



# LefSe介绍 及云平台使用

# LefSe软件介绍

- LefSe分析是LDA Effect Size分析，是一种进行高维度数据(基因、通路和微生物分类单元等)的分析工具，可以进行两个或多个分组的比较，能够在组间找到具有统计学差异的biomarker。目前常用于微生物多样性分析。
- LefSe有多个版本，既可以下载本地运行，也可以在官网上运行，又可以利用云平台线上运行。

<https://bitbucket.org/nsegata/lefse/src/default/>

# LefSe安装

- LefSe在Linux安装时可以首先安装hg软件以后再下载LefSe到具体地址  
linux版本可以直接使用如下命令克隆，完成后文件如下：

```
hg clone https://bitbucket.org/nsegata/lefse
```

- LefSe还需安装requirement.txt里面的R和Python所需要的相关安装包

- **R libraries:**

splines, stats4, survival,  
mvtnorm, modeltools, coin, MASS

- **python libraries:**

rpy2, **numpy**, **matplotlib**,  
argparse

```
example
format_input.py
__init__.py
lefse2circlader.py
lefsebiom
lefse.py
lefse.pyc
license.txt
plot_cladogram.py
plot_features.py
plot_res.py
qiime2lefse.py
requirements.txt
run_lefse.py
```

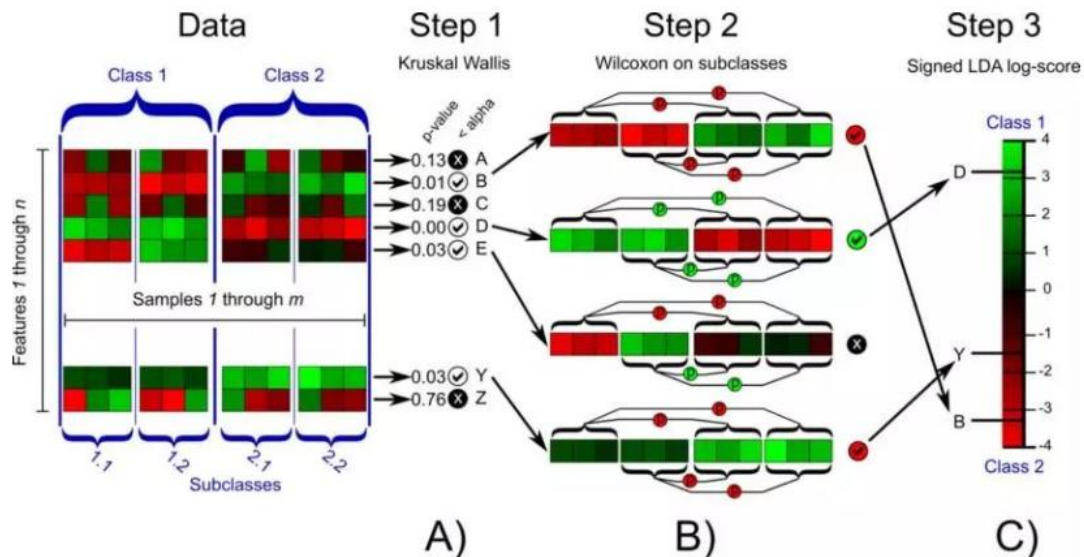
# LefSe使用

**输入数据：** 该软件支持不同类型的带有分组信息的高维度数据，一般数据包括样品的具体名称，分组信息，亚组分组信息(可以无)以及具体生物标记在不同样品中的丰度等数据信息。

Group	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B
OTU ID	L1S105	L1S140	L1S208	L1S257	L1S281	L1S57	L1S76	L1S8	L2S155	L2S175	L2S204	L2S222	L2S240	L2S309	L2S357	L2S382
k_Bacteria;p_Firm	0	0	12	2	0	5	0	30	386	933	133	148	659	230	566	670
k_Bacteria;p_Firm	914	722	521	256	589	700	906	130	0	7	19	12	0	18	10	0
k_Bacteria;p_Bact	4572	4661	4628	3043	3113	4570	4837	4581	13	10	31	66	3	37	14	0

# LefSe原理

首先在多组样品中采用Kruskal-Wallis秩和检验和Wilcoxon秩和检验检测不同分组间丰度差异显著的biomarker及其亚种是否趋同于同一分类级别，然后使用线性判别分析LDA对数据进行降维和评估差异显著biomarker的LDA值。



# LefSe参数设置

**参数：** 选择feature数据信息是按照行还是列； 分组信息所在行； 秩和检验的alpha值； 亚分组信息所在行； 样品所在行； 预设的LDA过滤值等；

\* set whether the features are on rows or on columns

\* set which feature use as class

\* set which feature use as subclass, -1 means no subclass

\* set which feature use as subject, -1 means no subject

\* set the normalization value

\* set the multiple testing correction options

\* set the alpha value for the Wilcoxon test

\* for multiclass tasks, set whether the test is performed in a more strict mode, 1 stands for more strict!! Default setting in LefSe website is "All-against-all (more strict)"

\* set the threshold on the absolute value of the logarithmic LDA score

\* set whether perform the wilcoxon test only among the subclasses with the same name

\* set the alpha value for the Anova test

\* number of label levels to be displayed

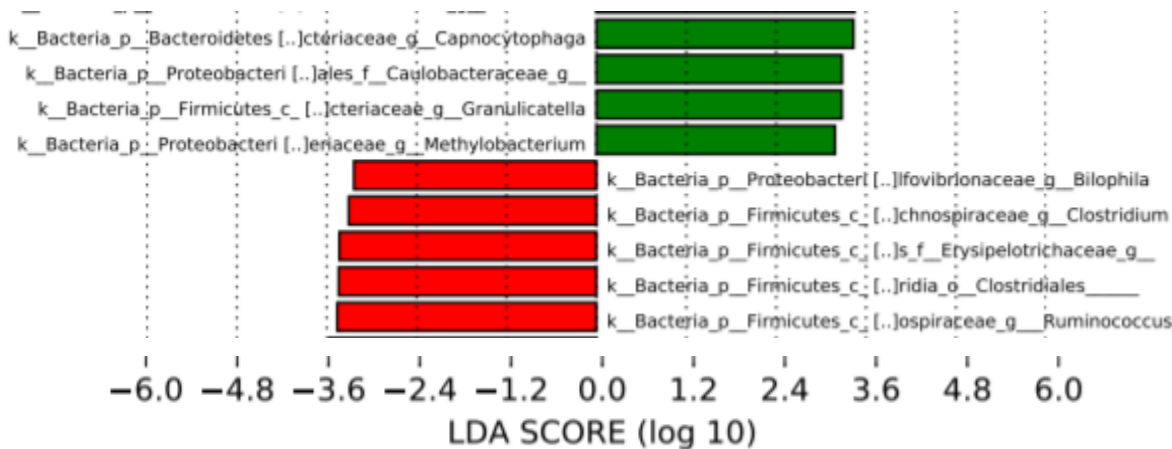
\* dpi for the plot

\* set the number of the feature to plot

set the name of the feature to plot

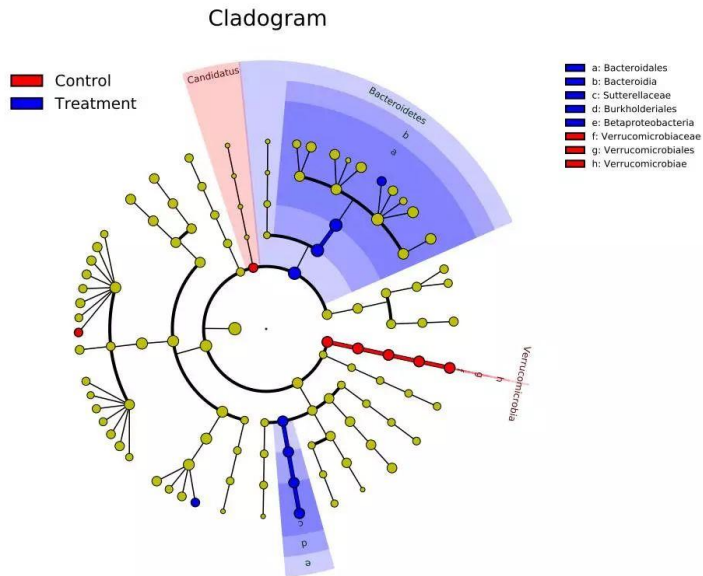
# LefSe结果

**1.LDA值分布柱状图：**该柱状图显示了LDA score大于预设值即具有统计学差异的的biomarker。LDA score是不同组间显著差异标记的影响程度。一般默认预设值为2，根据预设值的情况LDA绝对值大于预设值的会在柱状图中显示。柱状图的颜色代表组别，长短代表LDA score。



# LefSe结果

**2.进化分支图：**若LefSe进行显著差异物种分析，会显示进化分支图。由内向外的圆圈代表了由门到属的不同分类级别。每个分类级别的每个小圆圈代表该分类水平下的具体分类，小圆圈的直径代表其相对丰度。无显著差异的物种颜色为黄色，差异显著的物种根据其组别显示颜色。



# LefSe结果

**3.特征表：** 第一列是Biomarker的名称；第二列是其丰度平均值中最大值的log10；第三列是组别名；第四列是LDA值；第五列是秩和检验的P值，如果不是biomarker用-表示。

Biomarker	Logarithm value	Groups	LDA score	P value
k_Bacteria_p_Bacteroidetes_c_Bacteroidia_o_Bacteroidales_f_S24_7_g__	2.181826911			-
k_Bacteria_p_Verrucomicrobia_c_Verrucomicrobiae_o_Verrucomicrobiales_f_Verrucomicrobiaceae_g_Luteolibacter	1.959978162			-
k_Bacteria_p_Firmicutes_c_Clostridia_o_Clostridiales_f_Lachnospiraceae_g_Coprococcus	3.948834261	A	3.70365428	0.001194
k_Bacteria_p_Firmicutes_c_Clostridia_o_Clostridiales_f_Tissierellaceae_g_ph2	3.016069083			-
k_Bacteria_p_Proteobacteria_c_Betaproteobacteria_o_Burkholderiales_f_Oxalobacteraceae_g_Janthinobacterium	3.810906075	B	3.55334845	0.003776
k_Bacteria_p_Firmicutes_c_Clostridia_o_Clostridiales_f_Peptostreptococcaceae_g__	2.846500192			-
k_Bacteria_p_Firmicutes_c_Clostridia_o_Clostridiales_____	3.191528572	A	3.43954213	0.001476
k_Bacteria_p_Firmicutes_c_Clostridia_o_Clostridiales_f_Ruminococcaceae_g_Faecalibacterium	4.88806155	A	4.57418757	0.00075

# LefSe案例展示

**上传数据：**数据包括样品的分组信息，亚组分组信息(可以无)以及具体名称和具体生物标记在不同样品中的丰度等数据信息。

Group	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B
OTU ID	L1S105	L1S140	L1S208	L1S257	L1S281	L1S57	L1S76	L1S8	L2S155	L2S175	L2S204	L2S222	L2S240	L2S309	L2S357	L2S382
k_Bacteria;p_Firm	0	0	12	2	0	5	0	30	386	933	133	148	659	230	566	670
k_Bacteria;p_Firm	914	722	521	256	589	700	906	130	0	7	19	12	0	18	10	0
k_Bacteria;p_Bact	4572	4661	4628	3043	3113	4570	4837	4581	13	10	31	66	3	37	14	0

# LefSe案例演示

- 国家微生物数据科学中心官网进行云平台使用 (<http://www.nmdc.cn/>)，注意使用前先进行注册、登录



登录

注册

English

首页

数据资源

元数据

数据下载

分析工具

数据汇交

服务案例

关于我们

世界微生物  
数据中心

World Data Centre for Microorganisms



# LefSe案例演示

- 选择分析工具，搜索栏中输入LefSe

当前位置: 首页>分析工具

lelse

搜索示: SOAPdenovo2 / CANU / PfamScan / LefSe /

Blast分析工具 (5)

宏基因组分析流程 (3)

基因组拼接工具 (25)

基因组结构分析 (11)

基因组注释分析 (4)

元基因组分析 (20)

比较基因组分析 (13)

便捷分析工具 (7)

**LEfSe** HOT

原始链接: <https://bitbucket.org...>

工具介绍: LEfSe (线性判别分析效应量) 将统计显著性标...

关键字: biomarker discovery

文章: Segata, N., Izard, J., Waldron, L., Gevers, ...

[查看详情](#) [开始使用](#)

**LEfSe**

原始链接: <https://bitbucket.org...>

工具介绍: LEfSe (线性判别分析效应量) 将统计显著性标...

关键字: biomarker discovery

文章: Segata, N., Izard, J., Waldron, L., Gevers, ...

[查看详情](#) [开始使用](#)

# LefSe案例演示

- 上传选择需要分析的数据

## LEfSe

---

\* input\_file

Select File

# LefSe案例演示

- 流程参数主要选择feature数据信息是按照行还是列；哪一行是分组信息；秩和检验的alpha值；

\* set whether the features are on rows or on columns

 ▲

\* set the multiple testing correction options

 ▼

\* number of label levels to be displayed

\* set which feature use as class

 ▲

\* set the alpha value for the Wilcoxon test

 ▲

\* dpi for the plot

 ▼

# LefSe案例演示

- 流程参数主要选择哪一行是亚分组信息(无选择-1); 哪一行是样品所在行; 预设的LDA过滤值; 秩和检验的alpha过滤值等信息;

\* set which feature use as subclass, -1 means no subclass



\* set which feature use as subject, -1 means no subject



\* set the normalization value

\* for multiclass tasks, set whether the test is performed in a more strict mode, 1 stands for more strict!! Default setting in LefSe website is "All-against-all (more strict)"

\* set the threshold on the absolute value of the logarithmic LDA score



\* set whether perform the wilcoxon test only among the subclasses with the same name

\* set the alpha value for the Anova test



\* set the number of the feature to plot

set the name of the feature to plot

# LefSe案例演示

- LefSe的结果文件：  
bar\_plot.pdf, result.in, result.res

## Results File

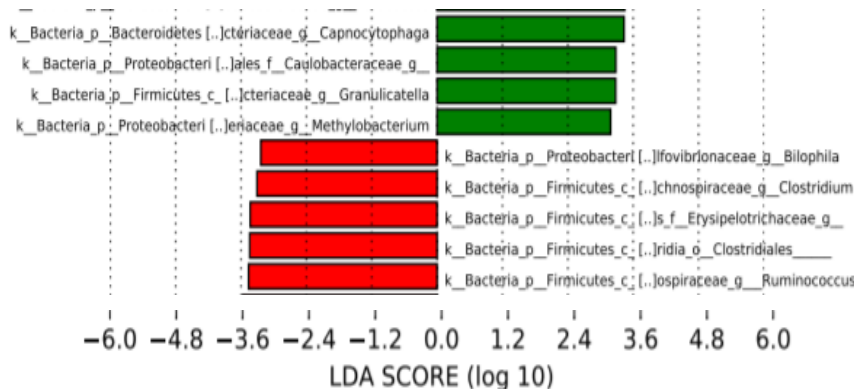
1. bar\_plot.pdf

2. result.in

3. result.res

# LefSe结果

- 主要结果文件： LDA值分布柱状图； 特征表；



Biomarker	Logarithm value	Groups	LDA score	P value
k_Bacteria_p_Bacteroidetes_c_Bacteroidia_o_Bacteroidales_f_S24_7_g_	2.181826911			-
k_Bacteria_p_Verrucomicrobia_c_Verrucomicrobiae_o_Verrucomicrobiales_f_Verrucomicrobiaceae_g_Luteolibacter	1.959978162			-
k_Bacteria_p_Firmicutes_c_Clostridia_o_Clostridiales_f_Lachnospiraceae_g_Coprococcus	3.948834261	A	3.70365428	0.001194
k_Bacteria_p_Firmicutes_c_Clostridia_o_Clostridiales_f_Tissierellaceae_g_ph2	3.016069083			-
k_Bacteria_p_Proteobacteria_c_Betaproteobacteria_o_Burkholderiales_f_Oxalobacteraceae_g_Janthinobacterium	3.810906075	B	3.55334845	0.003776
k_Bacteria_p_Firmicutes_c_Clostridia_o_Clostridiales_f_Peptostreptococcaceae_g_	2.846500192			-
k_Bacteria_p_Firmicutes_c_Clostridia_o_Clostridiales_____	3.191528572	A	3.43954213	0.001476
k_Bacteria_p_Firmicutes_c_Clostridia_o_Clostridiales_f_Ruminococcaceae_g_Faecalibacterium	4.88806155	A	4.57418757	0.00075

感谢观看!!!