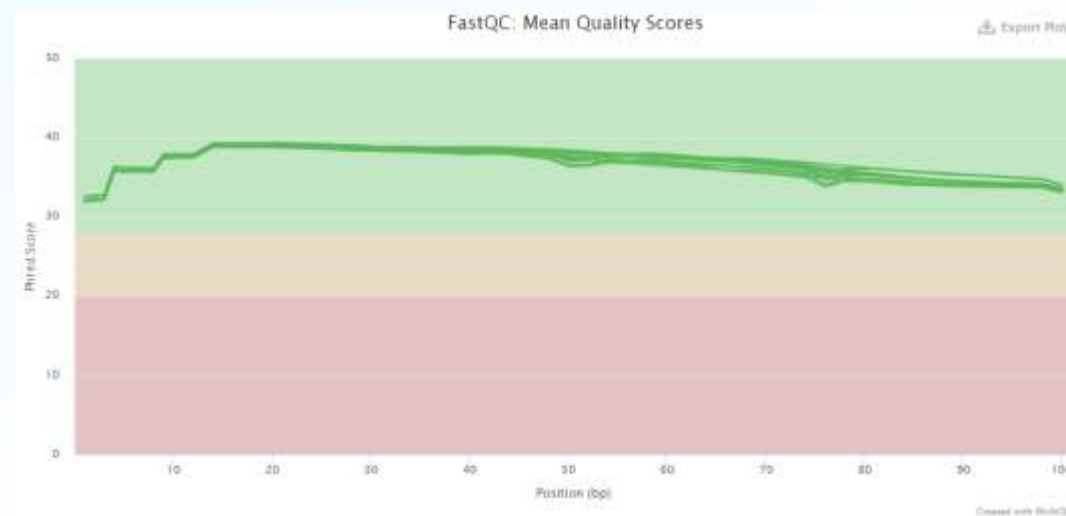


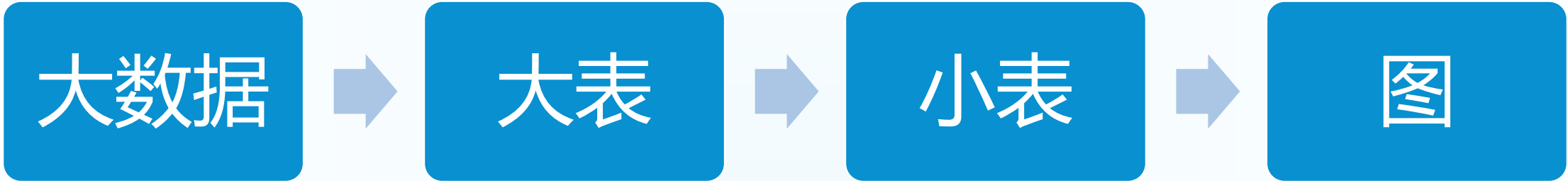


22易宏基因组  
质控和去宿主  
1理论+2安装+3实战

刘永鑫  
2025年11月29日



# 数据分析的基本思想——三步走



```
HHISEQ:549:HLNYBCXY:1:1101:1267:2220 1:N:0:CACTCAAT
TCGTCGCTCGAACAGGATTAGATACCTGGTAGTCCACGCTGTAACGTTGGCG
+
DDDDIHHHHIIIIIIHHIIIIIIIIHHIIIIIIIIHHIIIIIIIIHHIIIIIIII
HHISEQ:549:HLNYBCXY:1:1101:1887:2204 1:N:0:CACTCAAT
TACGAGTATGAACAGGATTAGATACCTGGTAGTCCACGCCCTAAACGATGTCTA
+
DDDDH~GHIIIIIIIIIIHHIIIIIIIIHHIIIIIIIIHHIIIIIIIIHHIIIIII
HHISEQ:549:HLNYBCXY:1:1101:2196:2168 1:N:0:CACTCAAT
TCGTCGCTCGAACAGGATTAGATACCTGGTAGTCCACGCCATAACGATGACAA
+
DDDDIIIIIIIIIIHHIIIIIIIIHHIIIIIIIIHHIIIIIIIIHHIIIIIIII
HHISEQ:549:HLNYBCXY:1:1101:2025:2183 1:N:0:CACTCAAT
ATATCGCGAGAACAGGATTAGATACCTGGTAGTCCACGCCGTAACGATGAGCG
+
DDDD@E@HGHIIIIHHFHHIIIIIFHHIIIIHHGHIIIIICHDEHHIIIIHGH
HHISEQ:549:HLNYBCXY:1:1101:2052:2198 1:N:0:CACTCAAT
CACGAGACAGAACAGGATTAGATACCTGGTAGTCCACGCTGTAACGATGGGTA
+
D@DD@H=7CCHIIIIIIIIIIHHIIIIIIIIHHIIIIIIIIHHIIIIIIIIHHIIII
```

序列:  $10^6 \sim 10^9$

ID	WT6	WT3	OE4	WT2	OE3	WT1
OTU_265	18	18	6	11	20	15
OTU_36	63	77	57	194	155	163
OTU_102	20	44	18	77	18	43
OTU_49	106	92	25	137	76	65
OTU_270	9	5	22	5	22	5
OTU_1865	0	3	0	0	0	2
OTU_58	77	75	28	84	53	64
OTU_1110	6	3	3	2	2	2
OTU_30	100	142	78	111	124	145
OTU_51	87	79	21	38	42	102
OTU_1353	0	1	2	0	1	1
OTU_1137	0	1	0	3	0	0
OTU_18	166	150	126	318	130	265
OTU_4	498	343	189	804	224	626
OTU_3	459	690	340	1039	568	580
OTU_704	3	14	12	8	9	4
OTU_14	176	283	110	314	169	232

特征表:  $10^{1\sim3} \times 10^{3\sim5}$

Sample	berger_parker	buzas_gibson	chaol
WT6	0.042	0.0381	1388.9
WT3	0.0453	0.0475	1474.9
OE4	0.0359	0.0414	1476.4
WT2	0.0642	0.0244	1203.0
OE3	0.0426	0.0396	1716.9
WT1	0.0586	0.0293	1317.0
WT4	0.0518	0.0359	1353.2
OE5	0.0361	0.0441	1622.8
OE2	0.0466	0.0472	1733.3
OE6	0.0432	0.0523	1759.5
WT5	0.0435	0.0252	1181.6
OE1	0.0374	0.0524	1591.2
K04	0.0558	0.0325	1474.1
K01	0.0552	0.0409	1651.6
K05	0.0732	0.025	1306.2
K02	0.0509	0.0445	1675.3
K03	0.0571	0.0329	1489.8
K06	0.0518	0.0334	1215.9

统计表:  $1\sim N \times 10^{1\sim3}$

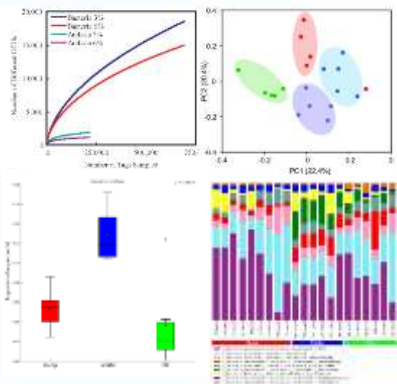


图:  $10^{1\sim3}$ 个点和统计信息

# 宏基因组有参分析基本思路

16S rRNA基因扩增子

宏基因组

U/VSEARCH

QIIME 2

MetaPhlAn4

Kraken 2

HUMAnN4

物种组成

	Sample 1	Sample 2	Sample 3
OTU_1	4	0	2
OTU_2	1	0	0
OTU_3	2	4	2

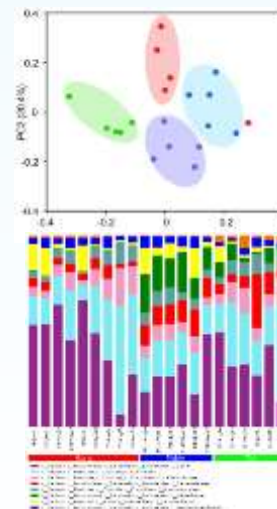
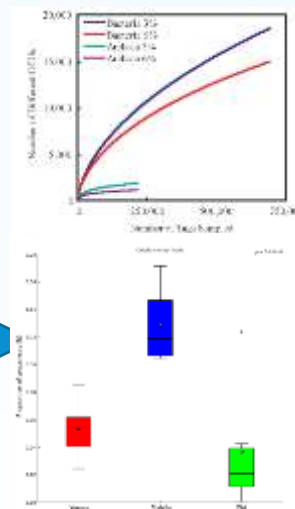
PICRUST2

Tax4Fun2

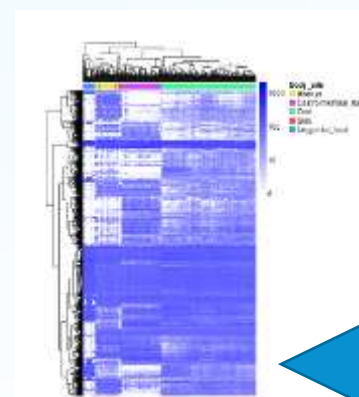
	Sample 1	Sample 2	Sample 3
K00001	20	15	18
K00002	1	2	0
K00003	4	5	4

功能组成

STAMP /  
LEfSe / R



STAMP /  
LEfSe / R



# 宏基因组实验分析流程

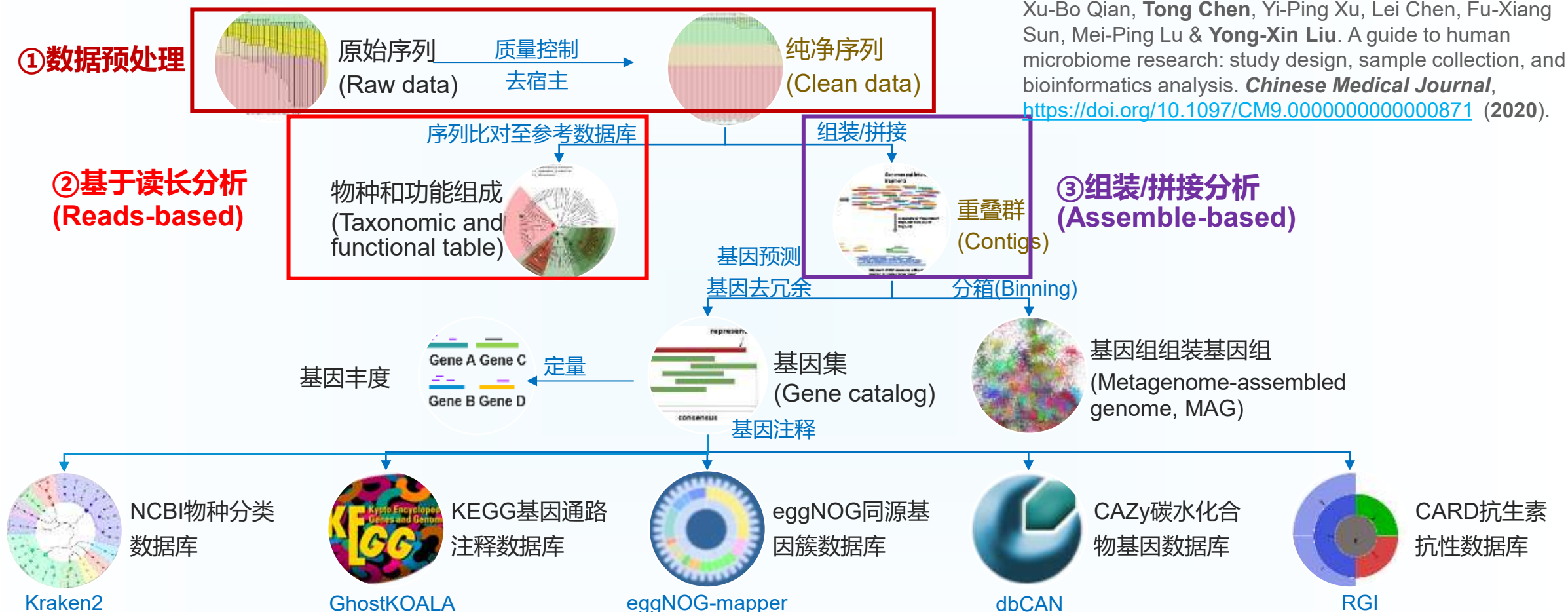
DNA提取

测序

质控, 比对/  
组装注释

物种功能  
组成分析

# 宏基因组分析流程



Xu-Bo Qian, Tong Chen, Yi-Ping Xu, Lei Chen, Fu-Xiang Sun, Mei-Ping Lu & Yong-Xin Liu. A guide to human microbiome research: study design, sample collection, and bioinformatics analysis. *Chinese Medical Journal*, <https://doi.org/10.1097/CM9.0000000000000871> (2020).

# 宏基因组测序技术可以回答的科学问题

回答3个科学问题：

## 1. 样品中有什么？

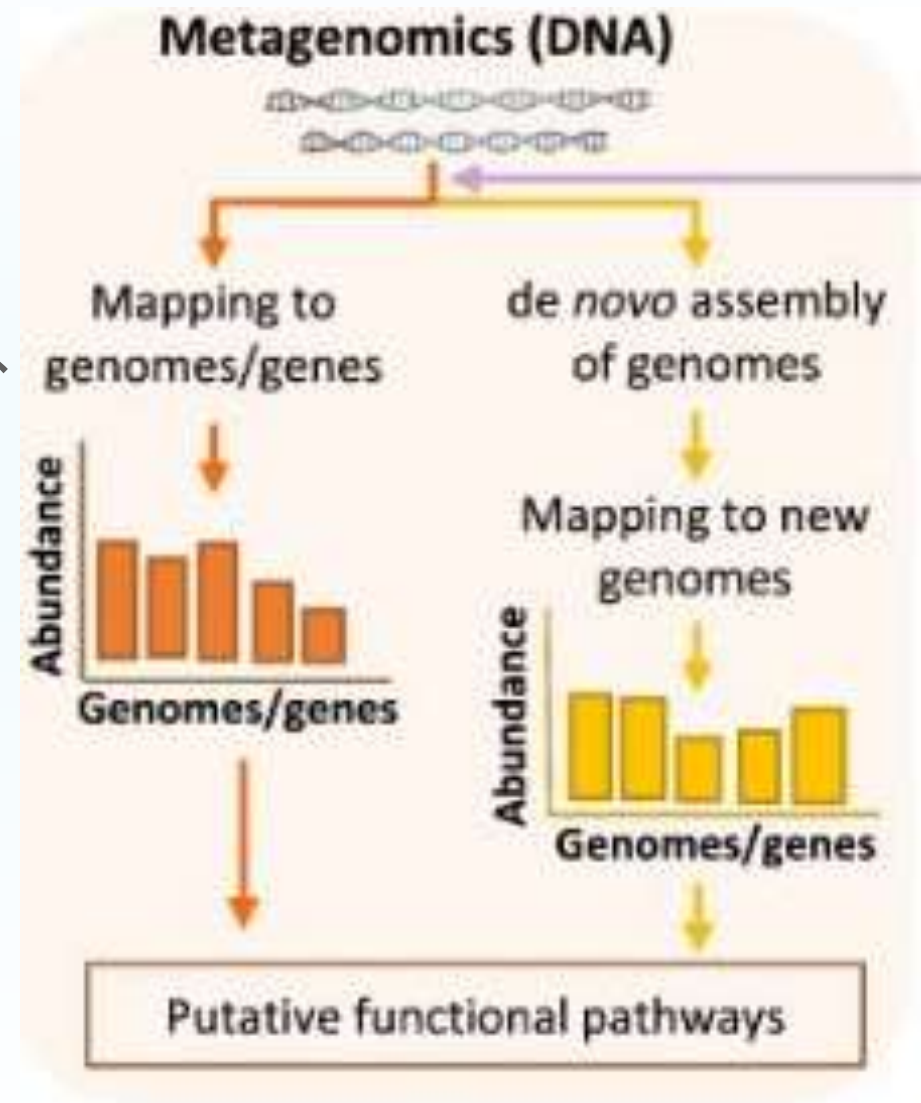
物种组成(包括宿主、细菌、真菌、病毒、原生动物等)

## 2. 样品中有哪些功能基因？

功能基因组成——潜在的功能

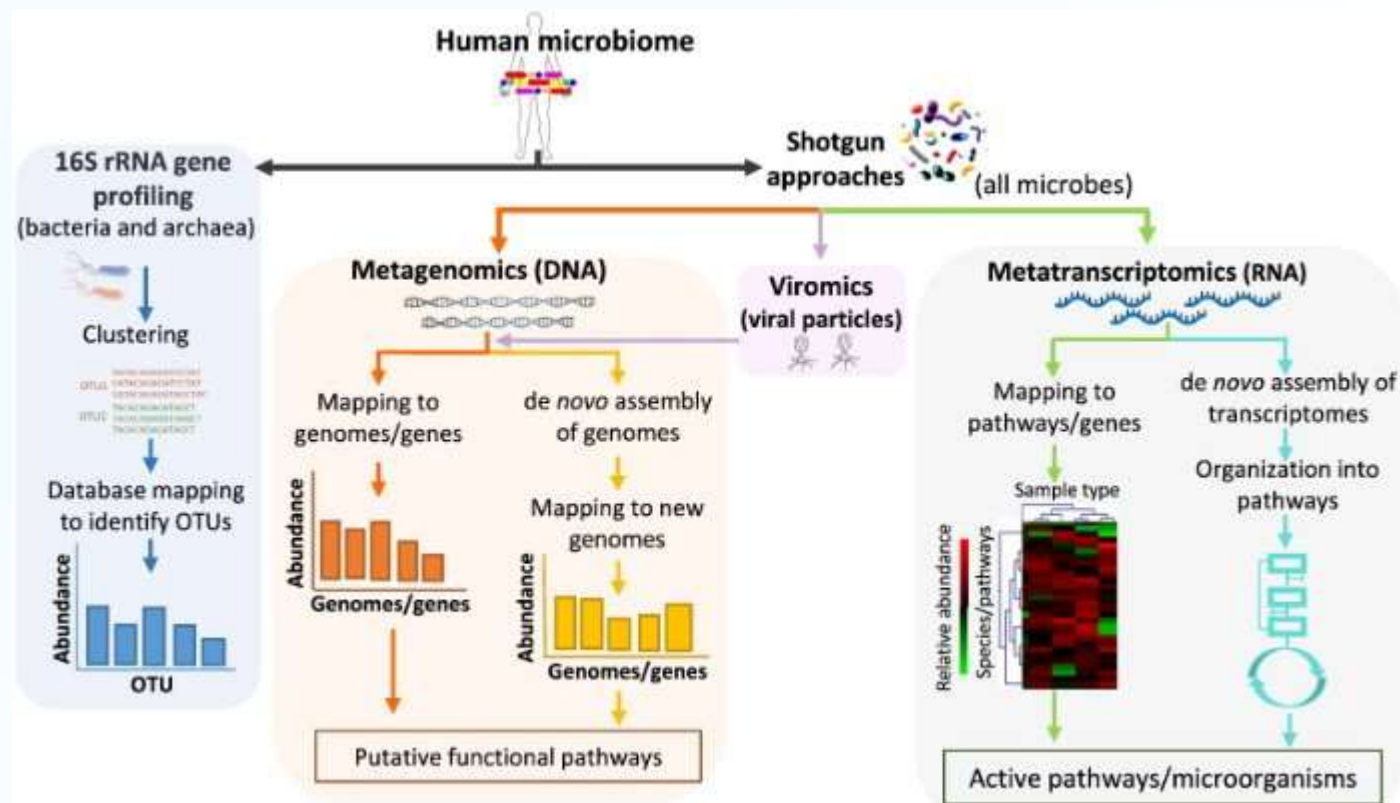
## 3. 组间物种和功能差异？

分组有关的物种分类(界/门/纲/目/科/属/种/株)和功能(通路/模块/同源簇/基因)



# 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去宿主
- 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



# 易宏基因组EasyMetagenome安装

## ○ Windows安装常用软件方便本地绘图

Windows用户 (占PC中87.5%) 为例, 快速下载[GitForWindows](#) [[图文视频教程](#)], [R](#), [RStudio](#) [[图文视频教程](#)], [STAMP v2.1.3](#)

## ○ Linux安装易扩增子 (EasyAmplicon)

EasyAmplicon流程(正对照) <https://github.com/YongxinLiu/EasyMetagenome>

EasyMicrobiome软件和数据库 <https://github.com/YongxinLiu/EasyMicrobiome>

方法 1. 访问上方链接, 点击 Code -- Download

方法2. 百度网盘: [https://pan.baidu.com/s/1lkd\\_47HHODOqC3Rcx6eJ6Q?pwd=0315](https://pan.baidu.com/s/1lkd_47HHODOqC3Rcx6eJ6Q?pwd=0315)

[Linux下所有软件和数据库安装参考EasyMetagenome中0Install.sh中代码运行即可](#)

# 常用生物信息小工具推荐(EasyMicrobiome)

## ○ 陈实富GitHub主页 <https://github.com/OpenGene>

fastp 1.0.1: Fastq序列质控

MutScan v1.14.1: 突变位置检测和可视化

repaq v0.3.0: Fastq序列高压压缩比快速解压

**Shifu Chen. 2023. fastp. *iMeta* 2: e107. <https://doi.org/10.1002/imt2.107>**

[iMeta | 海普洛斯陈实富发布新版fastp, 更快更好地处理FASTQ数据](#)



## ○ 沈伟GitHub主页 <https://github.com/shenwei356>

通用工具支持Windows / Linux / MacOS的32/64位系统, 支持下载或conda安装

seqkit 2.4: 序列处理

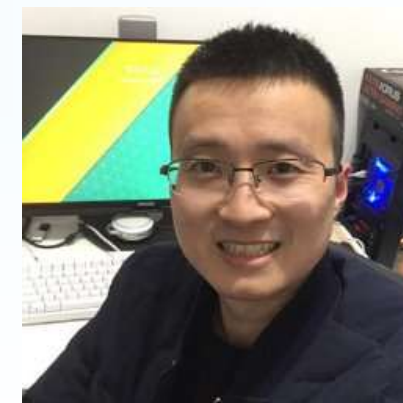
csvtk v0.25.0: 表格处理

taxonkit v0.14.1: NCBI物种信息查询和整理

rush v0.5.0: 任务并行管理软件

**Wei Shen, et al. 2024. SeqKit2. *iMeta* 3: e191. <https://doi.org/10.1002/imt2.191>**

[iMeta | 重医大沈伟等发布全能序列处理工具SeqKit2](#)



# fastp: fastq数据质量评估和质控

- 主页: <https://github.com/OpenGene/fastp>
- 安装 `conda install fastp -c bioconda`
- 下载 `wget http://opengene.org/fastp/fastp` 添加权限 `chmod a+x ./fastp`
- 示例: 适合单独质控或无需去宿主的环境样本, 分析速度极快  
`mkdir -p temp/qc`  
`i=C1`  
`fastp -i seq/${i}_1.fq.gz -o temp/qc/${i}_1.fastq -l seq/${i}_2.fq.gz -O temp/qc/${i}_2.fastq`
- 质控前后报告见 [fastp.html](#)

[iMeta高引论文 | fastp 1.0: FASTQ数据质量控制与预处理的超快速全能工具](#)

# seqkit: fastq数据基本统计和操作

- seqkit: 序列梳理神器-统计、格式转换、长度筛选、质量值转换、翻译、反向互补、抽样、去重、滑窗、拆分等30项全能
- 安装 `conda install seqkit -c bioconda`
- 可选在 <https://github.com/shenwei356/seqkit/releases> 发布页下载
- 样本批量统计 `seqkit stat seq/*.fq.gz`

file	format	type	num_seqs	sum_len	min_len	avg_len	max_len
seq/C1_1.fq.gz	FASTQ	DNA	1,000,000	150,000,000	150	150	150
seq/C1_2.fq.gz	FASTQ	DNA	1,000,000	150,000,000	150	150	150
seq/C2_1.fq.gz	FASTQ	DNA	1,000,000	150,000,000	150	150	150
seq/C2_2.fq.gz	FASTQ	DNA	1,000,000	150,000,000	150	150	150
seq/C3_1.fq.gz	FASTQ	DNA	1,000,000	150,000,000	150	150	150
seq/C3_2.fq.gz	FASTQ	DNA	1,000,000	150,000,000	150	150	150

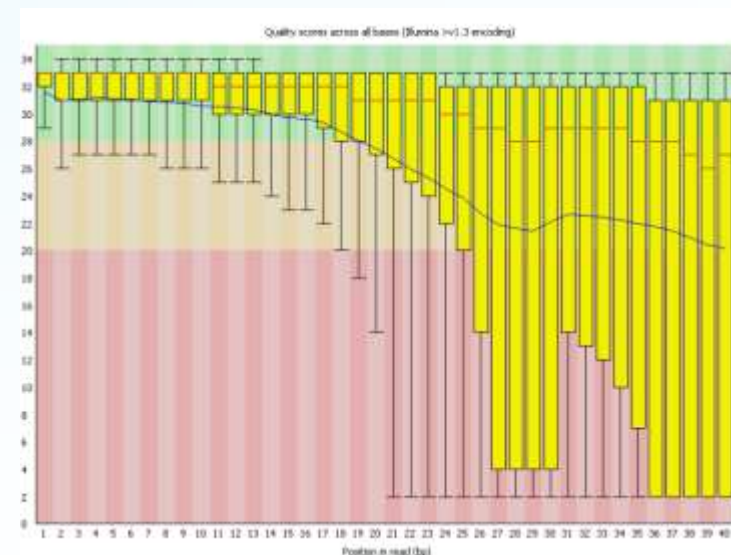
# 本节内容大纲

## 一. 软件安装和数据库部署

- Conda简介与安装
- 软件安装
- 数据库部署

## 二. KneadData去宿主

- FastQC评估和MultiQC汇总结果
- Fastp数据质控
- KneadData去宿主





- Conda是(Python, R, Java, C等)软件包和环境管理系统，用于安装多个版本的软件包及其依赖关系，并在它们之间轻松切换。
- 开源软件，支持Windows、MacOS和Linux(软件最多)三大主流系统
- 容易安装、升级软件及依赖包；
- 方便创建、保存、加载和切换不同的环境变量(如Python2/3)
- Conda由本地软件(Anaconda/Miniconda)和远程软件仓库组成
- 推荐安装Miniconda
- 生物软件安装必添加Bioconda频道

# 推荐Miniconda3

- 最流行的Python数据科学管理平台
- <https://conda.io/miniconda.html> 推荐下载Linux python3 64位版本

# 下载软件，可根据官网下载最新版本

```
wget -c https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
```

# 安装，如管理员推荐安装目录设为conda，普通用户根据个人喜好设定或使用默认值~/miniconda3，其它选项全yes

```
bash Miniconda3-latest-Linux-x86_64.sh -b -f
```

详细教程见：[Nature Method: Bioconda](#)解决生物软件安装的烦恼

- BioConda是conda系统的生物信息软件专用频道，包括4部分：
- 可用软件清单 [http://bioconda.github.io/conda-package\\_index.html](http://bioconda.github.io/conda-package_index.html)
- 软件部署系统，方便用户定制软件及依赖关系
- [8627个生物信息软件/包及多版本](#)，如收录fastqc就有29个版本
- 超千人添加、修改、升级和维护软件清单
- [2017年发布于bioRxiv](#)；[2018年以通讯发表于Nature Methods](#)，以后可以优雅的引用它(吃水不忘挖井人)，被引1000+次
- 添加频道：conda config --add channels bioconda  
[Nature Method: Bioconda解决生物软件安装的烦恼](#)  
<https://bioconda.github.io/> <https://doi.org/10.1038/s41592-018-0046-7>

# (可选)清华/北外维护的Anaconda镜像站加速下载

# 添加北外镜像加速下载

```
site= https://mirrors.bfsu.edu.cn/anaconda/
```

```
conda config --add channels ${site}/pkgs/free/
```

```
conda config --add channels ${site}/pkgs/main/
```

```
conda config --add channels ${site}/pkgs/r/
```

```
conda config --add channels ${site}/cloud/conda-forge/
```

```
conda config --add channels ${site}/cloud/bioconda/
```

# 如果不可用，请手动在conda配置文件 ~/.condarc 中手动删除

# 质控软件安装

## ○ # 质量评估软件fastqc

```
conda install fastqc
```

```
fastqc -v # FastQC v0.12.1
```

## ○ # 多样品评估报告汇总multiqc

```
conda install multiqc
```

```
multiqc --version # multiqc, version 1.14
```

## ○ # fastp质控和kneaddata去宿主，安装最新/指定版解决ID问题

```
conda install fastp
```

```
conda install kneaddata
```

```
kneaddata --version # 0.12.0
```

```
conda install kneaddata=0.12.0
```

**注意记录安装软件版本！**

**默认安装工作环境兼容的最新版，保证可运行且功能最全**

**有问题时安装指定版本，确保分析结果正确；**

# 质控相关数据库安装——人类基因组

- # 查看可用数据库

```
kneaddata_database
```

- # 包括人类基因组human\_genome bowtie2、转录组、小鼠基因组、核糖体SILVA128数据库

- # 如下载人类基因组bowtie2索引至指定数据目录

```
mkdir -p ~/db/kneaddata/human
```

```
kneaddata_database --download human_genome bowtie2 ~/db/kneaddata/human
```

- 其它物种可自行下载并使用bowtie2建索引，可参考代码或下方链接

# 自定义基因组构建bowtie2索引-Kneaddata去宿主

- 大多数基因组可在ensembl genome下载。此处以拟南芥为例，访问 <http://plants.ensembl.org/index.html>，选择Arabidopsis thaliana —— Download DNA sequence (FASTA)，选择toplevel右键复制链接

# 新建目录、进入并下载链接

```
mkdir -p ${db}/kneaddata/ath && cd ${db}/kneaddata/ath
```

```
wget -c http://ftp.ensemblgenomes.org/pub/plants/release-51/fasta/arabidopsis_thaliana/dna/Arabidopsis_thaliana.TAIR10.dna.toplevel.fa.gz
```

# 简化文件名并解压

```
mv Arabidopsis_thaliana.TAIR10.dna.toplevel.fa.gz tair10.fa.gz
```

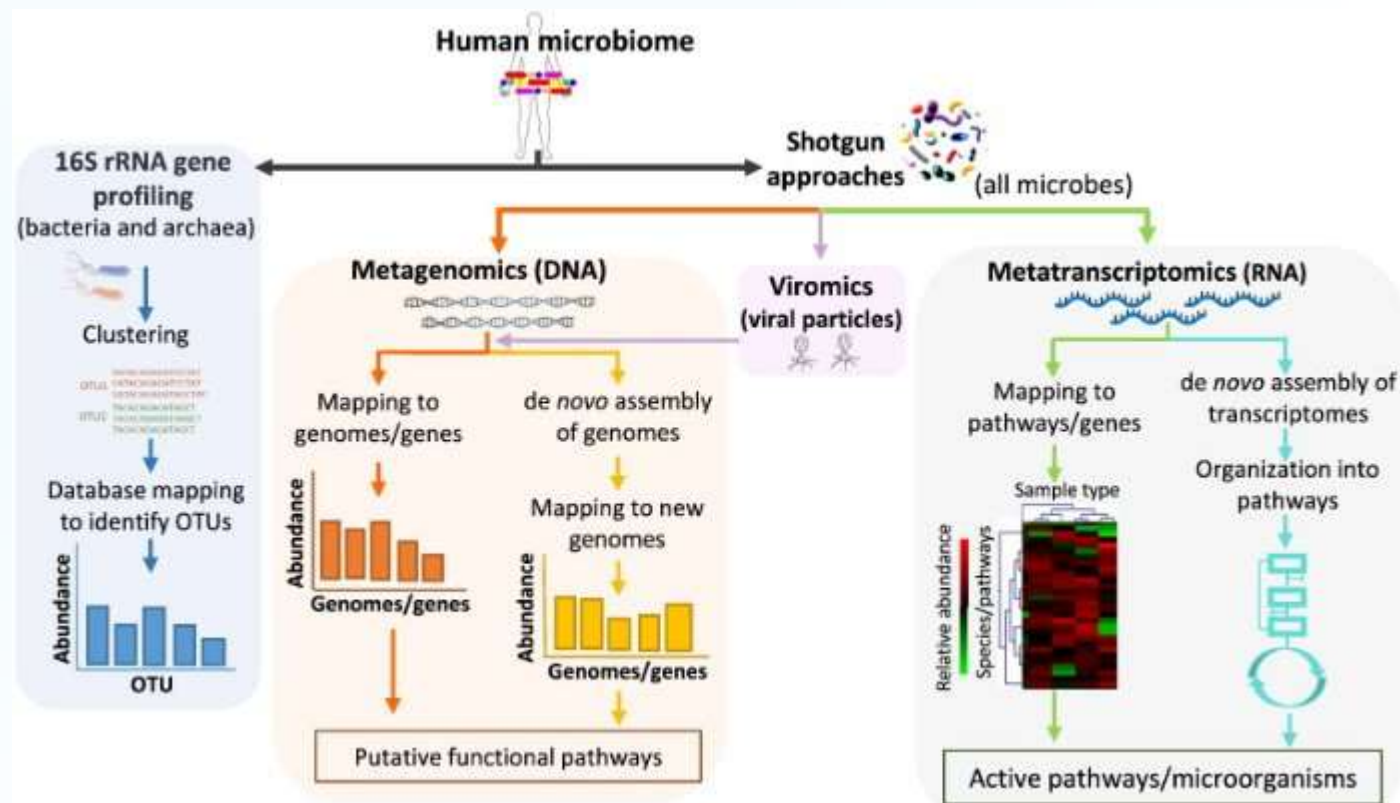
```
gunzip tair10.fa.gz
```

# bowtie 2建索引，输入文件，输出文件前缀，9线程2分

```
time bowtie2-build -f tair10.fa tair10 --threads 9 --seed 1
```

# 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去宿主**
- 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



# 分析开始前必须设置环境变量

- # 公共数据库database位置, 如db公用可能为/db, 而自己下载可能为~/db
- **db=~/db**
- # Conda软件software安装目录, 如db公用可能为/conda, 而自己下载可能为~/miniconda3
- **soft=~/miniconda3**
- # wd为项目工作目录work directory, 如meta
- **wd=~/meta**

# 了解宏基因组分析起始文件(上传到服务器)

- 测序数据：成对测序文件seq/\*.fq.gz，通常为压缩的gz格式

C1\_1.fq.gz C2\_1.fq.gz C3\_1.fq.gz Y1\_1.fq.gz Y2\_1.fq.gz Y3\_1.fq.gz  
C1\_2.fq.gz C2\_2.fq.gz C3\_2.fq.gz Y1\_2.fq.gz Y2\_2.fq.gz Y3\_2.fq.gz

```
@A00808:940:HYJTFDSX2:2:1101:2899:1063 1:N:0:GCCTTACAAC+CTCAAGACAC
CNGCTGGCTGAGCTGCCGACCAAGACGGCCGTCAAGCTCACCGCCGAAATCACCGGCGCCAACCGCAACACCTTGTATGA
+
F#FFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00808:940:HYJTFDSX2:2:1101:8847:1063 1:N:0:GCCTTACAAC+CTCAAGACAC
CNTCCGCAGAAATAATCGCCCTTCGCGAAAAGCAAAAACCCGGCTTGCCGGGTTTTTTTATGTCTGATGTGAGGTCCTA
+
, #FFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF, FFFFFF:FFFFFF:FFF:FFFFFF, FFFF:FFFFFF:, FFF:FFFFFF,
```

- 实验设计：样本名和分组 result/metadata.txt

SampleID	SampleIDOld	Group	Site	Cooperation	latitude longitude	Date	CRA	CRR
Y1	YA04	Young	ZhanJiang	Majorbio	21.19 N 110.24 E	2022/10/8	CRA032890	CRR2274860
Y2	YA07	Young	ZhanJiang	Majorbio	21.19 N 110.24 E	2022/10/8	CRA032890	CRR2274861
Y3	YB59	Young	ZhanJiang	Majorbio	21.19 N 110.24 E	2022/10/8	CRA032890	CRR2274862
C1	G18	Centenarians	ZhanJiang	Nanxin	21.19 N 110.24 E	2022/10/8	CRA032890	CRR2274863
C2	A15	Centenarians	ZhanJiang	Nanxin	21.19 N 110.24 E	2022/10/8	CRA032890	CRR2274864
C3	Z17	Centenarians	ZhanJiang	Nanxin	21.19 N 110.24 E	2022/10/8	CRA032890	CRR2274865

# 并行管理软件 rush

- 现实中是有一大堆样品，for可以单个或全部提交任务效率都很低，如何让服务器性能允许下并行加速分析，并有序管理队伍呢？
- 国人开发了跨平台的并行管理工具rush，[官网下载](https://github.com/shenwei356/rush)或 conda安装  
conda install rush

官网：<https://github.com/shenwei356/rush>

- 使用格式
- `echo sample1 sample2 | rush -j 2 "command"`
- `tail -n+2 result/metadata.txt | cut -f1 | rush -j 2 "command"`

# fastp批量数据质量评估和质控

# -j 2: 表示同时处理2个样本

```
time tail -n+2 result/metadata.txt|cut -f1|rush -j 2 \  
  "fastp -i seq/{1}_1.fq.gz -l seq/{1}_2.fq.gz \  
    -j temp/qc/{1}_fastp.json -h temp/qc/{1}_fastp.html \  
    -o temp/qc/{1}_1.fastq -O temp/qc/{1}_2.fastq \  
    > temp/qc/{1}.log 2>&1 "
```

# 质控后结果汇总

```
echo -e "SampleID\tRaw\tClean" > temp/fastp  
for i in `tail -n+2 result/metadata.txt|cut -f1`;do  
  echo -e -n "$i\t" >> temp/fastp  
  grep 'total reads' temp/qc/${i}.log|uniq|cut -f2 -d ':'|tr '\n' '\t' >> temp/fastp  
  echo "" >> temp/fastp  
done  
sed -i 's/ //g;s/\t$//' temp/fastp
```

# rush并行Kneaddata去宿主

- **-i输入文件, -o输出目录, -t线程数, -db 宿主基因组索引位置**

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

```
"kneaddata \
```

```
-i1 temp/qc/{1}_1.fastq -i2 temp/qc/{1}_2.fastq \
```

```
-o temp/hr/ \
```

```
--bypass-trim --bypass-trf --reorder \
```

```
--bowtie2-options '--very-sensitive --dovetail' \
```

```
-db ${db}/kneaddata/human/hg_39 \
```

```
--remove-intermediate-output -v -t 3"
```

# 检查结果格式是否正确配对

```
paste <(head -n40 temp/hr/`tail -n+2 result/metadata.txt|cut -f1|head -  
n1`_1.fastq|grep @) <(head -n40 temp/hr/`tail -n+2 result/metadata.txt|cut -  
f1|head -n1`_2.fastq|grep @)
```

```
@A01927:79:HFTKWDSX5:2:1101:4616:1000#0/1    @A01927:79:HFTKWDSX5:2:1101:4616:1000#0/2  
@A01927:79:HFTKWDSX5:2:1101:5466:1000#0/1    @A01927:79:HFTKWDSX5:2:1101:5466:1000#0/2  
@A01927:79:HFTKWDSX5:2:1101:7401:1000#0/1    @A01927:79:HFTKWDSX5:2:1101:7401:1000#0/2  
@A01927:79:HFTKWDSX5:2:1101:8865:1000#0/1    @A01927:79:HFTKWDSX5:2:1101:8865:1000#0/2  
@A01927:79:HFTKWDSX5:2:1101:9516:1000#0/1    @A01927:79:HFTKWDSX5:2:1101:9516:1000#0/2  
@A01927:79:HFTKWDSX5:2:1101:9842:1000#0/1    @A01927:79:HFTKWDSX5:2:1101:9842:1000#0/2  
@A01927:79:HFTKWDSX5:2:1101:10041:1000#0/1   @A01927:79:HFTKWDSX5:2:1101:10041:1000#0/2  
@A01927:79:HFTKWDSX5:2:1101:10276:1000#0/1   @A01927:79:HFTKWDSX5:2:1101:10276:1000#0/2  
@A01927:79:HFTKWDSX5:2:1101:10673:1000#0/1   @A01927:79:HFTKWDSX5:2:1101:10673:1000#0/2  
@A01927:79:HFTKWDSX5:2:1101:10800:1000#0/1   @A01927:79:HFTKWDSX5:2:1101:10800:1000#0/2
```

双端 ID 一致，仅结尾用/1或/2来区分为正确格式

# 质控去宿主 结果文件简化统一(与质控一致)

## ○ 实现简化名

`rename 's/_1_kneaddata_paired/' temp/hr/*.fastq # Ubuntu系统改名`

`rename '_1_kneaddata_paired' " temp/hr/*.fastq # CentOS系统改名`

## ○ 大文件清理，高宿主含量样本可节约>90%空间

`/bin/rm -rf temp/hr/*contam* temp/hr/*unmatched* temp/hr/reformatted*  
temp/hr/_temp*`

`ls -l temp/hr/`

# 确认去宿主结果后，可以删除质控后中间文件

`rm temp/qc/*.fastq`

# 质控结果汇总表

# 合并所有样本统计结果为表

```
kneaddata_read_count_table --input temp/hr -output temp/kneaddata.txt
```

```
csvtk -t pretty temp/kneaddata.txt
```

# 筛选重要的列，并查看结果

```
cut -f 1,2,5,6 temp/kneaddata.txt | sed 's/_1_kneaddata//' > result/qc/sum.txt
```

```
csvtk -t pretty result/qc/sum.txt
```

Sample	raw pair1	decontaminated hg_39 pair1	decontaminated hg_39 pair2
-----	-----	-----	-----
C1	986795.0	984103.0	984103.0
C2	987127.0	987120.0	987120.0
C3	986817.0	986804.0	986804.0
Y1	986797.0	986417.0	986417.0
Y2	985888.0	985798.0	985798.0
Y3	980721.0	980461.0	980461.0

# 质控结果统计和可视化

## ○ # 用R代码统计下质控结果

Rscript -e

```
"data=read.table('result/qc/sum.txt',  
header=T, row.names=1, sep='\t');  
summary(data)"
```

## ○ # R转换宽表格为长表格

```
Rscript -e "library(reshape2);  
data=read.table('result/qc/sum.txt',  
header=T,row.names=1, sep='\t');  
write.table(melt(data),  
file='result/qc/sum_long.txt',sep='\t',  
quote=F, col.names=T, row.names=F)"
```

Essential parameters

Legend variable

variable

Legend variable order

raw.pair1  
decontaminated.hg\_39.pair1  
decontaminated.hg\_39.pair2

Y-axis variable

value



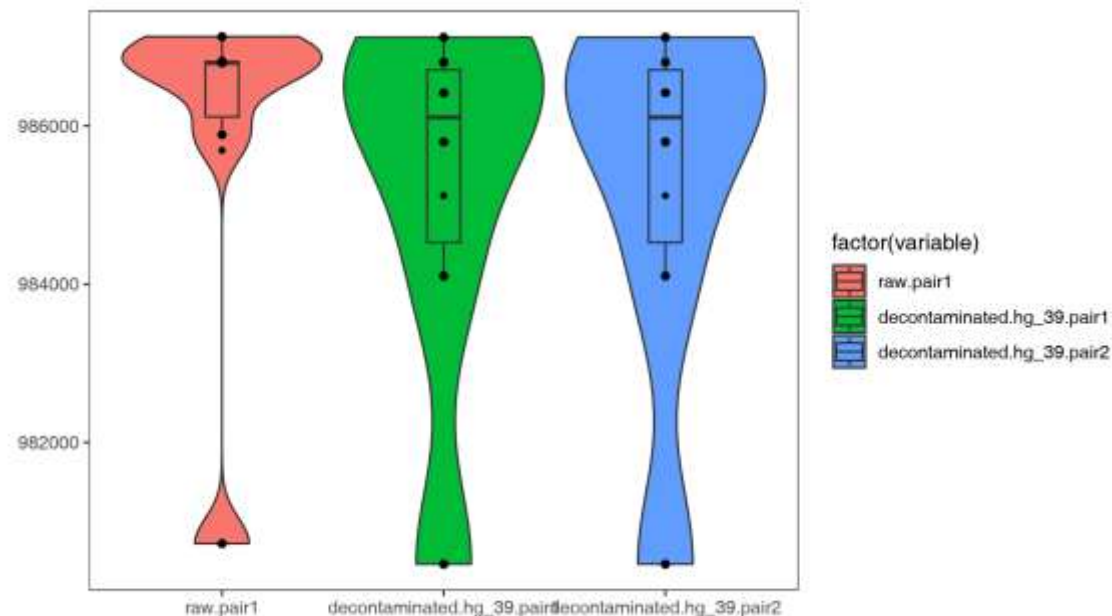
Box+Jitter+Violin plot

X-axis variable

variable

X-axis variable order

E.g. B, D, C, E



<https://www.bic.ac.cn/ImageGP/index.php/Home/Index/Boxplot.html>

# 总结

- EasyMetagenome是全能型宏基因组分析流程，提供质控、有参、无参、分箱、单菌等分析流程和可视化，
- 依赖EasyMicrobiome包括常用软件、脚本和人工整理数据表；
- Conda是软件安装和管理神器，Bioconda频道是生物学家的福音，8千多个生信软件及数十万个版本满足你各种需求；
- 有些软件依赖数据库需要下载，如基因组用于去除宿主来源序列；
- 多任务管理专家 rush，Fastp用于快速质控
- 去宿主流程KneadData，整合Bowtie 2等软件和宿主基因组数据库。

[iMeta | EasyMetagenome: 用户友好且灵活的宏基因组测序数据分析流程](#)  
[GigaScience | 基因组所刘永鑫组-短读长宏基因组数据的去宿主污染软件评估](#)

# 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