

16S 检测报告

2018-08-13

Pan

一、 样本信息

本次分析使用数据集为 NCBI SRA 编号 [SRP156211](#)。实验检测了来自严重急性营养不良儿童粪便样本的 16S, 18S 和反式 ITS 扩增子的原始序列。报告挑选了其中 4 个 16S 样本进行分析。

表 1-1 样本信息

SRA 编号	类型	上传时间	其他信息	检测平台
SRR7639091	16S-V4	2018-08-02	616	Illumina
SRR7639092	16S-V4	2018-08-02	617	Illumina
SRR7639093	16S-V4	2018-08-02	618	Illumina
SRR7639094	16S-V4	2018-08-02	619	Illumina

二、 数据质量统计

2.1 测序数据情况汇总

使用自编脚本（`gunzip -c fq.gz | awk '{s++}END{print s/4}'`）进行统计。

表 2-1 测序情况汇总

SRA ID	Sample Name	Row reads
SRR7639091	616	180979
SRR7639092	617	152168
SRR7639093	618	177523
SRR7639094	619	153823

注：

- (1) SRA ID: 数据在 SRA 数据库中的 ID;
- (2) Sample Name: 数据在实验中的命名;
- (3) Row reads: 原始数据 reads pair 数。

2.2 测序质量分布图

使用 DADA2 (Callahan BJ , et al.) 读入原始数据并进行统计。采取过滤阈值为 Q20。将过滤掉 forward 中循环数为 150 以上的 reads；过滤掉 reverse 中循环数为 130 以上的 reads。下图为序列测序质量分布图，其中横坐标为测序的循环数，纵坐标为 phred 值。

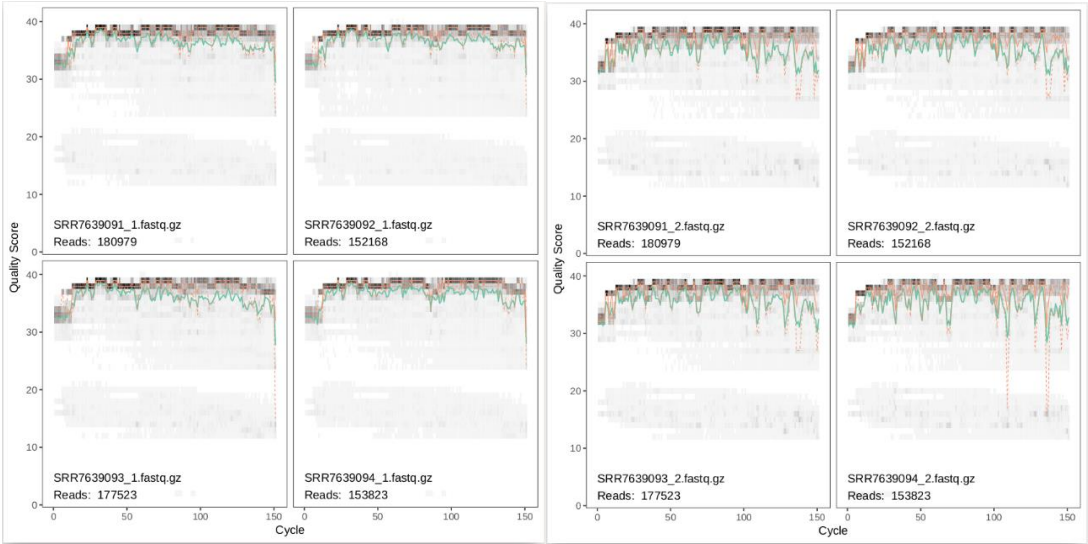


图 2-1 序列测序质量分布图

2.3 过滤后数据情况汇总

表 2-2 过滤后数据情况汇总

SRA ID	Sample Name	Row reads	Clean reads	Filter%
SRR7639091	616	180979	171855	94.96
SRR7639092	617	152168	144277	94.81
SRR7639093	618	177523	167065	94.11
SRR7639094	619	153823	145825	94.80

- 注：
- (1) SRA ID：数据在 SRA 数据库中的 ID；
 - (2) Sample Name：数据在实验中的命名；
 - (3) Row reads：原始数据 reads pair 数；
 - (4) Clean reads：过滤后的 reads pair 数；
 - (5) Filter%：过滤通过率。

三、 生成 ASVs

ASV (Amplicon Sequence Variants) 指的是“扩增子序列变体”。ASV 可以通过处理测序错误的方法，实现单个核苷酸分辨率的序列分类。ASV 被认为代表真正的生物序列（大部分修正和排序的序列错误）。DADA2 算法采取生成 ASVs 的方式来分析扩增子序列。

3.1 计算错误率

下图左为 F 端错误率统计图、图右为 R 端错误率统计图。其中横坐标为质量得分，纵坐标为错误率的 log10 对数。黑色线为错误率收敛线，红色线为期望错误率。

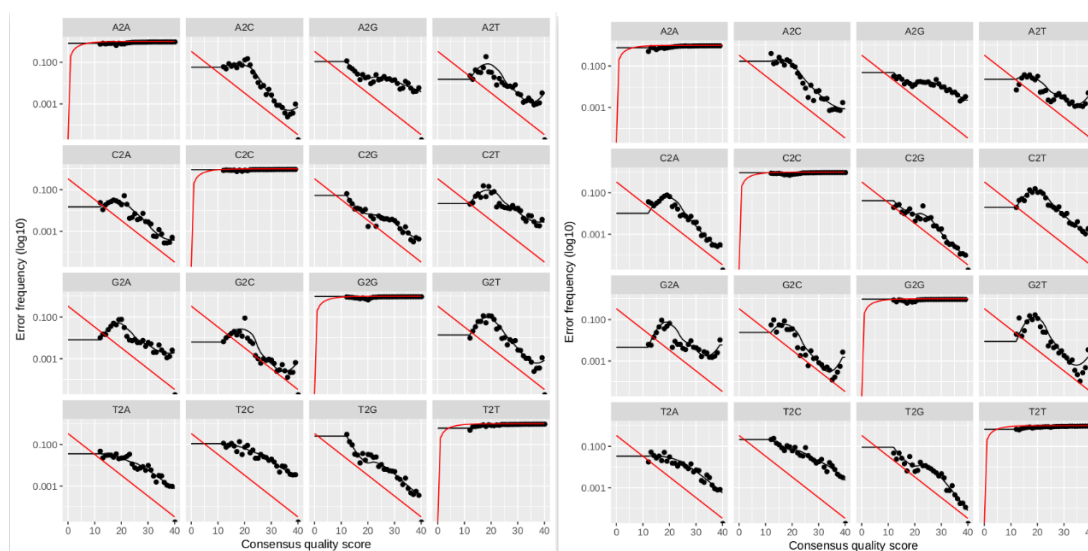


图 3-1 错误率统计图

3.2 生成 ASVs

使用 DADA2 根据错误率消除误差，计算统计出 610702 条物种 16S 序列，去重后得到 972 条。再进行去除嵌合体，得到 589107 条物种 16S 序列，通过率为 96.46%，去重后得到 652 条。将结果与 silva 数据库 v132 (www.arb-silva.de) 进行比对，生成 16S 序列统计表以及 16S 序列注释表。以下列出部分结果。

表 3-1 序列统计表

SRA ID	ASV_1	ASV_2	ASV_3	ASV_4
SRR7639091	29110	0	9523	304
SRR7639092	22455	0	2877	84
SRR7639093	26128	24	49625	0
SRR7639094	0	73979	469	52472

注：

- (1) SRA ID：数据在 SRA 数据库中的 ID；
- (2) ASV_X：ASV 序列，此处只是展示表格形式，并不代表只有 4 个 ASV。

表 3-2 序列注释表

ASV	Kingdom	Phylum	Class	Order	Family	Genus	Species
ASV_1	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella_9	copri
ASV_2	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	NA
ASV_3	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae	Escherichia/Shigella	NA
ASV_5	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae	Klebsiella	NA
ASV_6	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella_9	NA
ASV_7	Bacteria	Firmicutes	Negativicutes	Selenomonadales	Veillonellaceae	Megamonas	funiformis

注：

- (1) 第一列为 ASV 编号，只用于展示形式，不代表正式编号，与表 3-1 中 ASV 无关联；
- (2) 列名，由第二列开始为界、门、纲、目、科、属、种。

四、 其他分析

4.1 物种组成

使用 phyloseq (McMurdie PJ and Holmes S) 对“门”分类进行统计。横坐标为样本 SRA ID，纵坐标为 counts 数。图例为门分类下的物种种类。

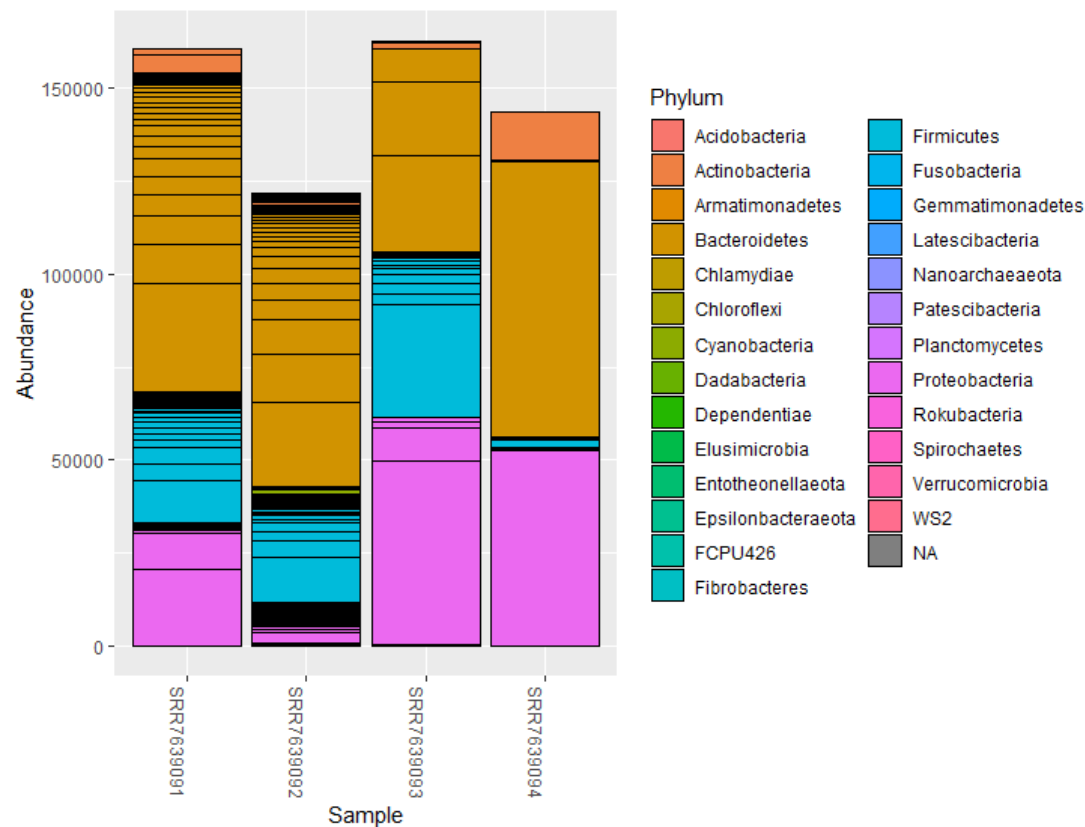


图 4-1 物种组成 (门) 图

4.2 Alpha 分析

Alpha 多样性用于度量群落生态单样本的物种多样性，是反映丰富度和均匀度的综合指标。菌群丰富度（Community richness）指数：Chao1 指数越大，说明菌群丰富度越高。菌群多样性（Community diversity）指数：Shannon 值越大，说明群落多样性越高。

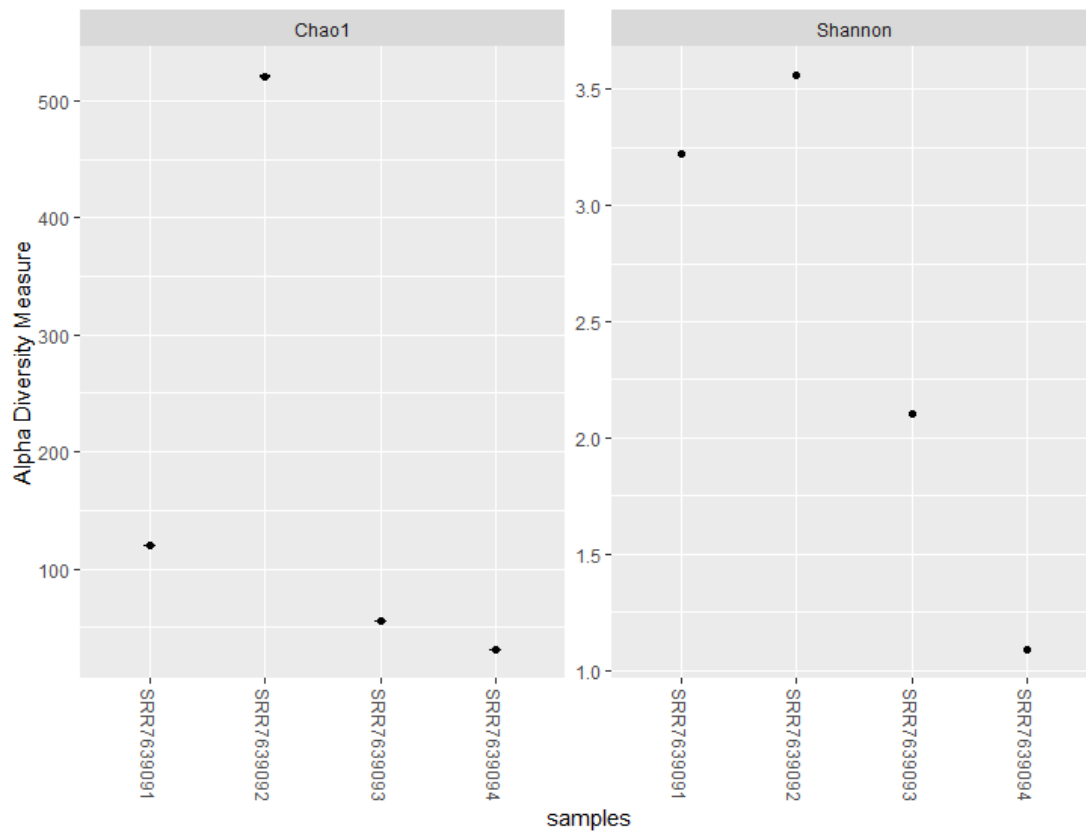


图 4-2 alpha 分析图

4.3 Beta 分析

Beta 多样性用于不同生态系统之间多样性的比较，利用各样本序列间的进化关系及丰度信息来计算样本间距离，反映样本（组）间是否具有显著的微生物群落差异。目前应用比较多的是 PCA、PCoA、NMDS 分析等。由于样本只有一个组别，因此不做 Beta 分析。

五、 参考文献

1. Deficiency of Adenosine Deaminase 2 (DADA2): Updates on the Phenotype, Genetics, Pathogenesis, and Treatment. Meyts I and Aksentijevich I J Clin Immunol. 2018 Jul;38(5):569-578. doi: 10.1007/s10875-018-0525-8. Epub 2018 Jun 27.
2. DADA2: High-resolution sample inference from Illumina amplicon data. Callahan BJ , et al. Nat Methods. 2016 Jul;13(7):581-3. doi: 10.1038/nmeth.3869. Epub 2016 May 23.
3. Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. Callahan BJ , et al. Version 2. F1000Res. 2016 Jun 24
4. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. Callahan BJ , et al. ISME J. 2017 Dec;11(12):2639-2643. doi: 10.1038/ismej.2017.119. Epub 2017 Jul 21. Evaluation Studies
5. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Quast C , et al. Nucleic Acids Res. 2013 Jan;41(Database issue):D590-6. doi: 10.1093/nar/gks1219. Epub 2012 Nov 28.
6. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Yilmaz P , et al. Nucleic Acids Res. 2014 Jan;42(Database issue):D643-8. doi: 10.1093/nar/gkt1209. Epub 2013 Nov 28.
7. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. McMurdie PJ and Holmes S PLoS One. 2013 Apr 22;8(4):e61217. doi: 10.1371/journal.pone.0061217. Print 2013.
8. Counting the uncountable: statistical approaches to estimating microbial diversity. Hughes JB , et al. Appl Environ Microbiol. 2001 Oct;67(10):4399-406.