

中国科学院微生物研究所基因组研究中心

地址: 北京市朝阳区北辰西路一号院3号
migenome@im.ac.cn

电话: 010-64807433

邮箱:

环境微生物组16s rDNA测序与分析报告

项目信息: 16S

客户单位: 宁波格鲁康

报告时间: 2018-09-05

项目编号: 16S18NBGLG65

1. 概述

- 1.1 实验流程

- 1.2 生物信息分析路线

2. 样品信息

3. 信息分析结果

- 3.1 数据质控与优化

- 3.1.1 测序数据格式

- 3.1.2 数据质量评估

- 3.1.3 序列拼接

- 3.1.4 数据过滤

- 3.1.5 嵌合体序列检测

- 3.1.6 高质量序列长度分布统计

- 3.2 OTU聚类及物种注释

- 3.2.1 OTU聚类分析

- 3.2.2 物种结构分析

- 3.2.3 物种丰度聚类热图

- 3.2.4 OTU分布Venn 图

- 3.2.5 单样本多级物种组成图

- 3.2.6 OTU物种注释结果统计

- 3.2.7 各分类水平的微生物类群统计

- 3.3 Alpha多样性分析

- 3.3.1 稀释曲线分析

- 3.3.2 Shannon-Wiener曲线

- 3.3.3 Simpson曲线

- 3.3.4 Rank-Abundance曲线

- 3.3.5 Specaccum物种累积曲线

- 3.3.6 多样性指数

- 3.4 Beta多样性分析

- 3.4.1 距离矩阵热图

- 3.4.2 PCoA分析

- 3.4.3 PCA分析

- 3.4.4 NMDS分析

- 3.4.5 UPGMA聚类树

- 3.4.6 PLS-DA偏最小二乘法判别分析

- 3.5 差异分析

- 3.5.1 LEfSe分析

- 3.5.2 MetaStats分析

1. 概述

2. 样品信息

3. 信息分析结果

4. 参考文献

3.5.5 ANOSIM相似度分析

3.5.6 随机森林分析

4. 参考文献

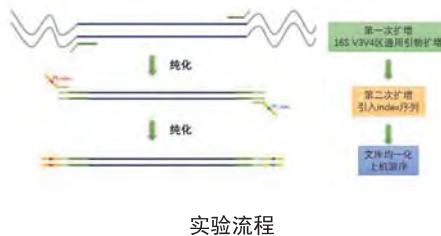
1. 概述

微生物是地球上生物多样性最为丰富的资源。探讨微生物多样性对于研究微生物与人类疾病、环境治理和微生物资源的开发利用有着重要的理论和现实意义。目前，因为研究手段的限制，纯培养方法获得的环境微生物多样性只占总量的0.1%~1%，严重阻碍了对于不同环境条件下微生物群落多样性和分布规律的研究。基于第二代高通量测序技术，不经过分离培养微生物，直接对样本中微生物的DNA进行测定。不断更新的生物信息学分析方法，不仅可以让科研工作者获得准确的不同环境中微生物群落的组成信息和多样性信息，还可以根据不同的实验设计，比较不同空间或/和时间上微生物结构发生的变化，寻找变化规律；与环境因子的关联性分析、共表达网络分析、功能基因预测等深入的数据挖掘为进一步认知和理解环境微生物群落提供了有力的研究工具。

16S rRNA基因普遍存在于原核细胞，在细菌基因组中位于核糖体小亚基（约1540 bp），该区域兼顾保守性和高变性，含有10个保守区域(Conserved Regions)和9个高变区域(variable Regions)，保守区可用于设计引物进行目的片段的扩增，而通过对高变区的分析可以辨别细菌种类。因此，16S rRNA基因被认为是最适于细菌系统发育学研究和物种分类鉴定。目前用于二代测序的区域主要有V3区，V6区，V1区，V3-V4区、和V4-V5区等。

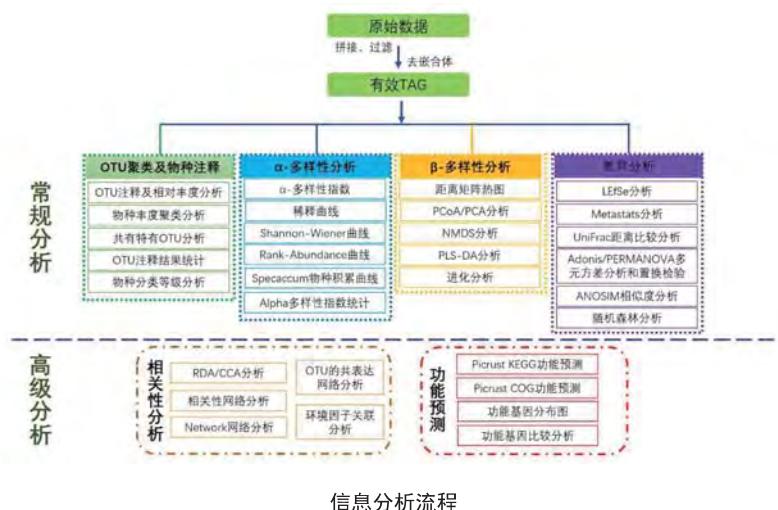
1.1 实验流程

将检测合格的环境微生物DNA样本对其指定区域进行PCR扩增、文库制备、文库质检、定量，使用设定的TAG序列进行样本区分。采用Illumina Hiseq 2500高通量测序平台对检测合格的文库进行测序。实验流程如下图所示：



实验流程

1.2 生物信息分析路线



信息分析流程

Hiseq/Miseq测序得到的PE reads根据overlap关系，将测序所得的双端序列拼接成一条目标区域序列，对目标序列进行质控过滤；过滤后的序列与参考数据库作比对，去除嵌合体序列得到最终得优化序列。基于优化序列进行OTU聚类分析和物种分类学(Taxonomy)注释。基于OTU聚类结果进行多样性指数分析。基于分类学信息进行物种结构分析和物种差异分析。

2. 样品信息

项目样品的详细信息如下表所示：

#SampleID	Group
◀	▶

Page 0 of 10 ▾

注 : Sample: 样品名称 ; Group: 样品分组信息 ;

查看原文件请点击:[Fasting Map.txt](#)

3. 信息分析结果

3.1 数据质控与优化

3.1.1 测序数据格式

根据barcode序列区分样本，提取出的数据以标准的fastq格式保存。以双端测序（PE：paired-end）数据为例，每一个样本有read1.fastq和read2.fastq两个文件，分别代表5'→3'和3'→5'的测序结果。Fastq文件格式如下：

第1行以@开头，为read的ID。第2行是该read的序列。第4行是该序列的测序质量，每个字符对应为第2行每个碱基的测序质量值，用Phred值+33后，所对应的ASCII字符来表示。Phred值的计算方法为：

$$Qphred = -10 * \log 10(e)$$

其中e是碱基错误率。质量评分：质量评分指的是一个碱基的错误概率的对数值。其最初在Phred拼接软件中定义与使用，其后在许多软件中得到使用。其质量得分与错误概率的对应关系见下表：

质量得分与错误概率对照表

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10000	99.99%

如上表所示，如果一个碱基的Q值为20，表示这个碱基可能测错的概率为1%。

3.1.2 数据质量评估

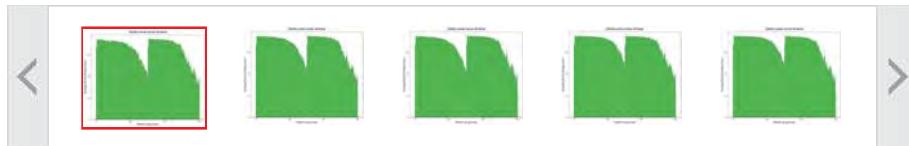
根据Fastq文件对测序样品进行数据质量评估。单个样品A3的碱基质量分布如下图所示。

1. 概述

2. 样品信息

3. 信息分析结果

4. 参考文献



注：横坐标为reads的碱基位置，纵坐标为单碱基平均Phred值。前250bp 为双端测序序列的Read1的碱基质量值分布情况，后250bp为read2 的碱基质量值分布情况。

所有样品图请点击:01.QC/PNG/

测序样品数据产出及质量评估统计如下表：

样品测序数据评估统计

样品测序数据评估统计				
Sample	PE_reads	Q20(%)	Q30(%)	GC(%)

Page 0 of 10

注：Sample:样品名称；PE_reads: 样品中成对reads数；Q20: Q值大于20的碱基比例；Q30: 样品中质量值高于30碱基比例；GC: 样品数据的GC含量

原始文件请点击:01.QC/QC_stat.txt

3.1.3 序列拼接

根据PE reads之间的overlap采用Flash[1]软件对数据进行拼接。样本的有效序列数据量统计结果见表 4：

3.1.4 数据过滤

由于测序过程中会引入错误或者不可靠碱基，严重影响后续分析结果准确性。因此，分析数据只保留高质量 (Q值 ≥ 25) 的碱基比例大于等于90%的reads，否则舍弃。过滤数据统计见表 4。

3.1.5 嵌合体序列检测

嵌合体在遗传学上用以指不同遗传性状嵌合或混杂表现的个体。嵌合体序列的出现会导致测序结果中产生一些实际并不存在的核酸序列，影响结果的可靠性。因此，采用QIIME[2]软件中的usearch8.0[3,4]软件进行嵌合体序列的检测及过滤。嵌合体序列检测结果见表4。

数据处理统计表

数据处理统计表					
Sample ID	PE reads	Tags	Filter	Dechime	Chime(%)

Page 0 of 10

1. 概述

2. 样品信息

3. 信息分析结果

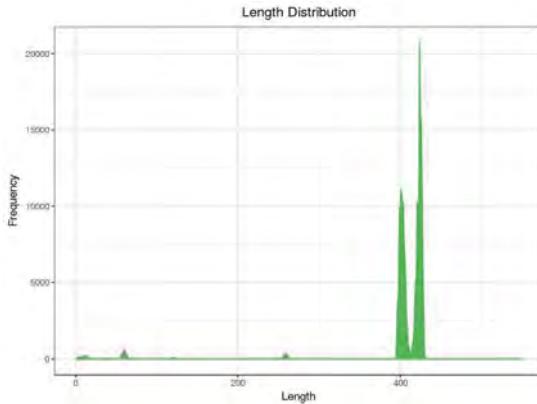
4. 参考文献

去除嵌合体后得到的序列数 ; Chime : 嵌合体比例 ;

原始文件请点击:01.QC/Data_process_stat.txt

3.1.6 高质量序列长度分布统计

所有样品的目标序列长度分布如下图所示



高质量序列长度分布图

注 : 横坐标为全部样本中高质量序列的长度分布 , 纵坐标则为各长度值所对应的序列总数。

数据处理结果目录 : 01.QC/ Readme

3.2 OTU聚类及物种注释

OTU (Operational Taxonomic Units , 操作分类单元) 是在系统发生分析或群体遗传研究中的一个假定的分类单元 , 通过一定的距离度量方法计算两两不同序列之间的距离度量或相似性 , 继而设置特定的分类阈值 , 获得同一阈值下的距离矩阵 , 进行聚类操作 , 形成不同的分类单元。

根据 97% 的序列相似性水平 , 利用 QIIME[2] 软件包中的 Uclust 方法进行 OTU 聚类分析。然后基于 Silva (Release128 <https://www.arb-silva.de/documentation/release-128/>) 参考数据库 , 利用 RDP Classifier[5] 软件 (confidence cutoff=0.8) 对每个样品的 OTUs 进行物种分类学 (Taxonomy) 注释。

3.2.1 OTU聚类分析

OTU聚类部分结果如下表所示 :

OTU聚类统计表

#OTU ID	B7	B8	A8	B3	B5	B4	A7	A3
GQ006259.1.1327	1.0	33.0	6.0	5.0	2.0	0.0	0.0	0.0
HE681263.1.1518	39.0	0.0	0.0	0.0	0.0	3.0	0.0	0.0
HW066318.1.1503	0.0	0.0	16.0	0.0	0.0	0.0	0.0	0.0
FJ658315.1.1400	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.0
AF132250.1.1452	0.0	0.0	0.0	14.0	0.0	0.0	0.0	2.0
JN012642.1.1446	0.0	0.0	0.0	0.0	9.0	0.0	0.0	0.0
HM438508.1.1493	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AB969652.1.1529	0.0	1.0	9.0	0.0	0.0	0.0	0.0	0.0
KJ192017.1.1477	0.0	0.0	13.0	0.0	0.0	0.0	0.0	0.0
HM308142.1.1356	0.0	0.0	6.0	11.0	0.0	0.0	0.0	0.0
HQ681741.1.1450	0.0	0.0	0.0	12.0	0.0	1.0	13.0	0.0
AB702728.1.1499	3.0	0.0	28.0	29.0	5.0	26.0	0.0	2.0
JF100816.1.1349	0.0	6.0	1.0	3.0	0.0	0.0	0.0	0.0
HW066284.1.1495	26.0	10.0	0.0	0.0	24.0	25.0	12.0	1.0
EU655709.1.1320	0.0	0.0	44.0	0.0	0.0	0.0	0.0	0.0
GQ138628.1.1373	18.0	8.0	6.0	11.0	9.0	27.0	5.0	7.0

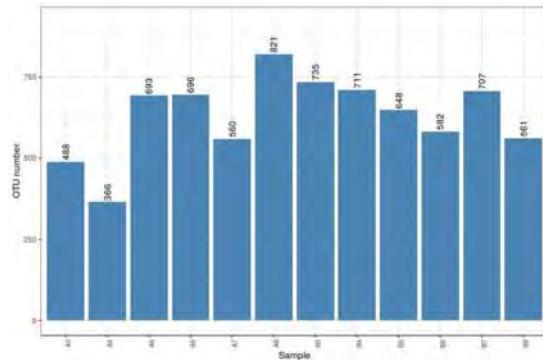
1. 概述

2. 样品信息

3. 信息分析结果

4. 参考文献

/02.OTU_taxa/Barplot/OTU_barplot.txt

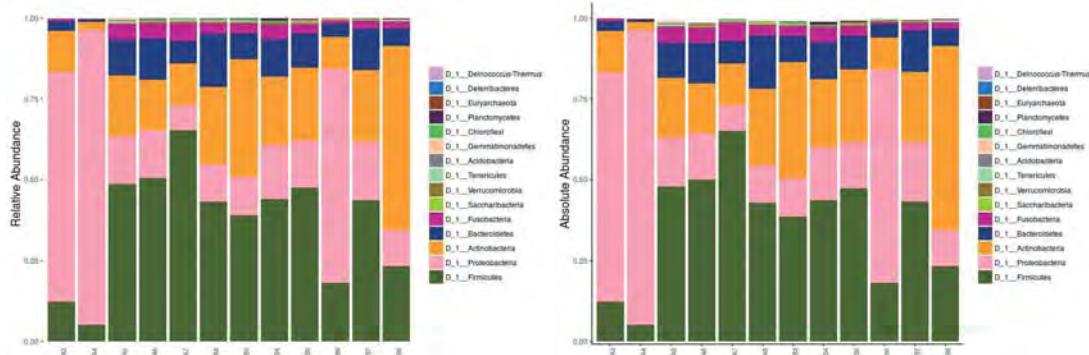


各样品OTU数据统计图

3.2.2 物种结构分析

统计所有样本在门、纲、目、科、属各层次上的分类结果。基于丰度前十五的物种采用累积柱状图比较样本间的物种组成差异，并进行列表展示。

全样本在门层次的群落结构分析结果如下图所示：



门层次菌落结构柱状图

注：上：仅展示前15个物种在每个样本内的比例关系；下：即展示前15个物种在每个样本中的比例关系，同时展示这15个物种在该样本中所占比例。每个柱子代表一个样本，不同色块代表相应层次的不同的类，色块的长度为该类序列占该样本的总序列数的比例

原图路径:02.OTU_taxa/Barplot/ Readme

部分样品物种注释统计表

ID	B7	B8	A8	B3	B5	B4	A7
D_1_Firmicutes	0.434935205184	0.234472952886	0.430610852008	0.387374618404	0.474491878214	0.436995995392	0.6515940551
D_1_Proteobacteria	0.181479481641	0.111743036257	0.114605846889	0.116659398168	0.142437352053	0.163750068572	0.0797662891
D_1_Actinobacteria	0.218412526998	0.567569314289	0.236833124787	0.360499345835	0.225450983593	0.210763069834	0.1282865481
D_1_Bacteroidetes	0.127699784017	0.0539649712402	0.163974519395	0.0805167902311	0.104481266835	0.114103900379	0.0692239291
D_1_Fusobacteria	0.0225701943844	0.0196471272539	0.0348083266978	0.031399912778	0.0308546241123	0.0462449942399	0.0583005201
D_1_Saccharibacteria	0.0011879049676	0.00206811865831	0.00591514048459	0.00218054949847	0.003591514354747	0.00153601404356	0.0021592781
D_1_Verrucomicrobia	0.00329373650108	0.000581658372649	0.0008531452622	0.00174443959878	0.00171414578402	0.00411432333096	0.0019052451
D_1_Tenericutes	0.00172786177106	0.000517029664577	0.00142190877033	0.00267117313563	0.00179577177373	0.0	0.0040645241
D_1_Acidobacteria	0.00151187904968	0.000840173204938	0.00244568308497	0.00163541212386	0.00195902375316	0.000274288222064	0.0015241961
D_1_Gemmatimonadetes	0.0	0.0039423511924	0.000910021613013	0.000545137374618	0.00138764182516	0.0	0.0
D_1_Chloroflexi	0.000647948164147	0.0	0.0	0.00419755778456	0.0	0.00499204564156	0.0020322621
D_1_Planctomycetes	0.0	0.00103405932915	0.0	0.000654164849542	0.00155089380459	0.00680234790718	0.0
D_1_Euryarchaeota	0.0011879049676	0.00090480191301	0.0	0.000218054949847	0.00261203167088	0.0	0.0
D_1_Deferrribacteres	0.0	0.00129257416144	0.00125127971789	0.0	0.0	0.0	0.0
D_1_Deinococcus-Thermus	0.0	0.0	0.0	0.0	0.00122438984573	0.0	0.0

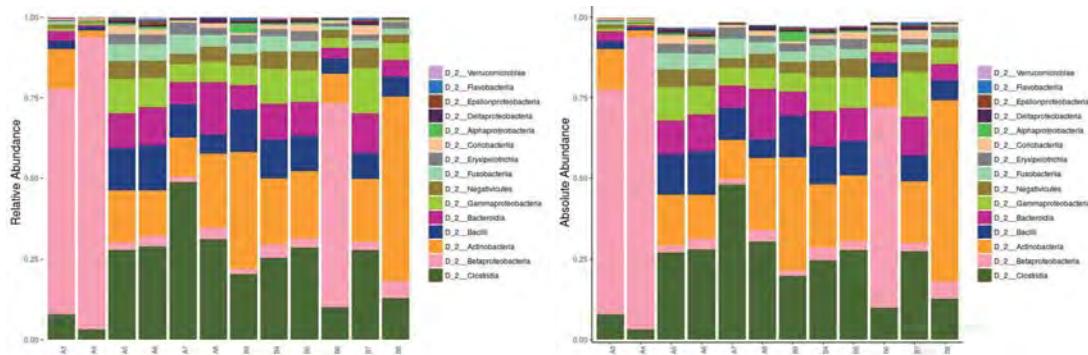
原始文件请点击:02.OTU_taxa/Barplot/L2_toptable.txt

1. 概述

2. 样品信息

3. 信息分析结果

4. 参考文献



纲层次菌落结构柱状图

注：上：仅展示前15个物种在每个样本内的比例关系；下：即展示前15个物种在每个样本中的比例关系，同时展示这15个物种在该样本中所占比例。每个柱子代表一个样本，不同色块代表相应层次的不同的类，色块的长度为该类序列占该样本的总序列数的比例

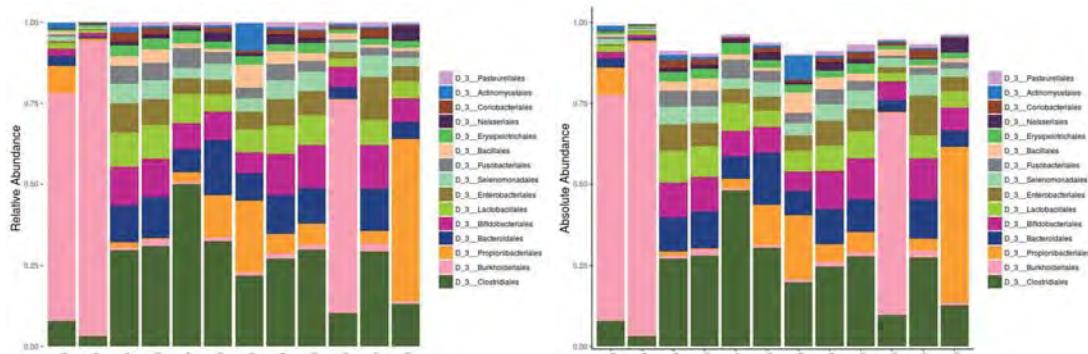
原图路径:02.OTU_taxa/Barplot/ Readme

部分样品物种注释统计表

ID	B7	B8	A8	B3	B5	B4	A7
D_2_Clostridia	0.274514038877	0.126672267821	0.304402229553	0.197993894461	0.278752754877	0.247133688079	0.48190016511
D_2_Betaproteobacteria	0.0260799136069	0.0526077683707	0.0363439881697	0.0170627998256	0.0283242184311	0.0416918097537	0.01663914646
D_2_Actinobacteria	0.19055075594	0.562980676016	0.222159026277	0.350795900567	0.201534568607	0.194360634154	0.11977645116
D_2_Bacilli	0.0791036717063	0.0616557875008	0.0578432487772	0.127507631923	0.107174924496	0.116298206155	0.10072399336
D_2_Bacteroidia	0.120680345572	0.0515090803335	0.158002502559	0.0745747928478	0.101869235164	0.108947281804	0.06833481519
D_2_Gammaproteobacteria	0.138660907127	0.0507981645447	0.0628483676487	0.0586022677715	0.0967267978124	0.105820396072	0.05525212752
D_2_Negativicutes	0.0624190064795	0.0252698248562	0.0448754407917	0.036360662887	0.056730062852	0.0513467551703	0.03162707986
D_2_Fusobacteriia	0.0225701943844	0.0196471272539	0.0348083266978	0.031399912778	0.0308546241123	0.0462449942399	0.05830052076
D_2_Erysipelotrichia	0.0184125269978	0.0201641569185	0.022238653168	0.0241495856956	0.0304464941637	0.0208459048768	0.03708878446
D_2_Coriobacteriia	0.02656587473	0.00400697990047	0.0146740985098	0.00697775839512	0.0207330013876	0.0140984146141	0.00609678646
D_2_Alphaproteobacteria	0.00183585313175	0.00484715310541	0.000910021613013	0.0296554731792	0.00604032323892	0.00537604915245	0.00190524576
D_2_Deltaproteobacteria	0.00955723542117	0.000840173204938	0.012797178933	0.0052331879634	0.00734633907436	0.00608919852982	0.00152419662
D_2_Epsilonproteobacteria	0.00534557235421	0.00264977703096	0.0017062905244	0.00610553859573	0.00399967349604	0.00477261506391	0.00444557341
D_2_Flavobacteriia	0.00658747300216	0.000710915788793	0.00227505403253	0.00490623637157	0.00228552771202	0.00186515991003	0.00088911465
D_2_Verrucomicrobiae	0.00253779697624	0.000581658372649	0.000284381754067	0.00174443959878	0.00171414578402	0.00411432333096	0.00190524576

原始文件请点击:02.OTU_taxa/Barplot/L3_toptable.txt

全样本在目层次的群落结构分析结果如下图所示：



目层次菌落结构柱状图

注：上：仅展示前15个物种在每个样本内的比例关系；下：即展示前15个物种在每个样本中的比例关系，同时展示这15个物种在该样本中所占比例。每个柱子代表一个样本，不同色块代表相应层次的不同的类，色块的长度为该类序列占该样本的总序列数的比例

原图路径:02.OTU_taxa/Barplot/ Readme

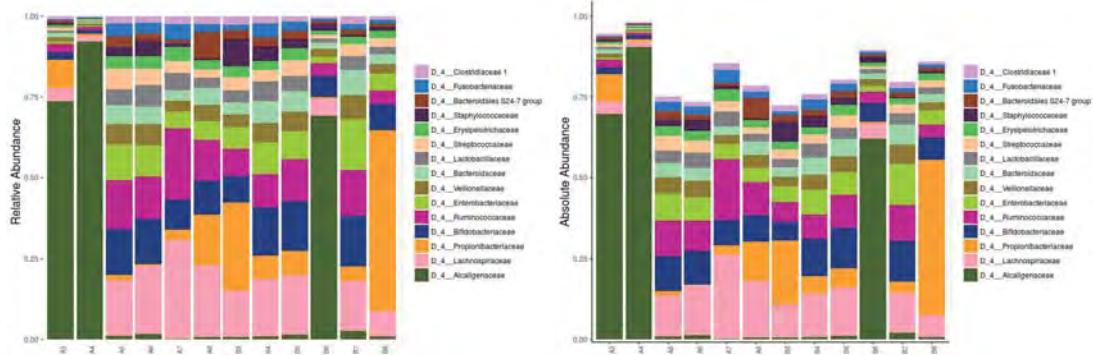
部分样品物种注释统计表

ID	B7	B8	A8	B3	B5	B4	A7
D_3_Clostridiales	0.274514038877	0.126672267821	0.304402229553	0.197993894461	0.278752754877	0.247133688079	0.481900165121

1. 概述		2. 样品信息		3. 信息分析结果		4. 参考文献	
D_3_Bacteroidales	0.120680345572	0.0515090803335	0.158002502559	0.0745747928478	0.101869235164	0.108947281804	0.0683348151912
D_3_Bifidobacteriales	0.125485961123	0.0697343760098	0.0818450688204	0.058547754034	0.124153130357	0.114707334467	0.0790041915407
D_3_Lactobacillales	0.0711123110151	0.0504103922963	0.0492549198043	0.0639446140427	0.0847277773243	0.079872730265	0.0856090435666
D_3_Enterobacteriales	0.124190064795	0.0449815808182	0.0445341826868	0.0488988225033	0.0685658313607	0.075758406934	0.0461069477963
D_3_Selenomonadales	0.0624190064795	0.0252698248562	0.0448754407917	0.036360662887	0.056730062852	0.0513467551703	0.0316270798933
D_3_Fusobacteriales	0.0225701943844	0.0196471272539	0.0348083266978	0.031399912778	0.0308546241123	0.0462449942399	0.0583005207672
D_3_Bacillales	0.00799136069114	0.0112453952045	0.00858832897281	0.0635630178805	0.0224471471717	0.0364254758901	0.0151149498285
D_3_Erysipelotrichales	0.0184125269978	0.0201641569185	0.022238653168	0.0241495856956	0.0304464941637	0.0208459048768	0.0370887844532
D_3_Neisseriales	0.00323974082073	0.0424610612034	0.0253668524627	0.00774095071958	0.0142029222104	0.0265510998958	0.00990727803887
D_3_Coriobacteriales	0.02656587473	0.00400697990047	0.0146740985098	0.00697775839512	0.0207330013876	0.0140984146141	0.00609678648546
D_3_Actinomycetales	0.00566954643629	0.00258514832289	0.00659765669435	0.0752289576973	0.00342829156804	0.0060343408854	0.00152419662136
D_3_Pasteurellales	0.0110151187905	0.00542881147806	0.00830394721875	0.0034343654601	0.0199167414905	0.0134401228811	0.00800203226216

原始文件请点击:02.OTU_taxa/Barplot/L4_toptable.txt

全样本在科层次的群落结构分析结果如下图所示：



科层次菌落结构柱状图

注：上：仅展示前15个物种在每个样本内的比例关系；下：即展示前15个物种在每个样本中的比例关系，同时展示这15个物种在该样本中所占比例。每个柱子代表一个样本，不同色块代表相应层次的不同的类，色块的长度为该类序列占该样本的总序列数的比例

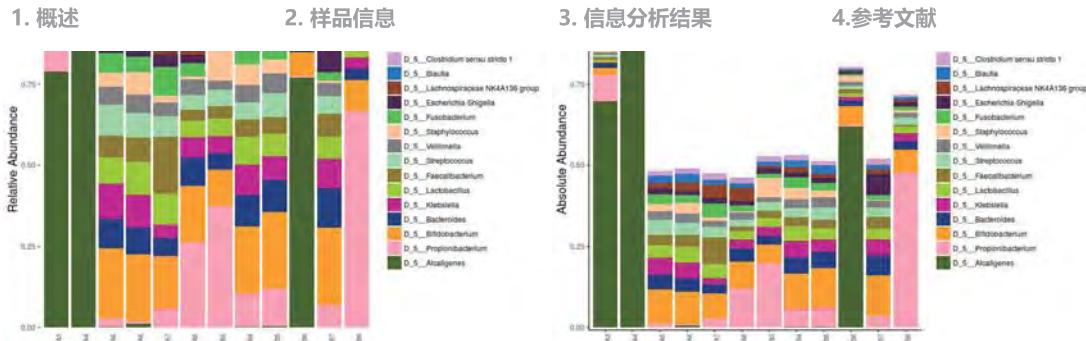
原图路径:02.OTU_taxa/Barplot/ Readme

部分样品物种注释统计表

ID	B7	B8	A8	B3	B5	B4	A7
D_4_Alcaligenaceae	0.0213282937365	0.00814321721709	0.00642702764191	0.00588748364588	0.0116725165293	0.00795435843985	0.00215927854693
D_4_Lachnospiraceae	0.122840172786	0.0682479157242	0.1748379024	0.103412559965	0.150028569096	0.135169235833	0.261780769719
D_4_Propionibacteriaceae	0.036555075594	0.479415756479	0.122568536003	0.197993894461	0.059586972492	0.0543090679686	0.027943604725
D_4_Bifidobacteriaceae	0.125485961123	0.0697343760098	0.0818450688204	0.058547754034	0.124153130357	0.114707334467	0.0790041915407
D_4_Ruminococcaceae	0.111609071274	0.037678536806	0.100159253782	0.0610008722198	0.102930373031	0.0757035492896	0.186333036962
D_4_Enterobacteriaceae	0.124190064795	0.0449815808182	0.0445341826868	0.0488988225033	0.0685658313607	0.075758406934	0.0461069477963
D_4_Veillonellaceae	0.0599352051836	0.0249466813158	0.0406665908315	0.0277474923681	0.0508529915925	0.0458609907291	0.0287057030357
D_4_Bacteroidaceae	0.0629049676026	0.025980740645	0.0402684563758	0.0269297863061	0.0508529915925	0.0515113281036	0.0270544900292
D_4_Lactobacillaceae	0.0343412526998	0.0199702707943	0.0250255943579	0.0279110335805	0.0367316953718	0.0459158483735	0.0452178331005
D_4_Streptococcaceae	0.0292656587473	0.0228785626575	0.0209304970993	0.0299280418666	0.0393437270427	0.0279773986505	0.0322621618189
D_4_Erysipelotrichaceae	0.0184125269978	0.0201641569185	0.022238653168	0.0241495856956	0.0304464941637	0.0208459048768	0.0370887844532
D_4_Staphylococcaceae	0.0036717062635	0.00859561817359	0.00409509725856	0.060673789795	0.0221206432128	0.0340117395359	0.0102883271942
D_4_Bacteroidales S24-7 group	0.0136069114471	0.0109222516642	0.0645546581731	0.0201700828609	0.0213043833156	0.0210653354545	0.00825606503239
D_4_Fusobacteriaceae	0.0139308855292	0.00962967750275	0.01888294847	0.0127017008286	0.0226103991511	0.0323660102035	0.0422964562429
D_4_Clostridiaceae 1	0.0188444924406	0.00859561817359	0.0189398248208	0.0179350196249	0.0132234103339	0.0171155850568	0.0201956052331

原始文件请点击:02.OTU_taxa/Barplot/L5_toptable.txt

全样本在属层次的群落结构分析结果如下图所示：



属层次菌落结构柱状图

注：上：仅展示前15个物种在每个样本内的比例关系；下：即展示前15个物种在每个样本中的比例关系，同时展示这15个物种在该样本中所占比例。每个柱子代表一个样本，不同色块代表相应层次的不同的类，色块的长度为该类序列占该样本的总序列数的比例

原图路径:02.OTU_taxa/Barplot/ Readme

部分样品物种注释统计表

ID	B7	B8	A8	B3	B5	B4	A7
D_5__Alcaligenes	5.39956803456e-05	0.000517029664577	0.000113752701627	0.000654164849542	0.00212227573259	0.00153601404356	0.0
D_5__Propionibacterium	0.036555075594	0.479415756479	0.122568536003	0.197993894461	0.059586972492	0.0543090679686	0.027943604725
D_5__Bifidobacterium	0.123866090713	0.0683771731403	0.0806506654533	0.0577845617096	0.121541098686	0.110209007625	0.0773529785342
D_5__Bacteroides	0.0629049676026	0.025980740645	0.0402684563758	0.0269297863061	0.0508529915925	0.0515113281036	0.0270544900292
D_5__Klebsiella	0.0473002159827	0.024171136819	0.028324422705	0.0280745747928	0.0368133213615	0.0493718799715	0.0195605233075
D_5__Lactobacillus	0.0343412526998	0.0199056420862	0.0249687180071	0.0279110335805	0.0355889315158	0.0456964177958	0.0452178331005
D_5__Faecalibacterium	0.0372030237581	0.00594584114264	0.015242862018	0.021369385085	0.0266100726471	0.0286356903835	0.082687666709
D_5__Streptococcus	0.0282397408207	0.0221676468687	0.0209304970993	0.0266572176188	0.0373847032895	0.0279773986505	0.0307379651975
D_5__Veillonella	0.0215982721382	0.0153816325212	0.022864293027	0.0148277365896	0.0311811280712	0.028306544517	0.0198145560777
D_5__Staphylococcus	0.0036717062635	0.00859561817359	0.00409509725856	0.060673789795	0.0221206432128	0.0340117395359	0.0102883271942
D_5__Fusobacterium	0.0139308855292	0.00962967750275	0.01888294847	0.0127017008286	0.0226103991511	0.0323660102035	0.0422964562429
D_5__Escherichia-Shigella	0.0655507559395	0.0129903703225	0.0116596519167	0.0137374618404	0.0165700759122	0.0156344286576	0.0180363266861
D_5__Lachnospiraceae NK4A136 group	0.0125809935205	0.00794933109287	0.0410647252872	0.00921282163105	0.0115908905395	0.0138789840364	0.0401371776959
D_5__Blautia	0.0131749460043	0.00891876171395	0.0143897167558	0.0119930222416	0.0251408048323	0.0209556201657	0.0163851136797
D_5__Clostridium sensu stricto 1	0.0188444924406	0.00859561817359	0.0189398248208	0.0173898822503	0.0132234103339	0.0171155850568	0.0191794741522

原始文件请点击:02.OTU_taxa/Barplot/L6_toptable.txt

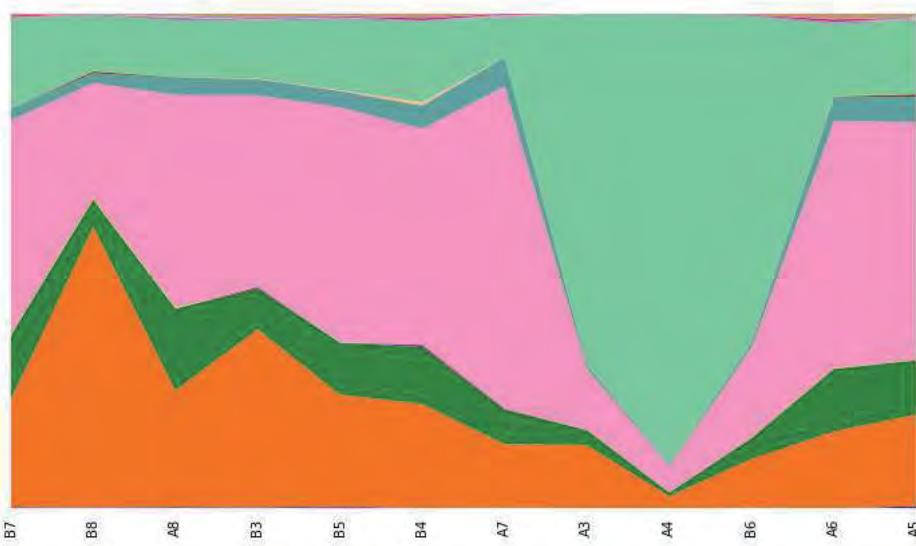
Qiime物种注释Area Charts:

1. 概述

2. 样品信息

3. 信息分析结果

4. 参考文献

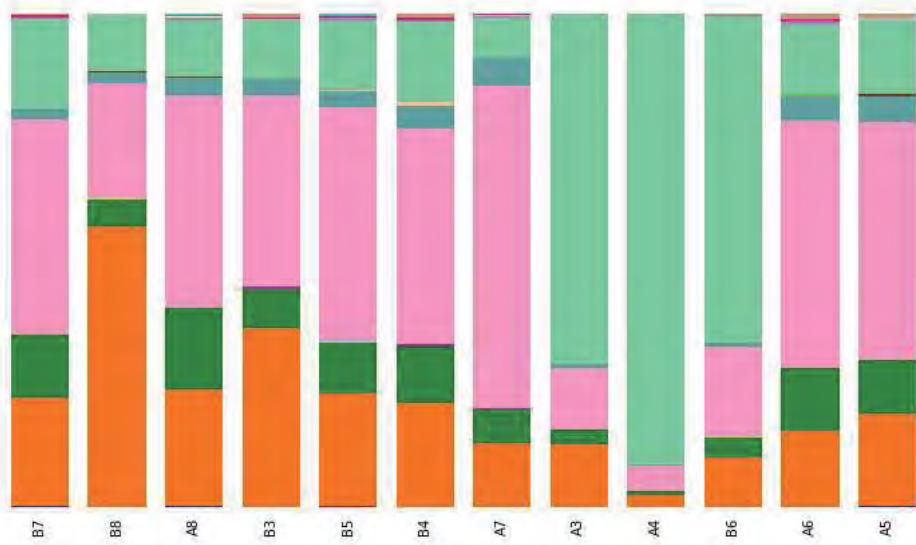
[View Figure \(.pdf\)](#) [View Legend \(.pdf\)](#)[View Table \(.txt\)](#)

Legend	Taxonomy	Total	B7	B8	A8	B3	B5	B4	A7	A3	A4	B6	A6	A5
		%	%	%	%	%	%	%	%	%	%	%	%	%
D_0_Archaea;D_1_Euryarchaeota	0.0%	0.1%	0.1%	0.0%	0.0%	0.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
D_0_Bacteria;D_1_Acidobacteria	0.1%	0.2%	0.1%	0.2%	0.2%	0.2%	0.0%	0.2%	0.0%	0.0%	0.0%	0.1%	0.0%	0.4%
D_0_Bacteria;D_1_Actinobacteria	21.2%	21.8%	56.8%	23.7%	36.0%	22.5%	21.1%	12.8%	12.8%	2.5%	9.9%	15.6%	18.6%	
D_0_Bacteria;D_1_Bacteroidetes	8.6%	12.8%	5.4%	16.4%	8.1%	10.4%	11.4%	6.9%	3.0%	0.7%	4.1%	12.6%	10.9%	

原始文件请点击:02.OTU_taxa/taxa_summary_plots/area_charts.html

Qiime物种注释Bar Charts:

Taxonomy Summary. Current Level:

[View Figure \(.pdf\)](#) [View Legend \(.pdf\)](#)[View Table \(.txt\)](#)

Legend	Taxonomy	Total	B7	B8	A8	B3	B5	B4	A7	A3	A4	B6	A6	A5
		%	%	%	%	%	%	%	%	%	%	%	%	%
D_0_Archaea;D_1_Euryarchaeota	0.0%	0.1%	0.1%	0.0%	0.0%	0.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
D_0_Bacteria;D_1_Acidobacteria	0.1%	0.2%	0.1%	0.2%	0.2%	0.2%	0.0%	0.2%	0.0%	0.0%	0.0%	0.1%	0.0%	0.4%
D_0_Bacteria;D_1_Actinobacteria	21.2%	21.8%	56.8%	23.7%	36.0%	22.5%	21.1%	12.8%	12.8%	2.5%	9.9%	15.6%	18.6%	
D_0_Bacteria;D_1_Bacteroidetes	8.6%	12.8%	5.4%	16.4%	8.1%	10.4%	11.4%	6.9%	3.0%	0.7%	4.1%	12.6%	10.9%	

原始文件请点击:02.OTU_taxa/taxa_summary_plots/bar_charts.html

[1. 概述](#)[2. 样品信息](#)[3. 信息分析结果](#)[4. 参考文献](#)

基于属水平上每个样品的物种丰度信息，选取丰度前50的属，根据各样品的丰度信息对样品和物种进行聚类，绘制热图。便于观察样品对应物种含量得高低。结果如下图所示：



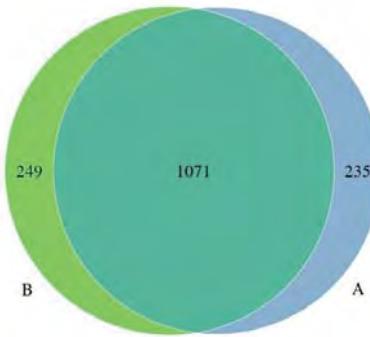
属水平的菌落丰度聚类热图

注：图中展示了物种丰度前50的菌落（除去未知分类的OTU）。横轴是样品名，纵轴是对应的属类。

原图路径:02.OTU_taxa/Heatmap/ Readme

3.2.4 OTU分布Venn 图

Venn图（维恩图），也称文氏图，用于显示元素集合重叠区域的图示。通常，选用相似水平为97%的OTU，展示多个样品中所共有的OTU数目，反映环境样品的相似性及重叠情况。



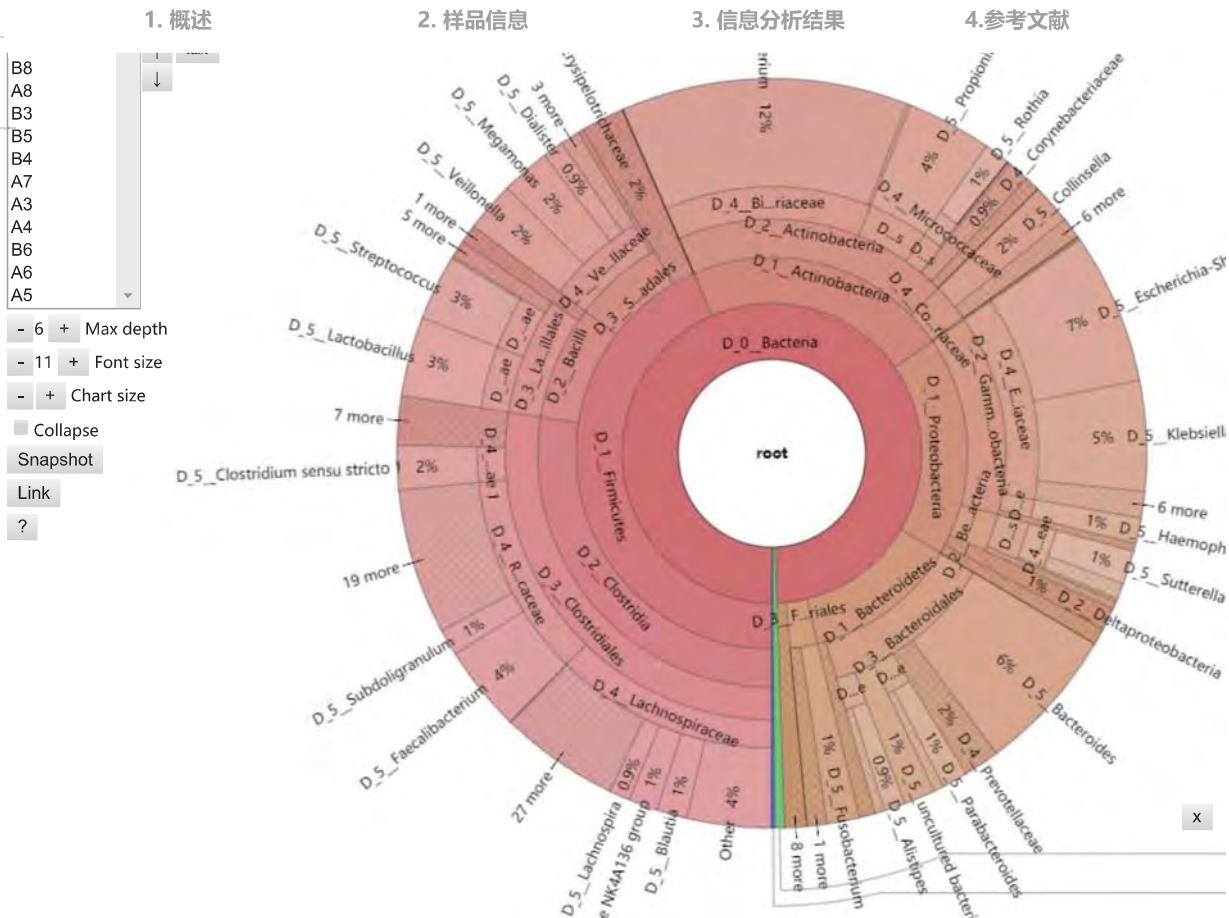
组间OTU韦恩图

注：不同颜色的圈代表不同的样本,如果两个圈重叠表示两个样本含有相同的OTU，重叠区域中的数字表示两个样本共有OTU的个数。

原图路径:02.OTU_taxa/venn/

3.2.5 单样本多级物种组成图

利用KRONA[6]软件对物种注释结果进行可视化展示，圆圈图得各层依次代表物种得分类级别，扇形的大小代表注释物种的比例如。



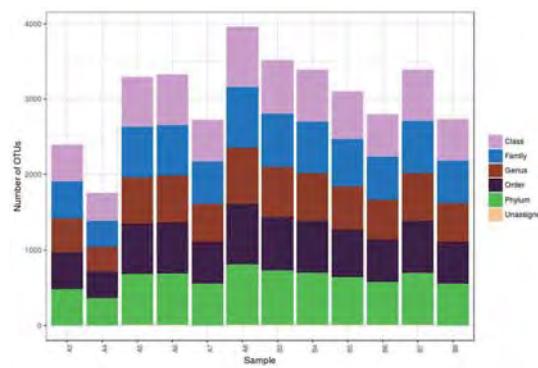
3.2.6 OTU物种注释结果统计

OTU物种注释结果统计表

OTU物种注释结果统计表						
Sample	Phylum	Class	Order	Family	Genus	Unassigned
◀						▶

原始文件请点击:02.OTU taxa/Taxa assignment stat/Taxa assignment stat.txt

注：“Phylum”、“Class”、“Order”、“Family”、“Genus”、“Species”分别对应各样本中能分类至门、纲、目、科、属、种的OTU数，而“Unassigned”指未能归属到任何已知分类单元的OTU数。



OTU物种注释结果统计图

注：横坐标依据样本名排列，纵坐标为各样本中能分类至门、纲、目、科、属、种各分类水平的OTU数。

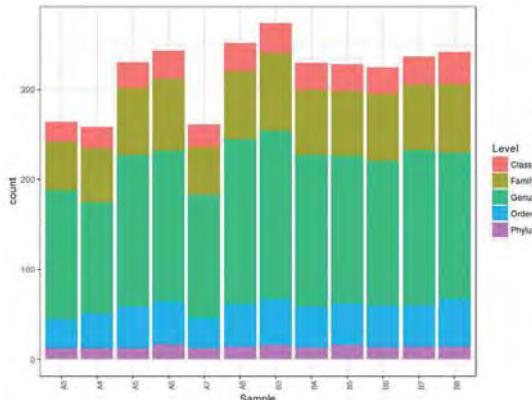
3.2.7 各分类水平的微生物类群统计

各分类水平的微生物类群数统计表

各分类水平的微生物类群数统计表					
Sample	Phylum	Class	Order	Family	Genus
◀					▶

[1. 概述](#)[2. 样品信息](#)[3. 信息分析结果](#)[4. 参考文献](#)

注：“Phylum”、“Class”、“Order”、“Family”、“Genus”、“Species”分别对应各样本在门、纲、目、科、属、种各分类水平的微生物类群数。



各分类水平的微生物类柱形图

注：横坐标依据样本名排列，纵坐标为门、纲、目、科、属、种六个分类水平各自含有的微生物类群数。

物种注释结果目录：[02.OTU_taxa/Readme](#)

3.3 Alpha多样性分析

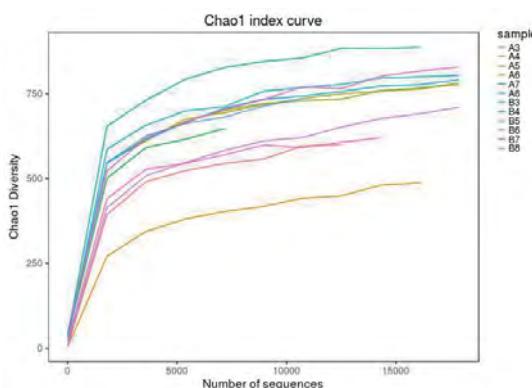
Alpha多样性是指一个特定区域或生态系统内的多样性。多样性指数是反映样本中微生物丰富度和均匀度的综合指标。

3.3.1 稀释曲线分析

采用随机抽样的方法，从样本中随机抽取一定数量的个体，统计出这些个体所代表物种数目，并以个体数与物种数来构建曲线，即稀释性曲线（rarefaction curve）。它可以用来比较测序数量不同的样本物种的丰富度，也可以用来说明样本的取样大小是否合理。稀释曲线是利用已测得16S rRNA序列中已知的各种OTU的相对比例，来计算抽取n个（n小于测得reads序列总数）reads时出现OTU数量的期望值，然后根据一组n值（一般为一组小于总序列数的等差数列）与其相对应的OTU数量的期望值（此处采用chao1算法估计），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。Chao1[7]：用chao1算法估计群落中含OTU数目的指数，在生态学中常用来估计物种总数，由Chao在1984年最早提出。计算公式如下：

$$chao1 = S_{obs} + \frac{F_1(F_1-1)}{2(F_1+1)}$$

其中chao1是估计的OTU数； S_{obs} 是实际观察的OTU数； F_1 是只有一条序列的OTU数目（如“singletons”）； F_2 是只有两条序列的OTU数目（如“doubletons”）。绘制各样本稀释曲线如下图所示：



各样本稀疏曲线

注：图中横坐标为随机抽取序列数，纵坐标为chao1丰富度指数（即估计的OTU数的期望值），每条曲线代表一个样本。上图不同颜色的曲线代表不同的样本，下图不同颜色的曲线代表不同分组。

原图路径：[03.AlphaDiversity/Curve/Readme](#)

由图可知，当测序量（横坐标）增加再多，对OTU数（纵坐标）的增长贡献不大，说明测序量已经足够，增加测序量不能提供更多的新的OTU。

1. 概述

2. 样品信息

3. 信息分析结果

4. 参考文献

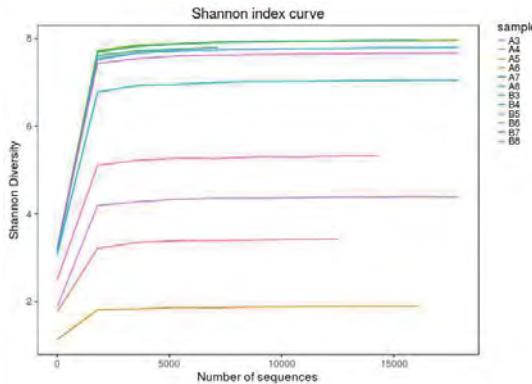
Shannon-Wiener指数[8]是由Shannon, C. E. and Weaver W在1948年提出。反映样品中微生物多样性的指数，利用各样品的测序量在不同测序深度时的微生物多样性指数构建曲线，以此反映各样本在不同测序数量时的微生物多样性。

计算Shannon指数公式如下：

$$H' = -\sum_{i=1}^R p_i \ln(p_i)$$

其中表示样品中属于第i种的个体的比例，如样品总个体数为N，第i种个体数为ni，则 $p_i = n_i/N$ 。各种之间，个体分配越均匀，H值就越大。如果每一个体都属于不同的种，多样性指数就最大；如果每一个体都属于同一种，则其多样性指数就最小。

绘制Shannon曲线如下图所示：



各样本Shannon指数曲线

注：图中横坐标是随机抽取序列数，纵坐标为Shannon指数，每条曲线代表一个样本。上图不同颜色的曲线代表不同的样本，下图不同颜色的曲线代表不同分组。

原图路径:03.AlphaDiversity/Curve/

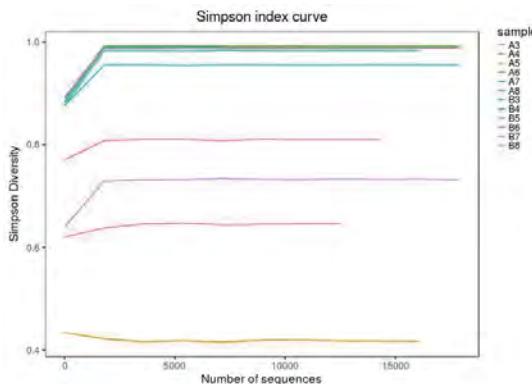
由图可知，测序量（横坐标）增加再多，对样本中物种多样性（纵坐标）的增长贡献不大，说明测序量足够多，当前测序量可以反映样品中绝大多数的微生物物种信息。

3.3.3 Simpson曲线

Simpson指数[9]由Edward H. Simpson于1949年提出，在生态学中常常用来定量描述一个区域的生物多样性。Simpson指数计算公式如下：

$$\text{Simpson index} = 1 - \text{dominance}, \text{ dominance 为 } D = \sum (n_i/n)^2$$

其中 n_i 表示样品中含有i条序列的OTU数，n表示样品总序列数。Simpson指数取值区间是0~1，当样品中只有一个物种时（即所有序列都代表同一个物种），值为0；当样品中各物种均匀分布时，值为1，此时样本中物种多样性最大。Simpson指数值越大，越均匀，群落多样性越高。simpson曲线如下图所示：



simpson指数曲线

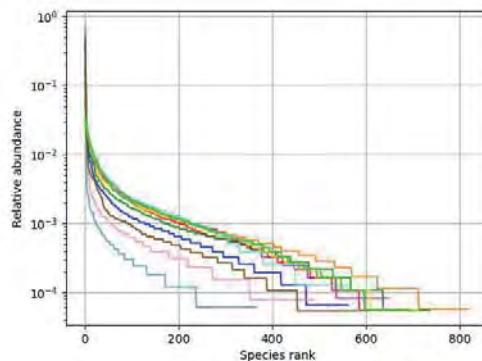
注：图中横坐标是随机抽取序列数，纵坐标为Simpson指数，每条曲线代表一个样本。上图不同颜色的曲线代表不同的样本，下图不同颜色的曲线代表不同分组。

原图路径:03.AlphaDiversity/Curve/

由图可知，测序量（横坐标）增加再多，对样本中物种多样性（纵坐标）的增长贡献不大，说明测序量足够多，当前测序量可以反映样品中绝大多数的微生物物种信息。

[1. 概述](#)[2. 样品信息](#)[3. 信息分析结果](#)[4. 参考文献](#)

将样品中OTUs按丰度(所含有的序列条数)由大到小等级排序,再以OTU等级为横坐标,以每个OTU中所含的序列数(也可用OTU中序列数的相对百分含量)为纵坐标做图,即Rank-Abundance曲线。用于解释样品所含物种的丰富程度和均匀程度,物种的丰富程度由曲线在横轴上的长度来反映,曲线越宽,表示物种的组成越丰富;物种组成的均匀程度由曲线的形状来反映,曲线越平坦,表示物种组成的均匀程度越高。绘制Rank-Abundance曲线下图所示:



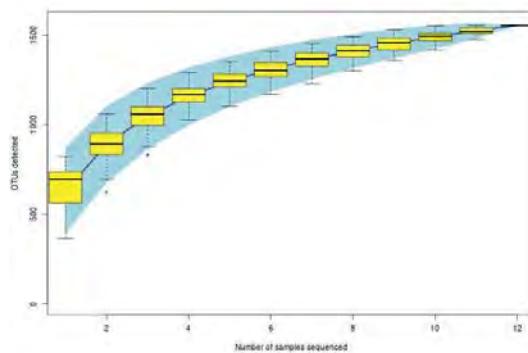
Rank-Abundance曲线

注:上图中每条曲线对应一个样本,横坐标为OTU相对丰度含量按等级降序排列(OTU Rank),纵坐标为该OTU中序列数的相对百分含量(Relative Abundance)。

原图路径:03.AlphaDiversity/Curve/

3.3.5 Specaccum物种累积曲线

物种累积曲线(species accumulation curves)是用于描述随着样品数量的加大而物种增加的状况,是调查样品的物种组成和预测样品中物种丰度的有效工具,被广泛用于样品量是否充分的判断以及物种丰富度(species richness)的估计。因此,通过物种累积曲线不仅可以判断样品量是否充分,在样品量充分的前提下,运用物种累积曲线还可以对物种丰富度进行预测(默认在样品量大于10个分析)。



Specaccum物种累积曲线

注:横坐标:样本量;纵坐标:抽样后的OTU数目,

原图路径:03.AlphaDiversity/Curve/

3.3.6 多样性指数

Alpha多样性指数,包括物种丰富度指数(chao1、ACE)、综合体现物种丰度度和均匀度的多样性指数(shannon、simpson)、以及反映样本低丰度OTU覆盖率的Good's Coverage和OTU数量的observed_species等。其中丰富度衡量单个样本中物种种类,通过估算OTU的数量来衡量。多样性指数衡量群落的异质性。多样性指数分析结果如下表所示。

多样性指数表

多样性指数表						
Sample	observed_species	chao1	shannon	simpson	ace	goods_coverage
4						

Page 0 of 10 ▼

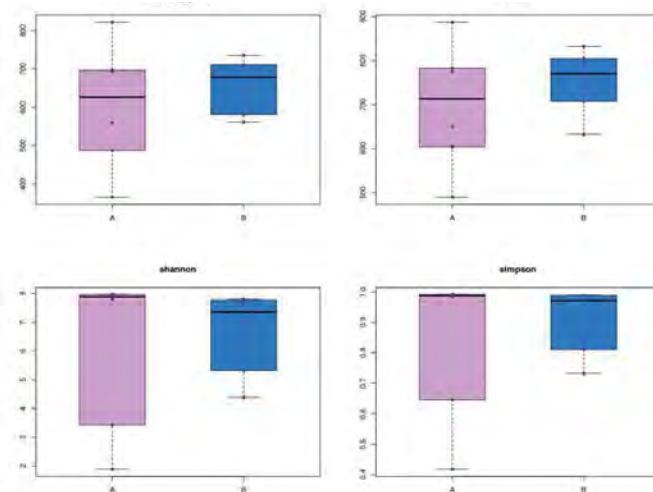
原始文件请点击:03.AlphaDiversity/All_indexs.txt

1. 概述

2. 样品信息

3. 信息分析结果

4. 参考文献



Alpha多样性指数分组比较箱形图

原图路径:03.AlphaDiversity/boxplot/

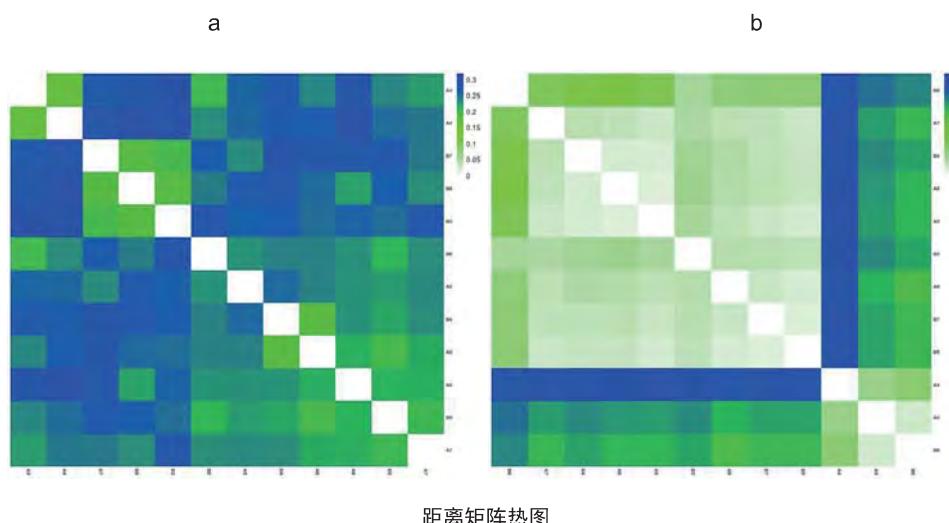
Alpha多样性结果目录 : 03.AlphaDiversity/ Readme

3.4 Beta多样性分析

Beta Diversity是对不同样本/不同组间样本的微生物群落构成进行比较分析。可以用OTUs的丰度信息进行样本间距离计算，也可以用OTUs之间的系统发生关系进行计算。UniFrac[10]是用于比较生物群落的一种距离度量，它在计算过程中包含了遗传距离 (phylogenetic distances) 的信息，根据构建的进化树的长度计量两个不同环境样品之间的差异。UniFrac 分析分为weighted unifrac 和unweighted unifrac 两种度量方法，两者之间差异在于是否计入不同环境样品的序列相对丰度。weighted unifrac 算法在计算树的长度时将序列的丰度信息进行加权计算，因此unweighted unifrac 可以检测样品间变化的存在，而weighted unifrac 可以更进一步定量的检测样品间不同谱系上发生的变异。基于OTU丰度信息表计算Unifrac距离 (Unweighted Unifrac) 和Weighted Unifrac距离。距离矩阵可用于后续的Beta多样性分析和可视化统计分析，从中发现不同样品 (组) 间的差异。

3.4.1 距离矩阵热图

Unweighted Unifrac距离矩阵热图如图a，Weighted Unifrac距离矩阵热图如图b。



注：图a是基于Unweighted unifrac距离，图b是基于Weighted unifrac距离。图样品两两之间的相异系数用颜色梯度进行表示，根据颜色得差异可判断样品间距离得大小

原图路径:04.BetaDiversity/Heatmap/

3.4.2 PCoA分析

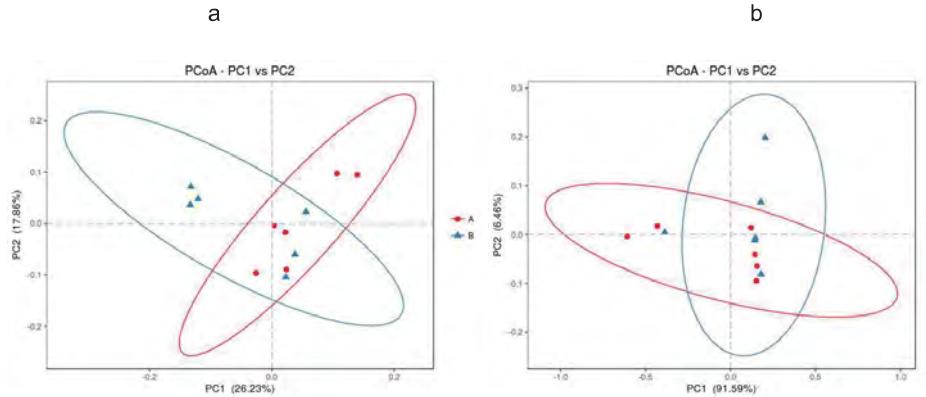
主坐标分析 (Principal Coordinate Analysis , PCoA) 将多维数据进行降维，是一种研究数据相似性或差异性的可视化方法，基于距离矩阵来寻找主坐标，通过一系列的特征值和特征向量进行排序后，选择主要排在前几位的特征值，有效地找出数据中

1. 概述

2. 样品信息

3. 信息分析结果

4. 参考文献



基于Unweight unifrac距离 PCoA图

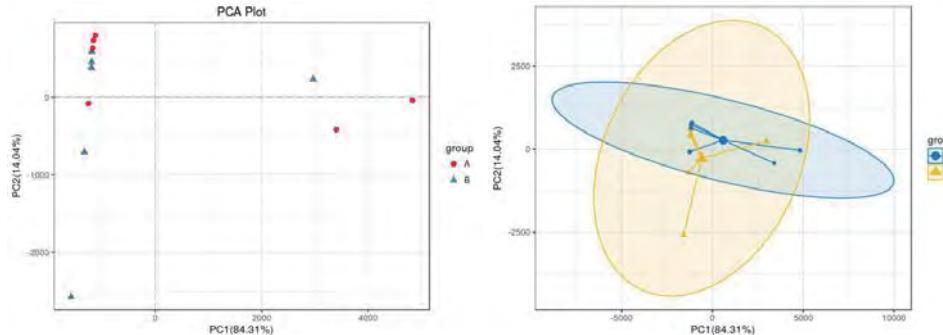
注：图22a是基于Unweighted unifrac距离，图22b是基于Weighted unifrac距离。图中每个点代表一个样本，横轴表示主坐标分析中的第一主坐标，纵轴表示第二主坐标。图中点之间的距离越近，表示对应的样品间组成越相似。

原图路径:04.BetaDiversity/PCoA/ Readme

说明：样本数量越多，该分析意义越大；反之样本数量过少，会产生个体差异，导致PCoA分析成图后形成较大距离的分开，建议多组样品时，每组不少于5个，不分组时样品不少于10个。

3.4.3 PCA分析

主成分分析PCA (Principal component analysis) 是一种研究数据相似性或差异性的可视化方法，通过一系列的特征值和特征向量进行排序后，选择主要的前几位特征值，采取降维的思想，PCA 可以找到距离矩阵中最主要的坐标，结果是数据矩阵的一个旋转，它没有改变样品点之间的相互位置关系，只是改变了坐标系统。PCA 可以观察个体或群体间的差异。下图每一个点代表一个样本，相同颜色的点来自同一个分组，两点之间距离越近表明两者的群落构成差异越小。



PCA图

注：图中每个点代表一个样本，x轴表示主坐标分析中的第一主成分，y轴表示第二主成分，z轴代表第三主成分，坐标轴名称中的数值代表相应主坐标对于整体的贡献度。

原图路径:04.BetaDiversity/PCA/ Readme

3.4.4 NMDS分析

NMDS (Non-metric Multidimensional scaling) 非度量多维尺度分析是一种将多维空间的研究对象（样本或变量）简化到低维空间进行定位、分析和归类，同时又保留对象间原始关系的数据分析方法。适用于无法获得研究对象间精确的相似性或相异性数据，仅能得到他们之间等级关系数据的情形。其基本特征是将对象间的相似性或相异性数据看成点间距离的单调函数，在保持原始数据次序关系的基础上，用新的相同次序的数据列替换原始数据进行度量型多维尺度分析。其特点是根据样品中包含的物种信息，以点的形式反映在多维空间上，而对不同样品间的差异程度，则是通过点与点间的距离体现的，最终获得样品的空间定位点图。

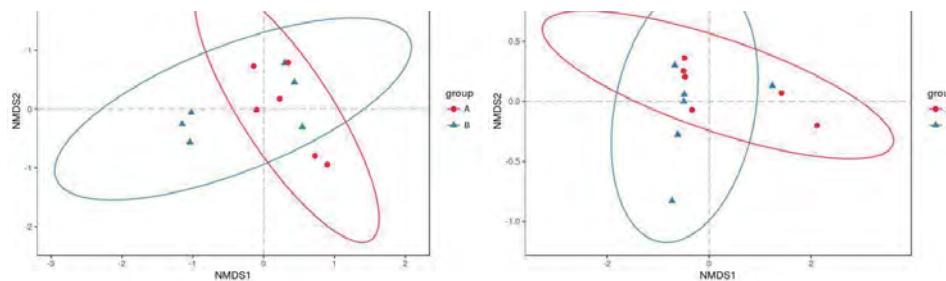


1. 概述

2. 样品信息

3. 信息分析结果

4. 参考文献



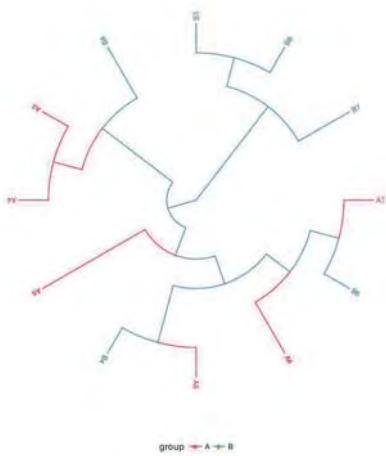
NMDS图

注：图24a是基于Unweighted unifrac距离，图24b是基于Weighted unifrac距离。下图每一个点代表一个样本，相同颜色的点来自同一个分组，两点之间距离越近表明两者的群落构成差异越小。

原图路径:04.BetaDiversity/NMDS/ Readme

3.4.5 UPGMA聚类树

树状图可以通过树枝结构比较直观地描述和比较多个样品间的相似性和差异关系。基于距离矩阵，使用UPGMA (Unweighted Pair Group Method with Arithmetic mean) 聚类方法进行分析构建进化树，结果如下图所示。

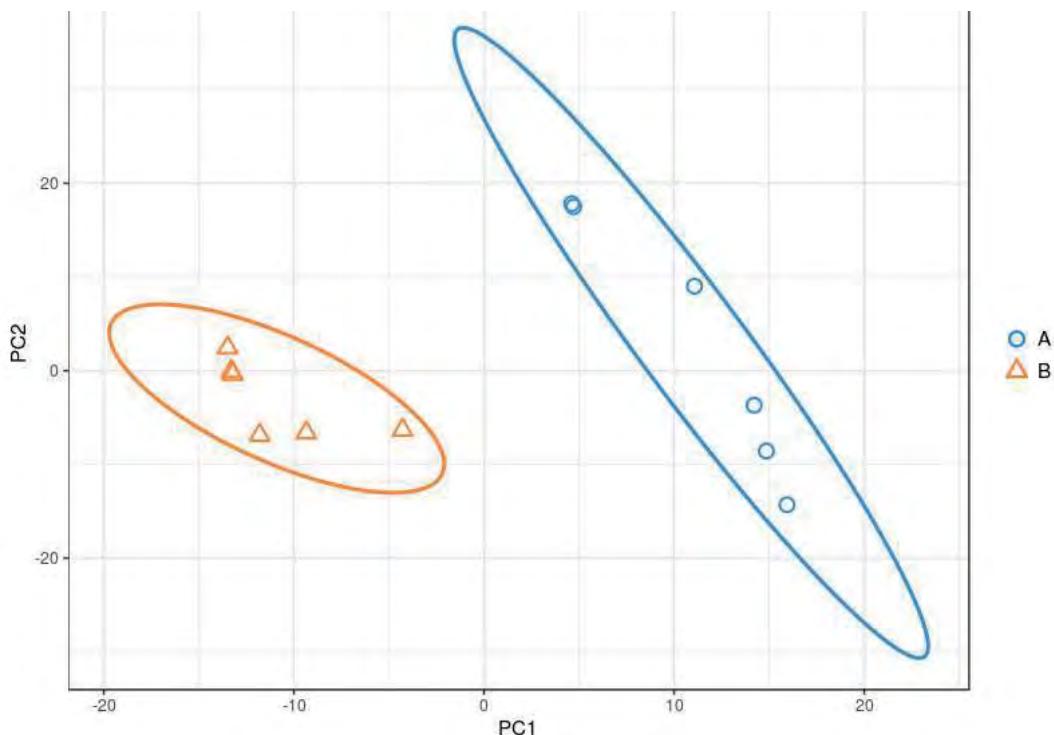


1. 概述

2. 样品信息

3. 信息分析结果

4. 参考文献



PLS-DA判别分析图

原图路径:04.BetaDiversity/plsda/

Beta多样性结果目录：04.BetaDiversity/ Readme

3.5 差异分析

3.5.1 LEfSe分析

LEfSe分析即LDA Effect Size分析，是一种用于发现高微生物标识和揭示基因组特征的分析，使用non-parametric factorial Kruskal-Wallis (KW) sum-rank test (非参数因子克鲁斯卡尔—沃利斯和秩检验)检测具有显著丰度差异特征，并找到与丰度有显著性差异的类群。最后，LEfSe采用线性判别分析 (LDA) 来估算每个组分 (物种) 丰度对差异效果影响的大小。能够在组与组之间寻找具有统计学差异的生物标识 (Biomarker)，即组间差异显著的物种。

1.LDA值分布柱状图：

通过LDA分析 (线性回归分析) 统计不同组别中有显著作用微生物类群的LDA分值，展示了LDA 分值大于设定阈值的物种，即具有统计学差异的biomarker。柱状图的长度代表显著差异物种的影响大小；

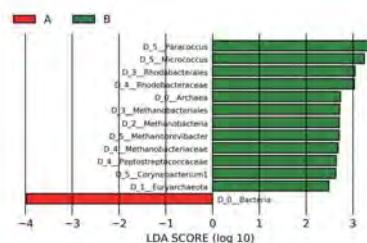


图28. LDA分值分布图

2.biomarker在不同组各样本中的丰度比较图：

将biomarker丰度最高的样本的丰度设定为1，其他样品中该 biomarker 的丰度为相对于丰度最高样品的相对值。

1. 概述

2. 样品信息

3. 信息分析结果

4. 参考文献

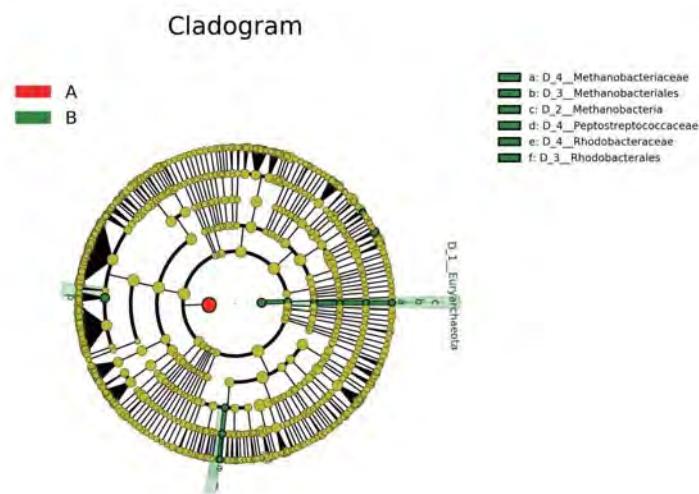


biomarker相对丰度图

3. 进化分支图：

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。

着色原则：无显著差异的物种统一着色为黄色，差异物种 Biomarker 跟随组进行着色，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，其它圈颜色意义类同。图中英文字母表示的物种名称在右侧图例中进行展示。



物种聚类图

注：图同心圆由内到外分别代表不同的分类水平，同心圆上的小圆圈代表个分类的名称，圆圈大小代表丰度的高低，黄色表示各组间无显著差异，其它颜色为同色组的代表物种。

LEfSe结果路径:05.Diff_analysis/Lefse/groupall/ Readme

3.5.2 MetaStats分析

使用Metastats (White et al., 2009)的统计学算法，对门和属水平的各个分类单元在样本（组）之间的序列量（即绝对丰度）差异进行两两比较检验。

组间Metastats分析结果统计表

level	B-A
class	0
family	0
genus	0
order	0
phylum	0

原始文件请点击:05.Diff_analysis/Metastats/stat.txt

门水平组间具有显著差异的分类单元丰度比较箱线图

原图路径:05.Diff_analysis/Metastats/ Readme

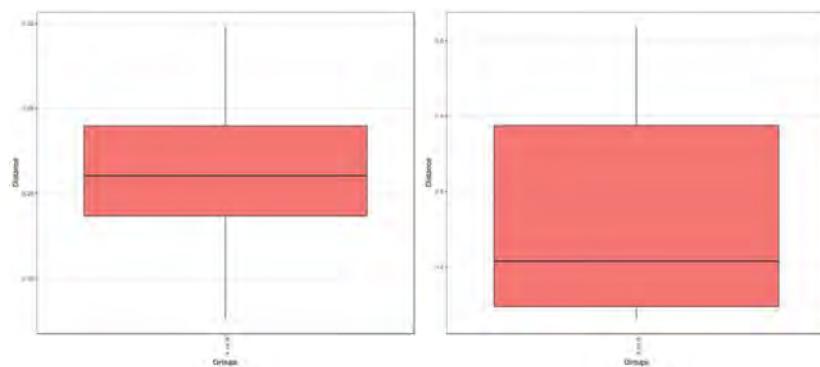
3.5.3 基于距离的多组比较箱线图

1. 概述

2. 样品信息

3. 信息分析结果

4. 参考文献



基Unweighted Unifrac距离的多组比较箱线图

基Weighted Unifrac距离的多组比较箱线图

原图路径:05.Diff_analysis/distance_boxplot/

3.5.4 Adonis/PERMANOVA多元方差分析和置换检验

基于置换的PERMANOVA (Permutational multivariate analysis of variance) 分析(McArdle and Anderson, 2001)借鉴了ANOVA方差分析多组间差异的统计检验思路，通过对距离矩阵进行置换检验，将原始样本的组间差异与多次随机置换打乱后获得的距离矩阵进行比较，从而评价原始样本组间差异的大小及其统计学显著性。而Adonis分析则可视为PERMANOVA分析的扩展版本，能够对样本检测指标（连续型变量）进行类似的统计检验。

Adonis/PERMANOVA分析将给出一个P值，以表征组间差异是否具有统计学意义。值得注意的是，与ANOVA一样，在多组别的相互比较时，只要其中一组与其它各组中的任意一组具有统计学差异，Adonis/PERMANOVA即给出具有显著性意义的P值。

Adonis分析结果

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
map	1	0.08312	0.083121	0.88943	0.08168	0.163
Residuals	10	0.93455	0.093455		0.91832	
Total	11	1.01767			1.00000	

3.5.5 ANOSIM相似度分析

ANOSIM (Analysis of similarities) 分析也是一种非参数检验方法，与NMDS分析类似，通过对样本距离等级排序来判断样本组内和组间差异的大小，并通过置换检验评价原始样本组间差异的统计学显著性(Clarke, 1993; Warton et al., 2012)。

method name	ANOSIM
test statistic name	R
sample size	12
number of groups	2
test statistic	-0.04444444444444287
p-value	0.5779999999999996
number of permutations	999

注：R值为ANOSIM 的统计量，数值介于-1和1之间，表征组间差异与组内差异的差值大小。R值越接近1，表明组间差异越大，同时组内差异越小，分组效果越好；如果R=0，表明样本的分组效果等同于随机分配，各样本分组之间不具有可观测的统计学差异；如果R为负值，则表明组内差异超过了组间差异的大小，预示分组效果较差。P值则反映了ANOSIM 分析结果的统计学显著性，P值越小，表明各样本分组之间的差异显著性越高。

3.5.6 随机森林分析

随机森林 (Random Forests) (Breiman, 2001)是一种基于决策树 (Decision tree) 的经典高效的机器学习算法，属于非线性分类器 (Non-linear classifier)，能够深入挖掘变量之间复杂的非线性相互依赖关系，对于经常呈现离散、不连续分布的微生物群落数据而言尤其适用，近几年已有研究证明这一算法能够对微生物群落样本进行有效且准确的分类(Yatsunenko et al., 2012)。

衡量随机森林模型效果的关键指标

Model	Random Forest
Error type	out-of-bag
Estimated error	0.66667

1. 概述

2. 样品信息

3. 信息分析结果

4. 参考文献

Number of trees

500

注：第一行表明分析方法为随机森林；第二行为衡量模型效果和分类准确率的具体方法，一般选用10倍交叉验证；第三行表明建模估计的预报错误率（Estimated error）；第四行为基线错误率，第五行对应两者的比值。表格最后一行表明建模使用的决策树数量，这一参数值越高，建模效果越好。

各分组中被错误分类的样本数列表

True\Predicted	A	B	Class error
A	3	3	0.5
B	5	1	0.833333333333333

注：上表从第二行开始，每一行代表一种真实的样本分组。第二列开始的每一列则反映了对应分组的样本被正确/错误分类到各分组中的样本数，据此可以计算每组的预报错误率大小，具体数值在最后一列中列出。

差异分析结果目录：[05.Diff_analysis/](#) Readme

4. 参考文献

- [1] Tanja Magoč, Steven L. Salzberg. (2011).FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* (2011) 27 (21): 2957-2963.
- [2] Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *NatureMethods*, 7: 335–336.
- [3] Edgar RC1, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 2011 Aug 15;27(16):2194-200.
- [4] Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST, *Bioinformatics*, 26, 2460–2461.
- [5] Wang, Q., Garrity, G. M., Tiedje, J. M., Cole, J. R. (2007). Naïve bayesian classifier for rapid assignment of rrna sequences into the new bacterial taxonomy. *Applied and environmental microbiology*, 73:5261-5267.
- [6] Ondov BD, Bergman NH, and Phillippy AM. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*. 2011 Sep 30; 12(1):385.
- [7] Chao, A. 1984. Non-parametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics* 11, 265-270.
- [8] Shannon, C. E. and Weaver W. (1948) A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–423 and 623–656.
- [9] Simpson, E. H. (1949). "Measurement of diversity". *Nature* 163: 688.
- [10] Lozupone, C., Knight, R. (2005).UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and environmental microbiology*, 71:8228-8235.