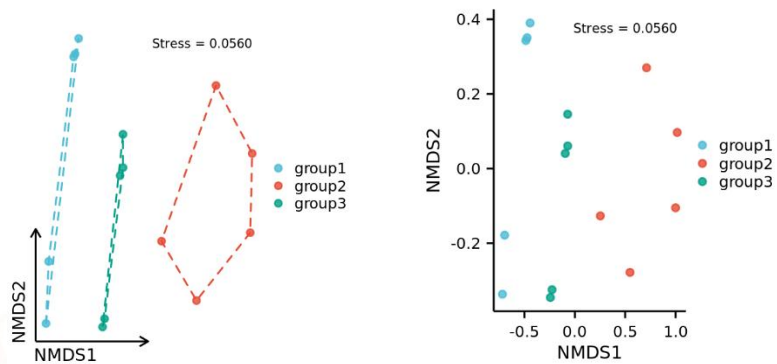


差异表达 - NMDS 图



网址: <https://www.xiantao love>



更新时间: 2023.09.11

目录

基本概念	3
应用场景	3
主要结果	4
数据格式	5
参数说明	6
分析参数	6
点	7
外圈连线	8
中心点	8
箱	9
应力值标注	10
样本标注	11
标题	11
图注(Legend)	12
风格	13
图片	14
结果说明	15
主要结果	15
补充结果	16
方法学	18
如何引用	19
常见问题	20

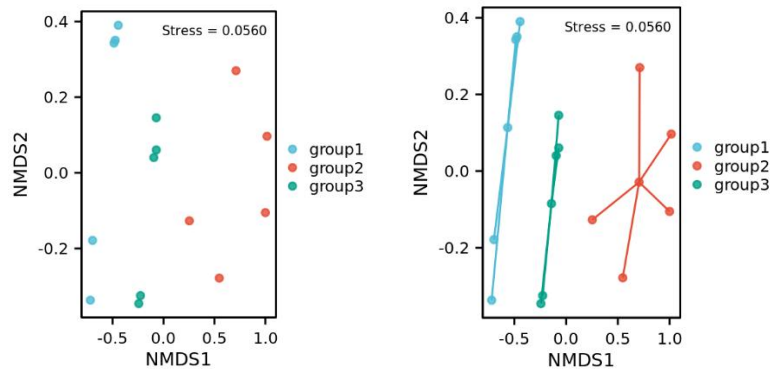
基本概念

- NMDS（非度量多维尺度分析）：通过降维计算样本间的距离来展示差异程度。常用于微生物群落研究的 β 分析，根据 OTU/ASV 丰度、物种组成或者功能丰度，NMDS 都可以帮助揭示样本之间的相似性和差异性，为微生物群落研究提供重要的数据可视化和解读手段。
- 计算距离方法：
 - Bray-Curtis, 以该统计指标的提出者 J. Roger Bray 和 John T. Curtis 的名字命名的，是可以转化为相似性指数的距离指数。在宏基因组和微生物多样性测序分析中使用最多，主要基于 OTUs 或 ASV 等的计数统计，比较两个群落微生物的组成差异。
 - Jaccard, 又称为 jaccard similarity coefficient, 用于比较有限样本集之间的相似性和差异性。在处理离散型（分类型）变量的相识度时非常有用。
 - Euclidean, 欧式距离，在 N 维空间中两个点之间的真实距离，或者向量的自然长度（即该点到原点的距离）。

应用场景

- 可以用来研究样本（群落组成）的相似性或相异性。
 - 微生物多样性测序数据中样本之间的群落组成差异，样本间距离越近代表微生物群落结构越相似。

主要结果



典型的 NMDS 图以点图形式展示。

- NMDS 是距离值的秩次（数据排名）信息的评估，图形上样本信息仅反映样本间数据秩次信息的远近，而不反映真实的数值差异，横纵坐标轴并无权重意义。
- 图中每个点代表每个样本在 NMDS1 和 NMDS2 中对应的映射位置信息，点与点（样本与样本）间的距离情况能体现样本间的差异。
- 图中不同的颜色表征不同样本所属的组，这部分来自上传数据中的 #注释头部内容，具体可见数据格式说明。
- 右图中给样本不同组增加了中心点连线。

数据格式

#group	group1	group1	group1	group1	group1	group2	group2	group2	group2	group2
OTU_ID	CK1	CK2	CK3	CK4	CK5	T1	T2	T3	T4	T5
1	17	22	0	3	2	23	10	1	8	1
2	143	137	223	186	222	69	56	5	14	9
3	272	214	8	8	6	20	16	80	58	7
4	62	59	6	1	3	0	0	0	0	0
5	13	12	0	0	1	7	7	0	1	0
6	28	34	66	57	86	0	0	0	0	0
7	208	257	122	45	119	45	33	3	2	5
8	73	62	23	4	29	22	20	1	0	1
9	198	210	386	242	322	191	346	214	277	322
10	106	125	203	171	200	142	107	56	70	81
11	1361	1287	280	239	295	245	183	73	99	43
12	122	142	371	507	473	22	14	1	1	1
13	73	58	23	18	29	19	15	5	6	5
14	2	6	59	47	58	0	2	6	1	41
15	59	54	10	7	19	1	3	0	0	0
16	47	59	74	106	110	21	10	3	5	23
17	598	532	127	89	94	44	46	44	42	24
18	121	91	29	15	26	1	0	0	0	2

数据要求：

➤ 头部注释行（以#开头）：

- 用于表征每个样本所属的分组，至少需要提供 1 行样本的注释信息，每行的分组最多是 10 个。注意，注释行不能超过 4 行。

➤ 主体部分：

- 数据至少有 4 列以上，至少需要 5 行数据。
- 主体的第一行为样本编号（如图中的第 2 行），这一行不能含有缺失、重复及特殊字符。
- 主体的第一列为变量名（示例是微生物多样性测序数据中的 OTU id，也可以是 ASV id 或物种等）。
- 主体的其他部分为样本在各个维度对应的数值，不能含有非数值内容。
- 数据不能含有整列之和为 0 的样本。
- 数据中不能含有负数值。

➤ 最多支持 500 列，5000 行。若验证数据时返回报错，需要在上传数据内进行相应的调整，然后再上传数据。

参数说明

(说明：标注了颜色的为常用参数。)

分析参数

分析参数	▼
距离算法	Bray-Curtis ▼
种子号	2023

- **距离算法**：计算距离方法默认为 Bray-Curtis，计算距离方法的选择可以参考“基本概念”中计算距离方法的说明。
- **种子号**：设置种子数可以保证统计检验分析结果可重复，默认为 2023，此参数请输入非零整数。

点

点

填充色

描边色

样式 圆形 ×

大小 1

不透明度 0.8

- **填充色**：点的填充色颜色选项，取决于上传数据中的头部注释行信息，有多少个分组会提取多少个颜色，最多支持修改 10 个颜色。受配色方案全局性修改。
- **描边色**：点的描边色颜色选项，取决于上传数据中的头部注释行信息，有多少个分组会提取多少个颜色，最多支持修改 10 个颜色。受配色方案全局性修改。
- **样式**：点的样式类型，可选择 圆形、正方形、菱形、三角形、倒三角，默认为圆形。多选，多选后不同的分组/分类中的点的类型也会有相应变化，循环取该参数值。
- **大小**：点的大小。
- **不透明度**：点的透明度。0 为完全透明，1 为完全不透明。

外圈连线

外部连线

是否展示

线条类型

实线

线条粗细

0.75pt

- 是否展示：是否需要圈住分组的不同分类连线，是由各个组最外层的点连接而成，最少两个样本才能展示。
- 线条类型：外圈连线的线条样式类型，可选择 实线、虚线，默认为实线。单选，选择类型后所有外圈连线的线条都统一改变。
- 线条粗细：外圈连线的线条粗细，默认为 0.75pt。

中心点

中心点

是否展示

中心点大小

1

线条类型

实线

线条粗细

0.75pt

- 是否展示：是否需要展示每个分组的中心点，是由中心点与同组其他点连接而成，最少两个样本。
- 中心点大小：中心点的大小。
- 线条类型：中心点连线的线条类型，可选择 实线、虚线，默认为实线。单选，选择类型后所有中心点连线的线条都统一改变。
- 线条粗细：中心点连线的粗细，默认为 0.75pt。

箱



- 是否展示：是否添加边际箱线图来展示组间差异性。默认不展示。

应力值标注

应力值标注 

是否展示 

图注位置 右上 

字的大小 5pt 

- **是否展示**：是否需要标注应力值。默认为标注。
- **图注位置**： 可选右上、右下、左上、左下，默认为右上。
- **标注大小**：控制图中需标注的应力值大小，默认为 5pt。



样本标注

样本标注

选择

不标注

特定样本

可以输入想要标注的样本, 1行1个

标注大小

5pt

- 类型选择：是否需要标注样本编号信息。可选择 不标注、标注全部样本、标注下面特定样本，默认为不标注。
- 特定样本：当上一个参数选择了“标注下面特定样本”时，将根据此参数输入的样本编号在图上进行标注，一行一个。注意样本编号是否与上传数据的样本信息保持一致。
- 标注大小：控制图中需标注的文字大小，默认为 5pt。

标题

标题

大标题

大标题内容

x轴标题

x轴标题内容

y轴标题

y轴标题内容

- 大标题：大标题文本

- x 轴标题: x 轴标题文本
- y 轴标题: y 轴标题文本
- 补充: 在要换行的中间插入\n。选择经典类型风格时如果需要上标, 可以用两个英文输入法下的大括号括住, 比如 {{2}}; 如果需要下标, 可以用两个英文输入法下的中括号括住, 比如 [[2]]。

图注(Legend)



图注

是否展示 ☒

图注标题 图注标题内容

图注标签 图注标签内容

图注位置 默认

- 是否展示: 是否展示图注。
- 图注标题: 可以添加图注标题。
- 图注标签: 可以修改图注中分组标签的名字, 如果有多个名字要修改, 则需要把这些名字以逗号的形式合并成一个, 类似 A, B。
- 图注位置: 可选右、上, 至少有两个分组时默认为右。

风格

风格

坐标样式 经典类型

边框 ☐

网格 ☐

文字大小 7pt

- 坐标样式：无边框的情况下，坐标轴的样式。可选择经典类型、指向类型，默认为经典类型。
- 边框：是否添加外框。
- 网格：是否添加网格。
- 文字大小：针对图中所有文字整体的大小控制，默认为 7pt。

图片

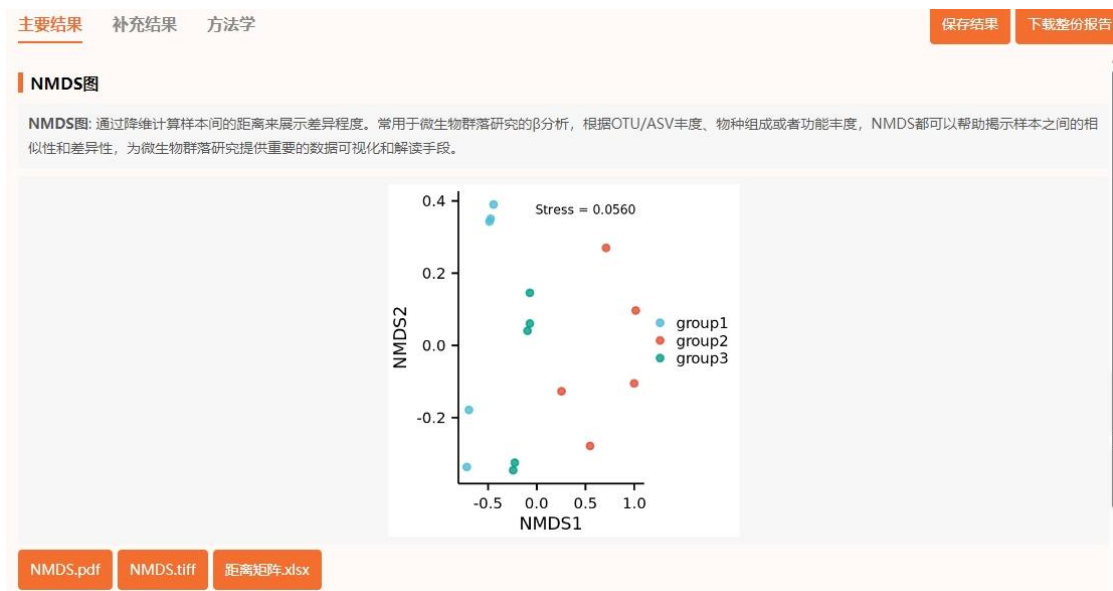
图片	▼
宽度 (cm)	5
高度 (cm)	5
字体	Arial ▼

- 宽度：图片横向长度，单位为 cm。
- 高度：图片纵向长度，单位为 cm。
- 字体：可以选择图片中文字的字体。



结果说明

主要结果



主要结果格式为图片格式, 提供 PDF、TIFF 格式下载, 结果报告可以下载包括 pdf 以及说明文本的内容。

	A	B	C	D	E	F	G	H	I	J	K
1	sample	CK1	CK2	CK3	CK4	CK5	T1	T2	T3	T4	T5
2	CK1	0	0.10891682	0.51293394	0.53783748	0.51600683	0.67980489	0.61637133	0.80040538	0.79638155	0.80515252
3	CK2	0.10891682	0	0.46674778	0.47998668	0.46476021	0.67789479	0.60211107	0.78702945	0.78912933	0.78359892
4	CK3	0.51293394	0.46674778	0	0.1837912	0.0997272	0.699854	0.59870782	0.751449	0.76251236	0.64748841
5	CK4	0.53783748	0.47998668	0.1837912	0	0.12541007	0.69401286	0.58416247	0.7314925	0.75125971	0.61913659
6	CK5	0.51600683	0.46476021	0.0997272	0.12541007	0	0.6985809	0.59487376	0.74448836	0.75529024	0.63206497
7	T1	0.67980489	0.67789479	0.699854	0.69401286	0.6985809	0	0.34735003	0.48168628	0.354238	0.46773088
8	T2	0.61637133	0.60211107	0.59870782	0.58416247	0.59487376	0.34735003	0	0.49591856	0.53025556	0.46988091
9	T3	0.80040538	0.78702945	0.751449	0.7314925	0.74448836	0.48168628	0.49591856	0	0.22717166	0.32853817
10	T4	0.79638155	0.78912933	0.76251236	0.75125971	0.75529024	0.354238	0.53025556	0.22717166	0	0.33515177
11	T5	0.80515252	0.78359892	0.64748841	0.61913659	0.63206497	0.46773088	0.46988091	0.32853817	0.33515177	0

另外, 提供样本间的距离矩阵结果表格 .xlsx 下载, 第一列和第一行均为样本名, 中间数据为上下三角对称的样本间距离。

补充结果

位置坐标

sample	NMDS1	NMDS2
CK1	-0.71937	-0.33677
CK2	-0.69743	-0.17869
CK3	-0.48657	0.343
CK4	-0.44459	0.39001
CK5	-0.47453	0.35047
T1	0.54771	-0.27835
T2	0.25342	-0.12701
T3	1.0169	0.096626
T4	1.0004	-0.10521
T5	0.71202	0.26998
B1	-0.24322	-0.34568
B2	-0.22659	-0.32475
B3	-0.071355	0.14556

位置坐标.xlsx

表中, 第一列为数据中的样本, 第二和第三列为样本对应的NMDS1和NMDS2值, 对应于NMDS图中各个点的位置坐标, NMDS1代表横坐标, NMDS2代表纵坐标

此表格提供位置坐标表格 [xlsx](#) 下载, 每个样本对应 NMDS1 和 NMDS2 的位置信息。

多元方差分析 (ADONIS)

ADONIS 为多元方差分析, 亦可称为非参数多元方差分析, 它通过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, 并使用置换检验对其统计学意义进行显著性分析, 从而判断不同分组之间是否有差异, 差异是否显著

统计检验方法	自由度	总方差	F检验值	R2	Pr(>F)
Adonis	2	1.3476	9.876	0.62207	0.0099

· R2 表示该分组方式对样品间差异的解释度, R2越大说明该分组方案对差异的解释度越高

· Pr 表示P值, 小于0.05说明本次检验的可信度高, 表明该分组水平在统计学上有显著性的差异

▶ adonis2函数 - Adonis检验分析-原始输出 [点击展开](#)

此表格为多元方差分析 (ADONIS) 结果, 用于判断不同分组之间是否有差异, 差异是否显著。 (只有数据中的#注释头部内容包含两个分组及以上, 且分组内包含 3 个样本及以上时才分析)

相似性分析 (ANOSIM)

ANOSIM 为非参数检验方法, 它通过检验组间的差异是否显著大于组内差异, 从而评判分组是否有意义

统计检验方法	统计量(R)	显著性(P)
ANOSIM	0.84533	0.0099

· 统计量为组间相关性, R 值在 -1 和 1 之间, R值越接近-1表示组内差异越大大于组间差异, R值越接近1表示组间差异越大大于组内差异, R值接近0则表示不同分组样本为随机分布, 组间和组内没有明显差异

· 显著性 p 值, 小于0.05表示统计具有显著性

▶ anosim函数 - ANOSIM检验分析-原始输出 [点击展开](#)

此表格为相似性分析（ANOSIM）结果，它通过检验组间的差异是否显著大于组内差异，从而评判分组是否有意义。（只有数据中的#注释头部内容包含两个分组及以上，且分组内包含 3 个样本及以上时才分析）

多元响应置换分析（MRPP）

MRPP分析，主要是用于分析高维度数据组间相似性的统计方法。

统计检验方法	统计量(A)	显著性(P)
MRPP	0.36029	0.0099

· 统计量 A 值用于评价组间与组内的差异， $A > 0$ ，表示组间差异大于组内差异； $A < 0$ ，表示组内差异大于组间差异

· 显著性 p 值，小于0.05表示统计具有显著性

此表格为多元响应置换分析（MRPP）结果，主要是用于分析高维度数据组间相似性的统计方法，用于判断组间数据是否具有显著性差异。（只有数据中的#注释头部内容包含两个分组及以上，且分组内包含 3 个样本及以上时才分析）



方法学

所有分析和可视化均在 R 4.2.1 中进行

涉及的 R 包: ggplot2 包 (用于可视化)、vegan (用于分析)

处理过程:

- (1) 将清洗后的数据, 通过 `vegdist` 函数计算样本间的距离矩阵, 再对距离矩阵使用 `metaMDS` 函数进行 NMDS 分析。
- (2) 使用 `ggplot2` 包对结果进行可视化。



如何引用

生信工具分析和可视化用的是 R 语言，可以直接写自己用 R 来进行分析和可视化即可，可以无需引用仙桃，如果想要引用仙桃，可以在致谢部分 (Acknowledge) 致谢仙桃学术 (www.xiantao love)。

方法学部分可以参考对应说明文本中的内容以及一些文献中的描述。



常见问题

1. 同样是降维排序，与 PCA、PCoA 等的区别？

答：

- PCA、PCoA 与 NMDS 都是以降维思想为核心的排序分析方法。PCA 分析是对输入的 OTU 丰度原始数据的降维，而 PCoA 与 NMDS 则是基于各类型样本相似性距离的降维。
- 不同于 PCoA 分析，NMDS 弱化了对实际距离数值的依赖，更加强调数值间的排名（秩次），例如三个样本的两两相似性距离，（1,2,3）和（10,20,30）在 NMDS 分析上的排序一致，所呈现的效果是相同的。
- NMDS 主要是通过 [Stress 值判断模型的优劣](#)。

