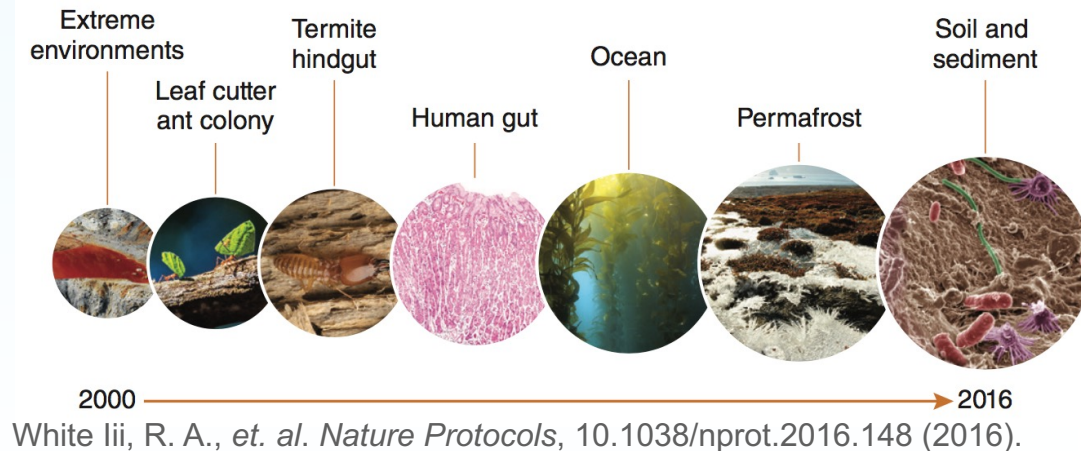


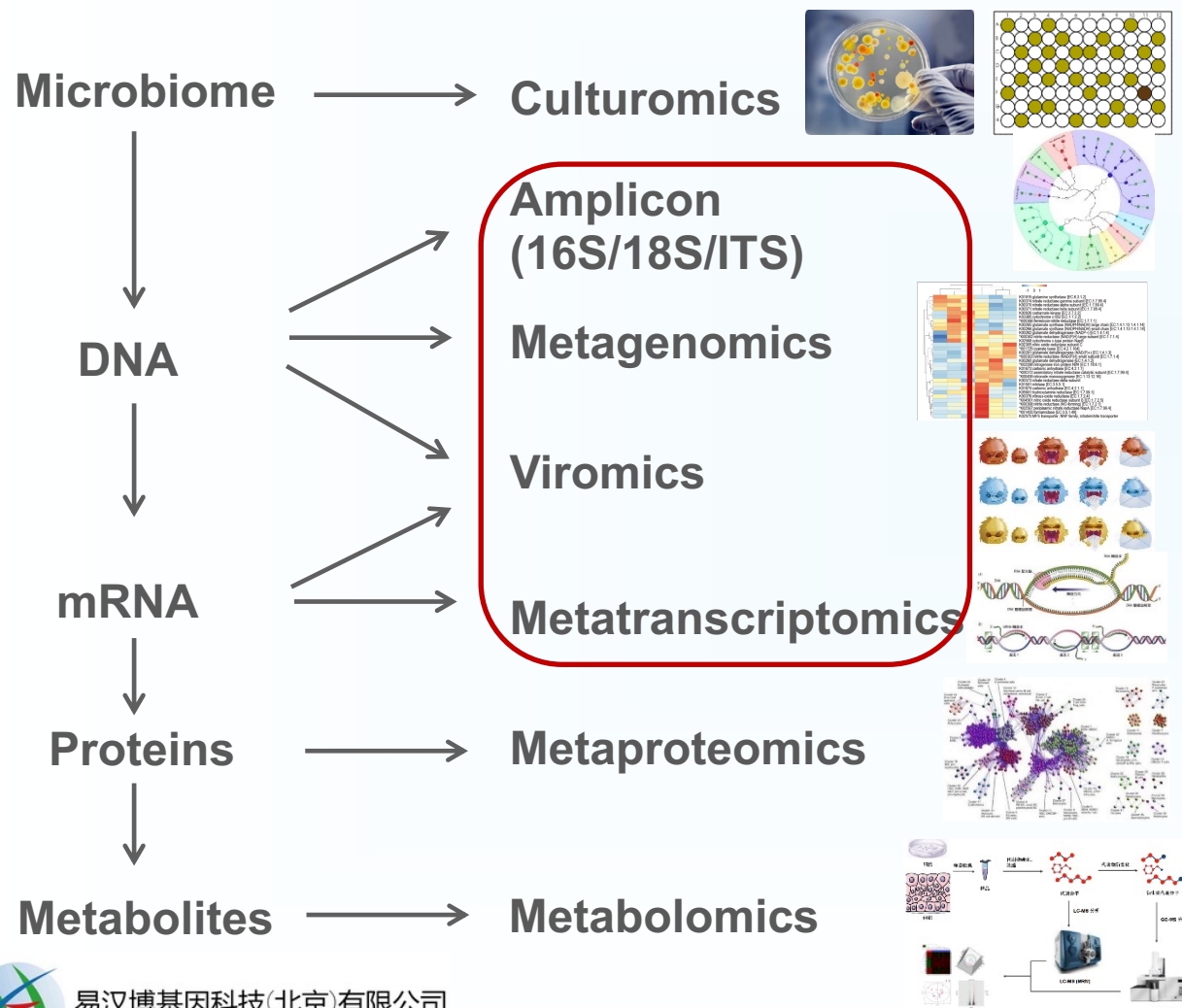


## 11 扩增子16S分析流程介绍

易生信  
2024年4月12日

### Microbiome complexity and multi-omics analysis timeline





## 特点

## 局限

• 高通量，高成本

• 无法分离难培养的物种

• 研究物种多样性

• PCR有偏好性  
• 只能研究分类信息和丰度

• 研究物种多样性和功能  
• 发现新基因

• 无法得到基因表达信息  
• 信息分析复杂

• 同时研究DNA和RNA病毒

• 实验成本高  
• 信息分析复杂

• 研究基因表达信息，寻找样本间表达差异

• 提取和纯化复杂，  
• 数据库信息不完善

• 获得蛋白表达谱  
• 比较不同环境样本的蛋白差异

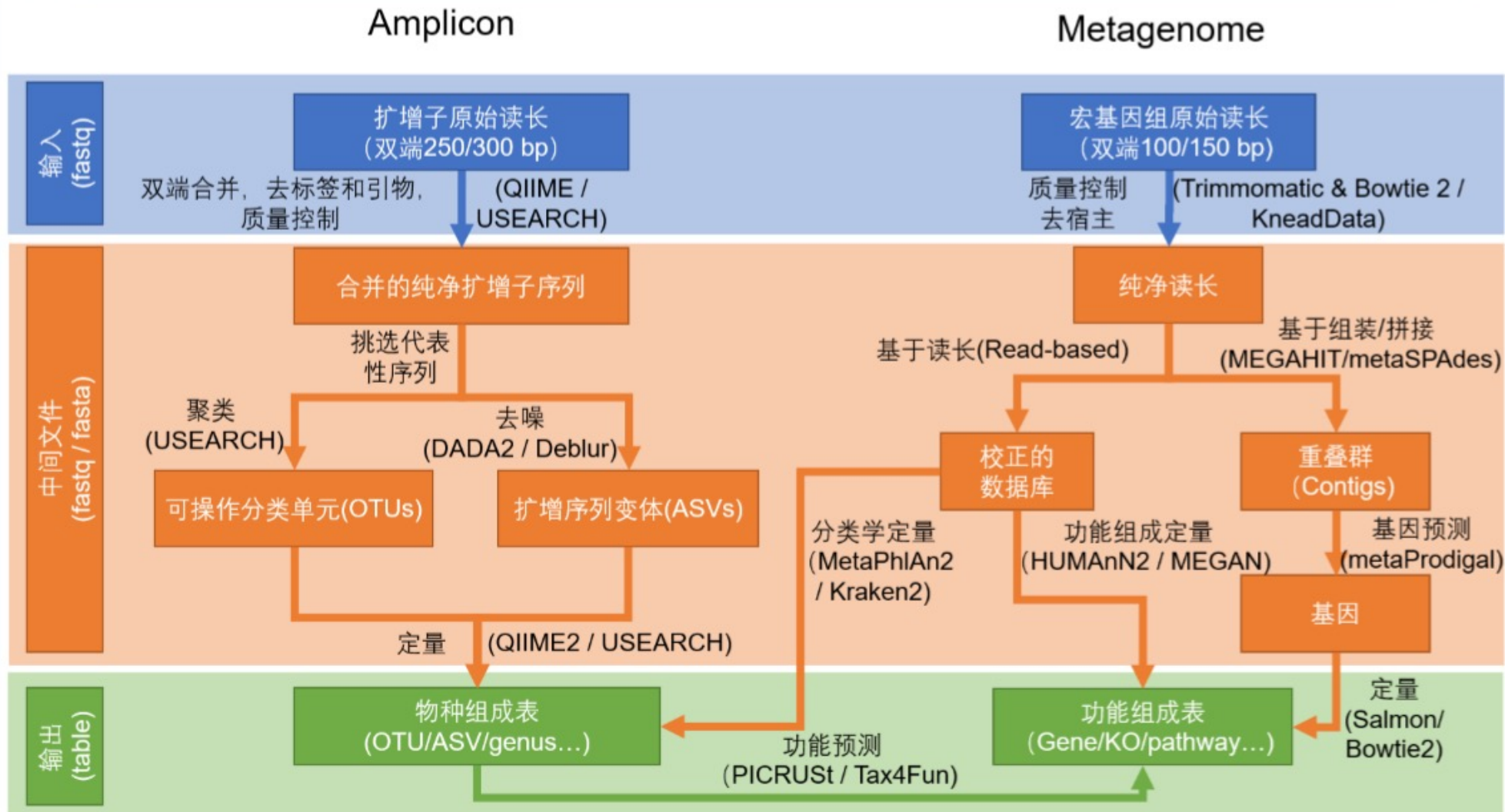
• 数据库信息不完善  
• 总蛋白提取困难

• 获得代谢图谱  
• 比较不同环境样本的代谢物差异

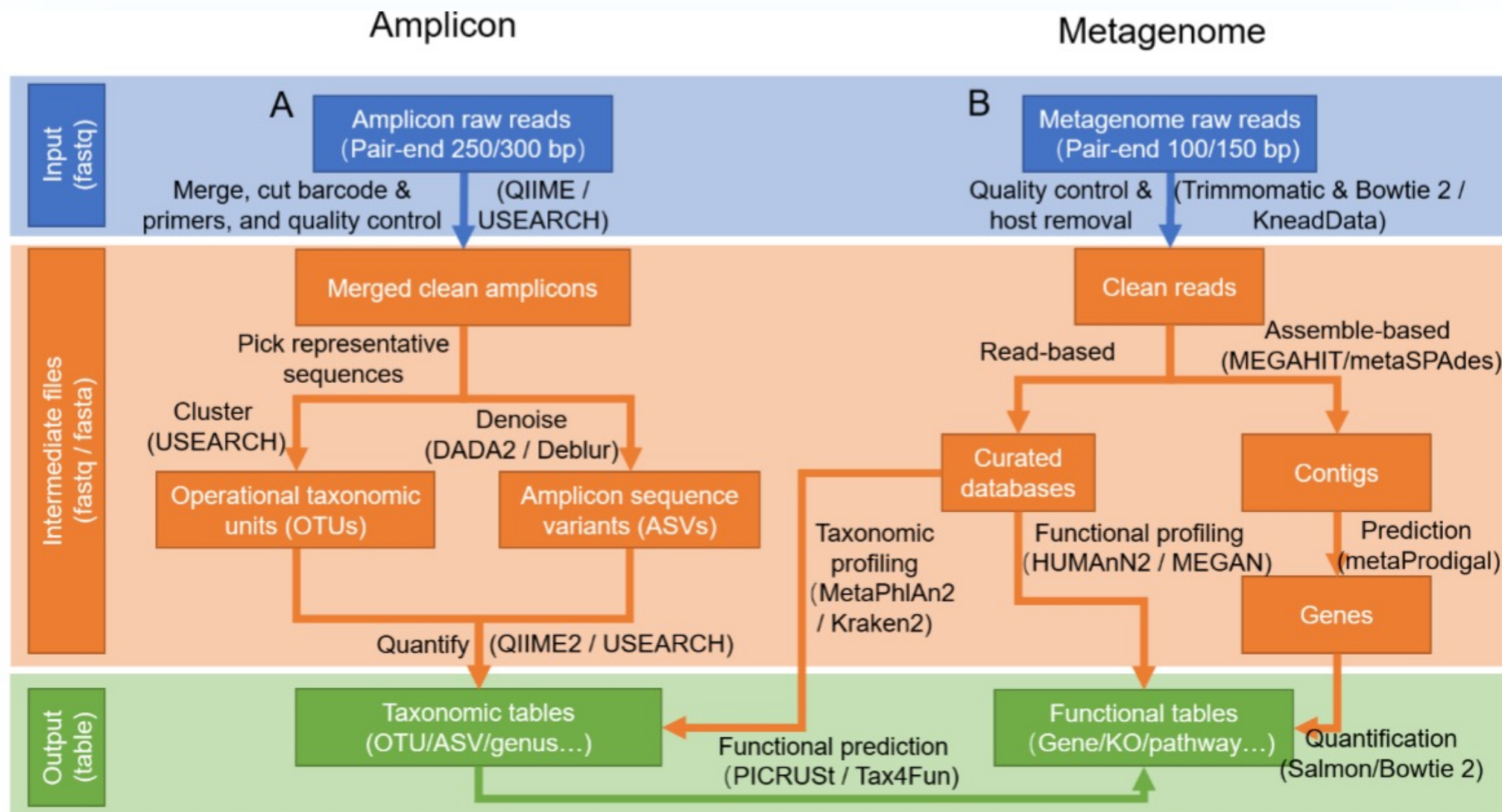
• 数据库信息不完善  
• 难区分宿主和微生物的代谢物



# 扩增子和宏基因组分析思路比较



# 扩增子和宏基因组分析思路比较





# 新版流程 EasyAmplicon 已在 iMeta 发表

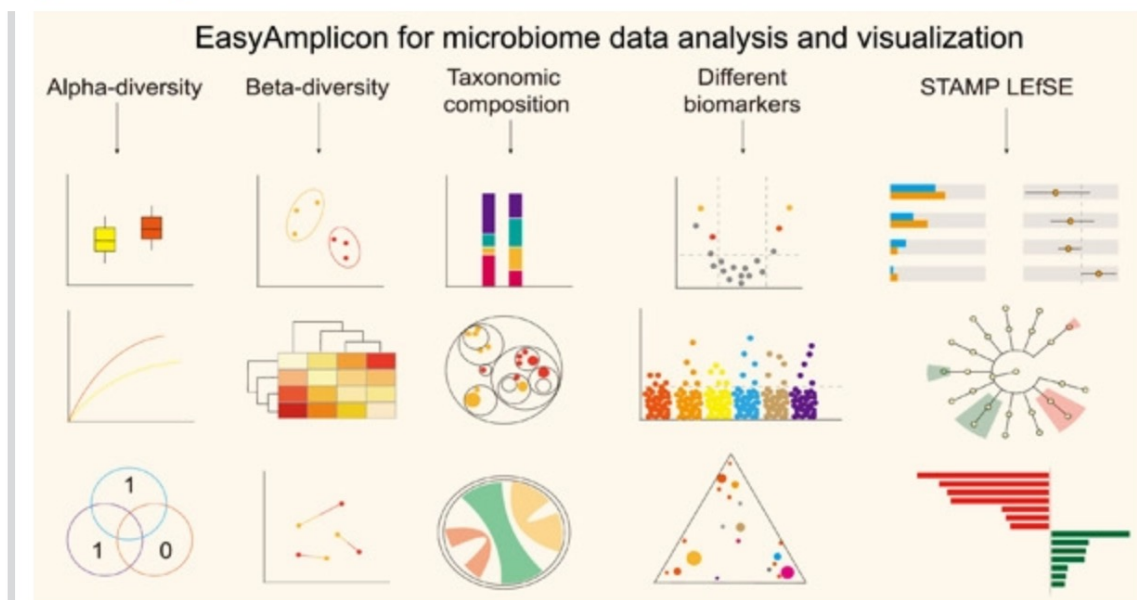
Open Access

## EasyAmplicon: An easy-to-use, open-source, reproducible, and community-based pipeline for amplicon data analysis in microbiome research

Yong-Xin Liu, Lei Chen, Tengfei Ma, Xiaofang Li, Maosheng Zheng, Xin Zhou, Liang Chen, Xubo Qian, Jiao Xi, Hongye Lu, Huiluo Cao, Xiaoya Ma, Bian Bian, Pengfan Zhang, Jiqiu Wu, Ren-You Gan, Baolei Jia, Linyang Sun, Zhicheng Ju, Yunyun Gao, Tao Wen, Tong Chen

e83 | Version of Record online: 27 January 2023

<https://doi.org/10.1002/imt2.83>

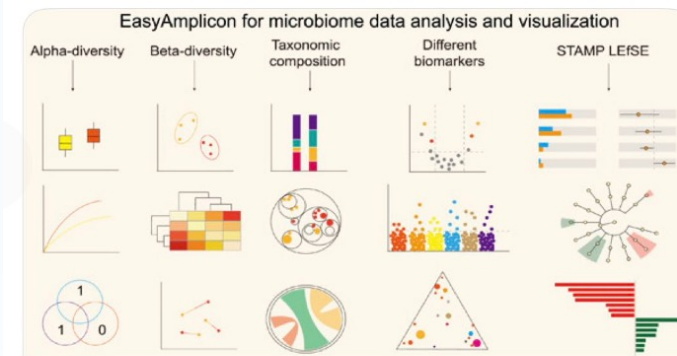


EasyAmplicon is a user-friendly, cross-platform, and community-supported pipeline for amplicon data analysis. It has most of the modules for data processing and visualization in microbiome research. The pipeline is maintained and updated regularly. We encourage users to contribute appropriate code.



Pierre Hohmann  
@plant\_holobiont

EasyAmplicon: An easy-to-use, open-source, reproducible, and community-based pipeline for amplicon data analysis in microbiome research



[onlinelibrary.wiley.com](https://onlinelibrary.wiley.com)

EasyAmplicon: An easy-to-use, open-source, reproducible, and community-based pipeline f...

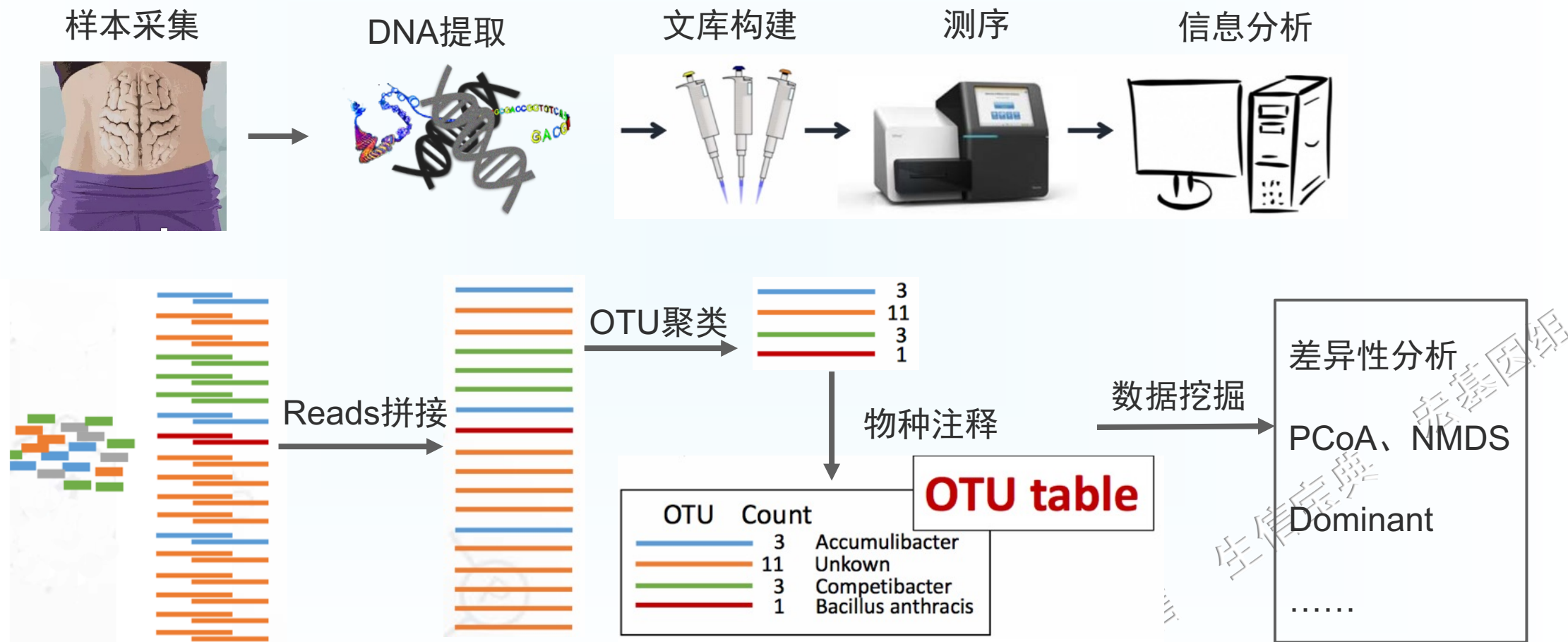
22:28 · 28/01/2023 · 2.7万 次查看

295 喜欢 72 转推 4 引用

[Abstract](#) | [Full text](#) | [PDF](#) | [References](#) | [Request permissions](#)

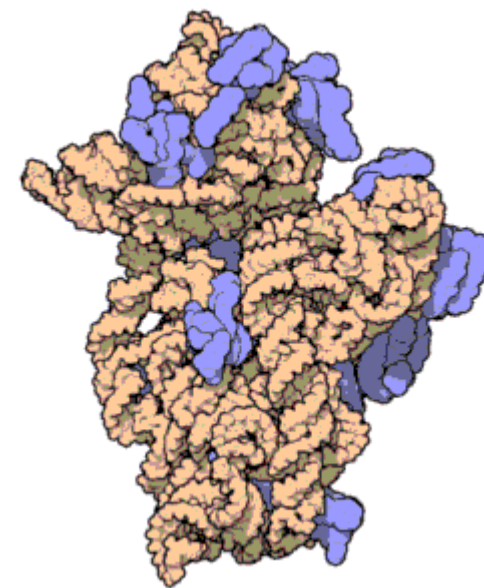
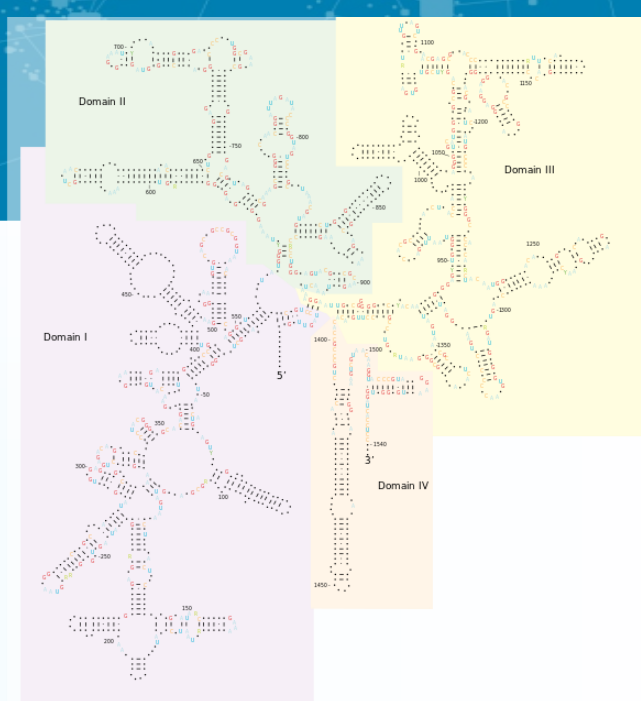
<https://doi.org/10.1002/imt2.83>

# 扩增子实验和分析流程



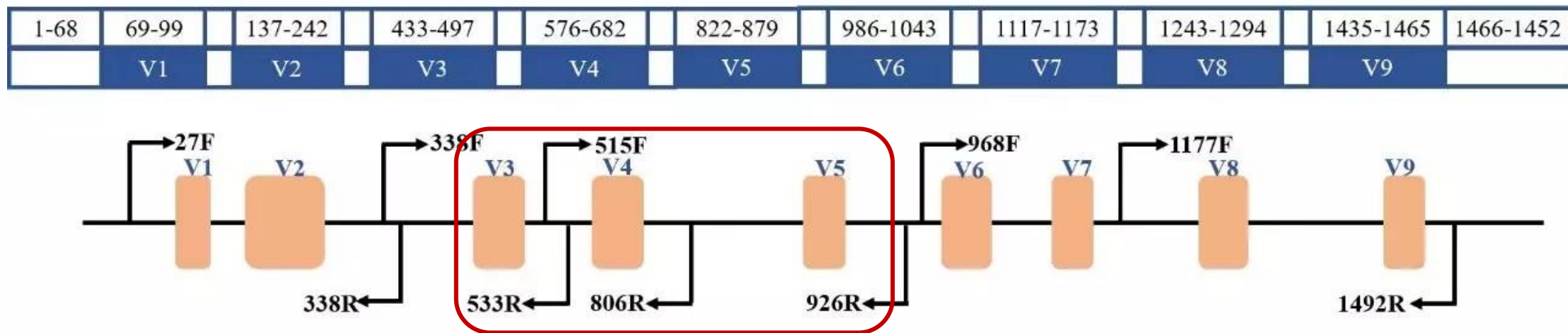
# 16S rDNA / rRNA gene

- **16S rRNA**为核糖体的RNA的一个亚基，16S rDNA就是编码该亚基的基因。细菌**rRNA**（核糖体RNA）按沉降系数分为3种，分别为**5S**、**16S**和**23S rRNA**。**16S rDNA**是细菌染色体上编码**rRNA**相对应的**DNA**序列，存在于所有细菌染色体基因中。
- 16S rDNA是细菌的系统分类研究中最有用的和最常用的分子钟，其种类少，含量大(rRNA约占细菌RNA含量的80%)，分子大小适中，存在于所有的生物中，其进化具有良好的时钟性质，在结构与功能上具有高度的保守性，素有“细菌化石”之称。
- 在大多数原核生物中rDNA都具有多个拷贝，5S、16S、23S rDNA的拷贝数相同。
- 16S rDNA由于大小适中，长度约**1.5Kb**左右，既能体现不同菌属之间的差异，又能利用测序技术较容易地得到其序列，故被细菌学家和分类学家接受。





# 16S rDNA结构与引物的选择



- 所有活着的生物都有核糖体RNA
- rRNAs在蛋白翻译中起至关重要的作用
- rRNAs相对保守，且较少发生水平转移
- 有分子钟的特征，进化分析中非常有用

易生信  
基因库



# 其它常用Marker基因（流程适用各种Marker基因，但注释数据库需要自己选择对应的）

- 真核生物(原生生物, 真菌)  
18S (<http://www.arb-silva.de>)  
ITS (<https://unite.ut.ee/>)
- 细菌  
rRNA转录间隔区ITS (Martiny, Env Micro 2009)  
Chaperonins伴侣蛋白 CPN60 (<http://www.cpndb.ca/>)  
同源重组RecA / 固氮基因NifH
- 病毒  
T4类噬菌体Gp23  
小核糖核酸病毒RdRp

快速进化的Marker基因可以用于菌株水平的分化研究



# 扩增子分析基本流程

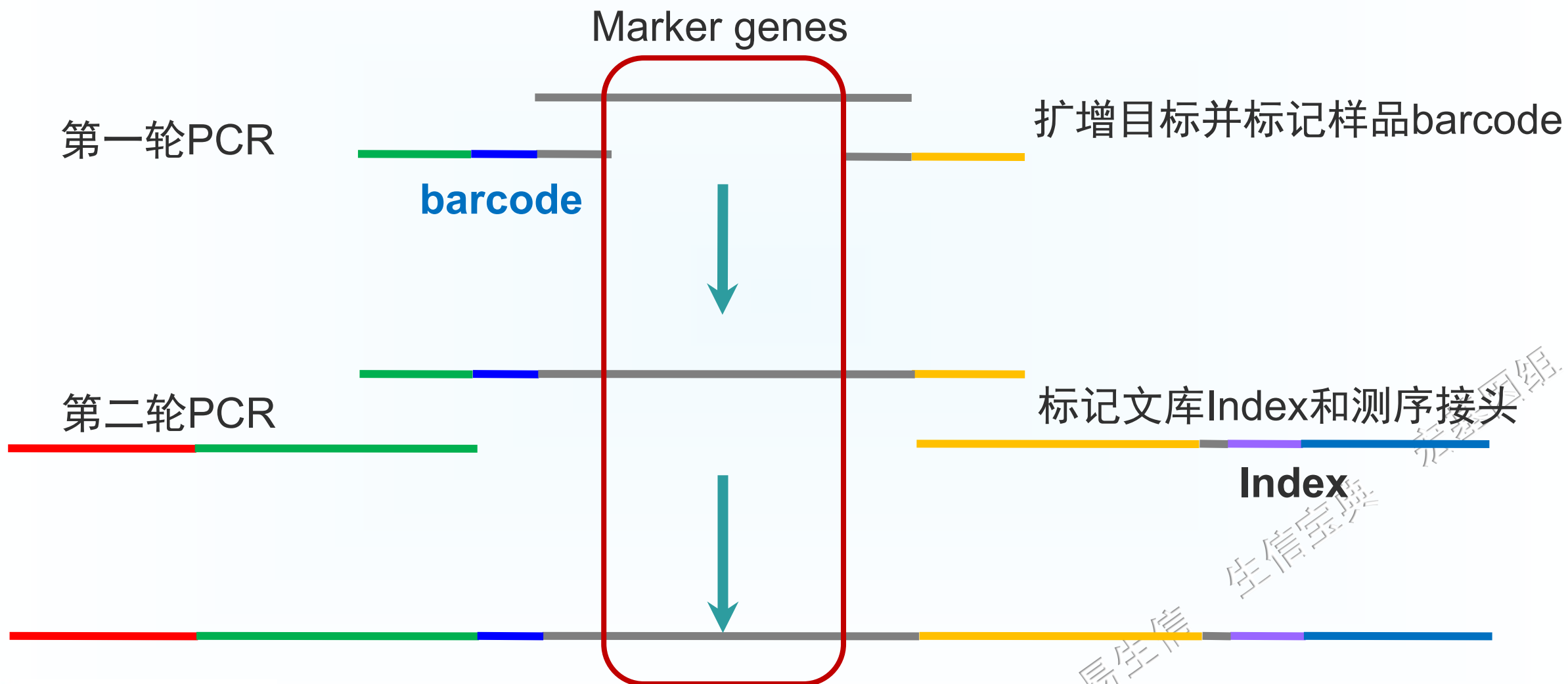
提取DNA

扩增  
测序

质控、(聚类)  
去噪、定量

多样性分析

# 建库原理：主要以两轮PCR为主



Microbiome: HiSeq平台16S扩增子文库构建方法

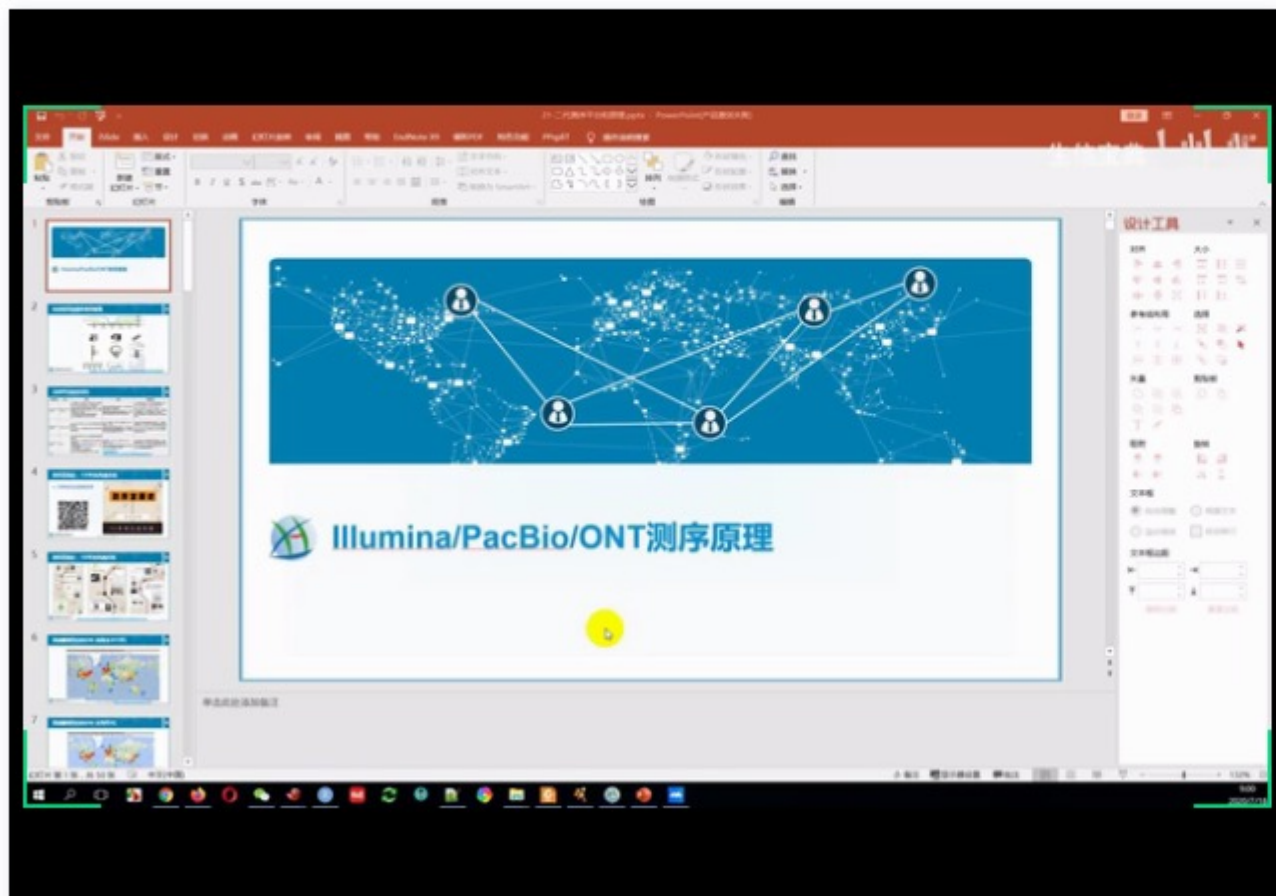
双向Barcode大大增加混样数量，如 $48 \times 96 = 4608$ 个样本



# 补充知识：测序原理

易生信转录组高级课程系列节选

3720播放 · 2弹幕 · 2020-08-31 23:55:34



生信宝典和宏基因组 发消息

欢迎大家关注公众号《生信宝典》和《宏基因...

为TA充电

+ 关注 7776

弹幕列表

展开



0基础学剪辑，在家接单赚钱养活自己！

广告 0基础学剪辑

视频选集 (6/9)

自动连播

P1 转录组概述-基本概念-PPT1-11页 19:12

P2 01-2 转录组概述-批次鉴定和校正-PPT12... 22:34

P3 01-3 转录组概述-实验设计生物重复-PPT... 25:23

P4 05-1-富集分析理论-GO富集分析理论和可... 26:00

P5 06-1-富集分析理论-GSEA理论和结果解读- 31:27

P6 07-1-高通量测序原理-文库构建-1-22... 22:31

P7 17-01-WGCNA理论-理论概述 36:53



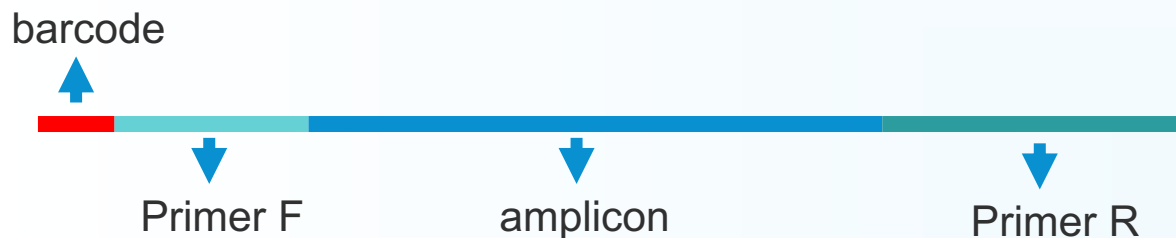
易汉博基因科技(北京)有限公司  
EHBIO Gene Technology (Beijing) co., LTD

易生信，毕生缘；培训版权所有。

生信宝典的B站

# 扩增子结构

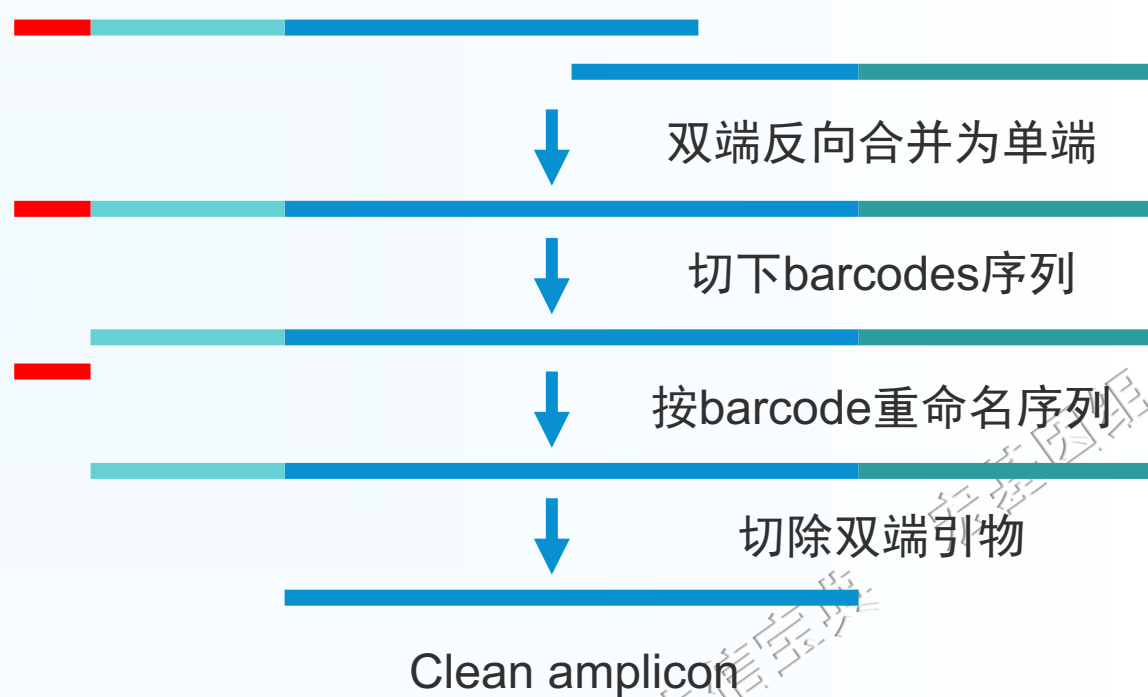
Barcode: 样品标签, 混池测序后区分序列来自那个样本



Primer: 在16S/ITS/18S保守区设计的引物, 用于扩增 rDNA的部分高变区

Amplicon: 扩增的部分 rDNA

Illumina Seq系列下机双端序列



原始数据到扩增子数据分析流程



**It's QUESTION TIME !!**



- OTU (Operational Taxonomic Units) 是在系统发生学研究或群体遗传学研究中, 为了便于进行分析, 人为给某一个分类单元 (品系, 种, 属, 分组等) 设置的同一标志。在生物信息分析中, 一般来说, 测序得到的每一条序列来自一个菌。要了解一个样品测序结果中的菌种、菌属等数目信息, 就需要对序列进行归类操作 (cluster)。通过归类操作, 将序列按照彼此的相似性分归为许多小组, 一个小组就是一个OTU。
- 通常在97%的相似水平下聚类生成OTU, 选择每个聚类群中最高丰度序列作为代表性序列
- 近期讨论发现100%更合理, 即不聚类的ASV (Amplicon Sequence Variants), 更容易实现跨研究比较

16S测序, 不知道OTU你就out了!  
扩增子分析还聚OTU就真OUT了



# OTU 和菌的对应关系

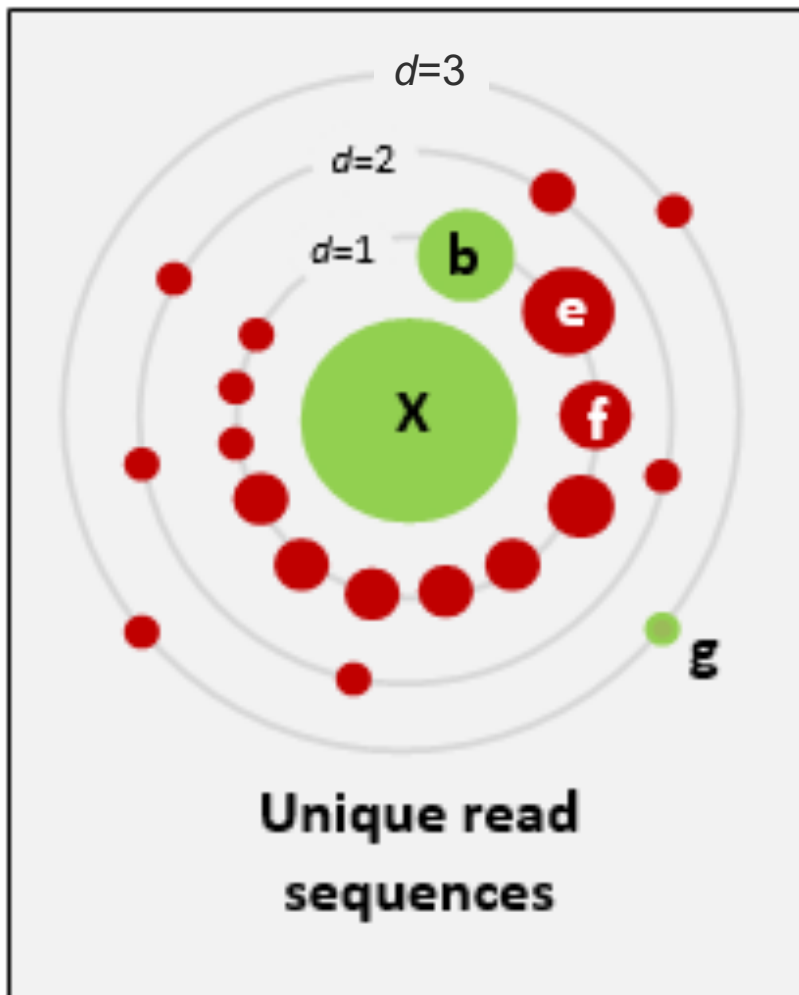
- 在数据分析时，可以理解为一个 OTU 为一个菌
- 扩增子分析中一个 OTU 实际为一条序列
- 一条 OTU 序列能注释到哪个层级取决于：
  - 序列长度
  - 序列本身的区分能力（序列复杂度）
  - 数据库的覆盖度

易生信 生信宝典 宏基因组



# OTU / ASV 鉴定原理

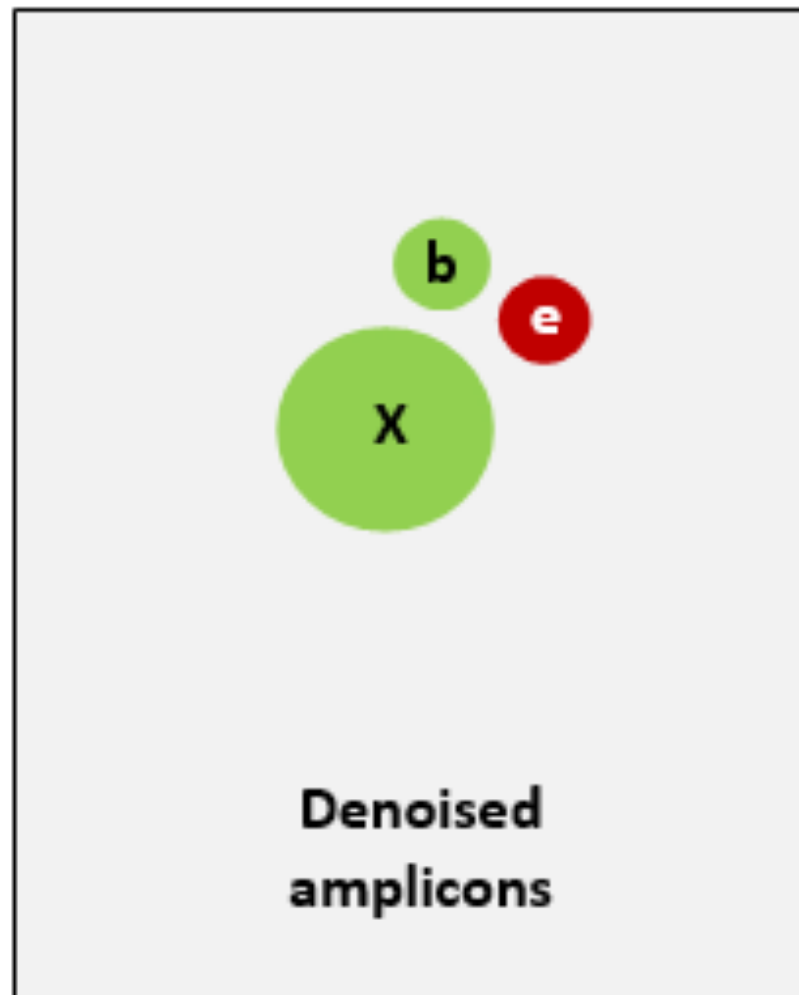
UPARSE



Cluster OTUs

VS

Denoise



unoise3

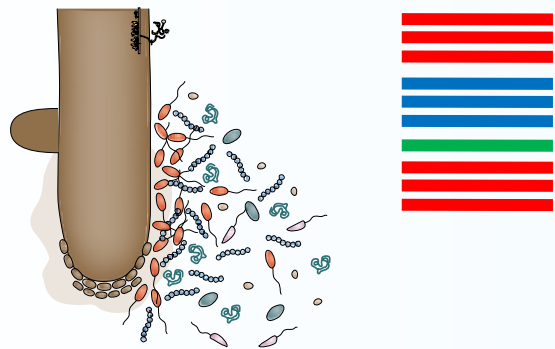
deblur

dada2

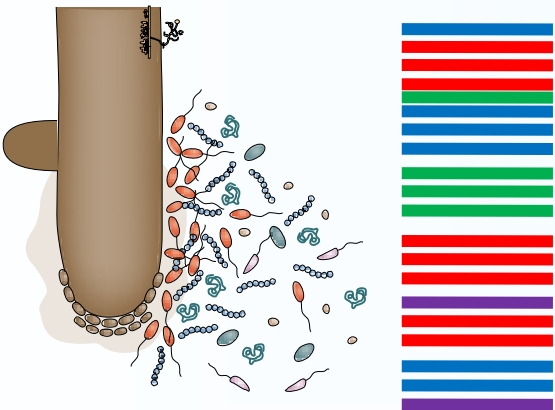


# 示例说明扩增子分析原理

SampleA PCR后挑单克隆测序



SampleB



统计序列特征表(OTU/ASV表)

	SampleA	SampleB
BacRed	6	8
BacGreen	1	4
BacBlue	3	6
BacPurple	0	2

等量重抽样：比较物种多样性

	SampleA	SampleB
BacRed	6	4
BacGreen	1	2
BacBlue	3	3
BacPurple	0	1

多样性指数：A的丰富度为3，而B为4

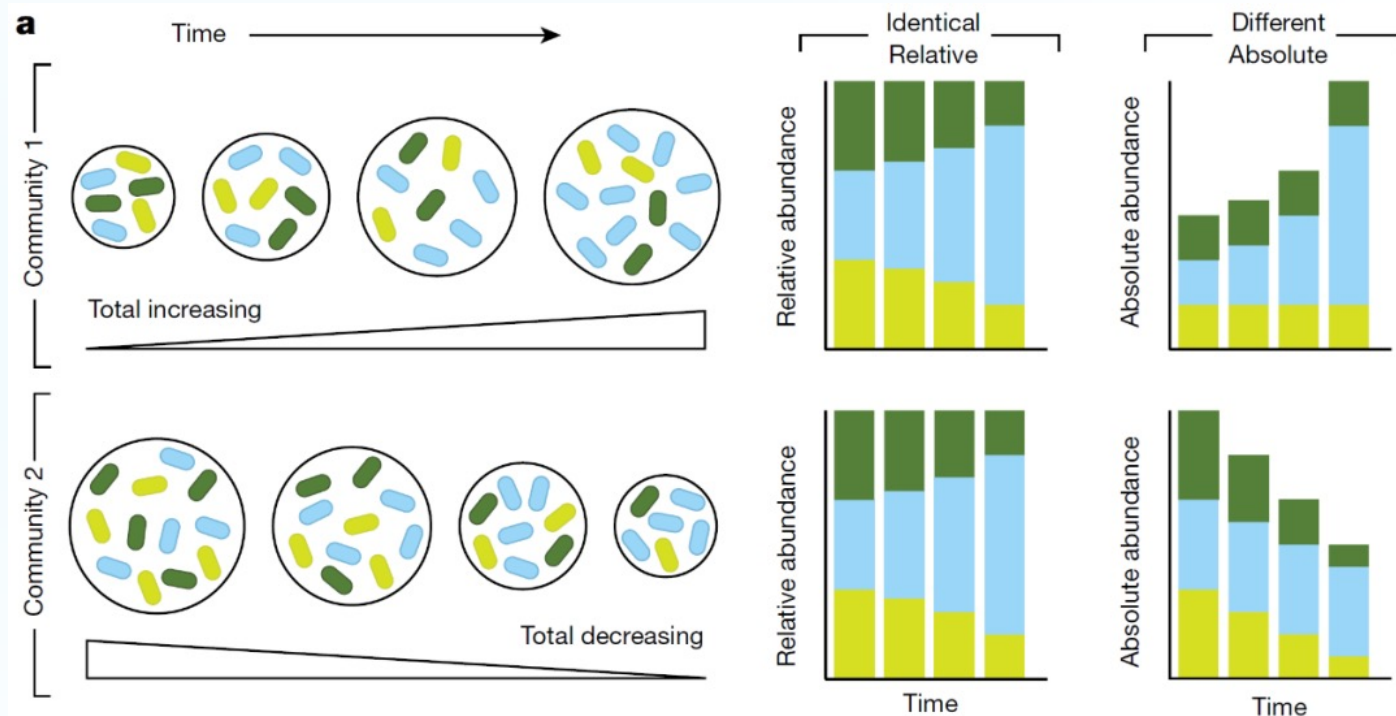
相对丰度：比较相比比例多少

	SampleA	SampleB
BacRed	60%	40%
BacGreen	10%	20%
BacBlue	30%	30%
BacPurple	0	10%

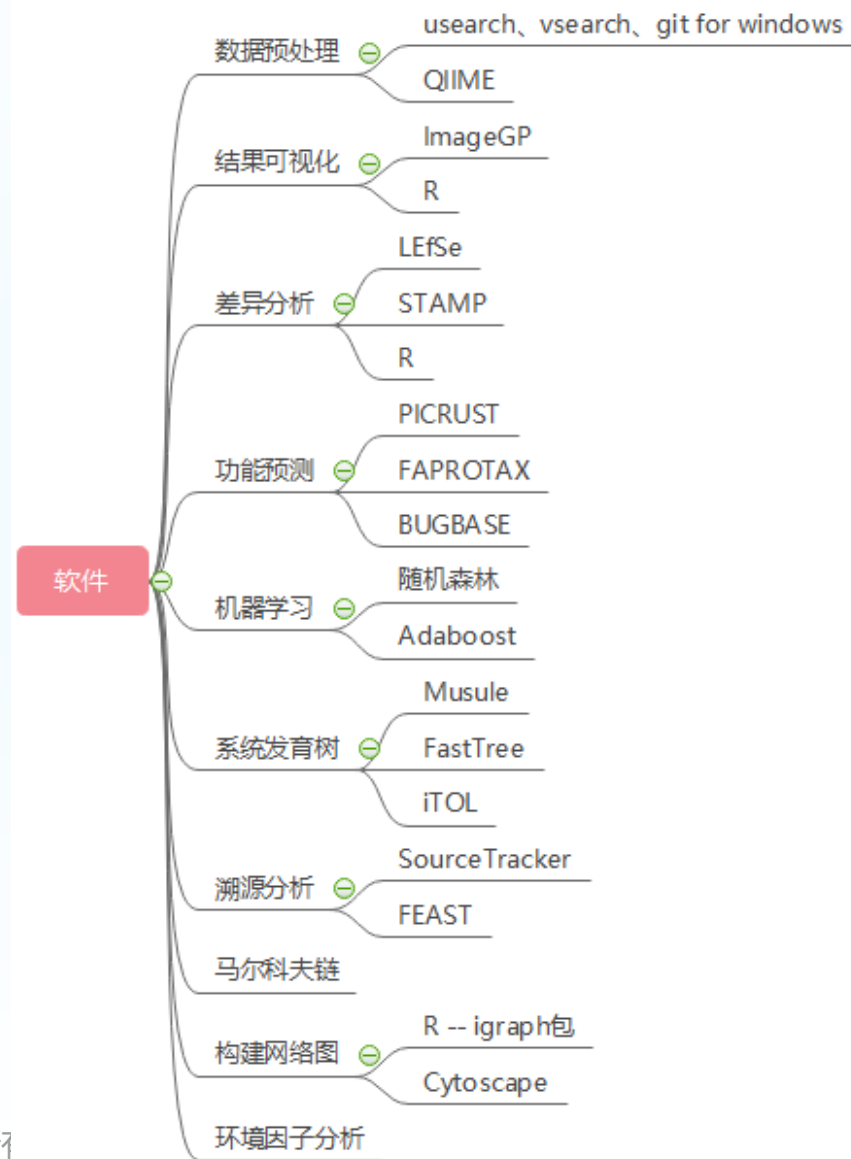
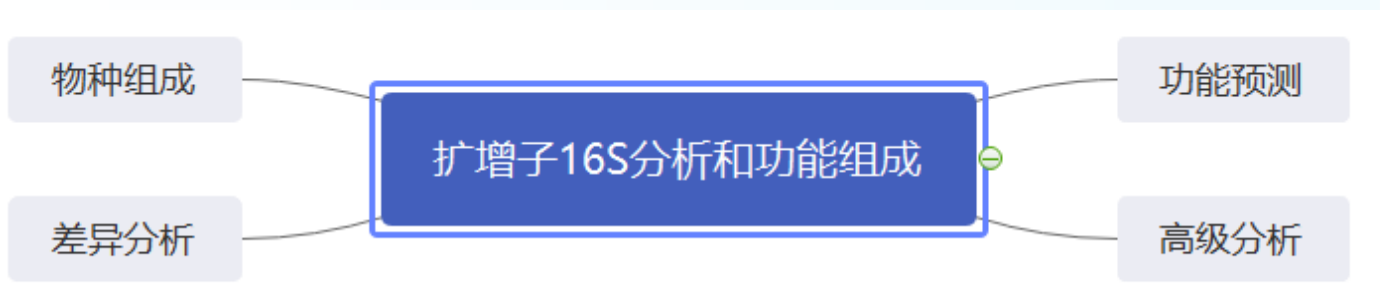
样品间菌相关丰度存在差异

# 相对定量的局限性

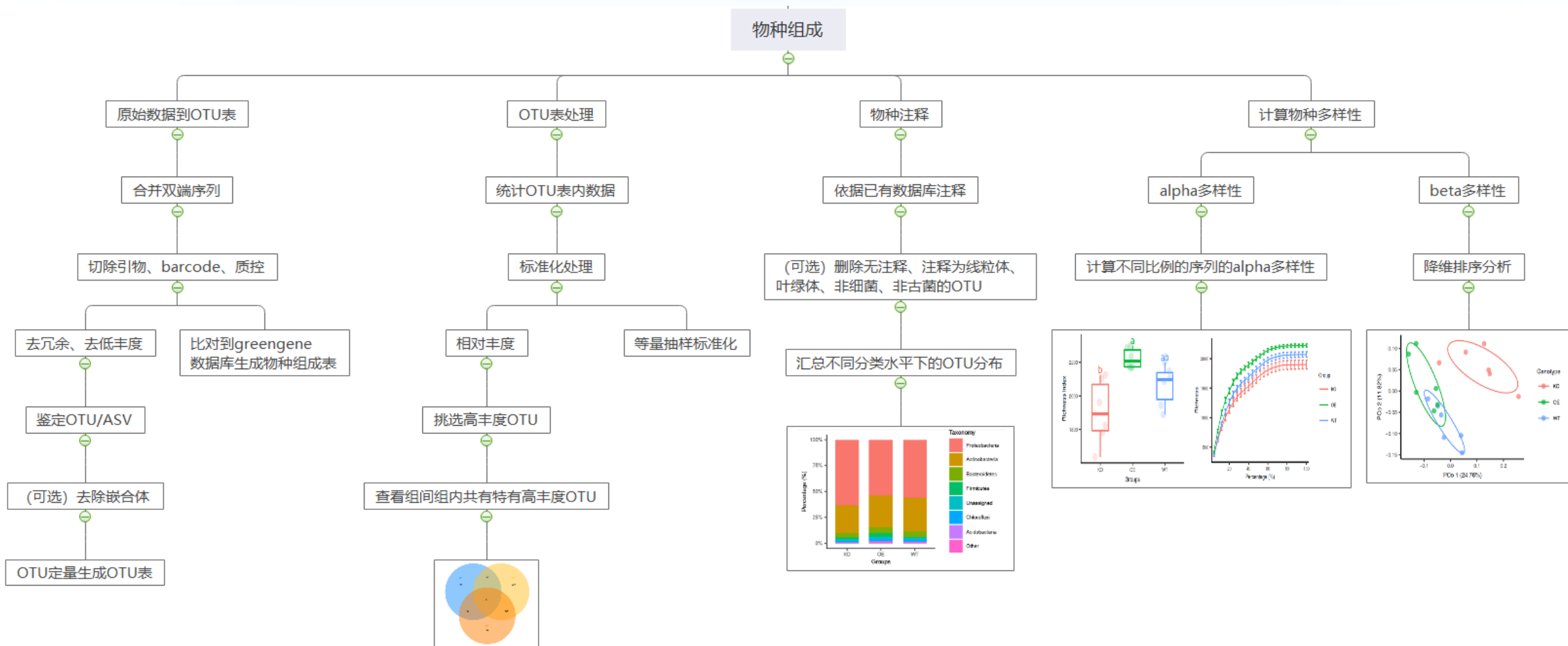
- 内参法
- 16S rRNA拷贝数法: 如果给定的分类单元占样品中16S rRNA基因区段的1%, 则将该1% 乘以16S rRNA 基因扩增子的总数 (例如来自使用相同引物、相同DNA样本量和 PCR循环数的qPCR方法), 可以提供该分类群绝对丰度的估计值。
- 已知单拷贝基因如rpoB的拷贝数法。



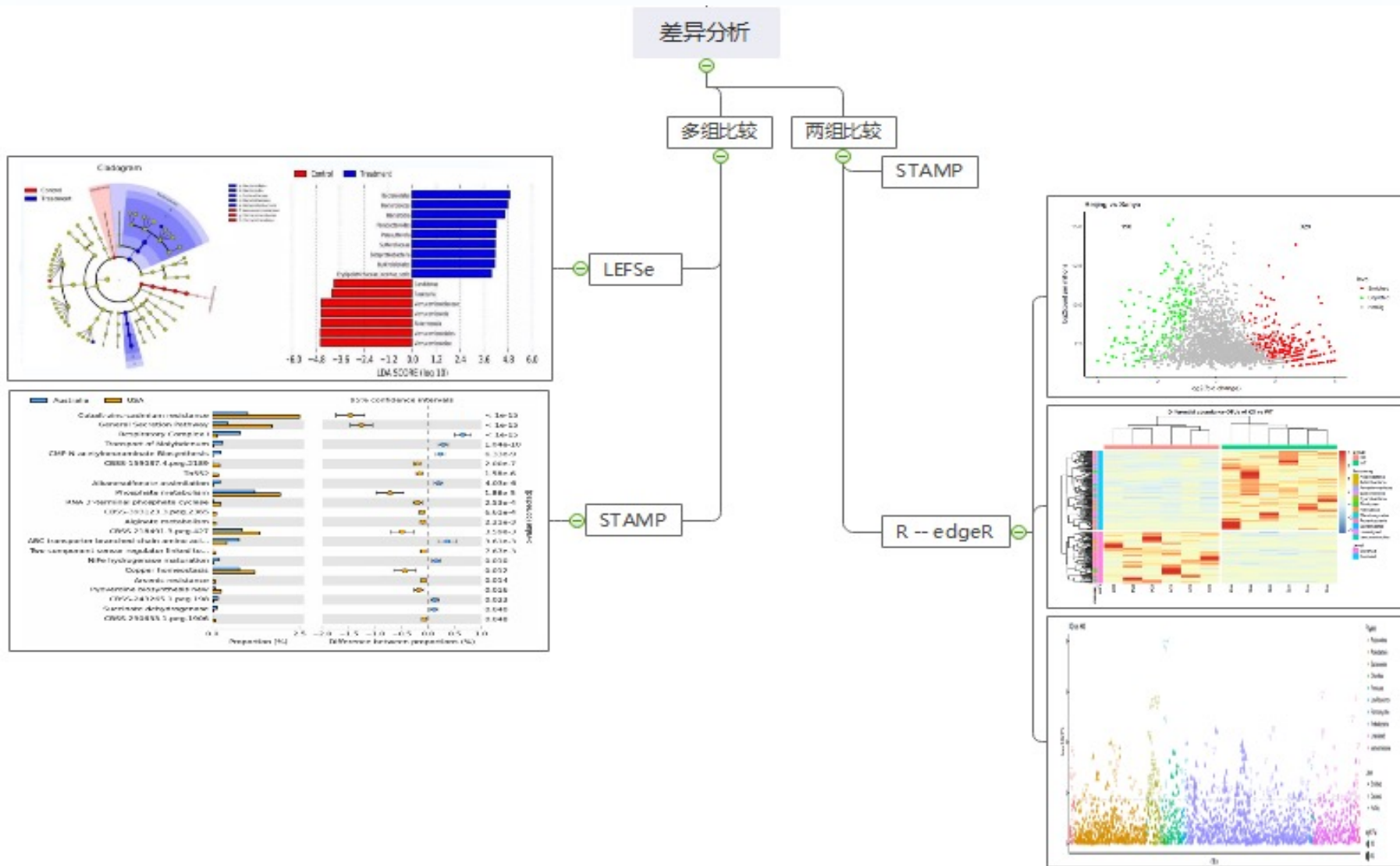
易生信







# 差异分析



基因组

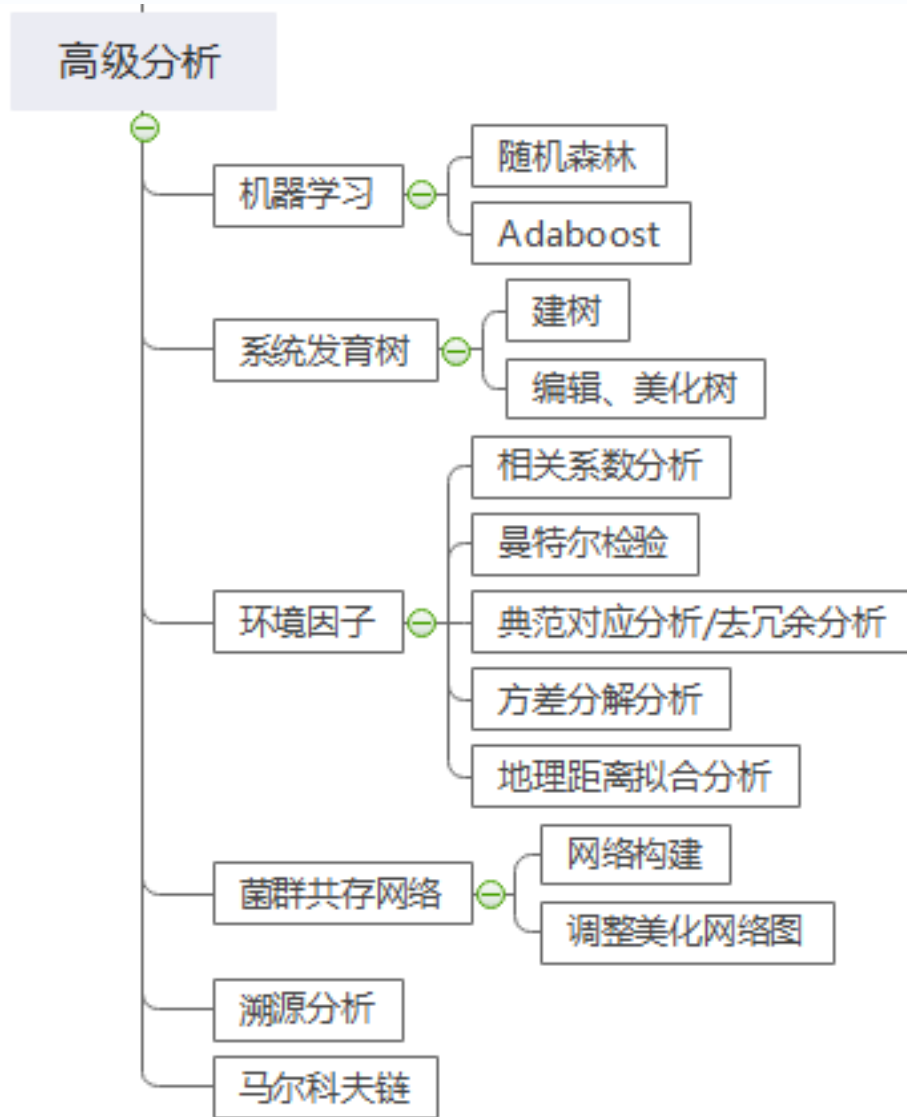


## 功能预测

PICRUST: 对菌群代谢功能进行预测

FAPROTAX: 预测微生物群落功能

BUGBASE: 预测微生物表型



易生信，毕生缘；培训版权所有。

易生信 生信宝典 宏基因组





# 进一步阅读

- [扩增子分析01-背景介绍](#)
- [扩增子分析02-真菌引物选择](#)
- [宏基因组分析教程2扩增子-微生物群落多样性](#)
- [扩增子图表解读-理解文章思路](#)
- [扩增子分析流程-把握分析细节](#)
- [扩增子统计绘图-冲击高分文章](#)
- [QIIME2中文教程-把握分析趋势](#)
- [16S信息分析流程软件和数据库合集](#)

易生信  
生信宝典  
宏基因组





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

