23 MetaPhlAn4物种和HUMAnN4功能组成
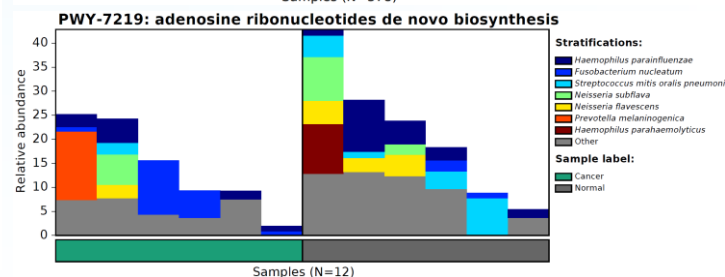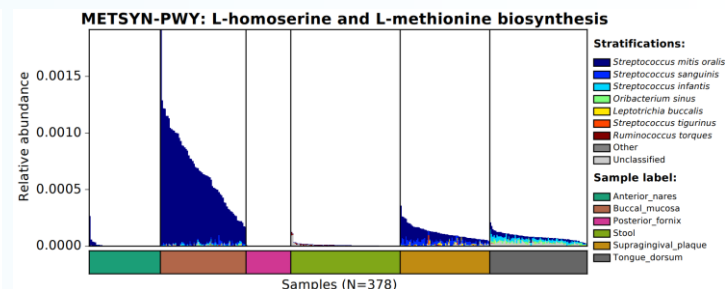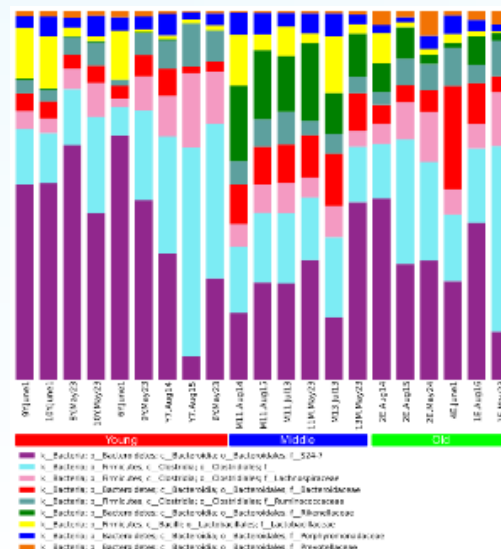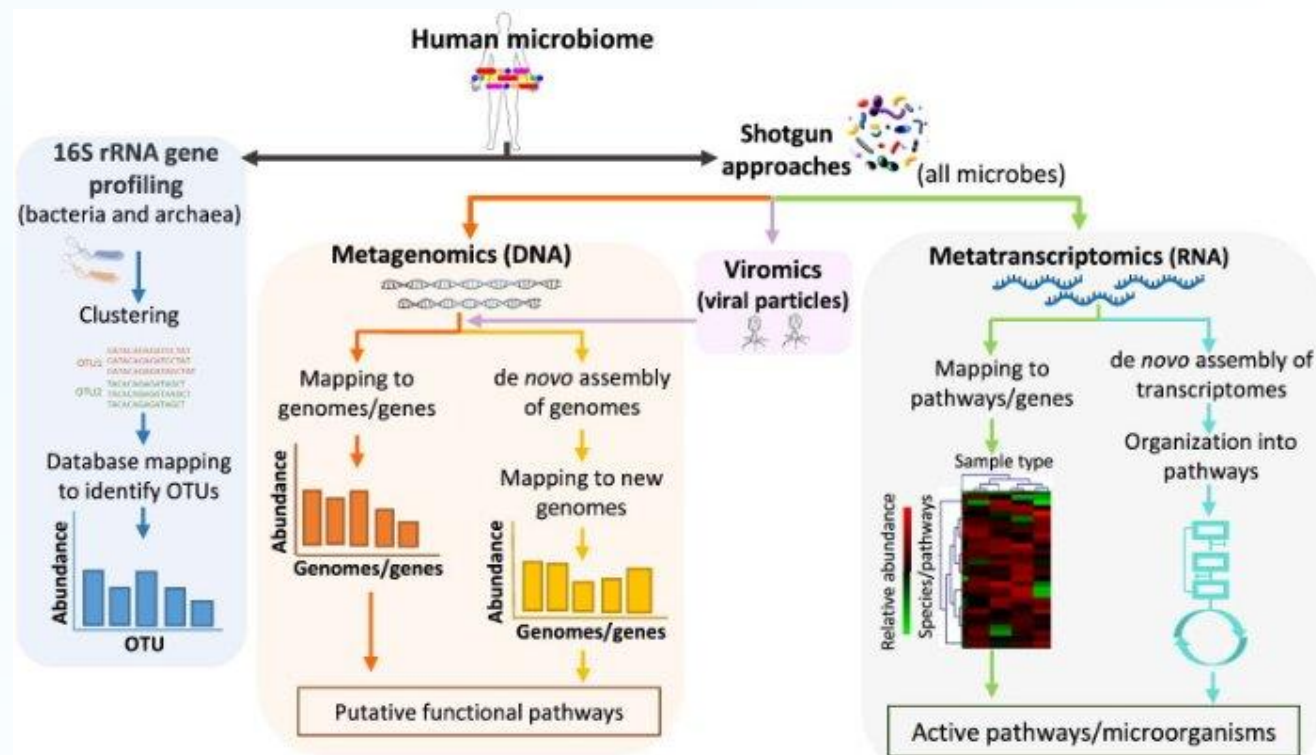
刘永鑫
2025年11月29日

宏基因组基于读长(Read-based)的分析流程

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功能组成分析

MetaPhlAn4 (metagenomic phylogenetic analysis，宏基因组系统发育分析，2012-至今)

○ MetaPhlAn4是分析微生物群落(细菌、古菌、真核生物和病毒)组成的工具，只需一条完命令即可获得微生物的物种丰度信息。同时提供脚本可进一步统计和可视化。

○ 主页：http://segatalab.cibio.unitn.it/tools/MetaPhlAn4/

○ Nicola Segata, Levi Waldron, Annalisa Ballarini, Vagheesh Narasimhan, Olivier Jousson, Curtis Huttenhower. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods* 9: 811. https://doi.org/10.1038/nmeth.2066

○ Duy Tin Truong, Eric A. Franzosa, Timothy L. Tickle, Matthias Scholz, George Weingart, Edoardo Pasolli, Adrian Tett, Curtis Huttenhower, Nicola Segata. 2015. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature Methods* 12: 902-903. https://doi.org/10.1038/nmeth.3589

○ Aitor Blanco-Míguez, et al. 2023. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nature Biotechnology* https://doi.org/10.1038/s41587-023-01688-w
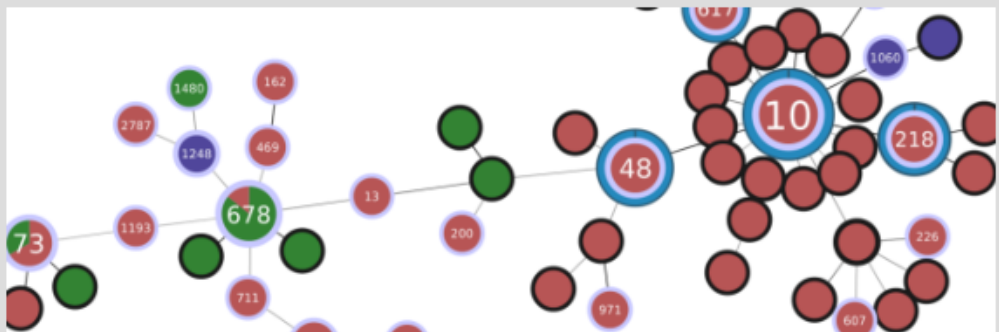
MetaPhlAn4基于多标记基因的宏基因组物种组成定量 文章解读 软件使用

意大利特伦托大学Nicola Segata组——宏基因组软件
(10-12年哈佛公共卫生学院Huttenhower组博后)

Segata lab http://segatalab.cibio.unitn.it/
Computational Metagenomics

Home People Publications Tools Contacts

In the context of our ERC Starting grant 2016 and other projects, we have **openings for four computational scientists** for human microbiome projects. Take a look at the **call for expressions of interest!**

Meta'omics for hacking the human microbiome

MetaMLST strain analysis (Zolfo et al, NAR, 2016)

Cell：宏基因组分箱15万人体微生物基因组

开发维护和参与的软件

- MetaPhlAn4 (2023) 宏基因组物种组成
- curatedMD (2017) 人类微生物组数据库R包
- MetaMLST (2016)宏基因组多位点序列分型
- StrainPhlAn (2016)菌株水平群体基因组分析
- MetAML (2016)微生物与表型关联预测
- PanPhlAn (2016)泛基因组和转录活性
- MetaPhlAn2 (2015) 宏基因组物种组成
- GraPhlAn (2015) 物种或进化树圈图美化
- ShortBRED (2015)蛋白聚类和功能定量
- MicroPITA (2014)宏基因样本挑选
- MetaRef (2014)微生物类特异基因数据库
- PhyloPhlAn (2013)微生物基因建树
- HUMAnN (2012) 宏基因组功能组成定量
- LEfSe (2011)生物标志物挖掘

Nicola Segata

Department CIBIO, University of Trento
Verified email at unitn.it - Homepage

Human Microbiome   Computational Biology   Metagenomics   Microbial Genomics
Machine Learning

Cited by		VIEW ALL
	All	Since 2020
Citations	111348	88861
h-index	103	95
i10-index	240	231

TITLE	CITED BY	YEAR
Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2 E Bolyen, JR Rideout, MR Dillon, NA Bokulich, CC Abnet, GA Al-Ghalith, ... Nature biotechnology 37 (8), 852-857	22630	2019
Metagenomic biomarker discovery and explanation N Segata, J Izard, L Waldron, D Gevers, L Miropolsky, WS Garrett, ... Genome biology 12 (6), R60	15276	2011
Structure, function and diversity of the healthy human microbiome nature 486 (7402), 207-214	11838	2012
A framework for human microbiome research nature 486 (7402), 215-221	2782	2012
MetaPhlAn2 for enhanced metagenomic taxonomic profiling DT Truong, EA Franzosa, TL Tickle, M Scholz, G Weingart, E Pasolli, ... Nature methods 12 (10), 902-903	2317	2015
Shotgun metagenomics, from sampling to analysis C Quince, AW Walker, JT Simpson, NJ Loman, N Segata Nature Biotechnology 35, 833–844	2285	2017
Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis JU Scher, A Sczesnak, RS Longman, N Segata, C Ubeda, C Bielski, ... elife 2, e01202	2250	2013
Metagenomic microbial community profiling using unique clade-specific marker genes	2067	2012

QIIME 2

LEfSe

MetaPhlAn 2

MetaPhlAn 1

Public access VIEW ALL
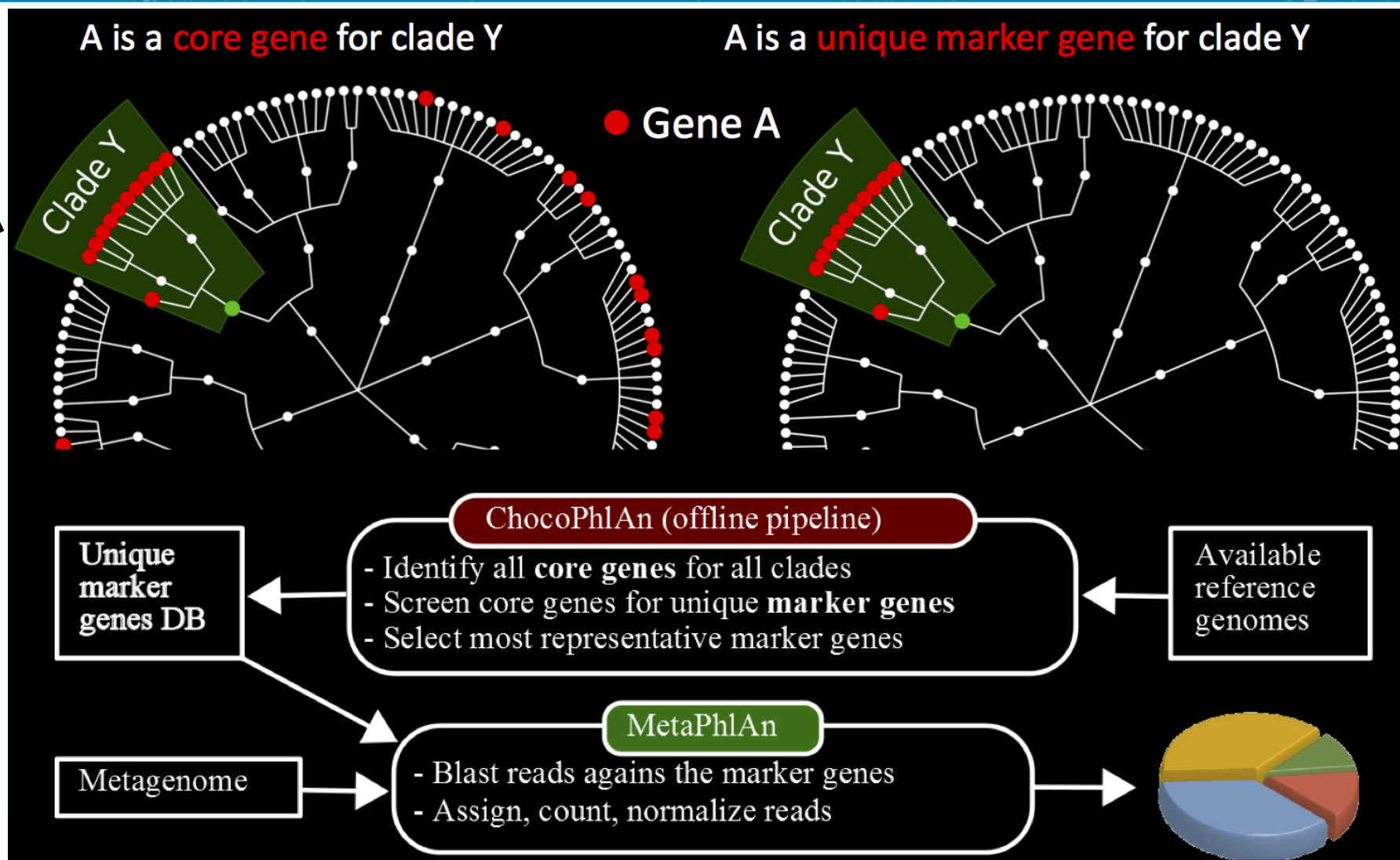
0 articles   259 articles

not available   available

Based on funding mandates

Co-authors VIEW ALL

Curtis Huttenhower   5
Department of Biostatistics, Harv...

类核心
基因

类特异
基因

泛基因组
数据库

最新的MetaPhlAn4

○ HUMAnN4.0+可基于MetaPhlAn4结果继续运行

○ 整合了分离培养基因组+1百万宏基因组组装基因组(MAGs)

○ 26,970种水平MAGs，其中4,992个未定义的种

○ 默认移除低质量结果

○ 安装 conda install -c bioconda metaphlan=4.0.6

○ 使用 https://github.com/biobakery/biobakery/wiki/metaphlan4

○ Aitor Blanco-Míguez, et al. 2023. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nature Biotechnology* https://doi.org/10.1038/s41587-023-01688-w

MetaPhlAn4的使用

conda activate humann4

mkdir -p temp/metaphlan4

time metaphlan --input_type fastq temp/hr/Y1_1.fastq \

  temp/metaphlan4/Y1.txt --nproc 8 --bowtie2db ${db}/metaphlan4 --index mpa_vOct22_CHOCOPhlAnSGB_202403 –offline

输入fastq单个文件，输出物种组成表，可选bowtie2比对中间文件

○ 软件安装代码见1soft_db.sh或2pipeline.sh

○ 除非只关注物种组成，否则MetaPhlAn很少单独使用

○ HUMAnN4整合了MetaPhlAn4软件，可实现一条命令完成物种、功能、以及功能对应物种组成三个文件，多角度挖掘宏基因组数据。

宏基因组基于读长(Reads-based)的分析流程

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. 功能组成统计分析
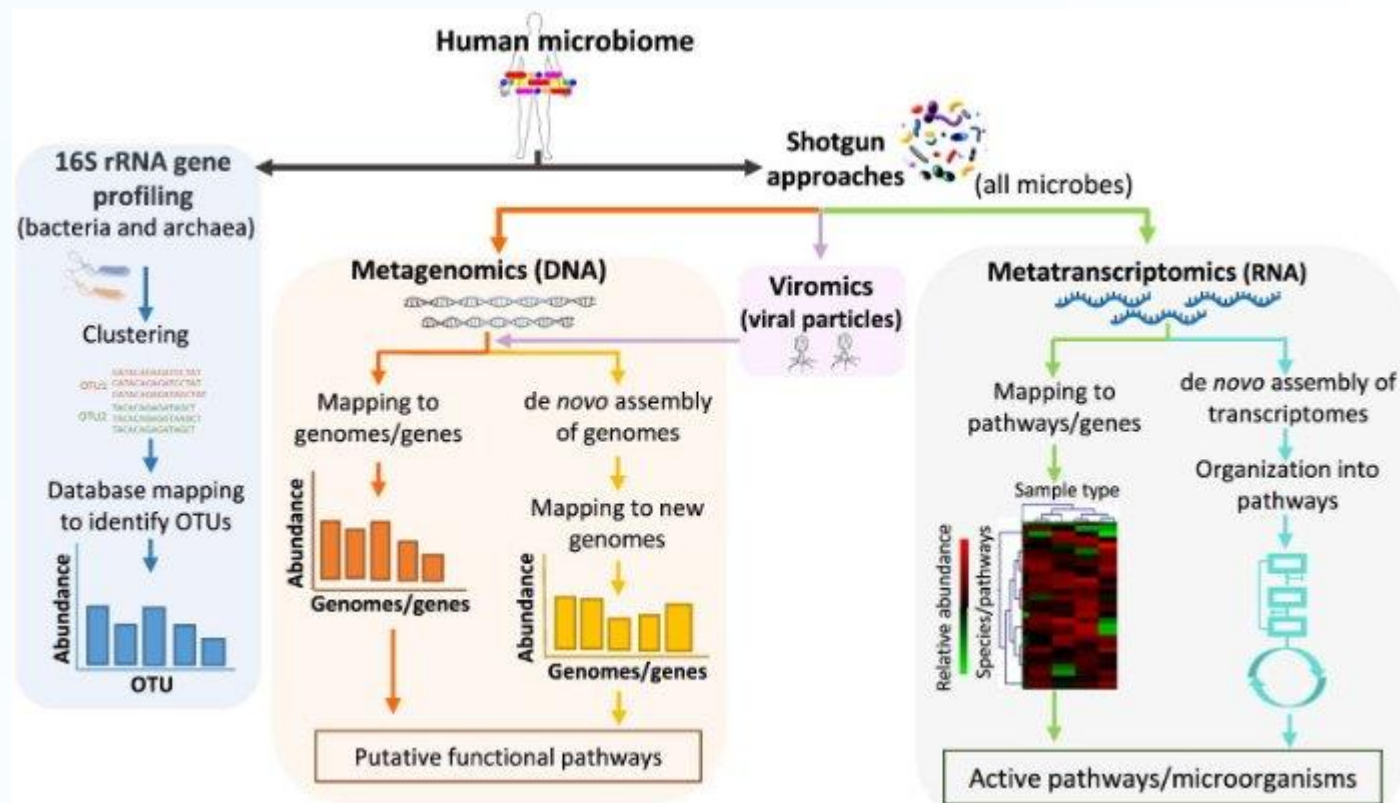
- **HUMAnN4**: The HMP Unified Metabolic Analysis Network 4, HUMAnN是基于**宏基因组、宏转录组数据分析**微生物通路丰度的有效工具。这一过程称为功能谱，目的是描述群体成员的代谢潜能。可以**回答微生物群体成员可能干什么，或在干什么的问题。**

Metabolic reconstruction for metagenomic data and its application to the human microbiome
S Abubucker, N Segata, J Goll, AM Schubert, J Izard, BL Cantarel, B Rodriguez-M
PLoS computational biology, 2012 · journals.plos.org

Microbial communities carry out the majority of the biochemical activ
they play integral roles in processes including metabolism and immu
human microbiome. Shotgun sequencing of such communities' meta
information complementary to organismal abundances from taxonon
resulting data typically comprise short reads from hundreds of differe
at best challenging to assemble comparably to single-organism gene

SHOW MORE ∨

☆ Save � Cite Cited by 1264 Related articles All 32 versions »

Species-level functional profiling of metagenomes and metatranscriptomes
EA Franzosa, LJ McIver, G Rahnavard, LR Thompson, M Schirmer, G Weingart, KS Lipson...
Nature methods, 2018 · nature.com

Abstract
Functional profiles of microbial communities are typically generated using comprehensive
metagenomic or metatranscriptomic sequence read searches, which are
prone to spurious mapping, and often limited to community-level quanti
developed HUMAnN2, a tiered search strategy that enables fast, accur
resolved functional profiling of host-associated and environmental com
HUMAnN2 identifies a community's known species, aligns reads to the

SHOW MORE ∨

☆ Save 🔗 Cite Cited by 1653 Related articles All 11 versions

Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3
F Beghini, LJ McIver, A Blanco-Míguez, L Dubois... - elife, 2021 - elifesciences.org
... To leverage these, we present bioBakery 3, a set of integrated, improved methods for ...,
MetaPhlAn 3 increases the accuracy of taxonomic profiling, and **HUMAnN** 3 improves that of ...
☆ Save 🔗 Cite Cited by 2005 Related articles All 16 versions

Abubucker, S. *et al.* Metabolic Reconstruction for Metagenomic Data and Its Application to the Human Microbiome. *PLOS Computational Biology* **8**, e1002358, https://doi.org/10.1371/journal.pcbi.1002358 (2012).

Franzosa, E. A. *et al.* Species-level functional profiling of metagenomes and metatranscriptomes. *Nature Methods* **15**, 962-968, https://doi.org/10.1038/s41592-018-0176-y (2018).

Beghini, Francesco, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *elife* 10 https://doi.org/10.7554/eLife.65088 (2021)

Nature Methods | HUMAnN2 实现宏基因组和宏转录组种水平功能组成分析

HUMAnN4的特点

o 可对已知和未知生物分析群体功能谱
MetaPhlAn4和ChocoPhlAn泛基因组数据库

o 可获得基因组、基因和通路层面的结果
UniRef基因家族，MetaCyc基因通路，MinPath定义最小通路集

o 简单的使用界面(单行命令实现全部工作流程)
用户只需提供质控后的宏基因组或宏转录组数据

o 加速序列比对
采用Bowtie2加速核酸水平比对

采用Diamond加速核酸翻译蛋白水平比对

HUMAnN4工作原理

•File Type 1 (a quality-controlled metagenome or metatranscriptome)
  • fastq (fastq.gz)
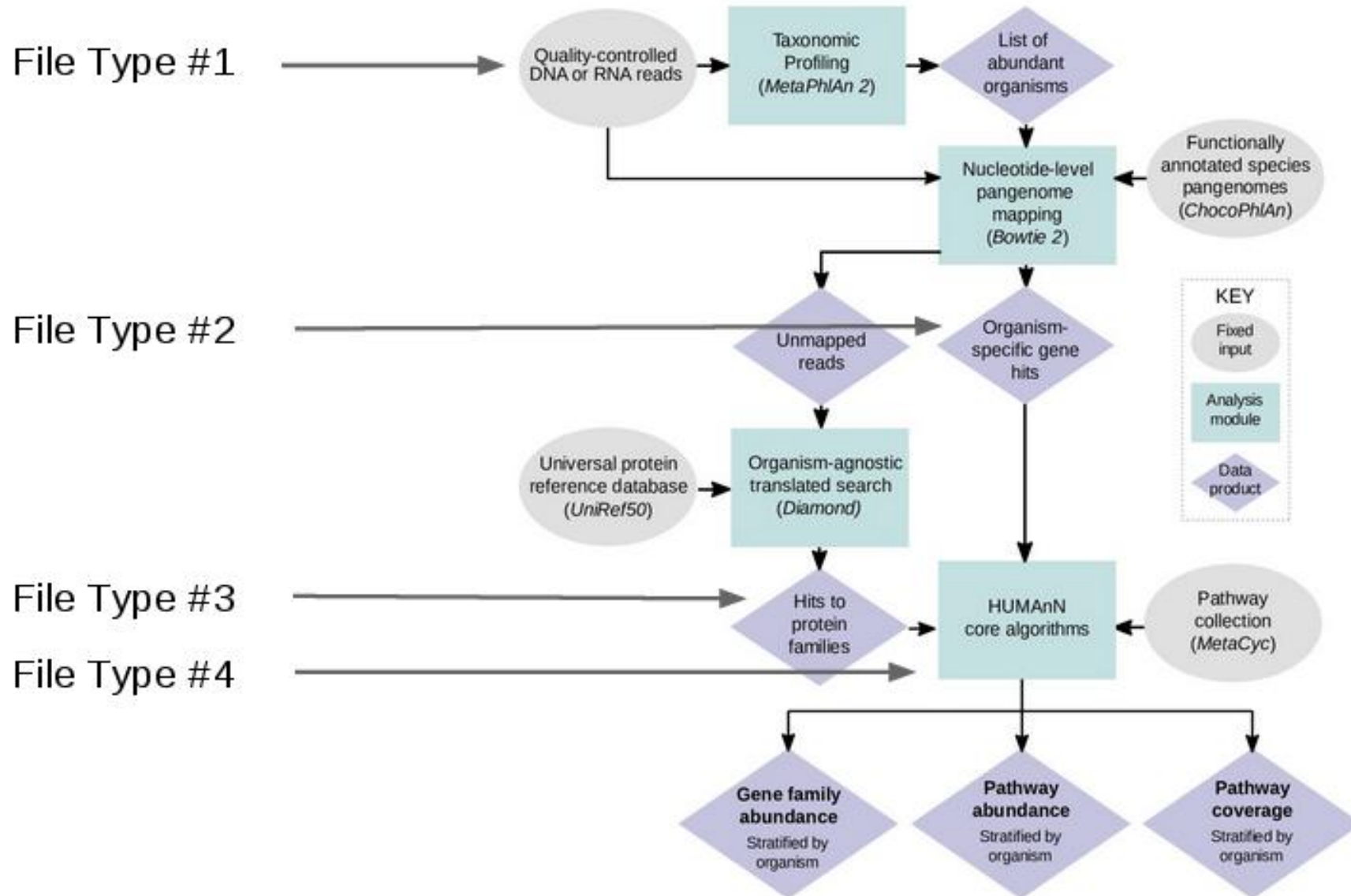 • fasta (fasta.gz)

•File Type 2 (alignment results type 1)
  • sam
  • Bam

•File Type 3 (alignment results type 2)
  • blast-like tsv

•File Type 4 (gene table)
  • tsv
  • biom

软件安装和数据库下载

○ **humann有参宏基因组物种和功能定量流程**

conda install humann

○ 查看可用数据库并设置下载位置

humann_databases # 显示可用数据库

○ 微生物物种核心基因full完整版42G, 可选ec_filtered过滤版6.5G

humann_databases --download chocophlan ec_filtered ~/db/humann4

○ 功能基因diamond索引| 893M

humann_databases --download uniref uniref90_diamond ~/db/humann4

○ 输助比对数据库 2.7G

humann_databases --download utility_mapping full ~/db/humann4

http://huttenhower.sph.harvard.edu/humann_data/uniprot/uniref_annotated/
目前数据为2016版，2019版为humann的数据库，数据量增大一倍

(可选)HUMAnN4物种和功能数据库手动下载

HUMAnN4数据库无法自动下载，备用链接/NMDC/百度云下载安装

mkdir -p ${db}/humann4 && cd ${db}/humann4

mkdir -p chocophlan chocophlan_ec uniref utility_mapping

wget -c
http://huttenhower.sph.harvard.edu/humann_data/chocophlan/chocophlan_EC_FILTERED.v4_alpha.tar.gz

wget -c
http://huttenhower.sph.harvard.edu/humann_data/uniprot/uniref_ec_filtered/uniref90_annotated_v4_alpha_ec_filtered.tar.gz

wget -c http://huttenhower.sph.harvard.edu/humann_data/full_mapping_v4_alpha.tar.gz

安装、解压

time tar xvzf chocophlan_EC_FILTERED.v4_alpha.tar.gz -C ${db}/humann4/chocophlan_ec

time tar xvzf uniref90_annotated_v4_alpha_ec_filtered.tar.gz -C ${db}/humann4/uniref

time tar xvzf full_mapping_v4_alpha.tar.gz -C ${db}/humann4/utility_mapping

配置数据库配置

○ 查看参数和数据库位置是否正确

humann_config --print

○ 常用修改线程数、核酸、蛋白库和多种功能注释数据库位置

humann_config --update run_modes threads 8

humann_config --update database_folders utility_mapping ~/db/humann4/utility_mapping

humann_config --update database_folders nucleotide ~/db/humann4/chocophlan_ec

humann_config --update database_folders protein ~/db/humann4/uniref

○ 有参宏基因组不考虑双端，将双端文件直接合并为一个文件

创建目录存放合并后的序列

mkdir -p temp/concat

for循环调用cat合并每一个样品

for i in `tail -n+2 result/metadata.txt | cut -f 1`;do \

    cat temp/hr/${i}*_?.fastq > temp/concat/${i}.fq; done

查看样品数量和大小

ls -sh temp/concat/*.fq

大文件加速(只用单端和/或截取，牺牲精度换速度)

○ 方法1. 仅链接左端为输入文件(提速50%，节省1倍文件空间)

```
for i in `tail -n+2 result/metadata.txt|cut -f1`;do

  ln -sf `pwd`/temp/hr/${i}_1.fastq temp/concat/${i}.fq

done
```

○ 方法2. 控制标准样比对时间。测序数据量通常为6~50G，同一样本分析时间可达10h~100h，严重浪费时间而浪费硬盘空间。可用head对单端分析截取20M序列，即3G，则为80M行

```
for i in `tail -n+2 result/metadata.txt|cut -f1`;do

  head -n80000000 temp/hr/${i}_1.fastq  > temp/concat/${i}.fq

done
```

2.2 HUMAnN4计算物种和功能组成

○ 输入文件和输出目录，经rush管理批量任务队列

```
 mkdir -p temp/humann4

time tail -n+3 result/metadata.txt | cut -f1 | rush -j 2 \

     "humann --input temp/concat/{1}.fq --threads 8 \

      --metaphlan-options '--input_type fastq --bowtie2db ${db}/metaphlan4 --index
mpa_vOct22_CHOCOPhlAnSGB_202403 --offline --nproc 8' \

      --output temp/humann4/
```

核心步骤，测序数据2任务 X 8核 = 16线程，用时1h，6G要几小时

mkdir -p result/metaphlan4

○ # 样品结果合并

merge_metaphlan_tables.py
temp/humann/*_humann_temp/*_metaphlan_bugs_list.tsv | \
 sed 's/_metaphlan_bugs_list//g' > result/metaphlan4/taxonomy.tsv

○ # 转换为spf格式方便stamp分析

metaphlan_to_stamp.pl result/metaphlan4/taxonomy.tsv \
    > result/metaphlan4/taxonomy.spf

○ metaphlan_to_stamp.pl脚本来自microbiome_helper项目，整合至 EasyMicrobiome

Alpha多样性

○ 计算多样性

```
Sample_ID observed_species  shannon simpson invsimpson  Pielou_evenness
c1  59  2.64265954506613  0.882515813523708 8.51178384081335  0.64810184613625
c2  63  2.96603418555888  0.91148414305896  11.2974108205956  0.715891319359086
```

Rscript $sd/metaphlan4_alpha.R \

-i metaphlan4/taxonomy.tsv -g metadata.txt \

-t 7 -o metaphlan4/alpha

○ 绘图

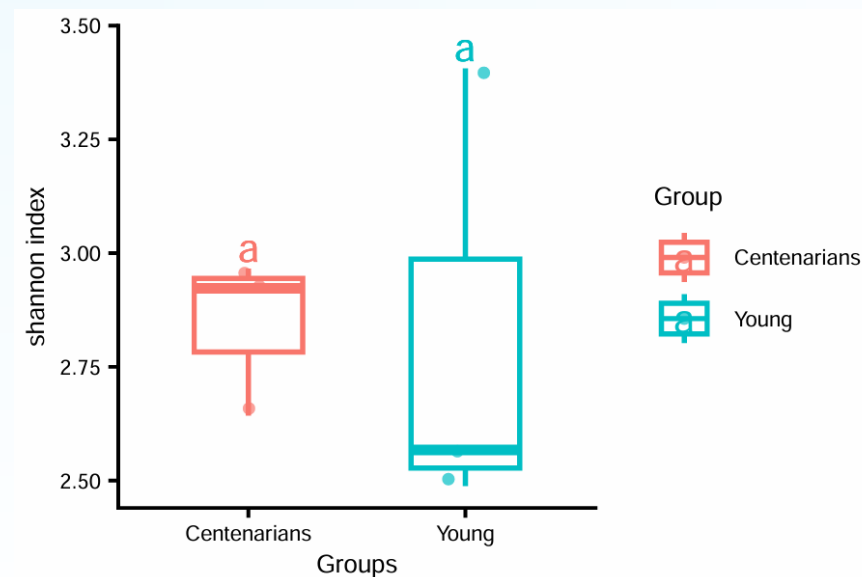Rscript $sd/alpha_boxplot.R \

-i metaphlan4/alpha.txt -a shannon \

-d metadata.txt -n Group \

-o metaphlan4/ -w 89 -e 59
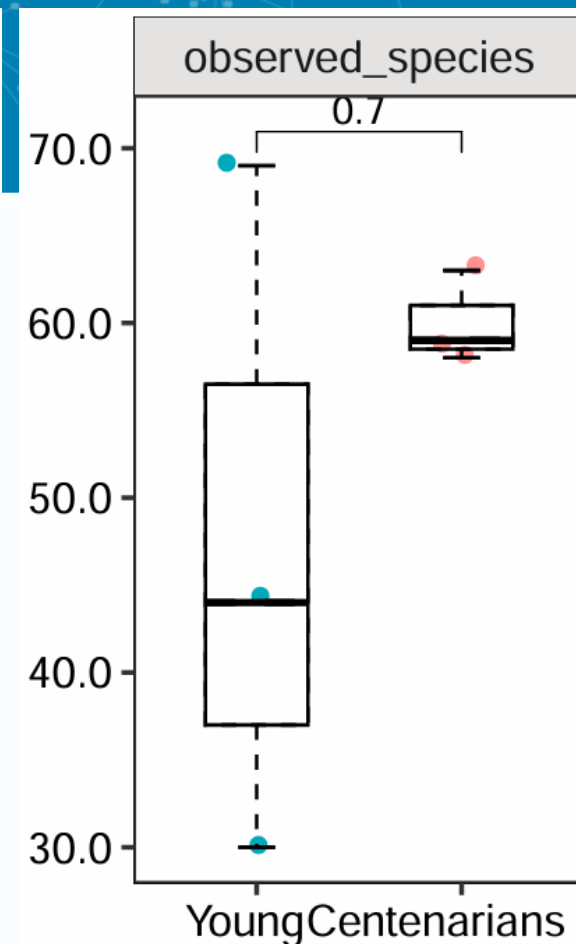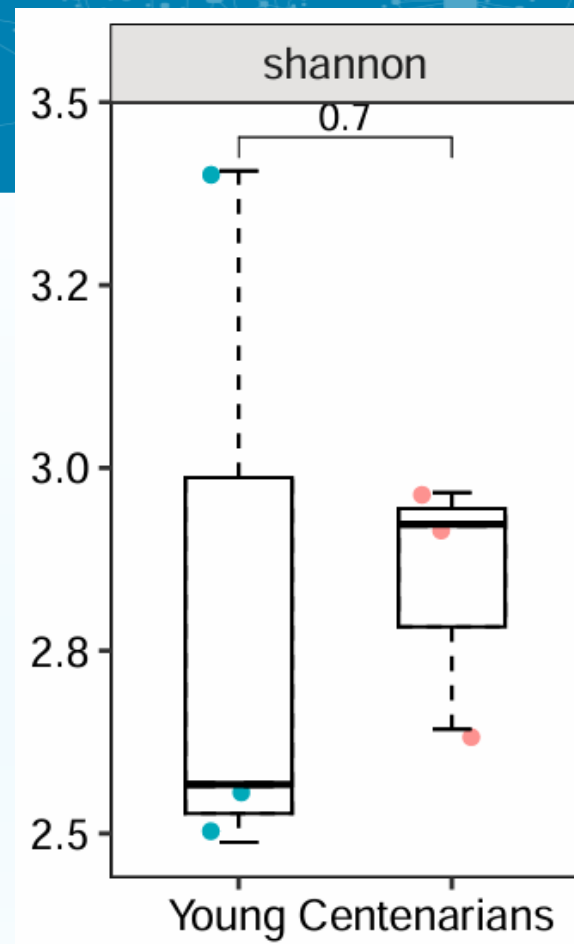
R脚本推荐在windows下运行，详见 2StatPlot.sh

箱线图+P值

Rscript $sd/alpha_boxplot_new.R \
  -i metaphlan4/alpha.txt \
  -a shannon \
  -d metadata.txt \
  -n Group \
  -o metaphlan4/ \
  -w 49 -e 79

```
# 6 Alpha diversity boxplot
for i in `head -n1 metaphlan4/alpha.txt|cut -f 2-`;do
Rscript $sd/alpha_boxplot_new.R -i metaphlan4/alpha.txt -a ${i} \
  -d metadata.txt -n Group -w 49 -e 79 \
  -o metaphlan4/
done
```
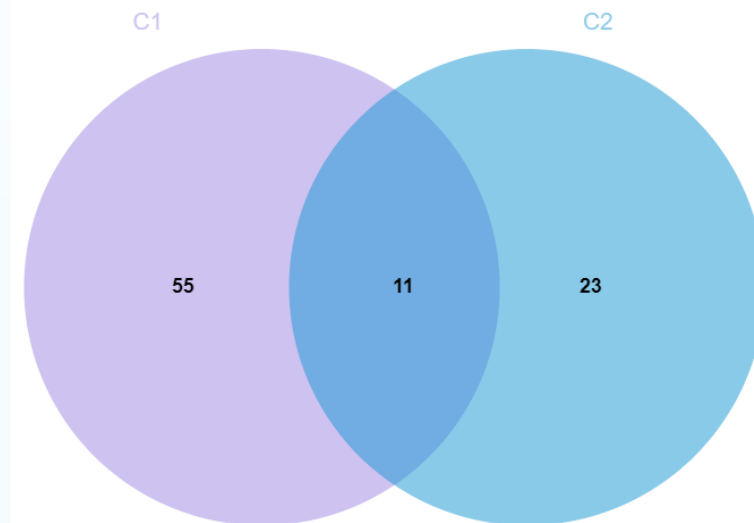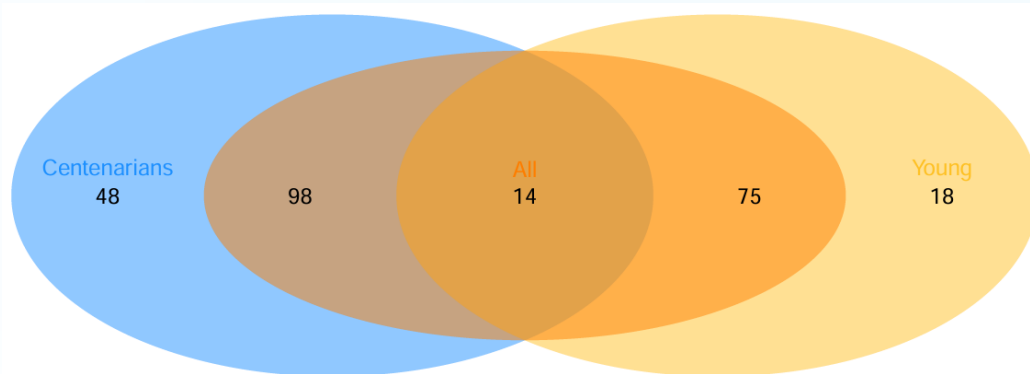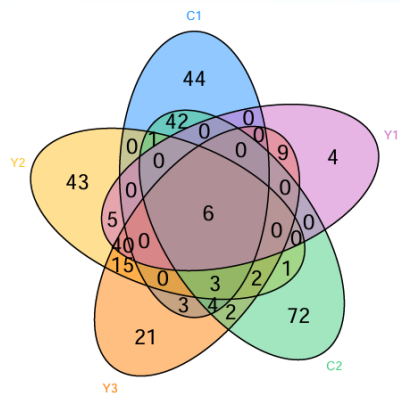
○ # (可选)筛选>0.5%的分类，绘制维恩图

```
awk 'BEGIN{OFS=FS="\t"}{if(FNR==1)
{for(i=2;i<=NF;i++) a[i]=$i;} \
    else {for(i=2;i<=NF;i++) if($i>0.5) print $1, a[i];}}' \
result/metaphlan4/taxonomy.tsv \
> result/metaphlan4/taxonomy_high.tsv
wc -l result/metaphlan4/taxonomy_high.tsv
```


Click on a venn diagram figure to display the linked elements:

Common elements in C1 C2 :
k__Bacteria
k__Bacteria|p__Bacteroidetes
k__Bacteria|p__Bacteroidetes|c__Bacteroidia
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Prevotellaceae
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Prevotellaceae|g__Prevotella
k__Bacteria|p__Firmicutes
k__Bacteria|p__Proteobacteria
k__Bacteria|p__Proteobacteria|c__Betaproteobacteria
k__Bacteria|p__Proteobacteria|c__Betaproteobacteria|o__Neisseriales

22

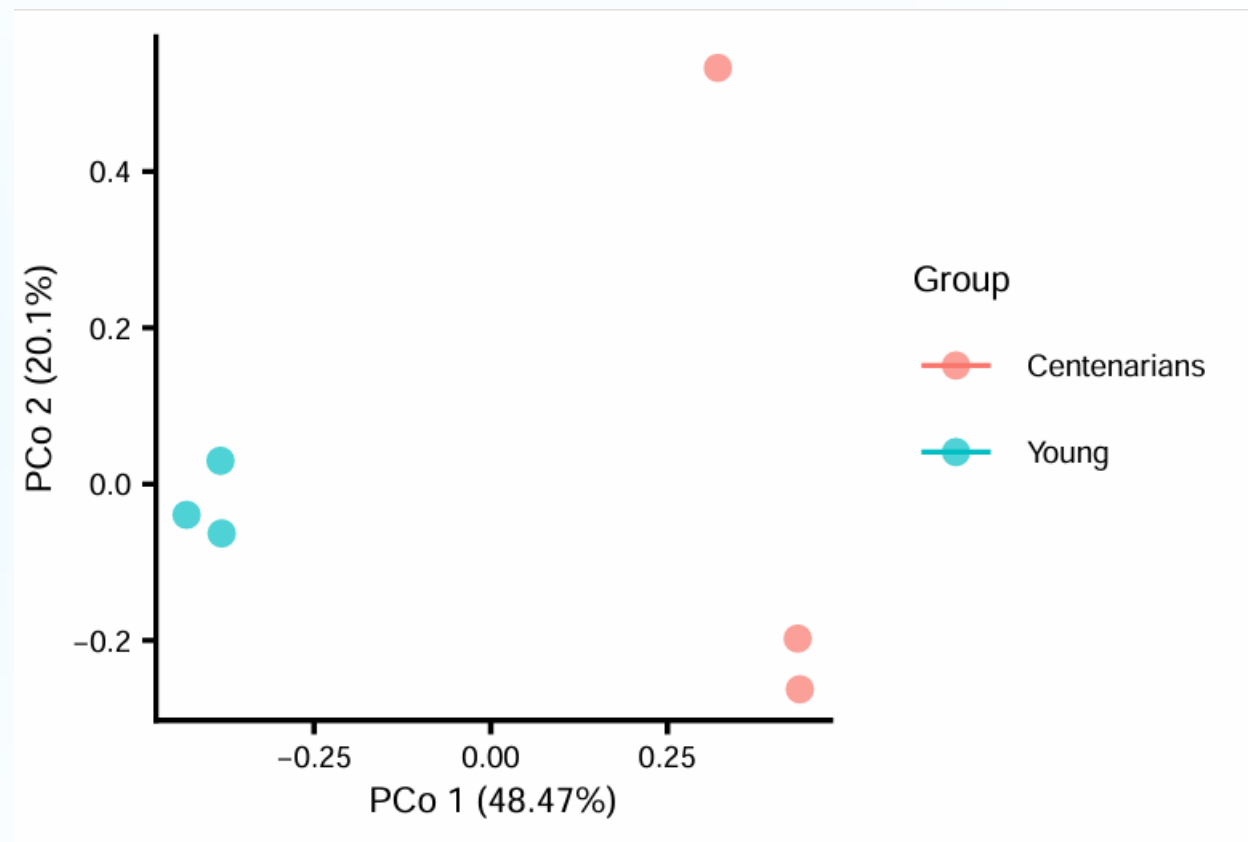# Beta多样性

```
# 距离矩阵诗雅
Rscript $sd/metaphlan4_beta.R \
  -i metaphlan4/taxonomy.tsv \
  -g metadata.txt \
  -t 7 \
  -m bray \
  -o metaphlan4/beta
# PCoA
Rscript $sd/beta_pcoa.R \
  --input metaphlan4/beta_bray.txt \
  --design metadata.txt \
  --group Group \
  --width 89 --height 59 \
  --output metaphlan4/pcoa.bray_curtis.pdf
```
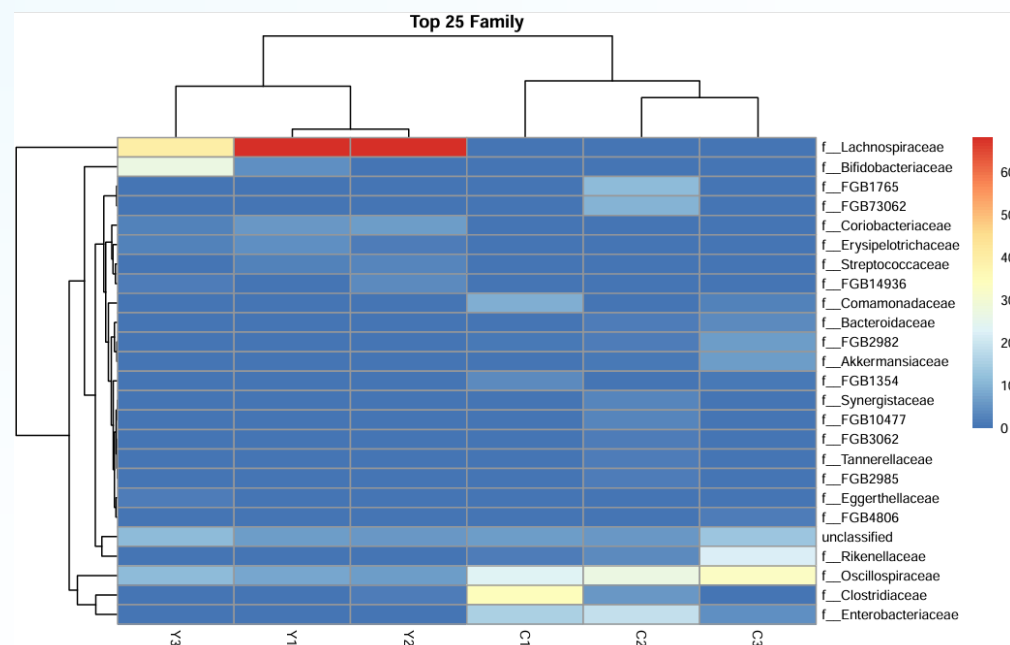
评估样本物种组成和分组聚类——更灵活绘制热图

- 方法1. metaphlan_hclust_heatmap.py 脚本服务器绘制热图，依赖关系和环境变量复杂
- 方法2. Excel筛选 metaphlan4/taxonomy.tsv 并在线绘制热图
- 方法3. R数据筛选taxonomy.spf并用pheatmap绘制热图，可指定分类级别、物种数量

```
Rscript $sd/metaphlan_hclust_heatmap.R \
    -i result/metaphlan4/taxonomy.spf \
    -t Order \
    -n 25 \
    -o result/metaphlan4/heatmap_Order
```


Top 25 Family

f__Lachnospiraceae
f__Bifidobacteriaceae
f__FGB1765
f__FGB73062
f__Coriobacteriaceae
f__Erysipelotrichaceae
f__Streptococcaceae
f__FGB14936
f__Comamonadaceae
f__Bacteroidaceae
f__FGB2982
f__Akkermansiaceae
f__FGB1354
f__Synergistaceae
f__FGB10477
f__FGB3062
f__Tannerellaceae
f__FGB2985
f__Eggerthellaceae
f__FGB4806
unclassified
f__Rikenellaceae
f__Oscillospiraceae
f__Clostridiaceae
f__Enterobacteriaceae

Y3 Y1 Y2 C1 C2 C3

Top 20 Order

Top 30 Genus

箱线图 Boxplot

○ 各级别级别、物种数量

 for tax in Phylum Family Genus Species; do
Rscript $sd/metaphlan_boxplot.R \
  -i metaphlan4/taxonomy.spf \
  -t ${tax} \
  -n 30 \
  -o metaphlan4/boxplot_${tax}
done
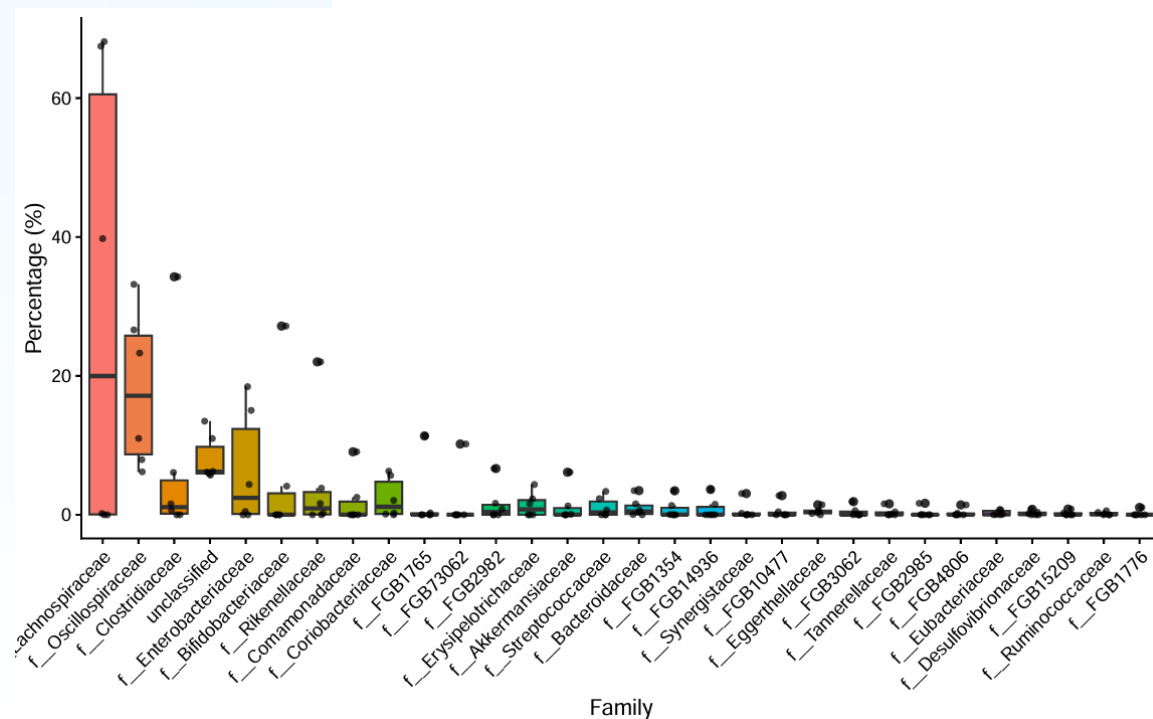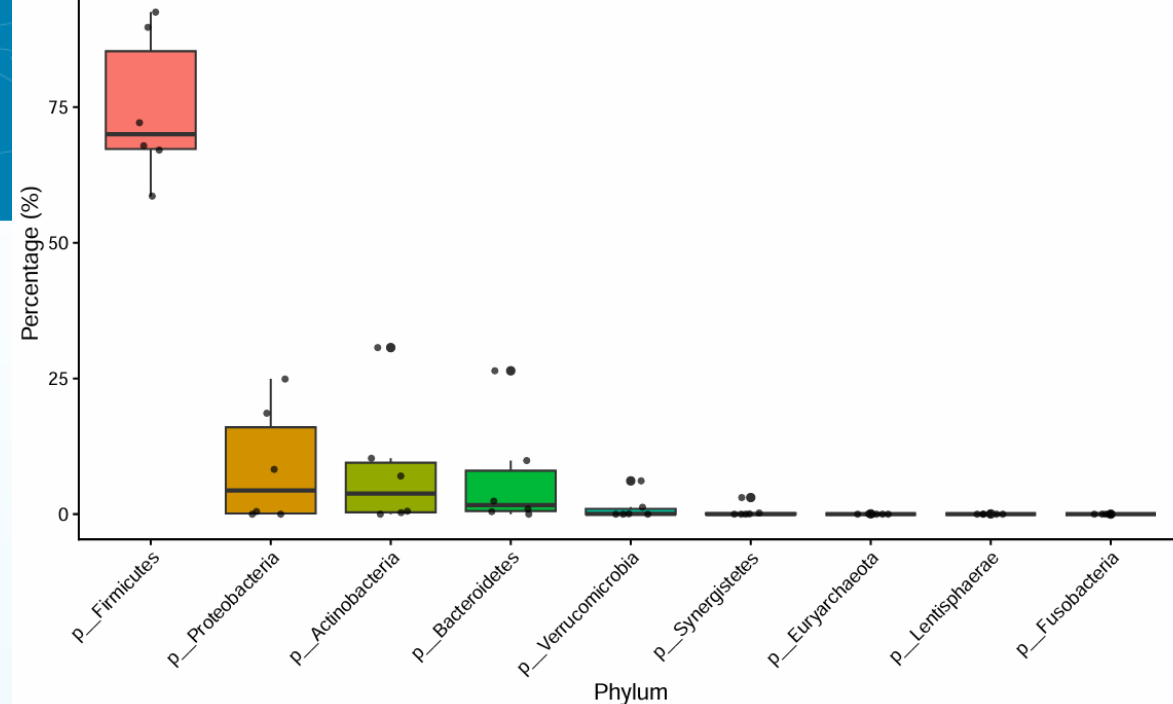
比较箱线图 Compare boxplot

○ # 组间比较箱线图 *_compare.pdf

for tax in Phylum Family Genus Species; do

Rscript $sd/metaphlan4_boxplot_compare.R \

    -i metaphlan4/taxonomy.spf \

    -t ${tax} \

    -n 30 \

    -o metaphlan4/boxplot_${tax};done
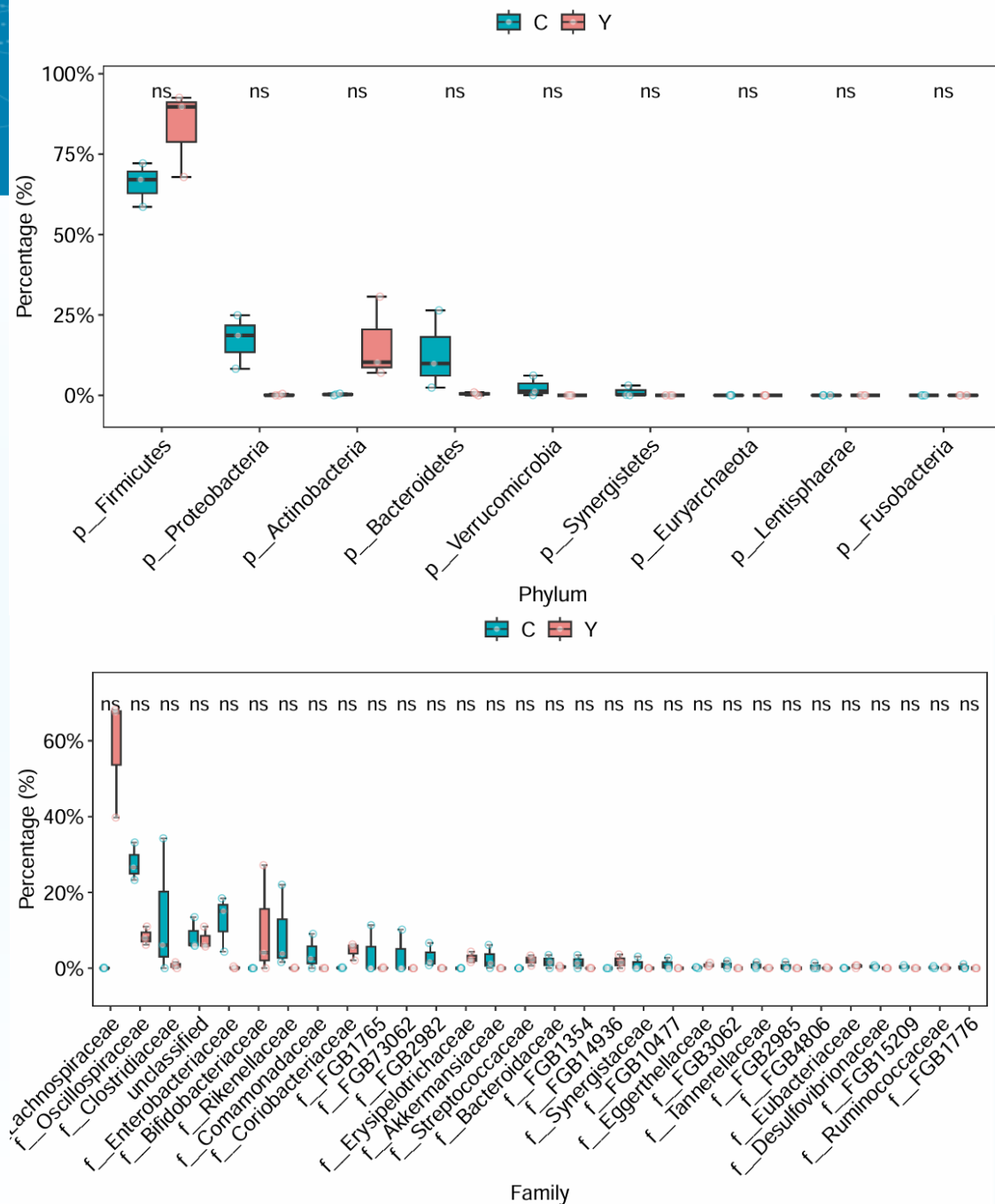
绘制不同分类层级堆叠柱状图

for tax in Phylum Genus Species; do

Rscript ${sd}/tax_stackplot.R \

    --input metaphlan4/${tax}.txt --design metadata.txt \

    --group Group --output metaphlan4/${tax}.stackplot \

    --legend 10 --width 120 --height 70; done
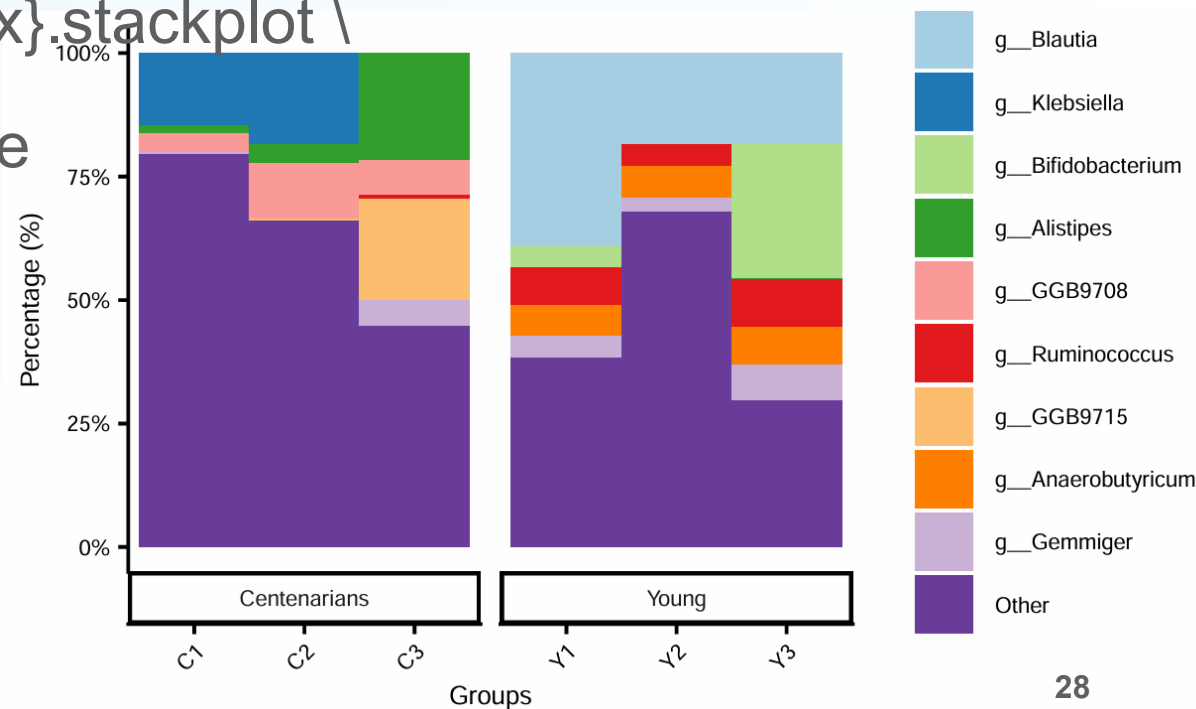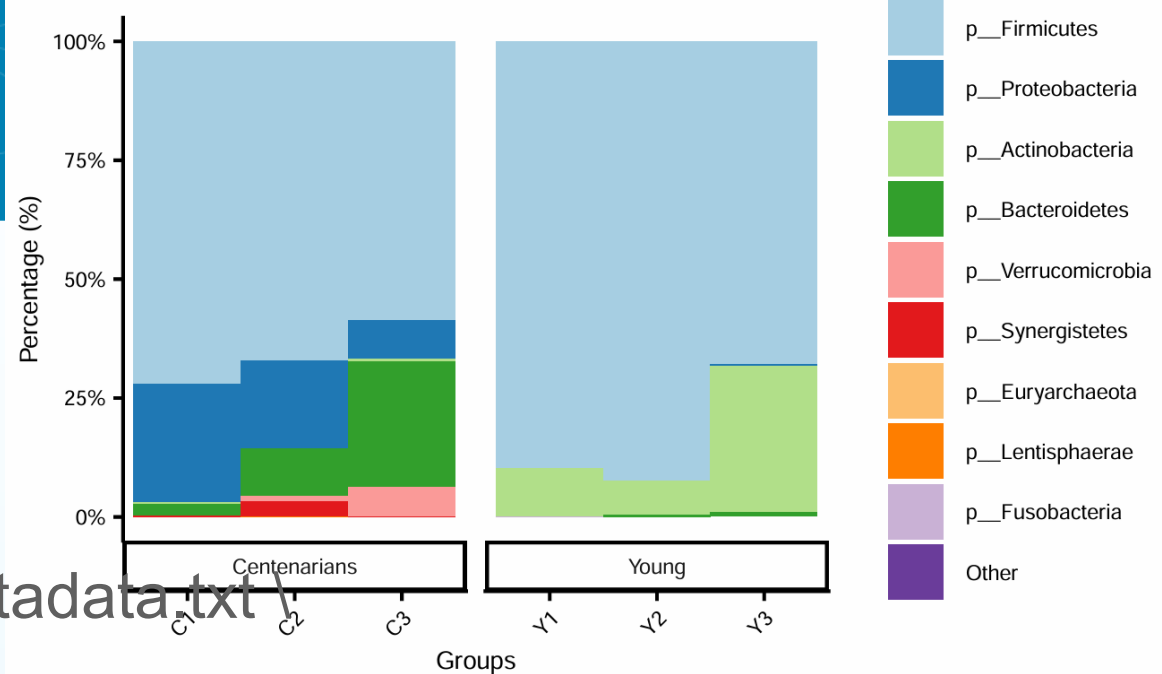
○ # 合并所有样品

humann_join_tables --input temp/humann --file_name pathabundance \

  --output result/HUMAnN4/pathabundance.tsv

sed -i 's/_Abundance//g' result/HUMAnN4/pathabundance.tsv

○ # 标准化为相对丰度relab或百万分数cpm

humann_renorm_table --input result/HUMAnN4/pathabundance.tsv --units relab \

  --output result/HUMAnN4/pathabundance_relab.tsv

○ # 分层结果

humann_split_stratified_table --input result/HUMAnN4/pathabundance_relab.tsv \

  --output result/HUMAnN4/

2.4.2 添加分组和差异比较

Shell命令在表头下面添加分组行

# Pathway HUMAnN v4.0.0.alpha.1	C1	C2	C3	Y1	Y2	Y3			
Group	Centenarians	Centenarians	Centenarians	Young	Young	Young			
Group	Cancer	Cancer	Cancer	Cancer	Cancer	Normal	Normal	Normal	Normal	Normal	Normal
PWY-2221: Entner-Doudoroff pathway III (semi-phosphorylative)|unclassified	142.4960866379	150.5974406277	13.9875571992	0	95.0701339734	60.9457969738			
PWY-241: C4 photosynthetic carbon assimilation cycle, NADP-ME type	351.3595944155	246.1356615705	178.4807734690	93.5199929391	160.9925301329	280.4848524796			

KW差异比较：输入input、分组focal、分组类型type、分组行结果、FDR和输出output；结果包括特征对应各组均值和统计值

humann_associate --input result/humann4/path.pcl \

    --focal-metadatum Group --focal-type categorical \

    --last-metadatum Group --fdr 0.2 \

    --output result/humann4/associate.txt

官方教程 https://bitbucket.org/biobakery/biobakery/wiki/humann#rst-header-downstream-analyses
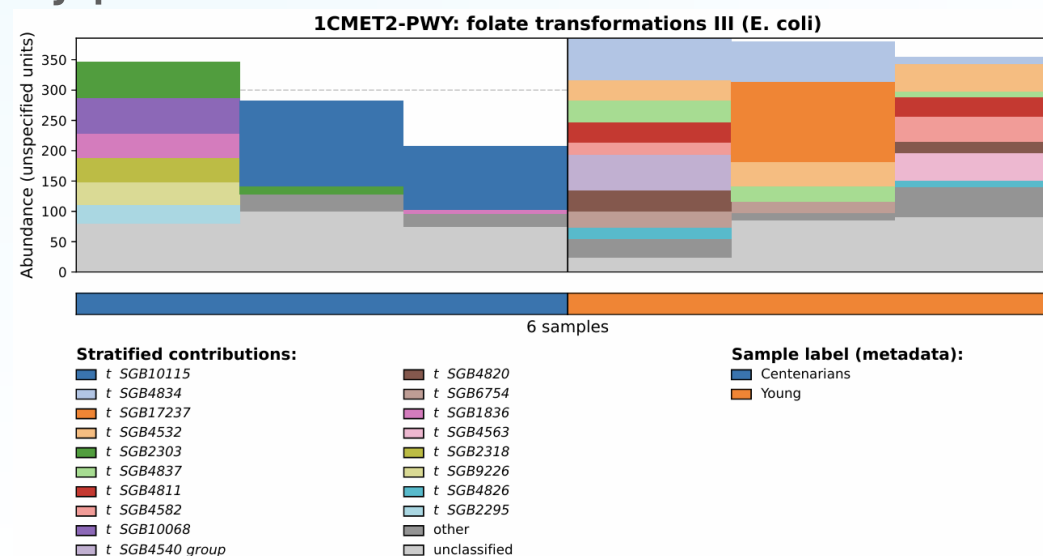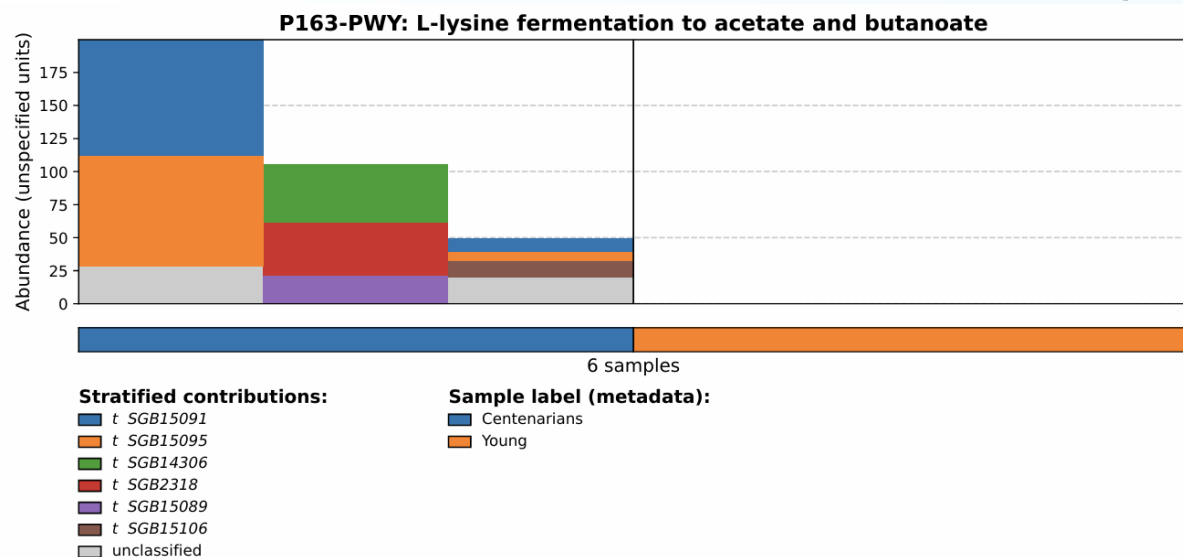
柱状图展示通路的物种组成

如：L-赖氨酸发酵生成乙酸和丁酸

path=P163-PWY

grep $path result/humann4/associate.txt humann_barplot \

    --input ${pcl} --focal-feature ${path} \

    --focal-metadata Group --last-metadata Group \

    --output result/humann4/barplot_${path}.pdf --sort sum metadata

使用HMP示例数据，样本名下面有分组信息，需要手动制作

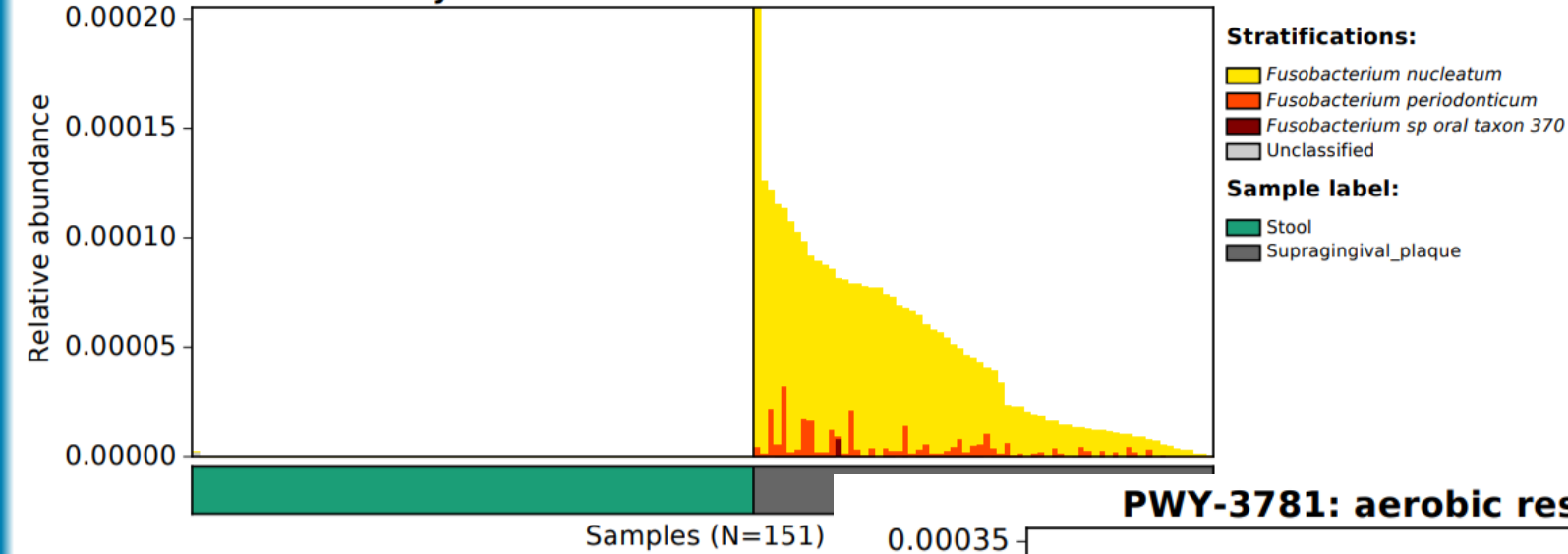| FEATURE \ SAMPLE	SRS011084	SRS011086	SRS011090
Group	Stool	Stool	Stool
1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis	0.000498359	0.000628096	0.000304951
1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis|g__Acidovorax.s__Acidovorax_ebreus	0	0	0

humann_associate --input hmp_pathabund.pcl --focal-metadatum Group --focal-type categorical --last-metadatum Group --fdr 0.05 --output associate.txt

# Feature	Level means (|ed)	P-value	Q-value
P163-PWY: L-lysine fermentation to acetate and butanoate	Stool:3.133e-08|Supragingival_plaque:5.385e-05	1.253e-30	4.261e-28
PWY-3781: aerobic respiration I (cytochrome c)	Stool:1.464e-07|Supragingival_plaque:0.0002634	4.076e-30	4.62e-28
PWY66-409: superpathway of purine nucleotide salvage	Stool:0|Supragingival_plaque:0.000106	2.986e-30	4.62e-28
PWY1F-823: leucopelargonidin and leucocyanidin biosynthesis	Stool:4.877e-09|Supragingival_plaque:4.695e-05	1.402e-29	1.192e-27

humann_barplot --input hmp_pathabund.pcl --focal-feature PWY-3781 --focal-metadata Group --last-metadata Group --output barplot_PWY-3781.pdf --sort sum metadata

官方教程 https://bitbucket.org/biobakery/biobakery/wiki/humann#rst-header-downstream-analyses

--sort sum metadata先按分组，再按丰度

P163-PWY: L-lysine fermentation to acetate and butanoate

Stratifications:
- *Fusobacterium nucleatum*
- *Fusobacterium periodonticum*
- *Fusobacterium sp oral taxon 370*
- Unclassified

Sample label:
- Stool
- Supragingival_plaque

PWY-3781: aerobic respiration I (cytochrome c)

Stratifications:
- *Neisseria sicca*
- *Neisseria macacae*
- *Neisseria flavescens*
- *Neisseria sp oral taxon 014*
- *Neisseria bacilliformis*
- *Neisseria subflava*
- *Campylobacter curvus*
- Other
- Unclassified

Sample label:
- Stool
- Supragingival_plaque
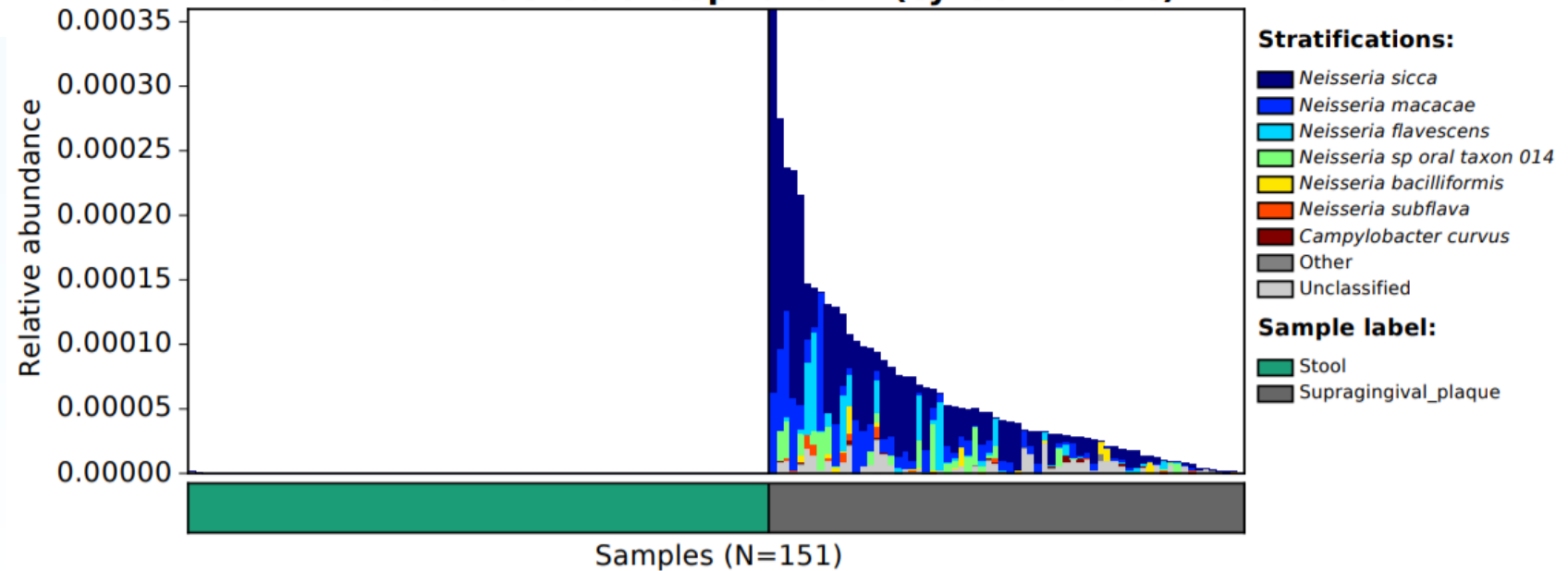
```
# 转换基因家族为KO(uniref90_ko)，可选eggNOG(uniref90_eggnog)或酶
(uniref90_level4ec)
for i in `tail -n+2 result/metadata.txt|cut -f1`;do
  humann_regroup_table \
    -i temp/HUMAnN4/${i}_genefamilies.tsv \
    -g uniref90_ko \
    -o temp/HUMAnN4/${i}_ko.tsv
done
# 合并，并修正样本名
humann_join_tables \
  --input temp/HUMAnN4/ \
  --file_name ko \
  --output result/HUMAnN4/ko.tsv
sed -i '1s/_Abundance-RPKs//g' result/HUMAnN4/ko.tsv
```

# Gene Family	KO1	WT1
K00029	0	23.999
K00029|g__Pseudomonas.s__Pseudomon	0	23.999
K00031	8.81	7.018
K00031|g__Agrobacterium.s__Agrobacter	0	1.802
K00031|unclassified	8.81	5.215
K00032	0	8.48
K00032|g__Pseudomonas.s__Pseudomon	0	8.48
K00033	0	12.989
K00033|g__Bacillus.s__Bacillus_megateriu	0	3.454
K00033|g__Pseudomonas.s__Pseudomon	0	9.535
K00035	0	1.285
K00035|g__Agrobacterium.s__Agrobacter	0	1.285
K00036	0	46.161
K00036|g__Agrobacterium.s__Agrobacter	0	0.754
K00036|g__Bacillus.s__Bacillus_megateriu	0	5.084
K00036|g__Pseudomonas.s__Pseudomon	0	40.324

总结

○ HUMAnN4调用MetaPhlAn4比对上万个物种的数据库，可快速、准确获得细菌、真菌、古菌、病毒、真核生物等的物种组成

○ merge_metaphlan_tables.py合并，metaphlan_to_stamp.pl生成STAMP输入文件

○ HUMAnN4比对UniRef数据库获得功能组成；注意数据库位置设置；了解其依赖关系以便解决依赖关系错误的问题

○ 结果包括功能通路丰度组成和**功能通路具体来源的物种**，提供join, norm, stratified等脚本实现合并、标准化和分层

○ associate, barplot脚本实现组间KW检验和通路组成可视化

○ humann_regroup_table脚本转换基因家族为KO/eggNOG/EC等注释

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