

**上海中科新生命**

**秉承中科，专“新”致“质”（始于2004年）**

**16S rDNA扩增子测序分析报告**

项目名称：

委托单位：

项目编号：

报告时间：

**目 录**

[1 项目概述 Project Introduction 5](#_Toc40259924)

[2 工作流程 Flowchart 5](#_Toc40259925)

[**2.1 实验流程** 5](#_Toc40259926)

[**2.2 数据分析流程** 6](#_Toc40259927)

[3 项目结果 Result 7](#_Toc40259928)

[**3.1 测序数据预处理及质控统计** 7](#_Toc40259929)

[**3.2 OTU分析和物种注释** 8](#_Toc40259930)

[**3.2.1 OTU聚类和物种注释** 8](#_Toc40259931)

[**3.2.2 基于OTU分布的Venn图/花瓣图** 9](#_Toc40259932)

[**3.3 物种组成分析** 9](#_Toc40259933)

[**3.3.1 单样本多级物种组成图** 9](#_Toc40259934)

[**3.3.2 特定物种分类树** 10](#_Toc40259935)

[**3.3.3 群落结构组分图** 11](#_Toc40259936)

[**3.3.4 UPGMA聚类树** 11](#_Toc40259937)

[**3.3.5 群落Heatmap图** 12](#_Toc40259938)

[**3.4 α多样性分析** 13](#_Toc40259939)

[**3.4.1 α多样性指数** 13](#_Toc40259940)

[**3.4.2 Rarefaction Curve** 14](#_Toc40259941)

[**3.4.3 Shannon** 14](#_Toc40259942)

[**3.4.4 Rank-Abundance** 15](#_Toc40259943)

[**3.4.5 Specaccum** 15](#_Toc40259944)

[**3.4.6 α多样性指数组间差异分析** 16](#_Toc40259945)

[**3.5 β多样性分析** 16](#_Toc40259946)

[**3.5.1 β多样性指数组间差异分析** 16](#_Toc40259947)

[**3.5.2 PCA分析** 17](#_Toc40259948)

[**3.5.3 PCoA分析** 18](#_Toc40259949)

[**3.5.4 NMDS分析** 18](#_Toc40259950)

[**3.6组间群落结构差异显著性检验** 19](#_Toc40259951)

[**3.6.1 Anosim** 19](#_Toc40259952)

[**3.6.2 Adonis** 20](#_Toc40259953)

[**3.7 组间差异物种分析** 20](#_Toc40259954)

[**3.7.1 LEfSe分析** 20](#_Toc40259955)

[**3.7.2 STAMP差异分析** 21](#_Toc40259956)

[**3.8 功能预测** 22](#_Toc40259957)

[**3.8.1 KEGG功能预测** 22](#_Toc40259958)

[**3.8.2 COG功能预测** 22](#_Toc40259959)

[**3.8.3 PCA分析** 23](#_Toc40259960)

[**3.8.4 功能分布组分图** 23](#_Toc40259961)

[**3.8.5 功能分布Heatmap图** 24](#_Toc40259962)

[**3.8.6 差异功能LEfSe分析** 24](#_Toc40259963)

[**3.8.7 差异功能STAMP分析** 26](#_Toc40259964)

[4 实验方法描述 Methods 26](#_Toc40259965)

[**4.1** **中文版实验方法** 26](#_Toc40259966)

[**4.2** **英文版实验方法（供参考）** 27](#_Toc40259967)

[**4.2.1 Sequencing** 27](#_Toc40259968)

[**4.2.2 Data analysis** 28](#_Toc40259969)

[5 16S报告阅读顺序建议 Reading Order Suggestion 29](#_Toc40259970)

[6 参考文献 References 30](#_Toc40259971)

[7 生信数据分析云平台 Data analysis platform 32](#_Toc40259972)

[8 组学期刊投稿指南（供参考）Publication Recommendation 33](#_Toc40259973)

[**8.1 第一档期刊（影响因子9 分以上）** 33](#_Toc40259974)

[**8.2 第二档期刊（影响因子4~9分）** 34](#_Toc40259975)

[**8.3 第三档期刊（影响因子4分以下）** 34](#_Toc40259976)

[9 部分合作发表文章 Cooperative Projects Papers 35](#_Toc40259977)

[**9.1 医学** 35](#_Toc40259978)

[**9.2 植物** 36](#_Toc40259979)

[**9.3 动医/动科/食品等其他** 36](#_Toc40259980)

1. **项目概述 Project Introduction**

16S rDNA指细菌基因组中编码核糖体16S rRNA分子对应的DNA序列,也就是16S rRNA的编码基因。16S序列包括10个保守区域（Conserved Regions）和9个高变区域（Hypervariable Regions），其中保守区在细菌间差异不大，高变区具有属或种的特异性，随亲缘关系不同而有一定的差异。因此，16S rDNA可以做作为揭示生物物种的特征核酸序列， 被认为是最适于细菌系统发育和分类鉴定的指标[1,2,3]。

16S rDNA扩增子测序（16S rDNA Amplicon Sequencing），通常是选择某个或某几个变异区域，利用保守区设计通用引物进行PCR扩增，然后对高变区进行测序分析和菌种鉴定，16S rDNA扩增子测序技术已成为研究环境样品中微生物群落组成结构的重要手段。根据扩增的16S区域特点，基于测序平台利用双末端测序（Paired-End）的方法，构建小片段文库并进行双末端测序。通过对Reads拼接过滤、OTUs（Operational Taxonomic Units）聚类、物种注释及丰度分析可以揭示样品物种构成，进一步也可以挖掘样品之间的差异，筛选差异菌群。

本项目采用Illumina Miseq测序平台对V3-V4可变区进行扩增和测序。一共total\_sample例样本，分为group\_num组样本，每组sample\_num份生物学重复样本，样品信息见下表。

**样品信息**

|  |  |  |
| --- | --- | --- |
| 样品分组 | 数量 | 样品状态 |
|  |  |  |

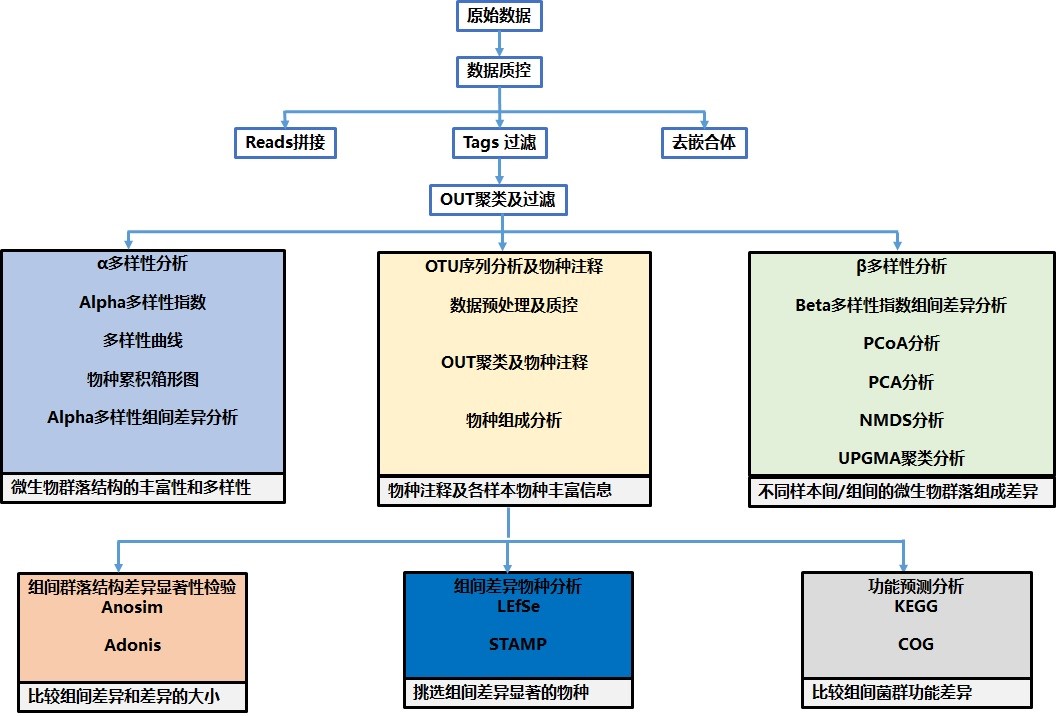
1. **工作流程 Flowchart**

**2.1 实验流程**

从DNA提取到上机测序，每一个环节都会对数据质量有影响，而数据质量又会直接影响后续信息分析的结果。因此在每一环节都要对样本质量进行严格控制，才能确保测序数据的真实可信。

**2.2 数据分析流程**

测序得到的原始数据（Raw Data），存在一定比例的干扰数据（Dirty Data），为了使信息分析的结果更加准确、可靠，首先对原始数据进行拼接、过滤及去嵌合体，得到有效数据（Clean Data）。然后基于有效数据进行OTUs（Operational Taxonomic Units）聚类和物种分类分析。基于OTU聚类分析结果，可以对OTU进行多种多样性指数分析以及对测序深度的检测。基于分类学信息，可以在各个分类水平上进行群落结构的统计分析。获得下机数据后的信息分析流程如图所示。



**数据分析流程图**

注：样品数小于3个，不能进行Beta Diversity分析、组间群落结构差异显著性检验和组间差异物种分析；若无分组信息，则不能进行组间群落结构差异显著性检验和组间差异物种分析。

1. **项目结果 Result**

**3.1 测序数据预处理及质控统计**

根据barcode序列区分各个样本的数据，提取出的数据以fastq格式保存，PE数据每个样本有fq1和fq2两个文件，分别为测序两端的reads，序列按顺序一一对应。

Fastq：Fastq是Solexa测序技术中一种反映测序序列的碱基质量的文件格式。每条read包含4行信息，其中第一行和第三行由文件识别标志和读段名（ID）组成（第一行以“@”开头，第三行以“+” 开头；第三行中 ID可以省略，但“+”不能省略），第二行为碱基序列，第四行是第二行的序列中每个碱基所对应的测序质量值。

如下所示：

@HWI-ST531R:144:D11RDACXX:4:1101:1212:1946 1:N:0:ATTCCT

ATNATGACTCAAGCGCTTCCTCAGTTTAATGAAGCTAACTTCAATGCTGAGATCGTTGACGACATCGAATGGG

+ HWI-ST531R:144:D11RDACXX:4:1101:1212:1946 1:N:0:ATTCCT

?A#AFFDFFHGFFHJJGIJJJIICHIIIIJJGGHIIJJIIJIIJIHGI@FEHIIJBFFHGJJIIHHHDFFFFDCCCCEDDCDDCDEACC

将测序得到的下机数据（Raw PE）进行拼接、质控及去嵌合体，得到Clean PE。数据处理过程中各步骤得到的统计结果见下表。

**质控数据统计分析结果**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sample ID** | **Raw PE** | **Clean PE** | **Base(nt)** | **AvgLen(nt)** | **Q20(%)** | **Q30(%)** | **GC(%)** | **Effective(%)** |
|  |  |  |  |  |  |  |  |  |

注：由于数据表较大，此处只列出部分数据，详见对应文件[QcStatic](file:///E:\Desktop\test16\Result\01_QC\).csv。

表中Raw PE表示原始下机的PE reads； Clean PE是Raw PE过滤低质量和短长度及去嵌合体后的序列；Base是Clean PE的碱基数目；AvgLen为Clean PE的平均长度；Q20和Q30是Clean PE中碱基质量值大于20（测序错误率小于1%）和30（测序错误率小于0.1%）的碱基所占的百分比；GC (%) 表示Clean PE中GC碱基的含量；Effective (%) 表示Clean PE的数目与Raw PE数目的百分比。

结果目录：

Result\01\_QC\QcStatic.csv

**3.2 OTU分析和物种注释**

**3.2.1 OTU聚类和物种注释**

为了研究样品的物种组成多样性，对所有样品的Clean reads进行聚类，以97%的一致性将序列聚类成为OTUs（Operational Taxonomic Units），然后对OTUs的代表序列进行物种注释。注释用的数据库为greengenes，注释结果示例见下表。

**OTU and Taxonomy**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **OTU ID** | **A** | **B** | **C** | **taxonomy** |
|  |  |  |  |  |

注：由于数据表较大，此处以图例形式列出部分数据，详见对应文件otu\_table\_tax.csv

OTU ID 为OTU编号；第二列至taxonomy列的前一列为各样本的序列在各个OTU中的分布情况。taxonomy列拉开可查看分类学系谱信息。各级分类水平以“；”隔开，分类学名称前的单个字母为分类等级的首字母缩写，以“\_\_”隔开。

结果目录：

Result\02\_OTU\_Taxa\otu\_table\_tax.csv

**3.2.2 基于OTU分布的Venn图/花瓣图**

根据OTUs聚类分析结果和研究需求，分析不同样品（组）之间共有、特有的OTUs，当样本（组）数小于5时，绘制成韦恩图（Venn Graph），当样本（组）数大于5时，绘制成花瓣图。

[venn.png]

**OTU Venn**

注：不同的颜色代表不同的样本（组），如果两个不同颜色圆圈重叠的区域标注有数字100，说明这两个样本（组）均有序列被划分入相同的OTU中，且这样的OTU有100个。

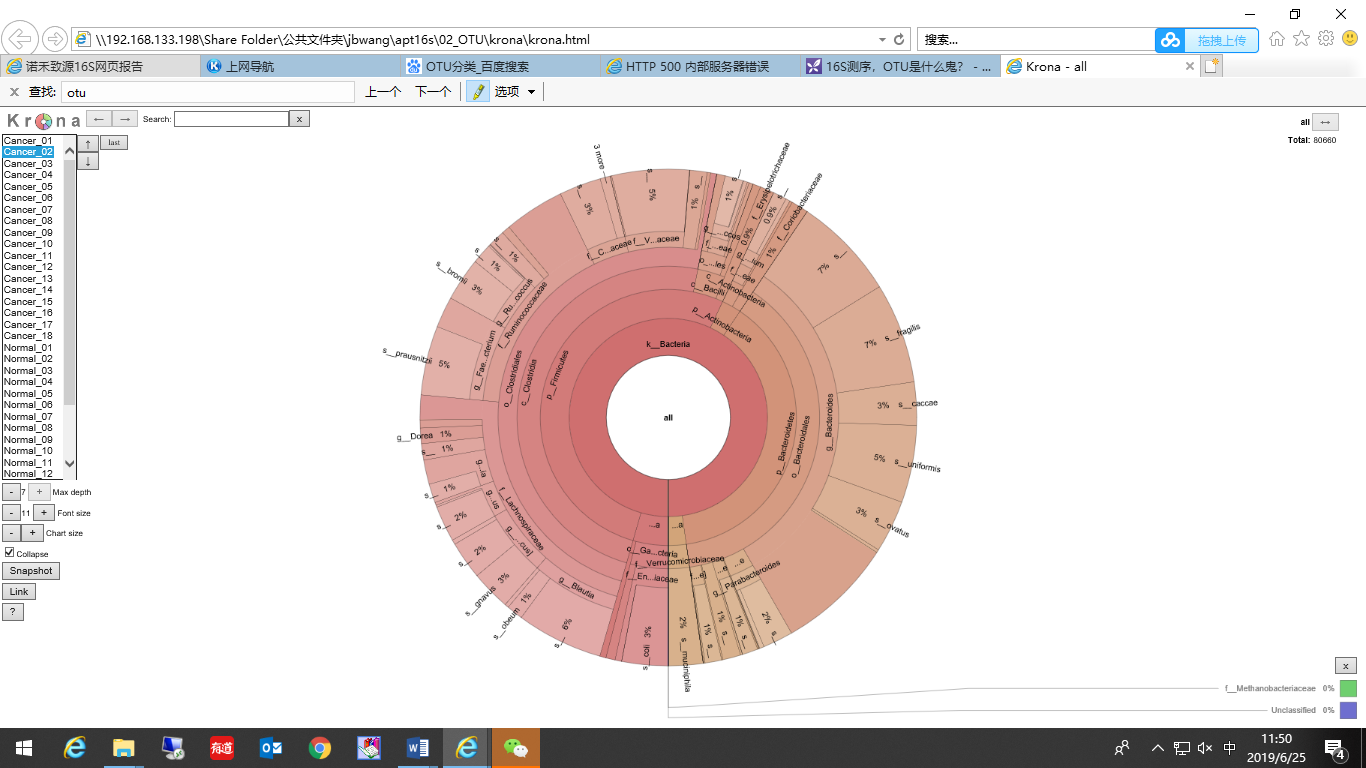
结果目录：

Result\02\_OTU\_Taxa \Venn

**3.3 物种组成分析**

**3.3.1 单样本多级物种组成图**

为了更直观展示各样品在不同分类水平的丰度情况，我们利用Krona软件[4]对单样品物种注释结果及丰度进行动态可视化展示。一个样品，做一张图，示例如图所示。



**物种注释结果KRONA展示**

结果目录：

Result\03\_Community\krona

**3.3.2 特定物种分类树**

对每个样品或分组的物种注释结果，筛选特别关注的物种（默认选择最大相对丰度前10的属）进行物种分类树统计[5]，其中单个样品的物种分类树及分组的物种分类树见下图。

[sample\_tree.png]

**单样品特定物种分类树**

[group\_tree.png]

**分组特定物种分类树**

注：图 4-1中不同颜色的圆圈表示分类水平，对应左侧图例；大小代表相丰度大小。图 4-2中圆圈不同颜色的扇形表示分组，对应左侧图例；大小表示在分类上相对丰度的比例大小。圆圈下的数字，第一个表示只比对到该分类（不能比对到该分类等级以下的分类水平）的序列数，第二个数字表示共有多少序列比对到该分类。

结果目录：

各样品的物种分类树见：Result\03\_Community\taxa\_tree\sample\_tree

分组的物种分类树见：Result\03\_Community\taxa\_tree\group\_tree

**3.3.3 群落结构组分图**

群落结构组分图可以展示每个样本或分组在不同分类水平上的群落结构。根据物种注释结果，选取每个样品或分组在各分类水平（Phylum、Class、Order、Family、Genus、Species）上最大丰度排名前10的物种，生成物种相对丰度柱形累加图，以便直观查看在不同分类水平上，相对丰度较高的物种及其比例。以门水平为例，结果如下图所示。

[sample\_barplot.png]

**各样品门水平物种相对丰度柱状图**

[group\_barplot.png]

**各分组门水平物种相对丰度柱状图**

注：不同颜色代谢不同的物种，对应右侧图例；横轴代表不同的样本或分组，纵轴代表不同物种的相对丰度。

结果目录：

各样本的群落结构组成图见：Result\03\_Community\community\Top10\_sample;

各分组的群落结构组成图见：Result\03\_Community\community\Top10\_group, 包括门纲目科属种（Phylum、Class、Order、Family、Genus、Species）6个分类级别的结果

**3.3.4 UPGMA聚类树**

为了研究不同样品间的相似性，可以通过对样品进行聚类分析，构建样品的聚类树。在环境生物学中，UPGMA（Unweighted Pair-group Method with Arithmetic Mean）是一种较为常用的聚类分析方法，是最早用来解决分类问题的分析方法。

UPGMA 的基本思想是：首先将距离最小的2个样品聚在一起，并形成一个新的节点（新的样品），其分支点位于2个样品间距离的1/2 处；然后计算新的“样品”与其它样品间的平均距离，再找出其中的最小2个样品进行聚类；如此反复，直到所有的样品都聚到一起，最终得到一个完整的聚类树。

基于Weighted Unifrac距离矩阵和Unweighted Unifrac距离矩阵做UPGMA聚类分析，并将聚类结果与各样品在门水平上的群落结构柱状图整合展示，结果如下图所示。

[weight\_upgma.png]

**基于Weighted Unifrac距离的UPGMA聚类树及群落结构柱状图**

[unweight\_upgma.png]

**基于Unweighted Unifrac距离的UPGMA聚类树及群落结构柱状图**

注：左边是样本间基于群落组成的层次聚类分析，右边是各样本在门水平的群落结构柱状图。Unweighted Unifrac距离是基于微生物序列间的进化信息计算得到的[6,7]，而Weighted Unifrac距离是

基于微生物序列间的进化信息和OTUs的丰度信息计算得到的[8]。

结果目录：

基于Weighted Unifrac距离矩阵的UPGMA聚类分析见：Result\03\_Community\UPGMA\weighted\_unifrac

基于Unweighted Unifrac距离矩阵的UPGMA聚类分析见：Result\03\_Community\UPGMA\unweighted\_unifrac

**3.3.5 群落Heatmap图**

Heatmap可以用颜色变化来反映二维矩阵或表格中的数据信息，它可以直观地将数据值的大小以定义的颜色深浅表示出来。根据物种注释结果，在各分类水平（Phylum、 Class、 Order、 Family、 Genus）上进行物种层面的聚类，将聚类后数据表示在heatmap图上，便于直观发现哪些物种在哪些样品或分组中聚集较多或聚集较少。以门水平为例，结果如下图所示。

[sample\_heatmap.png]

**各样品门水平物种丰度聚类热图**

[group\_heatmap.png]

**各分组门水平物种丰度聚类热图**

注：横轴表示不同的样本，纵轴表示不同的物种，颜色的深浅与物种的丰度有关，颜色越深，丰度越高。

结果目录：

各样本的物种丰度聚类热图见：Result\03\_Community\taxa\_heatmap\cluster\_sample;各分组的物种丰度聚类热图见：Result\03\_Community\\taxa\_heatmap \cluster\_group, 包括门纲目科属种（Phylum、Class、Order、Family、Genus）5个分类级别的结果。

**3.4 α多样性分析**

**3.4.1 α多样性指数**

Alpha Diversity 用于分析样品内（Within-community）的微生物群落多样性。各样品的α多样性指数见下表。

**α 多样性指数**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sample ID | 0.97 | | | | | | |
| **Shannon** | **Simpson** | **ACE** | **Coverage** | **Chao1** | **Observed\_species** | **PD\_whole\_tree** |
|  |  |  |  |  |  |  |  |

注：由于数据表较大，此处只列出部分数据，详见对应文件alpha\_diversity\_index.csv。

Observed\_species为OTU数目；Shannon和Simpson为菌群多样性指数；Chao 1 和ACE为菌群丰度指数；Coverage为测序深度指数；PD\_whole\_tree为系统发育多样性指数。

结果目录：

Result\04\_Alpha\_diversity\ alpha\_diversity\_index.csv

**3.4.2 Rarefaction Curve**

Rarefaction Curve是从样品中随机抽取一定测序量的数据，统计它们所代表物种数目，以抽取的测序数据量与对应的物种数来构建曲线，结果如下图所示。稀释曲线可直接反映测序数据量的合理性，间接反映样品中物种的丰富程度。当曲线趋向平坦时，说明测序数据量合理，更多的数据量只会产生少量新的OTU，反之则表明继续测序还可能产生较多新的OTU。

[rarefaction.png]

**Rarefaction Curve**

注：横坐标：随机抽取的测序数据量；纵坐标：观测到的OTU数量。

结果目录：

Result\04\_Alpha\_diversity\Rarefaction\_Curve

**3.4.3 Shannon**

Shannon曲线是根据各样本的测序量在不同测序深度时的微生物多样性指数来构建曲线，结果如下图所示。当曲线趋向平坦时，说明测序数据量足够大，可以反映样本中绝大多数的微生物信息。

[shannon.png]

**Shannon Curve**

注：横坐标：随机抽取的测序数据量；纵坐标：Shannon指数

结果目录：

Result\04\_Alpha\_diversity\Shannon

**3.4.4 Rank-Abundance**

Rank-abundance曲线可用来解释多样性的两个方面，即物种丰度和物种均匀度，结果如下图所示。在水平方向，物种的丰度由曲线的宽度来反映，物种的丰度越高，曲线在横轴上的范围越大；曲线的形状（平滑程度）反映了样本中物种的均度，曲线越平缓，物种分布越均匀[9]。

[rank.png]

**Rank-Abundance**

注：横坐标：OTU等级，横坐标轴上数字，例如“200”代表样本中按照丰度排列第200位的OUT，“400”代表样本中按照丰度排列第400位的OUT，依次类推；纵坐标：该等级OTU中序列数的相对百分含量，即属于该OTU的序列数除以总序列数，纵坐标轴上数字，例如“100”代表相对丰度为100%，“10”代表相对丰度为10%，依次类推。

结果目录：

Result\04\_Alpha\_diversity\Rank\_Abundance

**3.4.5 Specaccum**

物种累积曲线( species accumulation curves)是描述随着样本量的增加物种多样性增加的分析，是调查样本的物种组成和预测样本中物种丰度的有效工具，在生物多样性和群落调查中，被广泛用于样本量是否充分的判断以及物种丰富度（species richness）的估计。因此，通过物种累积箱形图不仅可以判断样本量是否充分，在样本量充分的前提下，运用物种累积箱形图还可以对物种丰富度进行预测（默认在样本量大于10个时分析）。结果如下图所示。

[specaccum.png]

**Species Accumulation Curves**

注：横坐标为样本量；纵坐标为抽样后 OTU 数目。结果反映了持续抽样下新 OTU（新物种）出现的速率。在一定范围内，随着样本量的加大，若箱形图位置表现为急剧上升则表示群落中有大量物种被发现；当箱形图位置趋于平缓，则表示此环境中的物种并不会随样本量的增加而显著增多。物种累积箱形图可以作为对样本量是否充分的判断,箱形图位置急剧上升表明样本量不足，需要增加抽样量；反之，则表明抽样充分，可以进行数据分析。

结果目录：

Result\04\_Alpha\_diversity\Specaccum

**3.4.6** **α多样性指数组间差异分析**

α多样性指数组间差异分析中，箱形图可以直观的反应组内物种多样性的中位数、离散程度、最大值、最小值、异常值。同时，通过T-test检验，wilcox秩和检验和Tukey检验（当数据呈正态分布时，只有2个分组时选用T-test，分组大于2时进行Tukey检验，当数据不呈正态分布时选用wilcox）分析组间物种多样性差异是否显著。以Observed\_Species指数为例，其组间差异分析的箱形图如下图所示：

[alpha\_diff.png]

**Observed\_Species指数组间差异箱形图**

注：横坐标为不同组别；纵坐标为α多样性指数

结果目录：

Result\04\_Alpha\_diversity\Alpha\_div\_diff，包括7种Alpha多样性指数（Observed\_Species、Shannon、Simpson、Chao 1 、ACE、Coverage、 PD\_whole\_tree）组间差异的箱形图

**3.5** **β多样性分析**

**3.5.1 β多样性指数组间差异分析**

β多样性指数组间差异分析中，箱形图可以直观的反应组内物种多样性的中位数、离散程度、最大值、最小值、异常值。同时，通过T-test检验，wilcox秩和检验和Tukey检验（当数据呈正态分布时，只有2个分组时选用T-test，分组大于2时进行Tukey检验，当数据不呈正态分布时选用wilcox）分析组间物种多样性差异是否显著。基于Weighted Unifrac 和Unweighted UnifracBeta距离的β多样性组间差异分析的结果如下图所示。

[weight\_beta\_diff.png]

**基于Weighted Unifrac距离的β多样性组间差异分析**

[unweight\_beta\_diff.png]

**基于Unweighted Unifrac距离的β多样性组间差异分析**

注：图中第一[四分位数](http://zhidao.baidu.com/search?word=%E5%9B%9B%E5%88%86%E4%BD%8D%E6%95%B0&fr=qb_search_exp&ie=utf8) (Q1)，又称“下[四分位](http://zhidao.baidu.com/search?word=%E5%9B%9B%E5%88%86%E4%BD%8D&fr=qb_search_exp&ie=utf8" \t "_blank)数”，等于该样本中所有数值由小到大排列后第25%的数字。第二[四分位](http://zhidao.baidu.com/search?word=%E5%9B%9B%E5%88%86%E4%BD%8D&fr=qb_search_exp&ie=utf8" \t "_blank)数 (Q2)，又称“[中位数](http://zhidao.baidu.com/search?word=%E4%B8%AD%E4%BD%8D%E6%95%B0&fr=qb_search_exp&ie=utf8)”，等于该样本中所有数值由小到大排列后第50%的数字。 第三[四分位](http://zhidao.baidu.com/search?word=%E5%9B%9B%E5%88%86%E4%BD%8D&fr=qb_search_exp&ie=utf8)数 (Q3)，又称“上四[分位数](http://zhidao.baidu.com/search?word=%E5%88%86%E4%BD%8D%E6%95%B0&fr=qb_search_exp&ie=utf8" \t "_blank)”，等于该样本中所有数值由小到大排列后第75%的数字。

结果目录：

基于Weighted Unifrac距离的β多样性组间差异分析见: Result\05\_Beta\_diversity\Beta\_div\_diff\weighted\_unifrac

基于Unweighted Unifrac距离的β多样性组间差异分析见: Result\05\_Beta\_diversity\Beta\_div\_diff\unweighted\_unifrac

**3.5.2 PCA分析**

主成分分析（PCA，Principal Component Analysis），是一种对数据进行简化分析的技术，对多维数据进行降维，从而提取出数据中最主要的元素和结构的方法[10]。应用PCA分析，能够提取出最大程度反映样品间差异的两个坐标轴，从而将多维数据的差异反映在二维坐标图上，进而揭示复杂数据背景下的简单规律。样品的群落组成越相似，则它们在PCA图中的距离越接近。PCA分析结果如下图所示。

[pca.png]

**PCA**

注**：**横坐标表示第一主成分，纵坐标表示第二主成分，百分比表示对样品差异的贡献值。

结果目录：

Result\05\_Beta\_diversity\PCA

**3.5.3 PCoA分析**

主坐标分析（PCoA，Principal Co-ordinates Analysis），与PCA分析类似，用来研究样本群落组成的相似性或差异性。如果样品距离越接近，表示物种组成结构越相似，因此群落结构相似度高的样品倾向于聚集在一起，群落差异很大的样品则会远远分开。与PCA分析的不同点在于：PCA是基于OUT丰度表的欧式距离进行分析，PCoA则是基于Weighted Unifrac 和Unweighted UnifracBeta距离进行分析。PCoA 分析结果如下图所示。

[weight\_pcoa.png]

**基于Weighted Unifrac 距离的PCoA分析**

[unweight\_pcoa.png]

**基于Unweighted Unifrac距离的PCoA分析**

注**：**横坐标表示第一主成分，纵坐标表示第二主成分，百分比表示对样品差异的贡献值。

结果目录：

Result\05\_Beta\_diversity\PCoA

**3.5.4 NMDS分析**

非度量多维尺度法（NMDS，Non-Metric Multi-Dimensional Scaling）统计是一种适用于生态学研究的排序方法。NMDS是非线性模型，其设计目的是为了克服线性模型（包括PCA、 PCoA）的缺点，更好地反映生态学数据的非线性结构[11]。

应用NMDS分析，根据样本中包含的物种信息，以点的形式反映在多维空间上，而对不同样本间的差异程度，则是通过点与点间的距离体现，能够反映样本的组间和组内差异等。NMDS分析结果如下图所示。

[nmds.png]

**NMDS**

注：每个点表示一个样品，点与点之间的距离表示差异程度，同一个组的样品使用同一种颜色表示。

结果目录：

Result\05\_Beta\_diversity\NMDS

**3.6组间群落结构差异显著性检验**

**3.6.1 Anosim**

Anosim分析是一种非参数检验，基于Bray-Curtis算法，检验组间的差异是否显著大于组内差异，从而判断分组是否有意义，分析结果如下表所示。

**表4-4 Anosim 分析**

|  |  |  |
| --- | --- | --- |
| Group | R-value | p-value |
|  |  |  |

注：R值介于（-1，1）之间，R值大于0，说明组间差异显著。R值小于0，说明组内差异大于组间差异，统计分析的可信度用 P-value 表示，P< 0.05 表示统计具有显著性。

对Anosim的分析结果，基于两两样本之间的距离值排序获得的秩（组间的为between，组内的为within），这样任一两两组的比较可以获得三个分类的数据，并进行箱线图的展示，如下图所示。

[anosim.png]

**Anosim组间差异**

注：横坐标：Between为两组之间的结果，其他两个为各自组内的结果。纵坐标为样品间距离的秩。

结果目录：

Result\05\_Beta\_diversity\anosim

**3.6.2 Adonis**

Adonis又称置换多因素方差分析（permutational MANOVA）或非参数多因素方差分析（nonparametric MANOVA）。该方法可分析不同分组因素对样品差异的解释度，并使用置换检验对分组的统计学意义进行显著性分析[12,13,14,15]，分析结果如下表所示。

**Adonis analysis**

|  |  |  |
| --- | --- | --- |
| Group | R2-value | p-value |
|  |  |  |

注： R2 表示不同分组对样品差异的解释度，即分组方差与总方差的比值，R2 越大表示分组对差异的解释度越高；P 值小于0.05 说明本次检验的可信度高。

结果目录：

Result\05\_Beta\_diversity\adonis

**3.7 组间差异物种分析**

**3.7.1 LEfSe分析**

LEfSe(LDA Effect Size)是一种用于发现高维生物标识和揭示基因组特征的软件。包括基因，代谢和分类，用于区别两个或两个以上生物条件（或者是类群）。该算法强调的是统计意义和生物相关性，让研究人员能够识别不同丰度的特征以及相关联的类别。

LEfSe分析根据分类学组成对样本按照不同的分组条件进行线型判别分析（LDA），找出对样本划分产生显著性差异影响的群落或物种[16]。LEfSe的分析结果如下图所示，包括进化分支图（系统发育分布），LDA值分布柱状图和组间具有统计学差异的Biomarker在不同组中丰度比较图。

**[cladogram.png]**

**Cladogram（进化分支图）**

注：进化分支图中由内至外辐射的圆圈代表了由门至属（或种）的分类级别。在不同分类级别上的每一个小圆圈代表该水平下的一个分类，小圆圈直径大小与相对丰度大小呈正比。红色区域和绿色区域表示不同分组，树枝中红色节点表示在红色组别中起到重要作用的微生物类群，绿色节点表示在绿色组别中起到重要作用的微生物类群，黄色节点表示的是在两组中均没有起到重要作用的微生物类群。图中英文字母表示的物种名称在右侧图例中进行展示。

**[Lda\_score.png]**

**LDA Score**

注：LDA值分布柱状图中红色区域和绿色区域表示不同分组，树枝中红色节点表示在红色组别中起到重要作用的微生物类群，绿色节点表示在绿色组别中起到重要作用的微生物类群。图中仅展示了LDA Score大于设定值（默认设置为2）的物种，柱状图的长度代表LDA值的大小。图中英文字母表示的物种名称在右侧图例中进行展示。

默认LDA值大于2的物种为组间具有统计学差异的Biomarker。各Biomarker在不同组中丰度比较图详见附件，示例如下图所示。

[biomarker.png]

**Biomarker在各组样品中的相对丰度**

注：图中左右两侧表示不同的组别。横坐标表示样本，纵坐标表示菌的相对丰度。

结果目录：

Result\06\_Differential\_analysis\LefSe

**3.7.2 STAMP差异分析**

STAMP差异分析用于比较两组样品（Welch’s t-test 分析）或多组样品（ANOVA 分析，该软件只出Pvalue，无图片展示）之间物种的丰度，通过此分析可获得显著性差异物种。在属水平上的STAMP差异分析结果如下图所示。

**[stamp.png]**

**STAMP Result**

注：左图所示为不同物种在两个样品或者两组样品中的丰度比例，中间所示为95%置信度区间内的差异比例，最右边的值为p值，p值＜0.05，表示差异显著。

结果目录：

Result\06\_Differential\_analysis\STAMP

**3.8 功能预测**

使用PICRUSt软件通过比对16S测序数据获得的物种组成信息，推测样本中的功能基因组成，从而分析不同样本或分组之间在功能上的差异

**3.8.1 KEGG功能预测**

KEGG功能预测分析是研究群落样本为适应环境变化发生的代谢功能改变的有效手段。KEGG功能预测结果见下表。

**KEGG功能预测结果**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Orthology | A1 | A2 | A3 | A4 |
|  |  |  |  |  |

注：由于数据表较大，此处以图例形式列出部分数据，详见对应文件ko\_prediction.csv

结果目录：

Result\07\_FunctionPrediction\KEGG\ko\_prediction.csv

**3.8.2 COG功能预测**

COG即原核生物同源蛋白簇数据库，是原核生物常用的蛋白功能分类数据库，与KEGG互补，更为全面揭示菌群的功能组成。COG功能预测结果见下表。

**COG功能预测结果**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Orthology | A1 | A2 | A3 | A4 |
|  |  |  |  |  |

注：由于数据表较大，此处以图例形式列出部分数据，详见对应文件cog\_prediction.csv

结果目录：

Result\07\_FunctionPrediction\COG\cog\_prediction.csv

**3.8.3 PCA分析**

PCA分析是在整体层面上揭示不同组别的菌群功能的相似度或差异性。在PCA图中的距离越接近，反映样品的群落功能越相似。对KEGG和COG功能预测结果进行PCA分析，结果如下图。

**[pca12.png]**

**KEGG PCA COG PCA**

注**：**横坐标表示第一主成分，纵坐标表示第二主成分，百分比表示对样品差异的贡献值。

结果目录：

Result\07\_FunctionPrediction\KEGG\PCA

Result\07\_FunctionPrediction\COG\PCA

**3.8.4 功能分布组分图**

群落结构组分图可以展示每个样本或分组的功能组成及相对丰度。分别选择TOP 10的KEGG和COG功能条目，生成相对丰度柱形累加图，结果如下图所示。

**[barplot12.png]**

**KEGG功能条目相对丰度柱状图 COG功能条目相对丰度柱状图**

注：不同颜色代谢不同的功能条目，对应右侧图例；横轴代表不同的样本或分组，纵轴代表各功能条目的相对丰度。

结果目录：

Result\07\_FunctionPrediction\KEGG \Barplot

Result\07\_FunctionPrediction\COG \Barplot

**3.8.5 功能分布Heatmap图**

Heatmap可以用颜色变化来反映二维矩阵或表格中的数据信息，可以直观的将功能丰度值用定义的颜色深浅表示出来。对KEGG和COG功能预测结果进行Heatmap分析，结果如下图所示。

**[heatmap12.png]**

**KEGG heatmap analysis COG heatmap analysis**

注：横轴表示不同的样本，纵轴表示不同的功能条目，颜色的深浅与功能条目的丰度有关，颜色越深，丰度越高。

结果目录：

Result\07\_FunctionPrediction\KEGG \ Heatmap

Result\07\_FunctionPrediction\COG \ Heatmap

**3.8.6 差异功能LEfSe分析**

LEfSe分析根据分类学组成对样本按照不同的分组条件进行线型判别分析（LDA），找出对样本划分产生显著性差异影响的功能条目。

（1）KEGG

基于KEGG功能预测的LEfSe分析结果如下图所示，包括LDA值分布柱状图和组间具有统计学差异的功能条目在不同组中丰度比较图。

**[kegg1.png]**

**LDA Score**

注：LDA值分布柱状图中红色区域和绿色区域表示不同分组，树枝中红色节点表示在红色组别中起到重要作用的功能条目，绿色节点表示在绿色组别中起到重要作用的功能条目。图中仅展示了LDA Score大于设定值（默认设置为2）的功能条目，柱状图的长度代表LDA值的大小。

默认LDA值大于2为组间具有统计学差异的功能条目。各功能条目在不同组中丰度比较图详见附件，示例如下图所示。

**[kegg2.png]**

**Biomarker在各组样品中的相对丰度**

注：图中左右两侧表示不同的组别。横坐标表示样本，纵坐标表示功能条目的相对丰度。

（2）COG

基于COG功能预测的LEfSe分析结果如下图所示，包括LDA值分布柱状图和组间具有统计学差异的功能条目在不同组中丰度比较图，如下图所示。

**[cog1.png]**

**LDA Score**

注：LDA值分布柱状图中红色区域和绿色区域表示不同分组，树枝中红色节点表示在红色组别中起到重要作用的微生物类群，绿色节点表示在绿色组别中起到重要作用的微生物类群。图中仅展示了LDA Score大于设定值（默认设置为2）的功能条目，柱状图的长度代表LDA值的大小。

默认LDA值大于2为组间具有统计学差异的功能条目。各功能条目在不同组中丰度比较图详见附件，示例如下图所示。

**[cog2.png]**

**Biomarker在各组样品中的相对丰度**

注：图中左右两侧表示不同的组别。横坐标表示样本，纵坐标表示功能条目的相对丰度。

结果目录：

Result\07\_FunctionPrediction\KEGG \ LefSe

Result\07\_FunctionPrediction\COG \ LefSe

**3.8.7 差异功能STAMP分析**

基于PICRUSt功能二级分类结果，采用STAMP差异分析比较两组样品（Welch’s t-test 分析）或多组样品（ANOVA 分析，该软件只出Pvalue，无图片展示）之间功能的丰度，通过此分析可获得显著性差异功能。对KEGG和COG功能预测结果进行stamp分析，结果如下图所示。

**[stamp1.png]**

**STAMP Result of KEGG**

**[stamp2.png]**

**STAMP Result of COG**

注：左图所示为不同功能条目在两个样品或者两组样品中的丰度比例，中间所示为95% 置信度区间内的差异比例，最右边的值为 p 值，p 值＜0.05，表示差异显著。

结果目录：

Result\07\_FunctionPrediction\KEGG \ STAMP

Result\07\_FunctionPrediction\COG \ STAMP

1. **实验方法描述 Methods**
   1. **中文版实验方法**

采用CTAB或SDS方法对样本的基因组DNA进行提取，并对DNA的纯度和浓度进行检测。根据测序区域的选择，使用带Barcode的特异引物和高保真DNA聚合酶对选定的V3-V4可变区进行PCR扩增。PCR产物用2%琼脂糖凝胶电泳进行检测，并对目标片段进行切胶回收，胶回收采用AxyPrepDNA凝胶回收试剂盒（AXYGEN公司）。参照电泳初步定量结果，对PCR扩增回收产物用QuantiFluor™ -ST蓝色荧光定量系统（Promega公司）进行检测定量，按照每个样本的测序量要求，进行相应比例的混合。使用NEB Next® Ultra™DNA Library Prep Kit建库试剂盒进行文库构建。构建好的文库通过Agilent Bioanalyzer 2100和Qubit进行质检，文库质检合格后进行上机测序。

* 1. **英文版实验方法（供参考）**

**4.2.1 Sequencing**

**（1）Extraction of genome DNA**

Total genome DNA from samples was extracted using CTAB/SDS method. DNA concentration and purity was monitored on 1% agarosegels. According to the concentration, DNA was diluted to 1ng/μl using sterile water.

**（2）Amplicon Generation**

Primer：16S V3-V4: 341F-806R, 18S V9: 1380F-1510R, ITS1: ITS1F- ITS2R. 16S /18S rRNA genes were amplified used the specific primer with the barcode. All PCR reactions were carried out in 30μL reactions with 15μL of Phusion®High-Fidelity PCR Master Mix (New England Biolabs); 0.2μM of forward and reverse primers, and about 10 ng template DNA.Thermal cycling consisted of initial denaturation at 98 ℃ for 1 min, followed by 30 cycles of denaturation at 98℃ for 10 s, annealing at 50 ℃ for 30 s, and elongation at 72 ℃ for 60 s. Finally 72 ℃ for 5 min.

**（3）PCR Products quantification and qualification**

Mix same volume of 1X loading buffer (contained SYB green) with PCR products and operate electrophoresis on 2% agarose gel fordetection. Samples with bright main strip between 400-450bp were chosen for further experiments.

**（4）PCR Products Mixing and Purification**

PCR products was mixed in equidensity ratios. Then, mixture PCR products was purified with AxyPrepDNA Gel Extraction Kit (AXYGEN).

**（5）Library preparation and sequencing**

Sequencing libraries were generated using NEB Next®Ultra™DNA Library Prep Kit for Illumina (NEB, USA) followingmanufacturer’s recommendations and index codes were added. The library quality was assessed on the Qubit@ 2.0 Fluorometer (Thermo Scientific) and Agilent Bioanalyzer 2100 system. At last, the library was sequenced on an Illumina Miseq/HiSeq2500 platform and 250bp/300bp paired-end reads were generated.

**4.2.2 Data analysis**

**（1）Paired-end reads assemblies**

Paired-end reads from the original DNA fragments were merged using FLASH, a very fast and accurate analysis tool, which was designed to merge paired-end reads when at least some of the reads overlap the read generated from the opposite end of the same DNA fragment. Paired-end reads was assigned to each sample according to the unique barcodes.

**（2）OTU cluster and Species annotation**

Sequences analysis were performed by UPARSE software package using the UPARSE-OTU and UPARSE-OTUref algorithms. In-house Perl scripts were used to analyze alpha (within samples) and beta (among samples) diversity. Sequences with ≥97% similarity were assigned to the same OTUs. We pick a representative sequences for each OTU and use the RDP classifier to annotate taxonomic information for each representative sequence. In order to compute Alpha Diversity, we rarify the OTU table and calculate three metrics: Chao1 estimates the species abundance; Observed Species estimates the amount of unique OTUs found in each sample, and Shannon index. Rarefaction curves were generated based on these three metrics.

**（3）Phylogenics distance and community distribution**

Graphical representation of the relative abundance of bacterial diversity from phylum to species can be visualized using Krona chart. Cluster analysis was preceded by principal component analysis (PCA), which was applied to reduce the dimension of the original variables using the QIIME software package. QIIME calculates both weighted and unweighted unifrac distance, which are phylogenetic measures of beta diversity. We used unweighted unifrac distance for Principal Coordinate Analysis (PCoA) and Unweighted Pair Group Method with Arithmetic mean (UPGMA) Clustering. PCoA helps to get principal coordinates and visualize them from complex, multidimensional data. It takes a transformation from a distance matrix to a new set of orthogonal axes. By which the maximum variation factor is demonstrated by first principal coordinate, and the second maximum one by the second principal coordinate, and so on. UPGMA Clustering is a type of hierarchical clustering method using average linkage and can be used to interpret the distance matrix.

**（4）Statistical analysis**

To confirm differences in the abundances of individual taxonomy between the two groups, STAMP software was utilized. LEfSe was used for the quantitative analysis of biomarkers within different groups. This method was designed to analyze data in which the number of species is much higher than the number of samples and to provide biological class explanations to establish statistical significance, biological consistency, and effect-size estimation of predicted biomarkers. To identify differences of microbial communities between the two groups, ANOSIM and ADONIS were performed based on the Bray-Curtis dissimilarity distance matrices.

1. **16S报告阅读顺序建议 Reading Order Suggestion**

**Step1. 测序数据是否合格？**

测序数据是否合格主要有两个判定指标，一是测序下机数据是否合格，可以根据4.1 测序数据预处理及质控统计查看Q20，Q30等质控指标。二是测序数据量是否足够，主要参考4.4.2 Rarefaction Curve以及4.4.3 Shannon，它们的区别在于Rarefaction Curve是评估OTU数量是否趋近饱和，Shannon是评估α-多样性指数是否趋近饱和。

**Step2. 了解各组内的菌群组成情况**

查看一组样本内的菌群组成，可以通过宏观的方式，即4.4.1 α多样性指数观察组内的菌群多样性；也可以通过4.2.1 OTU聚类和物种注释查看组内的特定物种组成。4.3 物种组成分析中的各项分析是对物种组成的进一步统计及展示。

**Step3. 判断各样本整体上的差异情况？**

（1）通过4.5.2 PCA分析，4.5.3 PCoA分析以及4.5.4 NMDS分析对各样本的差异情况进行总体评估。这三项分析的主要区别在于：PCA分析利用原始OTU数据表进行分析，是对测序数据的最客观还原；PCoA分析使用特定的距离矩阵，从而在OTU数据的基础上引入不同的加权值。例如基于Weighted Unifrac的PCoA分析中，两种进化距离较近的物种的丰度差异对样本的整体差异影响就会远小于PCA分析；NMDS分析是对OTU数据表进行秩次排序后，利用排序结果进行主成分分析，主要适用于菌群差异较大的样本。

（2）可以使用统计检验手段对组间差异进行评估。主要分析有基于多样性指数的差异检验4.4.6 α多样性指数组间差异分析、4.5.1 β多样性指数组间差异分析；判断组间差异是否大于组内差异的4.6.1 Anosim检验；判断分组是否有意义的4.6.2 Adonis检验。

（3）可以以聚类树的形式展现样本间的差异情况，可见4.3.5 UPGMA 聚类树。

**Step4 差异分组间具体有哪些物种差异？**

组间的差异物种可以通过4.7.1 LEfSe分析以及4.7.2 STAMP差异分析进行分析及检验。前者为多维统计，通过LDA值的表征物种丰度差异对组间差异的贡献程度，再用Kruskal-Wallis秩和检验显著性；后者为单维统计，针对特定物种进行组间的差异检验。在进行菌群biomarker初筛时，通常用LEfSe分析。

**Step5 通过菌群信息进行功能预测**

基于数据库中菌群与功能基因的对应关系，以及测序中获得的菌群信息，可对样本中的功能基因进行预测，主要参考4.8 功能预测中各项分析。由于16S rDNA测序获得的功能基因只是预测，如需深入研究样本中的功能基因，建议后续再进行宏基因组测序。

1. **参考文献 References**

[1] Caporaso, J. Gregory, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences 108.Supplement 1 (2011): 4516-4522.  
[2] Youssef, Noha, et al. Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. Applied and environmental microbiology 75.16 (2009): 5227-5236.

[3] Hess, Matthias, et al. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331.6016 (2011): 463-467.

[4] Ondov, Brian D., Nicholas H. Bergman, and Adam M. Phillippy. Interactive metagenomic visualization in a Web browser. BMC bioinformatics 12.1 (2011): 385.

[5] DeSantis, T. Z., et al. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic acids research 34.suppl 2 (2006): W394-W399.

[6] Lozupone, Catherine, and Rob Knight. UniFrac: a new phylogenetic method for comparing microbial communities. Applied and environmental microbiology 71.12 (2005): 8228-8235.  
[7] Lozupone, Catherine, et al. UniFrac: an effective distance metric for microbial community comparison. The ISME journal 5.2 (2011): 169.  
[8] Lozupone, Catherine A., et al. Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. Applied and environmental microbiology 73.5 (2007): 1576-1585.

[9] Lundberg, Derek S., et al. Practical innovations for high-throughput amplicon sequencing.Nature methods 10.10 (2013): 999-1002.

[10] Avershina, Ekaterina, Trine Frisli, and Knut Rudi. De novo Semi-alignment of 16S rRNA Gene Sequences for Deep Phylogenetic Characterization of Next Generation Sequencing Data. Microbes and Environments 28.2 (2013): 211-216.  
[11] Magali Noval Rivas, PhD, Oliver T. Burton, et al. A microbita signature associated with experimental food allergy promotes allergic senitization and anaphylaxis. The Journal of Allergy and Clinical Immunology.Volume 131, Issue 1, Pages 201-12212, January 2013.

[12] Anderson, M.J. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecology, 26: 32-46.  
[13] McArdle, B.H. and M.J. Anderson. 2001. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. Ecology, 82: 290-297.  
[14] Warton, D.I., Wright, T.W., Wang, Y. 2012. Distance-based multivariate analyses confound location and dispersion effects. Methods in Ecology and Evolution, 3, 89-101.  
[15] Zapala, M.A. and N.J. Schork. 2006. Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. Proceedings of the National Academy of Sciences, USA, 103:19430-19435.  
[16] Segata, Nicola, et al. Metagenomic biomarker discovery and explanation. Genome Biol 12.6 (2011): R60.

1. **生信数据分析云平台 Data analysis platform**

登录中科新生命APT-BioCloud分析云平台，您可自行进行统计学、火山图、聚类分析、富集分析等多种生物信息学分析，快速便捷输出个性化的生信分析图。注册与登录地址如下，如有疑问可联系中科新生命当地销售。

注册网址：[http//cloud.aptbiotech.com/#/register](http/cloud.aptbiotech.com/#/register) 登录网址：[http//cloud.aptbiotech.com/#/login](http/cloud.aptbiotech.com/" \l "/login)



1. **组学期刊投稿指南（供参考）Publication Recommendation**

**8.1 第一档期刊（影响因子9 分以上）**

组学期刊：Nature Genetics，Genome Research， Microbiome，Molecular Systems Biology等

综合性期刊： Cell，Nature，Science，Nature Communications，PNAS，Cell Metabolism，Molecular Cell，Cell Research等

医学专业期刊：Nature Medicine，Cancer Cell，Cancer Research，European Heart Journal，Circulation，Journal of the American College of Cardiology，Gut，Gastroenterology，Hepatology，Diabetes Care，Molecular Psychiatry，Immunity 等

农林专业期刊：Nature Plants，Molecular Plant，Plant Cell等

**组学思路**：在这类高水平期刊中，组学驱动的研究主要包括：1.大队列研究，筛选临床生物标志物以及进行疾病分子分型等。2. 组学+后续功能机制研究。3. 结合时下的热点，如肠道微生物进行相关研究，阐明肠道微生物与疾病的相关性及相互作用机制。

**8.2 第二档期刊（影响因子4~9分）**

组学期刊：Molecular & Cellular Proteomics，Metabolism，Metallomics，Gigascience，Proteomics & Bioinformatics等

综合期刊： Scientific Data，Scientific Reports，Cell Reports，Progress in Lipid Research，Frontiers in Microbiology，mSystems等

医学专业期刊：Advances in Cancer Research，International Journal of Cancer，Diabetologia，Thyroid，Diabetes，Oncogene，mBio等

农林领域专业期刊：New Phytologist，Plant Biotechnology Journal，Plant Physiology，Plant Journal，Plant Cell and Environment，Food Chemistry，Journal of Experimental Botany等

**组学思路**：发表在该分数段期刊的组学文章，1. 对于蛋白组学，修饰组学而言，要求组学+功能验证与机制研究（如表达量验证，基因敲除，点突变，位点特异性抗体验证，蛋白活性验证，后续功能实验，动物模型等）。2. 对于脂质组学是个例外，研究基础浅，新颖性高，纯脂质组学也能发表文章，无需后续验证。

**8.3 第三档期刊（影响因子4分以下）**

组学期刊：BMC Genomics， Proteomics，Journal of Proteomics，Journal of Proteome Research，Metabolomics等

综合期刊：Plos One等

农林领域专业期刊: BMC Plant Biology，Journal of Agricultural and Food Chemistry，Frontiers in Plant Science等

**组学思路**：这类文章一般是最常见也是最简单的组学研究思路。1. 蛋白质组学方向：组学+差异蛋白WB或PRM验证。2.蛋白质翻译后修饰组学方向：则以修饰组学数据分析为主。3. 代谢组学方向：组学（+差异代谢物MRM验证）。

1. **部分合作发表文章 Cooperative Projects Papers**

**9.1 医学**

[1] Lysozyme-Assisted Photothermal Eradication of Methicillin-Resistant Staphylococcus aureus Infection and Accelerated Tissue Repair with Natural Melanosome Nanostructures. **ACS Nano**

[2]One-Carbon Metabolism Supports SAdenosylmethionine and Histone Methylation to Drive Inflammatory Macrophages. **Molecular Cell**

[3] Early Prediction of Developing Type 2 Diabetes by Plasma Acylcarnitines: Population-Based Study. **Diabetes Care**

[4] Hepatocellular Carcinoma-Associated Protein TD26 Interacts and Enhances Sterol Regulatory Element-Binding Protein 1 Activity to Promote Tumor Cell Proliferation and Growth. **Hepatology**

[5] Polyunsaturated fatty acids metabolism, purine metabolism and inosine as potential independent diagnostic biomarkers for major depressive disorder in children and adolescents. **Mol Psychiatry**

[6] CLOCK Acetylates ASS1 to Drive Circadian Rhythm of Ureagenesis. **Mol Cell**

[7] EGFR modulates monounsaturated fatty acid synthesis through phosphorylation of SCD1 in lung cancer. **Molecular Cancer**

[8] Protein phosphatase 2A has an essential role in promoting thymocyte survival during selection. **Proc Natl Acad Sci U S A**

[9] O-GlcNAcylation destabilizes the active tetrameric PKM2 to promote the Warburg effect. **Proc Natl Acad Sci U S A**

[10] Proteomic analysis of solid pseudopapillary tumor of the pancreas reveals dysfunction of the endoplasmic reticulum protein processing pathway. **Molecular & Cellular Proteomics**

**9.2 植物**

[1] Natural alleles of a proteasome α2 subunit gene contribute to thermotolerance and adaptation of African rice. **Nat Genetics**

[2]Maize Oxalyl-CoA Decarboxylase1 Degrades Oxalate and Affects the Seed Metabolome and Nutritional Quality. **Plant Cell**

[3] OsSPL3, a SBP-Domain Protein, Regulates Crown Root Development in Rice. **Plant Cell**

[4] RRM Transcription Factors Interact with NLRs and Regulate Broad-Spectrum Blast Resistance in Rice. **Mol Cell**

[5] Global analysis of sumoylation function reveals novel insights into development and appressorium-mediated infection of the rice blast fungus.**New Phytologist**

[6] Quantitative Phosphoproteomic and Metabolomic Analyses Reveal GmMYB173 Optimizes Flavonoid Metabolism in Soybean under Salt Stress. **Molecular & Cellular Proteomics**

[7] Label-Free Quantitative Proteomics of Lysine Acetylome Identifies Substrates of Gcn5 in Magnaporthe oryzae Autophagy and Epigenetic Regulation.**mSystems**

[8] Phototrophy and starvation-based induction of autophagy upon removal of Gcn5-catalyzed acetylation of Atg7 in Magnaporthe oryzae. **Autophagy**

[9] Integrative Transcriptome and Proteome Analysis Identifies Major Metabolic Pathways Involved in Pepper Fruit Development. **J Proteome Res**

[10] Proteomics integrated with metabolomics: analysis of the internal causes of nutrient changes in alfalfa at different growth stages. **BMC Plant Biology**

**9.3 动医/动科/食品等其他**

[1] Fecal Microbiota Transplantation Beneficially Regulates Intestinal Mucosal Autophagy and Alleviates Gut Barrier Injury. **mSystems**

[2] Quantitative Phosphoproteomic Analysis among Muscles of Different Color Stability using Tandem Mass Tag Labeling.**Food Chem**

[3] Comparative Analysis of Whey N-Glycoproteins in Human Colostrum and Mature Milk Using Quantitative Glycoproteomics.**J Agric Food Chem**

[4] N-glycosylation proteomic characterization and cross-species comparison of milk fat globule membrane proteins from mammals. **Proteomics**

[5] Metabolomics Reveals that Crossbred Dairy Buffaloes Are More Thermotolerant than Holstein Cows under Chronic Heat Stress. **Journal of agricultural and food chemistry**

[6] The Synergistic Effect of Exogenous Glutamine and Rifampicin Against Mycobacterium Persisters. **Frontiers in Microbiology**