

# 大豆根系微生物的基因鉴定与基因功能预测

——竞赛单元“2-1 基因组”

团队名称：水哥微生物小队

指导老师：郑金水

团队成员：张敦彪，张子栋，颜旭，姚代洪

## 摘要

高通量培养与鉴定<sup>[1]</sup>可用于从给定环境和植物物种的本地根微生物组样本中建立一个分类学上全面的细菌培养收集，通过 16S rRNA 基因扩增子测序对根样品的细菌多样性进行评价，筛选出具有相应 16S rRNA 基因序列的培养菌。通过对大豆根系微生物的高通量培养筛选可以鉴定其基因的类型，进而对其基因功能进行预测。

关键字：高通量培养，基因鉴定，功能预测

## 1 引言

### 1.1 根系微生物群

根系微生物是指生活在植物根系附近的微生物群落，包括细菌、真菌和古菌等。它们与植物根系形成一种共生关系，相互作用并互利共生。根系微生物可以对植物生长和健康产生积极影响。研究根系微生物对于促进植物生长健康、提高作物产量质量、维护土壤生态系统健康以及发掘新的生物资源具有重要意义。

传统的微生物培养方法是基于单个菌落的分离和培养，但是由于绝大部分微生物无法在常规培养条件下生长，很多微生物资源依然无法被发现和利用。高通量培养是一种用于微生物的培养和筛选的技术，通过该技术可以大大提高微生物的培养效率和筛选速度，同时也有助于扩展微生物资源库，挖掘未知微生物代谢产物等领域的研究。而高通量培养可以利用微型板或微流控芯片等设备将微生物单个分离后进行多重并行培养，以大幅提高微生物的培养效率。此外，高通量培养还可以借助自动化解决方案，实现对大量微生物的快速筛选。

本文从新鲜大豆根系（根际、根内）中高通量分离培养细菌，使用梯度稀释的方法增加获得单一细菌的比例，采用双侧标签 PCR 扩增法高通量鉴定分离培养细菌的 16S rRNA 基因。同时开发了一套从二代测序下机数据到基因鉴定个基因功能分析的全套流程。

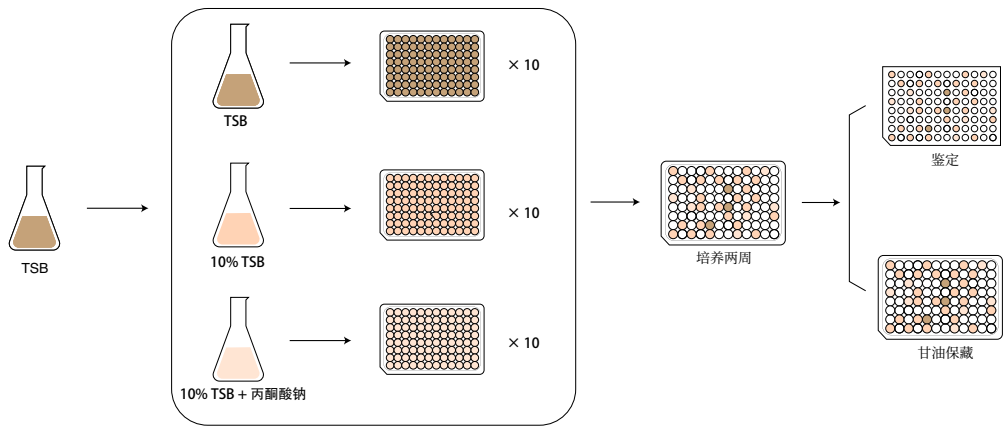


图 1 高通量培养流程

2 材料与方法

2.1 培养基的制作

实验中需要三种不同的培养基（三种培养基均高压蒸汽灭菌）：

表 1 实验所需培养基

培养基类型	配置方法
TSB 培养基	30 g/L TSB，蒸馏水定容
10% TSB 培养基	TSB 培养基按照体积分数 10% 进行稀释
10% TSB 培养基 + 丙酮酸钠	30 g/L TSB，10 mM/L 丙酮酸钠，蒸馏水定容

2.2 大豆根系菌群的获取

在华中农业大学 BMB 课题组大豆孢囊线虫大棚中摘取自然条件下生长、表型正常的未开花的大豆，在取样前不宜过多浇水，以免造成大豆根系附着较多土壤，不易除去。用手轻轻拨去大豆根系的土壤以去除大量土壤团聚体，此后将大豆根系置于手掌虎口处轻轻抖动，去除碎土。

(1) 根际样本

超净工作台中将大豆根系置于无菌滤纸上，使用无菌刀片截取其下方约 3cm 的根系，总共截取 3g 的大豆根系样本。拿取 10x PBS 溶液，按照体积分数的 10% 稀释为 1x PBS 溶液约 100 mL 置于两个 50 mL 的离心管中。将截取的根系样本置于 50 mL 的 1x PBS 溶液中来回水平

摇匀，将其以 1800 rpm 的速度震荡 20 min，取出其中的根系，将溶液过 500 目筛，获得根际样本。

## (2) 根内样本

按照上述方法新截取约 3 g 的大豆根系，使用无菌水持续冲洗，直至除去所有的土壤颗粒。将根系转移至一个新的含有 50 mL 的 1x PBS 溶液的离心管中，以 1800 rpm 的速度震荡处理 20 min，从离心管中取出根系，使用无菌滤纸吸干，从根表面除去 PBS 溶液，使用无菌水冲洗三次，将根系置于含有 50 mL 无菌水的离心管中，进行超声处理 5 min，将根系取出置于含有 50 mL 的 75% 体积分数的酒精溶液的离心管中，震荡 1 min，取出根系，使用无菌水冲洗三次，将根系悬浮在 50 mL 的 10 mmol/L 的无菌  $MgCl_2$  溶液中，使用无菌研钵研碎根系直至其均匀，将研钵内液体过 500 目筛获取根内样本<sup>[2]</sup>。

根内样本与根际样本均置于 4°C 储存。

## 2.3 分液与流式细胞分选

用分液仪将液体培养基分装到 60 个 384 孔板，再用流式细胞分选仪将根际与根内样本转移到每个孔中并做好标记。

## 2.4 高通量培养

高通量培养是一种用于快速筛选和培养大量微生物的方法。它利用自动化设备和高效的培养技术，可以同时处理多个微生物样品，并快速判断微生物的生长状况和特性。本实验中通过流式细胞分选仪连续地将多个样本注入仪器中进行分析 and 分选，从而实现快速、连续地处理大量样本，提高分选效率。对比需要一个月时间传统的分菌方法，高通量培养只需要一天的时间便可完成。

## 2.5 PCR 扩增与电泳鉴定

将得到的 PCR 管每个孔中按表 2 比例配置反应体系，按照对应程序将 PCR 管放入 PCR 仪按照表 3 程序进行扩增。

实验采上下游 PCR 引物设计，下游引物是针对目标基因（或序列）末端的引物，而上游引物则与目标基因（或序列）起始点相匹配。用 F 与 R 引物测定了培养菌全长的 16S rRNA 基因序列。

通过 PCR 后的产物电泳结果与 DNA Marker 进行对比来检验 PCR 产物是否合格。

表 2 PCR 反应体系		表 3 PCR 程序设定			
成分	体积 (μL)	循环数	流程	温度	时间
F 引物	1	1	预变性	98℃	10 min
R 引物	1		变性	95℃	30 s
ddH <sub>2</sub> O	7	2-36	复性	55℃	30 s
Mix	10		延伸	72℃	90 s
原样液	1	37	终延伸	72℃	5 min

2.6 产物回收并测序

(1) 产物回收

短暂离心 PCR 产物后用移液枪测体积，并转移至灭菌的 2 mL 离心管中，加入 Buffer DP，混匀 15 S 后加入加入 2.5 倍体积 Buffer DP 和 1 倍体积异丙醇，离心收集管壁液滴。将 HiPure 柱子套在收集管中，混合液转移至柱子，10000 × g 离心 60 s，重复上述步骤 5-6 次直到混合体积小于 700 μL 后倒掉滤液。柱子套到回收管中，加入 500 μL Buffer DW2，10000 × g 离心 60 s 后倒掉滤液，套在回收等中，10000 × g 离心 2 min。将柱子套在 2 mL 的离心管中，加入 25 μL Elution Buff 至柱子的膜中央，静置 1 min 后 10000 × g 离心 1 min，最后得到产物。

(2) 浓度测定

使用微量分光光度计来测定得到产物的浓度。

(3) 测序

为了让来自不同板的菌体的 DNA 总量大概一致，按浓度混合后再送往测序公司进行测序。

2.7 扩增子分析

(1) 16S rRNA 菌种鉴定

16S rRNA 序列由高度保守的不可变区和相对可变的可变区组成，不可变区域是 16S rRNA 序列中相对较短的区域，其在细菌和古菌中高度保守，即基本保持相同的序列。

16S rRNA 是一个长序列，经过多年的研究和数据库的积累，已经建立了广泛和全球性的 16S rRNA 数据库<sup>[3]</sup>。这些数据库中包含了大量的细菌和古菌的 16S rRNA 序列，使得通过比对新测序的 16S rRNA 序列与数据库中的序列进行比较成为可能。通过将新测序的 16S rRNA 序列与数据库中已知的序列进行比对，可以根据比对结果确定其与已知物种或菌株的相似性，并进一步进行物种或菌株的鉴定。

培养组（Culturome）流程<sup>[4]</sup>采用 Shell/R/Perl 语言编写脚本，整合了微生物组分析常用软件 QIIME、VSEARCH、GraPhlAn 等，定制完成的高通量培养物鉴定流程，可实现双端标签化扩增子文库的分析和可视化全流程。

## (2) 质量分析

使用 FastQC 进行质量分析。测序结果较好，均符合预期。

结果保存至 `result\fastqc.test\BeanRoot_fastqc.html`。

## (3) 数据预处理

在使用 QIIME2 之前，要先使用 R 手动编写并导出符合 QIIME2 要求<sup>[5]</sup>的 Metadata mapping file。由于二代测序下机数据为双端测序序列，在开始数据分析之前，需要使用 VSEARCH 进行双端测序合并。然后从序列中拆分出用于识别菌孔的 Barcode 和引物，得到扩增子。随后利用拆分的 Barcode 与 Mapping file 将序列重命名为板和孔的位置。

同时我们也完成了对每板每孔数据读长、读长数量的统计（如图 2）。

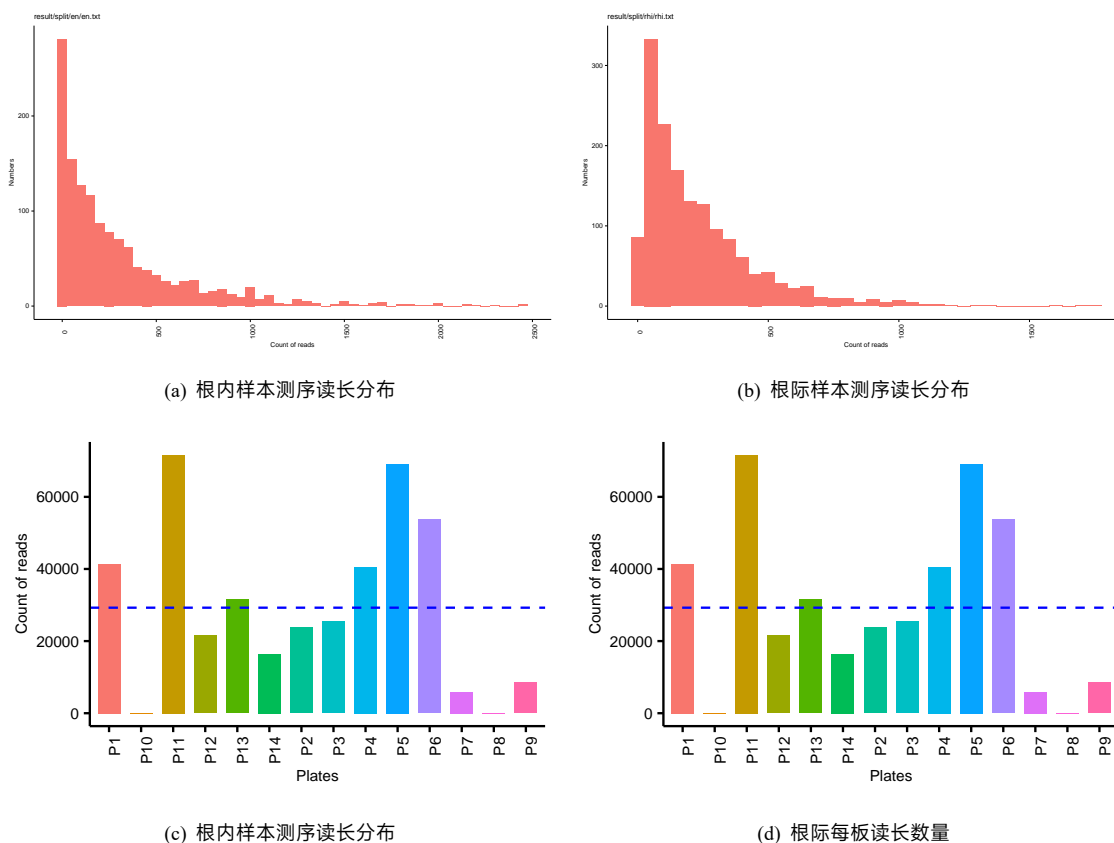


图 2 测序读长分布

## 2.8 代表性序列鉴定和物种注释

### (1) 挑选代表性序列

挑选出代表性序列作为物种的代表是扩增子分析的关键步骤，主要包括聚类生成 OTU 和去噪生成 ASV 两类方法。UPARSE 算法将具有 97% 相似性的序列聚类为 OTU，但此方法可能无法检测“种”或“株”之间的细微差异。DADA2 是最近开发的一种去噪算法，可挑选出更准确的代表性序列 ASV。QIIME 流程中有两种去噪方法可选，即 DADA2 插件的 `denoise-paired` 和 Deblur 插件的 `denoise-16S`，此外 USEARCH 中的 `-unoise3` 也可用于高速去噪并挑选 ASV。最后，可以通过量化每个样本中特征序列的频率来获得特征表，即 OTU 或 ASV 表。同时，可以对特征序列进行分类，通常在界，门，纲，目，科，属和种的层级上进行分类，从而为微生物群提供了更多级别的降维视角。本文使用 ASV 方法，相比 OUT 聚类精度更高，结合去噪算法去除噪音序列，能全包分菌保留更高的多样性。

在 PCR 反应中，延伸阶段由于不完全延伸，会导致嵌合体序列的出现。通常在 PCR 过程中，大概有 1% 的几率会出现嵌合体序列，而要在 16S/18S/ITS 扩增子测序的分析中，系统相似度极高，嵌合体可达 1%-20%，需要去除嵌合体序列。我们使用 VSEARCH 通过序列自身比对 *de novo* 去除嵌合体。

### (2) 物种注释

采用 USEARCH 的 `sintax` 命令基于 RDP 数据库进行物种注释，结果整理为制表符分隔的表格（附件 `result/*_ASV_table.txt`）。随后鉴定每个 ASV 序列对应的候选菌株。此步包括整合 ASV 和物种注释表、评估 ASV 多样性的饱和度、ASV 或属水平种类的分布、以及每个孔中培养菌的纯度。输出结果主要包括两个表：ASV 列表（`isolate_ASV.txt`）包括 5 个最佳候选菌孔位置信息、纯度和测序量；孔列表（`isolate_well.txt`）包括 ASV 的纯度、测序量和物种注释信息，随后运行脚本进行可视化（如图 3、4）。

## 2.9 功能预测分析与结果可视化

### (1) 功能预测

基于 PICRUST2<sup>[7]</sup> 开展基因功能预测。相比 PICRUST，PICRUST2 包含一个更新的，更大的基因家族和参考基因组数据库，可与任何可操作的分类单位 (OTU) 筛选或去噪算法互操作，并能够进行表型预测。PICRUST2 运行简单，但负载较大，需要 16 GB+ 内存，仅需 ASV 序列和丰度两个文件作为输入，输入基因家族和途径丰度参数即可开始工作流。

### (2) 结果可视化

使用 R 包 `ggpicrust2` [8] 进行结果可视化。`ggpicrust2` 直接接收 PICRUST2 的输出文件作为输入，并进行差异丰度分析。但介于软件仍处于开发早期，如果通路显著的数量较多（大

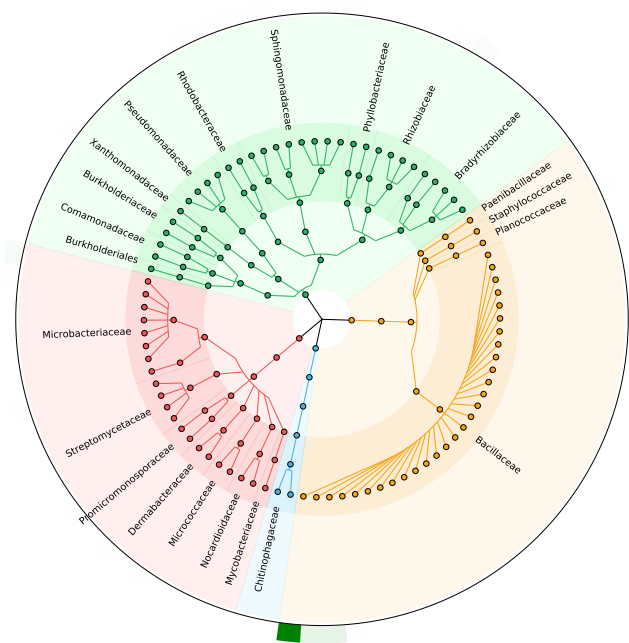


图 3 培养菌进化分支树（根内）

p\_Bacteroidetes

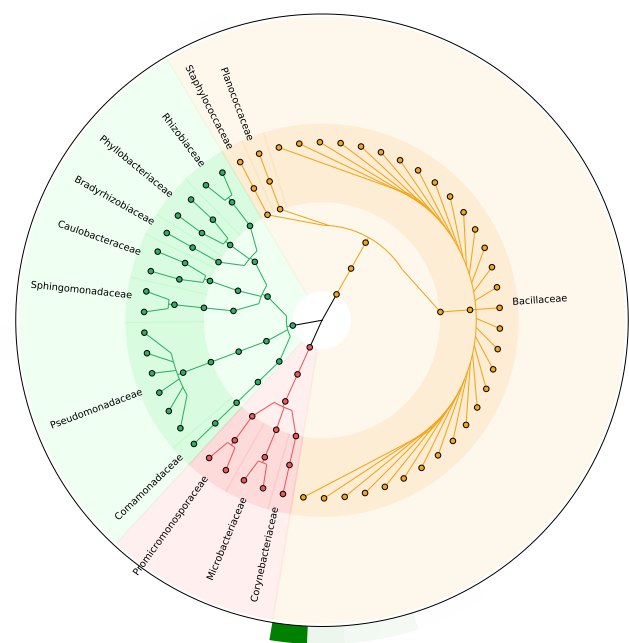


图 4 培养菌进化分支树（根内）

于 30)，则无法生成可视化图像，所以本文中每次仅输入 30 个通路进行分析，得到根内与根际的功能丰度柱状图和 PCA 图（如图 5、6，其他图像见附件）。

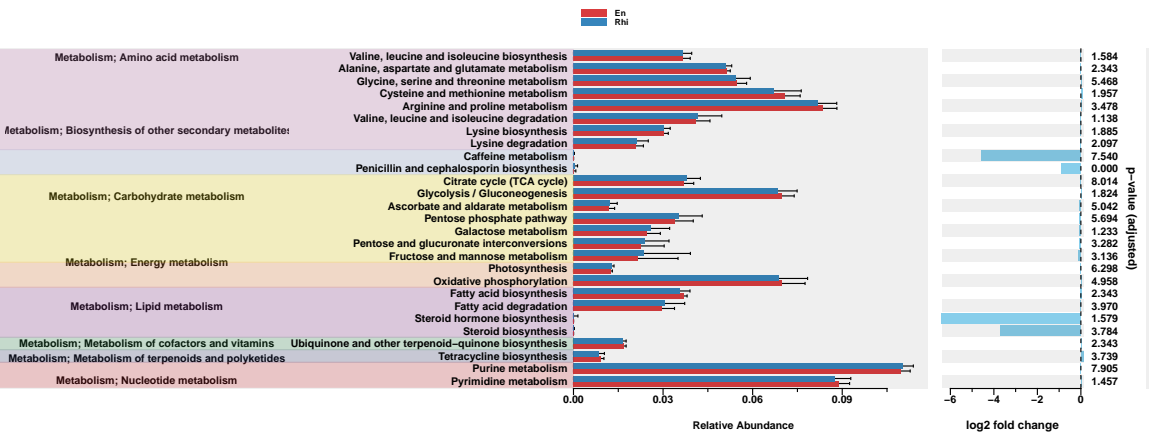


图 5 根内、根际微生物功能丰度柱状图

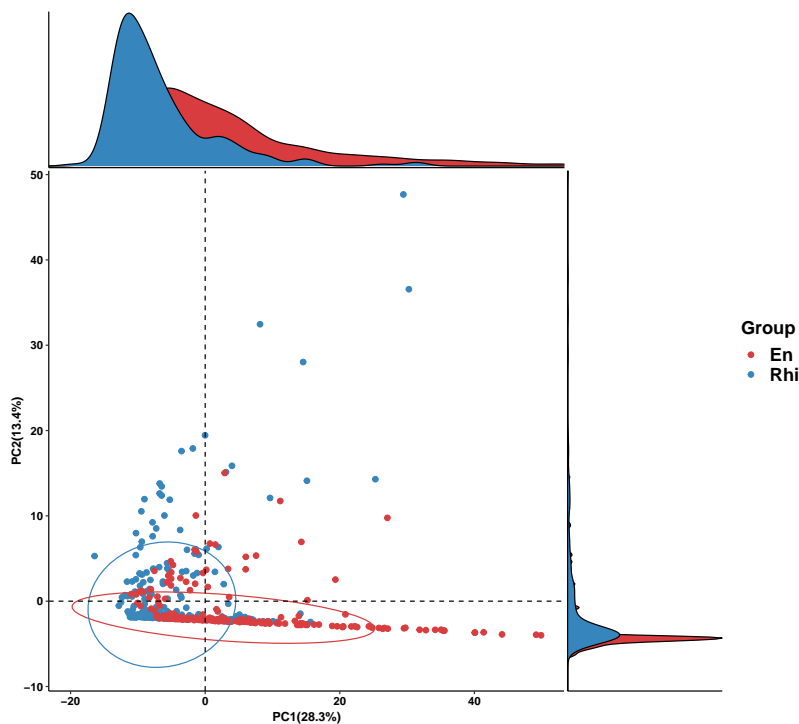


图 6 根内、根际微生物功能丰度 PCA 图

### 3 总结与展望

以大豆为生物材料，我们采集了大豆的根系样品，分为根际、根内两组进行高通量分离培养，并设置三种不同比例的培养基，培养两周后，对 30 个 96 孔细胞培养板每个孔中的细菌进行鉴定。目前基于二代 16S rRNA 扩增子测序结果的分析流程已经比较完善，但由于读长限制，



二代扩增子挑选可变区是一个主要问题,这意味着部分信息的丢失。而三代测序读长长的优点,可以轻松覆盖 16S rRNA 全长约 1500 bp 共 9 个可变区,最大程度保留了物种鉴定的可能性。但在软件和分析流程方面仍可完善,例如当前的 Phred33/64 评分标准不适用于三代测序结果,从而无法在 fastq 文件内存入有效的质量信息,而目前的分析软件均接受 fastq 文件作为输入,如 QIIME。所以基于三代测序的 16S rRNA 全长数据分析仍有很大的研究潜力。

## 参考文献

- [1] Jingying Z, Xin Y L, Xiaoxuan G, et al. High-throughput cultivation and identification of bacteria from the plant root microbiota[J]. Nature Protocols,2021,16(2). <https://doi.org/10.1038/s41596-020-00444-7>
- [2] 李亨玉,李太元. 益生菌发酵工艺中蜡样芽孢杆菌的分离鉴定 [J]. 当代畜禽养殖业, 2019(6):2. CNKI:SUN:DAYZ.0.2019-06-006.
- [3] Nadav B, Grant G, H C W, et al. Genome-wide prediction of disease variant effects with a deep protein language model.[J]. Nature genetics,2023. <https://doi.org/10.1038/s41588-023-01465-0>
- [4] Jingying Zhang,Yong-Xin Liu,Xiaoxuan Guo,Yuan Qin, Ruben Garrido-Oter,Paul Schulze-Lefert,Yang Bai. 2021. High-throughput cultivation and identification of bacteria from the plant root microbiota. Nature Protocols 16: 988-1012. <https://doi.org/10.1038/s41596-020-00444-7>
- [5] Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciulek T, Kreps J, Langille MGI, Lee J, Ley R, Liu YX, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Priesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, and Caporaso JG. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology 37: 852–857. <https://doi.org/10.1038/s41587-019-0209-9>
- [6] QIIME 2 development team. Input Files - Homepage(访问于 2023 年 9 月 18 日) [EB/OL][http://qiime.org/documentation/file\\_formats.html](http://qiime.org/documentation/file_formats.html)

- [7] Douglas, G.M., Maffei, V.J., Zaneveld, J.R. et al. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol* 38, 685–688 (2020). <https://doi.org/10.1038/s41587-020-0548-6>
- [8] Chen Yang and others. (2023). ggpicrust2: an R package for PICRUSt2 predicted functional profile analysis and visualization. [J]. *Bioinformatics*, btad470. <https://doi.org/10.1093/bioinformatics/btad470>