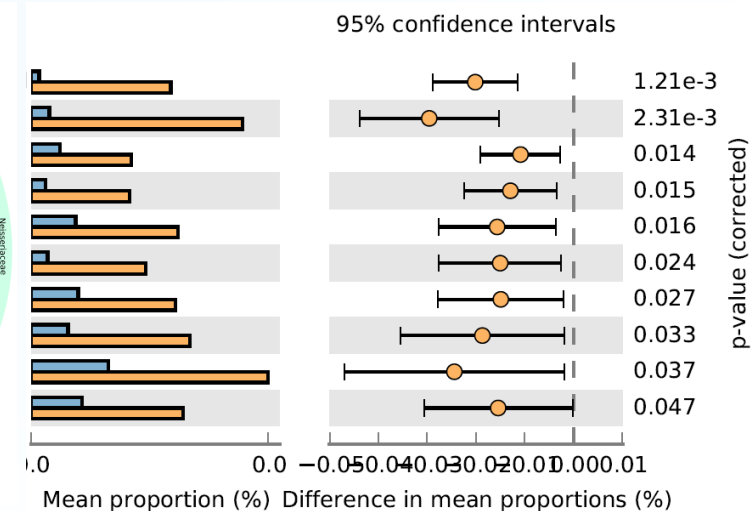
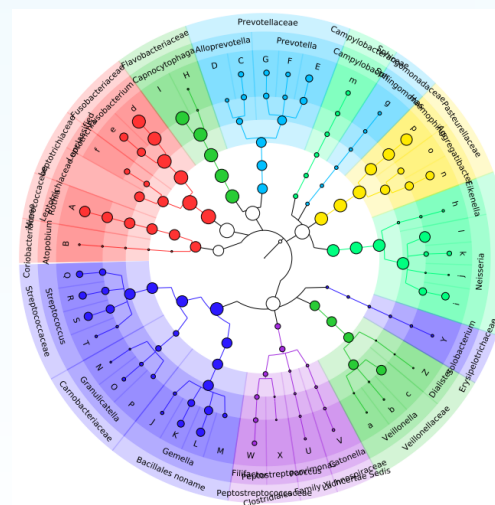




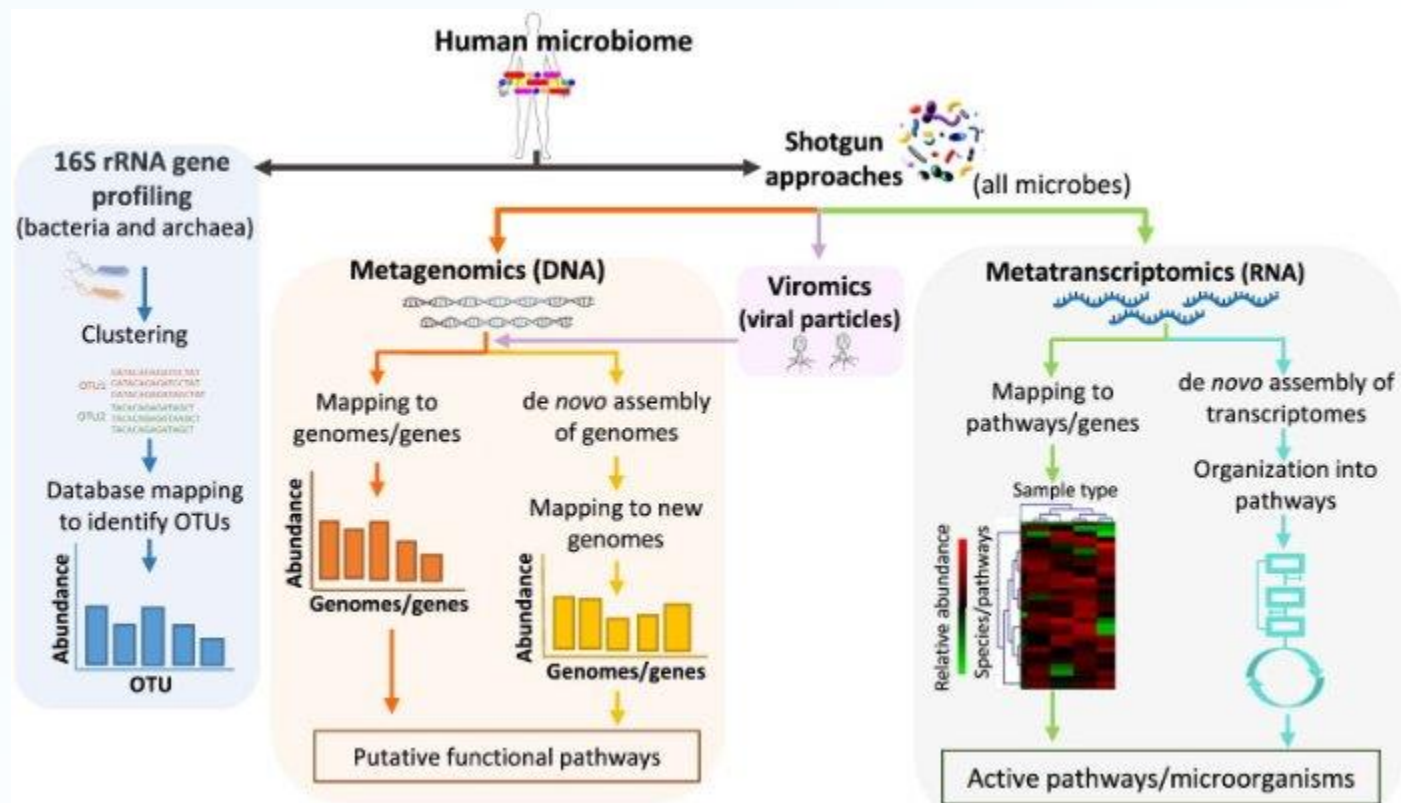
24物种和功能组成的 差异比较及可视化

易生信
2022年3月26日



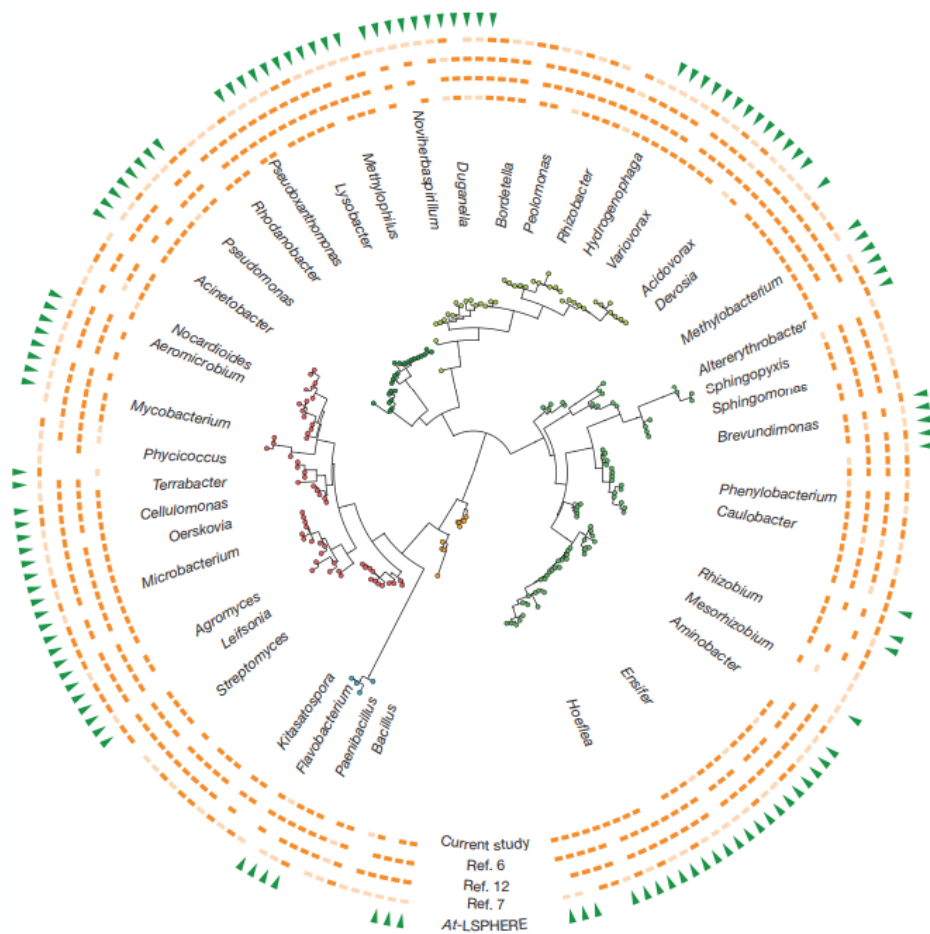
宏基因组有参流程——实战分析大纲

- 一. 软件安装和数据库部署
- 二. KneadData质控
- 三. MetaPhlAn2物种组成
- 四. HUMAnN2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

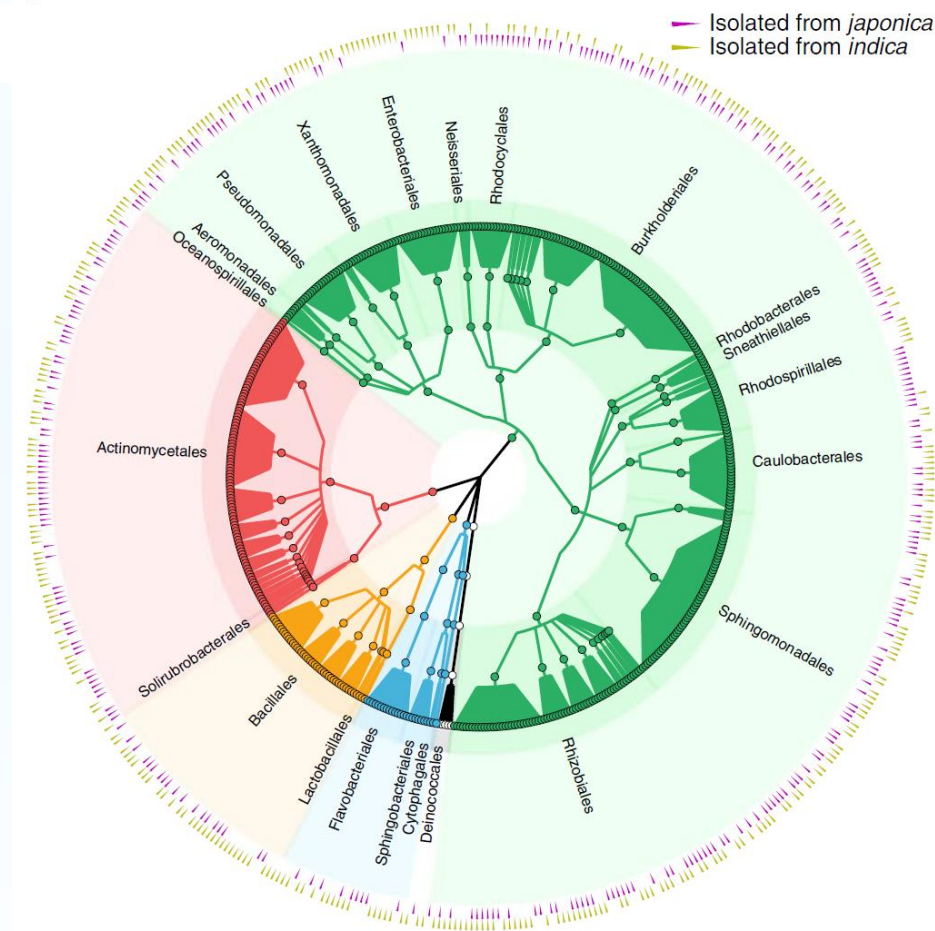


我们经常看到这样的图

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



- 软件由Huttenhower实验室编写，发布于2012年，15年发表

[HTML] Compact graphical representation of phylogenetic data and metadata with GraPhlAn

F Asnicar, G Weingart, TL Tickle, C Huttenhower... - PeerJ, 2015 - peerj.com

The increased availability of genomic and metagenomic data poses challenges at multiple analysis levels, including visualization of very large-scale microbial and microbial community data paired with rich metadata. We developed **GraPhlAn** (Graphical ...

☆ 被引用 340 次 相关文章 所有 15 个版本

Asnicar F, Weingart G, Tickle T L, et al. Compact graphical representation of phylogenetic data and metadata with GraPhlAn[J]. PeerJ, 2015, 3: e1029. <http://huttenhower.sph.harvard.edu/GraPhlAn>

- 它的引用被严重低估，两方面原因：发布3年没有引文，二是被众多软件内嵌，如LEfSe 2011发表已经引用5千次，容易被忽略。

[HTML] Metagenomic biomarker discovery and explanation

N Segata, J Izard, L Waldron... - **Genome** ..., 2011 - genomebiology.biomedcentral.com

... Search all BMC articles Search. **Genome Biology**.

☆ 被引用 4961 次 相关文章 所有 24 个版本



- <http://huttenhower.sph.harvard.edu/GraPhlAn>
- GraPhlAn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
- # GraPhlAn核心程序包
conda install graphlan
- # GraPhlAn输入文件制作程序，如转换LEfSe、MetaPhlAn2结果格式化为GraPhlAn输入文件
conda install export2graphlan



绘制物种树(Cladogram)——GraPhlAn

转换MetaPhlAn2结果为GraPhlAn输入

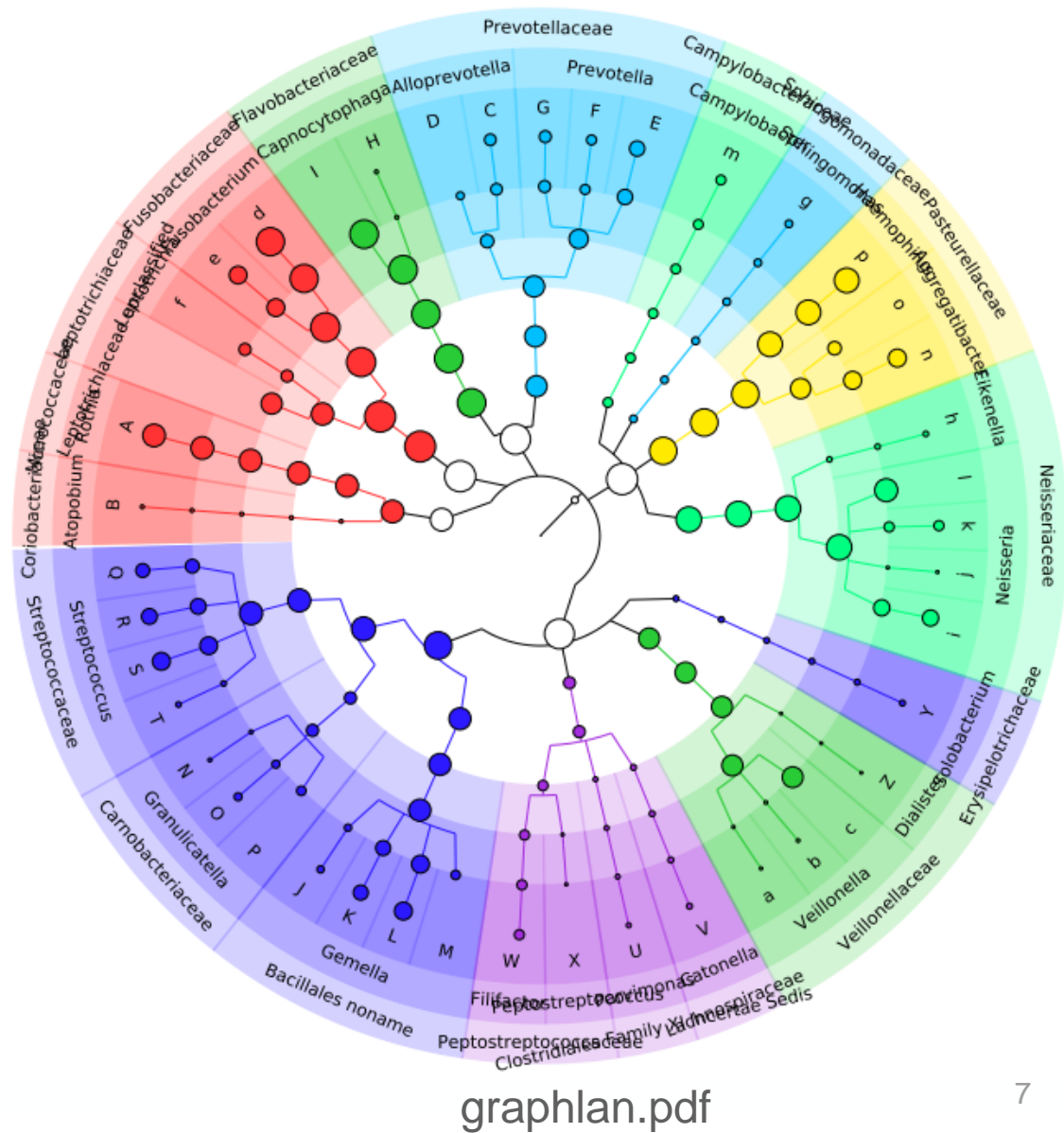
```
export2graphlan.py --skip_rows 1,2 -i result/metaphlan2/taxonomy.tsv \  
--tree temp/merged_abundance.tree.txt \  
--annotation temp/merged_abundance.annot.txt \  
--most_abundant 1000 --abundance_threshold 1 --least_biomarkers 10 \  
--annotations 5,6 --external_annotations 7 --min_clade_size 1
```

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

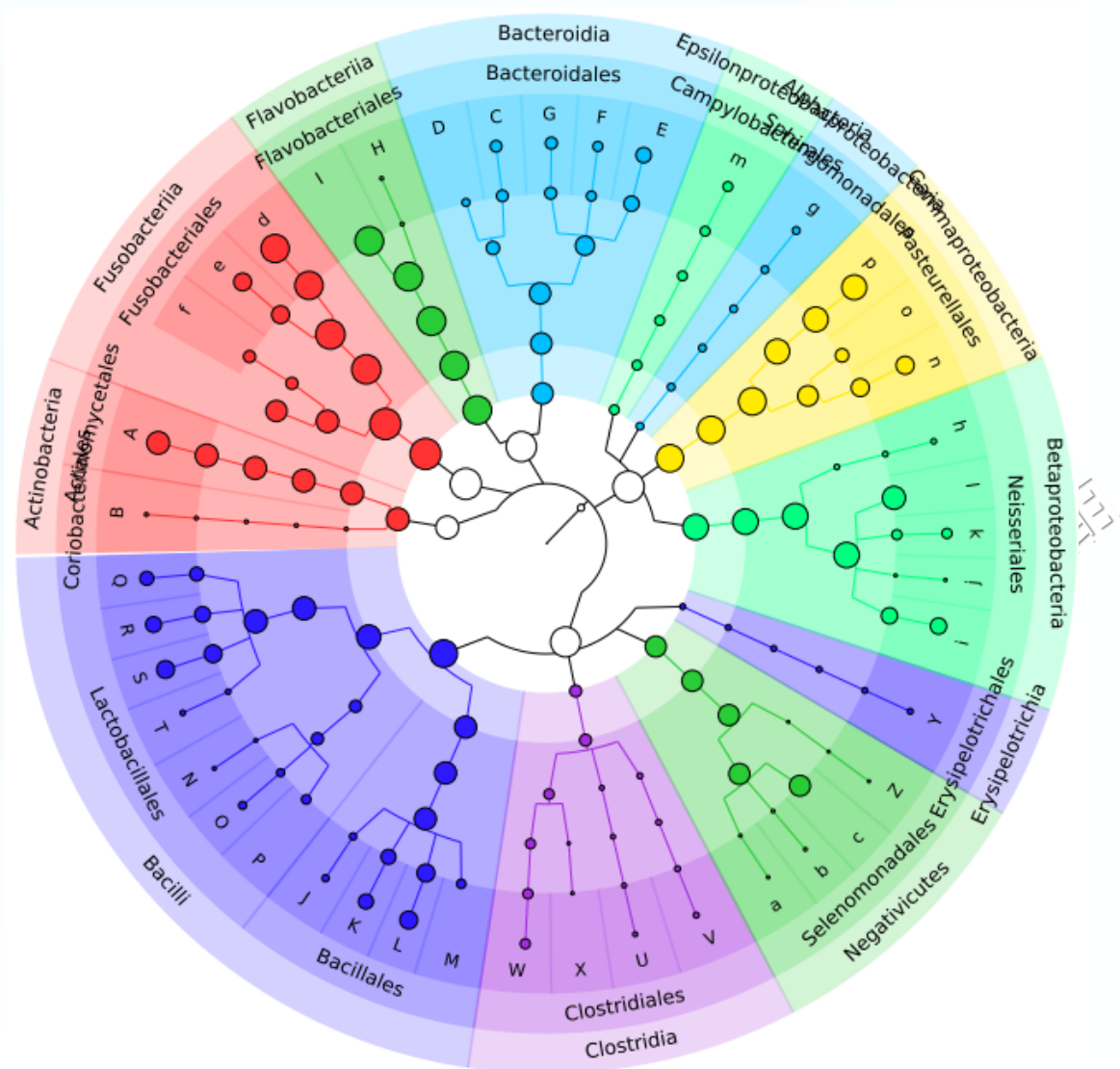
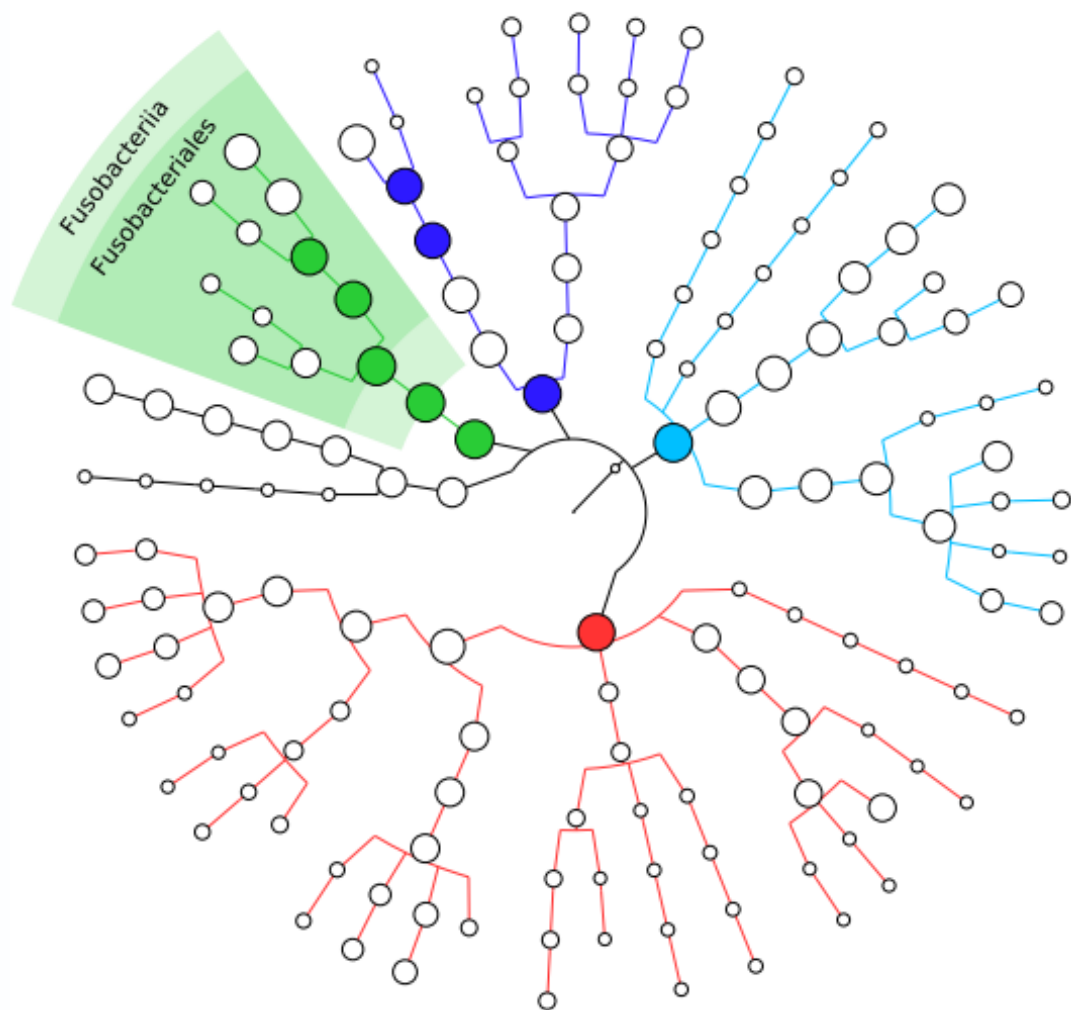
数据准备完成，一键出图

```
# 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
```

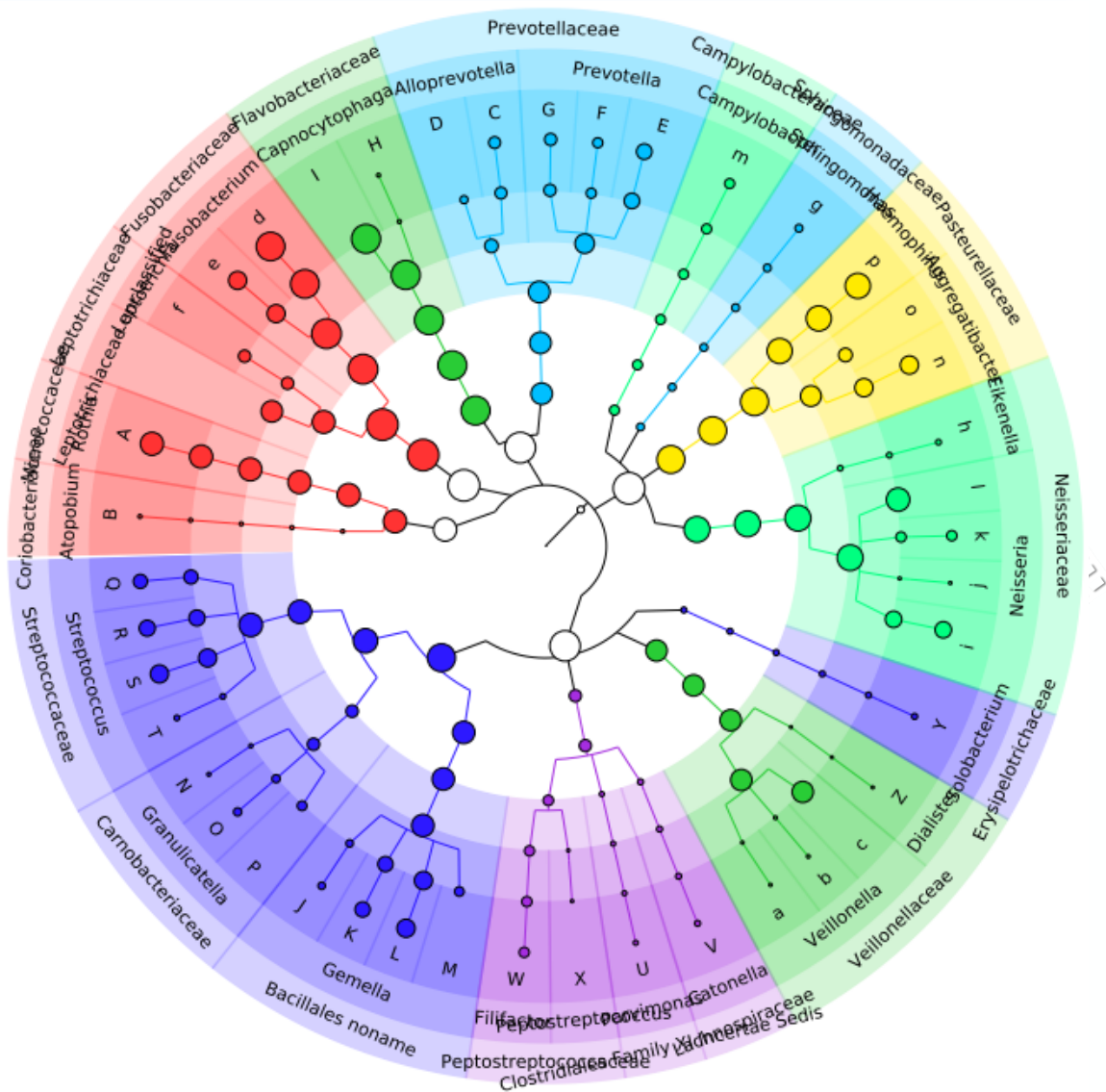
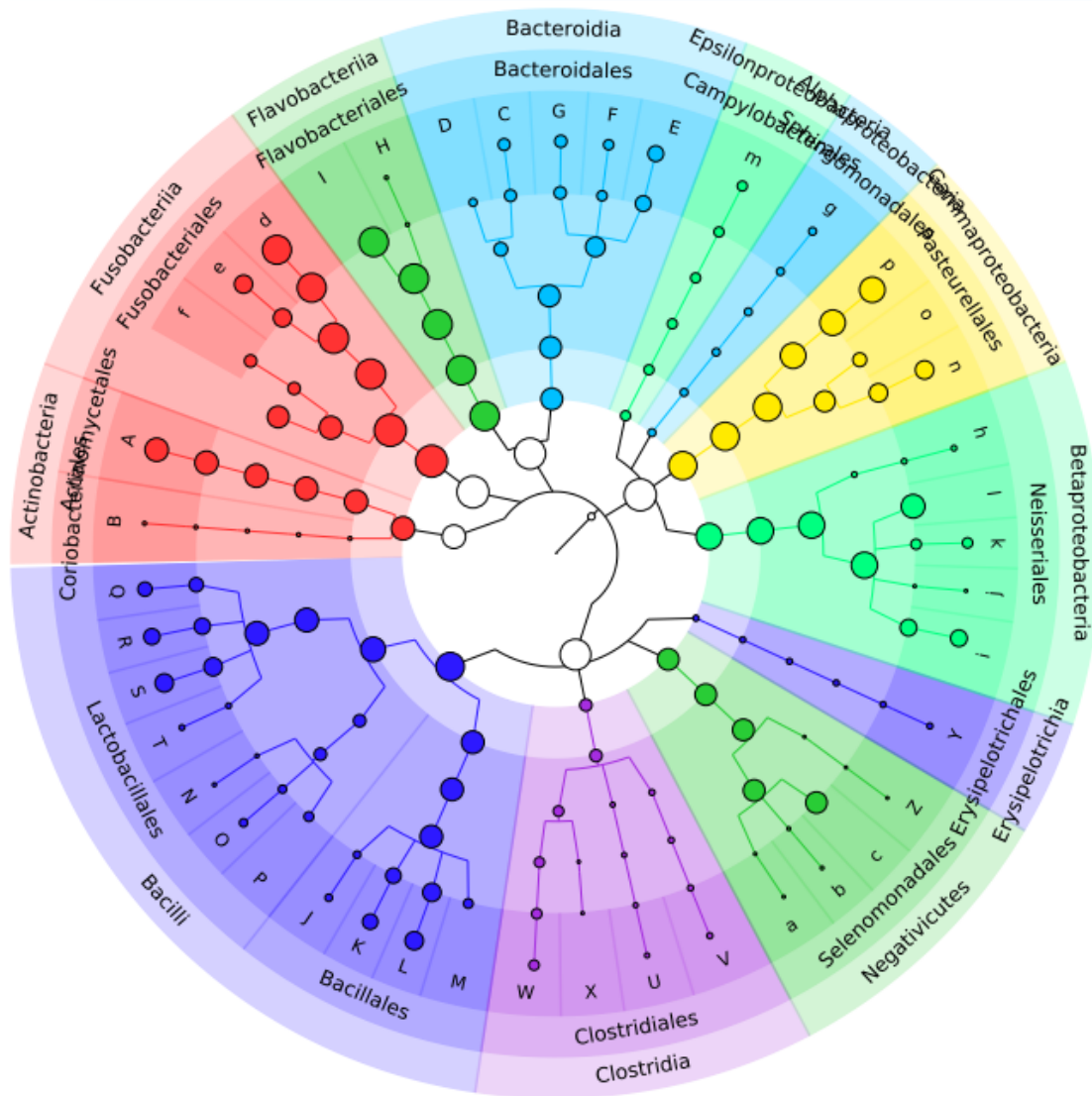


--most_abundant为注释类数量，默认10，推荐1000



--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

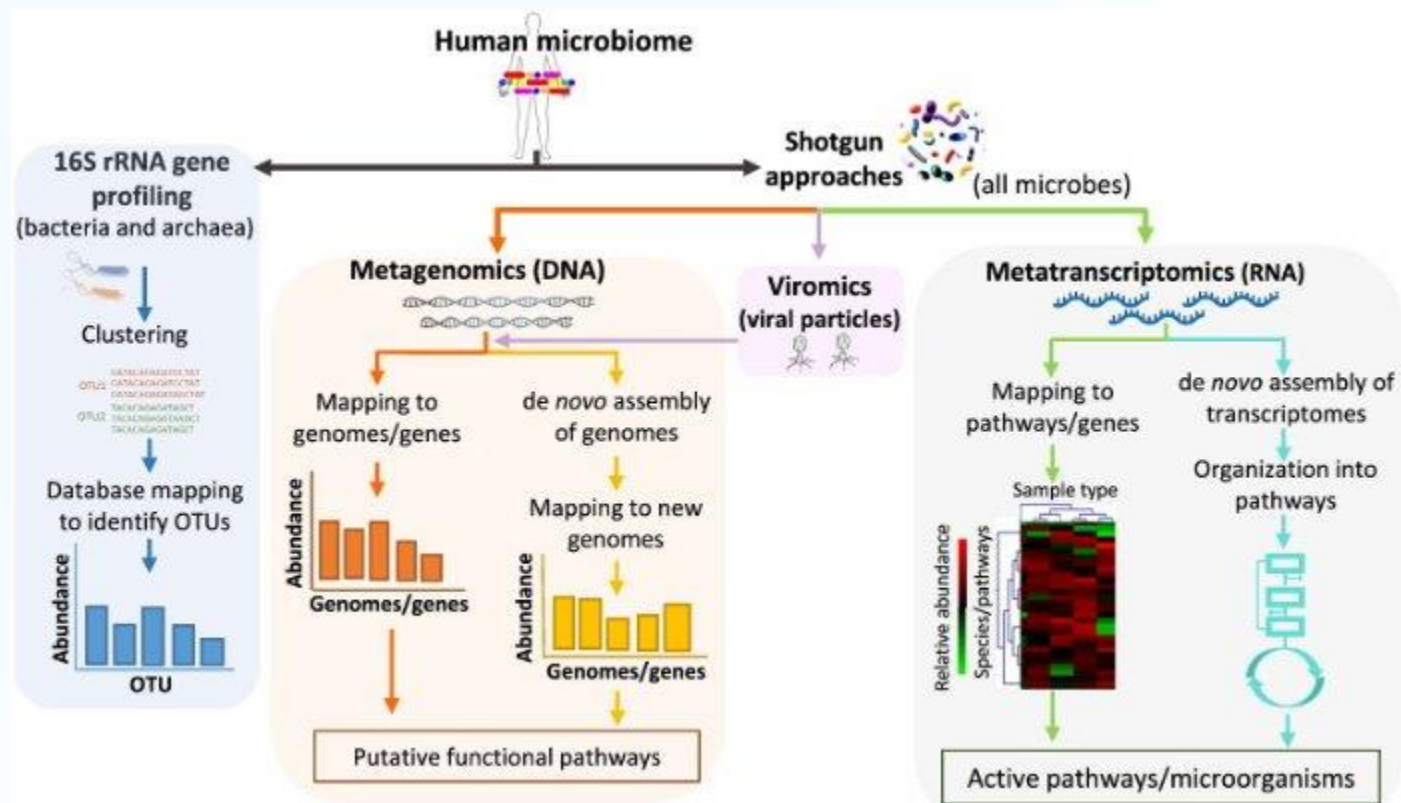
A: Rothia mucilaginosa
 B: Atopobium parvulum
 C: Alloprevotella tannerae
 D: Alloprevotella unclassified
 E: Prevotella melaninogenica
 F: Prevotella nanceiensis
 G: Prevotella nigrescens
 H: Capnocytophaga gingivalis
 I: Capnocytophaga unclassified
 J: Gemella haemolysans
 K: Gemella morbillorum
 L: 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: Parvimonas 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 flavescens
 j: Neisseria meningitidis
 k: Neisseria subflava
 l: Neisseria unclassified
 m: Campylobacter showae
 n: Aggregatibacter segnis
 o: Aggregatibacter unclassified
 p: Haemophilus parainfluenzae

● ACTINOBACTERIA
 ● ALPHAPROTEOBACTERIA
 ● BACILLI
 ● BACTEROIDIA
 ● BETAPROTEOBACTERIA
 ● CLOSTRIDIA
 ● EPSILONPROTEOBACTERIA
 ● ERYSIPELOTRICHIA
 ● FLAVOBACTERIIA
 ● FUSOBACTERIIA
 ● GAMMAPROTEOBACTERIA
 ● NEGATIVICUTES

根据需求在Adobe Illustrator中整合排版

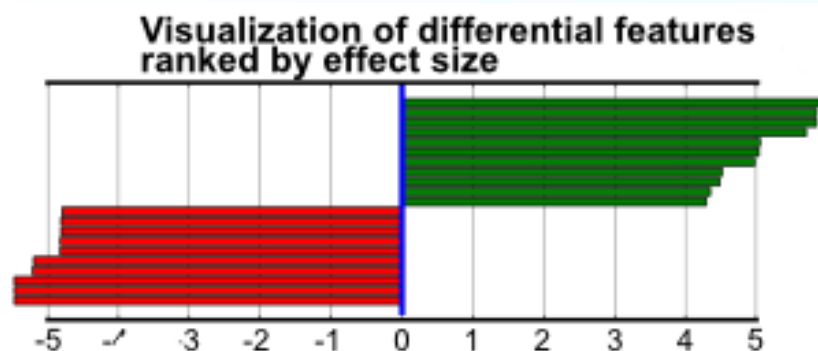
宏基因组有参流程——实战分析大纲

- 一. 软件安装和数据库部署
- 二. KneadData质控
- 三. MetaPhlAn2物种组成
- 四. HUMAnN2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

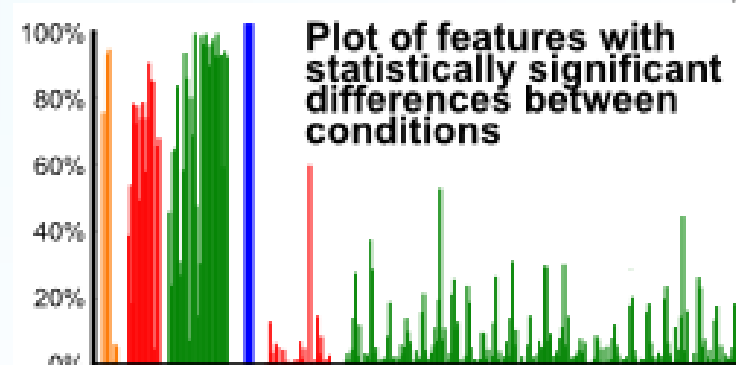
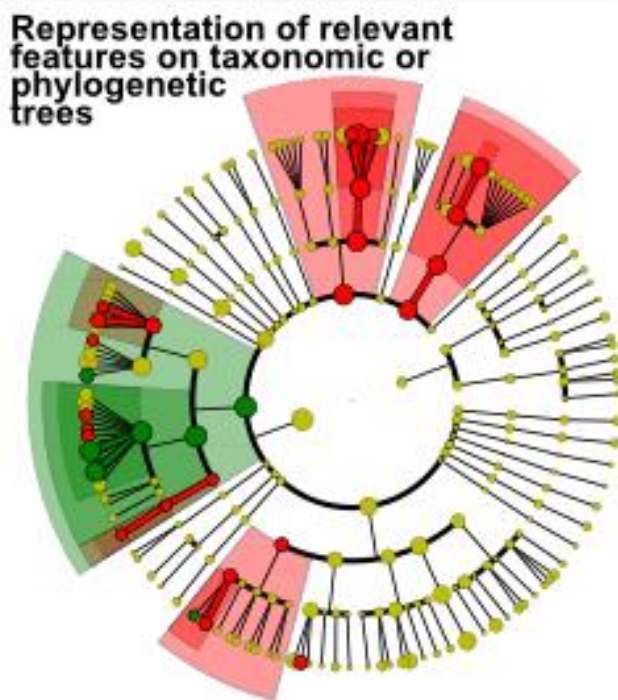


Linear discriminant analysis **Effect Size** (LEfSe)

- 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)。

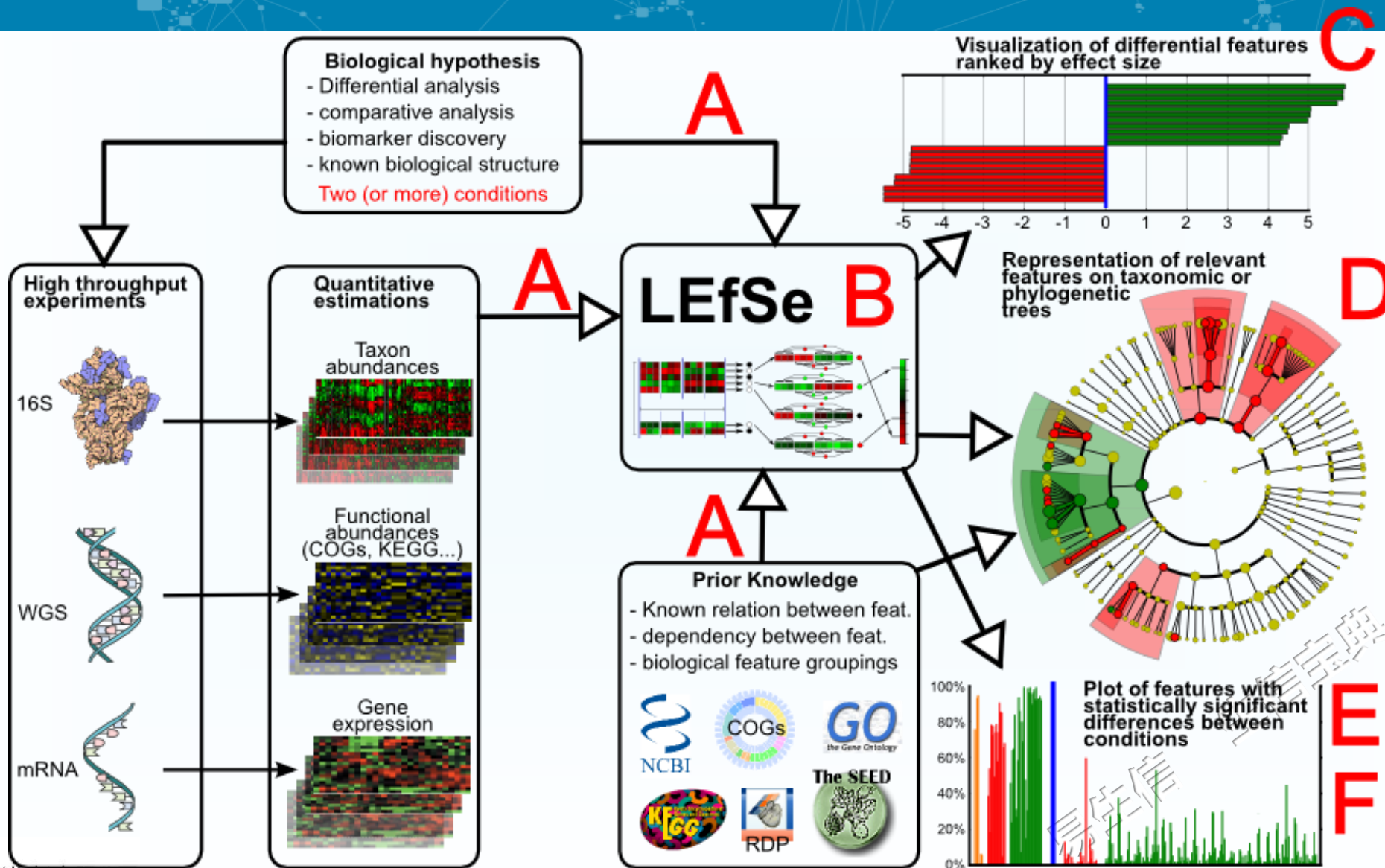


Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome biology*, 12(6), R60.



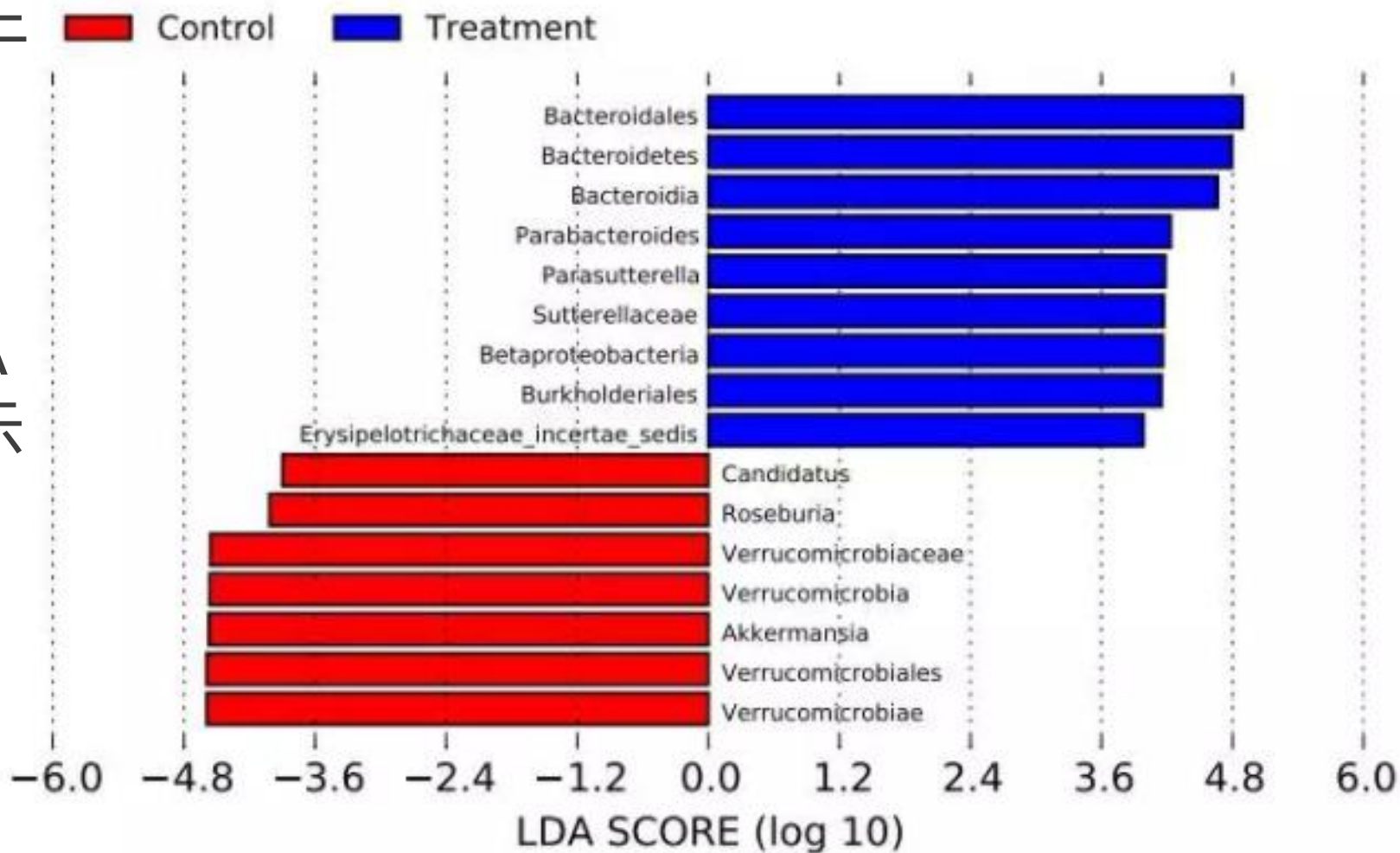
- A) Format Data for LEfSe:** 数据格式转换, 将OTU表转换为LEfSe专用格式
- B) LDA Effect Size (LEfSe):** 对A的结果进行线性分歧分析, 并生成用于下面可视化的数据(C, D, E, F)
- C) Plot LEfSe Results:** 绘制发现生物标志(Biomarkers)的效应大小
- D) Plot Cladogram:** 采用GraPhlAn绘制物种树, 并展示生物标志
- E) Plot One Feature:** 绘制单个生物标志在组间的差异
- F) Plot Differential Features:** 绘制所有的Features

LEfSe分析流程



1. LDA值分布柱状图

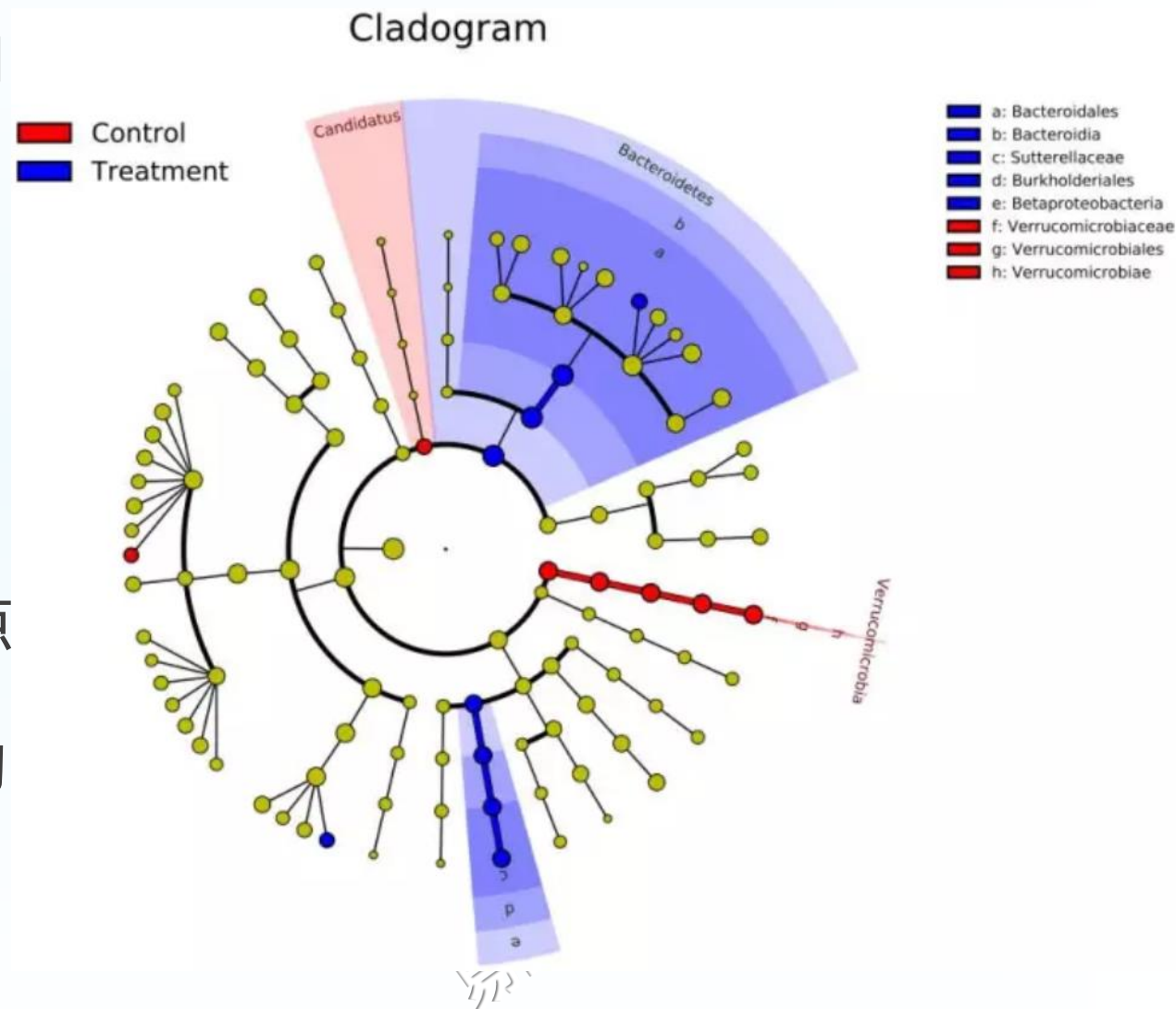
条形图展示LDA score大于预设值的显著差异物种，即具有统计学差异的Biomaker，默认预设值为**2.0**（看横坐标，只有LDA值的绝对值大于2才会显示在图中）；柱状图的颜色代表各自的组别，长短代表的是**LDA score**，即不同组间显著差异物种的影响程度。



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）。不同分类级别上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）。



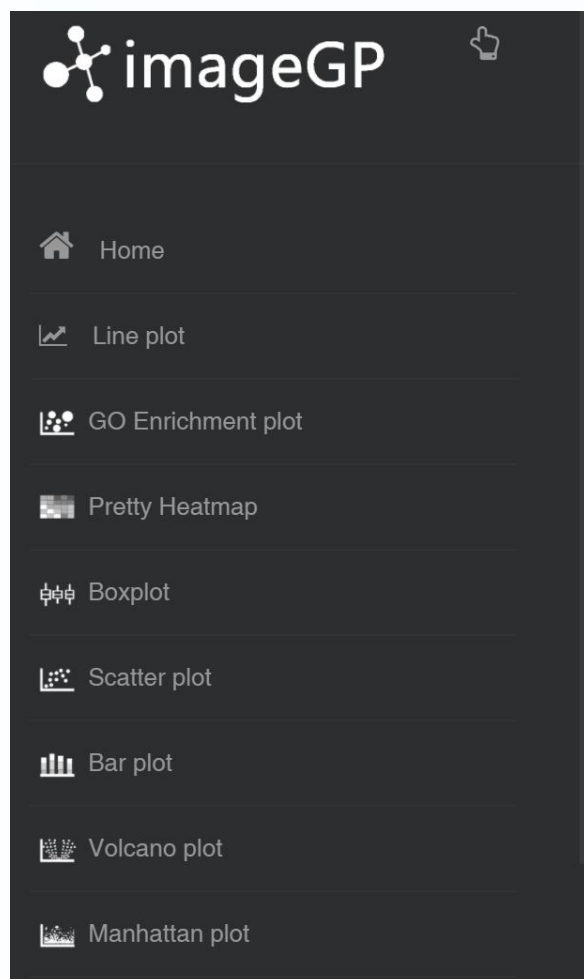
3. 特征表

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——LEfSe——一键分析



Make plot be easy and exquisite woodcorpse@163.com

LEfSe

添加新功能, 请扫码拉您入群

Please refer to [LEfSe](#) for input formats, results explanation and citation information.

如果操作中遇到问题

LEfSe format as described [here](#).

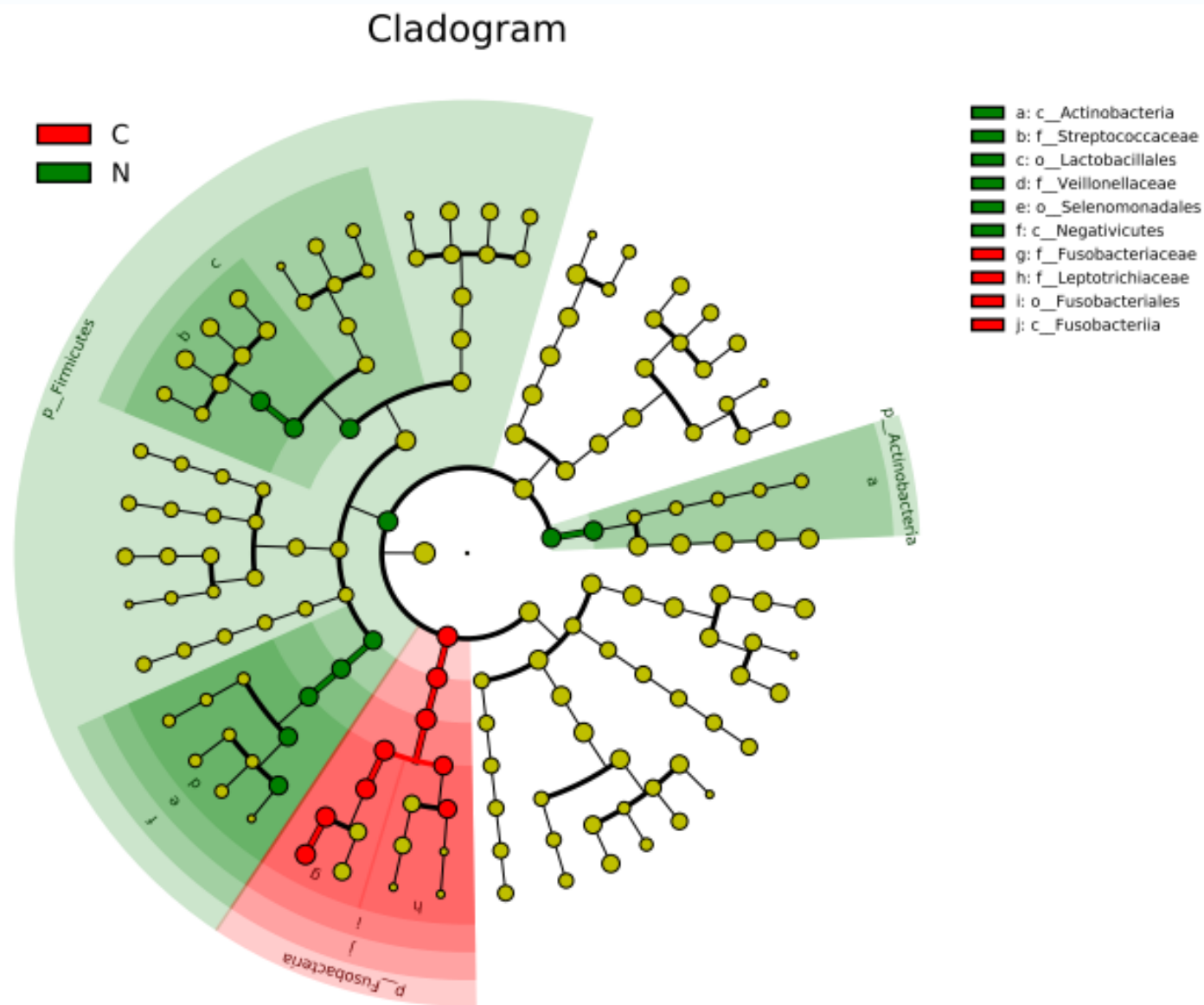
Class	A	A	A	A	A	A	B
B	B	B	B	B	C	C	C
C	C	C					
k__Bacteria	91.5407662395614	92.4308234365034	93.6857097642032				
94.3880277926243	95.4549126255419	92.8378073539364	88.1886073180639				
89.4735168817956	91.731142643764	91.5340040753948	90.4902050951899				
87.5034954171198	93.0211707542804	94.1412145660108	91.4397666170919				
92.3633648295471	93.581570955651	92.2621261983116					
k__Bacteria p__Acidobacteria	0.172117039586919	0.139719952616712					
0.126602310492166	0.143463951165996	0.105010904978594	0.197383261899391				
0.474955472924413	0.24162268709862	0.316654219566841	0.156011207335711				
0.2697814115739293	0.466055615970172	0.283433595557612	0.164649707746769				

<http://www.ehbio.com/ImageGP/index.php/Home/Index/LEfSe.html>

Tong Chen, Yong-Xin Liu, Luqi Huang. 2022. ImageGP: An easy-to-use data visualization web server for scientific researchers. *iMeta* 1: e5. <https://doi.org/10.1002/imt2.5>

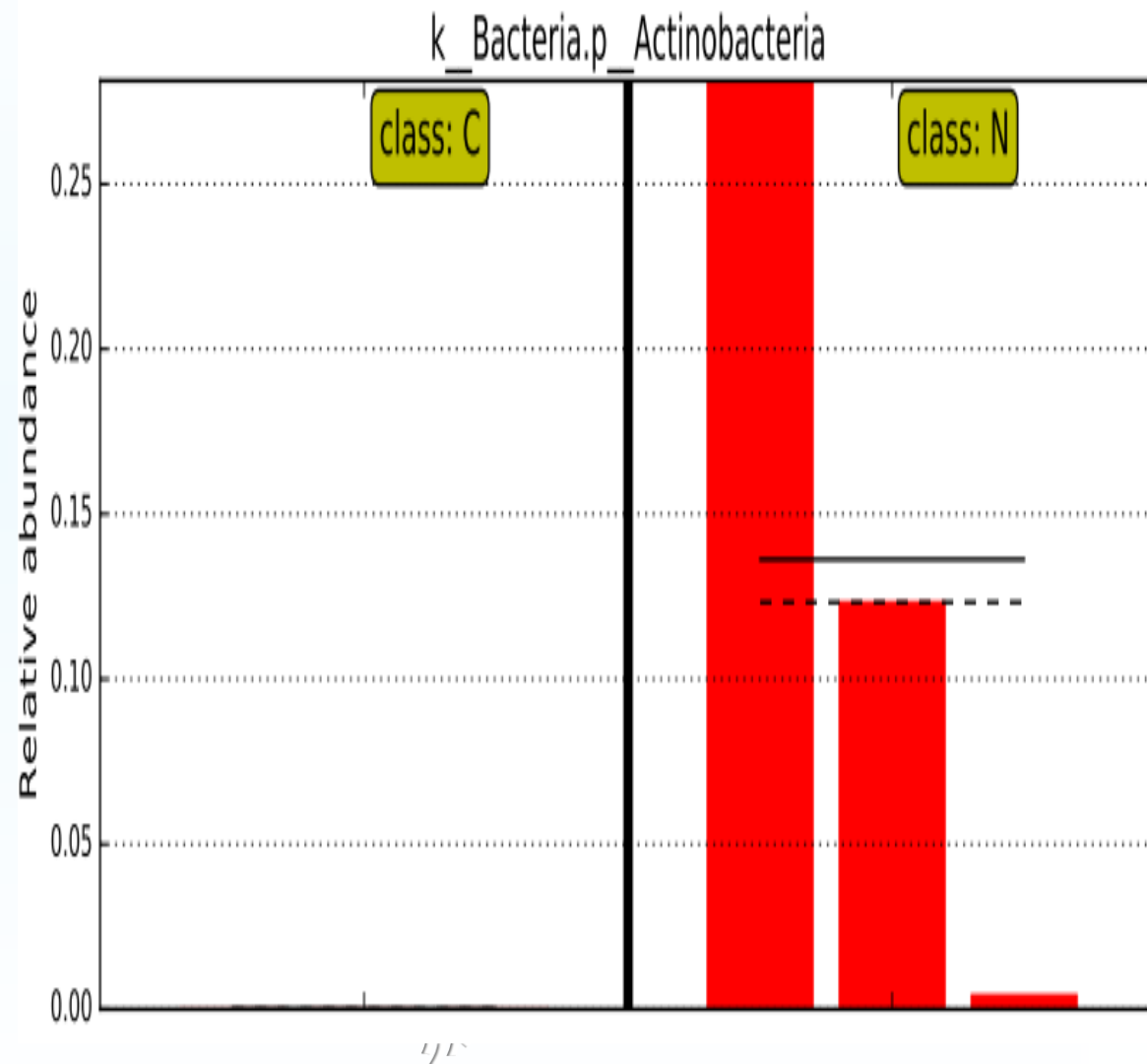
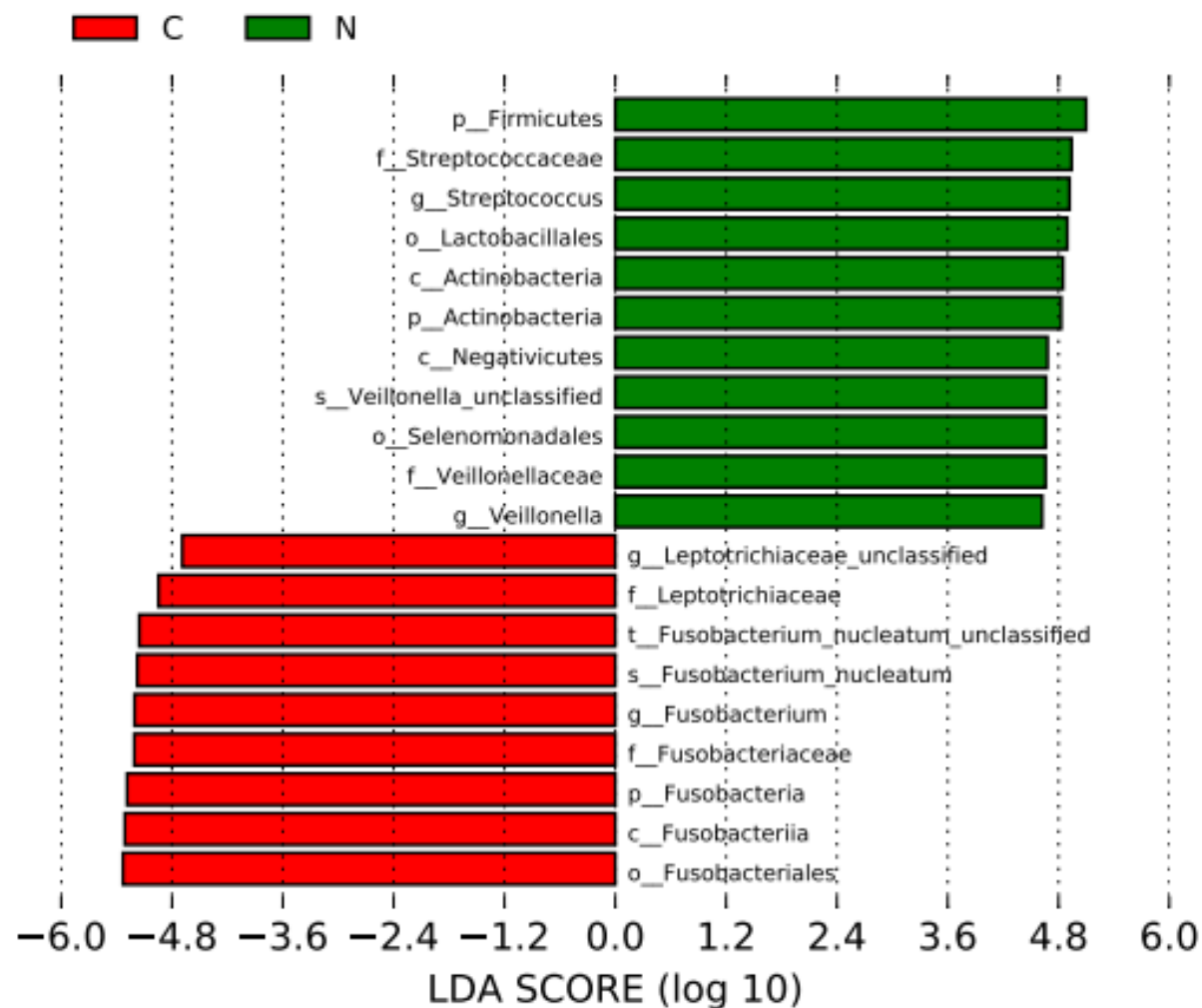


物种树展示各级别差异



信信與
宏基因组

差异features柱状图(总体-个体)



(选学)Linux服务器运行LEfSe代码 ——LEfSe差异分析和绘制Cladogram

- # 修改样本名为组名

```
sed '1 s/p[0-9]*//g' metaphlan2/taxonomy.tsv | grep -v '#' > metaphlan2/lefse.txt
```

- # 格式转换为lefse内部格式

```
lefse-format_input.py metaphlan2/lefse.txt temp/input.in -c 1 -o 1000000
```

- # 运行lefse

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

- # 绘制物种树注释差异

```
lefse-plot_cladogram.py temp/input.res \  
    metaphlan2/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 metaphlan2/lefse_res.pdf --format pdf --dpi 600
```

- # 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 --feature_name "k__Bacteria.p__Firmicutes" \  
--format pdf temp/input.in temp/input.res metaphlan2/lefse_Firmicutes.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 metaphlan2/lefse_
```

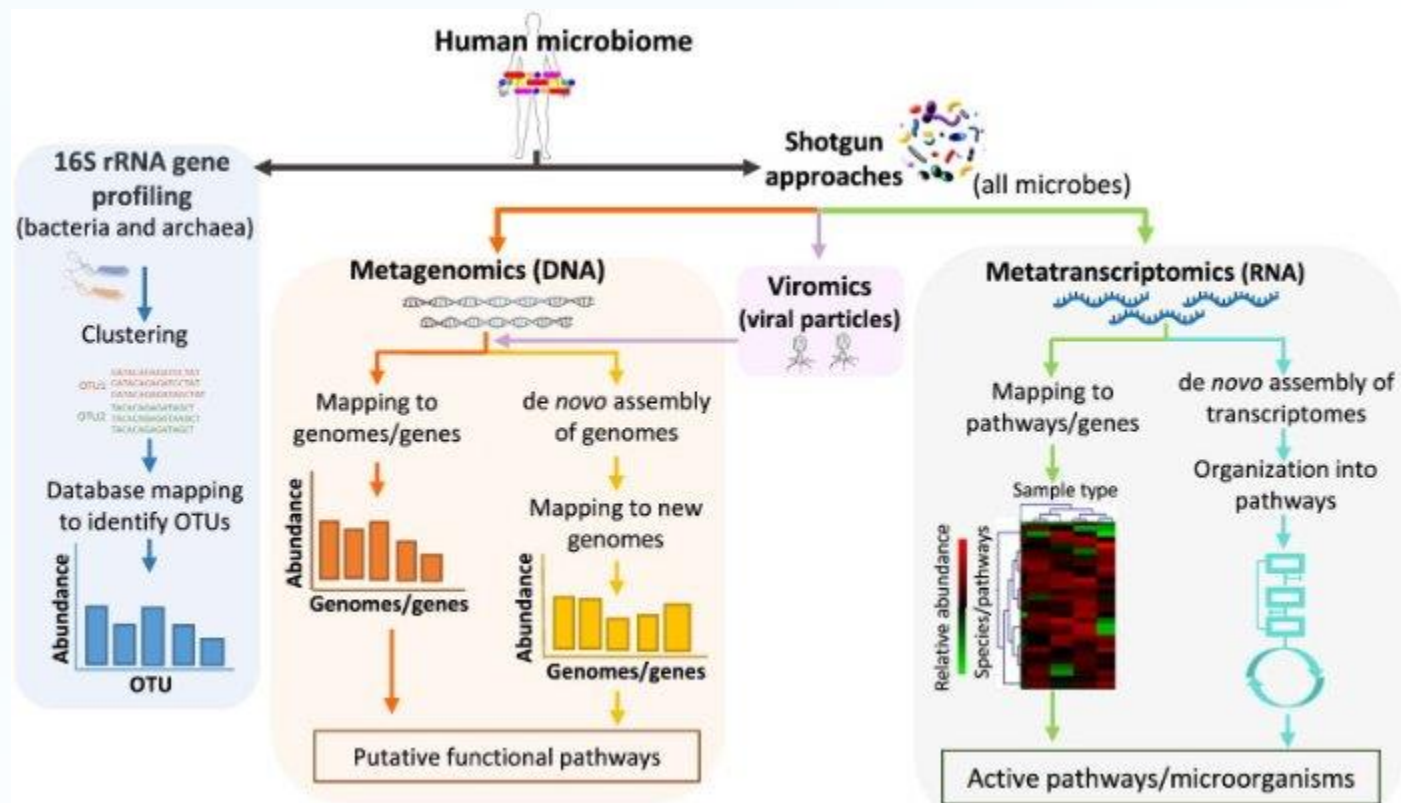
LEfSe官网在线使用——图文教程见文末附录

- 主页：<http://huttenhower.sph.harvard.edu/galaxy>，不仅有LEfSe工具，还包括PICRUST、GraPhlAn、MetaPhlAn、MetaPhlAn2、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. MetaPhlAn2物种组成
- 四. HUMAnN2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**



HUMAnN2获得功能组成表

- 通路丰度表pathabundance.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



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 supported 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>

Identifying biologically relevant differences between metagenomic communities [DH Parks](#), [RG Beiko](#) - Bioinformatics, 2010 - [academic.oup.com](#)

Motivation: Metagenomics is the study of genetic material recovered directly from environmental samples. Taxonomic and functional differences between metagenomic samples can highlight the influence of ecological factors on patterns of microbial life in a wide range of habitats. Statistical hypothesis tests can help us distinguish ecological influences from sampling artifacts, but knowledge of only the P-value from a statistical hypothesis test is insufficient to make inferences about biological relevance. Current ...

☆ Cited by 753 Related articles All 15 versions

STAMP: statistical analysis of taxonomic and functional profiles

[DH Parks](#), [GW Tyson](#), [P Hugenholtz](#), [RG Beiko](#) - Bioinformatics, 2014 - [academic.oup.com](#)

STAMP is a graphical software package that provides statistical hypothesis tests and exploratory plots for analysing taxonomic and functional profiles. It supports tests for comparing pairs of samples or samples organized into two or more treatment groups. Effect sizes and confidence intervals are provided to allow critical assessment of the biological relevancy of test results. A user-friendly graphical interface permits easy exploration of statistical results and generation of publication-quality plots. Availability and implementation ...

☆ Cited by 1594 Related articles All 12 versions



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。

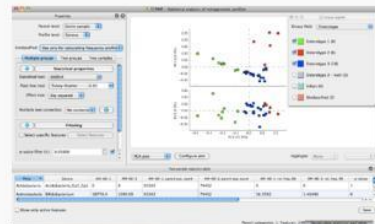
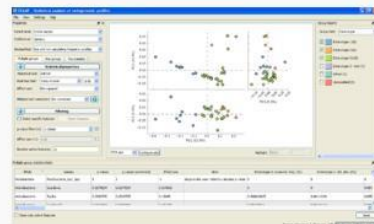
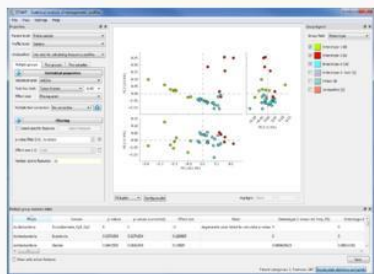
# Pathway	C1	C2	C3
1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis	3.30	0.00	0.00
ANAEROFRUCAT-PWY: homolactic fermentation	4.88	0.00	3.56
ANAGLYCOLYSIS-PWY: glycolysis III (from glucose)	12.78	3.38	3.42
ANAGLYCOLYSIS-PWY: glycolysis III (from glucose) g_Fusobacterium.s	5.00	0.00	0.00
ANAGLYCOLYSIS-PWY: 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

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

特征表 pathabundance.tsv

元数据(分组信息) metadata.txt

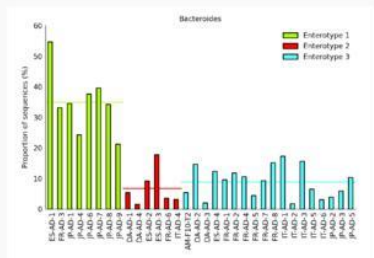
STAMP可用的图表类型



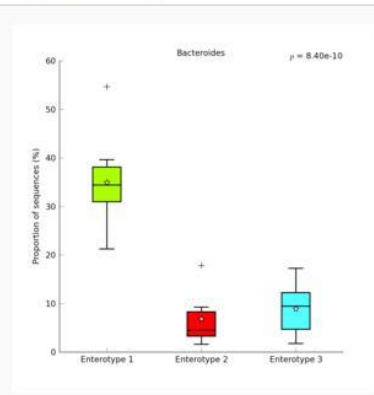
STAMP v2.0.0 on Microsoft Windows 7 (x64).

STAMP v2.0.0 on Microsoft Windows XP.

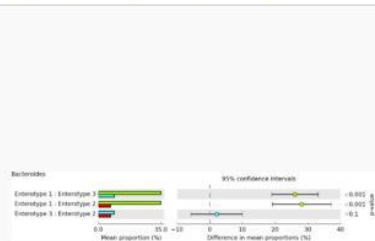
STAMP v2.0.0 on Apple's Mac OS X Leopard.



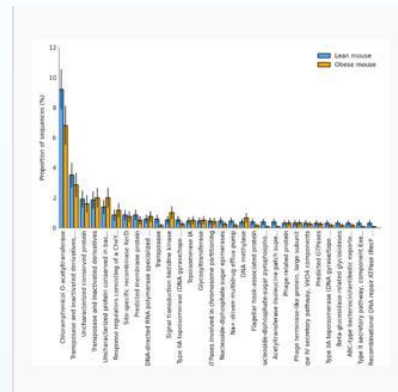
Bar plot showing the abundance of *Bacteroides* within the gut microbiota of individuals assigned to the 3 enterotypes proposed by Arumugam and colleagues (data described in Arumugam et al., 2011).



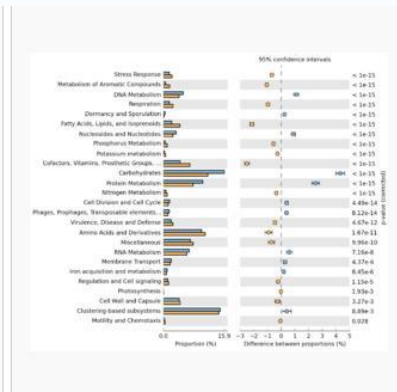
Box plot showing the abundance of *Bacteroides* within the gut microbiota of individuals assigned to the 3 enterotypes proposed by Arumugam and colleagues (data described in Arumugam et al., 2011).



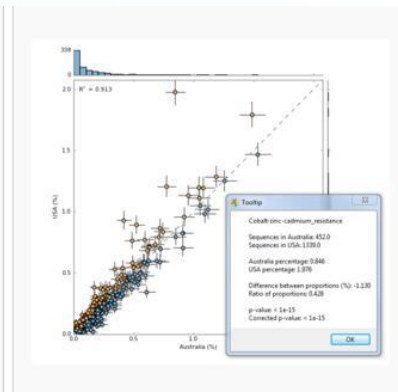
Post-hoc plot indicating that *Bacteroides* is significantly over-represented in enterotype 1 compared to the 2 other enterotypes proposed by Arumugam and colleagues (data described in Arumugam et al., 2011).



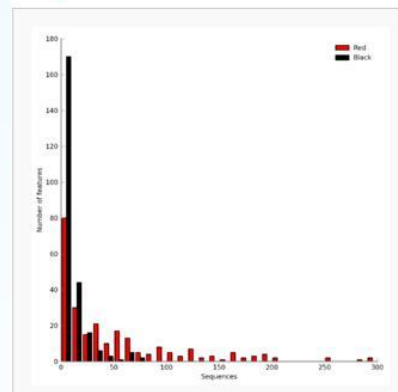
Functional profile plot for an obese and a lean mouse microbiome (data described in Turnbaugh et al., 2006).



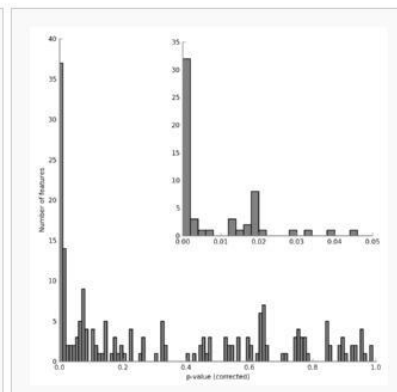
Extended error bar plot for a pair of bovine rumen microbiomes (data described in Brulic et al., 2009).



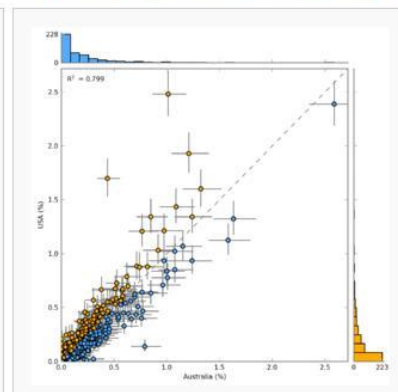
SEED subsystem scatter plot for a pair of enhanced biological phosphorus removal (EBPR) sludge metagenomes (data described in Martin et al., 2006).



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



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



Scatterplot with histograms showing the functional profile of *A.phosphatis* sequences from two EBPR communities (data described in Parks and Beiko, 2010).

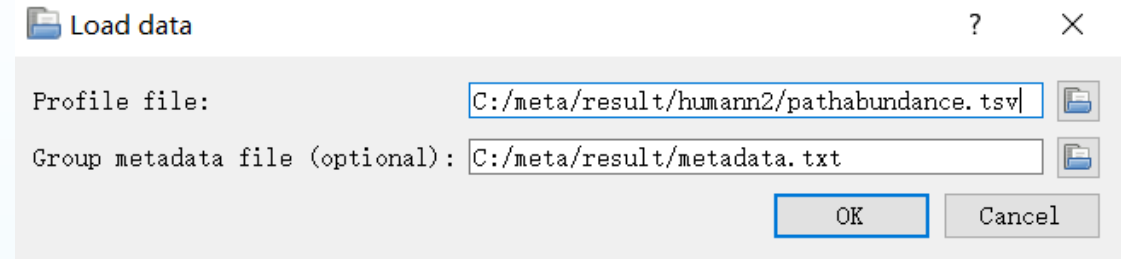
1. STAMP安装和运行

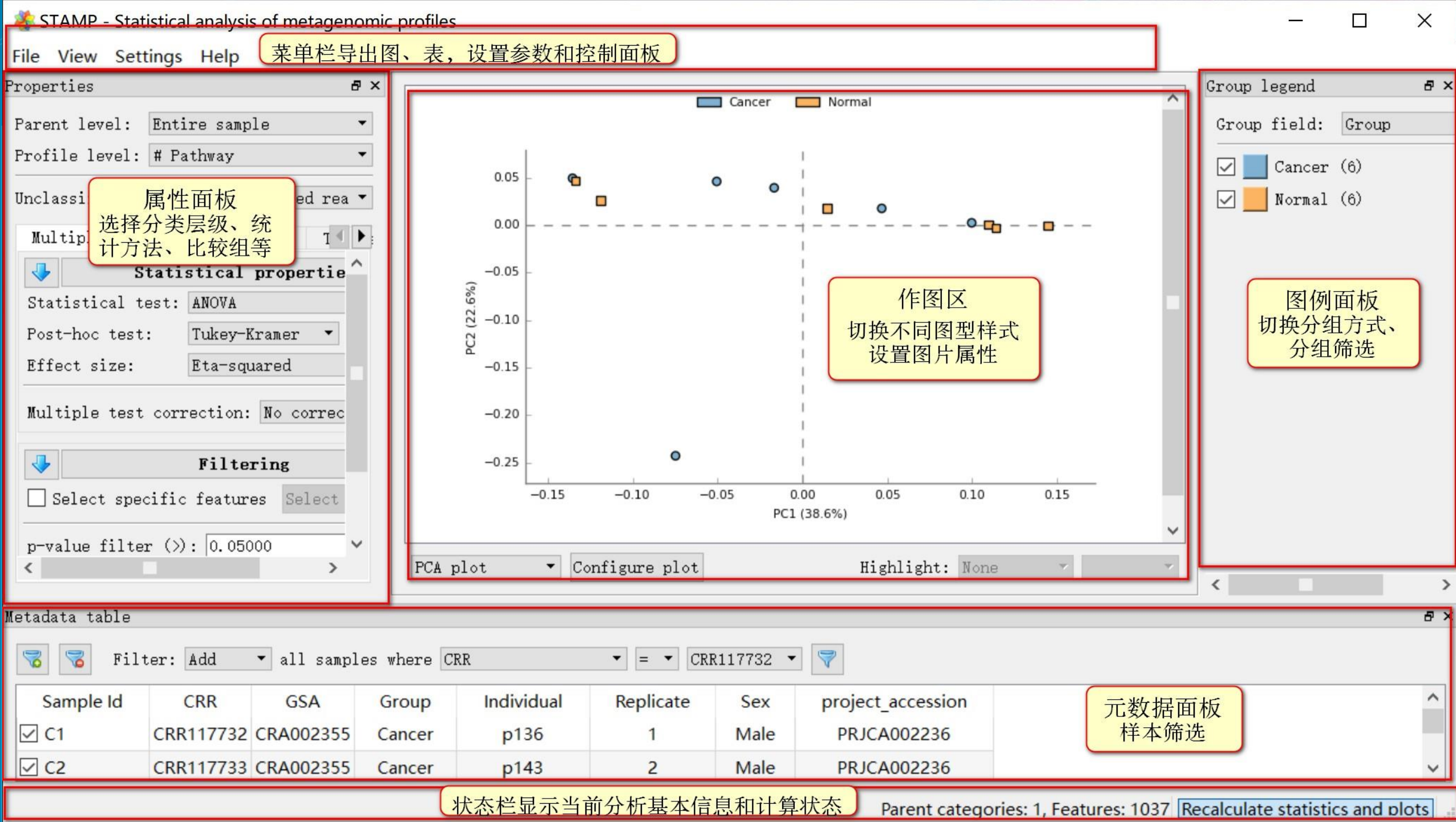
- <http://kiwi.cs.dal.ca/Software/STAMP> 下载系统对应版本;
- Windows下按要求一路安装即可(安装目录不要出现中文);
- 安装后程序名为STAMP的花瓣图标, 双击运行;
- Mac下安装麻烦



2. 打开文件 —— 默认主成分分析

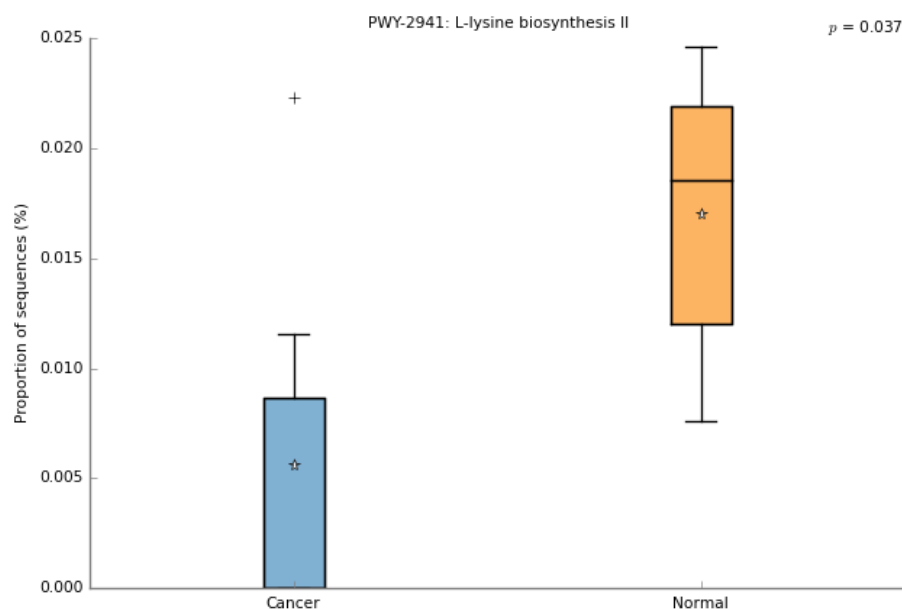
- File – Load data
- 选择**特征表**和元数据，点OK
- 左侧Properties属性面板选择分类层级、统计方法、比较组等；
- 右侧Group legend图例面板切换分组方式、分组筛选；
- 底部Metadata table元数据面板可以进行样本筛选；
- 中间作图区，可切换不同图型样式，设置图片属性；
- 最上面菜单栏可以导出图、表，设置参数和控制面板开关；
- 最下面状态栏显示当前分析基本信息和计算状态。





3. 多组比较 —— 箱线图

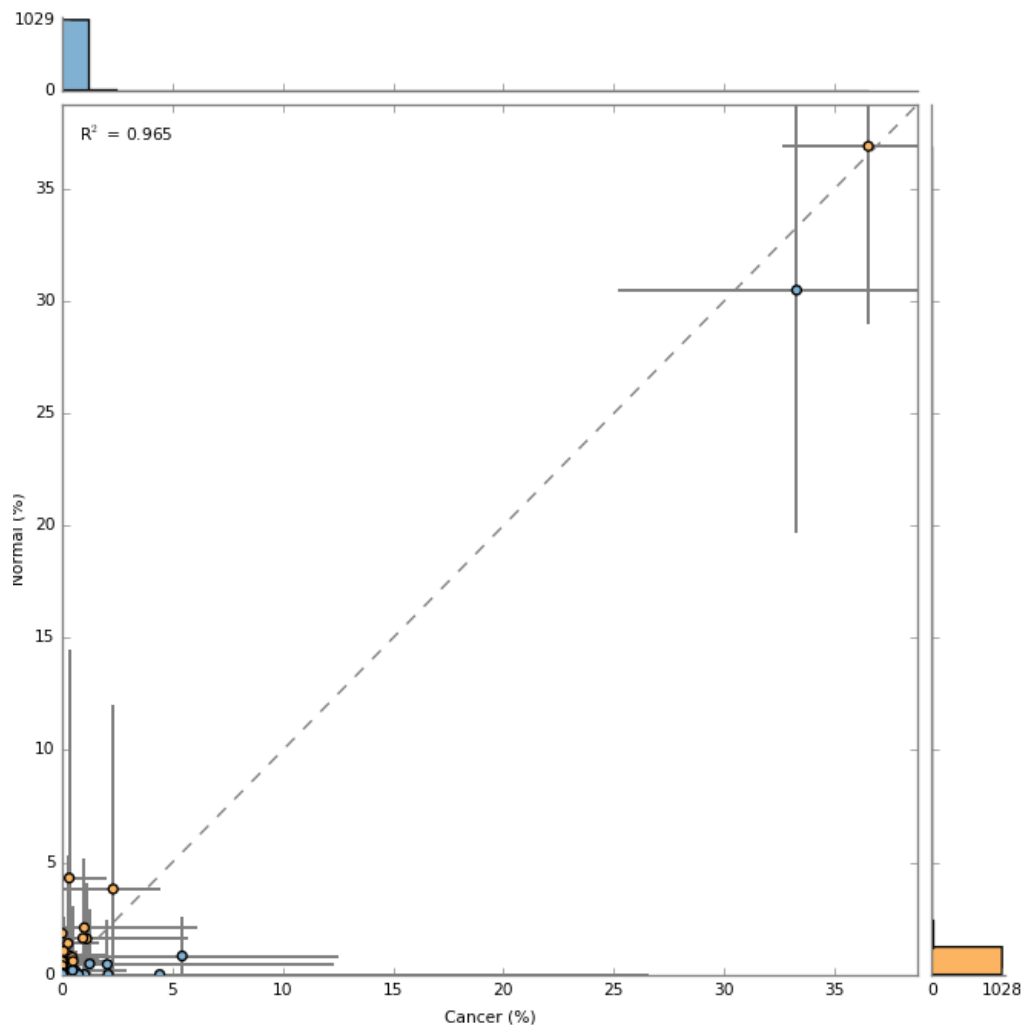
- 左侧多组比较 (Multiple groups) – 校正多组检验 (Multiple test correction) – None修改为Storey FDR(无差异可使用None或提高P值)
- 图片下方类型的主坐标轴分析(PCA plot)改为 箱线图(Box plot), 右侧勾选只显示差异(Show only active features), 选择具体条目展示如下



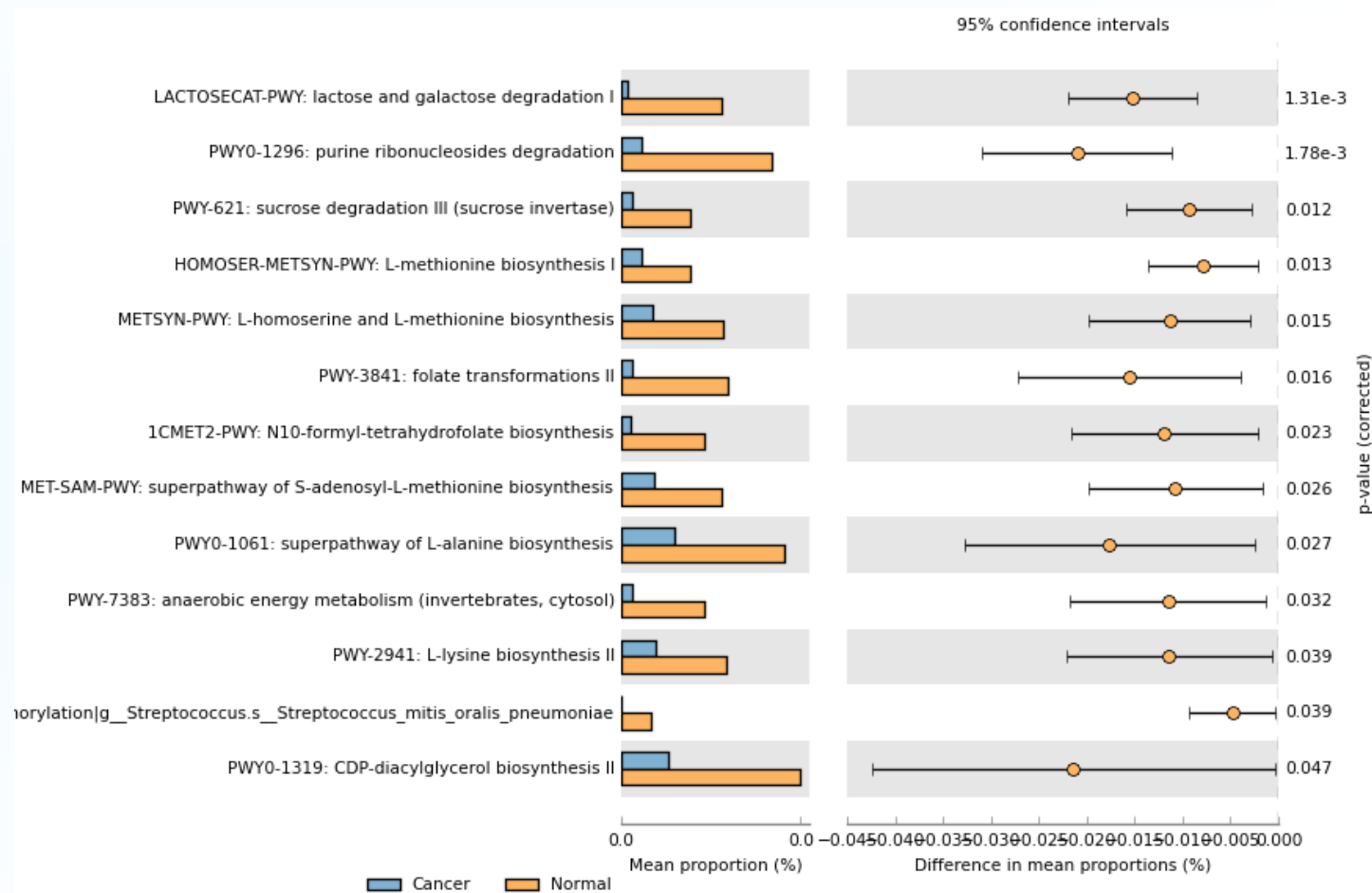
- 1CMET2-PWY: N10-formyl-tetr...
 - HOMOSER-METSYN-PWY: L-me
 - LACTOSECAT-PWY: lactose and
 - MET-SAM-PWY: superpathway
 - METSYN-PWY: L-homoserine ar
 - NAD-BIOSYNTHESIS-II: NAD sal
 - P163-PWY: L-lysine fermentatic
 - PWY-2941: L-lysine biosynthesi**
 - PWY-3841: folate transformati
 - PWY-621: sucrose degradation
 - PWY-7197: pyrimidine deoxyrib
 - PWY-7383: anaerobic energy m
 - PWY0-1061: superpathway of L
 - PWY0-1296: purine ribonucleo: v
- ☒ Show only active features

宏基因组

4. 两组比较——常用结果类型

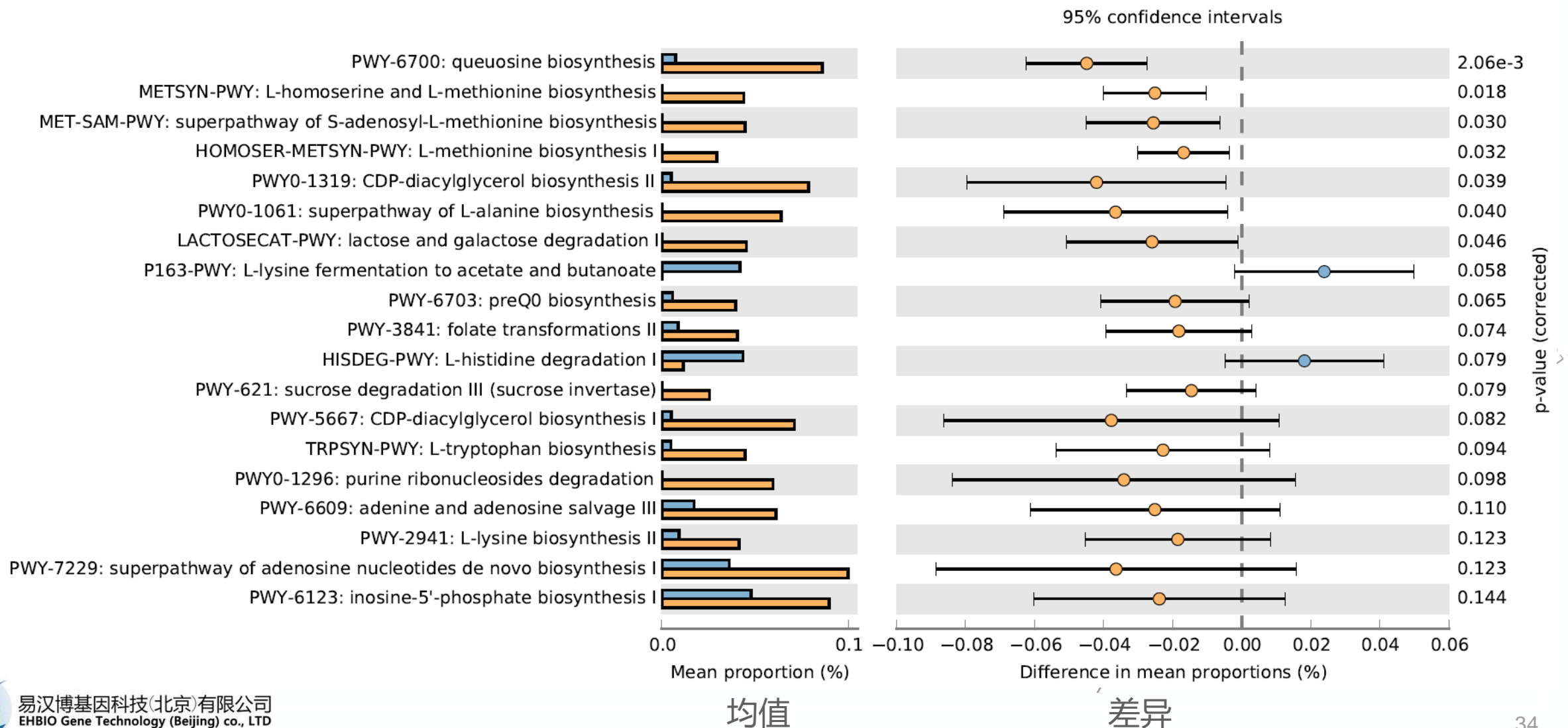


散点图(Scatter plot)——相关分析



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。

- GraPhlAn可绘制最美树状图，参数和输入格式复杂，被众多软件引用如LEfSe，配合export2graphlan使用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；
- HUMAnN2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 + Extended 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。

- [宏基因组公众号文章目录](#) [生信宝典公众号文章目录](#)
- [科学出版社《微生物组数据分析》——50+篇](#)
- [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

