

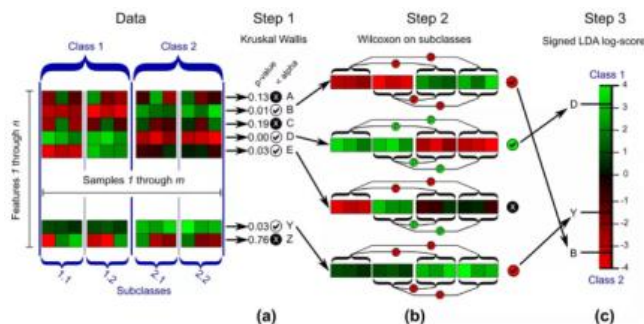
LefSe云工具操作手册

目录

1.LefSe分析简介.....	2
2.打开云工具.....	2
3.上传数据.....	3
4.Lda Effective Size (LEfSe)差异计算.....	4
5.进化分支图的绘制.....	5
6.分布柱状图的绘制.....	6
7.文献应用.....	7

1.LEfSe分析简介

LEfSe分析即Linear discriminant analysis Effect Size分析，是一种用于发现和解释高维度数据生物标志物的分析工具，可以进行两个或多个分组的比较，它强调统计意义和生物相关性，能够在组与组之间寻找具有统计学差异的生物标志物（Biomarker）。



图LEfSe分析原理

A.首先在多组样本中采用的非参数因子Kruskal-Wallis秩和检验检测不同分组间丰度差异显著的物种；B.再利用Wilcoxon秩和检验检查在显著差异物种类中的所有亚种比较是否都趋同于同一分类级别；C.最后用线性判别分析（LDA）对数据进行降维和评估差异显著的物种的影响力（即LDA score）。

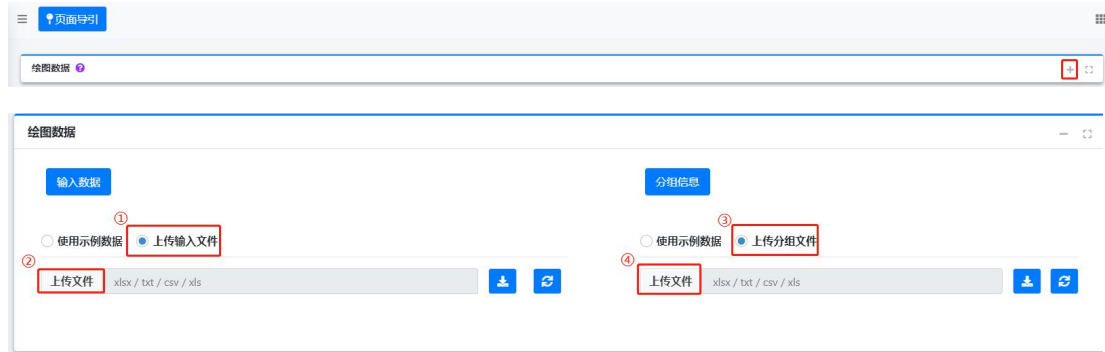
2.打开云工具

首先打开联川生物云平台circos云工具：<https://www.omicstudio.cn/tool/>选择LefSe分析工具，然后进入分析页面，选择左侧**开始绘制**按钮。



3.上传数据

打开工具是默认使用示例数据，供测试和参考，如果需要使用自己的数据进行图片绘制，需要打开页面上方绘图数据的+，展开数据上传模块，然后分别上传输入数据和分组信息共两种表格。



输入数据表格格式如下：

第一列为序列ID，中间列为样本丰度，最后列为物种注释信息，在联川生物进行16S/ITS 扩增子测序的同学可以直接使用结果文件中feature_table_with_taxonomy.xlsx 表格。

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	feature-ID	A_1	A_2	A_3	A_4	A_5	A_6	Taxonomy													
2	f8e9e27a6f81c83d22b011b6f52238	66	182	1239	426	153	104	Fungi;ip__Ascomycotax__Dothideomycetes;o__Pleosporales;f__Pleosporales;Incertae_sedis;g__Phomas__Phoma_radicina													
3	3417521ae3b49cc4051edd6b4a65aa1	9092	10185	11124	17528	7253	5776	Fungi;ip__Ascomycotax__Dothideomycetes;o__Capnodiales;f__Mycosphaerellaceae;g__Davidiella__Davidiella_undclassified													
4	bf151ecd46746906d571de57d8085c1	36	36	456	427	138	27	d__Fungi;ip__Ascomycotax__Leotiomycetes;o__Helotiales;f__Helotiales;Incertae_sedis;g__Tetradiadiums__Tetradiadium_undclassified													
5	5649b47f432703d6472e14fc72d82a1	6382	8153	3574	8040	4632	4284	d__Fungi;ip__Ascomycotax__Dothideomycetes;o__Pleosporales;f__Pleosporales;f__Pleosporaceae;g__Alternaria;f__Alternaria_alternata													
6	c210b262ebcf0ff95959d393ec3613e	394	4908	5914	4715	4668	18522	d__Fungi;ip__Ascomycotax__Dothideomycetes;o__Pleosporales;f__Didymellaceae;g__Atradiidymella;f__Didymella_rabiei													
7	772d5b8ec29ebd7c753163992fc85a0	1625	3092	3962	1697	2322	2859	d__Fungi;ip__Ascomycotax__Sordariomycetes;o__Hypocerales;f__Nectriaceae;g__Gibberellias__Fusarium_tricinctum													
8	e017eed7488625568b6236ce4a954	17090	19369	4110	799	9447	5980	d__Fungi;ip__Ascomycotax__Dothideomycetes;o__Pleosporales;f__Pleosporales;fam__Incertae_sedis;g__Phaeomycoentrospora;f__Phaeomyco													
9	ce60e0d48a1f87cb42ebc084c89f356	6472	3384	5269	2965	4459	4436	d__Fungi;ip__Ascomycotax__Sordariomycetes;o__Hypocerales;f__Nectriaceae;g__Gibberellias__Fusarium_of_equiseti_MY_2011													
10	d54a99defb7864aa704890e0302a722	951	5762	1715	3348	1394	5657	d__Fungi;ip__Ascomycotax__Dothideomycetes;o__Capnodiales;f__Mycosphaerellaceae;g__Davidiella__Davidiella_undclassified													
11	7e20b055f45912a8028807602e898c9	420	1399	0	2394	4614	8422	d__Fungi;ip__Ascomycotax__Dothideomycetes;o__Dothideomycetes_undclassified;f__Dothideomycetes_undclassified;g__Dothideomycetes_undclas													
12	3c3280c119b3d6eaf954a9731377af2	13	37	212	138	112	34	d__Fungi;ip__Ascomycotax__Leotiomycetes;o__Helotiales;f__Helotiales;Incertae_sedis;g__Tetradiadiums__Tetradiadium_maxilliforme													
13	db8962e6d13e3fba7a3e90af60a5f	102	347	465	206	641	972	d__Fungi;ip__Ascomycotax__Dothideomycetes;o__Pleosporales;f__Leptosphaeriaceae;g__Paraphomas__Paraphoma_chrysanthemicola													
14	db3c1bb38cf53af8885ced979b8eac	1054	375	768	579	397	3064	d__Fungi;ip__Ascomycotax__Sordariomycetes;o__Sordariomycetidae;Incertae_sedis;f__Plectosphaerellaceae;g__Plectosphaerellias__Plectospha													

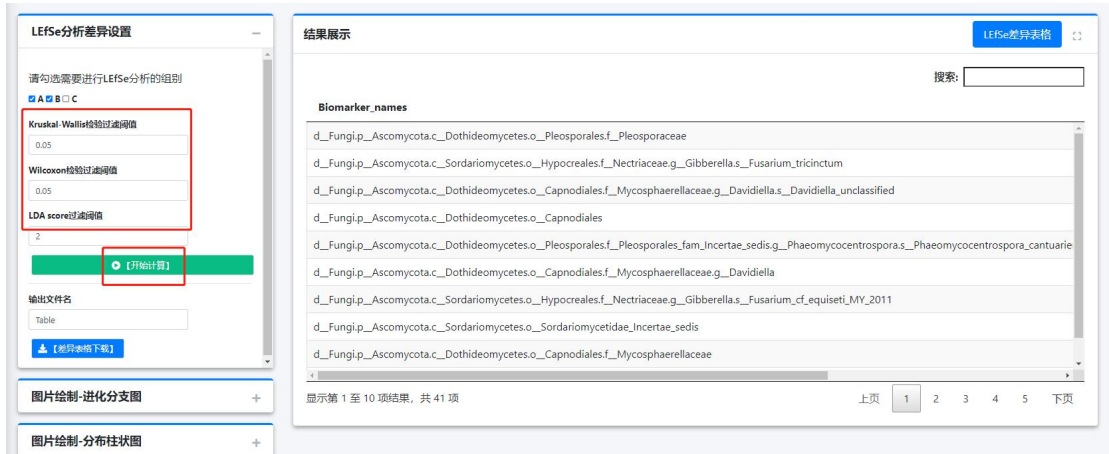
分组信息表格格式如下：

第一列为样本名称，第二列为分组名称，在联川生物进行16S/ITS 扩增子测序的同学可以直接使用结果文件中sample_map.xlsx 表格。

	A	B
1	#SampleID	Group
2	A_1	A
3	A_2	A
4	A_3	A
5	A_4	A
6	A_5	A
7	A_6	A
8	A_7	A
9	A_8	A
10	A_9	A
11	B_1	B
12	B_2	B
13	B_3	B
14	B_4	B

4.Lda Effective Size (LEfSe)差异计算

(1) 在左侧设置LEfSe差异分析的阈值，然后点击开始计算，即可得到LEfSe分析结果表格。



表头说明：

Biomarker_names: Biomarker名称；

Log_value: 各组分丰度平均值中最大值的log10，如果平均丰度小于10的按照10来计算；

Groups: 物种富集的组名；

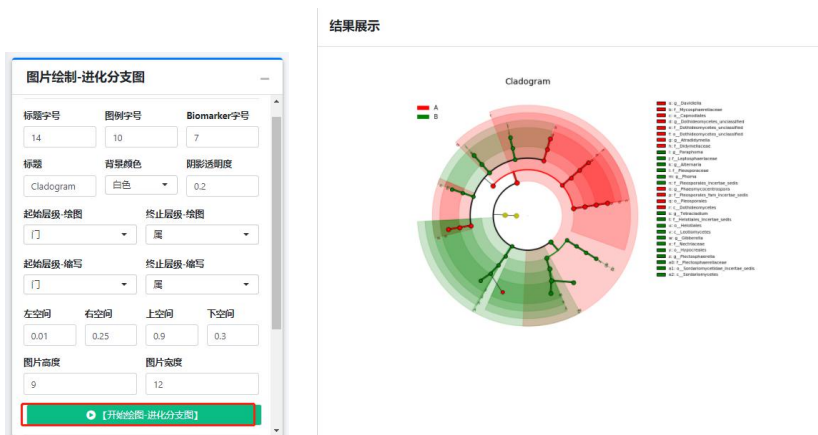
LDA_values: LDA值；

P_value: Kruskal-Wallis秩和检验的p值。

(2) 点击**差异表格下载**，即可下载此表。

5.进化分支图的绘制

(1) 设置左侧进化分支图的相关参数，然后点击**开始绘图**，稍作等待，即可完成图片绘制。



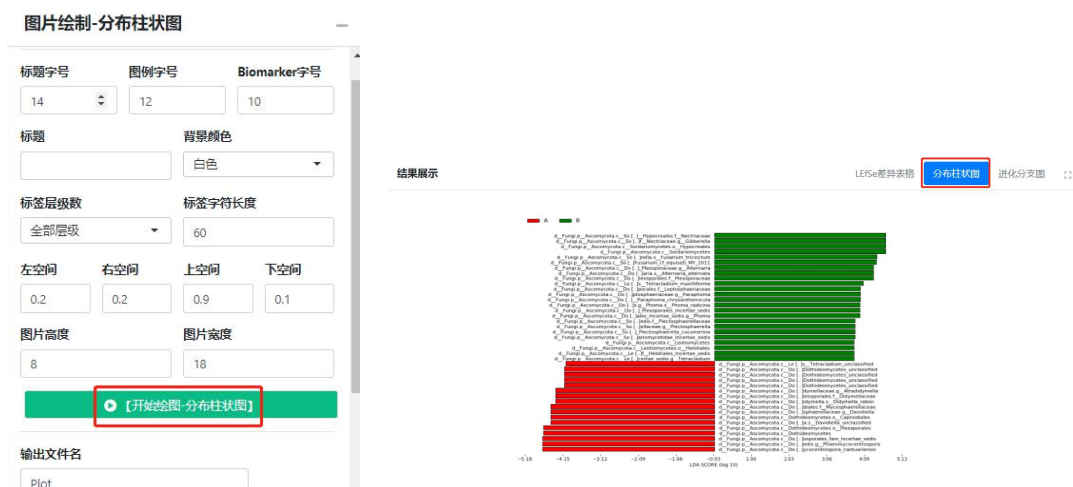
图片说明：图中由内至外辐射的圆圈代表了由界（单个圆圈）至属（或种）的分类级别。不同分类级别上的每一个小圆圈代表该水平下的一个分类，小圆圈直径大小与相对丰度大小呈正比；颜色：无显著差异的物种统一着色为黄色，差异显著的物种Biomarker跟随组别进行着色，红色节点表示在红色组别中起到重要作用的微生物类群，蓝色节点表示在蓝色组别中起到重要作用的微生物类群。未能在图中显示的Biomarker对应的物种名会展示在右侧，字母编号与图中对应。

(2) 点击下方图片下载，即可保存图片。



6.分布柱状图的绘制

(1) 设置左侧分布柱状图的相关参数，然后点击**开始绘图**，稍作等待，即可完成图片绘制。



图片说明：条形图主要展示了LDA值大于预设值的显著差异物种（less_strict设为2；more_strict 设为4），即具有统计学差异的Biomarker，默认预设值为3（看横坐标，只有LDA值的绝对值大于3才会显示在图中）；柱状图的颜色代表各自的组别，长短代表的是LDA值，即不同组间显著差异物种的影响程度。

(2) 点击下方图片下载，即可保存此图。



7.文献应用

Integrated microbiome and metabolome analysis reveals a novel interplay between commensal bacteria and metabolites in colorectal cancer

Yongzhi Yang^{1,2,3*}, Biswapriya B. Misra^{4*}, Lei Liang^{1,2*}, Dexi Bi⁵, Wenhao Weng^{6,7}, Wen Wu³, Sanjun Cai^{1,2}, Huanlong Qin³, Ajay Goel⁸, Xinxiang Li^{1,2}, Yanlei Ma^{1,2,9}

发表期刊: Theranostics

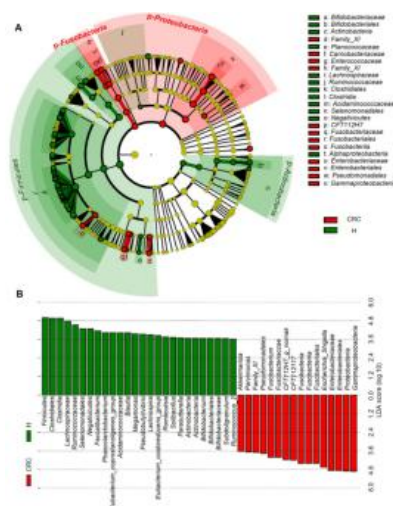
影响因子: 8.063

发表时间: 2019

研究内容: 鉴别健康志愿者和结直肠癌 (CRC) 患者的特征性微生物群以及具有疾病表型的相关代谢物

样本数量: 50 CRC vs 50 healthy volunteers; 粪便样本

实验方法: 16S rRNA基因测序+非靶向GC-MS代谢组



研究者使用LefSe生成进化分枝图 (A) 和LDA值分布柱状图 (B) 鉴定CRC特异的细菌。其中76个差别的OTUs可作为关键的鉴别指标。在CRC组中, 多种条件致病菌包括 *Gammaproteobacteria* (γ -变形菌纲, 属于变形菌门), *Enterobacteriaceae* (肠杆菌科, 属于肠杆菌目) 和 *Fusobacteriales* (梭杆菌目, 属于梭杆菌门) 均显著高丰度。在H组中, *Firmicutes* (厚壁菌门)、*Clostridiales* (梭菌目)、*Clostridia* (梭状芽孢杆菌)、*Lachnospiraceae* (毛螺菌科)、*Ruminococcaceae* (瘤胃球菌科)、*Selenomonadales*、*Negativicutes*和 *Faecalibacterium* (粪杆菌属) 是最丰富的微生物群。