

扩增子结果文件补充说明

04.BetaDiversity 【Beta 多样性分析结果】

1 Adonis 【Adonis 分析】

bray_adonis.txt 【Adonis 分析结果】

列数	列标题	含义
1	Vs_group	进行分析的组名称
2	Df	自由度
3	SumsOfSqs	总方差，又称离差平方和
4	MeanSqs	均方（差），即 SumsOfSqs/Df
5	F.Model	F 检验值
6	R2	表示不同分组对样本差异的解释度，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
7	Pr(>F)	表示 P 值，小于 0.05 说明本次检验的可信度高

说明：括号里面为残差项对应的值。

2 Amova 【Amova 分析】

(un)weighted_unifrac_amova.txt 【基于加权（非加权）Unifrac 距离的 Amova 结果】

列数	列标题	含义
1	vs_group	进行分析的组名称
2	SS	总方差，又称离差平方和
3	Df	自由度

4	MS	均方（差），即 SS/df
5	Fs	F 检验值
6	p-value	P 值，小于 0.05 说明组间差异显著

说明：括号里面为残差项对应的值。

3 Anosim 【Anosim 分析】

3.1 stat_anosim.txt 【Anosim 组间差异分析结果】

列数	列标题	含义
1	Group	进行分析的组名称
2	R-value	R-value 介于 (-1, 1) 之间，R-value 大于 0，说明组间差异显著。R-value 小于 0，说明组内差异大于组间差异
3	P-value	统计分析的可信度，P< 0.05 表示统计具有显著性

3.2 *.{pdf,png} 【Anosim 分析箱图结果】

4 Beta_div 【组间 Beta 多样性比较箱型图】

4.1 (un)weighted_unifrac.{png,pdf} 【基于(非)加权 unifrac 距离的 Beta 多样性箱型图】

4.2 (un)weighted_unifrac_TukeyHSD.txt 【基于(非)加权的 unifrac 距离的多组间方差分析】

列数	列标题	含义
1	diff	组间差值
2	lwr	置信区间上限

3	upr	置信区间下限
4	p adj	P 值

4.3 (un)weighted_unifrac_wilcox.txt 【基于(非)加权的 unifrac 距离的多组间非参数 wilcox 检验】

列数	列标题	含义
1	Difference	均值差
2	pvalue	P 值
3	sig.	是否显著，若 p 值<0.05,标*,p 值<0.01 标 **,p 值<0.001 标 ***
4	LCL	置信区间下限
5	UCL	置信区间上限

5 Environmen_factor 【环境因子分析】

5.1 mantel_test 【mantel_test 分析结果】

mantel_test_table.xls 【门纲目科属种以及 otu (p,c,o,f,g,s,otu) 水平上的分析结果】

列数	列标题	含义
1	Variable	环境因子信息
2	r	相关系数，r 值越大，说明该组环境因子与物种丰度信息相关性越大
3	P	显著性检验 p 值，P<0.05 表示统计具有显著性

5.2 multiCCA 【multiCCA 分析结果】

5.2.1 CCA 【包含所有环境因子的 CCA 和 RDA】

5.2.1.1 *.env.csv 【环境因子在各排序轴上的排序值】

5.2.1.2 *.sample.csv 【不同样方在各排序轴上的排序值】

5.2.1.3 *.sp.csv 【物种在各排序轴上的排序值】

其中，CCA1 表示对应的物种/样方/环境因子与第一排序轴的相关性，CCA2 表示与第二排序轴的相关性，等等。

5.2.1.4 *.envfit.csv 【envfit 函数检验每个环境因子的显著性分析结果】

列数	列标题	含义
1	CCA1/CCA2	两列所对应的值是环境因子箭头与排序轴夹角的余弦值，表示环境因子与排序轴的相关性
2	r2	环境因子对物种分布的决定系数，r2 越小，表示该环境因子对物种分布影响越小
3	Pr	表示相关性的显著性检验

5.2.1.5 dca.csv 【DCA 分析】

DCA 分析（除趋势对应分析，Detrended Correspondence Analysis）可判别用 CCA 还是 RDA 来进行环境因子对群落结构影响的展示。

物种沿着环境梯度的变化，其响应曲线可能是线性的或单峰（钟形），相对应的，RDA 分析是基于线性模型，CCA 是基于单峰模型。DCA 分析结果中前 4 个轴中最大的值如果大于 4.0，一般选用 CCA，如果 3.0-4.0 之间，选 RDA 和 CCA 均可，如果小于 3.0 一般认为 RDA 展示结果更好。

之后的 CCA_BioENV、CCA_envfit、CCA_VIF 分别是经 Bioenv、envfit、VIF 筛选后环境因子的 CCA 和 RDA，具体说明可参考 CCA。

5.2.2 BioENV.txt 【Bioenv 筛选后得到的环境因子组合】

列数	列标题	含义
1	size	各组合中包含的环境因子数量
2	correlation	各组环境因子与微生物群落的相关性

5.2.3 VIF.txt 【VIF 筛选后得到的环境因子组合】

每个环境因子所对应的 VIF 值

注意：对于 CCA/RDA 分析，环境因子个数需要小于样品个数。

6 LefSe 【LEfSe 分析结果】

6.1 LDA.*.res 【LEfSe 统计结果】

lefse 默认输出的格式，总共 5 列，第一列 biomarker 名称，第二列是平均丰度最大的 log10 的值，如果平均丰度小于 10 的按照 10 来计算，第三列是差异基因或物种富集的组名称，第四列是 LDA 值，第五列是 Kruskal-Wallis 秩和检验的 p 值，如果不是 biomarker 则用“-”表示。

6.2 LDA.*.txt 【每个 Biomaker 在各组中的分布情况】

至少两组处理，每组处理至少 3 个生物学重复（否则，无此分析）。lefse 软件默认的设置 LDA score 是 2，LDA score 的大小代表差异物种的影响大小，值越大，代表差异物种的影响越大。我们是以 LDA score 为 4 来做的，因为相对 2 来说更加严格，但是如果老师的样本组间差异并不是很大，找出的差异物种物种较少，达不到分析要求，可以降低 LDA score 来做，以期找到更多的差异物种。

7 MetaStat 【门纲目科属种（p,c,o,f,g,s）水平上的 MetaStat 分析结果，属水平目录下含有差异物种热图】

7.1 boxplot 【具有显著性差异物种的箱图结果】

7.2 PCA 【具有显著性差异物种的 PCA 结果】

7.3 cluster.*.diff.{pdf,png,t} 【具有显著性差异物种的 heatmap 热图分析结果和输入文档】

7.4 *.psig.xls 【从 MetaStat 分析结果中，筛选出的 p value \leq 0.05 的信息】

7.5 *.qsig.xls 【从 MetaStat 分析结果中，筛选出的 q value \leq 0.05 的信息】

7.6 *.test.xls 【门纲目科属 (p,c,o,f,g) 水平上 MetaStat 分析结果】

利用 Metastats 软件 (<http://metastats.cbcb.umd.edu/>) 对组间的物种丰度数据进行假设检验得到 p 值，通过对 p 值的校正，得到 q 值；最后根据 p 值或 q 值筛选具有显著性差异的物种。要求样品至少有两个分组，每个分组至少 3 个样品。

.psig.xls、.qsig.xls、.test.xls 表格说明如下：

列数	列标题	含义
1	Taxa	物种分类信息
2	mean (G1)	第一组的平均值
3	Variance (G1)	第一组的方差
4	standard error (G1)	第一组的标准差
5	mean (G2)	第二组的平均值
6	variance (G2)	第二组的方差
7	standard error (G2)	第二组的标准差
8	p value	假设检验的 p 值
9	q value	p value 校正的 q 值

8 MRPP 【MRPP 分析】

stat_mrpp.txt 【MRPP 分析结果】

列数	列标题	含义
1	Group	比较的两组
2	A	大于 0 说明组间差异大于组内差异，小于 0 说明组内差异大于组间差异
3	observed-delta	值越小说明组内差异小
4	expected-delta	值越大说明组间差异大
5	Significance	Significance 值小于 0.05 说明差异显著

9 Network 【Network 分析结果】

cytoplot.csv

列数	列标题	含义
1	source	源物种
2	source_group	源物种所在的门
3	target	目标物种
4	target_group	目标物种所在的门
5	coefficient	相关系数

edge.xls

列数	列标题	含义
1	source	源物种
2	target	目标物种

3	value	相关系数
---	-------	------

node.xls

列数	列标题	含义
1	name	network 点文件参与物种
2	group	network 点文件参与物种所在的门
3	score	network 点文件参与物种系数

Network 分析，默认每个分组不少于六个生物学重复时方可出图。

10 NMDS 【NMDS 分析结果】

10.1 NMDS.{pdf,png} 【NMDS 分析结果图】

10.2 NMDS_2.{pdf,png} 【未标样品名称的 NMDS 图】

10.3 NMDS_with_cluster.{pdf,png} 【带有置信椭圆标注样品名称的 NMDS 图】

10.4 NMDS_with_cluster2.{pdf,png} 【带有置信椭圆未标样品名称的 NMDS 图】

10.5 NMDS_scores.txt 【各样品在前两个主成分轴上的位置坐标】

11 PCA 【PCA 分析结果】

11.1 PCA12.{pdf,png} 【PCA 分析结果图】

11.2 PCA12_2.{pdf,png} 【未标样品名称的 PCA 图】

11.3 PCA12_with_cluster.{pdf,png} 【带有置信椭圆标注样品名称的 PCA 图】

11.4 PCA12_with_cluster2.{pdf,png} 【带有置信椭圆未标样品名称的 PCA 图】

11.5 pca.csv 【PCA 作图数据】

列数	列标题	含义
----	-----	----

1	样本名称	各样本名称
2	Axis1	1 维坐标
3	Axis2	2 维坐标
4	Axis3	3 维坐标
5	Axis4	4 维坐标
6	Axis5	5 维坐标

12 PCoA 【PCoA 分析】

12.1 unweighted_unifrac 【基于非加权 unifrac 距离的 PCoA 分析结果】

12.1.1 emperor_pcoa_plots 【三维 PCoA 网页版展示】

12.1.2 PCoA12-2.{pdf,png} 【标有样品名称 PCoA 图】

12.1.3 .{pdf,png} 【未标样品名称的 PCoA 图】

12.1.4 .{pdf,png} 【带有置信椭圆标有样品名称 PCoA 图】

12.1.5 .{pdf,png} 【带有置信椭圆未标样品名称的 PCoA 图】

12.1.6 PCoA.txt 【各个主成分的分析结果】

列数	列标题	含义
1	pc	各样本名称
2	V1	第一主成分对对样本差异的贡献百分比
3	V2	第二主成分对对样本差异的贡献百分比
4	V3	第三主成分对对样本差异的贡献百分比

12.2 weighted_unifrac 【基于加权 unifrac 距离的 PCoA 分析结果】

相关补充可参考 12.1

12.3 (un)weighted_unifrac_dm.txt 【用于 PCoA 分析的(un)weighted unifrac 距离矩阵】

12.4 (un)weighted_unifrac_pc.txt 【PCoA 分析主成分信息】

13 t.test_bar_plot 【门纲目科属种 (p,c,o,f,g,s) 水平上的组间差异显著的物种分析】

13.1 *-vs-*.mat 【两样本 t-test 数据文件】

13.2 *-vs-*.psig.xls 【两样本 t-test 数据结果】

列数	列标题	含义
1	Taxa	物种
2	avg(Z1)	Z1 组均值
3	sd(Z1)	Z1 组标准差
4	avg(Z2)	Z2 组均值
5	sd(Z2)	Z2 组标准差
6	p.value	p 值
7	q.values	q 值
8	interval lower	置信区间
9	interval upper	

13.3 *-vs-*.psig.xls.{svg,png} 【组间差异显著的物种分析条形图】

13.4 *-vs-*.qsig.xls 【从组间差异显著的物种分析结果中 ,筛选出的 p value \leq 0.05 的信息】

13.5 *-vs-*.test.xls 【组间差异显著的物种分析结果】

14 Random_Forest 【门纲目科属种 (p,c,o,f,g,s) 水平上的随机森林】

此项分析要求样品至少有两个分组，每个分组至少 15 个样品（否则无此分析）。