

DOI:10.12113/j.issn.1672-5565.201804007

Windows 下 16S rRNA 基因扩增子测序数据分析的简易流程

方梅梅,王禹焯,王明月,郑和龙,张璐,向沙沙,张国庆,李余动*
(浙江工商大学 食品与生物工程学院,杭州 310018)

摘要:微生物组数据分析需要掌握 Linux 系统操作,这对缺乏计算机知识的生物研究人员是一个很大的障碍。为此我们设计了一套在 Windows 的 Linux 子系统(WSL)下分析 16S rRNA 基因扩增子高通量测序数据的简易流程。本流程整合常用的开源软件 VSEARCH 与 QIIME 等,能对 16S rRNA 测序数据进行质量控制、OTU 聚类、多样性分析及结果可视化呈现。以唾液微生物组分析为例,详细介绍从原始数据到多样性统计分析过程的参数和命令,及结果解读。教学实践证明,此流程易于学习,并有助于掌握微生物组的基本概念与方法。利用 Windows 系统最新的 WSL 功能,本流程方便 Windows 用户使用大量在 Linux 上运行的生物信息工具,有助于促进微生物组研究的发展。流程的安装程序与测序数据可从网址(<http://www.ligene.cn/win16s/>)免费下载使用。

关键词:16S rRNA;微生物组;Linux;QIIME;VSEARCH

中图分类号:Q752 **文献标志码:**A **文章编号:**1672-5565(2018)04-239-07

A simple pipeline to analyze 16S rRNA amplicon sequencing data in the Windows

FANG Meimei, WANG Yuxuan, WANG Mingyue, ZHENG Helong, ZHANG Lu, XIANG Shasha,
ZHANG Guoqing, LI Yudong*

(School of Food Science and Biotechnology, Zhejiang Gongshang University, Hangzhou 310018, China)

Abstract: Biologists are required to grasp the operation of Linux systems for microbiome analysis. To overcome this barrier, we developed an easy pipeline for microbiome analysis operating in the Windows Subsystem for Linux (WSL). The pipeline combines several open-access tools (VSEARCH and QIIME) for processing 16S rRNA data, including data quality filtering, OTU clustering, taxonomic analysis, and results visualization. We reported an application of this pipeline in salivary microbiome analysis, which brings the user from raw sequencing reads to diversity-related conclusions. Practical teaching revealed that the pipeline is helpful for beginners to learn the basic concepts and methods of microbiome. It is expected to promote the development of microbiome research by utilizing the latest WSL features of the Windows, which facilitates rapid access to the bioinformatic tools for Windows users. The package and sequencing reads are freely available on the website (<http://www.ligene.cn/win16s/>).

Keywords: 16S rRNA; Microbiome; Linux; QIIME; VSEARCH

地球上的各种环境中,如土壤、海洋或人体肠道等,都生存着成千上万的微生物,包含细菌、真菌、病毒和古菌等。微生物组(Microbiome)指某一特定生态环境中全部微生物及其遗传物质的总和^[1-2]。人类微生物组的组成与功能具有非常丰富的多样性,并会随着环境的变化而动态变化。如在人类口腔中

定殖的微生物就有超过 600 种,口腔微生物菌群失衡会引起龋齿、口臭等口腔健康问题^[3]。随着高通量测序技术的快速发展,宏基因组方法可以系统研究不同环境中微生物群落的组成与多样性变化^[4]。大量研究表明,微生物组结构变化与人类疾病,如糖尿病、癌症、消化不良等密切相关^[5]。微生物组学

收稿日期:2018-04-24;修回日期:2018-06-12.

基金项目:国家自然科学基金项目(No:31671836).

作者简介:方梅梅,女,本科生,研究方向:微生物组学。E-mail:15967149972@163.com.

*通信作者:李余动,男,副教授,研究方向:微生物基因组学。E-mail:lyd@zjsu.edu.cn.

已成为最热门的研究方向之一,目前有多项国际合作大的研究项目正在开展,如人类微生物组计划(HMP)、地球微生物组计划(EMP)和中国微生物组计划等^[2,6]。

16S rRNA 基因全长约 1 542 bp,因在结构和功能上都具有高度的保守性,存在于所有的细菌中,因此常被用于微生物分类鉴定的标志物。16S rRNA 基因有 9 个高变区(V1~V9)中,V4 区是最准确且能识别到属水平的可变区,而 V3~V4 区测序长度(约 465 bp)比 V4 区更长(约 290 bp),相较单独 V4 区测序更加准确^[7]。新一代高通量测序平台 Illumina MiSeq 克服了传统高通量测序方法的读长较短与误差较大的缺陷,因而用于 16S rRNA 的 V3~V4 区测序可获得更准确的微生物物种分类信息。

高通量测序项目一般会产生海量的测序原始数据,对于原始数据的过滤筛选,加工以及统计分析,对计算机的软硬件要求较高,使没有计算机背景的研究人员难以分析数据和解释研究结果^[8]。尤其是,常用的微生物组分析软件(如 QIIME、Mothur 等)需要在 Linux 系统下运行,而普通的生物研究人员完全掌握 Linux 系统的使用非常困难^[9]。因此,如何使大多数科研人员能进行生物信息分析是生物大数据时代的一个挑战^[10]。尽管现在已经开发有一些基于网页的数据分析平台,如 MG-RAST^[11], SEED^[12]等,但是这些平台往往有许多限制,如数据不能太大,网络传输速度慢,分析任务等待时间长等。尤其是网络分析平台不适合那些需要对样本数据严格保密的项目。

微软在 2016 年推出 Windows 的 Linux 子系统—Windows Subsystem for Linux (WSL),原生支持在 windows10 上运行 Linux 程序,包括大多数的 Linux 命令与工具,如 bash, grep, ssh 等。这样大量在 Linux 平台运行的生物信息软件都可以直接安装在 WSL 下运行。WSL 使 Linux 软件安装与使用变得容易,只需要在 Windows 系统下,学习 Linux 系统的基本命令,不需要专门在电脑上或用虚拟机(如 VirtualBox)安装 Linux 系统。

为方便研究人员分析微生物组数据,我们开发了一套在 WSL-Ubuntu 下分析样品 16S rRNA 基因扩增子高通量测序的简易流程。本流程基于 Win10 的 Linux 子系统,避免了安装 Linux 的复杂过程,并简化了 QIIME^[13]、VSEARCH^[14] 及相关软件的安装过程,只要运行一个安装脚本就自动安装并配置运行所需的环境变量,方便没有计算机背景的研究人员进行相关微生物组数据分析。

1 材料与方法

1.1 唾液样本采集与 DNA 测序数据

本实验数据来源于一项针对在校大学生的口腔微生物研究项目。实验简要说明如下:首先在实验的前期阶段,我们进行样本信息的搜集,筛选和整理,后送实验室提取 DNA 和质量检验,其中,唾液样本的采集参考陈瑜等人的方法^[15],DNA 的提取使用 QIAGEN DNA Mini KIT 试剂盒提取,经 PCR 扩增 16S DNA 的 V3~V4 区后,进行琼脂糖凝胶电泳和核酸蛋白定量检测来检测提取的 DNA 的纯度及浓度。对于检测合格的样本 DNA 送测序公司采用 Illumina HiSeq 测序平台进行 16S rDNA 双端(Paired-End)测序。公司返回的测序结果为 FASTQ 格式文件的压缩包,包含所有的样本正反两个方向的序列信息。

为方便学习本分析流程,本文只使用部分唾液样本(10 个)的微生物组数据,测序数据可从本流程网站下载(<http://www.ligene.cn/win16s/>)。整个分析过程只需要约 2 h。

1.2 电脑系统与软件安装

1.2.1 装有最新版 Windows 10 系统的笔记本电脑

为了运行本分析流程,首先需要一台 Windows 10 系统的电脑,并且将该系统更新到最新版本(版本 1 709 及以上)。升级过程可以选择电脑自动升级,也可以在微软网站下载升级程序,进行手动升级(Windows 易升软件:<https://www.microsoft.com/en-us/software-download/windows10>)。在系统升级完毕后,就可开启系统的 Windows Subsystem for Linux (WSL)功能:在控制面板—“应用”—“程序与功能”—“启用或关闭 Windows 功能”—“Windows Subsystem for Linux”中开启“Windows Subsystem for Linux”,从而开启 WSL 的功能。

1.2.2 WSL+Ubuntu 安装

完成 WSL 的开启之后,在 Windows10 自带的微软应用商店“Microsoft Store”中搜索并下载 Ubuntu。当 Ubuntu app 下载完毕之后,点击“Launch”启动安装。然后窗口会提示你创建一个 Linux 的账户,以后在 Linux 环境下进行数据分析都是在这个用户名下。这个用户名和密码与电脑微软系统的账户和密码没有任何关系,因此不必设置为相同的。安装完成之后,即可以通过开始菜单的 Ubuntu 程序图标来运行 Ubuntu Linux 系统。

1.2.3 Xming 安装

目前 WSL 不支持图形界面,为了实现数据分析

结果的可视化,需要安装一个 Xming 软件,通过 Xming 官网 (<http://sourceforge.net/projects/xming/>) 下载安装包(最新版本 6.9.0.31),进行安装即可。

1.2.4 安装 win16s 流程

此流程是基于 VSEARCH (v2.7.1) 与 QIIME (v1.9.1) 等开源软件,因此要在 WSL-Ubuntu 的命令终端中输入不同软件的安装指令,所有命令集见附件 1(流程安装)。为方便初学者安装,我们提供一个 bash 脚本(win16s_install.sh),下载此文件后,在 ubuntu bash 终端使用“bash win16s_install.sh”命

令运行此脚本,即可一次安装所有流程运行依赖软件,并能自动配置合适的环境变量。

Win16s 所有软件与相关脚本可以从网址下载:
<http://www.ligene.cn/win16s/>.

1.3 win16s 分析流程

本分析流程的基本步骤如图 1 所示。主要步骤为:样本双端测序运用 VSEARCH 软件进行序列拼接, OTU 聚类,后通过比对 RDP 分类数据库注释物种,最后用 QIIME 相关脚本进行 α/β 多样性分析(如 PCoA 等)。

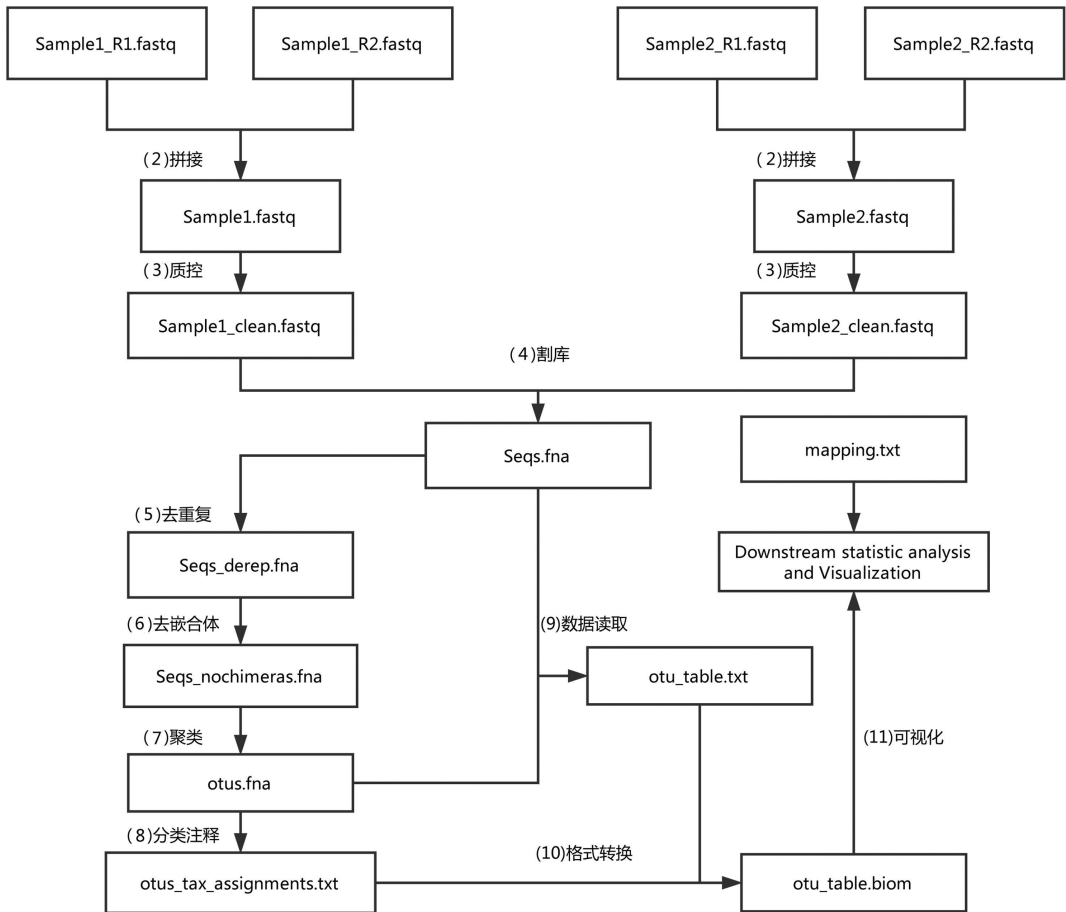


图 1 16S rRNA 基因扩增子测序数据分析流程图

Fig.1 Flowchart of 16S rRNA gene amplicon data analysis pipeline

2 结果与讨论

本结果是基于对 10 个唾液样本的 16S rRNA 基因测序数据分析的教学过程。首先,每个学生按流程分析样本数据,可以得到样本的 OTU 聚类与物种分类信息;在查看样品的细菌种类后,统计每个样品的差异菌种。然后,根据问卷调查结果对样本进行分组:荤组与素组(见表 1),通过微生物组多样性相关分析命令,统计不同组样品间的

菌群结构差异。

2.1 样本信息

为了分析不同饮食习惯大学生的唾液微生物的种群差异性,我们首选通过调查问卷来筛选出合适的志愿者,然后再统一组织开展取样工作。取样人群的基本信息,包括样本编号,性别,年龄, BMI 值,取样方法,所属组别(见表 1)。为了避免长时间的储存对样本产生的影响,对于取得的唾液样本将直接用于后续研究(具体参见材料与方法 1.1 部分)。

表 1 样本信息表

Table 1 Basic information of the samples

样品名称	性别	组别	年龄	身体质量指数	取样方法
SAF	女	荤食	22	19.23	Natural outflow method
SAG	女	荤食	20	17.85	Natural outflow method
SAJ	男	荤食	21	19.49	Natural outflow method
SAK	男	荤食	21	19.14	Natural outflow method
SAN	男	荤食	22	21.97	Natural outflow method
SAQ	女	素食	22	22.06	Natural outflow method
SAS	男	素食	20	23.26	Natural outflow method
SCB	男	素食	26	18.59	Natural outflow method
SAA	女	素食	22	31.62	Natural outflow method
SAB	女	素食	20	18.97	Natural outflow method

2.2 数据质控与预处理

为保证数据质量和避免错误结果,16s rDNA 扩增子测序数据需要质控及预处理后,再用于下游的分析。首先常用 FastQC 软件对测序数据进行质量评估,由于一般公司的测序报告文件会附带评估报告,质量太差会重测,此步非本实验必须步骤。然后,利用常用的 NGS 质控工具(如 Trimmomatic^[16])进行预处理,如过滤低质量 reads ($Q < 20$),切除 read 末端低质量碱基与未识别碱基 N、切除测序接头(Adapters)、去除长度太短的序列等。

本实验数据为 Illumina HiSeq2500 平台 Paired-end 测序(PE250),并已去除引物与 Barcodes 后的原始序列,每个样品有两个正反向测序 reads 数据文件(注意文件名书写格式,如 SAA_R1.fastq 与 SAA_R2.fastq)。正向 Reads (R1) 和反向 Reads (R2) 两段序列有部分重叠可进行拼接。一般默认用 fastq-join 方法拼接,所得序列再过滤低质量 reads 与切除末端低质量碱基。由于 16S rRNA 的 V3~V4 区长度为 465 bp,本实验过滤长度小于 380 bp 的序列片段。原始序列经过滤后,最终每个样本得到约 15 000 条有效 reads(见表 2)用于后续分析,好的质量控制可以得到准确的分类结果。

2.3 OTU 聚类

OTU(Operational Taxonomic Units)是在系统发生学研究中,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。要了解一个样品测序结果中的菌种、菌属等数目信息,就需要对序列进行聚类操作(Clustering)。通过聚类将序列按照彼此的相似性 97% 以上归为一个 OTU。一般建议采用开放参考(Open reference)聚类方式,但是 PICRUSt 预测需要以默认 Greengene 数据库的封闭参考(Closed reference)聚类分析结果。

本实验对样品数据进行质量控制、序列拼接及去除嵌合体后,选择序列相似度值 97% 进行 OTU 聚类。细菌的 16S rRNA 分类一般使用 97% 相似度,而真核生物的 18S rRNA 分类一般使用 96% 相似度,使用不同的算法或相似度阈值可能得到完全不同的 OTU 数^[17]。得到 OTU 后,采用 RDP 算法与 Greengenes rDNA 数据库比对,注释各 OTU 的分类单元。大多数 OTU 序列能注释到属水平,一部分 OTU 可以鉴定到种水平。

本实验共得到了 212 个 OTU,在门和属两个水平上对同类的 OTU 进行分类鉴定,归属于 12 个门,94 个属(见表 2)。通过 OTU 聚类及分类鉴定得到的结果,可以作为后续菌群统计分析的基础。

2.4 细菌种群的丰富度

不同人体的口腔微生物种群存在着一定的差异,通过分析可以得到不同分类水平的细菌种群结构及其丰富度,可以清晰地得到不同样本在同一分类水平的差异。分析流程在进行可视化操作后,在 taxa_summary 文件夹中生成子文件夹 taxa_summary_plots,打开其中的文件 bar_charts.html,即可观察到各样品自门水平(L2)到属水平(L6)的分类及其组成比例。图 2 是在门和属水平对菌群的差异性的结果。具体菌群分类与丰度信息可以在多样性分析输出目录("taxa_summary/"),查看在不同分类水平下的细菌种群的丰富度。由结果可知,部分菌种在所有样本中都存在;而少数菌种则只在部分样本中存在。

通过对不同样本分组的种群丰富度分析,本实验共检出优势菌涉及 12 个门(见图 2(a)),其中 Firmicutes、Proteobacteria、Bacteroidetes、Actinobacteria 不存在显著性差异,而 Fusobacteria 存在显著性组间差异(wilcox.test, $P < 0.05$)。本实验共检出优势菌

涉及 94 个属(见图 2(b)), 荤食组中 *Streptococcus* 减少(wilcox.test, $P < 0.05$)。比素食组显著增加,而 *Fusobacterium* 比素食组显著

表 2 唾液细菌测序的基本信息
Table 2 Basic sequencing information of bacteria in saliva

样品名称	有效 reads 数目/条	OTU 数目/个	不同分类阶元归类数目/个	
			门	属
SAA	14 356	421	8	82
SAB	16 772	456	12	90
SAF	14 441	463	11	81
SAG	14 377	403	9	62
SAJ	11 631	406	8	90
SAK	14 181	450	11	79
SAN	16 368	379	10	74
SAQ	18 430	418	12	81
SAS	14 832	389	11	67
SCB	18 749	450	11	87

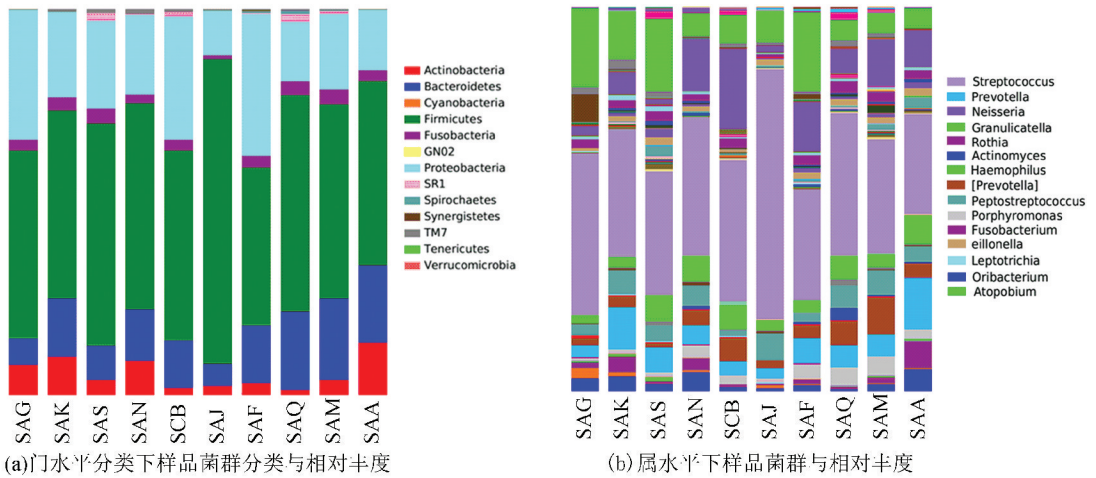


图 2 实验样品菌群分类与相对丰度

Fig.2 Taxonomy and relative abundance of the bacterial populations across samples

2.5 其它统计分析

最后得到的 biom 文件(otus_table_tax.biom)可用于后续统计分析。通常可以计算两类多样性指数:Alpha 多样性指数(α -diversity)与 Beta 多样性指数(β -diversity)。Alpha 多样性指数度量的是群落内部的物种多样性, Beta 多样性指数度量的是群落间的物种多样性。按照生态学的观点,多样性高的生态群落,抵抗力和稳定性都可能更强。这里以 QIIME 流程中的 core_diversity_analyses.py 脚本分析,并通过最小测序量样本的 reads(Counts)数作为 -e 参数(Rarefaction 抽样数)。core_diversity_analyses.py 分析运行一系列标准的 alpha 与 beta 多样性分析,可以查看分析输出文件夹(core_diversity/)下的文件查看相关结果。最后,可以将 OTU 表的 biom 格式文件转换成文本文

件,用于其它统计分析软件的分析,如 STAMP^[18]或 R 软件包 phyloseq, vena 等。

2.5.1 样本菌群 Alpha 多样性分析

本实验统计了样本 Alpha 多样性分析中的不同多样性指数。图 3(a)是系统发育多样性指数(PD),两组的多样性非常相似,虽然两条曲线(PD 值)区分明显,但误差棒有重叠,即没有显著统计差异($P > 0.05$)。另外,荤食组和素食组 Simpson 多样性指数分别为(0.94±0.04)与(0.95±0.02), Shannon 多样性指数分别为(5.54±0.63)与(5.78±0.30),可见两组样本有着相似的多样性指数。

2.5.2 样本菌群 Beta 多样性分析

本实验利用 QIIME 的脚本对得到的 OTU 表做差异距离矩阵,后用主坐标分析(Principal

coordinates analysis, PCoA) 将距离矩阵可视化。PCoA 结果如图 3(b) 所示, 其中一个点代表一个样本, 相同的颜色表示是同一组的样本, 两个点之间的距离越近, 说明两个样品的微生物群落差异越小。由 PCoA 分析结果(见图 3(b)) 可知两组样本

均比较分散, 主成分区分度不是特别明显, 但在各自象限均占主要优势, 说明两组样本的主成分有一定差异, 但没有显著差异。这可能与本研究的采样对象样本数量太少有关, 还需要扩大取样人群范围进一步研究。

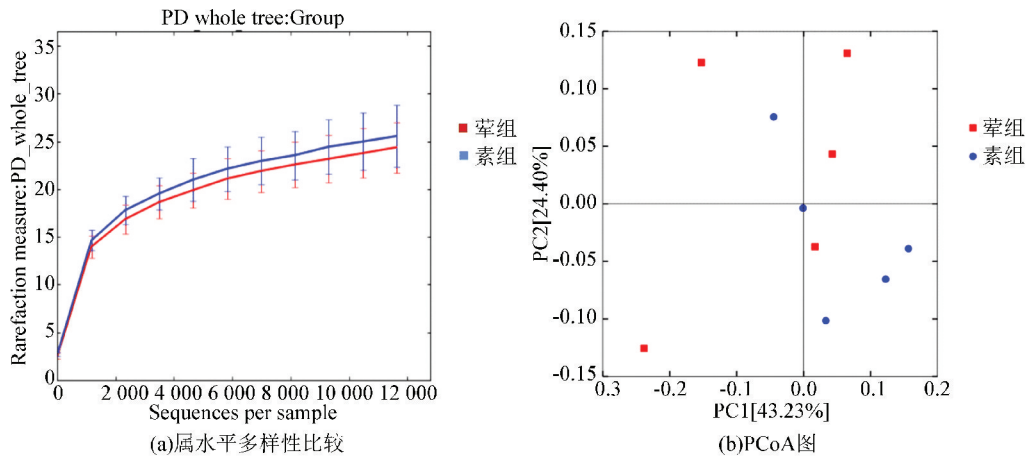


图 3 样本分组的菌群差异

Fig.3 Bacterial population difference between sample groups

生活习惯和饮食习惯的不同, 形成具有差异的口腔微生物群落结构, 存在不同的优势菌和劣势菌。它们影响着口腔的环境, 对人体造成不同的影响。通过分类统计分析研究不同组别的菌群组成差异, 进一步研究菌群差异与样本之间存在的关系, 是微生物组研究的意义和目的。

3 结论

通过整合开源软件 VSEARCH 与 QIIME 等分析流程, 我们开发了在 Windows 系统下分析 16S rRNA 基因扩增子测序数据的简易流程。本流程不仅适用于 16S rRNA 数据分析, 还可以用于 18S rRNA 或 ITS 等标记基因的分析。以后随着研究的深入, 我们将提供更多元化的分析脚本, 不断更新现有分析流程, 并同时在网上公布, 供初学者学习微生物组分析方法^[19]。

WSL 是微软在 Windows 10 中推出的 Linux 子系统功能, 允许 Linux ELF64 二进制文件运行在 Windows 系统上。随着微软对 WSL 的改进, 越来越多的生物信息软件将支持在 WSL 运行, 从而方便广大生物学背景的人员在 Windows 系统下进行生物信息数据分析^[20]。然而 WSL 也有一定局限, 如它只支持 64 位程序, 一些老的 32 位程序将不能在 WSL 下运行, 如 QIIME 默认的 uclust 程序不能在 WSL 下编译成功。好在目前大多数 32 位程序都有相应的

64 位替代程序, 如 uclust、USEARCH^[20] 可以用免费的 64 位程序 VSEARCH 代替。另外, 在 WSL 下, 程序对硬盘的读写效率还比较低, 程序运行的效率还是比在原生 Linux 系统下运行的速度慢。但 WSL 适合作为数据分析入门的学习环境, 当研究人员习惯了 Linux 命令操作, 可以很方便迁移到原生 Linux 计算机上进行实际项目数据的分析研究工作。

参考文献 (References)

- [1] WHITESIDE S A, RAZVI H, DAVE S, et al. The microbiome of the urinary tract—a role beyond infection[J]. *Journal of Urology*, 2015, 12(2): 81. DOI: 10.1016/j.juro.2015.09.053.
- [2] 刘双江, 施文元, 赵国屏. 中国微生物组计划: 机遇与挑战[J]. *中国科学院院刊*, 2017, 32(3): 241–250. DOI: 10.16418/j.issn.1000–3045.2017.03.004. LIU Shuangjiang, SHI Wenyuan, Zhao Guoping. China microbiome initiative: Opportunity and challenges[J]. *Bulletin of Chinese Academy of Sciences*, 2017, 32(3): 241–250. DOI: 10.16418/j.issn.1000–3045.2017.03.004.
- [3] 周学东, 徐健, 施文元. 人类口腔微生物组学研究: 现状、挑战及机遇[J]. *微生物学报*, 2017, 57(6): 806–821. DOI: 10.13343/j.cnki.wsxb.20170063. ZHOU Xuedong, XU Jian, SHI Wenyuan. Human oral microbiome: Progress, challenge and opportunity[J]. *Acta Microbiologica Sinica*, 2017, 57(6): 806–821. DOI: 10.13343/j.cnki.wsxb.20170063.
- [4] WOOLEY J C, GODZIK A, FRIEDBERG I. A primer on

- metagenomics [J]. *PLoS Computational Biology*, 2010, 6(2):e1000667. DOI:10.1371/journal.pcbi.1000667.
- [5] DEWEERDT S. Microbiome: Microbial mystery [J]. *Nature*, 2015, 521(7551):S10. DOI:10.1038/521S10a.
- [6] DUBILIER N, MCFALL-NGAI M, ZHAO L P. Create a global microbiome effort[J]. *Nature*, 2015, 526: 631–634. DOI:10.1038/526631a.
- [7] WU X, ZHANG H, CHEN J, et al. Comparison of the fecal microbiota of dholes high-throughput Illumina sequencing of the V3-V4 region of the 16S rRNA gene[J]. *Applied Microbiology and Biotechnology*, 2016, 100(8): 3577–3586. DOI:10.1007/s00253-015-7257-y.
- [8] WANG J T, DALY J N, WILLNER D L, et al. Do you kiss your mother with that mouth? An authentic large-scale undergraduate research experience in mapping the human oral microbiome [J]. *Journal of Microbiology and Biology Education*, 2015, 16(1): 50–60. DOI: 10.1128/jmbe.v16i1.816.
- [9] STEVENS J L, DEHORITY R, GOLLER C C. Using QIIME to interpret environmental microbial communities in an upper level metagenomics course [J]. *CourseSource*, 2017, (4): 1–6. DOI:10.24918/cs.2017.3.
- [10] MARX V. The big challenges of big data [J]. *Nature*, 2013, 498(7453): 255–260. DOI:10.1038/498255a.
- [11] MEYER F, PAARMANN D, D'SOUZA M, et al. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes [J]. *BMC Bioinformatics*, 2008, 9(1): 386–386. DOI: 10.1186/1471-2105-9-386.
- [12] VETROVSKY T, BALDRIAN P, MORAIS D. SEED 2: A user-friendly platform for amplicon high-throughput sequencing data analyses [J]. *Bioinformatics*, 2018, DOI: 10.1093/bioinformatics/bty071.
- [13] ROGNES T, FLOURI T, NICHOLS B, et al. VSEARCH: A versatile open source tool for metagenomics [J]. *PeerJ*, 2016, 4(10):e2584. DOI:10.7287/peerj.preprints.2409v1.
- [14] CAPORASO J G, KUCZYNSKI J, STOMBAUGH J, et al. QIIME allows analysis of high-throughput community sequencing data [J]. *Nature Methods*, 2010, 7(5): 335–336. DOI: 10.1038/nmeth.f.303.
- [15] 陈瑜, 刘婉薇, 李良芳, 等. 功能性消化不良患者唾液菌群的研究 [J]. *重庆医学*, 2017, 46(13): 1789–1796. DOI: 10.3969/j.issn.1671-8348.2017.13.020.
- CHEN Yu, LIU Wanwei, LI Liangfang, et al. Analysis on saliva microbiome in patients with functional dyspepsia [J]. *Chongqing Medicine*, 2017, 46(13): 1789–1796. DOI: 10.3969/j.issn.1671-8348.2017.13.020.
- [16] BOLGER A M, LOHSE M, USADEL B. Trimmomatic: A flexible trimmer for Illumina sequence data [J]. *Bioinformatics*, 2014, 30(15): 2114–2120. DOI:10.1093/bioinformatics/btu170.
- [17] 杨晨雪, 季吟秋, 王晓阳, 等. 基于18S rDNA的 metabarcoding 技术分析土壤小型动物多样性 3 种方法的比较 [J]. *中国科学: 生命科学*, 2012, 42(12): 993–1001. DOI: 10.1360/zc2012-42-12-993.
- YANG Chenxue, JI Yinqiu, WANG Xiaoyang, et al. Testing three pipelines for 18S rDNA-based metabarcoding of soil faunal diversity [J]. *Scientia Sinica Vitae*, 2012, 42(12): 993–1001. DOI: 10.1360/zc2012-42-12-993.
- [18] PARKS D H, TYSON G W, HUGENHOLTZ P, et al. STAMP: Statistical analysis of taxonomic and functional profiles [J]. *Bioinformatics*, 2014, 30(21): 3123–3124. DOI: 10.1007/978-1-4614-6418-1_780-1.
- [19] MORAIS D, ROESCH L, REDMILE-GORDON M, et al. BTW-bioinformatics through windows: An easy-to-install package to analyze marker gene data [J]. *PeerJ*, 2018, 6: e26581v1. DOI:10.7287/peerj.preprints.26581v1.
- [20] EDGAR R C. Search and clustering orders of magnitude faster than BLAST [J]. *Bioinformatics*, 2010, 26(19): 2460–2461. DOI:10.1093/bioinformatics/btq461.