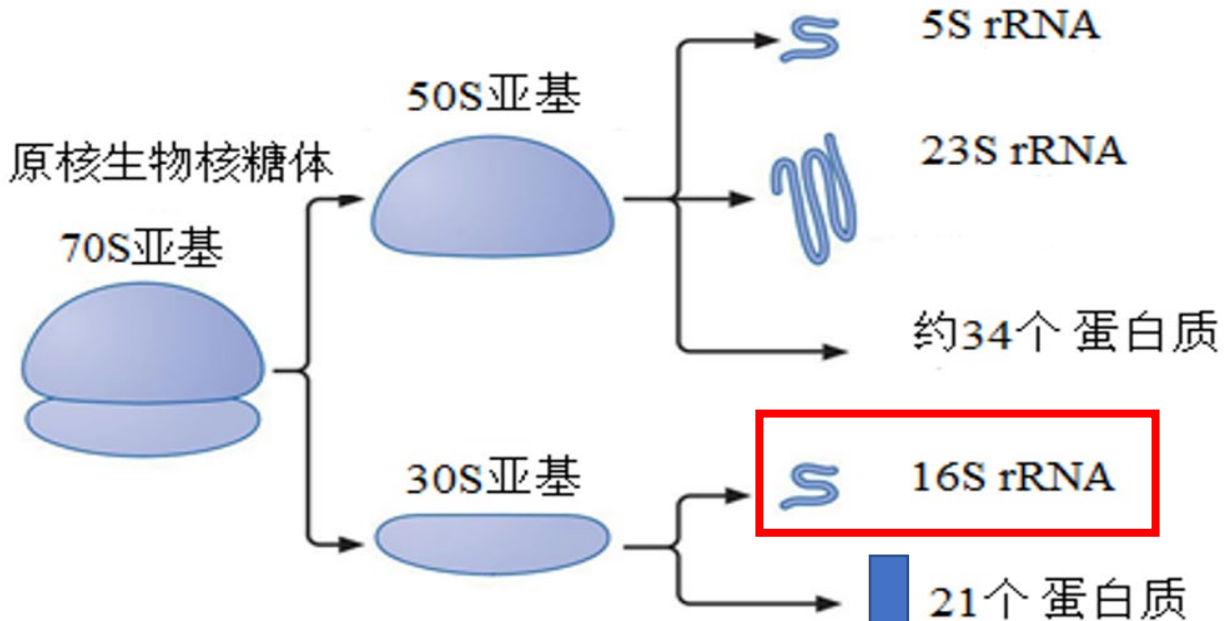
北京大学
PEKING UNIVERSITY

基于16s测序技术解析
肠易激综合征和健康人群的肠道微生物差异

汇报人：陈俞竹

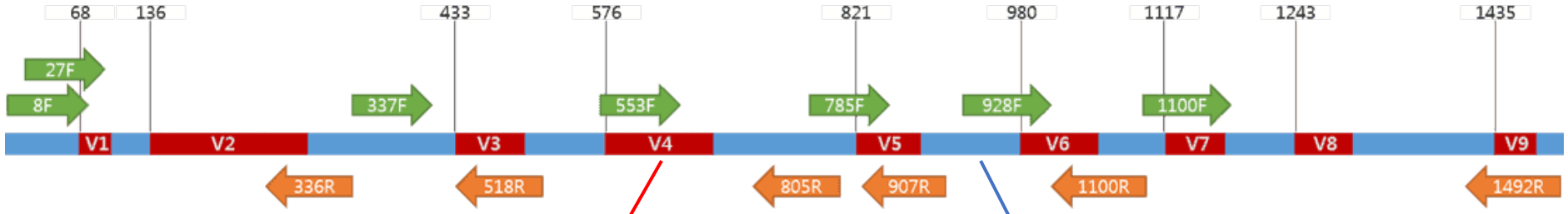
小组成员：齐俊添　张欣玥　兰天龙

16s rDNA



DNAsegment coding for rRNA is called either rRNA gene or **rDNA**

16S rDNA

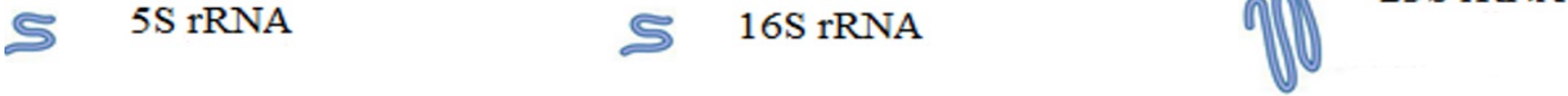


可变区　　保守区

16s rDNA可作为分子钟的原因

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。

- 16S rDNA的长度适宜



太短，信息量不够

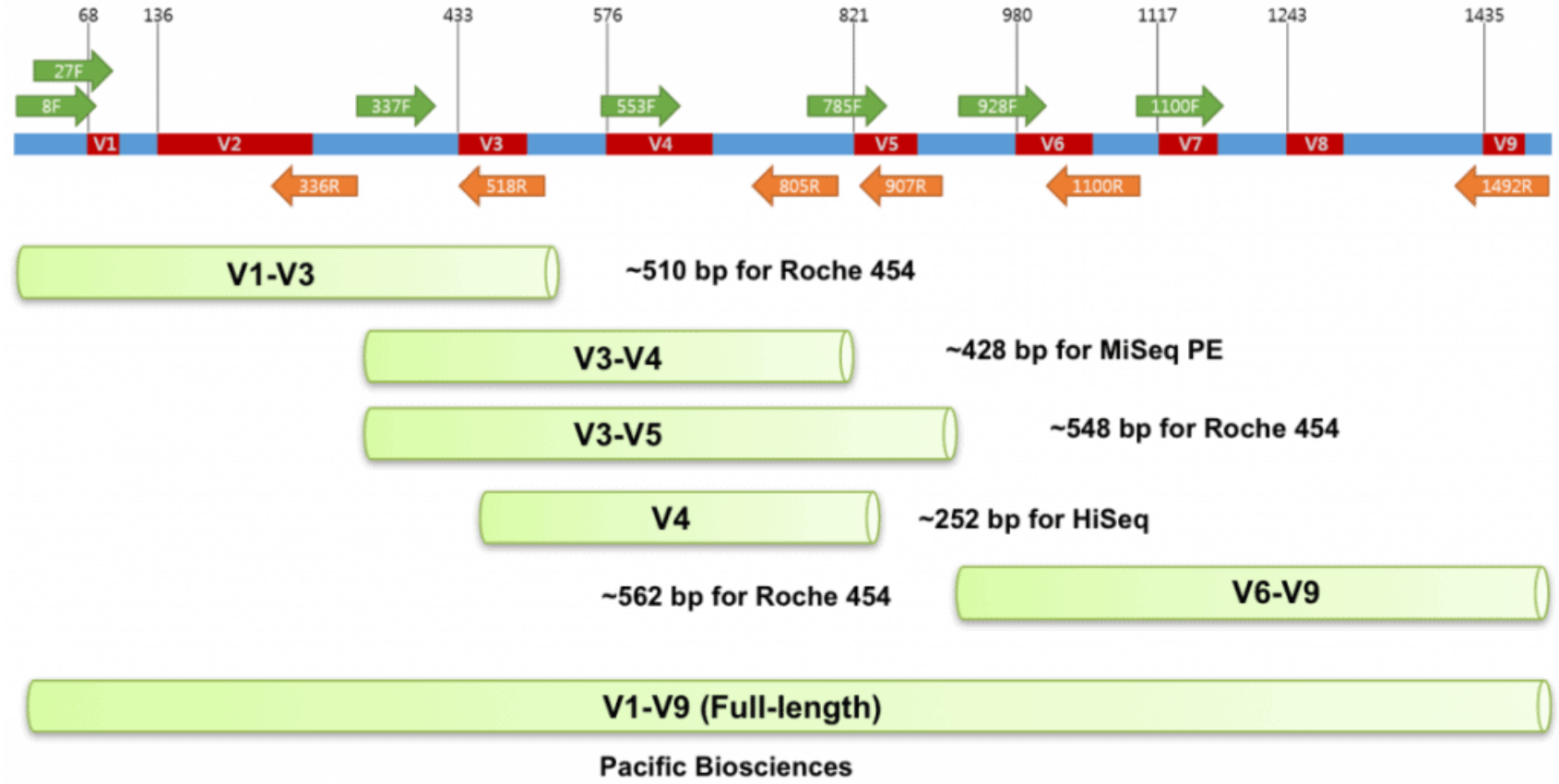
长度适中，信息量大且容易分析

太长，分析困难

- 16S rDNA具有良好的进化保守性

可变区的选择

原核16S rRNA没有任何一个可变区可以准确的将所有种类的细菌从域到种进行明确分类。不同的V区会对原核微生物群落结构的分析结果产生明显的影响。



- V3-V4（长度464bp），illumina Miseq平台（PE300），该平台可完整覆盖该区域，对细菌的覆盖率较高。
- V4-V5（长度~303bp），illumina Miseq平台（PE250），其特异性较好，具有很好“捕捉”细菌多样性的能力，数据库信息全，是细菌多样性分析注释的最佳选择。

16s rDNA的缺点

北京大学 PEKING UNIVERSITY

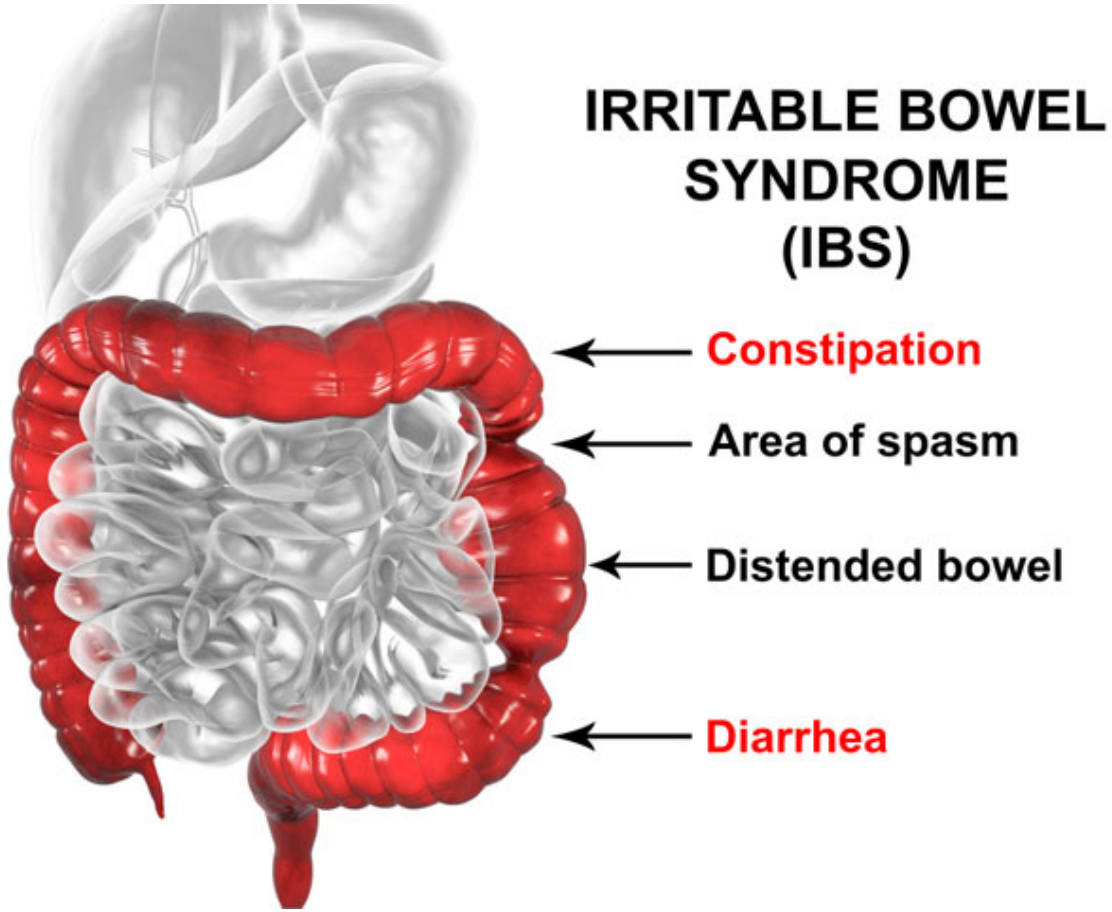
由于16S rDNA测序是通过扩增某个或某几个高变区来检测，一般可精确到“属”水平，少数可鉴定到“种”。对于某些菌，高变区中序列的相似度非常高；或者区分不同菌的序列片段不在我们的扩增区域内，均会导致无法鉴定到“种”。

IBS

肠易激综合征(irritable bowel syndrome，IBS) 是一种临床常见的功能性肠病，最主要的临床表现为持续或间歇发作的腹痛或腹部不适及排便习惯和粪便性状的改变。

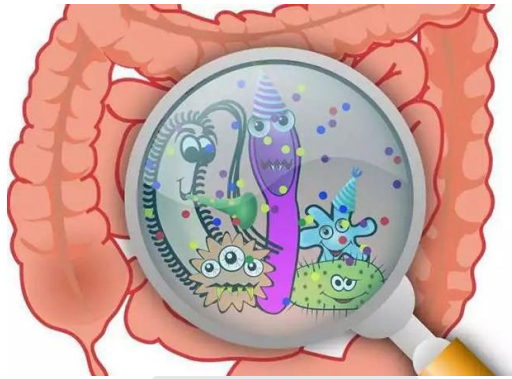
该病发病原因和机制尚未完全明确，目前有以下几个可能的原因：

- 胃肠运动障碍；
- 内脏的高敏感性；
- 脑肠轴调节异常；
- 肠道感染；
- **肠道菌群失调；**
- 遗传因素、性别、饮食及心理社会因素等

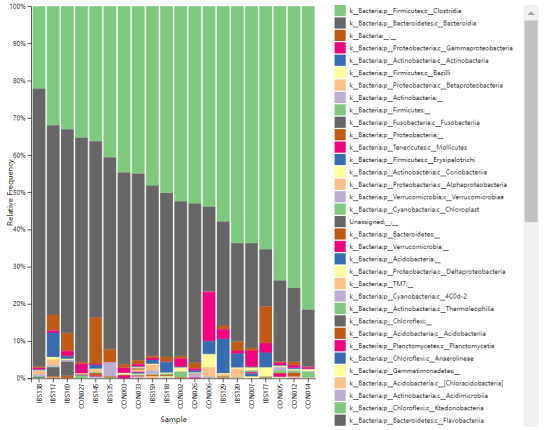


菌群失调

人体肠道内栖息着约 400 种微生物，其中双歧杆菌、乳酸杆菌、链球菌等有益菌占 98%以上，另外 2%是处于中间状态的细菌和有害菌。当肠道菌群失调时，致病菌占了优势，就会引发各种肠道疾病。



生信分析通用流程



样本测序数据以及元数据

数据库

切除引物与质控

序列去冗余

特征挑选

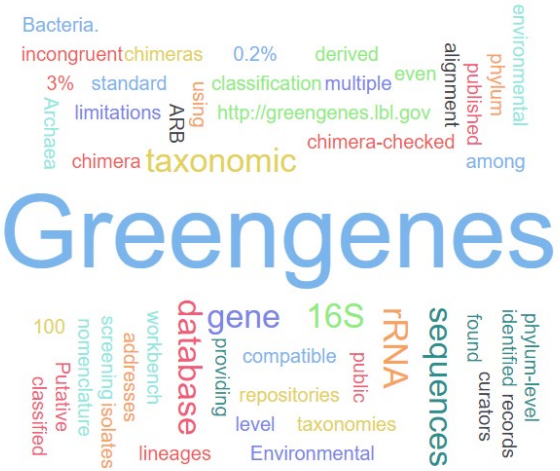
嵌合体检测

OTU聚类及物种注释

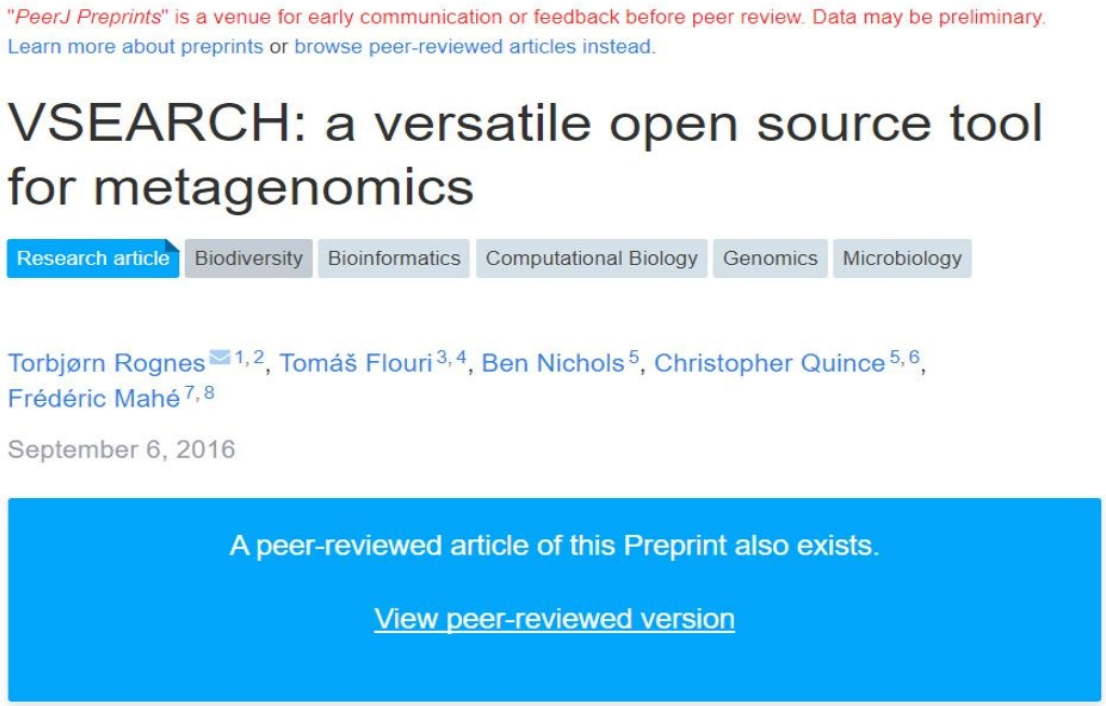
alpha多样性分析

beta多样性分析

差异分析



VSEARCH



序列双端合并

去除两端接头

Fastq质量检测

序列去重复

嵌合体检测

OTU聚类

分类信息注释

优点：

矢量化---精确

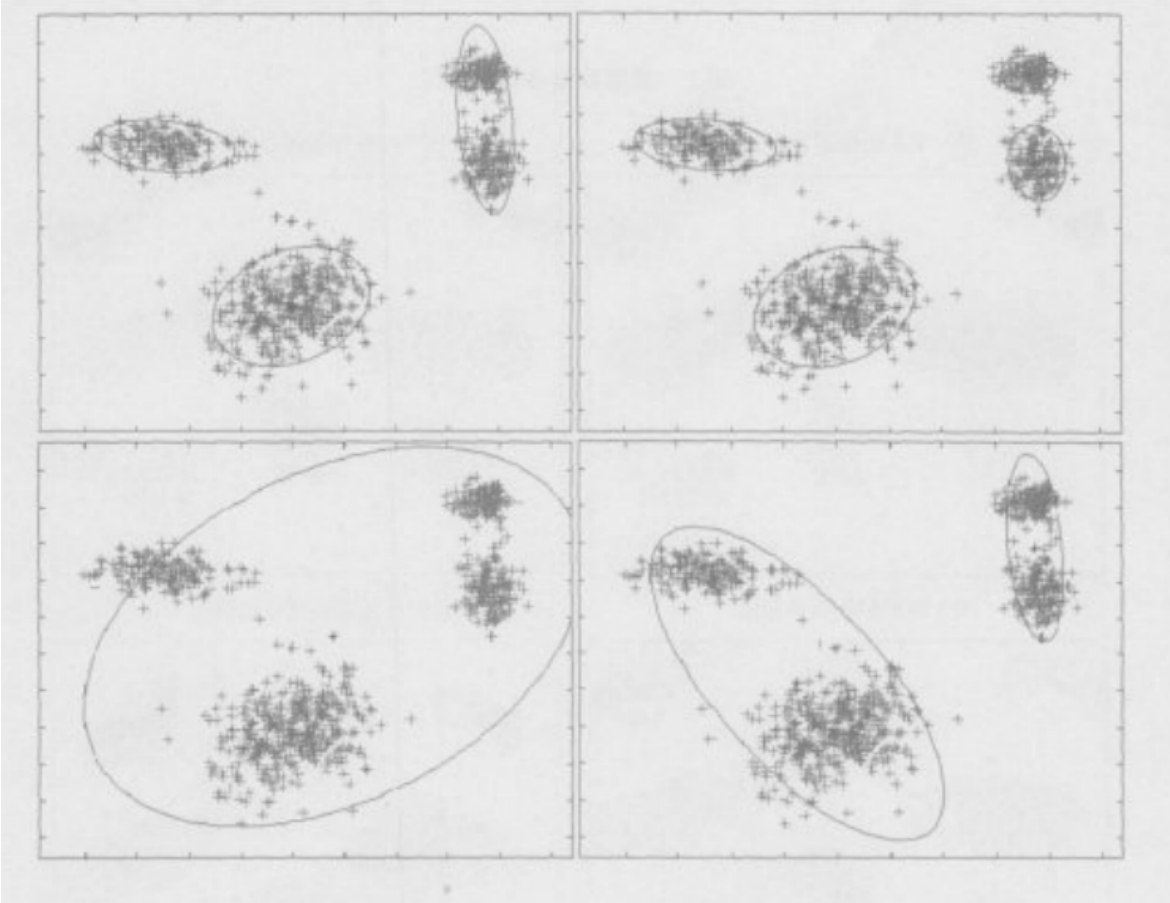
并行-----高速

序列比对算法：

vsearch	usearch
Needleman-Wunsch	heuristic seed and extend aligner

VSEARCH

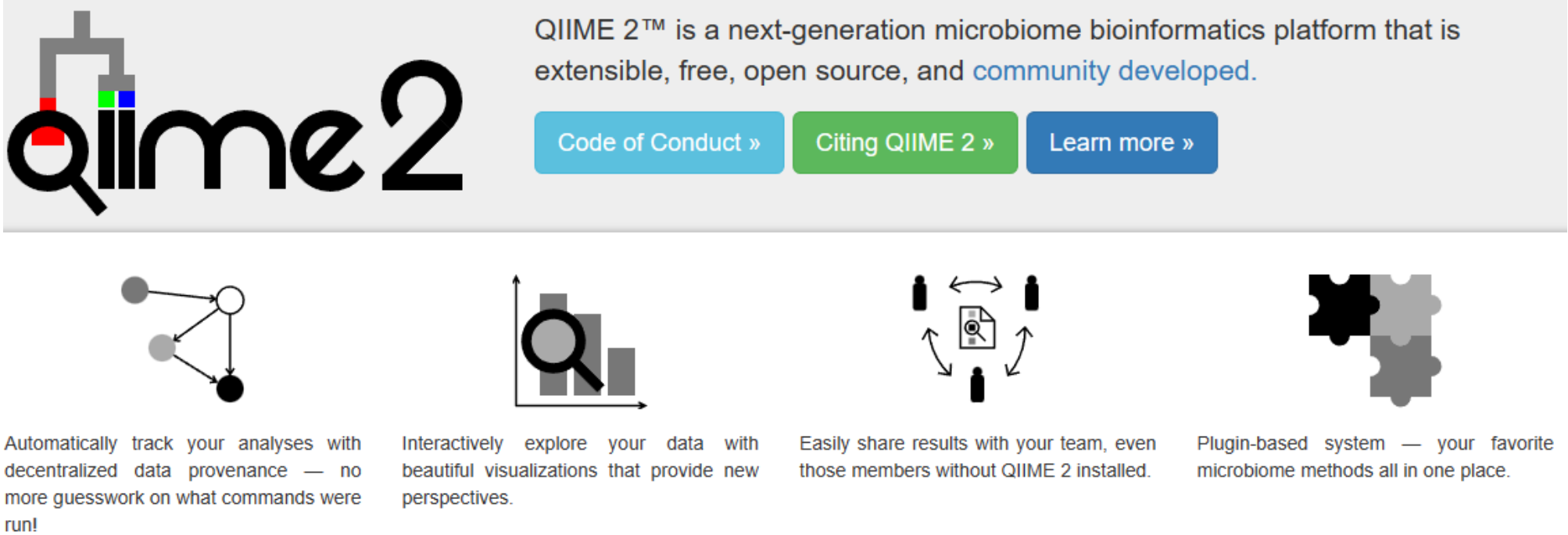
- 单路、贪心中心聚类算法
- 局部最优解，并非全局最优解
- 影响物种分类的显著性
- 结果误差来源之一



QIIME 2

QIIME 2是一款强大、可扩展的微生物组分析平台，

强调数据分析透明，但实践体验中使用步骤较为繁琐。



Evan, et al., *Nature Biotechnology,* 2019

QIIME 2

北京大学 PEKING UNIVERSITY

blast

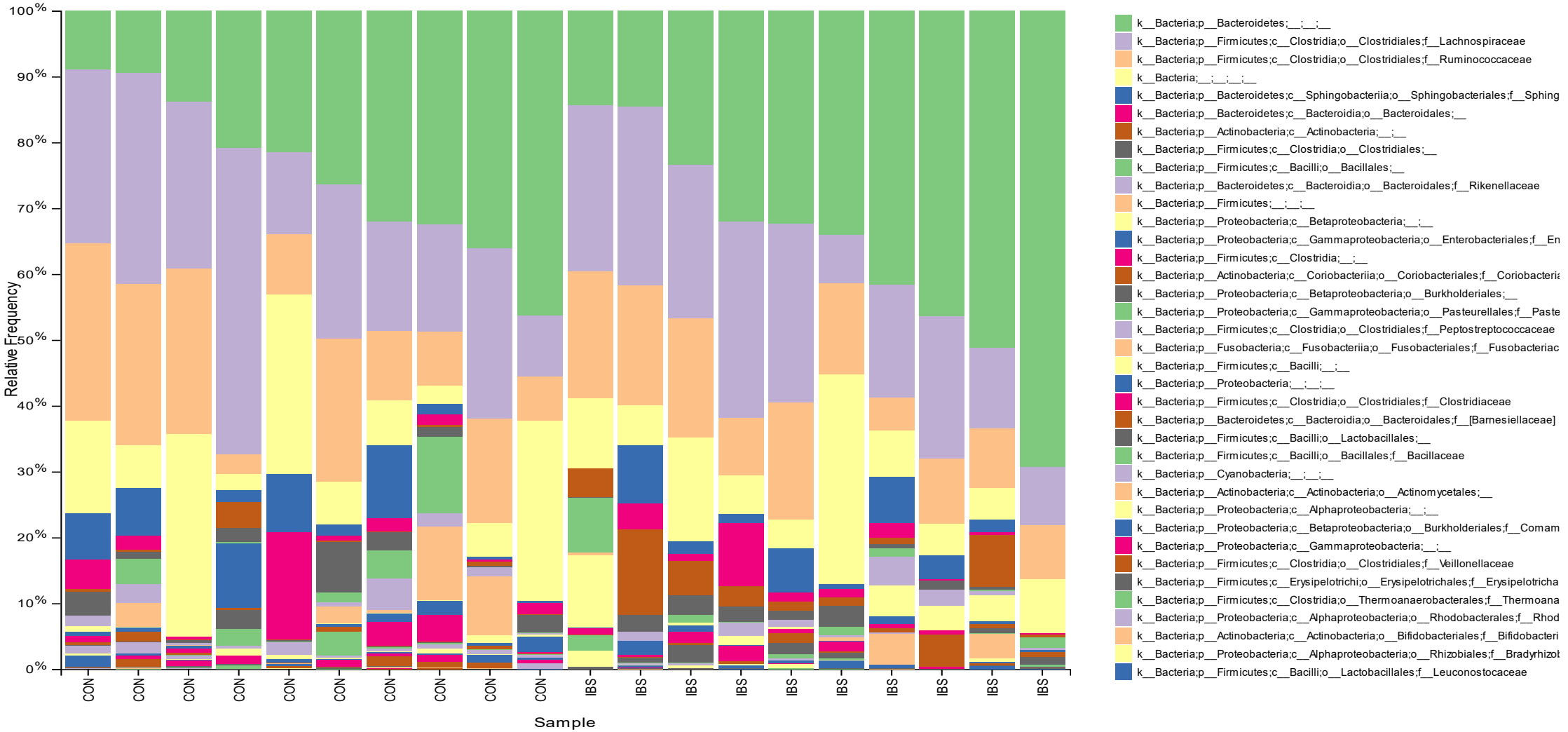
vsearch

基于序列对齐的方法，在比对结果找到合适的注释信息

classify-sklearn

基于数据训练产生**机器学习（朴素贝叶斯机器学习）**分类器，使用分类器进行注释。

QIIME 2



小样本量结果展示

两种序列比对方法的对比

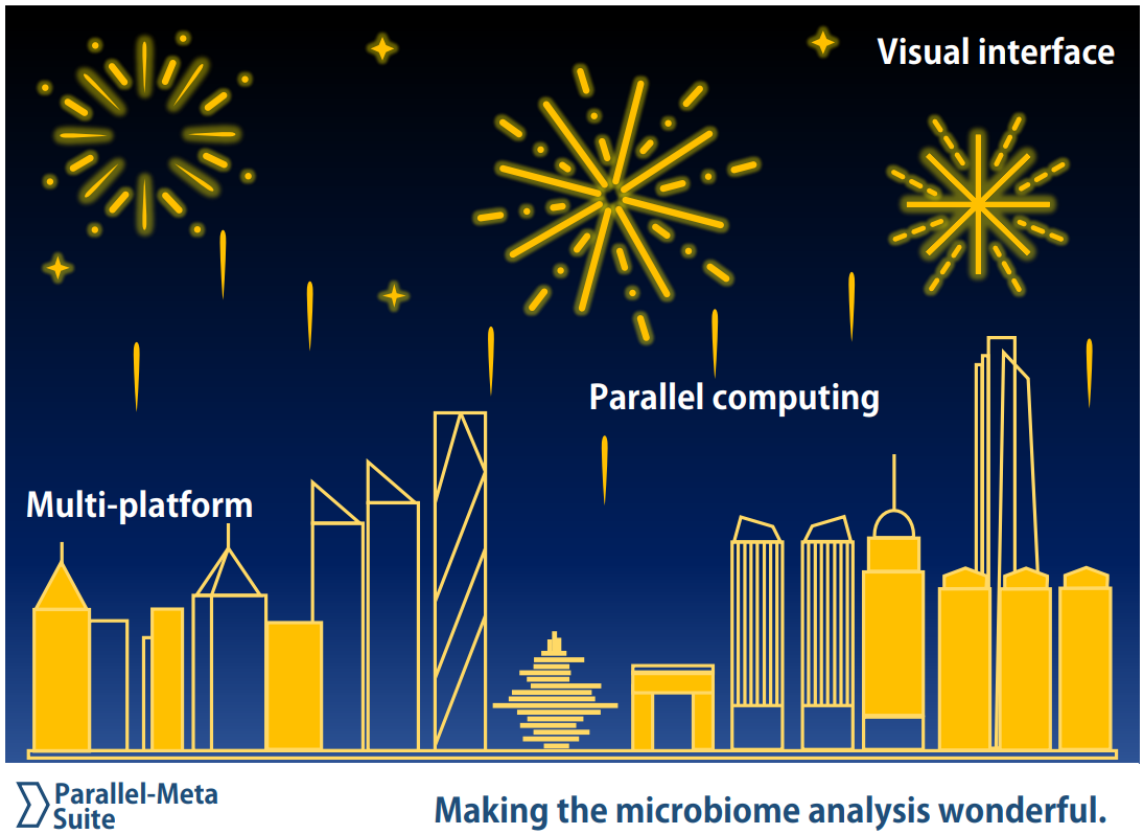
北京大学 PEKING UNIVERSITY

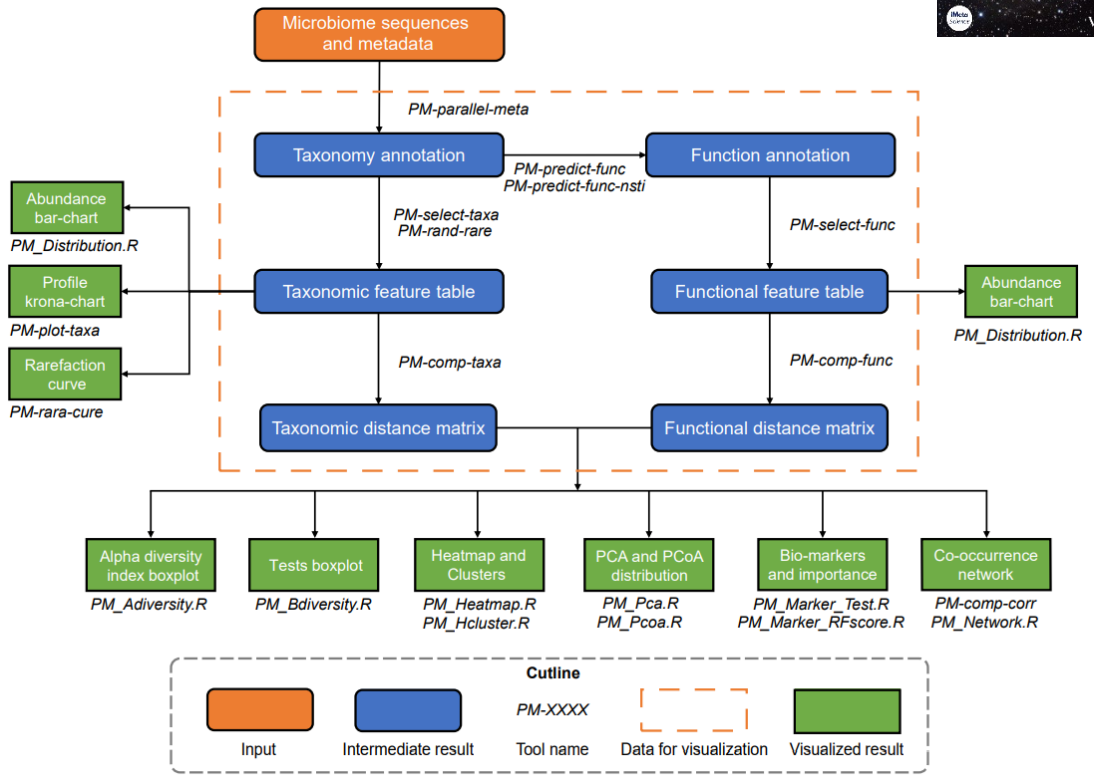
	vsearch	classify-sklearn
优点	应用较为广泛，结果可重复性高。	充分训练的分类器在16S rRNA基因和真菌ITS序列物种注释的精确度和严谨性方面一般优于其他的标准分类方法
缺点	比对后产生多条比对结果，常选取得分最高、e值最小的一条为目标序列的比对结果。有可能出现“假阳性”的注释结果，无法反映群落的真实特征。	依赖分类器的训练，不同分类器（类群、扩增引物和序列读长不同等）注释的结果可能存在差异。

分析工具

北京大学 PEKING UNIVERSITY

Parallel-Meta Suite：跨平台、可交互的微生物组快速分析工具





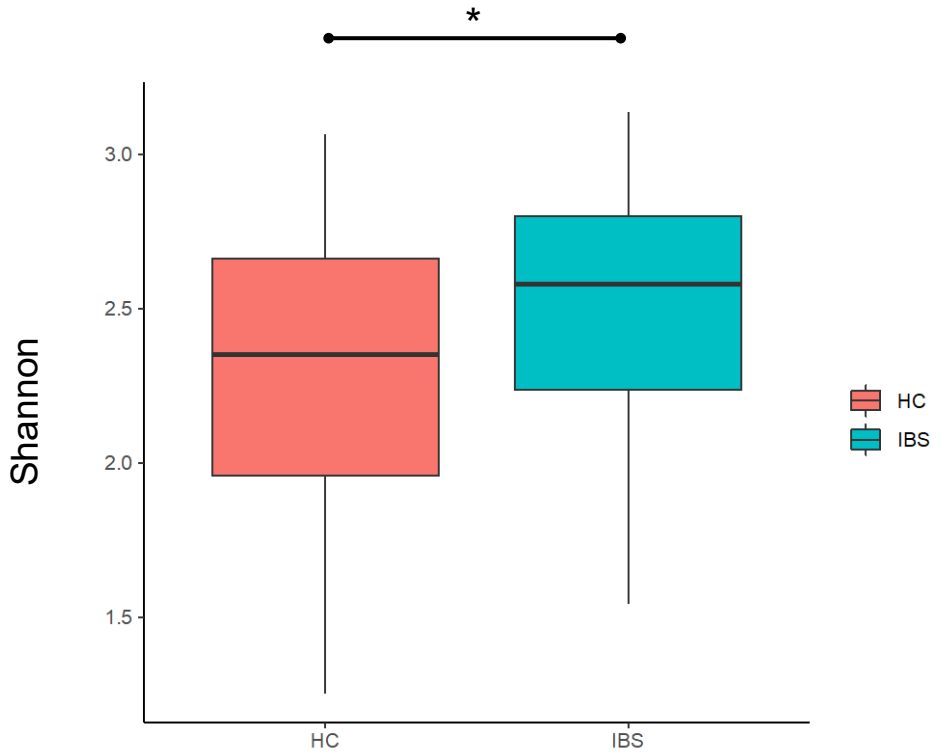
PMS使用了各种最先进的算法和分析策略，实现了从DNA序列到可视化结果的完整流程

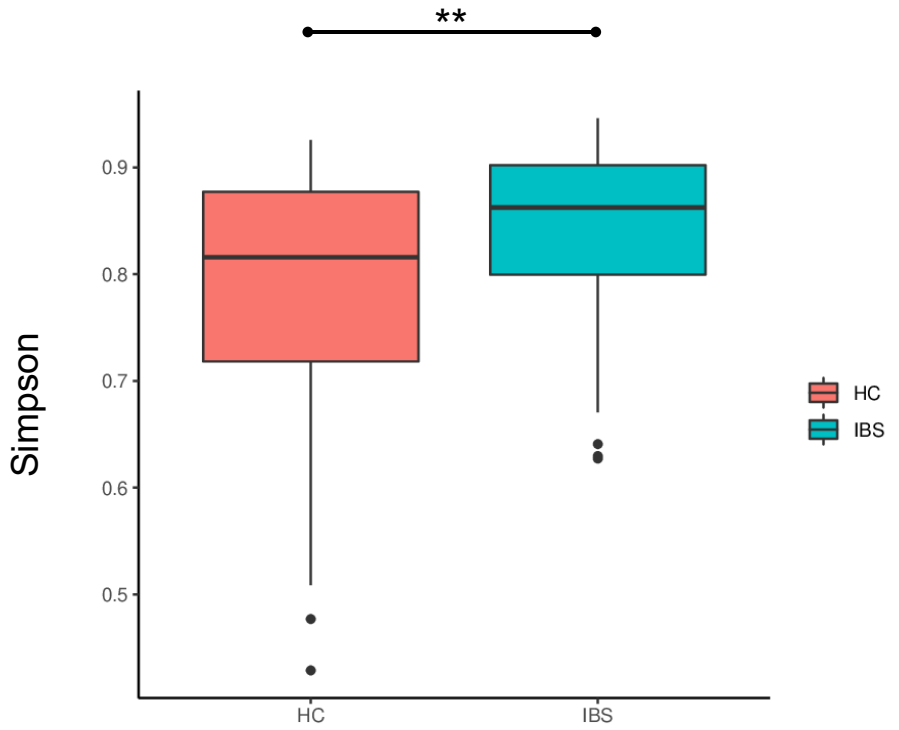
Chen, et al., *iMeta,* 2021

群落多样性 - alpha多样性

北京大学 PEKING UNIVERSITY

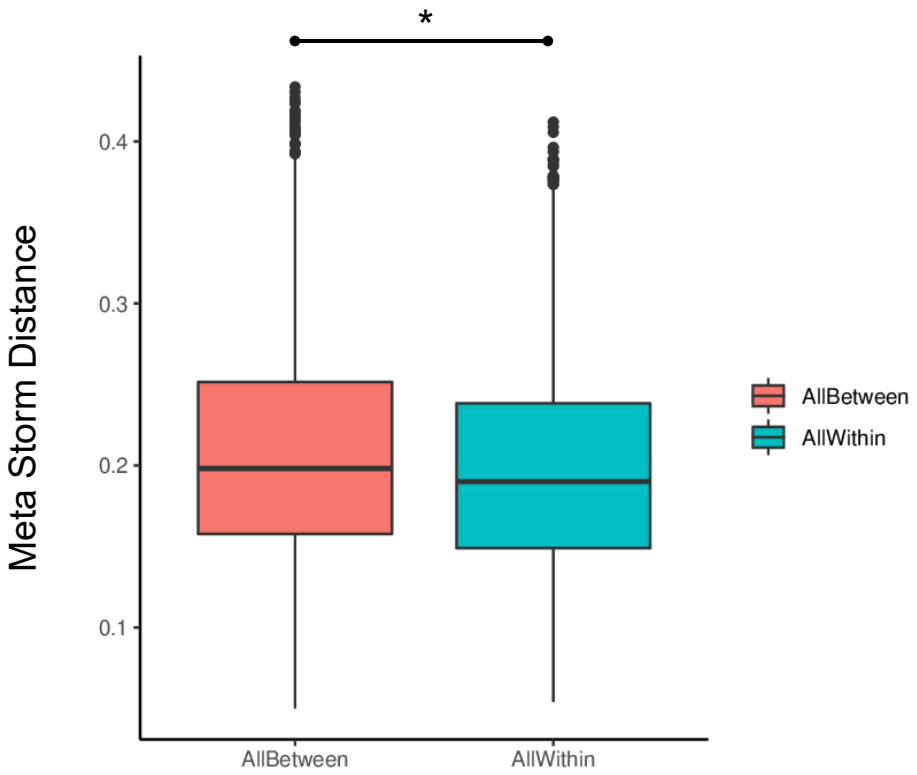
* Wilcox-test $p < 0.05$
** Wilcox-test $p < 0.01$



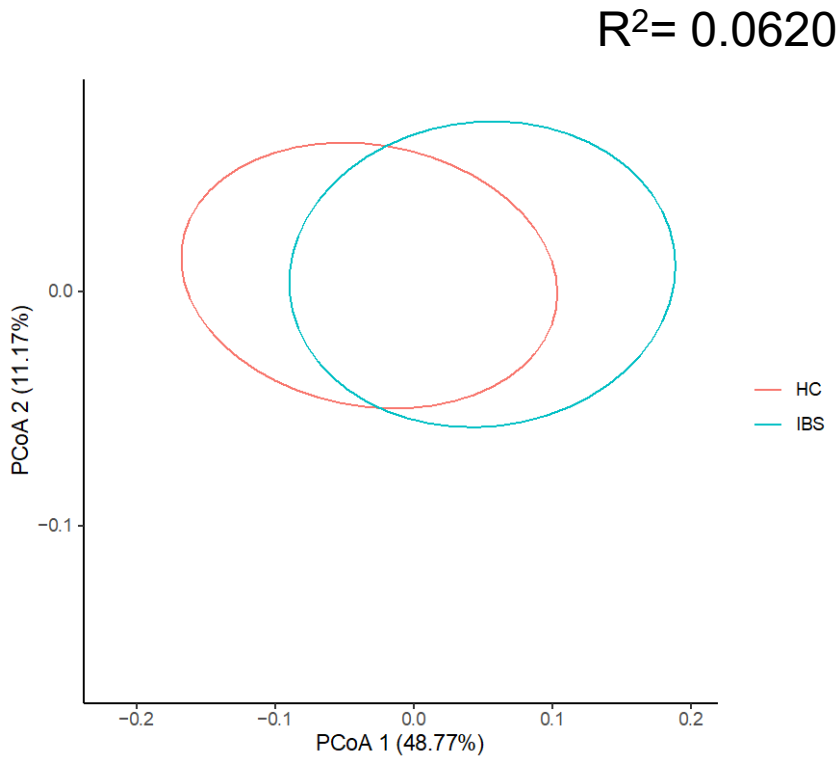


- α 多样性分析显示：HC、IBS组Shannon指数和Simpson指数差异有统计学意义(p<0.05)
- IBS组与健康对照组相比，群落多样性更高

群落多样性 - beta多样性



* Wilcox-test $p < 0.05$



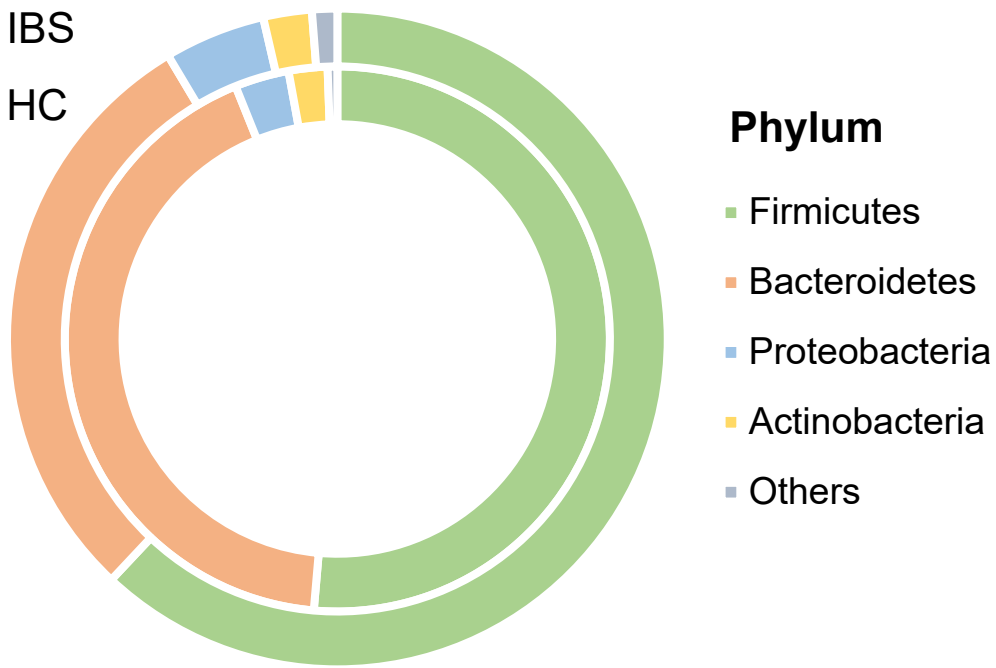
PERMANOVA $p<0.001$

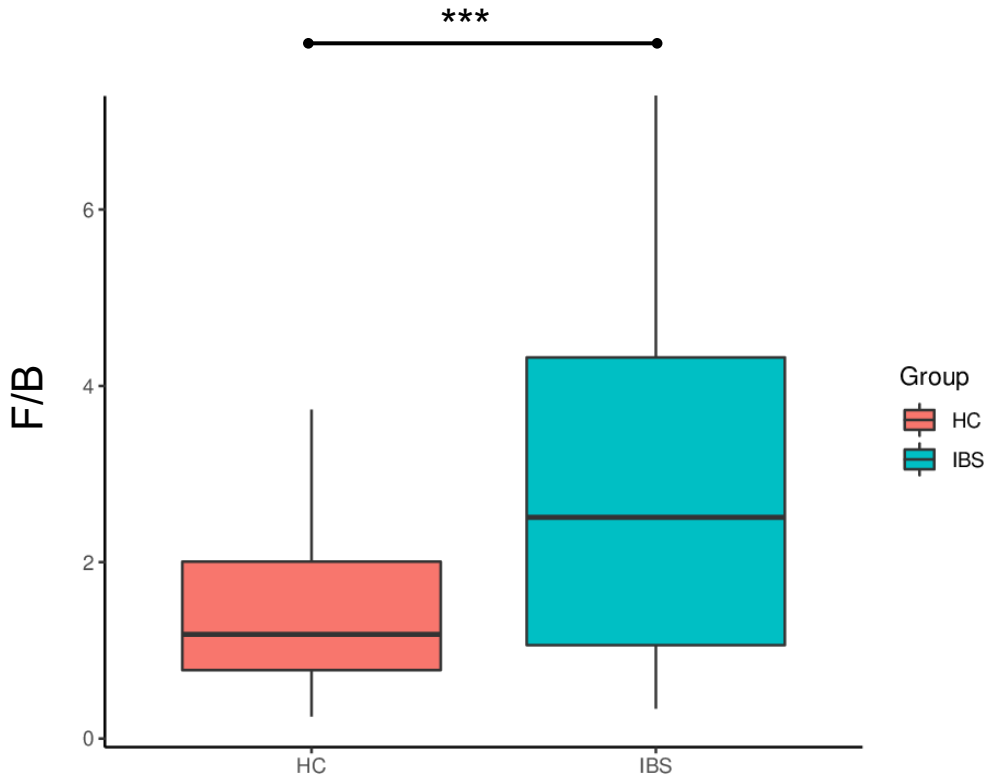
- 组间距离大于组内距离，健康状态（HC或IBS）会显著影响菌群群落结构变化（$p<0.05$）
- PCoA分析可见，肠道菌群能够显著区分不同健康状态（$p<0.001$）

厚壁菌门与拟杆菌门的比例

北京大学 PEKING UNIVERSITY

*** Wilcox-test $p < 0.005$





IBS组厚壁菌门(Firmicutes)丰度显著增高，拟杆菌门(Bacteroidetes)丰度显著降低，

F/B与健康对照组相比显著升高，与其他以往IBS研究结果相同

IBS治疗方法

- **微生态制剂**

  可纠正肠道菌群失调，对腹胀、腹泻等症状均有良好的疗效

- **粪菌移植**

  帮助纠正紊乱的肠道菌群和重建正常的肠道微生态

- **三氧疗法**

  三氧又称臭氧，为广谱有效的强杀菌剂，对包括大肠杆菌在内的各种病原菌均有较好的杀灭效能

如何找数据？

北京大学 PEKING UNIVERSITY

Applied Bioinformatics Course

ABC | People | Docs | Databases | Tools | PDB | UniProt | ExPASy | EBI | NCBI | BIGD

Welcome to ABC - the Applied Bioinformatics Course!

As inspired by Walter Gilbert's 1991 Nature article *"We must hook our individual computers into the worldwide network that gives us access to daily changes in the database and also makes immediate our communications with each other"*, this course is designed for wet-lab experimental graduate students in biology.

What we learn

We'll learn, step by step, the ABCs of:

- How to find literature papers from PubMed efficiently
- How to search databases such as UniProt and RefSeq in an Advance mode
- How to align your DNA and protein sequences
- How to make a good Blast search to obtain your results with less false positive and false negative
- How to analyze your own DNA or protein sequences with various tools
- How to construct a phylogenetic tree for a set of sequences at your hand
- How to predict the three dimensional structure of your favorite protein

And lots more!

How we learn

We will run the course in a training room. Each student will have a PC connected into the Internet.

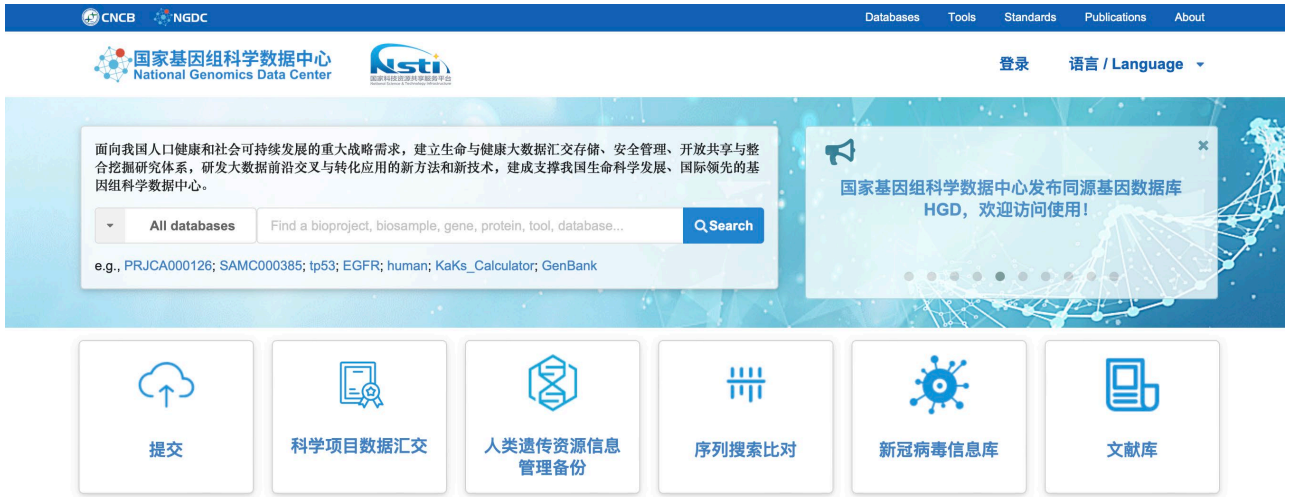
We start with introducing the international bioinformatics resources around the world, for example, NCBI and EBI. We then use the bioinformatics platform to do hands-on practice. We will do a lot of practice for sequence alignment, database similarity search, motif finding, gene prediction, as well as phylogenetic tree construction and molecular modeling.

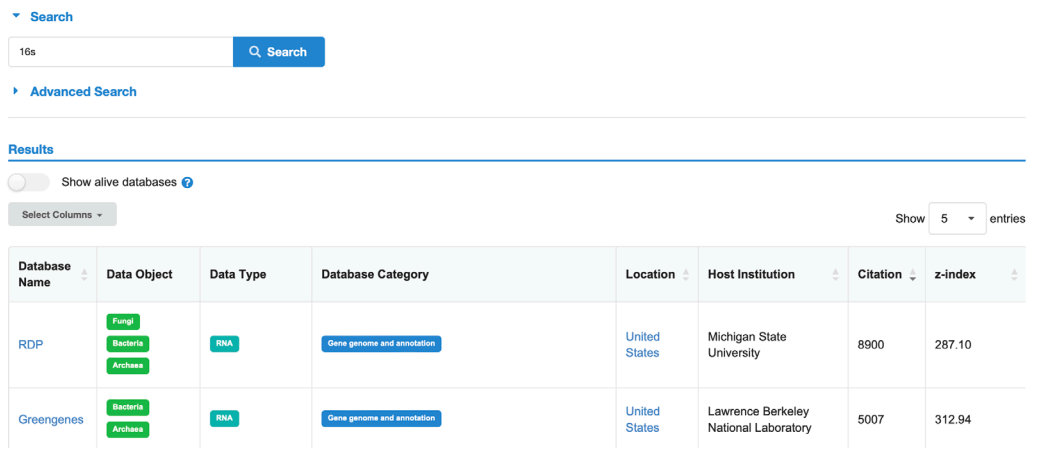
Finally, we will focus on several projects to solve real biological problems. You are encouraged to bring your own problems to discuss and, hopefully, to solve during the course!

You may read a brief introduction and the outline (in Chinese) of the course, and read the article Teaching the ABC of Bioinformatics (in English), or ABC Examples (in Chinese) for more details.

What you need before the course

- A desktop PC or laptop hooked to the Internet. You may also use iPad or mobile phone to browse most of the pages.
- A good background of biochemistry and molecular biology - you may try the Biotest_Cn (in Chinese) or BioTest_En (in English) to see how good at it you are.
- An ability to read the English text such as this page, and understand the meaning.





总结

北京大学 PEKING UNIVERSITY

1. 通过PubMed查阅相关文献了解到
    a. 16SrDNA能够做为分子钟的优点和缺点；
    b. IBS疾病可能产生的原因和解决方法；
2. 通过课上了解到Needleman-Wunsch、BLAST等序列比对算法，在该课题分析过程中，更加深入研究并比较了基于序列比对和基于机器学习物种组成分析算法。
    a. 分析中采用的vsearch使用矢量并行算法，数据处理过程高速执行，比对结果精确，采用Needleman-Wunsch全局比对，召回率高；
    b. 基于机器学习的物种注释，精确度和严谨性方面一般优于其他的标准分类方法，但是特别依赖分类器的训练；
3. 分析IBS患者与健康对照的肠道菌群发现，IBS组群落多样性更高，不同健康状态（IBS or HC）会显著影响肠道菌群群落结构组成；两组间的特有菌及共有差异菌分析显示，IBS组独有或共有富集的菌属普遍与腹痛腹泻、肠炎病史等相关，与疾病症状谱相符；
4. 除了自己收集样本外，也可以利用课上讲过的BIGD(http://bigd.big.ac.cn)来搜索公共数据库，查找符合要求的数据。

北京大学
PEKING UNIVERSITY

欢迎大家批评指正！