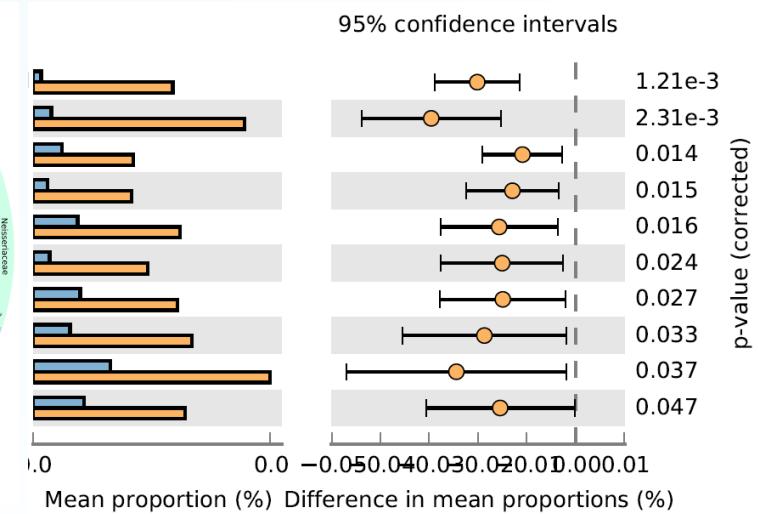
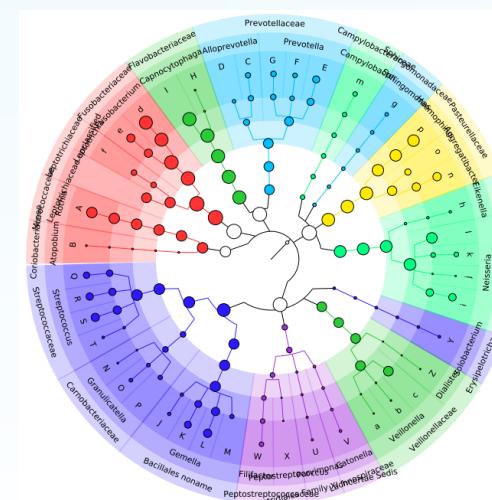


# 微生物组—宏基因组分析专题研讨会第24期



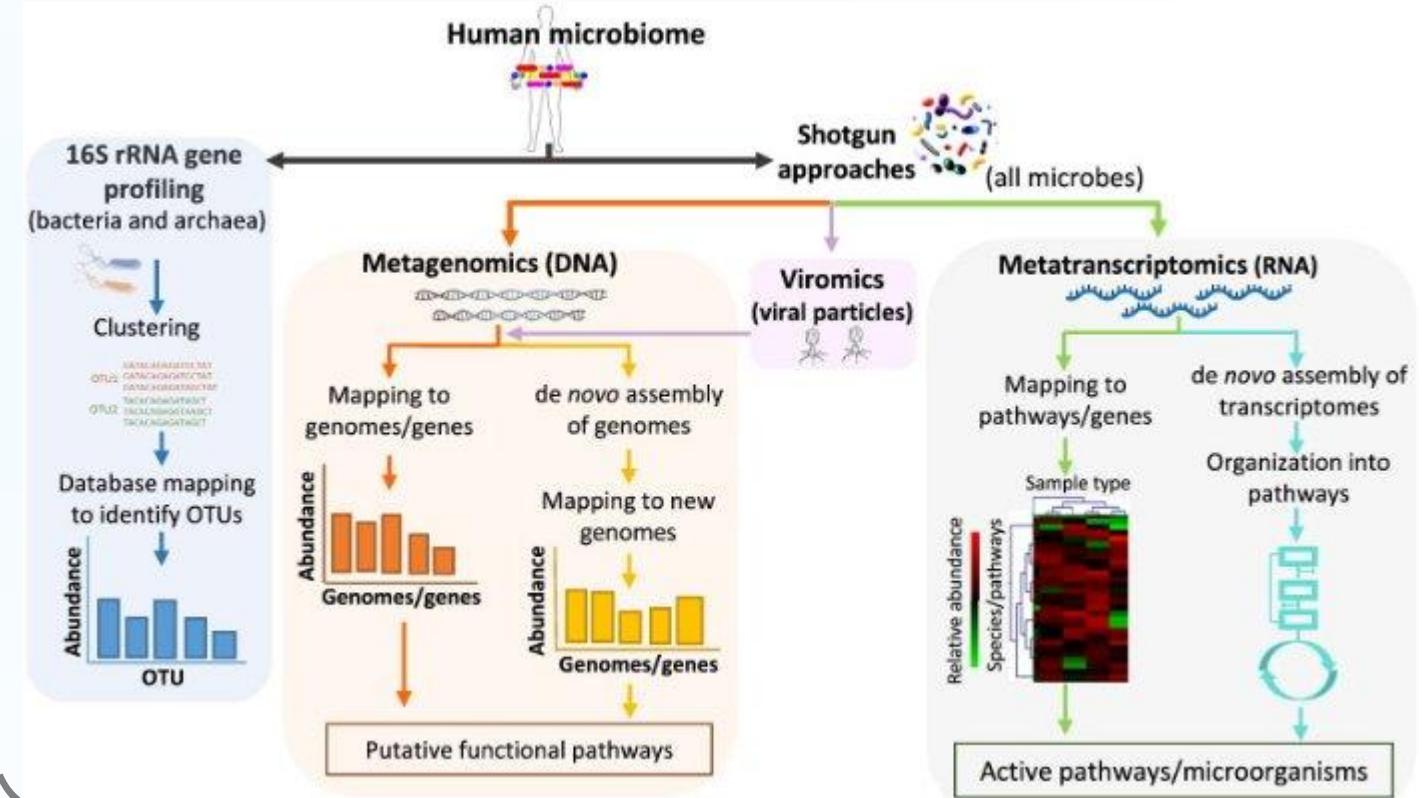
## 24物种和功能组成的 差异比较及可视化

刘永鑫  
2025年11月29日



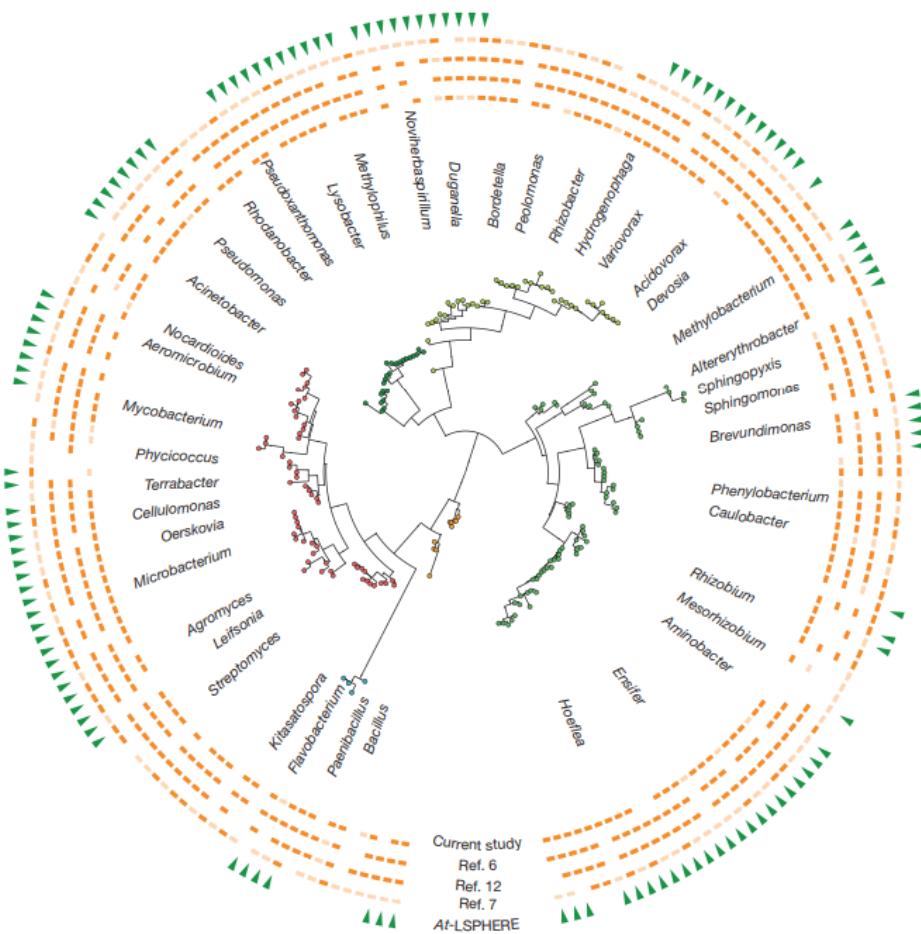
# 宏基因组有参流程——实战分析大纲

- 一. 软件安装和数据库部署
- 二. KneadData质控
- 三. MetaPhlAn4物种组成
- 四. HUMAnN4功能组成
- 五. **GraPhlAn可视化物种**
- 六. LEfSe分析物种差异
- 七. STAMP功能组成统计分析

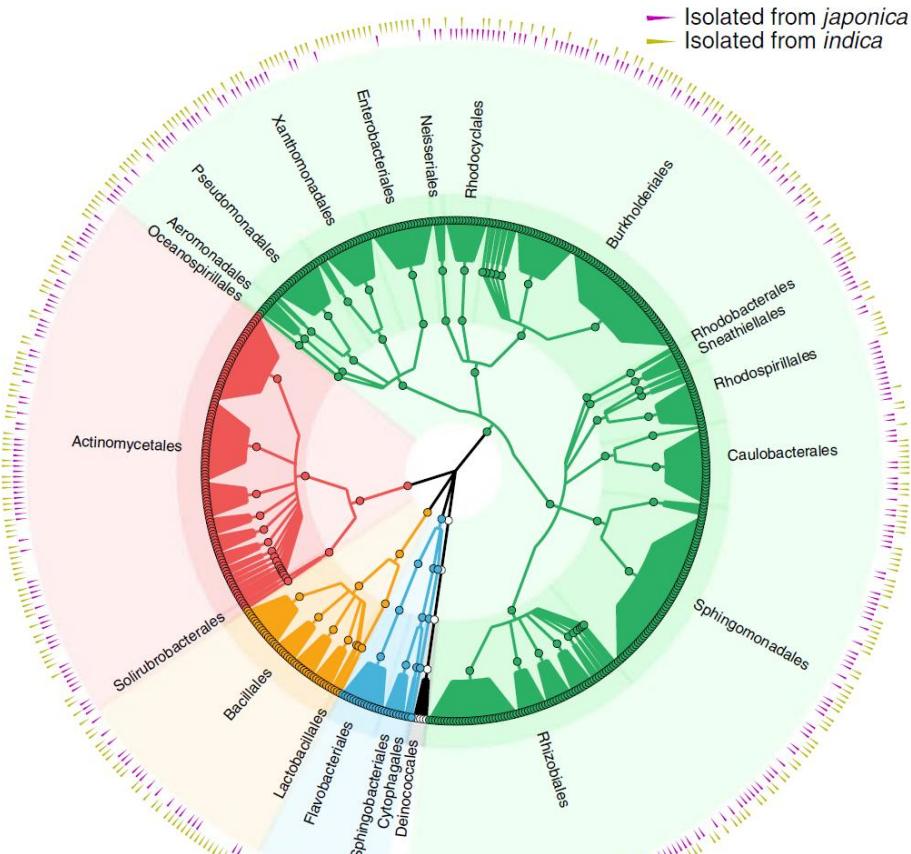


# 我们经常看到这样的图

Nature: 高通量细菌分离培养鉴定



Nat. Biotechnol.: 水稻根系微生物组参与氮利用



# 软件安装

- <http://huttenhower.sph.harvard.edu/GraPhIAn>
- GraPhIAn软件使用Python脚本，可用git或wget下载添加环境变量即可，但依赖biopython、 matplotlib包需用户自行安装
- 推荐使用Conda安装
- # GraPhIAn核心程序包  
conda install graphlan
- # GraPhIAn输入文件制作程序，如转换LEfSe、 MetaPhIAn4结果格式化为GraPhIAn输入文件  
conda install export2graphlan

# 绘制物种树(Cladogram)——GraPhIAn

# 转换MetaPhlAn4结果为GraPhIAn输入

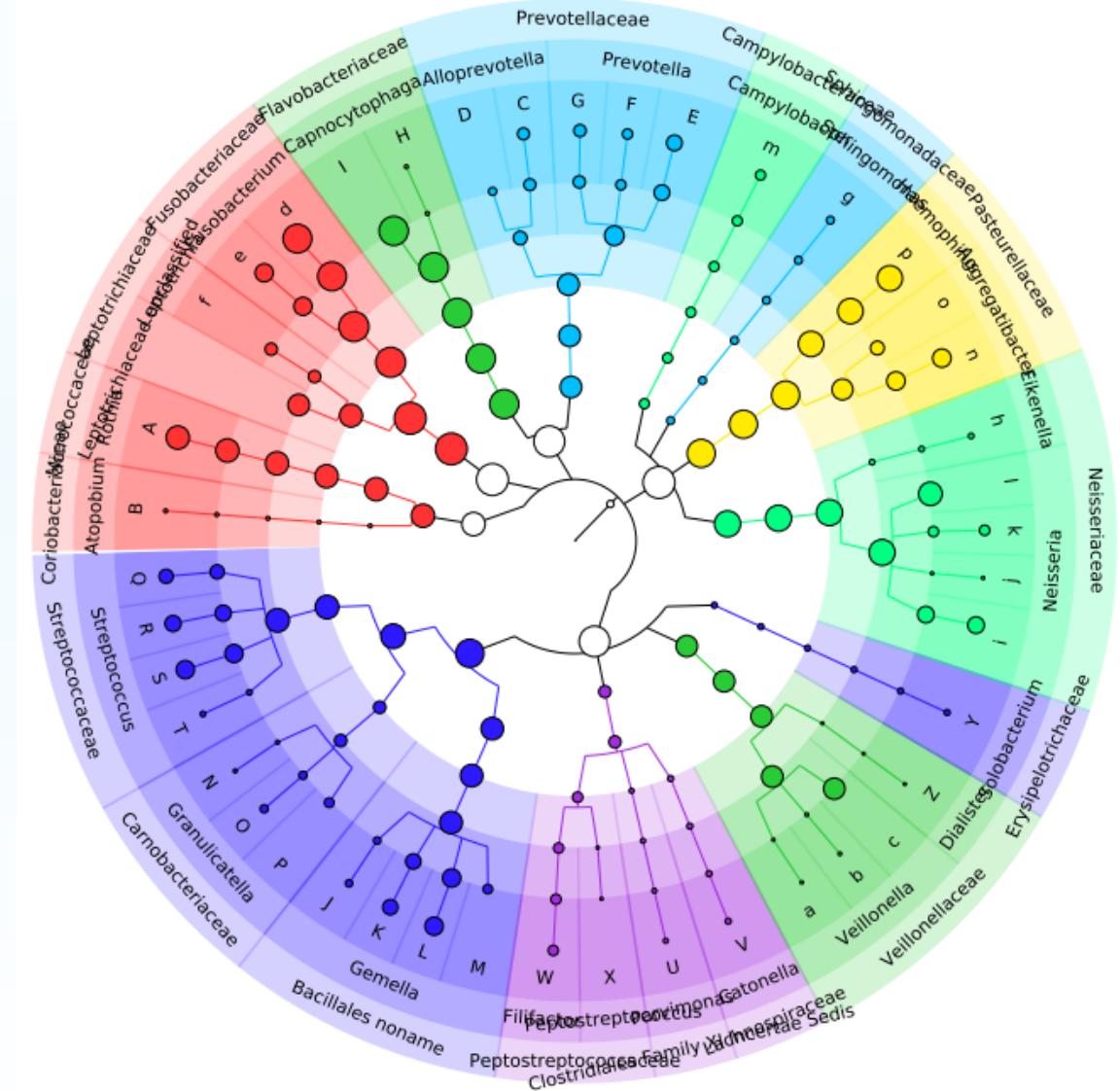
```
export2graphlan.py --skip_rows 1,2 -i result/metaphlan4/taxonomy.tsv \
--tree temp/merged_abundance.tree.txt \
--annotation temp/merged_abundance.annot.txt \
--most_abundant 1000 --abundance_threshold 20 --least_biomarkers 10 \
--annotations 3,4 --external_annotations 7
```

# export2graphlan.py -h参数说明：--skip\_rows是跳过表头，--tree输出树文件，--annotation输出注释文件；--most\_abundant为注释枝数量，默认为10，推荐填写1000，否则会有枝无色无字；--abundance\_threshold最小注释的丰度值，默认20，改为1可全部注释；--least\_biomarkers 目水平最少注释阈值；--annotations 树上文字注释的级别，7级分类

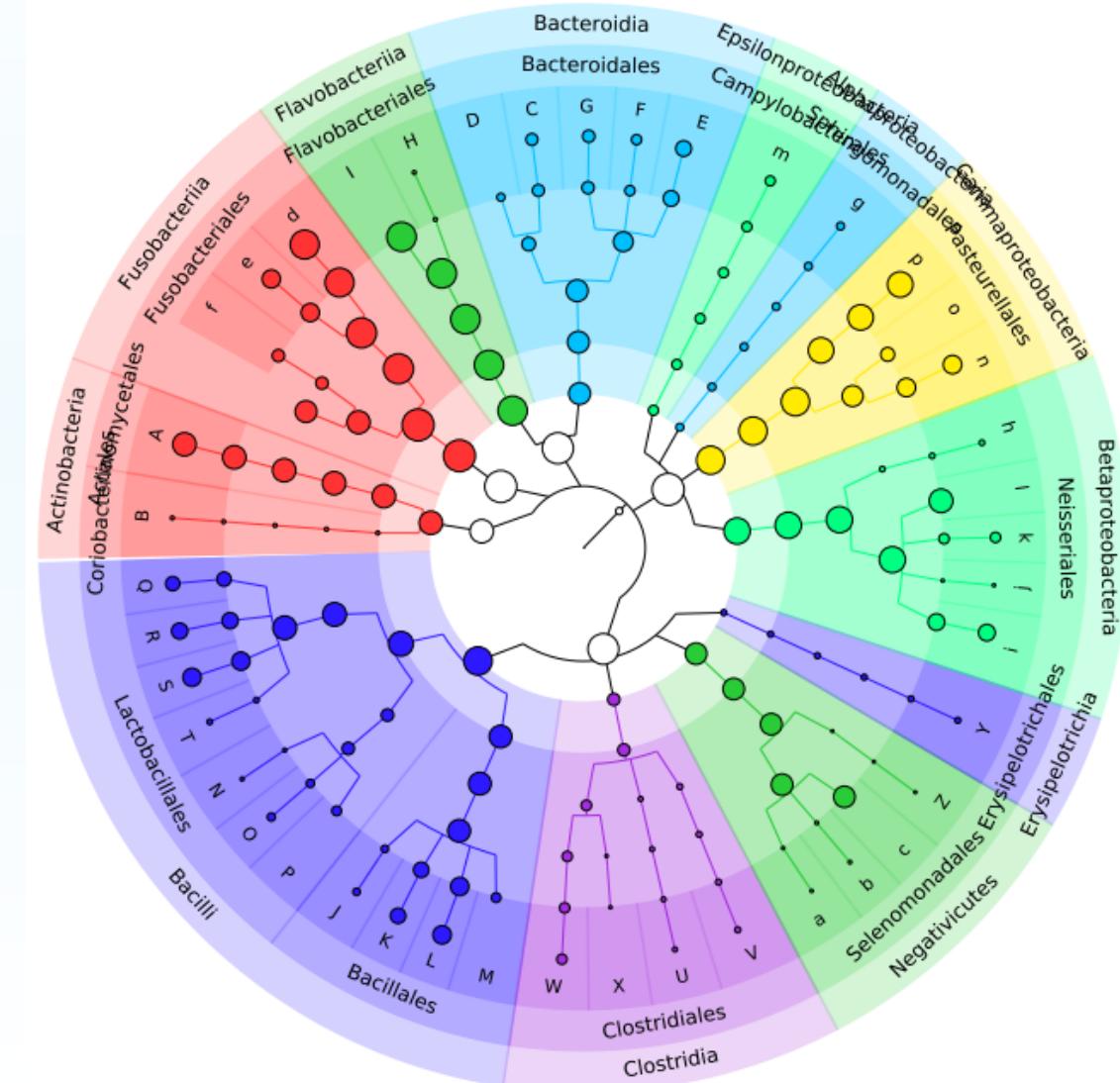
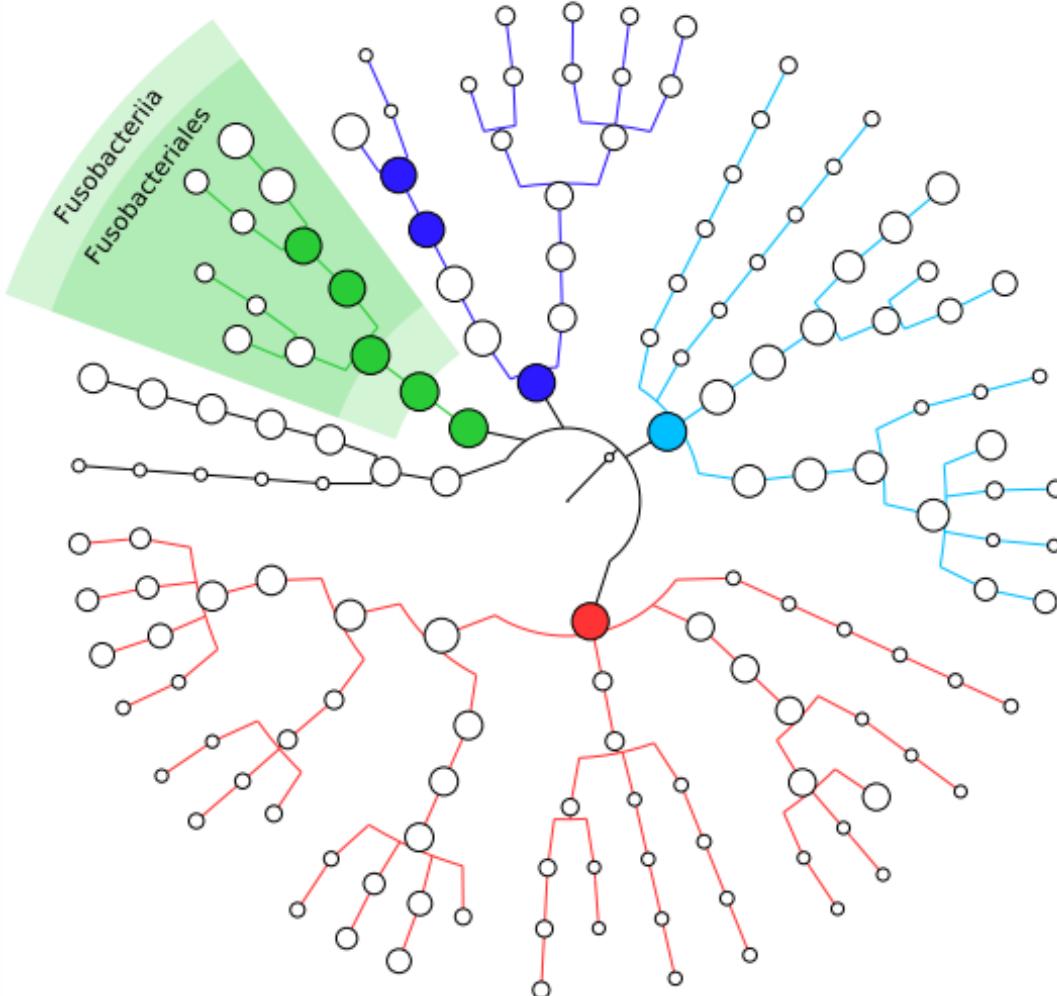
# 数据准备完成，一键出图

```
# 整理树和注释文件为绘图脚本xml格式  
graphlan_annotate.py --annot \  
    temp/merged_abundance.annot.txt \  
    temp/merged_abundance.tree.txt \  
    temp/merged_abundance.xml
```

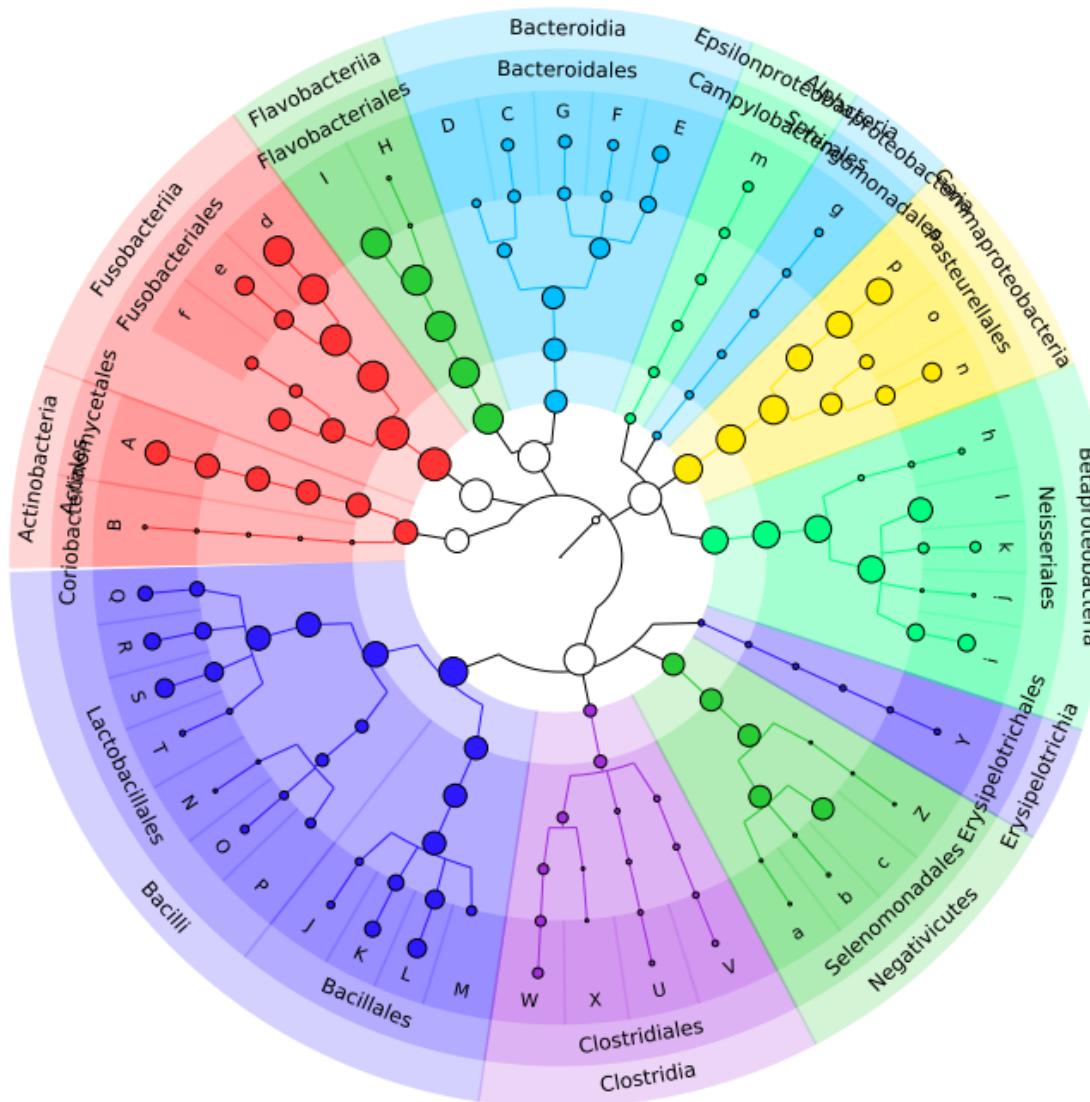
```
# 生成PDF图片和图例  
graphlan.py \  
    temp/merged_abundance.xml \  
    metaphlan2/graphlan.pdf \  
    --external_legends
```



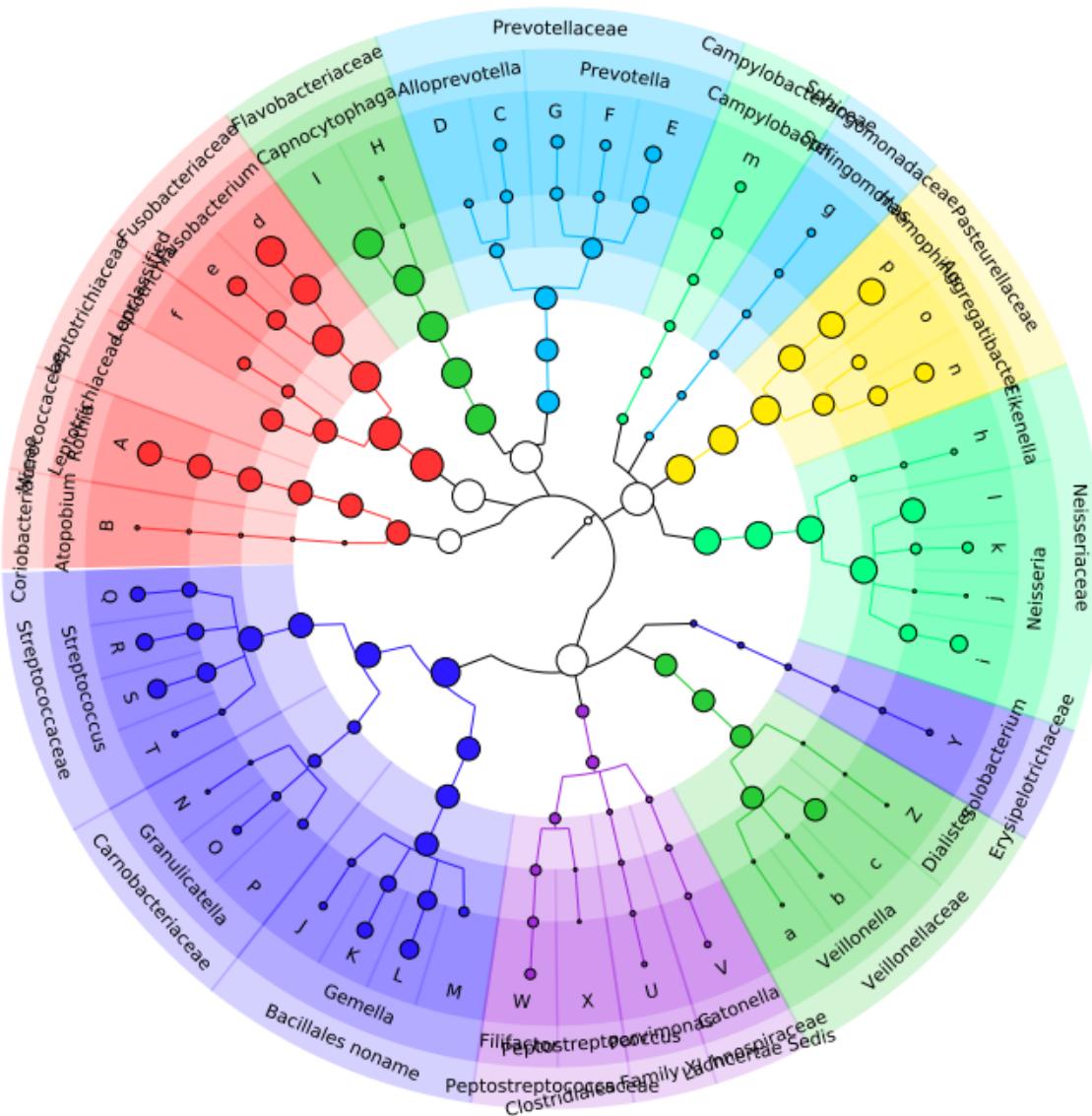
--most\_abundant为注释类数量, 默认10, 推荐1000



## --annotations 3,4(纲目)

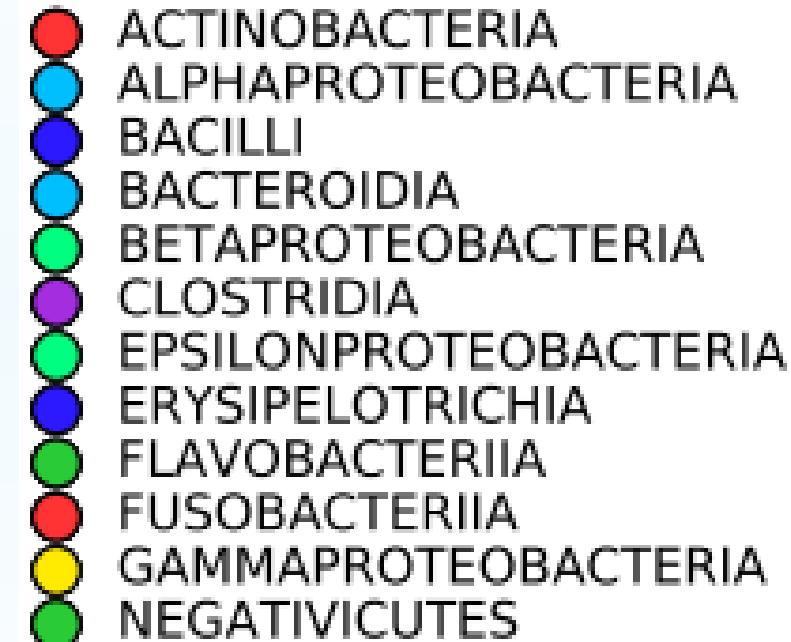


## --annotations 5,6(科属)



# 结果还有扩展注释(anno)和图例(legend)文件

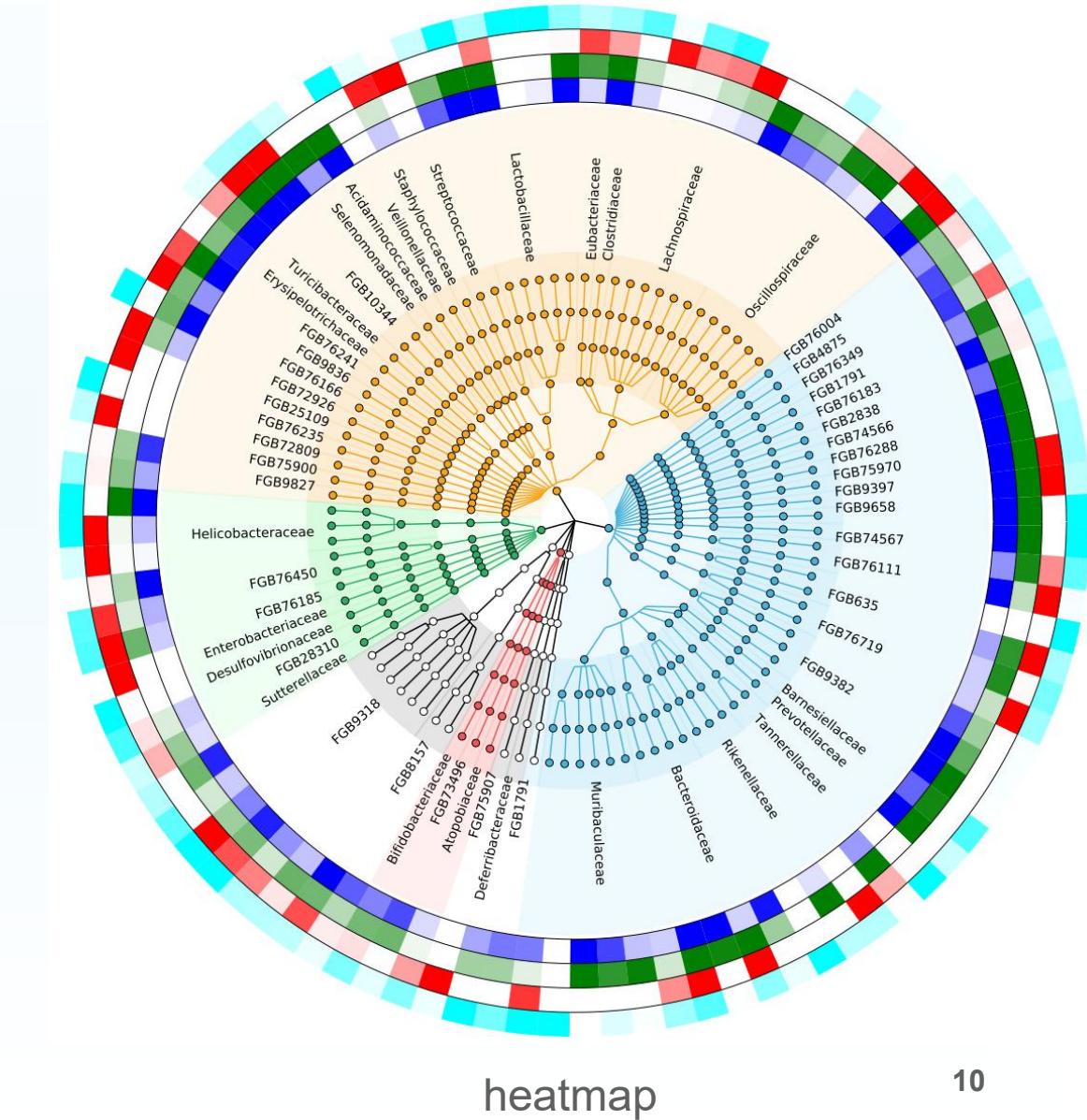
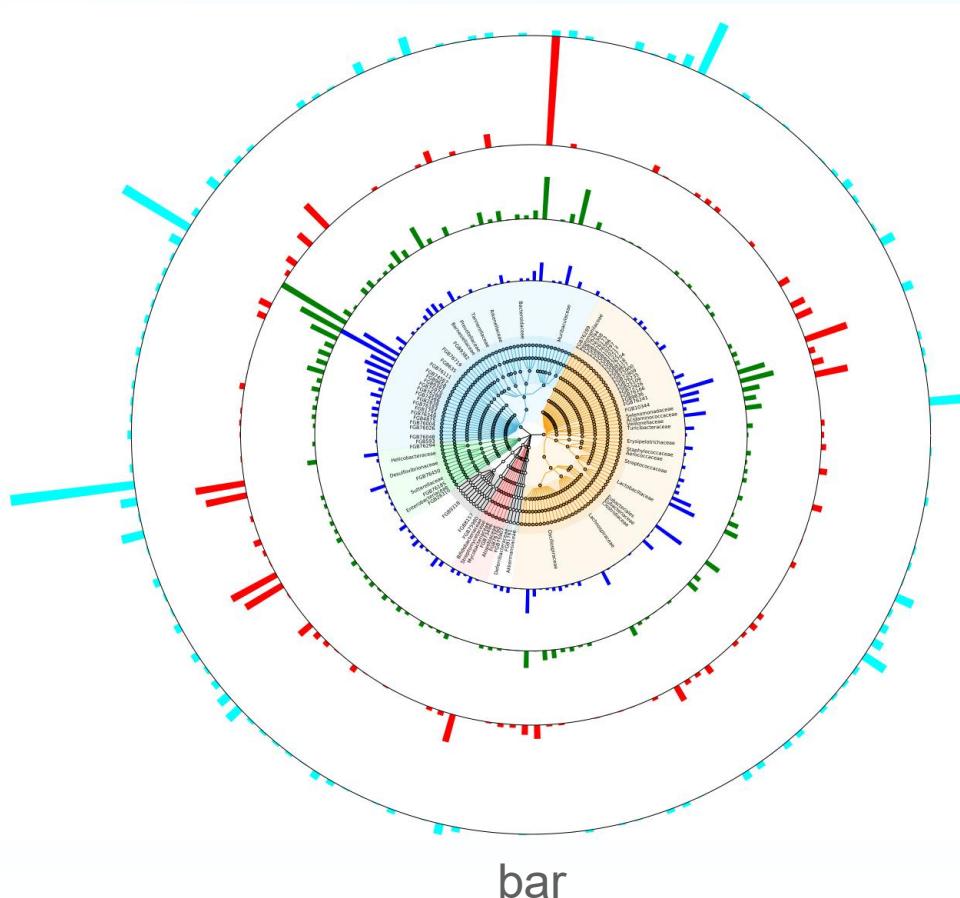
A: Rothia mucilaginosa  
B: Atopobium parvulum  
C: Alloprevotella tannerae  
D: Alloprevotella unclassified  
E: Prevotella melaninogenica  
F: Prevotella nanceiensis  
G: Prevotella nigrescens  
H: Capnocytophaga gingivalis  
I: Capnocytophaga unclassified  
J: Gemella haemolysans  
K: Gemella morbillorum  
L: Gemella sanguinis  
M: Gemella unclassified  
N: Granulicatella adiacens  
O: Granulicatella elegans  
P: Granulicatella unclassified  
Q: Streptococcus australis  
R: Streptococcus infantis  
S: Streptococcus mitis oralis pneumoniae  
T: Streptococcus parasanguinis  
U: Parimonas micra  
V: Catonella morbi  
W: Filifactor alocis  
X: Peptostreptococcus unclassified  
Y: Solobacterium moorei  
Z: Dialister invisus  
a: Veillonella dispar  
b: Veillonella parvula  
c: Veillonella unclassified  
d: Fusobacterium nucleatum  
e: Fusobacterium periodonticum  
f: Leptotrichia unclassified  
g: Sphingomonas melonis  
h: Eikenella corrodens  
i: Neisseria flavaescens  
j: Neisseria meningitidis  
k: Neisseria subflava  
l: Neisseria unclassified  
m: Campylobacter showae  
n: Aggregatibacter segnis  
o: Aggregatibacter unclassified  
p: Haemophilus parainfluenzae



根据需求在Adobe Illustrator中整合排版

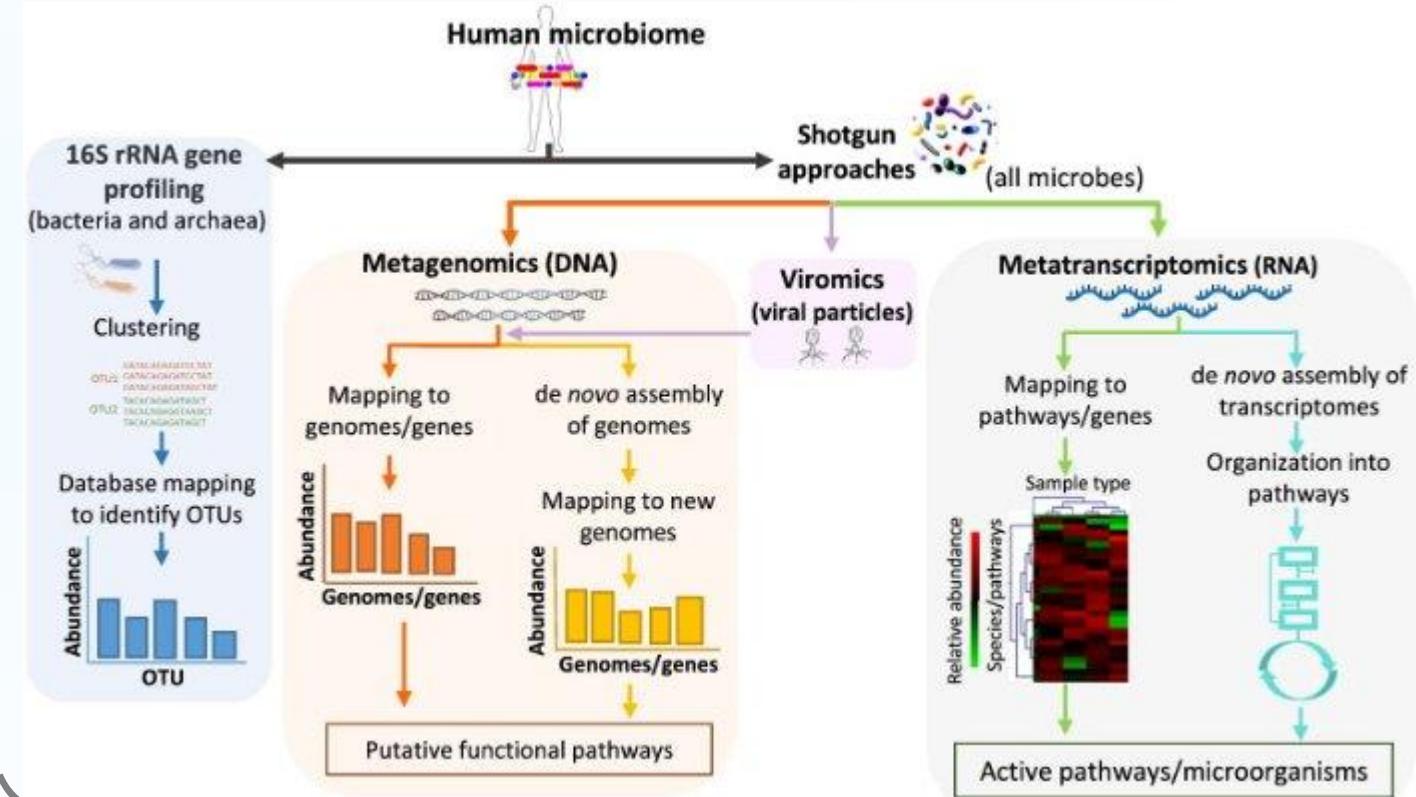
# 新开发的GraPhIAn Plot

```
graphlan_plot.r --input result/metaphlan4/taxonomy.spf \
--design result/metadata.txt --number 100 \
--group all --type heatmap \
--output result/metaphlan4/heatmap
```



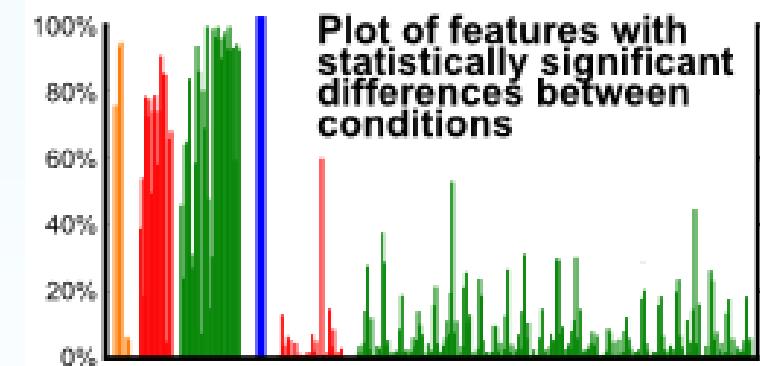
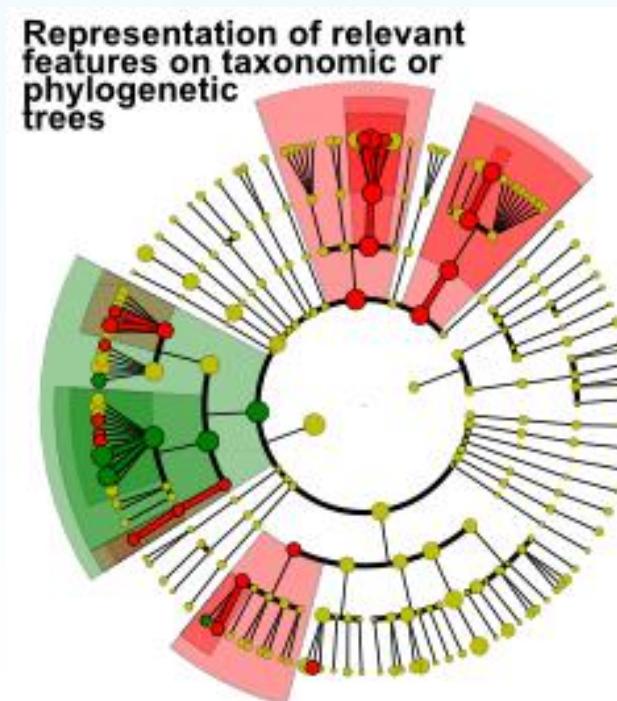
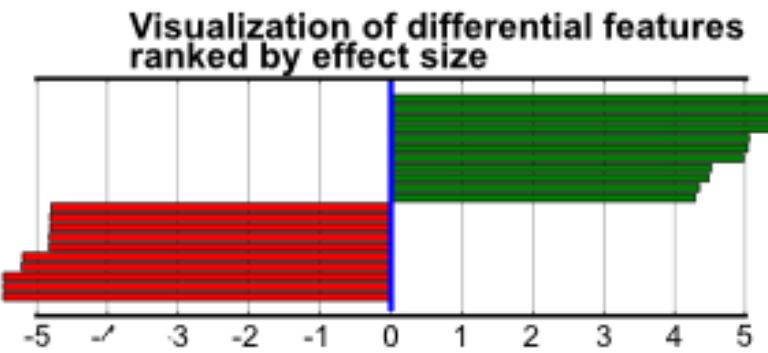
# 宏基因组有参流程——实战分析大纲

- 一. 软件安装和数据库部署
- 二. KneadData质控
- 三. MetaPhlAn4物种组成
- 四. HUMAnN3功能组成
- 五. GraPhlAn可视化物种
- 六. LEfSe分析物种差异
- 七. STAMP功能组成统计分析



# Linear discriminant analysis Effect Size (LEfSe)

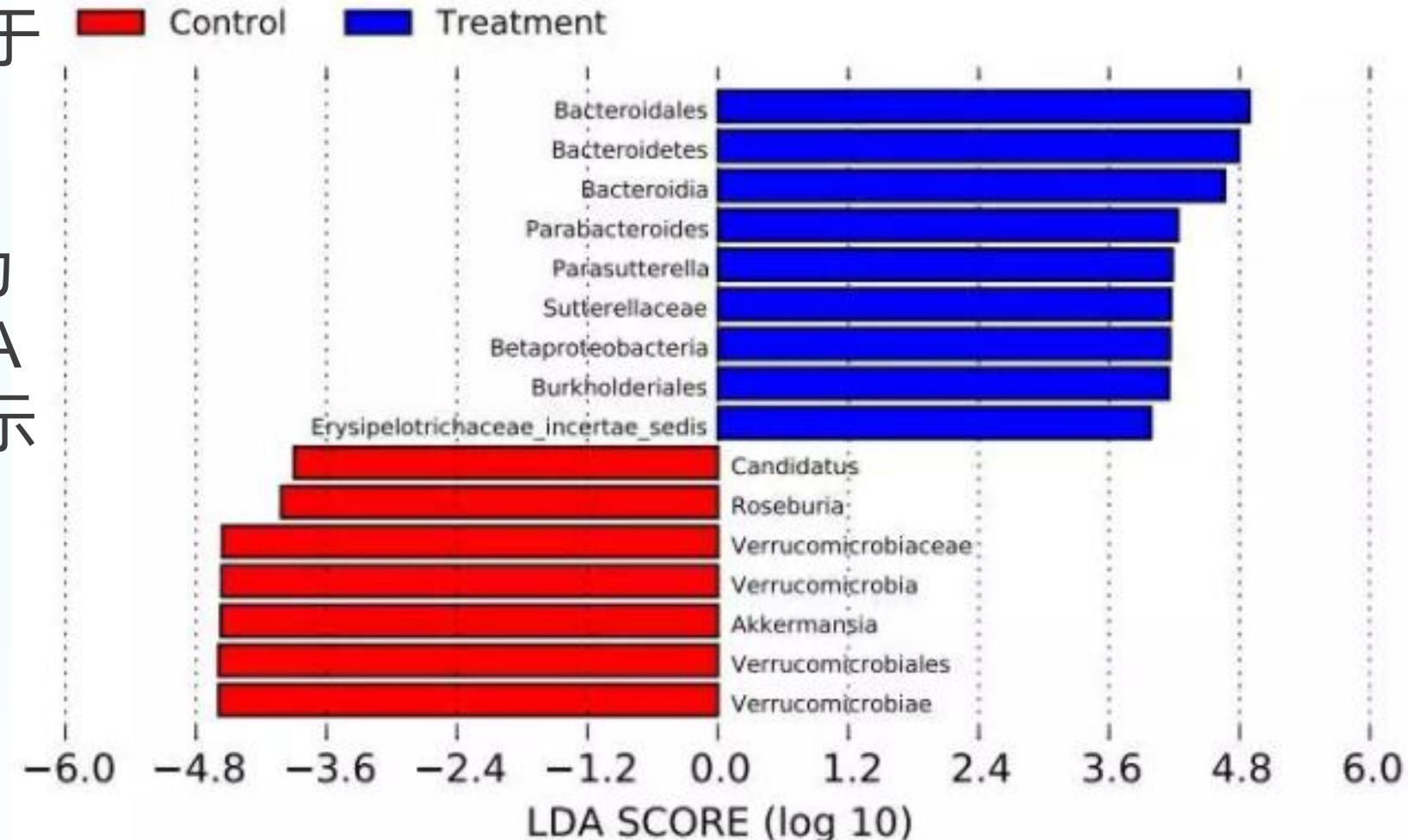
- LEfSe分析即LDA Effect Size分析，是一种用于发现和解释**高维度数据生物标志**(基因、通路和分类单元等)的分析工具，可以进行两个或多个分组的比较，它强调**统计意义和生物相关性**，能够在组与组之间寻找具有统计学差异的生物标志 (Biomarker)。



Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome biology*, 12(6), R60.

# 1. LDA值分布柱状图

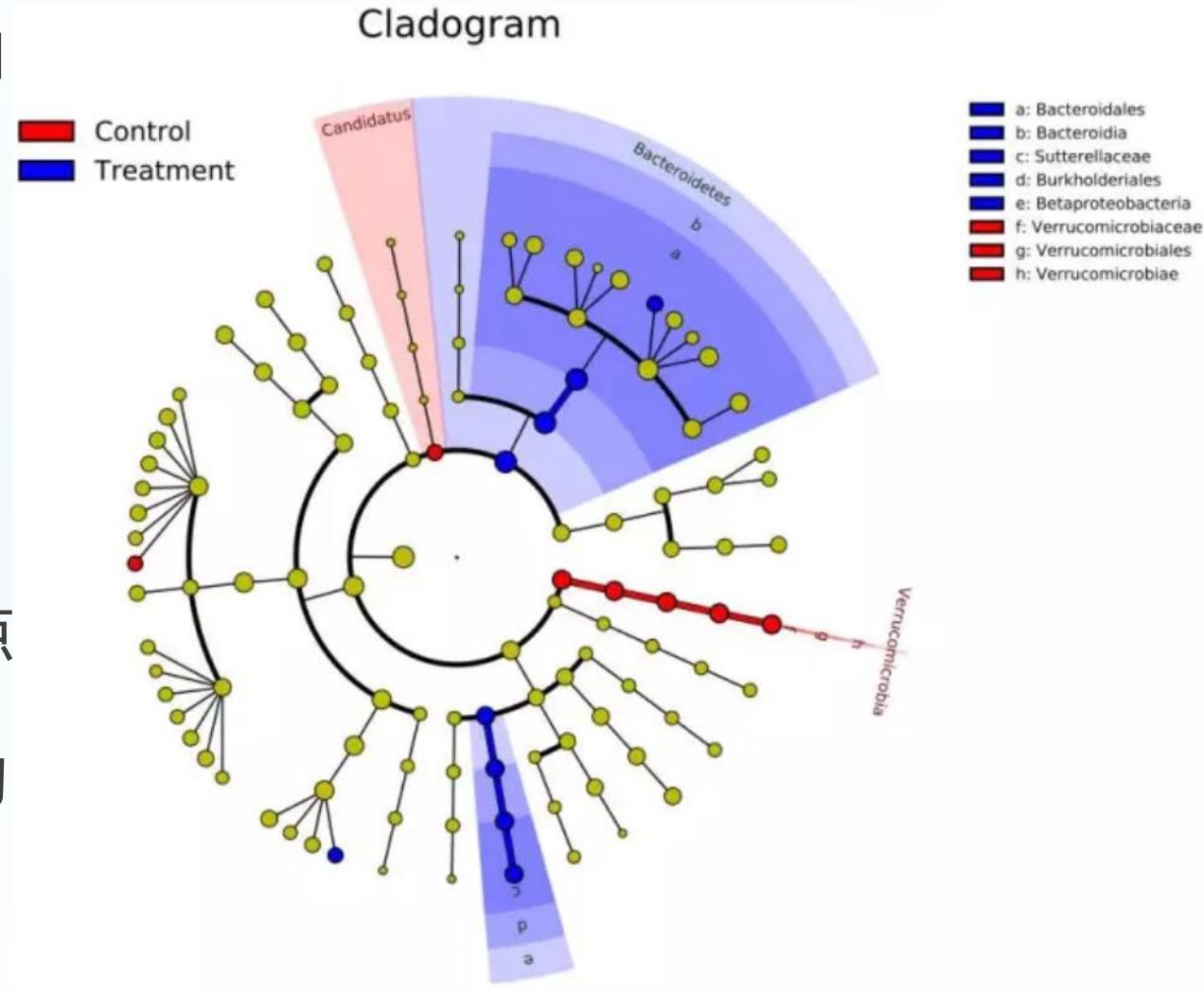
条形图展示LDA score大于预设值的显著差异物种，即具有统计学差异的 Biomaker， 默认预设值为 **2.0**（看横坐标，只有LDA值的绝对值大于2才会显示在图中）；柱状图的颜色代表各自的组别，长短代表的是**LDA score**，即不同组间显著差异物种的影响程度。



## 2. 分类/进化分支图

**小圆圈:** 图中由内至外辐射的圆圈代表了由门至属的分类级别（最里面的那个黄圈圈是界）。不同分类级别的每一个小圆圈代表该水平下的一个分类，**小圆圈的直径大小代表了相对丰度的大小。**

**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(显著富集)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（为了美观，右侧默认只显示**门到科**的差异物种）。



### 3. 特征表

第一列：Biomarker名称；  
第二列：各组分丰度平均值百万比RPM中最大值的log10，如果平均丰度小于10的按照10来计算；  
第三列：差异基因或物种富集的组名；  
第四列：LDA值；  
第五列：Kruskal-Wallis秩和检验的p值，若不是Biomarker用“-”表示。

Biomarker_names	Logarithm value	Groups	LDA_value	P_value
Bacteria.Actinobacteria.Actinobacteria.Coriobacteriales.Coriobacteriaceae	3.910677755			-
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Bacteroidaceae	2.638655621			-
Bacteria.Firmicutes.Bacilli.Lactobacillales.Lactobacillaceae.Lactobacillus	2.394743447	Treatment	3.604999464	0.014574297
Bacteria.Bacteroidetes.Flavobacteriia.Flavobacteriales.Flavobacteriaceae	4.384327808			-
Bacteria.Firmicutes.Clostridia.Clostridiales.Family_XIII.Incertae_Sedis	3.651261857	Treatment	3.383377953	0.015158439
Bacteria.Firmicutes.Bacilli.Bacillales	3.601154267			-
Bacteria.Candidate_division_SR1.norank.norank	3.375794209	Control	4.012050844	0.004274734
Bacteria.Firmicutes.Bacilli.Bacillales.Family_XI	3.601154267			-
Bacteria.BD1_5.norank.norank.norank.norank	3.854913022			-
Bacteria.Firmicutes.Clostridia.Clostridiales.Family_XIII	3.909279604	Control	3.494705494	0.022270979
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Prevotellaceae.Prevotella	5.356981083			-
Bacteria.Candidate_division_SR1.norank	3.375794209			-
Bacteria.Proteobacteria.Epsilonproteobacteria	4.464722976	Treatment	4.012050844	0.004274734
Bacteria.Cyanobacteria	2.306884355			-
Bacteria.Firmicutes.Bacilli.Lactobacillales	5.099727491			-
Bacteria.Proteobacteria.Gammaproteobacteria.Pseudomonadales.Moraxellaceae	4.274596542	Control	3.990287269	0.014556334
Bacteria.Cyanobacteria.Cyanobacteria.norank.norank	2.306884355			-
Bacteria.Proteobacteria.Betaproteobacteria.Burkholderiales.Burkholderiaceae	3.302330929			-

运行结果: temp/input.res, ImageGP中 \*.txt.lefse

# ImageGP/BIC——LEfSe一键分析

The screenshot shows the ImageGP web interface. On the left is a vertical sidebar with icons for file operations. The main area has a title "LEfSe" and text "两个国内在线分析网站备选". Below this are two URLs: "BIC: <https://www.bic.ac.cn/#/analysis?page=b%27MzY%3D%27>" and "ImageGP: <http://www.ehbio.com/ImageGP/index.php/Home/Index/LEFSe.html>". The central part contains a "For full LEFSe analysis" section with a table of taxonomy abundance data and a "Tab separated taxonomy abundance file (pay attention to the colorful symbols or words)" section with a table of abundance data. A large green arrow labeled "Wait for several minutes" points from the analysis sections to a "1 LEfSe result" section on the right, which lists five items: "1.1 LEfSe Feature Table", "1.2 Cladogram", "1.3 Cladogram (only significant organisms)", "1.4 LDA Score Distribution Bar plot for marker species", and "1.5 Download all results".

For full LEFSe analysis

Class	KO	KO	KO	OE	OE	OE	WT	WT	WT
k_Bacteria	41.5	40.0	45.8	41.3	39.6	33.3	46.0	49.0	42.7
k_Bacteria p_Actinobacteria	12.8	20.0	13.8	15.1	15.9	12.3	20.3	21.5	16.3
k_Bacteria p_Bacteroidetes	0.8	0.7	4.8	3.1	1.6	1.2	2.8	3.0	3.9
k_Bacteria p_Firmicutes	0.3	0.3	0.1	0.9	0.3	1.7	0.2	0.1	0.2
k_Bacteria p_Bacteroidetes c_Flavobacteriia	0.8	0.7	4.8	3.1	1.6	1.2	2.8	3.0	3.9
k_Bacteria p_Firmicutes c_Bacilli	0.3	0.3	0.1	0.9	0.3	1.7	0.2	0.1	0.2
k_Bacteria p_Proteobacteria c_Alphaproteobacteria	4.1	4.6	7.4	3.8	4.9	4.7	3.9	5.1	4.2
k_Bacteria p_Actinobacteria c_Actinobacteria o_Kineosporiales	0.3	0.6	0.3	0.6	0.5	0.3	0.8	0.6	0.4
k_Bacteria p_Actinobacteria c_Actinobacteria o_Kineosporiales f_Kineosporiaceae	0.3	0.6	0.3	0.6	0.5	0.3	0.8	0.6	0.4

Tab separated taxonomy abundance file (pay attention to the colorful symbols or words)

Class	KO	KO	KO	OE	OE	OE	WT	WT	WT
k_Bacteria	41.5	40.0	45.8	41.3	39.6	33.3	46.0	49.0	42.7
k_Bacteria p_Actinobacteria	12.8	20.0	13.8	15.1	15.9	12.3	20.3	21.5	16.3
k_Bacteria p_Bacteroidetes	0.8	0.7	4.8	3.1	1.6	1.2	2.8	3.0	3.9
k_Bacteria p_Firmicutes	0.3	0.3	0.1	0.9	0.3	1.7	0.2	0.1	0.2
k_Bacteria p_Bacteroidetes c_Flavobacteriia	0.8	0.7	4.8	3.1	1.6	1.2	2.8	3.0	3.9
k_Bacteria p_Firmicutes c_Bacilli	0.3	0.3	0.1	0.9	0.3	1.7	0.2	0.1	0.2
k_Bacteria p_Proteobacteria c_Alphaproteobacteria	4.1	4.6	7.4	3.8	4.9	4.7	3.9	5.1	4.2
k_Bacteria p_Actinobacteria c_Actinobacteria o_Kineosporiales	0.3	0.6	0.3	0.6	0.5	0.3	0.8	0.6	0.4
k_Bacteria p_Actinobacteria c_Actinobacteria o_Kineosporiales f_Kineosporiaceae	0.3	0.6	0.3	0.6	0.5	0.3	0.8	0.6	0.4

Wait for several minutes →

1 LEfSe result

- 1.1 LEfSe Feature Table
- 1.2 Cladogram
- 1.3 Cladogram (only significant organisms)
- 1.4 LDA Score Distribution Bar plot for marker species
- 1.5 Download all results

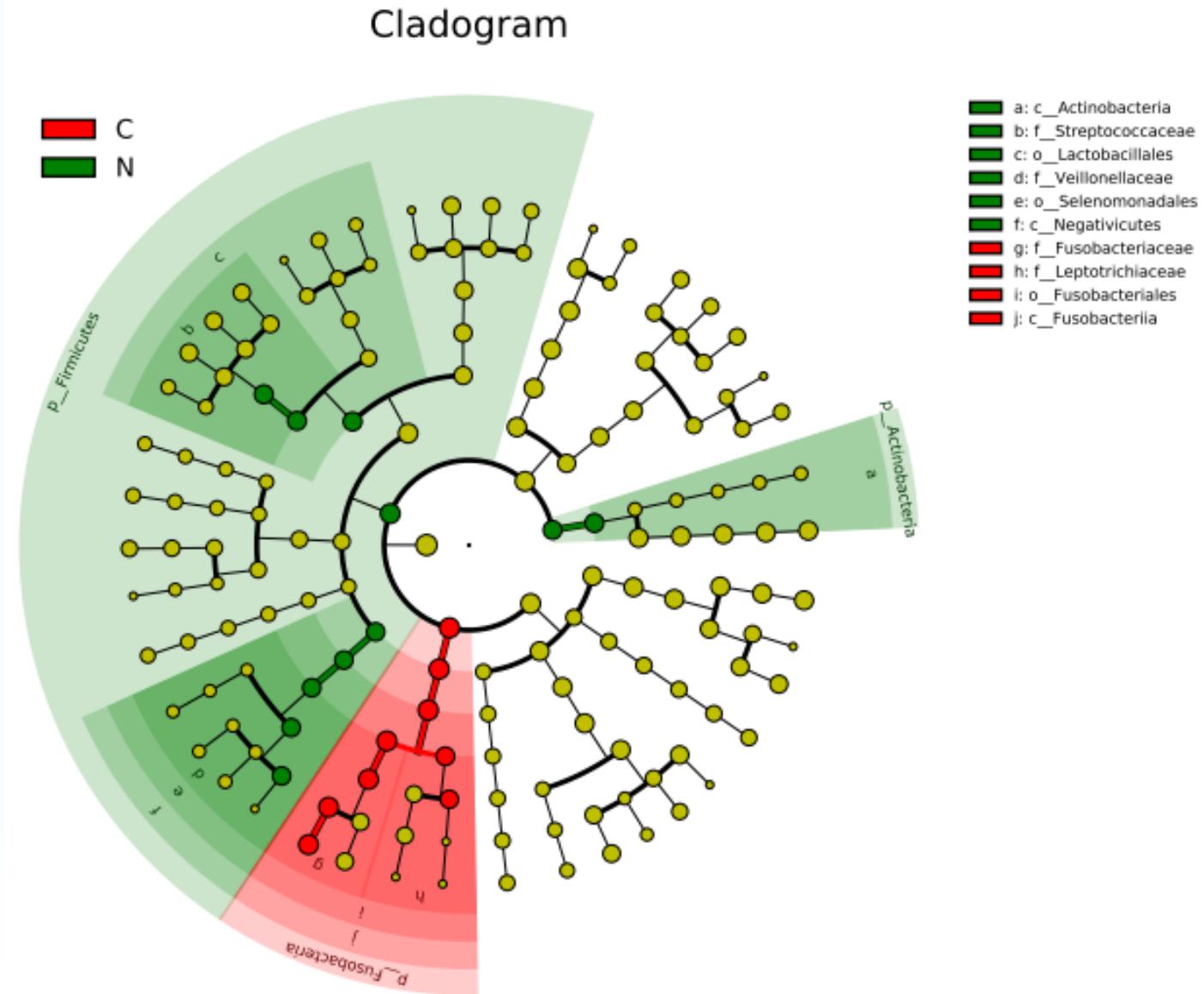
Thanks for citing: [ImageGP: An easy-to-use data visualization web server for scientific researchers. iMeta 1: e5.](#)

Download citation in [RIS](#) format.

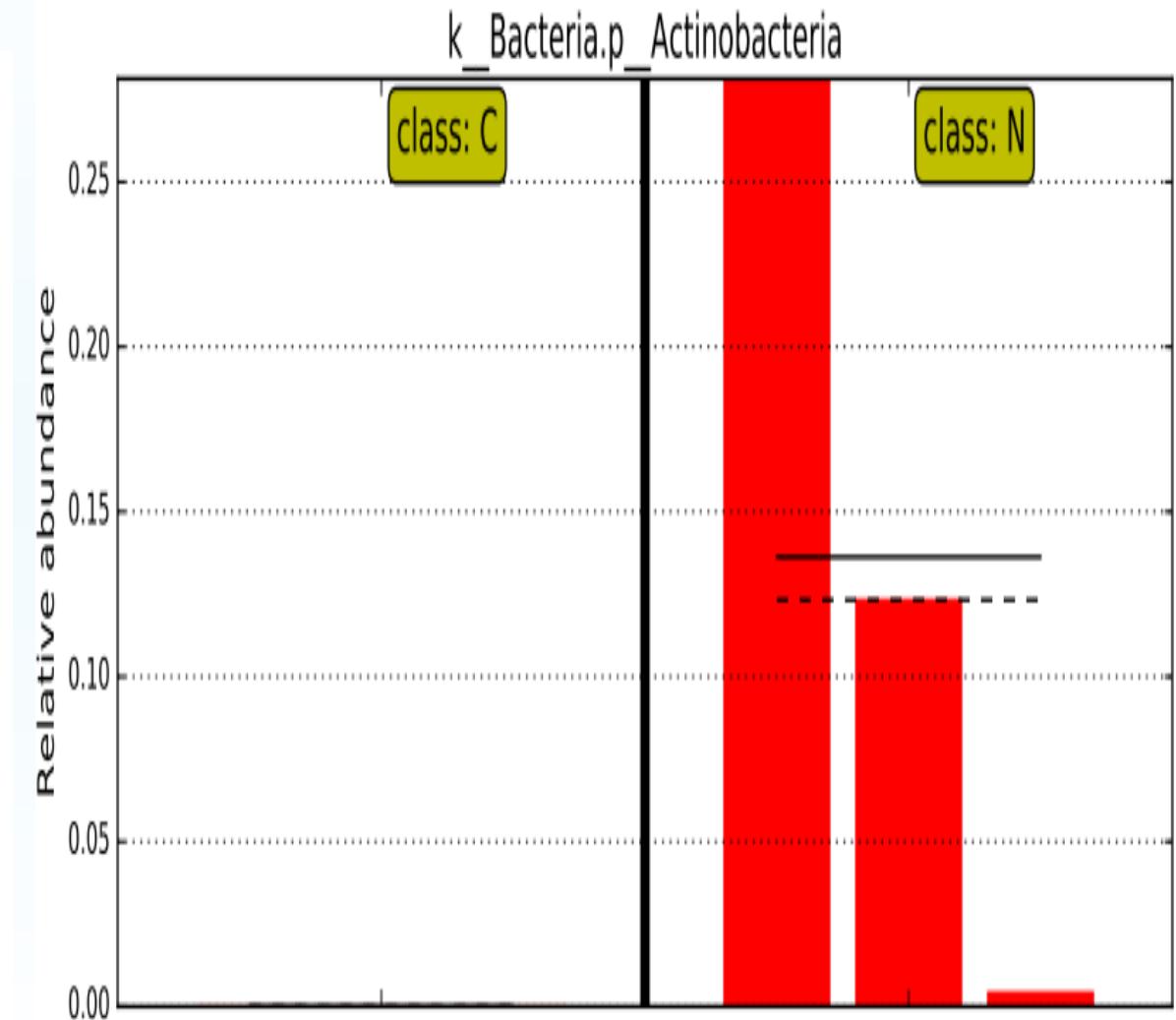
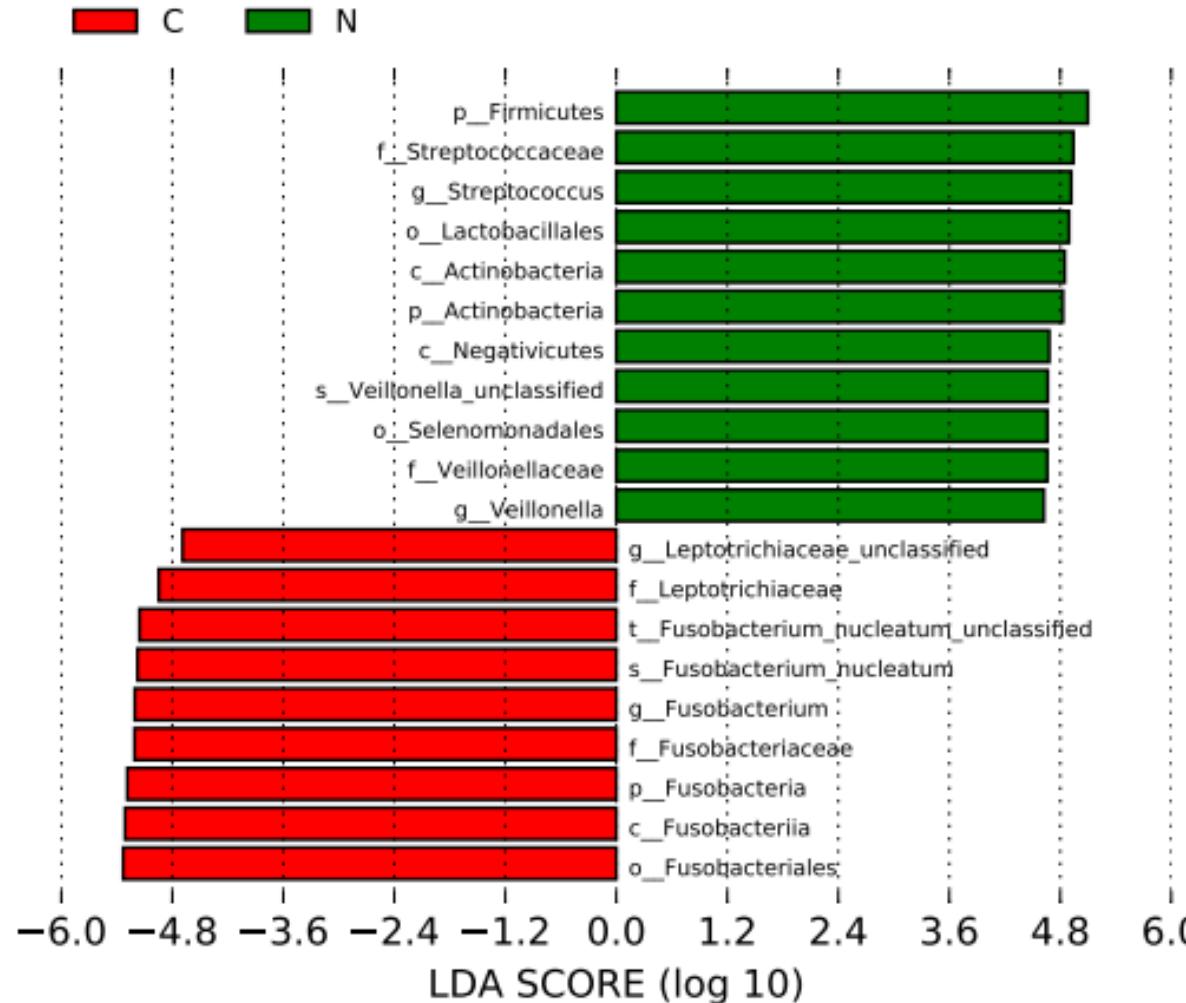
Chen, et al. 2024. "ImageGP 2 for enhanced data visualization and reproducible analysis in biomedical research."

*iMeta* 3: e239. <https://doi.org/10.1002/imt2.239>

# 物种树展示各级别差异



# 差异features柱状图(总体-个体)



# Linux服务器运行LEfSe代码 ——LEfSe差异分析和绘制Cladogram

- # 格式转换为lefse内部格式

```
lefse_format_input.py $result/metaphlan4/lefse.txt \
temp/input.in -c 1 -s 2 -o 1000000
```

- # 运行lefse

```
lefse_run.py temp/input.in temp/input.res
```

- # 绘制物种树注释差异

```
lefse_plot_cladogram.py temp/input.res \
$result/metaphlan4/lefse_cladogram.pdf --format pdf
```

# 绘制差异features柱状图

- # 绘制所有差异features柱状图

```
lefse_plot_res.py temp/input.res $result/metaphlan4/lefse_res.pdf --format pdf
```

- # 绘制单个features柱状图(同STAMP中barplot)

- # sort -k3,3n temp/input.res |less -S # 查看差异features列表

```
lefse_plot_features.py -f one --format pdf --feature_name  
"k_Bacteria.p_Actinobacteria" temp/input.in temp/input.res  
$result/metaphlan4/lefse_Actinobacteria.pdf
```

- # 批量绘制所有差异features柱状图

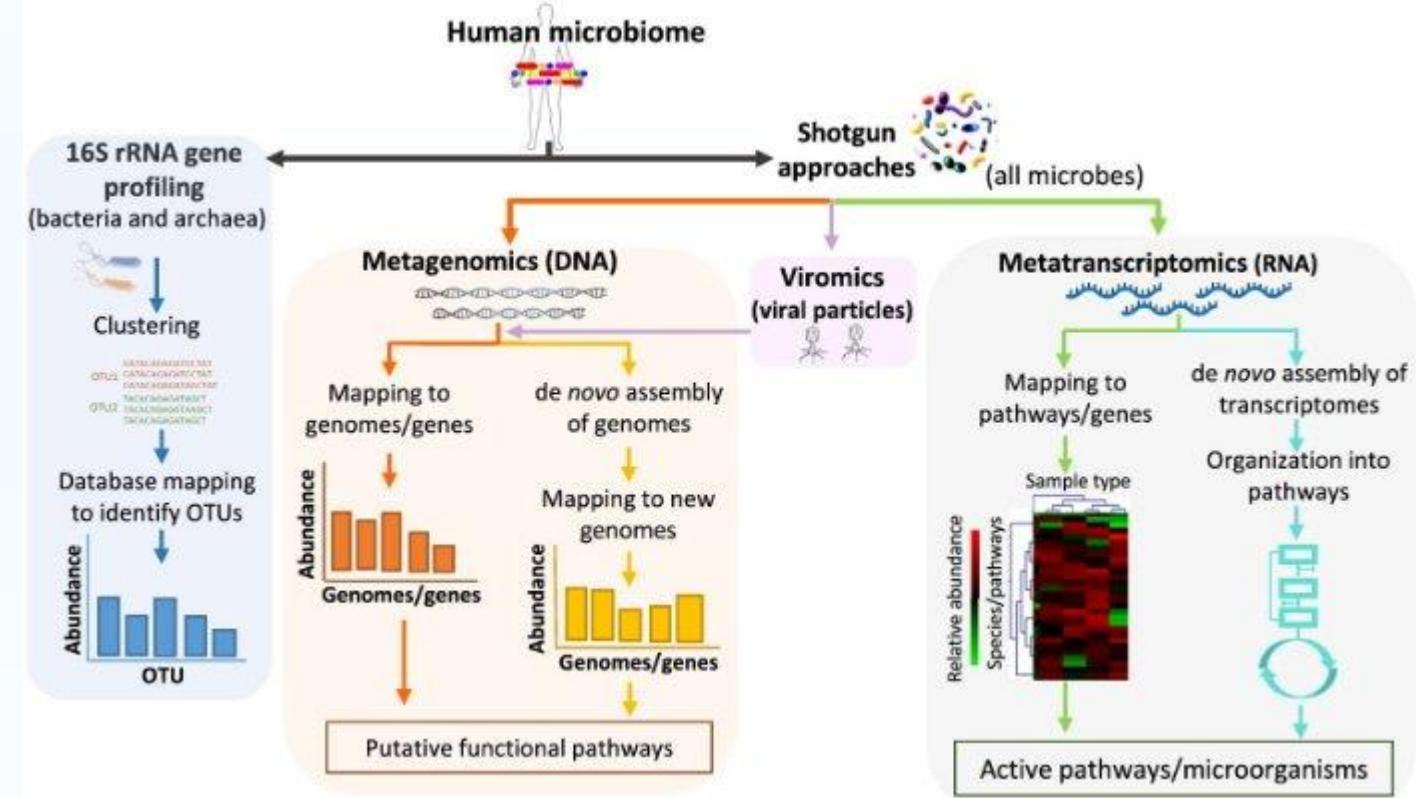
```
lefse_plot_features.py -f diff --archive none --format pdf temp/input.in temp/input.res  
$result/metaphlan4/lefse_
```

# LEfSe官网在线使用

- 主页：<http://huttenhower.sph.harvard.edu/galaxy>，不仅有LEfSe工具，还包括PICRUSt、GraPhlAn、MetaPhlAn、MetaPhlAn4、microPITA等微生物组分析工具。
- 最好先从顶部User菜单中注册一个账号并登陆，可保存工作记录，方便下次修改。
- 通过左向角向上箭头上传数据，将文件托入上传窗口中，文件类型选择tabular，其它默认即可。
- 注：上传后经常需要排除等待处理，可能几分钟或几小时。后面每一步操作都可能要排队很久，但支持提交未完成任务的后续任务。

# 宏基因组有参流程——实战分析大纲

- 一. 软件安装和数据库部署
- 二. KneadData质控
- 三. MetaPhlAn4物种组成
- 四. HUMAnN3功能组成
- 五. GraPhlAn可视化物种
- 六. LEfSe分析物种差异
- 七. STAMP功能组成分析



# HUMAnN4获得功能组成表

- 通路丰度表path.tsv，包括功能通路和物种组成

# Pathway	C1	C2	C3	C4	C5	C6	N1	N2	N3	N4	N5	N6
1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis	3.30	0.00	0.00	0.00	0.00	0.00	2.87	0.00	9.96	13.82	3.61	11.19
ANAEROFRUCAT-PWY: homolactic fermentation	4.88	0.00	3.56	7.05	8.69	4.73	0.00	1.80	8.06	11.20	7.17	9.90
ANAGLYCOLYSIS-PWY: glycolysis III (from glucose)	12.78	3.38	3.42	15.43	12.74	8.63	0.00	3.54	18.65	16.27	16.96	14.22
ANAGLYCOLYSIS-PWY: glycolysis III (from glucose) g_Fusobacterium.s	5.00	0.00	0.00	0.00	4.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ANAGLYCOLYSIS-PWY: glycolysis III (from glucose) g_Haemophilus.s	0.00	0.00	0.00	3.55	0.00	2.08	0.00	0.00	2.18	0.00	6.60	0.00
ANAGLYCOLYSIS-PWY: glycolysis III (from glucose) g_Streptococcus.s	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.13	0.00	0.00	0.00
ARGSYNBSUB-PWY: L-arginine biosynthesis II (acetyl cycle)	0.00	0.00	0.00	7.59	0.00	0.00	0.00	0.00	6.40	17.81	2.90	0.00
ARGSYNBSUB-PWY: L-arginine biosynthesis II (acetyl cycle) g_Neisseri	0.00	0.00	0.00	1.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ARGSYN-PWY: L-arginine biosynthesis I (via L-ornithine)	0.00	4.00	0.00	9.67	0.00	0.00	0.00	3.55	5.65	17.30	8.42	6.18
ARGSYN-PWY: L-arginine biosynthesis I (via L-ornithine) g_Haemophi	0.00	0.00	0.00	2.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ARGSYN-PWY: L-arginine biosynthesis I (via L-ornithine) g_Neisseria.s	0.00	0.00	0.00	1.98	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

- 还可从通路覆盖度pathwaycoverage、基因家族genefamilies分析
- 如何统计分析与可视化呢？

# STAMP简介

## STAMP



STAMP is a software package for analyzing taxonomic or metabolic profiles that promotes ‘best practices’ in choosing appropriate statistical techniques and reporting results. Statistical hypothesis tests for pairs of samples or groups of samples is support along with a wide range of exploratory plots. STAMP encourages the use of effect sizes and confidence intervals in assessing biological importance. A user friendly, graphical interface permits easy exploration of statistical results and generation of publication quality plots for inferring the biological relevance of features in a metagenomic profile. STAMP is open source, extensible via a plugin framework, and available for all major platforms.

### Announcements

- June 26, 2015: STAMP v2.1.3 released. Minor bug fix to scatter plot to properly handle profiles contained a single feature.
- June 15, 2015: STAMP v2.1.2 released. Minor enhancements to extended error bar, heatmap, and profile bar plots.
- June 7, 2015: STAMP v2.1.1 released. Resolves issue with v2.1.0 installation.
- June 4, 2015: STAMP v2.1.0 released. Resolves the \_hierarchy\_wrap issue. Requires numpy >= 1.9.1, scipy >= 0.15.1, matplotlib >= 1.4.2.
- Previous announcements

### Documentation

- Quick Installation instructions (Microsoft Windows, Linux, Apple's Mac OS X)
- User's Guide
- Google Group
- FAQs
- Version history

### Downloads

Please *uninstall previous versions of STAMP before installing a new release*.

- STAMP v2.1.3 setup package for Microsoft Windows (~42MB)
- Linux and OS X users can follow the instructions to *install from source*
- STAMP GitHub Repository
- Previous versions

主页: <http://kiwi.cs.dal.ca/Software/STAMP>

Donovan H. Parks, Robert G. Beiko. 2010. Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* 26: 715-721.

<https://doi.org/10.1093/bioinformatics/btq041>

Donovan H. Parks, Gene W. Tyson, Philip Hugenholtz, *et al.* 2014. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30: 3123-3124.

<https://doi.org/10.1093/bioinformatics/btu494>

# 输入文件：特征表 + 实验设计(分组信息等元数据)

○ 特征表为物种、功能通路等丰度矩阵；实验设计为样本分组信息。

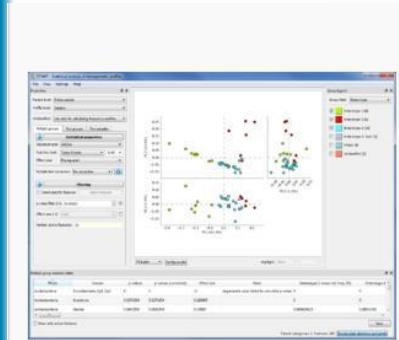
# Pathway	C1	C2	C3
1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis	3.30	0.00	0.00
ANAEROFRUCAT-PWY: homolactic fermentation	4.88	0.00	3.56
ANAGLYCOLYSIS-PWY: glycolysis III (from glucose)	12.78	3.38	3.42
ANAGLYCOLYSIS-PWY: glycolysis III (from glucose) g_Fusobacterium.s	5.00	0.00	0.00
ANAGLYCOLYSIS-PWY: glycolysis III (from glucose) g_Haemophilus.s_	0.00	0.00	0.00
ANAGLYCOLYSIS-PWY: glycolysis III (from glucose) g_Streptococcus.s_	0.00	0.00	0.00
ARGSYNBSUB-PWY: L-arginine biosynthesis II (acetyl cycle)	0.00	0.00	0.00
ARGSYNBSUB-PWY: L-arginine biosynthesis II (acetyl cycle) g_Neisseri	0.00	0.00	0.00
ARGSYN-PWY: L-arginine biosynthesis I (via L-ornithine)	0.00	4.00	0.00
ARGSYN-PWY: L-arginine biosynthesis I (via L-ornithine) g_Haemophi	0.00	0.00	0.00
ARGSYN-PWY: L-arginine biosynthesis I (via L-ornithine) g_Neisseria.s	0.00	0.00	0.00
ARGSYN-PWY: L-arginine biosynthesis I (via L-ornithine) g_Veillonella	0.00	0.00	0.00
ARO-PWY: chorismate biosynthesis I	0.00	1.64	0.00
ARO-PWY: chorismate biosynthesis II g_Haemophilus.s_Haemophilus_	0.00	0.00	0.00
ARO-PWY: chorismate biosynthesis II g_Haemophilus.s_Haemophilus_	0.00	0.00	0.00
ARO-PWY: chorismate biosynthesis II g_Neisseria.s_Neisseria_flavesce	0.00	0.00	0.00
ARO-PWY: chorismate biosynthesis II g_Neisseria.s_Neisseria_subflava	0.00	0.00	0.00

特征表 pathabundance.tsv

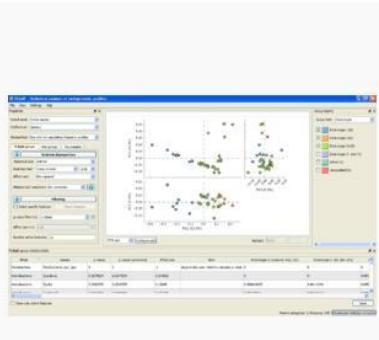
SampleID	Group	Replicate	Sex
C1	Cancer	1	Male
C2	Cancer	2	Male
C3	Cancer	3	Male
C4	Cancer	4	Male
C5	Cancer	5	Female
C6	Cancer	6	Female
N1	Normal	1	Male
N2	Normal	2	Male
N3	Normal	3	Male
N4	Normal	4	Male
N5	Normal	5	Female
N6	Normal	6	Female

元数据(分组信息) metadata.txt

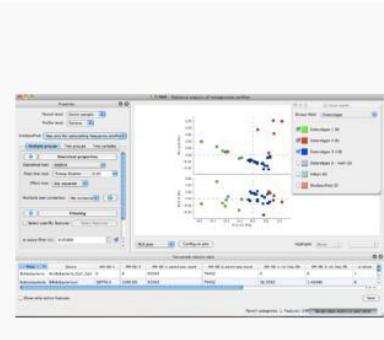
# STAMP可用的图表类型



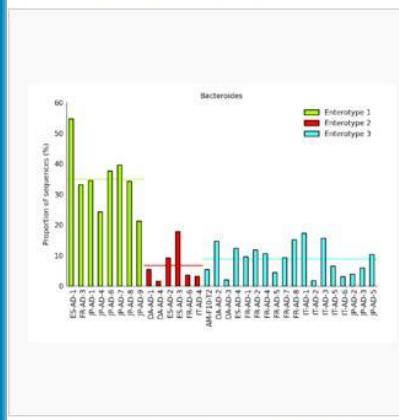
STAMP v2.0.0 on Microsoft Windows 7 (x64).



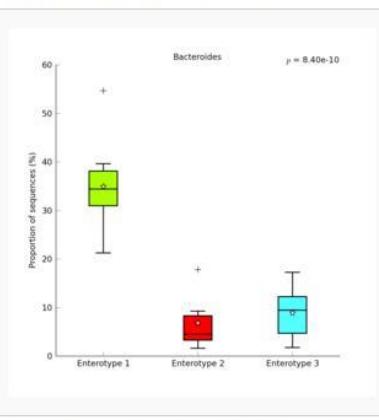
STAMP v2.0.0 on Microsoft Windows XP.



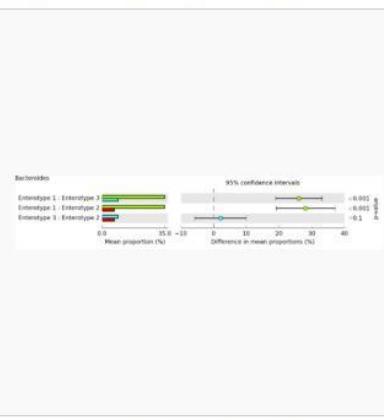
STAMP v2.0.0 on Apple's Mac OS X Leopard.



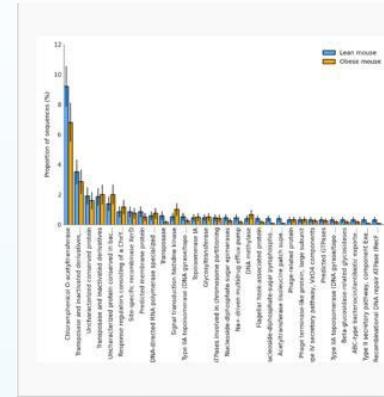
Bar plot showing the abundance of *Bacteroides* within the gut microbiota of individuals assigned to the 3 enterotypes proposed by Arumugam and colleagues (data described in Arumugam et al., 2011 [④](#)).



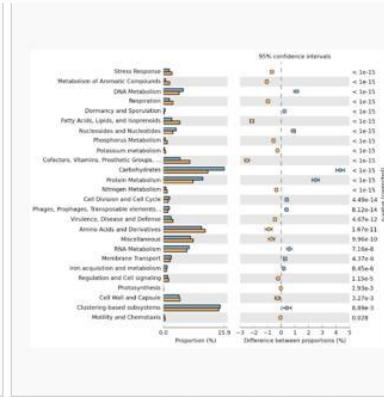
Box plot showing the abundance of *Bacteroides* within the gut microbiota of individuals assigned to the 3 enterotypes proposed by Arumugam and colleagues (data described in Arumugam et al., 2011 [④](#)).



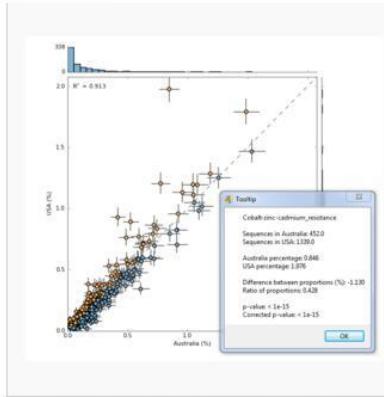
Post-hoc plot indicating that *Bacteroides* is significantly over-represented in enterotype 1 compared to the 2 other enterotypes proposed by Arumugam and colleagues (data described in Arumugam et al., 2011 [④](#)).



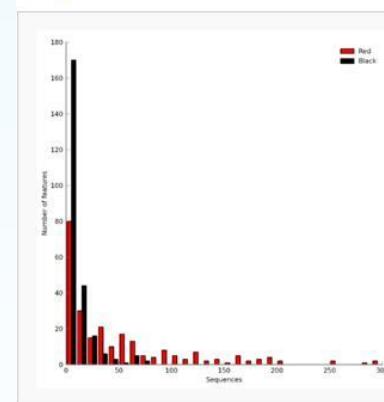
Functional profile plot for an obese and a lean mouse microbiome (data described in Turnbaugh et al., 2006 [④](#)).



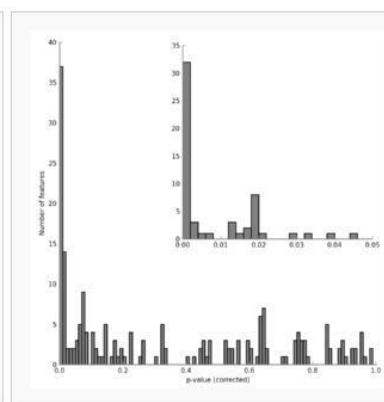
Extended error bar plot for a pair of bovine rumen microbiomes (data described in Brulc et al., 2009 [④](#)).



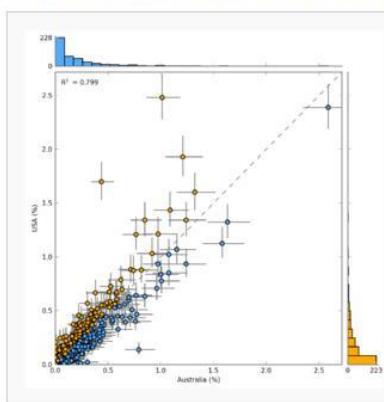
SEED subsystem scatter plot for a pair of enhanced biological phosphorus removal (EPR) sludge metagenomes (data described in Martín et al., 2006 [④](#)).



Sequence histogram for a functional profile of two iron mine metagenomes (data described in Edwards et al., 2006 [④](#)).



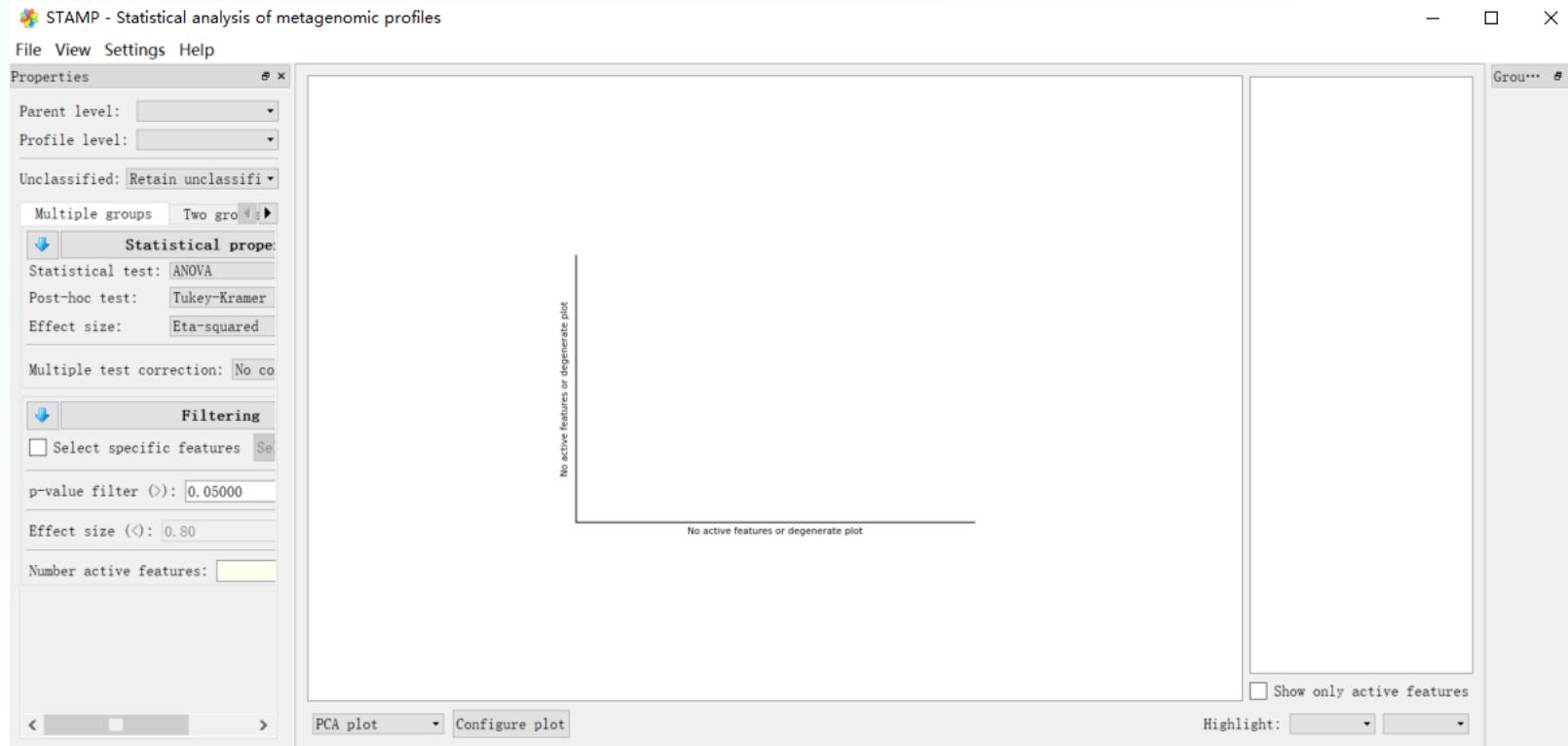
p-value histogram for two iron mine metagenomes (data described in Edwards et al., 2006 [④](#)).



Scatterplot with histograms showing the functional profile of A-phosphatase sequences from two EPR communities (data described in Parks and Beiko, 2010 [④](#)).

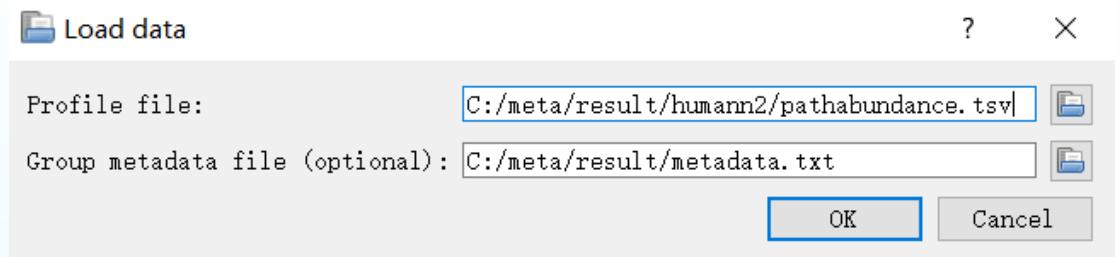
# 1. STAMP安装和运行

- <http://kiwi.cs.dal.ca/Software/STAMP> 下载系统对应版本；
- Windows下按要求一路安装即可(安装目录不要出现中文)；
- 安装后程序名为STAMP的花瓣图标，双击运行；
- Mac下安装麻烦



## 2. 打开文件 —— 默认主成分分析

- File – Load data
- 选择特征表和元数据，点OK
- 左侧Properties属性面板选择分类层级、统计方法、比较组等；
- 右侧Group legend图例面板切换分组方式、分组筛选；
- 底部Metadata table元数据面板可以进行样本筛选；
- 中间作图区，可切换不同图型样式，设置图片属性；
- 最上面菜单栏可以导出图、表，设置参数和控制面板开关；
- 最下面状态栏显示当前分析基本信息和计算状态。



File View Settings Help

菜单栏导出图、表，设置参数和控制面板

**Properties**

Parent level: Entire sample

Profile level: # Pathway

Unclassified

Multiplicity correction: Benjamini-Hochberg

**属性面板**  
选择分类层级、统计方法、比较组等

**Statistical properties**

Statistical test: ANOVA

Post-hoc test: Tukey-Kramer

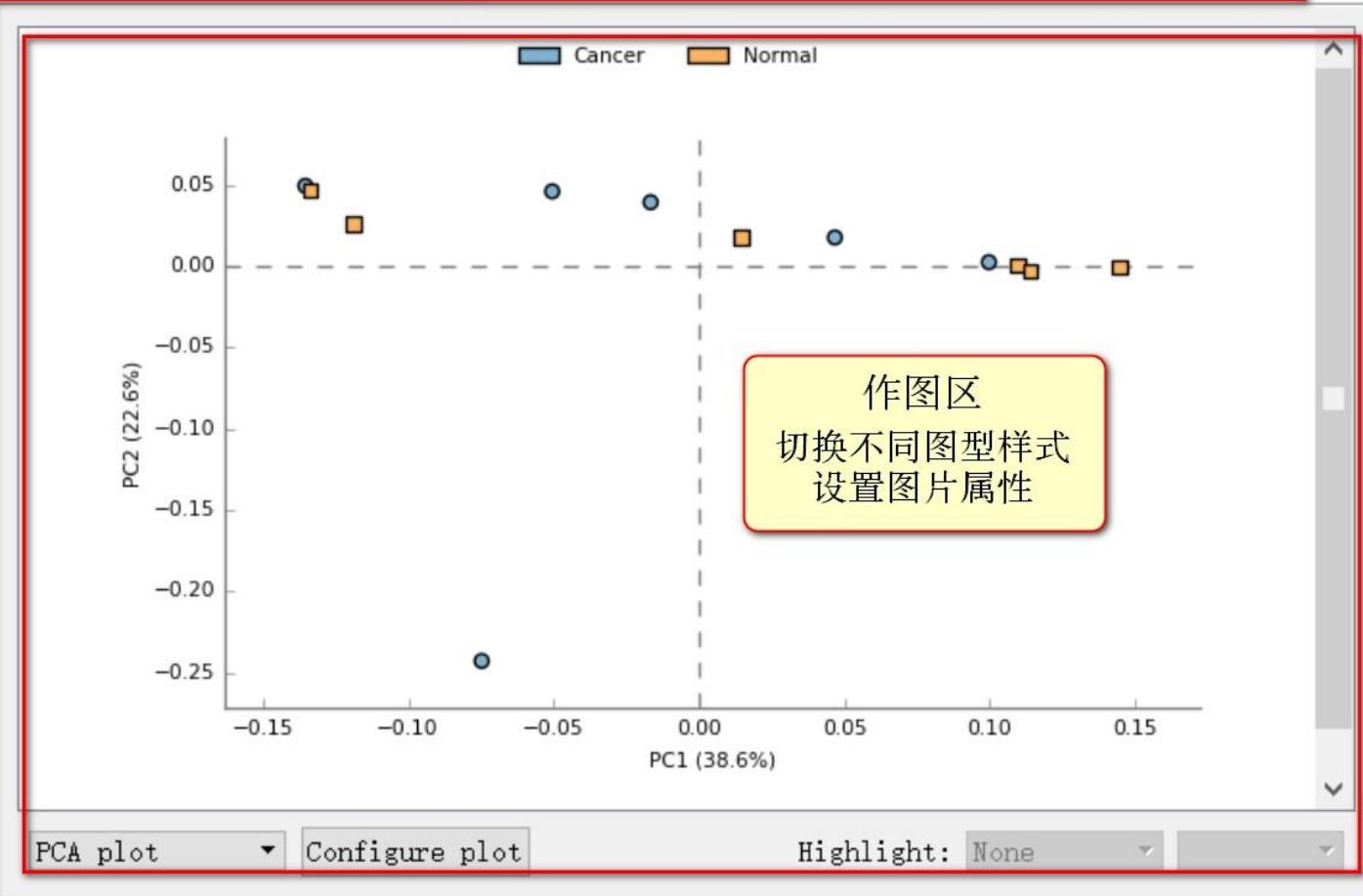
Effect size: Eta-squared

Multiple test correction: No correction

**Filtering**

Select specific features **Select**

p-value filter (>): 0.05000



**Group legend**

Group field: Group

Cancer (6)

Normal (6)

**图例面板**  
切换分组方式、分组筛选

**Metadata table**

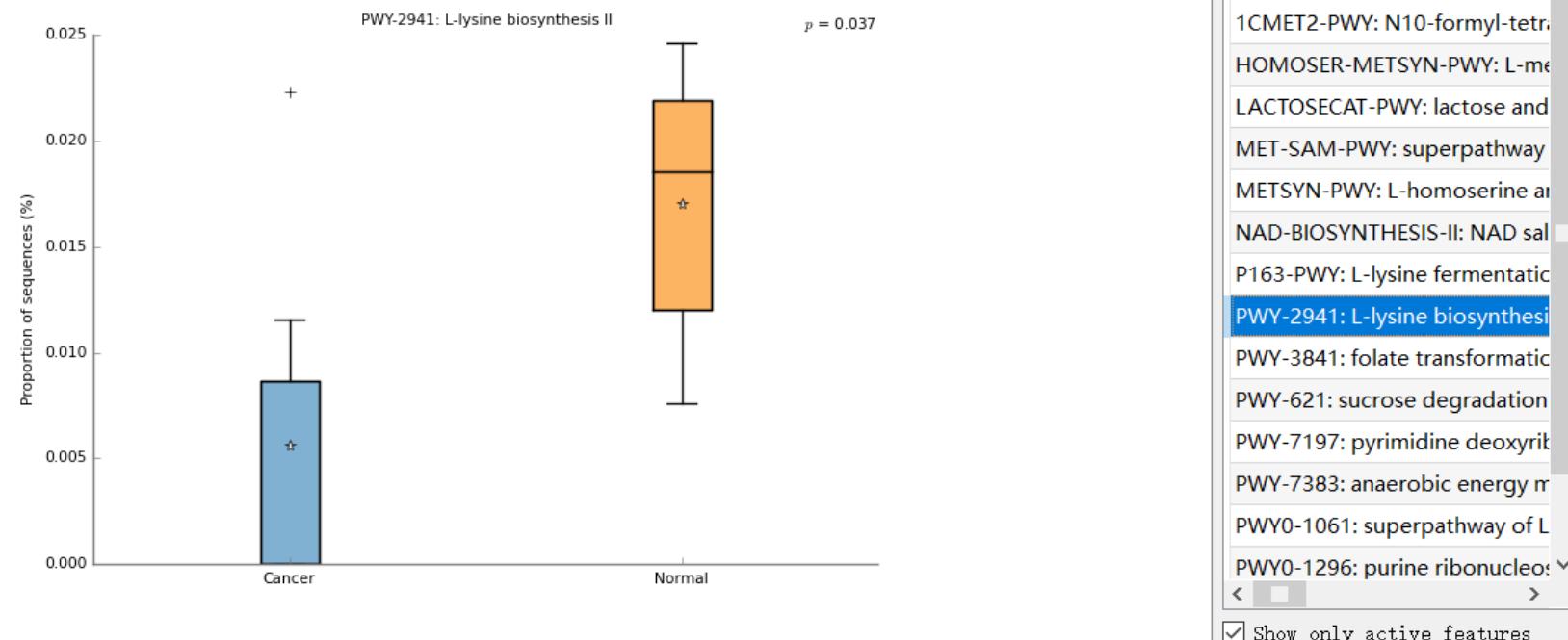
Filter: Add all samples where CRR = CRR117732

Sample Id	CRR	GSA	Group	Individual	Replicate	Sex	project_accession
<input checked="" type="checkbox"/> C1	CRR117732	CRA002355	Cancer	p136	1	Male	PRJCA002236
<input checked="" type="checkbox"/> C2	CRR117733	CRA002355	Cancer	p143	2	Male	PRJCA002236

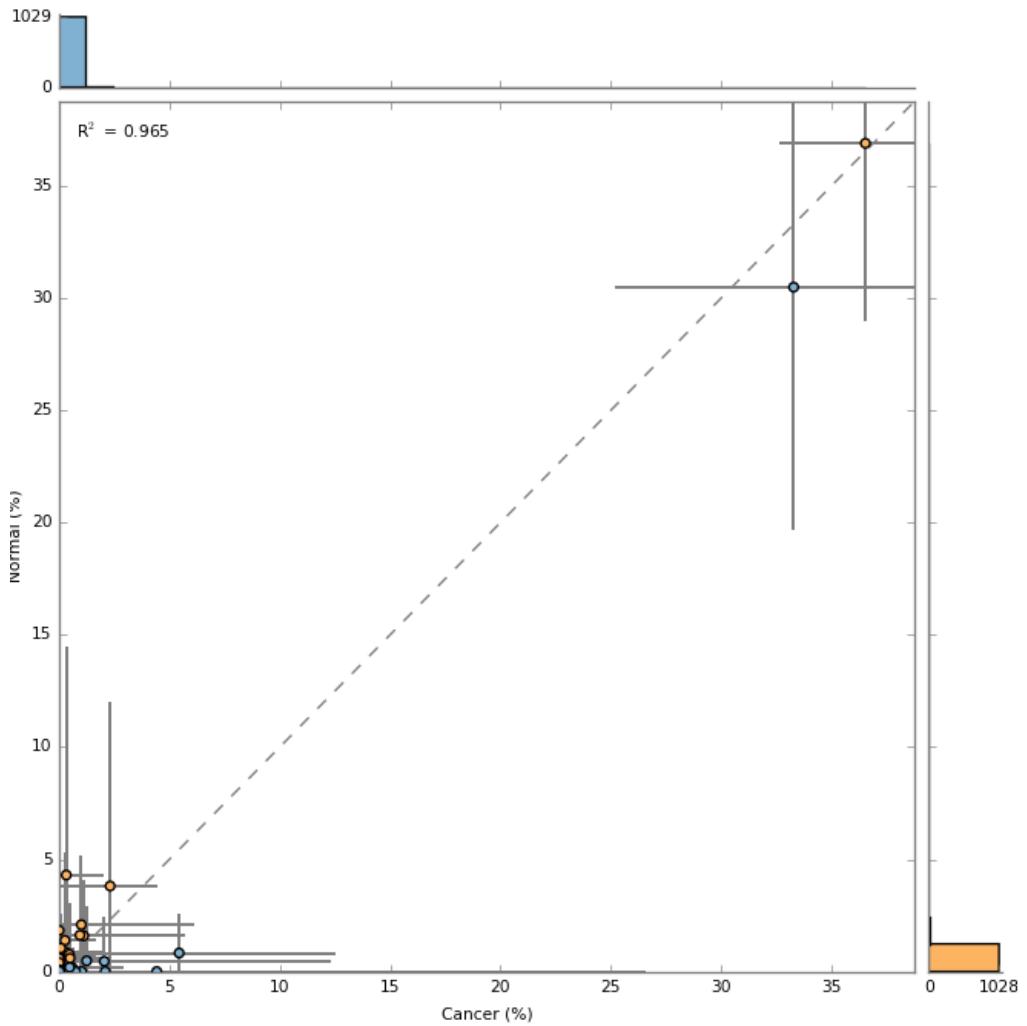
**元数据面板**  
样本筛选

### 3. 多组比较 —— 箱线图

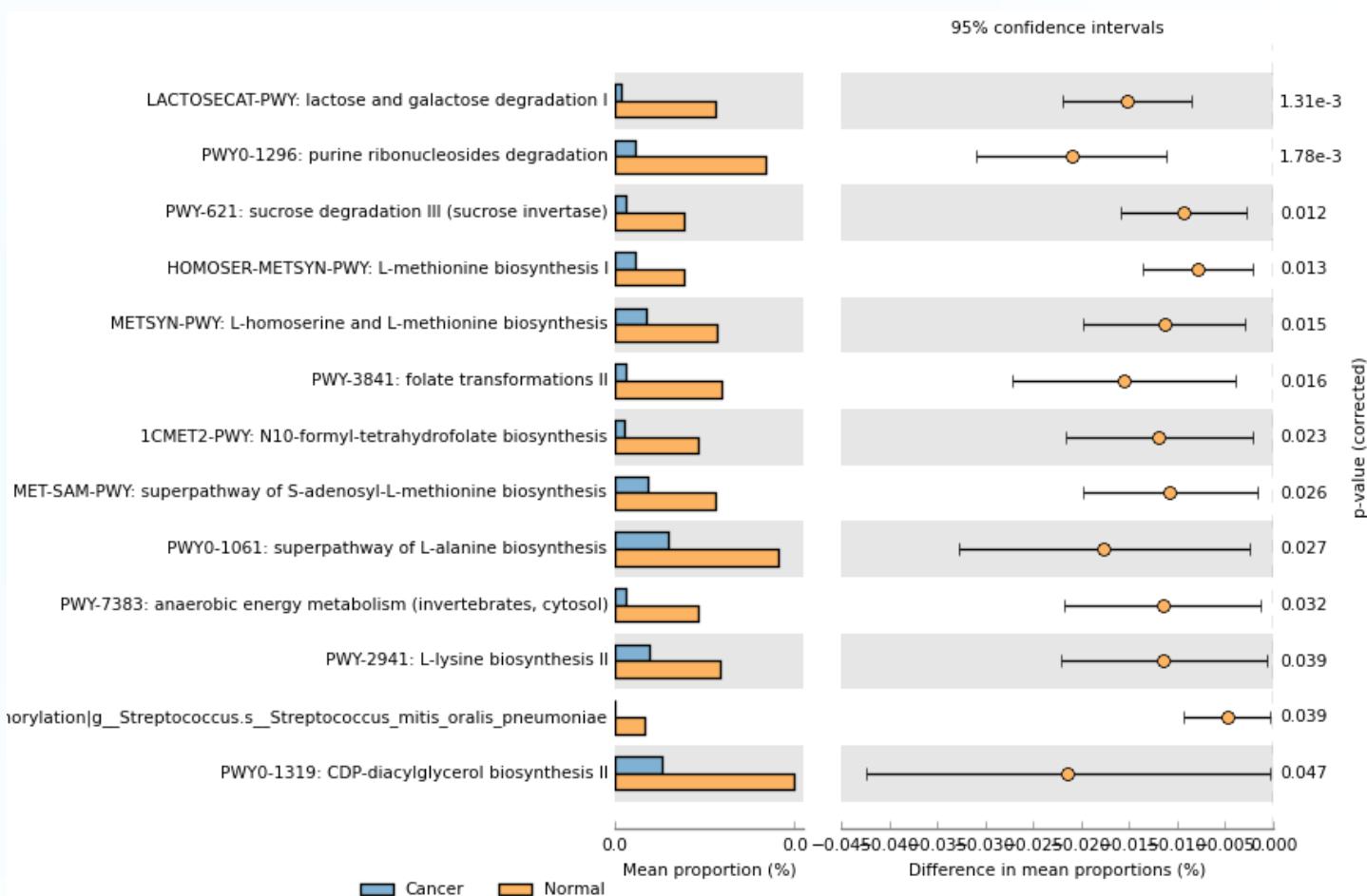
- 左侧多组比较(Multiple groups) – 校正多组检验(Multiple test correction) – None修改为Storey FDR(无差异可使用None或提高P值)
- 图片下方类型的主坐标轴分析(PCA plot)改为 箱线图(Box plot), 右侧勾选只显示差异>Show only active features, 选择具体条目展示如下



## 4. 两组比较——常用结果类型

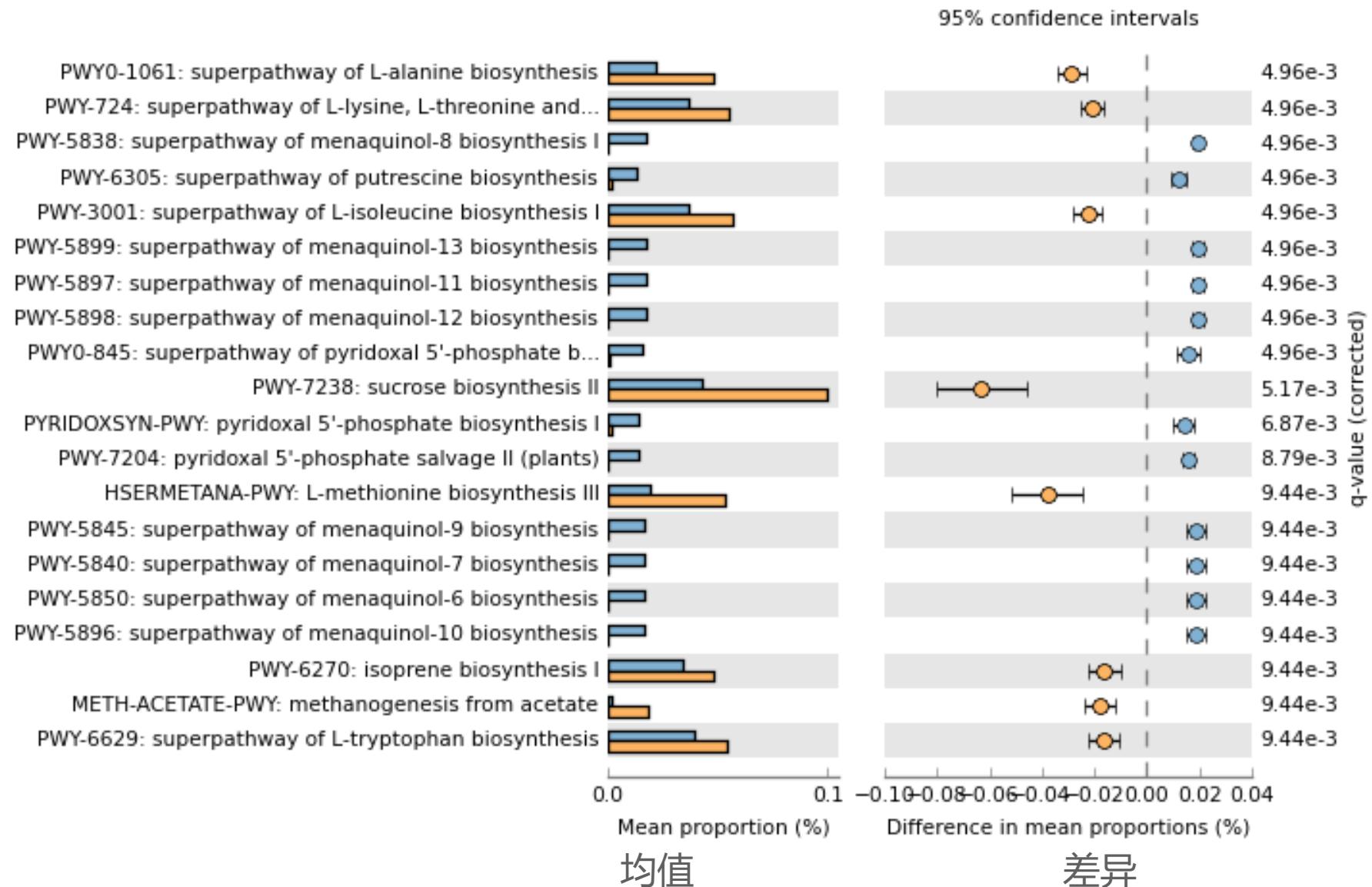


散点图(Scatter plot)——相关分析



扩展柱状图(Extended barplot)——所有差异

# STAMP结果组间差异功能扩展柱状图



## 5. 导出图表

- 图片在下方Configure plot中可设置图片大小、列宽、图例位置等细节；
- 注意bar / box / heatmap 中Field to plot选择百分比或原始数据
- 保存图片：File – Save plot (Ctrl + S)，保存类型默认png，推荐修改为pdf(矢量图更清晰、细节可用AI修改)
- 保存两组比较统计表：View – Two group statistics table (Ctrl + G)，菜单中默认保存所有，可勾选Show only active features只保存显著结果，格式有tsv, txt均可。
- 此外还有多组、两样品等统计，结果保存同两组类似。

# 总结

- GraPhlAn可绘制很美树状图，参数和输入格式复杂，被众多软件引用如LEfSe，配合export2graphlan、graphlan\_plot准备输出文件更方便；
- LEfSe分析物种组成差异，可以绘制非常漂亮的GraPhlAn图并标记差异，也可绘制差异特征柱状图；
- HUMAN3结果包括功能组成(非层化)和功能的具体物种组成(层化)，可用LEfSe、STAMP对各种结果表格进一步统计和可视化；
- STAMP安装要用默认目录，输入文件为丰度矩阵表+分组信息，路径不要有中文；两组比较推荐使用Welch's t-test + Extened barplot；导出图片为PDF格式方便AI修改；表格为tsv格式，方便Excel打开。

# 参考资源

- [宏基因组公众号文章目录](#) [生信宝典公众号文章目录](#)
- [iMeta | 易宏基因组\(EasyMetagenome\): 用户友好且灵活的宏基因组测序数据分析流程](#)
- [iMeta | MicrobiomeStatPlot 微生物组数据分析——50+篇](#)
- [iMetaOmics | 易扩增子\(EasyAmplicon\): 用户友好的扩增子测序数据分析指南](#)
- [Bio-protocol 《微生物组实验手册》——153篇](#)
- [Protein Cell: 扩增子和宏基因组数据分析实用指南](#)
- [CMJ: 人类微生物组研究设计、样本采集和生物信息分析指南](#)
- 加拿大生信网 <https://bioinformatics.ca/> 宏基因组课程中文版
- 美国高通量开源课程 <https://github.com/ngs-docs>
- Curtis Huttenhower <http://huttenhower.sph.harvard.edu/>
- Nicola Segata <http://segatalab.cibio.unitn.it/>



扫码关注生信宝典，学习更多生信知识



扫码关注宏基因组，获取专业学习资料

**易生信，没有难学的生信知识**