarefy	https://forum.qiime2.org/t/q2-repeat-rarefy-qiime2-plugin-for-generating-the-average-rarefied-table-for-library-size-normalization-using-repeated-rarefaction/18032	用于使用重复稀疏生成平均稀薄表，用于库大小规范化	[36]

Q2-data-augment	https://forum.qiime2.org/t/q2-data-augment-qiime2-plugin-for-data-augmentation-using-rarefaction-rarefy-for-augment/18094	用于使用 rarefaction 进行数据增强	[36]
Q2-FMT	https://q2-fmt.readthedocs.io/en/latest/tutorial/tutorial.html	用于评估粪菌移植（FMT）后微生物组的定植程度	[37]

常见问题

1. 外部导入特征表和代表序列

其他常用扩增子分析流程如 Mothur [38]、QIIME [9]、USEARCH [39]、VSEARCH [40]和 DADA2 [12] 等分析流程的特征表和代表序列结果也可以导入 QIIME 2 继续分析。需要准备特征表（otutab.txt）和代表序列（otus.fa）两个文件，示例在本文 github 中或以下链接下载。特征表通用的 BIOM 格式可以直接导入 QIIME 2 [10]，如果制表符分隔的纯文本格式需要使用 biom 命令转换为 BIOM 格式再导入 [41]。

```
wget -c http://210.75.224.110/github/MicrobiomeProtocol/e2.QIIME2/otutab.txt
wget -c http://210.75.224.110/github/MicrobiomeProtocol/e2.QIIME2/otus.fa
```

转换文本为 Biom1.0，注意 biom --version 2.1.5/8 可以，2.1.7 可能报错

```
biom convert -i otutab.txt -o otutab.biom \
--table-type="OTU table" --to-json
```

导入特征表

```
qiime tools import --input-path otutab.biom \
--type 'FeatureTable[Frequency]' --input-format BIOMV100Format \
--output-path table.qza
```

导入代表序列

```
qiime tools import --input-path otus.fa \
--type 'FeatureData[Sequence]' \
--output-path rep-seqs.qza
```

2. 导出特征表、代表序列和物种注释

导出特征表为 biom 格式

```
qiime tools export \
```

```
--input-path table.qza \
--output-path feature-table
```

转换 biom 格式特征表为 tsv 格式

```
biom convert -i feature-table/feature-table.biom \
-o feature-table/feature-table.txt \
--to-tsv
```

删除多余注释行

```
sed -i '/# Const/d' feature-table/feature-table.txt
```

导出代表序列

```
qiime tools export \
--input-path rep-seqs.qza \
--output-path rep-seqs
```

导出物种注释

```
qiime tools export \
--input-path taxonomy.qza \
--output-path taxonomy
```

致谢

该研究由国家自然科学基金(32470055, U23A20148)、中国博士后科学基金(2024M753580)资助和农业科学与技术创新项目(CAAS-ZDRW202308)。

参考文献

- GAO, G. F. and Chu, H. Y. (2020). Techniques and methods of microbiomics and their applications. *Chinese Journal of Plant Ecology* 44(4): 395–408. <https://doi.org/10.17521/cjpe.2019.0222>
- 刘双江, 施文元, 赵国屏 (2017). 中国微生物组计划_机遇与挑战. *中国科学院院刊*. <https://doi.org/10.16418/j.issn.1000-3045.2017.03.004>
- Aggarwal, N., Kitano, S., Puah, G. R. Y., Kittelmann, S., Hwang, I. Y. and Chang, M. W. (2022). Microbiome and Human Health: Current Understanding, Engineering, and Enabling Technologies. *Chem Rev.* 123(1): 31–72. <https://doi.org/10.1021/acs.chemrev.2c00431>

4. Singh, B. K., Trivedi, P., Egidi, E., Macdonald, C. A. and Delgado-Baquerizo, M. (2020). Crop microbiome and sustainable agriculture. *Nat Rev Microbiol.* 18(11): 601–602. <https://doi.org/10.1038/s41579-020-00446-y>
5. Xu, X. and Jiang, J. (2024). Engineering microbiomes for enhanced bioremediation. *PLoS Biol.* 22(12): e3002951. <https://doi.org/10.1371/journal.pbio.3002951>
6. Gao, Y., Li, D. and Liu, Y. X. (2023). Microbiome research outlook: past, present, and future. *Protein & Cell* 14(10): 709–712. <https://doi.org/10.1093/procel/pwad031>
7. Liu, Y. X., Qin, Y., Chen, T., Lu, M., Qian, X., Guo, X. and Bai, Y. (2020). A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein Cell* 12(5): 315–330. <https://doi.org/10.1007/s13238-020-00724-8>
8. 杨潇瀛, 张浩林, 韩莹莹, 翁强, 袁峥嵘 (2021) . 16s 扩增子分析中常用软件及数据库应用现状. *Bio-Protocol.* <https://doi.org/10.21769/BioProtoc.2003388>
9. Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 7(5): 335–336. <https://doi.org/10.1038/nmeth.f.303>
10. Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol.* 37(8): 852–857. <https://doi.org/10.1038/s41587-019-0209-9>
11. Yousuf, S., Luo, H., Zeng, M., Chen, L., Ma, T., Li, X., Zheng, M., Zhou, X., Chen, L., Xi, J., et al. (2024). Unveiling microbial communities with EasyAmplicon: A user-centric guide to perform amplicon sequencing data analysis. *iMetaOmics* 1(2): e42. <https://doi.org/10.1002/imo2.42>
12. Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A. and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods.* 13(7): 581–583. <https://doi.org/10.1038/nmeth.3869>
13. McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., Andersen, G. L., Knight, R. and Hugenholtz, P. (2011). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 6(3): 610–618. <https://doi.org/10.1038/ismej.2011.139>
14. Huang, A. C., Jiang, T., Liu, Y. X., Bai, Y. C., Reed, J., Qu, B., Goossens, A., Nützmann, H. W., Bai, Y., Osbourn, A., et al. (2019). A specialized metabolic network selectively modulates *Arabidopsis* root microbiota. *Science.* 364(6440): eaau6389. <https://doi.org/10.1126/science.aau6389>
15. Wang, Y., Song, F., Zhu, J., Zhang, S., Yang, Y., Chen, T., Tang, B., Dong, L., Ding, N., Zhang, Q. et al. (2017). *Genomics Proteomics Bioinformatics.* 15(1): 14–18.
16. Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biol Conserv.* 61(1): 1–10. [https://doi.org/10.1016/0006-3207\(92\)91201-3](https://doi.org/10.1016/0006-3207(92)91201-3)
17. Hamady, M., Lozupone, C. and Knight, R. (2009). Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J.* 4(1): 17–27. <https://doi.org/10.1038/ismej.2009.97>

引用格式: 杨海飞, 曾美尹, 高云云, 陈同, 刘永鑫 (2025). 使用 QIIME 2 2025.4 分析微生物组 16S rDNA 基因和 ITS 扩增子测序数据. *Bio-101* e1011032. Doi: 10.21769/BioProtoc.1011032. 32

How to cite: Yang, H. F., Zeng, M. Y., Gao, Y. Y., Chen, T. and Liu, Y. X. (2025). Using QIIME 2 2025.4 to Analyze Microbiome 16S rDNA Gene and ITS Amplicon Sequencing Data. *Bio-101* e1011032. Doi: 10.21769/BioProtoc.1011032. (in Chinese)

18. Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J. and Knight, R. (2010). UniFrac: an effective distance metric for microbial community comparison. *ISME J.* 5(2): 169–172. <https://doi.org/10.1038/ismej.2010.133>
19. Beals, E. W. (1984). Bray-Curtis Ordination: An Effective Strategy for Analysis of Multivariate Ecological Data. *Adv Ecol Res.* : 1–55. [https://doi.org/10.1016/s0065-2504\(08\)60168-3](https://doi.org/10.1016/s0065-2504(08)60168-3)
20. Mandal, S., Van Treuren, W., White, R. A., Eggesbø, M., Knight, R. and Peddada, S. D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb Ecol Health Dis.* 26: e27663. <https://doi.org/10.3402/mehd.v26.27663>
21. Caicedo, H. H., Hashimoto, D. A., Caicedo, J. C., Pentland, A. and Pisano, G. P. (2020). Overcoming barriers to early disease intervention. *Nat Biotechnol.* 38(6): 669–673. <https://doi.org/10.1038/s41587-020-0550-z>
22. Rivers, A. R., Weber, K. C., Gardner, T. G., Liu, S. and Armstrong, S. D. (2018). ITSxpress: Software to rapidly trim internally transcribed spacer sequences with quality scores for marker gene analysis. *F1000Research.* 7: 1418. <https://doi.org/10.12688/f1000research.15704.1>
23. Büyükalıy, K. (2022). IMPLEMENTATION OF KRONA INTO QIIME 2, Middle East Technical University.
24. Austin, G. I., Park, H., Meydan, Y., Seeram, D., Sezin, T., Lou, Y. C., Firek, B. A., Morowitz, M. J., Banfield, J. F., Christiano, A. M., et al. (2023). Contamination source modeling with SCRuB improves cancer phenotype prediction from microbiome data. *Nat Biotechnol.* 41(12): 1820–1828. <https://doi.org/10.1038/s41587-023-01696-w>
25. Shaffer, M., Thurimella, K., Sterrett, J. D. and Lozupone, C. A. (2022). SCNIC: Sparse correlation network investigation for compositional data. *Mol Ecol Resour.* 23(1): 312–325. <https://doi.org/10.1111/1755-0998.13704>
26. Ziernski, M., Adamov, A., Kim, L., Flörl, L. and Bokulich, N. A. (2022). Reproducible acquisition, management and meta-analysis of nucleotide sequence (meta)data using q2-fondue. *Bioinformatics* 38(22): 5081–5091. <https://doi.org/10.1093/bioinformatics/btac639>
27. Heidrich, V., Karlovsky, P. and Beule, L. (2021). ‘SRS’ R Package and ‘q2-srs’ QIIME 2 Plugin: Normalization of Microbiome Data Using Scaling with Ranked Subsampling (SRS). *Appl Sci.* 11(23): 11473. <https://doi.org/10.3390/app112311473>
28. Fouquier, J., Rideout, J. R., Bolyen, E., Chase, J., Shiffer, A., McDonald, D., Knight, R., Caporaso, J. G. and Kelley, S. T. (2016). ghost-tree: creating hybrid-gene phylogenetic trees for diversity analyses. *Microbiome.* 4(1): e1186/s40168–016–0153–6. <https://doi.org/10.1186/s40168-016-0153-6>
29. Kaehler, B. D., Bokulich, N. A., McDonald, D., Knight, R., Caporaso, J. G. and Huttley, G. A. (2019). Species abundance information improves sequence taxonomy classification accuracy. *Nat Commun.* 10(1): e1038/s41467–019–12669–6. <https://doi.org/10.1038/s41467-019-12669-6>
30. Gibbons, S. M., Duvallet, C. and Alm, E. J. (2018). Correcting for batch effects in case-control microbiome studies. *PLoS Comput Biol.* 14(4): e1006102. <https://doi.org/10.1371/journal.pcbi.1006102>
31. Martin, B. D., Witten, D. and Willis, A. D. (2020). Modeling microbial abundances and dysbiosis with beta-binomial regression. *Ann Appl Stat.* 14(1): e1214/19–aoas1283. <https://doi.org/10.1214/19-aoas1283>

32. Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. and Schmidt, T. M. (2014). rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.* 43: D593–D598. <https://doi.org/10.1093/nar/gku1201>
33. Fedarko, M. W., Martino, C., Morton, J. T., González, A., Rahman, G., Marotz, C. A., Minich, J. J., Allen, E. E. and Knight, R. (2020). Visualizing 'omic feature rankings and log-ratios using Qurro. *NAR Genomics Bioinf.* 2(2): e1093/nargab/lqaa023. <https://doi.org/10.1093/nargab/lqaa023>
34. Miranda, F. M., Azevedo, V. C., Ramos, R. J., Renard, B. Y. and Piro, V. C. (2024). Hitac: a hierarchical taxonomic classifier for fungal ITS sequences compatible with QIIME2. *BMC Bioinf.* 25(1): e1186/s12859–024–05839–x. <https://doi.org/10.1186/s12859-024-05839-x>
35. Titus Brown, C. and Irber, L. (2016). sourmash: a library for MinHash sketching of DNA. *The Journal of Open Source Software* 1(5): 27. <https://doi.org/10.21105/joss.00027>
36. Yao, X. (2021). q2-repeat-rarefy: QIIME2 plugin for generating the average rarefied table for library size normalization using repeated rarefaction, , GitHub repository, <https://github.com/yxia0125/q2-repeat-rarefy>.
37. Herman, C., Bolyen, E., Simard, A., Gehret, L. and Caporaso, J. G. (2025). Assessing microbiome engraftment extent following fecal microbiota transplant with q2-fmt. *PLoS Comput Biol.* 21(7): e1013299. <https://doi.org/10.1371/journal.pcbi.1013299>
38. Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., et al. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microbiol.* 75(23): 7537–7541. <https://doi.org/10.1128/aem.01541-09>
39. Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* 26(19): 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>
40. Rognes, T., Flouri, T., Nichols, B., Quince, C. and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ.* 4: e2584. <https://doi.org/10.7717/peerj.2584>
41. McDonald, D., Clemente, J. C., Kuczynski, J., Rideout, J. R., Stombaugh, J., Wendel, D., Wilke, A., Huse, S., Hufnagle, J., Meyer, F., et al. (2012). The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. *GigaScience.* 1(1): e1186/2047–217x–1–7. <https://doi.org/10.1186/2047-217x-1-7>