

# CMAP应用培训





## 微生物组数据分析云平台

### 分析总览

分析总览

分析前准备 <

Alpha多样性分析 <

Beta多样性分析 <

群落组成分析 <

环境因子分析 <

差异物种分析 <

>> LEfse分析

#### 分析前准备

数据上传  
数据筛选

#### $\alpha$ 多样性分析

Alpha多样性指数  
Alpha多样性组间比较  
稀释曲线

#### $\beta$ 多样性分析

PCA分析  
PCoA/NMDS分析  
NMDS分析  
Adonis分析  
Anosim分析  
样本层级聚类分析

#### 群落组成分析


Bar图

#### 环境因子分析

Maaslin分析

#### 差异物种分析

LEfse分析



CMAP是本题开发一个微生物组组学数据分析云平台，拥有功能强大且设计良好的交互式界面，综合了主流的微生物组学分析模块构建多样化的分析流程。

目的是为未经生物信息学系统培训的研究人员和临床医生能参考目前主流分析方法轻松挖掘微生物组学数据，包括数据预处理，统计分析和可视化。

## 1. MicrobiotaProcess

本课题组致力于微生物组学分析和可视化工具的开发，其中R包MicrobiotaProcess提供了功能强大的微生物组学数据的分析框架。CMAP借此实现了微生物组学的前期数据过滤处理，alpha与beta多样性分析，群落组成分析以及差异物种分析功能。

## 2. treeio + tidytree + ggtree + ggtreeExtra

CMAP借助上述4个R包实现了对微生物组学的系统发育进行更加方便灵活的探索以及结果展示。

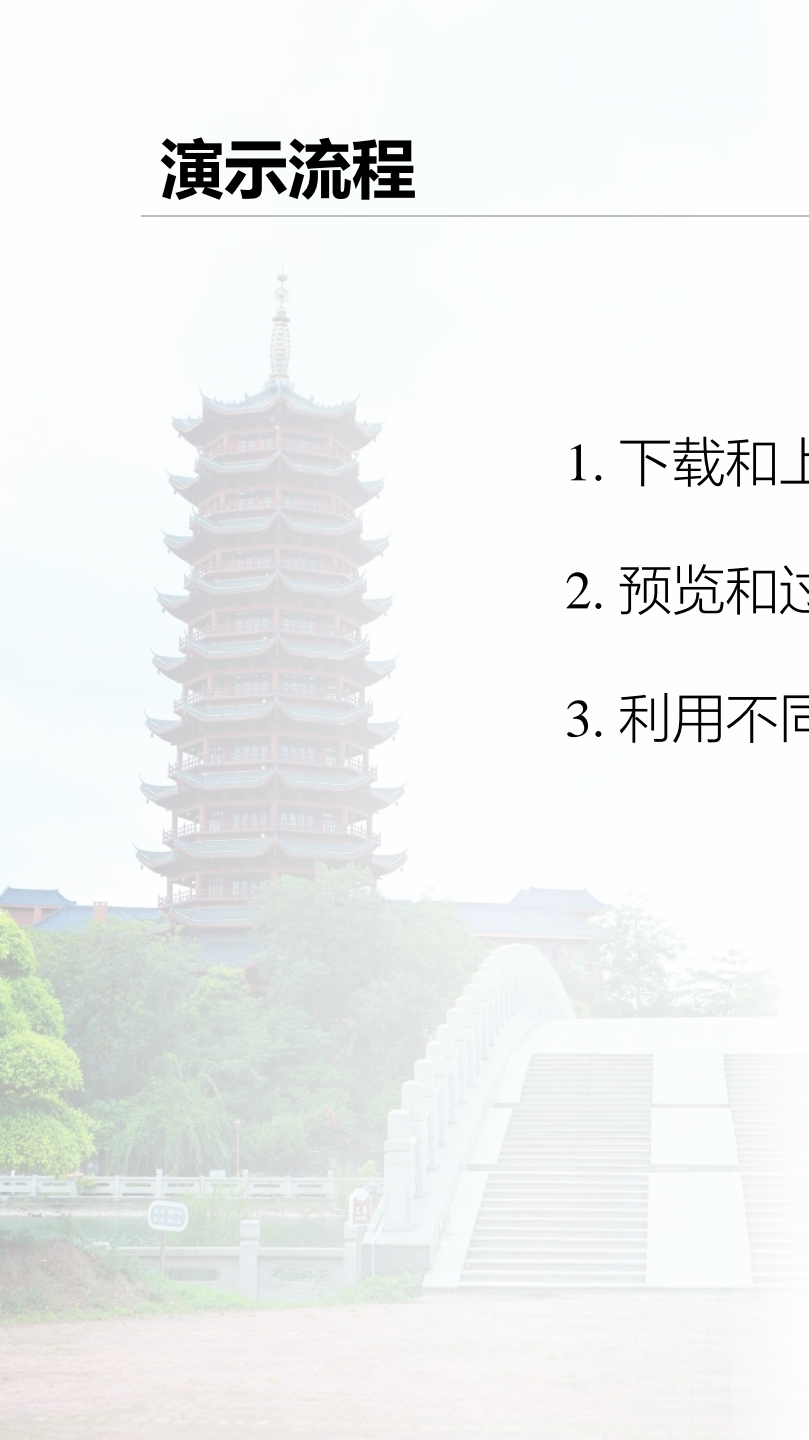
## 3. MaAsLin2

环境变量与微生物数据的关联分析R包

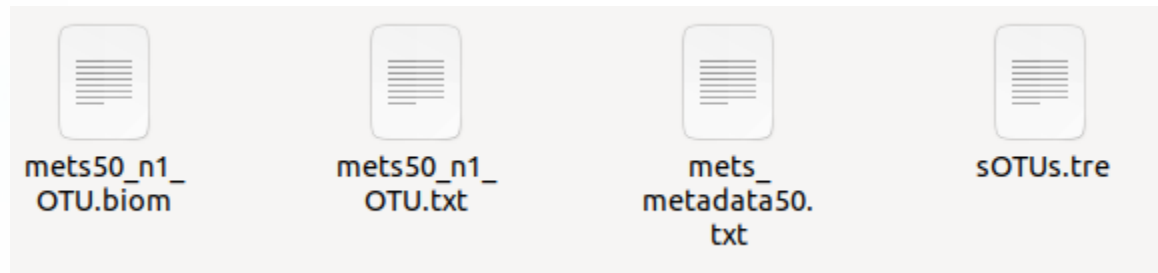
## 4. shiny + shinyWidgets + shinydashboard

构建ShinyApp提供友好的交互式页面



- 
1. 下载和上传示例数据
  2. 预览和过滤数据
  3. 利用不同的分析模块挖掘微生物组学数据

数据来源于GGMP（广东省肠道微生物组项目，Guangdong Gut Microbiome Project），从广东省随机抽样了14个地区，收集了7000例成年人的粪便，16S测序其V4区，并以问卷形式收集个人信息等数据。这里从中选择了25例代谢综合征患者，以及25例对照作为演示。



分析总览

分析前准备 <

» 数据上传

» 数据筛选

Alpha多样性分析 <

Beta多样性分析 <

群落组成分析 <

环境因子分析 <

差异物种分析 <

Upload your data

Abundance profile (.txt or .biom):

Browse...

No file selected

Metadata file (.txt):

Browse...


No file selected

(Optional) Phylogenetic tree (.tre or .nwk):

Browse...

No file selected

Submit

 Example Data

点击（Example Data），下载示例数据



分析总览

分析前准备

» 数据上传

» 数据筛选

Alpha多样性分析

Beta多样性分析

群落组成分析

环境因子分析

差异物种分析

## Upload your data

**Abundance profile (.txt or .biom):**

Browse...

mets50\_n1\_OTU.biom

**Metadata file (.txt):**

Browse...

mets\_metadata50.txt

**(Optional) Phylogenetic tree (.tre or .nwk):**

Browse...

sOTUs.tre

Submit

📄 Example Data

开始分析前，您需要上传微生物多样性分析过程中需要用到几个的文件，主要包括包含：

1. 物种注释的OTU/ASV表(必须)
2. 包含分组信息和/或环境因子信息的metadata（必须）
3. 进化树文件(可选)。

# 格式要求

1、OTU/ASV表为biom格式或txt格式的，OTU/ASV表的以样本名作为第一行(列名)，每一行为OTU/ASV的ID(行名)，最后一列的列名为taxonomy包含对应行OTU的注释信息，列与列之间必须以tab键分隔。（如图）

#OTU ID	G440606625	G441502228	G440305319	taxonomy
Seq4034	0	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__s__
Seq12622	214	14	2	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides; s__plebeius
Seq13264	122	273	255	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Faecalibacterium; s__prausnitzii
Seq8703	1	9	0	k__Bacteria; p__Actinobacteria; c__Coriobacteriia; o__Coriobacteriales; f__Coriobacteriaceae; g__Collinsella; s__aerofaciens
Seq4307	110	13	26	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Ruminococcus; s__bromii

## 格式要求

2、含分组信息和/或环境因子信息的metadata，制表符分割的文本格式文件  
第一列列名必须为#SampleID，且样本名不能重复。

#SampleID	district	age	gender	MetS
G440104002	G440104	78	m	y
G440104226	G440104	42	f	n
G440104592	G440104	55	f	y
G440205073	G440205	47	f	n
G440205303	G440205	67	f	y

3、进化树文件（可选），该文件一般以.tre、.nwk为后缀名，包含了物种的进化关系。若缺少该文件将不能进行涉及物种进化关系的分析（如计算 $\alpha$ 多样性的PD指数、 $\beta$ 多样性的Unifrac距离等）

Upload your data



Text summary



OTU number:	2076
OTU annotation:	Kingdom;Phylum;Class;Order;Family;Genus;Species
Total read counts:	5e+05
Phylogenetic tree uploaded:	FALSE
Number of samples in metadata:	50

Library size overview



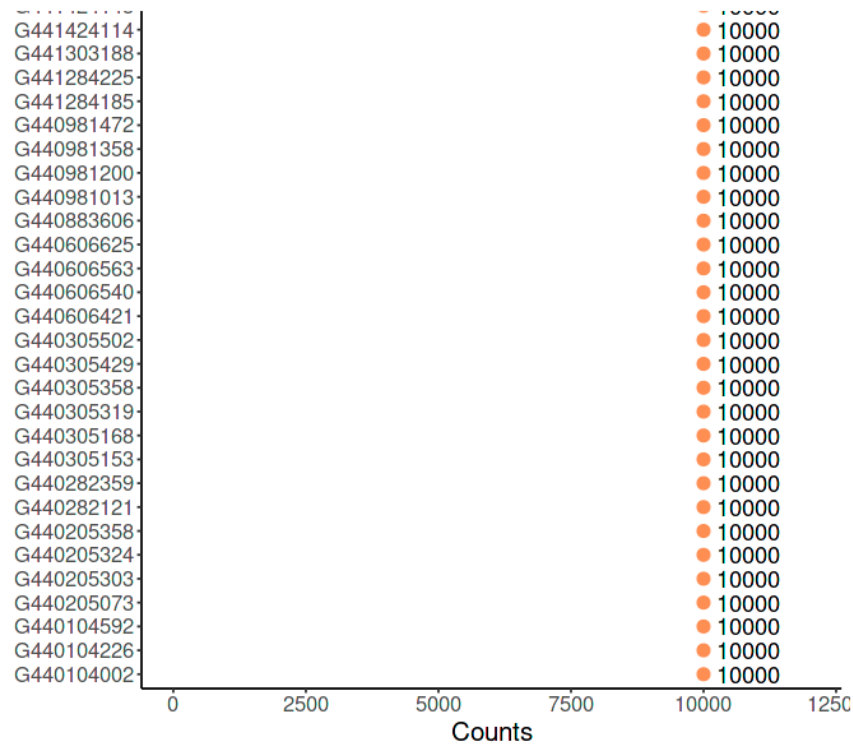
数据上传完毕，查看文本信息



Upload your data

Text summary

Library size overview



查看各样本的reads的总数



分析总览

分析前准备

» 数据上传

» 数据筛选

Alpha多样性分析

Beta多样性分析

群落组成分析

环境因子分析

差异物种分析

Feature filter

Minimum counts of features:

5

Prevalence in samples (%):

5

Data rarefying

☒ Minimum ☐ Default ☐ Do not

Submit

Taxonomy filter

Sample filter

1. 只保留至少在3 ( $0.05 * 50 \approx 3$ ) 个样本中序列数目都大于或等于5的特征。
2. 数据不抽平

分析总览

分析前准备

» 数据上传

» 数据筛选

Alpha多样性分析

Beta多样性分析

群落组成分析

环境因子分析

差异物种分析

Feature filter

+

Taxonomy filter

-

Kingdom

Nothing selected

Phylum

Nothing selected

Class

c\_\_Halobacteria, c\_\_Deinococci

Order

Nothing selected

Family

Nothing selected

Genus

g\_\_Mobiluncus

Species

Nothing selected

Submit

Sample filter

+

分析总览

分析前准备

» 数据上传

» 数据筛选

Alpha多样性分析

Beta多样性分析

群落组成分析

环境因子分析

差异物种分析

进化分析

功能预测

Feature filter

+

Taxonomy filter

+

Sample filter

-

Group:

occupation

Choose between:

j  
k  
l  
c  
a

You have selected:

j  
k

Submit

# Alpha多样性分析

分析总览

分析前准备

Alpha多样性分析

» a多样性指数比较

» 稀释曲线

Beta多样性分析

群落组成分析

环境因子分析

差异物种分析

## Alpha-diversity Index

Alpha-diversity Index:

Observe, Chao1, ACE

Statistical method:

T-test/ANOVA

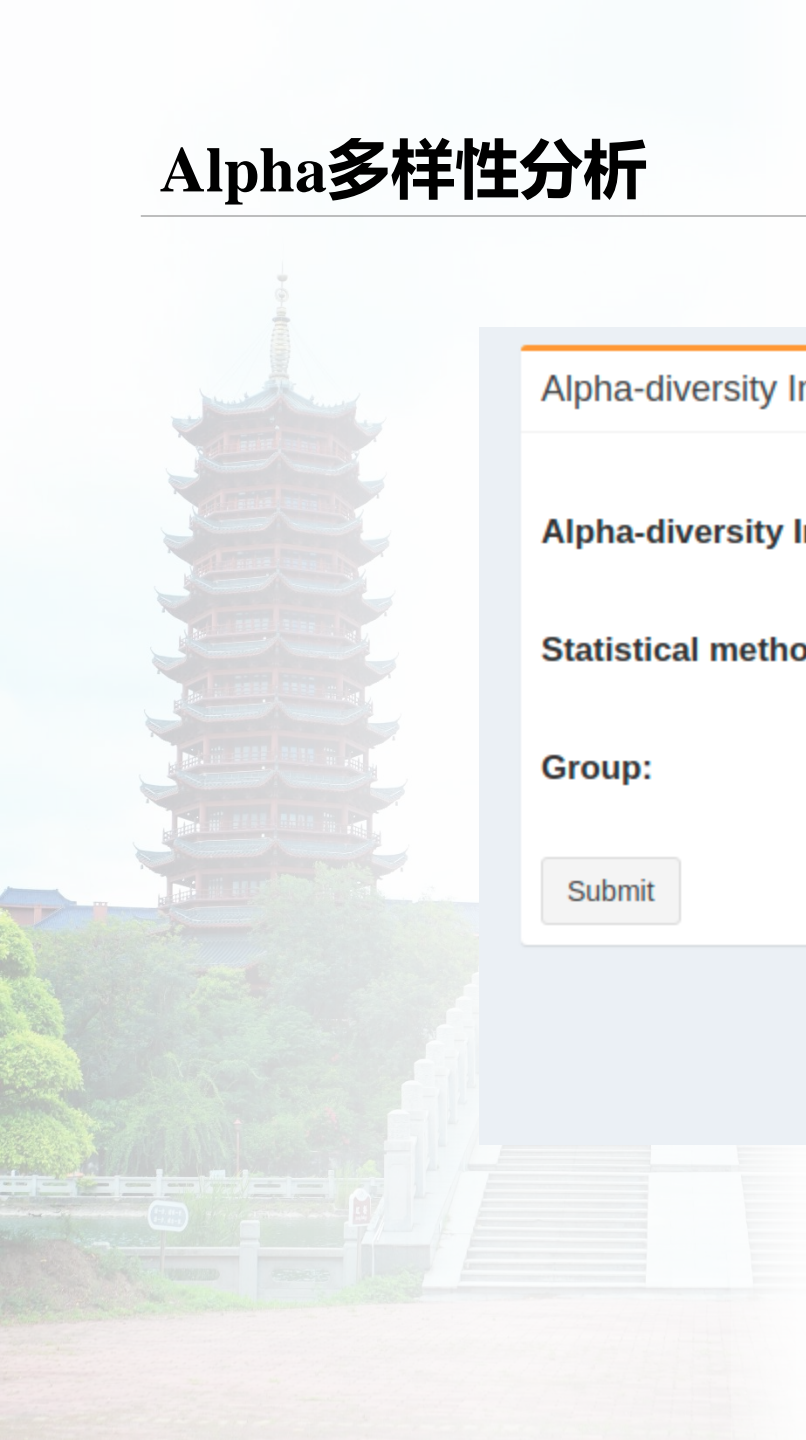
Group:

MetS

Submit

Alpha多样性是指一个特定区域或者生态系统内的多样性，常用的度量标准有chao、shannon、ace、simpson、observe等，在此功能模块，您可以通过观察各种指数值进而得到物种的多样性等信息，还可以对样本分组，运用统计学方法，检测每两组之间的指数值是否具有显著性差异。

# Alpha多样性分析



Alpha-diversity Index

Alpha-diversity Index: Observe, Chao1, ACE

Statistical method:

Group:

Submit

Select All	Deselect All
	<input checked="" type="checkbox"/>
	<input checked="" type="checkbox"/>
	<input checked="" type="checkbox"/>
	<input type="checkbox"/>
	<input type="checkbox"/>
	<input type="checkbox"/>
	<input type="checkbox"/>
	<input type="checkbox"/>

1. 选择Alpha多样性指数
2. 组间差异检验方法
3. 选择分组
4. 点击Submit，分析完成



# Alpha多样性分析

## Alpha-diversity Index

Alpha-diversity Index:

Observe, Chao1, ACE

Statistical method:

Wilcoxon-test/Kruskal-Wallis

T-test/ANOVA

Group:

Wilcoxon-test/Kruskal-Wallis

Submit

1. 选择Alpha多样性指数
2. 组间差异检验方法
3. 选择分组
4. 点击Submit，分析完成

# Alpha多样性分析

Alpha-diversity Index

Alpha-diversity Index: Observe, Chao1, ACE

Statistical method: Wilcoxon-test/Kruskal-Wallis

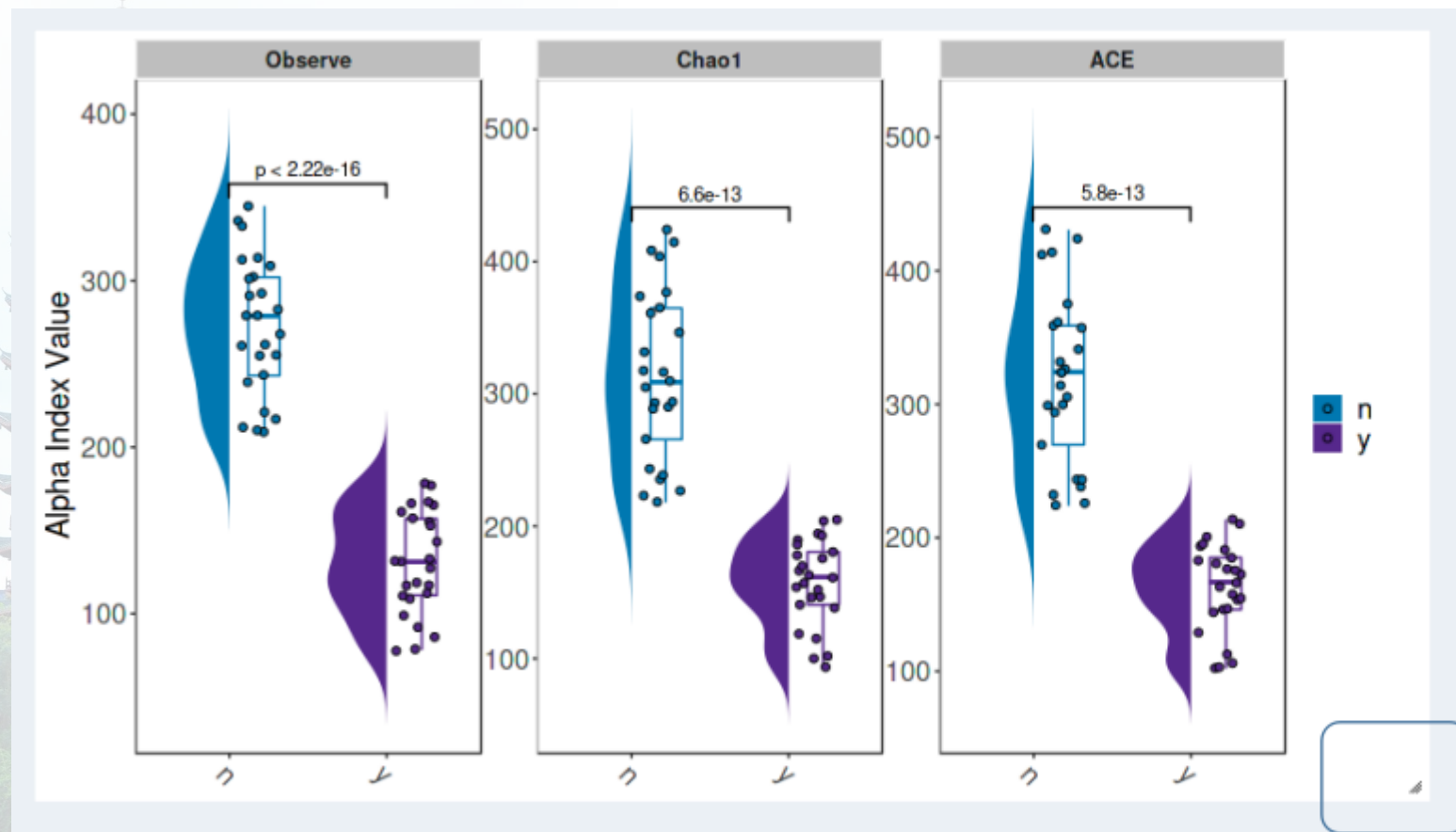
Group: MetS

Submit

- district
- age
- gender
- occupation
- smoke\_now
- height
- weight
- BMI
- waist
- Bristol\_stool\_type
- MetS**
- Obesity
- Overweight

1. 选择Alpha多样性指数
2. 组间差异检验方法
3. 选择分组
4. 点击Submit, 分析完成

# Alpha多样性分析



1. 选择Alpha多样性指数
2. 组间差异检验方法
3. 选择分组
4. 点击Submit, 分析完成

拖拽改变绘图尺寸

# 稀释曲线

分析总览

分析前准备

Alpha多样性分析

» a多样性指数比较

» 稀释曲线

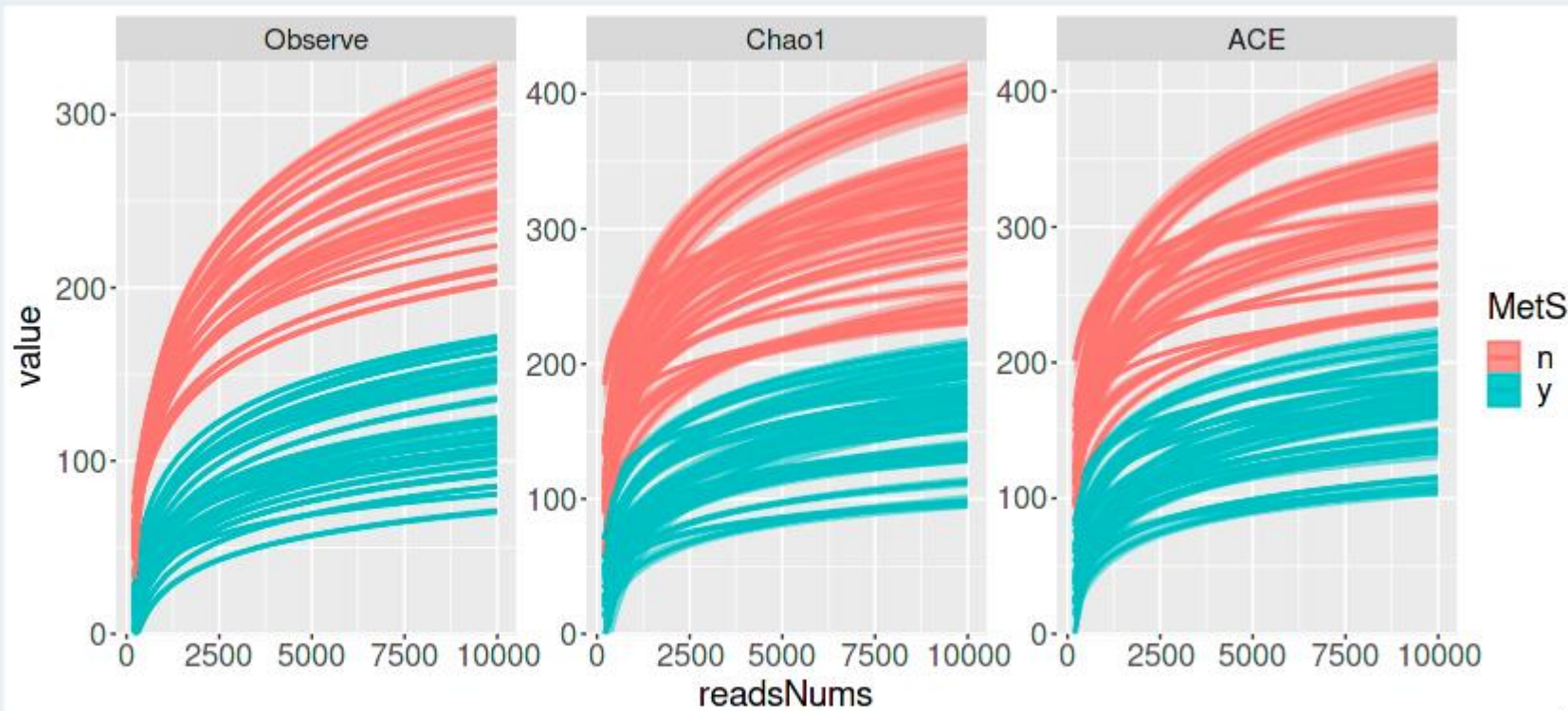
Beta多样性分析

群落组成分析

环境因子分析

差异物种分析

Rarefaction Curve





分析总览

分析前准备 <

Alpha多样性分析 <

Beta多样性分析 <

» PCA分析

» PCoA分析

» NMDS分析

» Adonis分析

» Anosim分析

» 样本层级聚类分析

群落组成分析 <

环境因子分析 <

差异物种分析 <

## PCA Analysis

Standardization method:

hellinger

Group:

MetS

Submit

1. 选择标准化方法

2. 选择分组

PCA分析（Principal Component Analysis），是一种有效去除噪音和冗余，将原有的复杂数据降维分析的方法，揭示隐藏在复杂数据背后的简单结构。



# PCA分析

分析总览

分析前准备

Alpha多样性分析

Beta多样性分析

» PCA分析

» PCoA分析

» NMDS分析

» Adonis分析

» Anosim分析

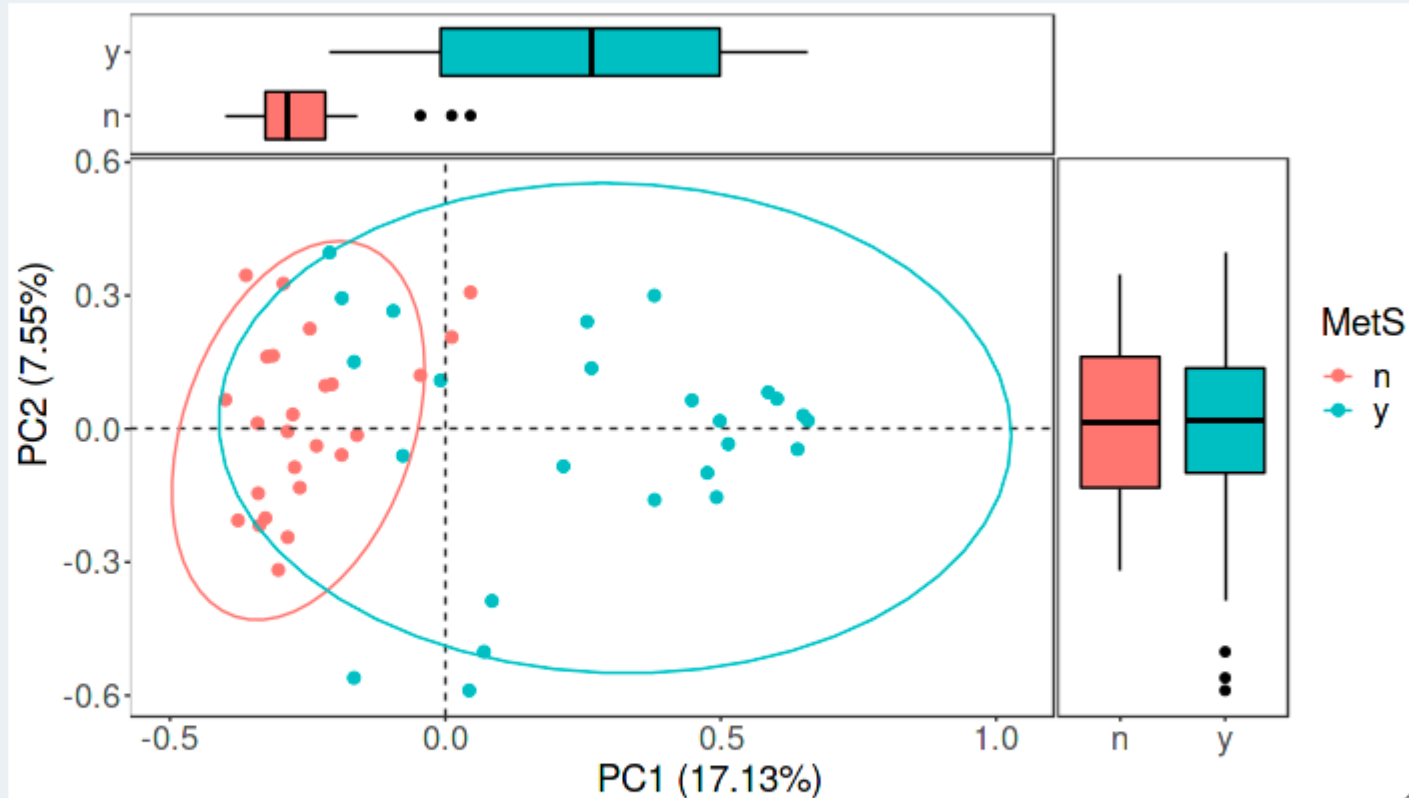
» 样本层级聚类分析

群落组成分析

环境因子分析

差异物种分析

PCA Analysis



基于PCA分析结果，将不同分组的样本在主成分轴上作箱线图和点图，直观地呈现不同分组样本在2个主成分的差异离散情况。

分析总览

分析前准备 <

Alpha多样性分析 <

Beta多样性分析 <

» PCA分析

» PCoA分析

» NMDS分析

» Adonis分析

» Anosim分析

» 样本层级聚类分析

群落组成分析 <

环境因子分析 <

差异物种分析 <

## PCoA Analysis

Standardization method:

hellinger

Distance method:

bray

Group:

MetS

Submit

1. 选择标准化方法

2. 选择计算距离方法

2. 选择分组

PCoA分析与PCA类似，二者都是通过降维找出影响样本群落组成差异的潜在主成分。主要区别在于，PCA利用物种丰度和基于欧氏距离直接作图，而PCoA是基于所选距离矩阵进行作图（请选择度量距离的方法）。

# PCoA分析

分析总览

分析前准备

Alpha多样性分析

Beta多样性分析

» PCA分析

» PCoA分析

» NMDS分析

» Adonis分析

» Anosim分析

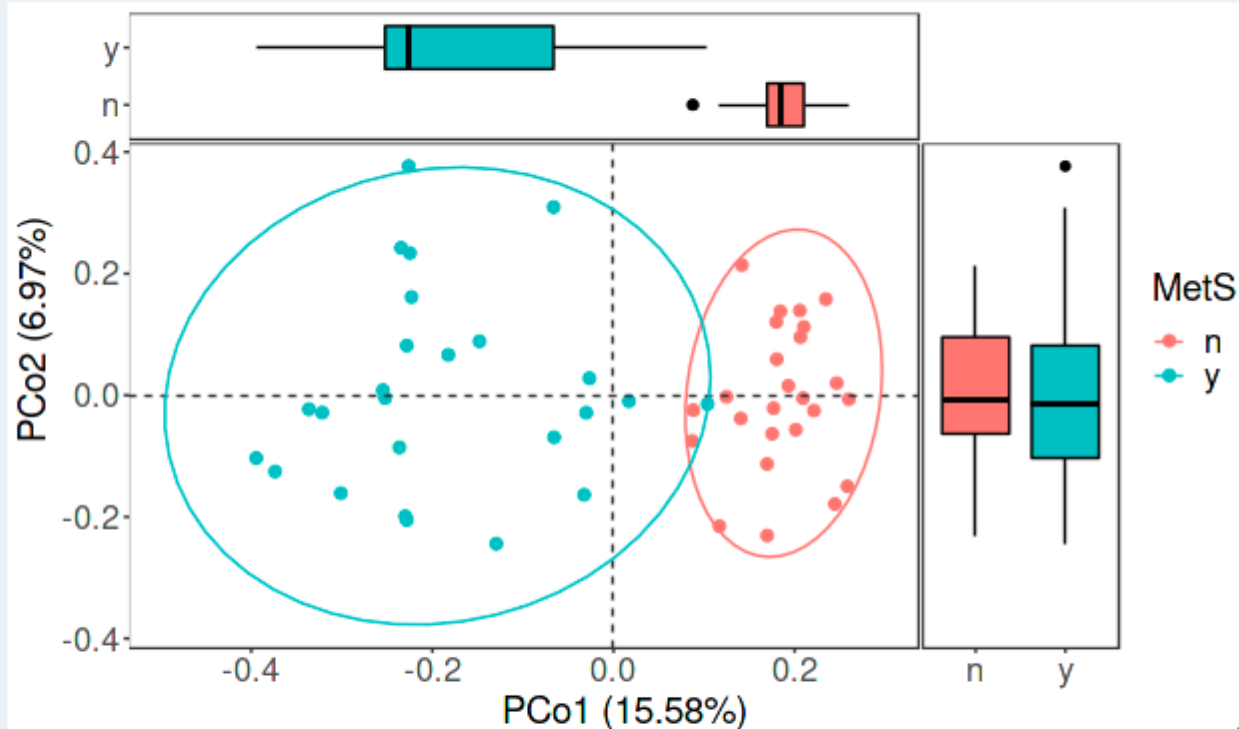
» 样本层级聚类分析

群落组成分析

环境因子分析

差异物种分析

PCoA Analysis



PCoA分析与PCA类似，二者都是通过降维找出影响样本群落组成差异的潜在主成分。主要区别在于，PCA利用物种丰度和基于欧氏距离直接作图，而PCoA是基于所选距离矩阵进行作图（请选择度量距离的方法）。

分析总览

分析前准备

Alpha多样性分析

Beta多样性分析

» PCA分析

» PCoA分析

» NMDS分析

» Adonis分析

» Anosim分析

» 样本层级聚类分析

群落组成分析

环境因子分析

差异物种分析

## NMDS Analysis

Standardization method:

hellinger

Distance method:

bray

Group:

MetS

Submit

NMDS分析，即非度量多维尺度分析(non-metric multi-dimensional scaling)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。

1. 选择标准化方法
2. 选择计算距离方法
2. 选择分组



# NMDS分析

分析总览

分析前准备

Alpha多样性分析

Beta多样性分析

» PCA分析

» PCoA分析

» NMDS分析

» Adonis分析

» Anosim分析

» 样本层级聚类分析

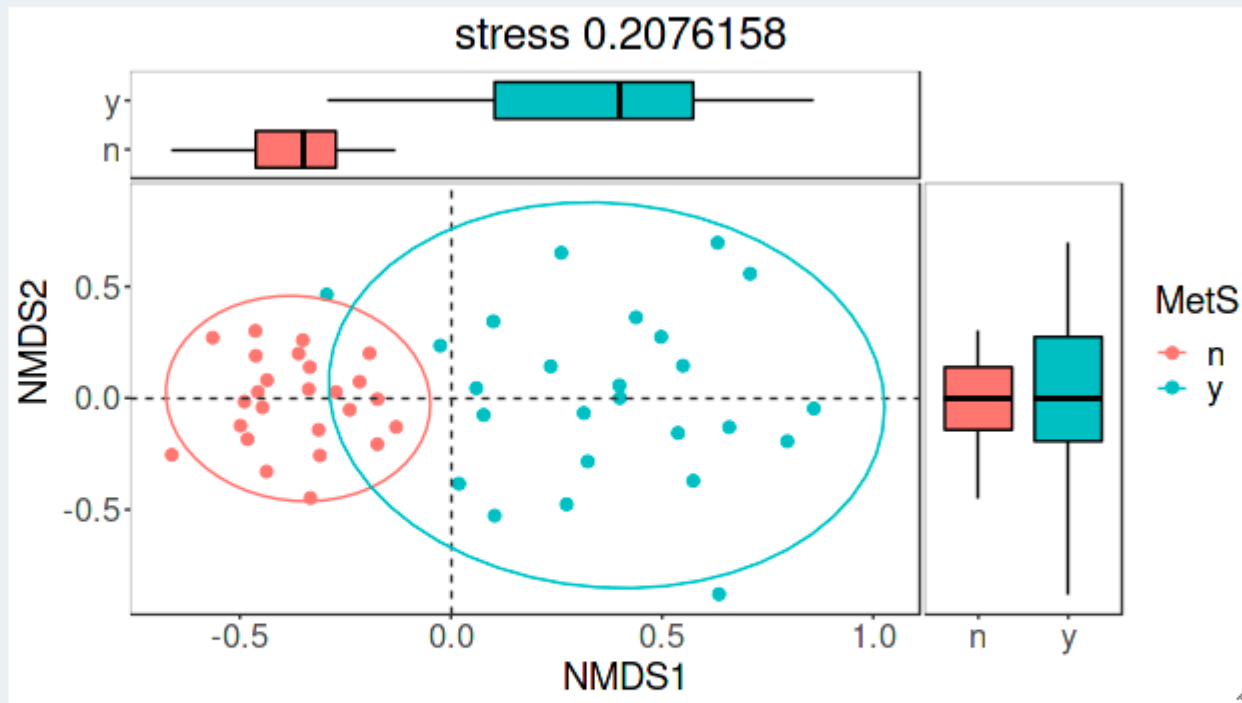
群落组成分析

环境因子分析

差异物种分析

NMDS Analysis

+



Stress值：检验NMDS分析结果的优劣。通常认为

1. 当stress < 0.2时可用NMDS的二维点图表示，其图形有一定的解释意义；
2. 当stress < 0.1时，可以认为是一个好的排序；
3. 当stress < 0.05时，则具有很好的代表性。



# Adonis分析

分析总览

分析前准备

Alpha多样性分析

Beta多样性分析

» PCA分析

» PCoA分析

» NMDS分析

» Adonis分析

» Anosim分析

» 样本层级聚类分析

群落组成分析

环境因子分析

差异物种分析

## Adonis Analysis

Standardization method:

hellinger

Distance method:

bray

Group:

MetS

Permutations:

999

Submit

Adonis分析利用距离矩阵对总方差进行分解，分析不同分组因素对样品差异的解释度，并使用置换检验（置换次数999）对划分的统计学意义进行显著性分析。

1. 选择标准化方法
2. 选择计算距离方法
3. 选择分组
4. 选择置换次数

# Adonis分析

分析总览

分析前准备

Alpha多样性分析

Beta多样性分析

» PCA分析

» PCoA分析

» NMDS分析

» Adonis分析

» Anosim分析

» 样本层级聚类分析

群落组成分析

环境因子分析

差异物种分析

## Adonis Analysis

Call:  
vegan::adonis(formula = .formula, data = sampled, permutations = permutatio

Permutation: free  
Number of permutations: 999

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
MetS	1	1.7774	1.77742	7.0688	0.12836	0.001 ***
Residuals	48	12.0694	0.25145		0.87164	
Total	49	13.8468			1.00000	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

1. R2值代表分组因素对样本差异的解释度，R2越大表示分组对差异的解释度越高；
2. Pr，为P值(=0.001\*\*\*)，说明本次检验的可信度高。

# ANOSIM分析

分析总览

分析前准备 <

Alpha多样性分析 <

Beta多样性分析 <

>> PCA分析

>> PCoA分析

>> NMDS分析

>> Adonis分析

>> Anosim分析

>> 样本层级聚类分析

群落组成分析 <

环境因子分析 <

差异物种分析 <

## Anosim Analysis

Standardization method:

hellinger

Distance method:

bray

Group:

MetS

Permutations:

999

Submit

1. 选择标准化方法
2. 选择计算距离方法
3. 选择分组
4. 选择置换次数

ANOSIM分析用来检验组间（两组或多组）的差异是否显著大于组内差异，从而判断分组是否有意义。

# ANOSIM分析

分析总览

分析前准备

Alpha多样性分析

Beta多样性分析

» PCA分析

» PCoA分析

» NMDS分析

» Adonis分析

» Anosim分析

» 样本层级聚类分析

群落组成分析

环境因子分析

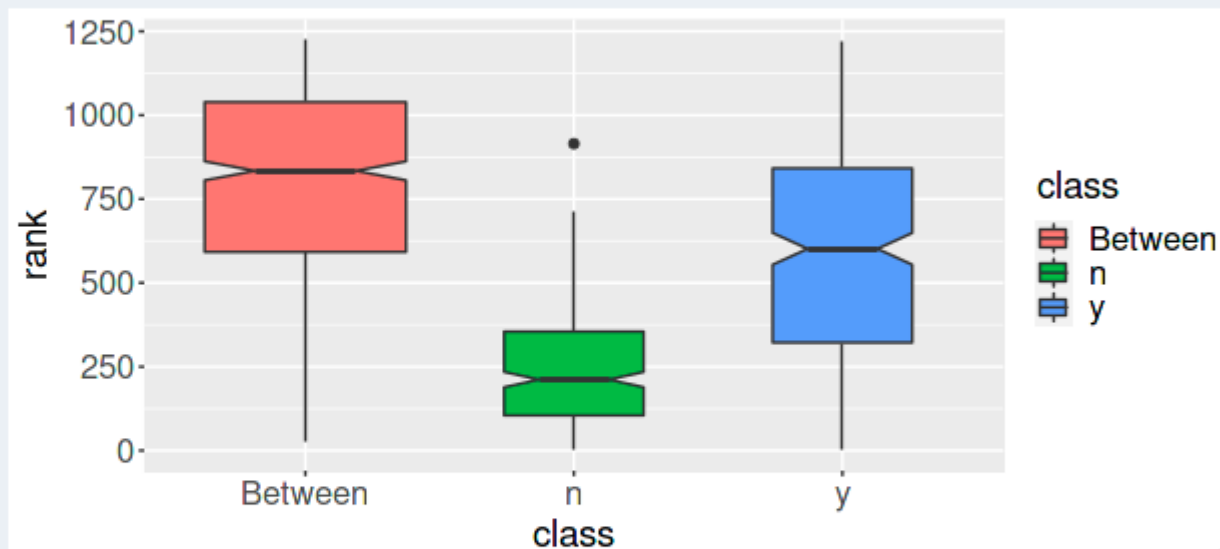
差异物种分析

## Anosim Analysis

ANOSIM statistic R: 0.6232

Significance: 0.001

Permutation: free Nu



R的范围为 $[-1, 1]$ ,  $R > 0$ 说明组间差异大于组内差异,  $R < 0$ 组间差异小于组内差异。P值则说明不同组间差异是否显著 (0.001\*\*\*, 差异显著)。



# 样本层次聚类分析

分析总览

分析前准备

Alpha多样性分析

Beta多样性分析

» PCA分析

» PCoA分析

» NMDS分析

» Adonis分析

» Anosim分析

» 样本层级聚类分析

群落组成分析

环境因子分析

差异物种分析

## Hierarchical clustering analysis

Standardization method:

hellinger

Distance method:

bray

Hierarchical clustering method:

average

Graph type:

rectangular

Group:

MetS

Submit

1. 选择标准化方法
2. 选择计算距离方法
3. 选择聚类的方法
4. 选择树的图层样式
5. 选择分组

对距离矩阵进行层级聚类（Hierarchical clustering）  
可以清楚地看出样本分支的距离远近。



# 样本层次聚类分析

分析总览

分析前准备

Alpha多样性分析

Beta多样性分析

» PCA分析

» PCoA分析

» NMDS分析

» Adonis分析

» Anosim分析

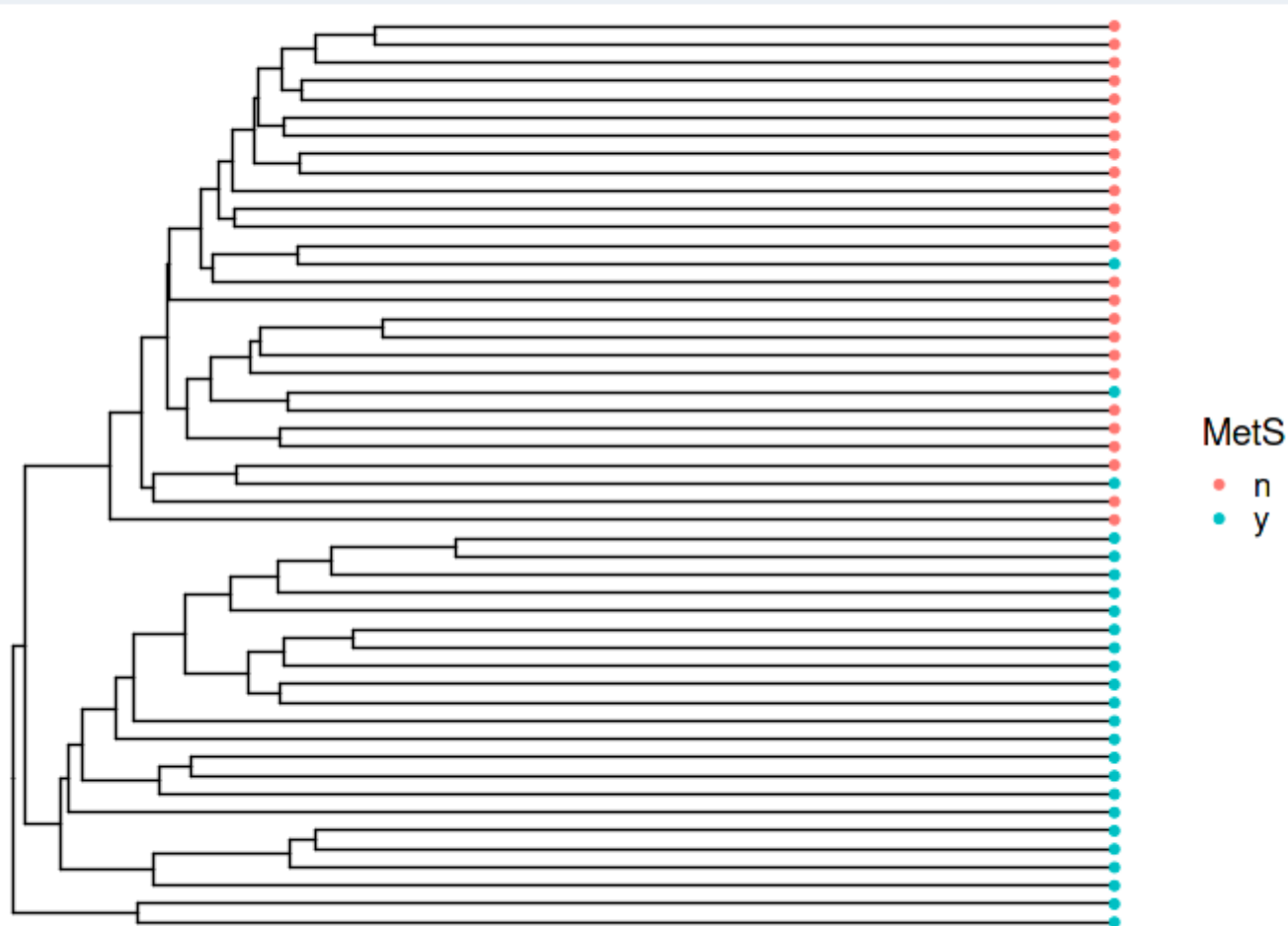
» 样本层级聚类分析

群落组成分析

环境因子分析

差异物种分析

Hierarchical clustering analysis



基于R包ggtree对结果可视化

# 样本层次聚类分析

分析总览

分析前准备 <

Alpha多样性分析 <

Beta多样性分析 <

>> PCA分析

>> PCoA分析

>> NMDS分析

>> Adonis分析

>> Anosim分析

>> 样本层级聚类分析

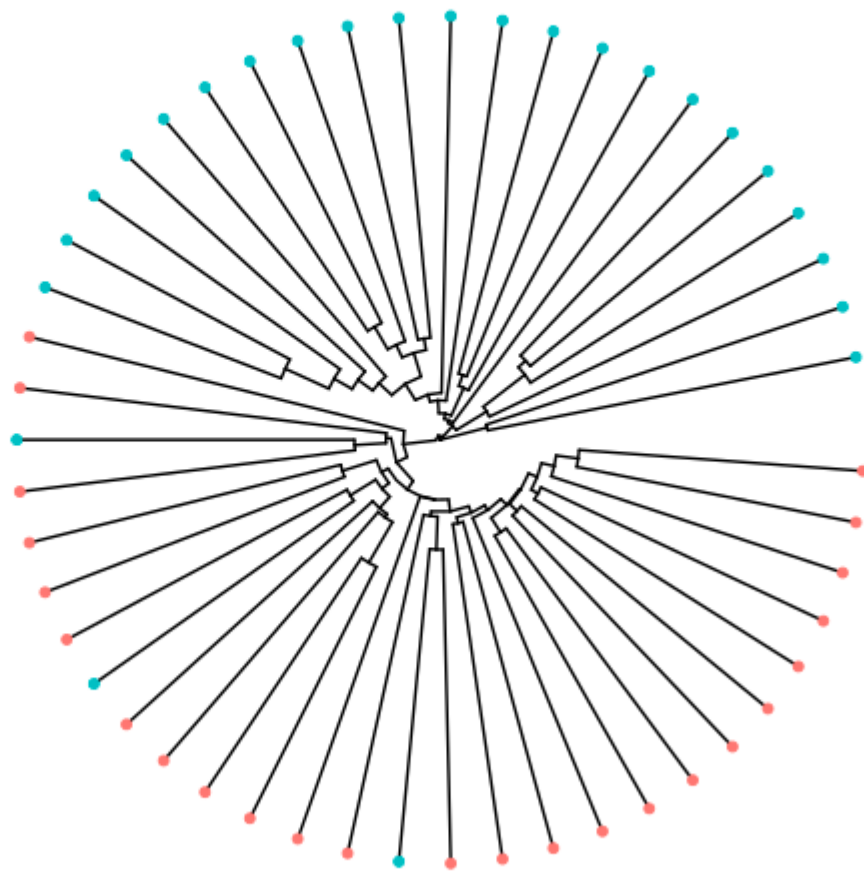
群落组成分析 <

环境因子分析 <

差异物种分析 <

Hierarchical clustering analysis

+



MetS

• n  
• y

# 群落组成分析

分析总览

分析前准备

Alpha多样性分析

Beta多样性分析

群落组成分析

» Bar图

环境因子分析

差异物种分析

## Taxonomy composition Analysis

Taxonomic level:

Class

Abundance type:

relative

Group:

MetS

Top most abundant:

5

Submit

1. 分类水平
2. 相对丰度或绝对数量
3. 选择分组
4. 优势物种的数目

根据群落组成分析，可以得知不同分组（或样本）在各分类水平（如域、界、门、纲、目、科、属、种）上的物种组成情况。

# 群落组成分析

分析总览

分析前准备

Alpha多样性分析

Beta多样性分析

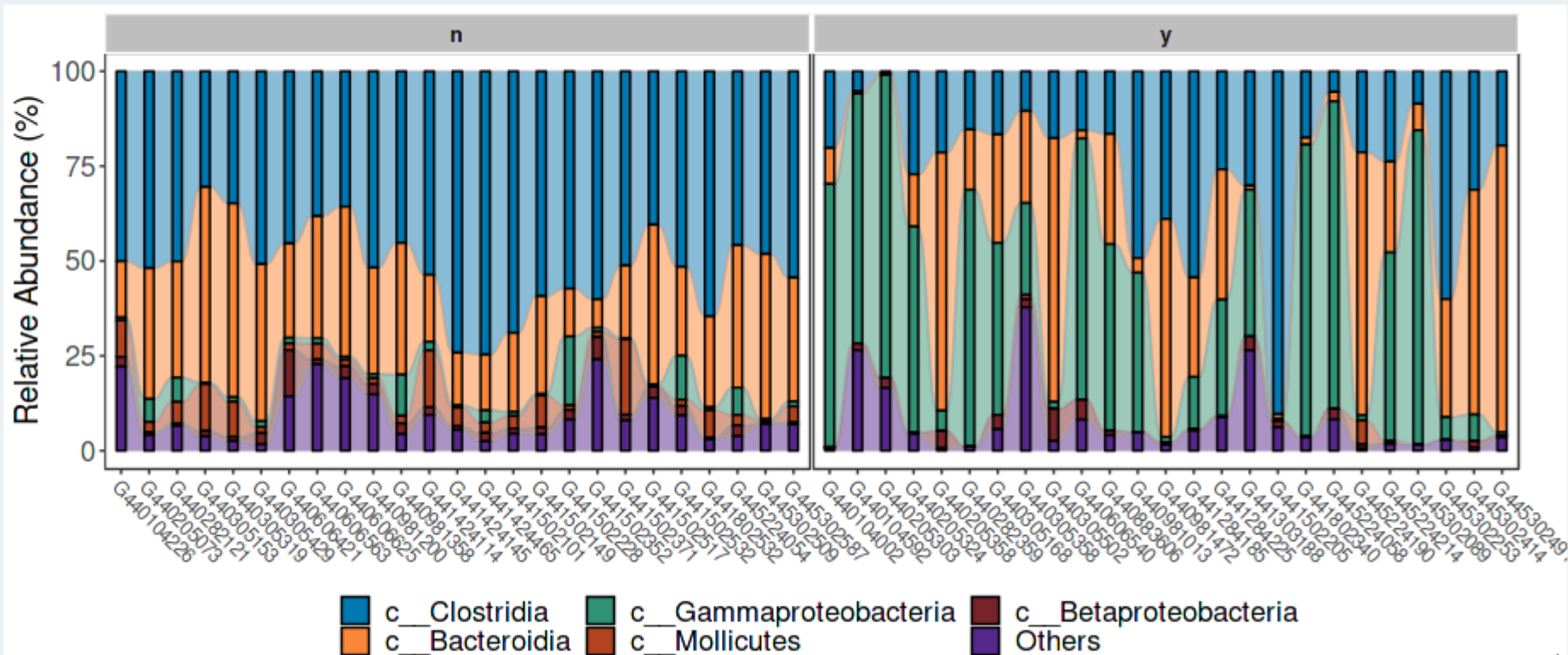
群落组成分析

» Bar图

环境因子分析

差异物种分析

Taxonomy composition Analysis





# Maaslin2分析

分析总览

分析前准备

Alpha多样性分析

Beta多样性分析

群落组成分析

环境因子分析

» Maaslin2分析

差异物种分析

## Maaslin2 Analysis

Min abundance:

0

Min prevalence:

0.1

Min variance:

0

Q-value cutoff:

0.25

Top features

50

Taxonomic level:

Class

Analysis method:

LM

Standardize(or not):

TRUE

Normalization method:

TSS

Transform method:

LOG

Submit

### Total effects

Sample  
district  
age  
gender  
occupation  
smoke\_now  
height  
weight  
BMI  
waist  
Bristol\_stool\_type  
MetS  
Obesity  
Overweight



### Fixed effects



### Random effects

**MaAslin (Multivariate Association with Linear Models)** 分析是一种通过线性模型挖掘环境因子（如临床数据、理化指标等）与微生物群落物种或功能相对丰度数据间相关性的分析。

分析总览

分析前准备

Alpha多样性分析

Beta多样性分析

群落组成分析

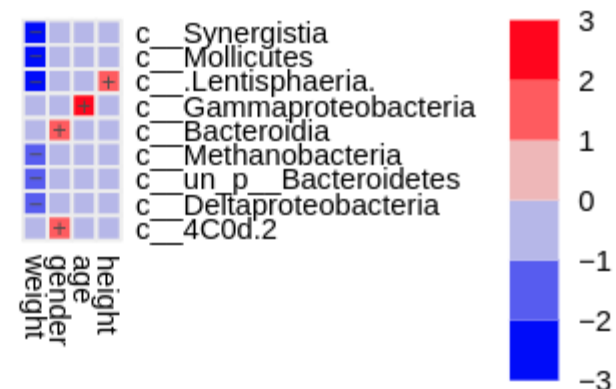
环境因子分析

» Maaslin2分析

差异物种分析

## Maaslin2 Analysis

### Significant associations ( $-\log(qval) * \text{sign}(coeff)$ )



# LEfSe分析

分析总览

分析前准备

Alpha多样性分析

Beta多样性分析

群落组成分析

环境因子分析

差异物种分析

» LEfSe分析

## LEfSe Analysis

LDA score threshold:

6

P value cutoff:

0.01

P value adjust:

fdr

Taxonomy level:

Class

Group:

MetS

Graph layout:



Box plot



Clade plot

Submit

1. LDA值
2. p值
3. p值校正
4. 分类水平
5. 选择分组
6. 选择图层样式

LEfSe分析首先使用非参数Kruskal-Wallis 秩和检测不同分组间丰度差异显著的物种，然后使用Wilcoxon秩和检验上一步的差异物种在不同组间子分组中的差异一致性，最后采用线性回归分析（LDA）来估算每个组分（物种）丰度对差异效果影响的大小。

分析总览

分析前准备

Alpha多样性分析

Beta多样性分析

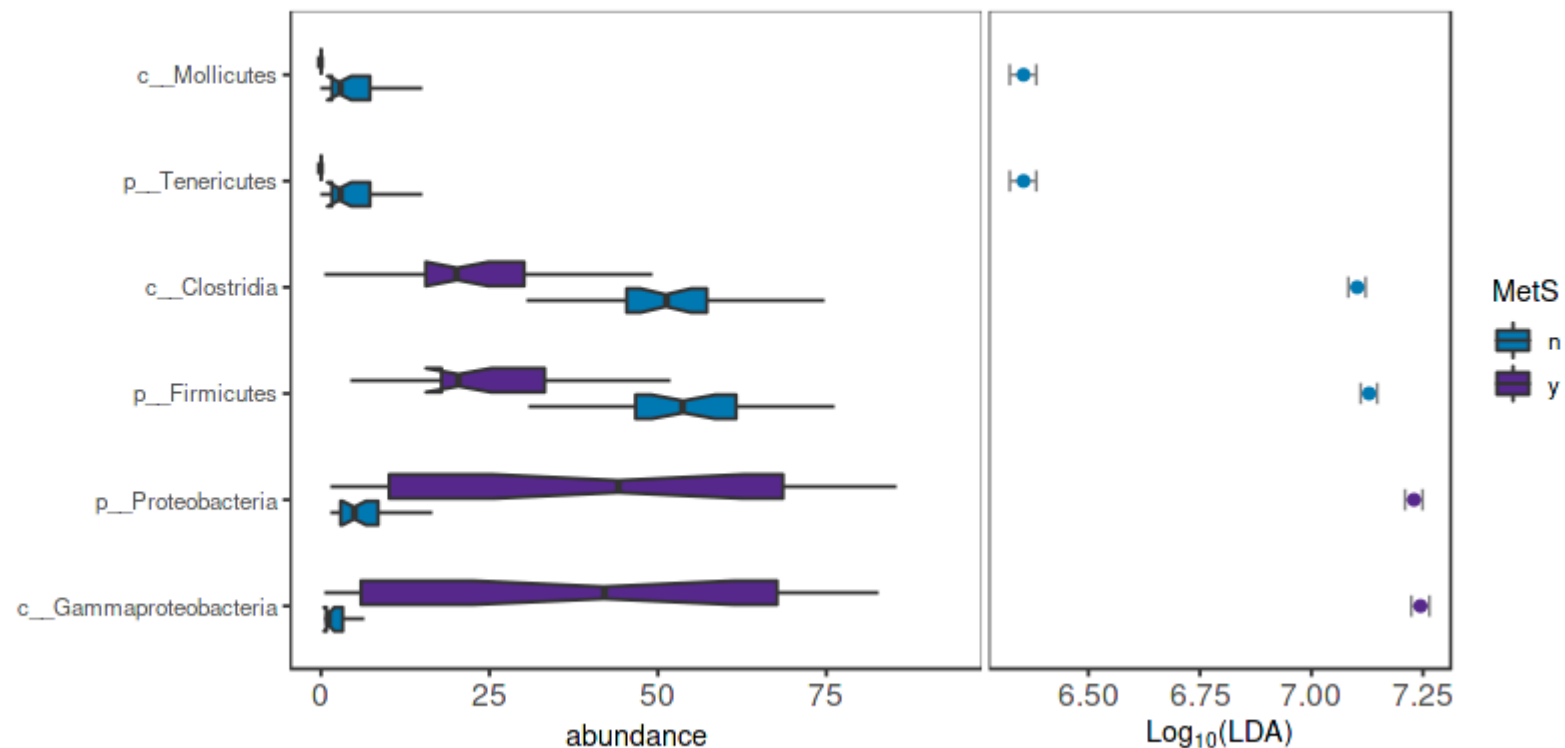
群落组成分析

环境因子分析

差异物种分析

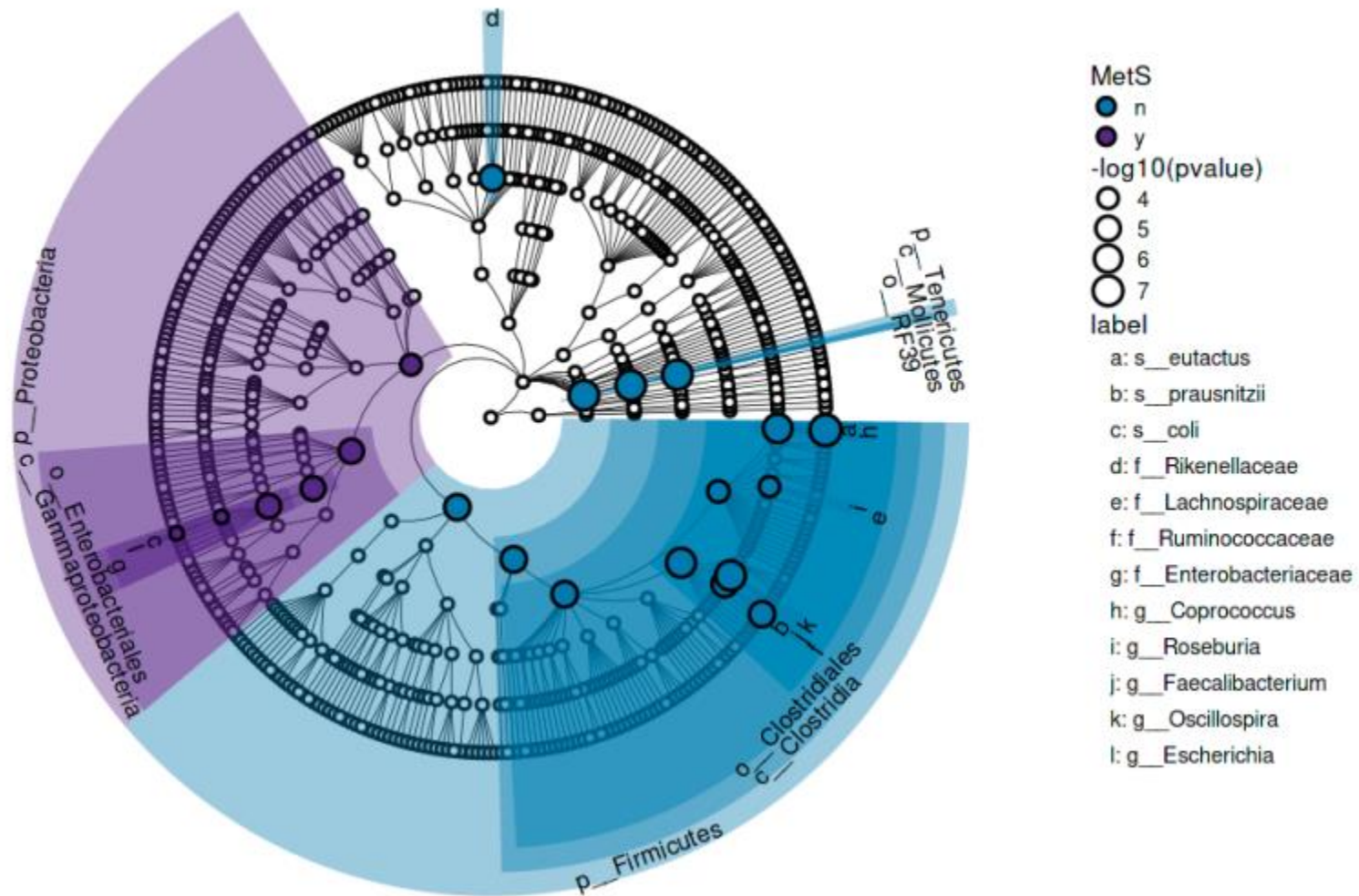
» LEfSe分析

## LEfSe Analysis





# LEfSe分析





**Thank You**