

LEfSe 在线分析教程

LEfSe (LDA Effect Size) 分析，可以用于两个或多个分组之间的比较，从而找到组间有显著性差异的物种（即 biomarker），分析步骤主要分为三步：

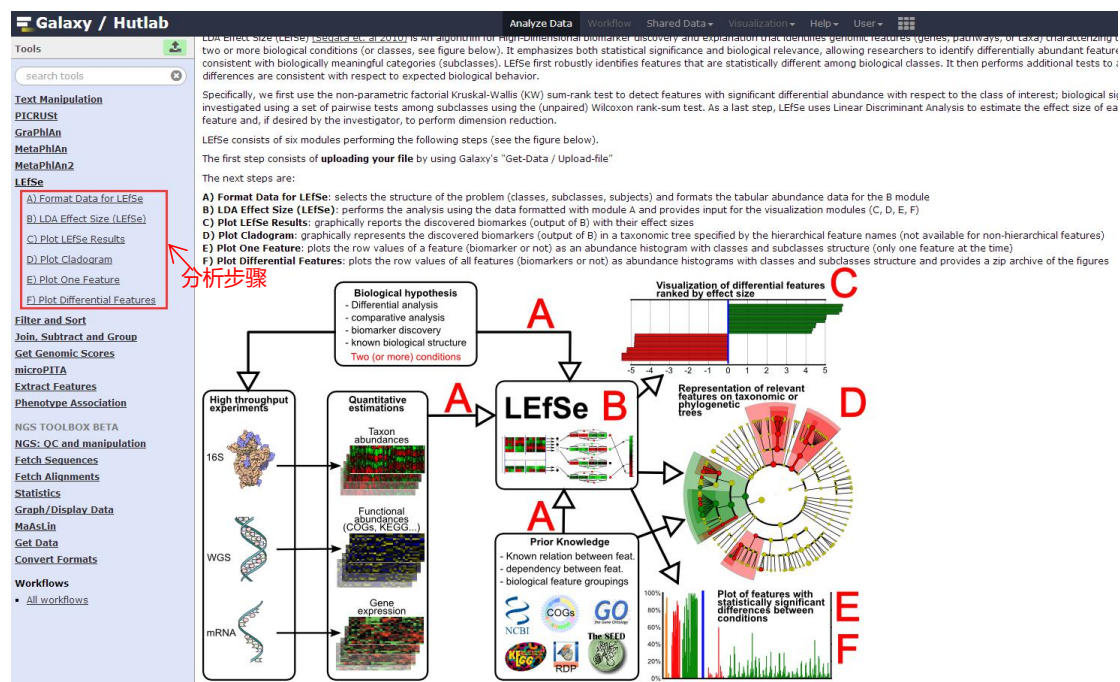
Step1：利用 Kruskal-Wallis 秩和检验检测所有的特征物种，通过检测不同组间的物种丰度差异，获得显著性差异物种。

Step2：再利用 Wilcoxon 秩和检验检测上步获得的显著性差异物种的所有亚种是否都趋于同一分类级别。

Step3：最后用线性判别分析（LDA），得到最终的差异物种（即 biomarker）。

LEfSe 在线分析网址：<http://huttenhower.sph.harvard.edu/galaxy/>

网页界面如下：



1、数据格式

一般微生物物种分析结果中都能得到 OTU 在不同分类水平的物种丰度结果，将数据格

式转化为下图所示的网站标准输入格式，即可进行后续的分析。

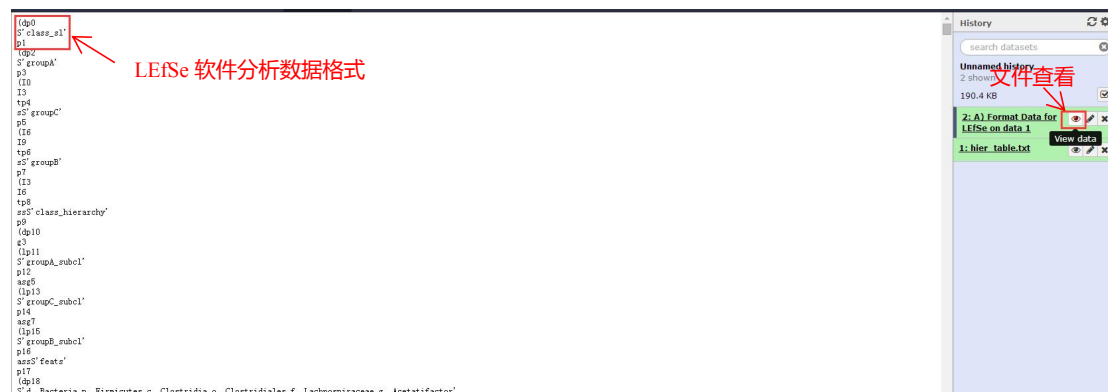
| Bodysite 样品分组行名 | | | | | | | | |
|--------------------------------|-----------|-------------|------------|-----------|-------------|-------------|-------------|--|
| body site | mucosal | mucosal | mucosal | mucosal | non_mucosal | non_mucosal | non_mucosal | |
| subsite | oral | oral | oral | gut | skin | nasal | skin | |
| id | 1023 | 1023 | 1672 | 1672 | 159005010 | 1023 | 1023 | |
| Bacteria | 0.99999 | 0.99999 | 0.999993 | 0.999989 | 0.999997 | 0.999977 | 0.999987 | |
| Bacteria Actinobacteria | 0.311037 | 0.000864363 | 0.00446132 | 0.0312045 | 0.000773642 | 0.359354 | 0.603002 | |
| Bacteria Bacteroidetes | 0.0689602 | 0.804293 | 0.00963343 | 0.0303561 | 0.859838 | 0.0195298 | 0.0212741 | |
| Bacteria Firmicutes | 0.494223 | 0.173411 | 0.715345 | 0.813046 | 0.124552 | 0.177961 | 0.189178 | |
| Bacteria Proteobacteria | 0.0914284 | 0.0180378 | 0.265664 | 0.109549 | 0.00941215 | 0.430869 | 0.0225884 | |
| Bacteria Firmicutes Clostridia | 0.090041 | 0.170246 | 0.00483188 | 0.0465328 | 0.122702 | 0.0402301 | 0.0460614 | |

2、上传数据

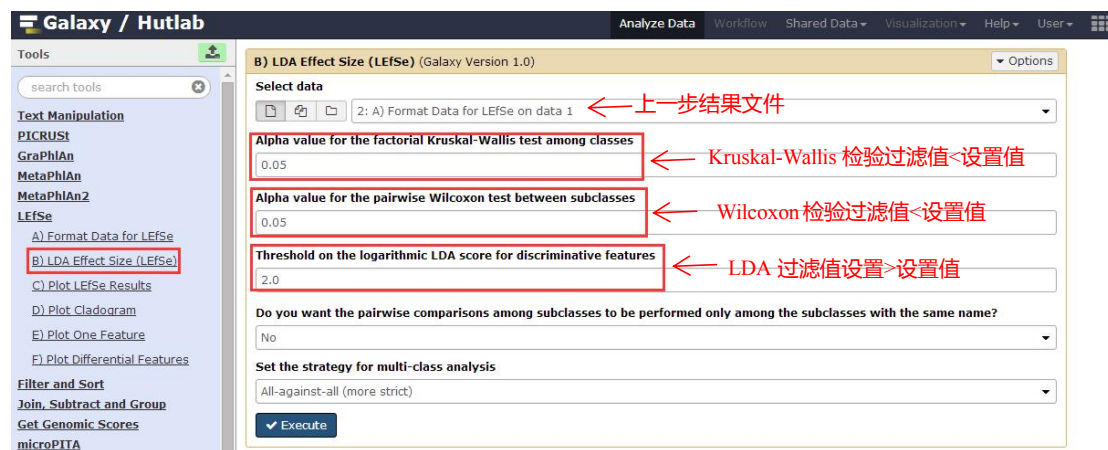
将标准格式的文件上传至网站：（Get Data->Upload File），具体设置如下图：

3、LEfSe 分析-步骤 A) Format Data for LEfSe 格式转化

输出文件格式（该格式文件只是中间文件，具体意义不需要详细理解）如下：



4、LEfSe 分析-步骤 B) LDA Effect Size (LEfSe) 差异计算与统计



输出结果：

```
d_Bacteria.p_Firmicutes.o_Bacilli.o_Lactobacillales.f_Lactobacillaceae.g_Lactobacillus.s_uncultured_Lactobacillus_sp_ 2.96913507904 -
d_Bacteria.p_Proteobacteria.o_Epsilonproteobacteria.o_Campylobacteriales.f_Helicobacteraceae.g_Helicobacter.s_Helicobacter_guayanae 1.90754877721 -
d_Bacteria.p_Proteobacteria.o_Bacteroidia.o_Bacteroidales.f_Prevotellaceae.g_Prevotella.s_Prevotella_unclassified 4.78542282878 -
d_Bacteria.p_Proteobacteria.o_Deltaproteobacteria.o_Desulfobibrionales.f_Desulfobibrionaceae.g_Desulfobibrion.s_uncultured_Desulfobibrion_sp_ 1.90627010197 -
d_Bacteria.p_Landisius.saccharibacteria.o_Saccharibacteria_genera_incertae_sedis.o_Saccharibacteria_genera_incertae_sedis.f_Saccharibacteria_genera_incertae_sedis 3.40003712696 groupB
0.0273315396 0.027332724473
d_Bacteria.p_Lentisphaerae 2.10308505866 -
d_Bacteria.p_Firmicutes.o_Clostridia.o_Clostridiales.f_Lachnospiraceae.g_Paraproteobacterium.s_Paraproteobacterium_unclassified 2.62164049461 -
d_Bacteria.p_Firmicutes.o_Clostridia.o_Clostridiales.f_Lachnospiraceae.g_Hespellia.s_uncultured_Hespellia_sp_ 2.66322086159 -
d_Bacteria.p_Proteobacteria.o_Gammaproteobacteria.o_Alteromonadales.f_Shewanellaceae 1.60651878154 -
d_Bacteria.p_Firmicutes.o_Clostridia.o_Clostridiales.f_Peptostreptococcaceae.g_Filifactor 3.3604965082 -
d_Bacteria.p_Firmicutes.o_Clostridia.o_Clostridiales.f_Lachnospiraceae.g_Acetatifactor.s_Acetatifactor_unclassified 3.91513410947 -
d_Bacteria.p_Bacteroidetes.o_Bacteroidia.o_Bacteroidales.f_Porphyromonadaceae.g_Odoribacter.s_Odoribacter_unclassified 2.69693290971 groupB 2.73424451532 0.0339856065141
d_Bacteria.p_Actinobacteria.o_Actinobacteria.o_Coriobacteriales.f_Coriobacteriaceae.g_Olsenella 1.77976382916 -
```

各列意义如下：

第一列：物种信息

第二列：平均丰度最大的 log10 的值，如果平均丰度小于 10 按照 10 来计算

第三列：差异物种富集的组名称

第四列：LDA 值

第五列：Kruskal-Wallis 秩和检验的 P 值，如果不是 biomarker，则用 '-' 表示

5、LEfSe 分析 C-F 步骤均为作图

步骤 C) Plot LEfSe Results

Galaxy / Hutlab | Analyze Data | Workflow | Shared Data | Visualization | Help | User

Tools

search tools

Text Manipulation

PICRUSt

GraPhlAn

MetaPhlAn

MetaPhlAn2

LEfSe

A) Format Data for LEfSe

B) LDA Effect Size (LEfSe)

C) Plot LEfSe Results

D) Plot Cladogram

E) Plot One Feature

F) Plot Differential Features

Filter and Sort

C) Plot LEfSe Results (Galaxy Version 1.0)

Select data: 3: B) LDA Effect Size (LEfSe) on data 2

Set text and label options (font size, abbreviations, ...): Default

Set some graphical options to personalize the output: Default

Output format: png

Set the dpi resolution of the output: 150

Execute

输入文件，第四步结果文件

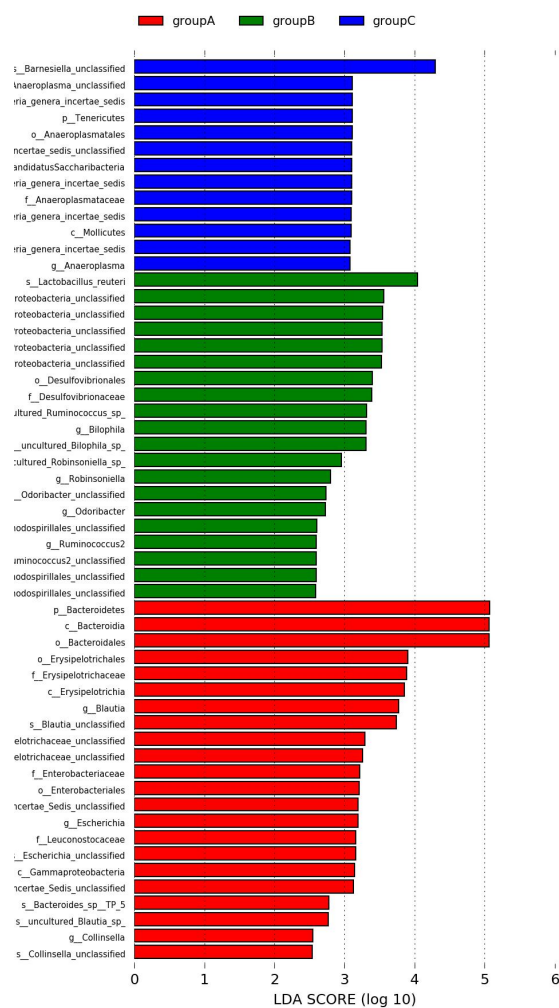
Text 和 label 参数 根据需要设置字体等，选择 Advanced 进行具体参数设置

图形参数设置

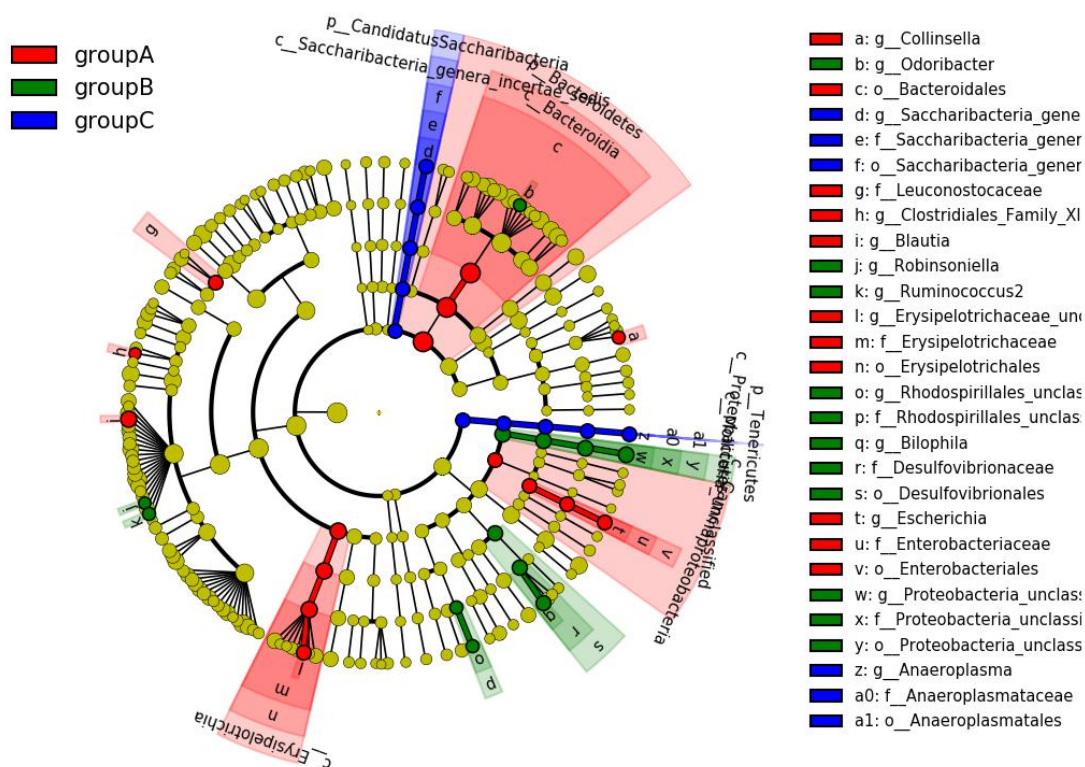
输出图形格式设置：png, pdf, svg

输出图形分辨率

输出图形结果如下（如果差异过多或过少，可重新设置第四步参数并运算，然后作图）：



步骤 D) Plot Cladogram (参数设置与步骤 C 类似)



图注：不同圆圈表示不同分类层级，从内至外，依次为门-纲-目-科-属。每个节点表示一种物种，黄色表示该物种在三组中无显著性差异，其他颜色，以 groupA 中的红色为例，如果节点颜色为红色则表示该物种在比较组中有显著性差异，且该物种在 groupA 中的丰度更高，其他颜色以此类推。具体差异的门和纲直接在左侧图中标出，其他差异以字母表示，具体代表的物种在右侧标出。

步骤 E) Plot One Feature (画其中一个物种在不同组样品中的柱状图)

Galaxy / Hutlab Analyze Data Workflow Shared Data Visualization Help User

Tools

search tools

Text Manipulation

PICRUSt

GraPhlAn

MetaPhlAn

MetaPhlAn2

LEfSe

A) Format Data for LEfSe

B) LDA Effect Size (LEfSe)

C) Plot LEfSe Results

D) Plot Cladogram

E) Plot One Feature

F) Plot Differential Features

Filter and Sort

Join, Subtract and Group

Get Genomic Scores

microPITA

Extract Features

Phenotype Association

E) Plot One Feature (Galaxy Version 1.0) Options

The formatted datasets

2: A) Format Data for LEfSe on data 1 ← 第三步结果文件

The LEfSe output

3: B) LDA Effect Size (LEfSe) on data 2 ← 第四步结果文件

Select the feature names among biomarkers or all features

Biomarkers only ← 只展示 biomarker 或展示所有

Select the feature to plot ← 选择其中一个 biomarker 进行作图

d__Bacteria.p__Acti.ceae.g__Collinsella

Set some graphical options to personalize the output

Default

Output format

png

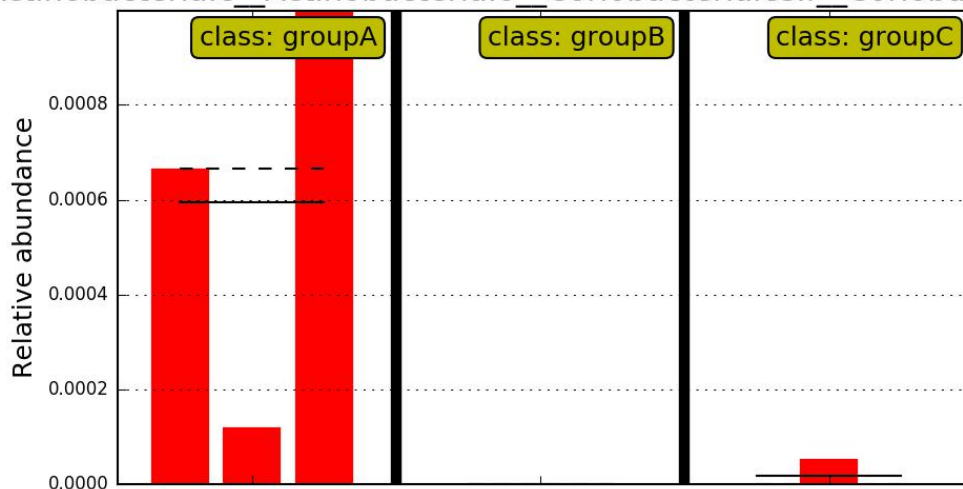
Set the dpi resolution of the output

150

☒ Execute

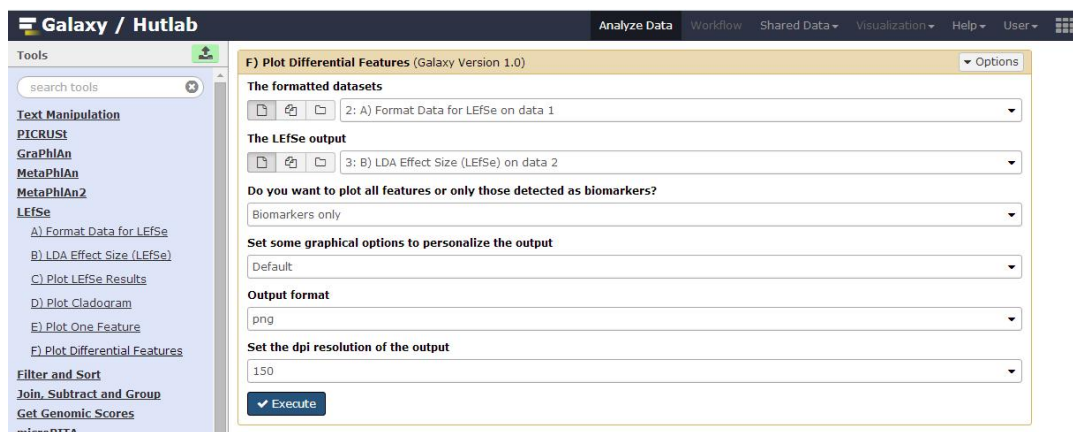
结果展示：

Actinobacteria.c_Actinobacteria.o_Coriobacteriales.f_Coriobacteriace



图注：图中不同分组用黑实线隔开，每组柱状图中的实线表示该组样品表达量的平均值，虚线代表该组样品表达量的中位值。

步骤 F) Plot Differential Features (绘制差异物种或基因柱状图，输入设置与 E 基本类似)



步骤 E 和 F 的区别在于，E 每次只能做一张图，F 可以绘制 biomarker（或所有的物种）柱状图。一般建议跳过步骤 E，直接做步骤 F，步骤 F 的结果一般以压缩包的形式，下载到本地解压后即可查看每个 biomarker 在不同样品中的表达柱状图。

6、结果下载

