



百迈客生物科技
BIOMARKER TECHNOLOGIES

为 世 界 创 造 新 的 可 能

微生物产品介绍

售前技术支持部
郑莉



目录

Contents

第一章

微生物研究背景

第二章

环境微生物

第三章

单菌基因组



百迈客生物科技
BIOMARKER TECHNOLOGIES

1

微生物研究背景

微生物组研究进展

nature reviews microbiology

土壤微生物组的生与死

Nat Rev Microbiol

2022 Feb 28.

DOI: 10.1038/s41579-022-00695-z.

nature

全球海洋微生物组的生物合成潜力

Nature

2022 Jun 22.

DOI: 10.1038/s41586-022-04862-3

Science

肠道菌群与人类的共同多样化

Science

2022 Sep 15.

DOI: 10.1126/science.abm7759

nature reviews microbiology

植物-微生物组相互作用

Nat Rev Microbiol

2020 Aug 12.

DOI: 10.1038/s41579-020-0412-1

nature

藻类微生物与光合作用

Nature

2022 Apr 27.

DOI: 10.1038/s41586-022-04662-9



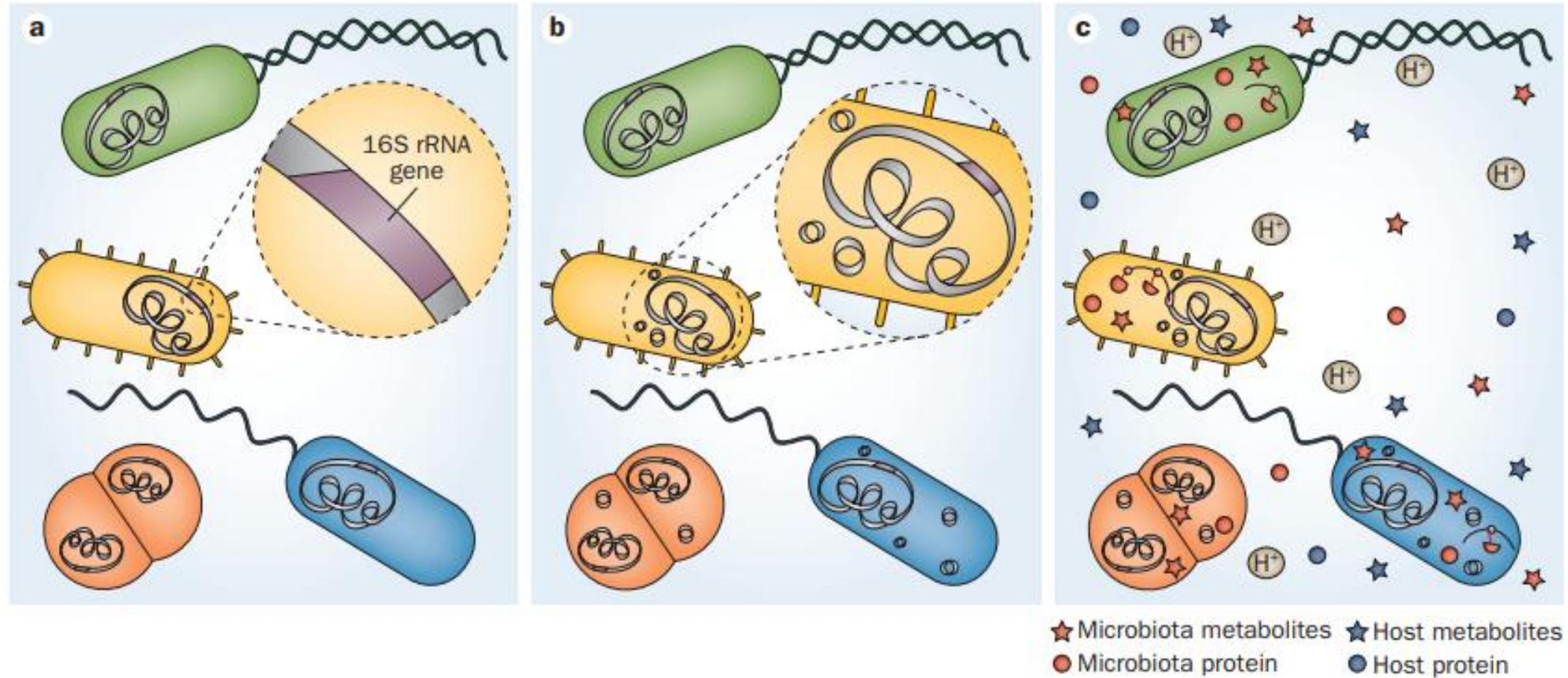
菌群研究中的单细胞方法

Cell

2022 Jul 21

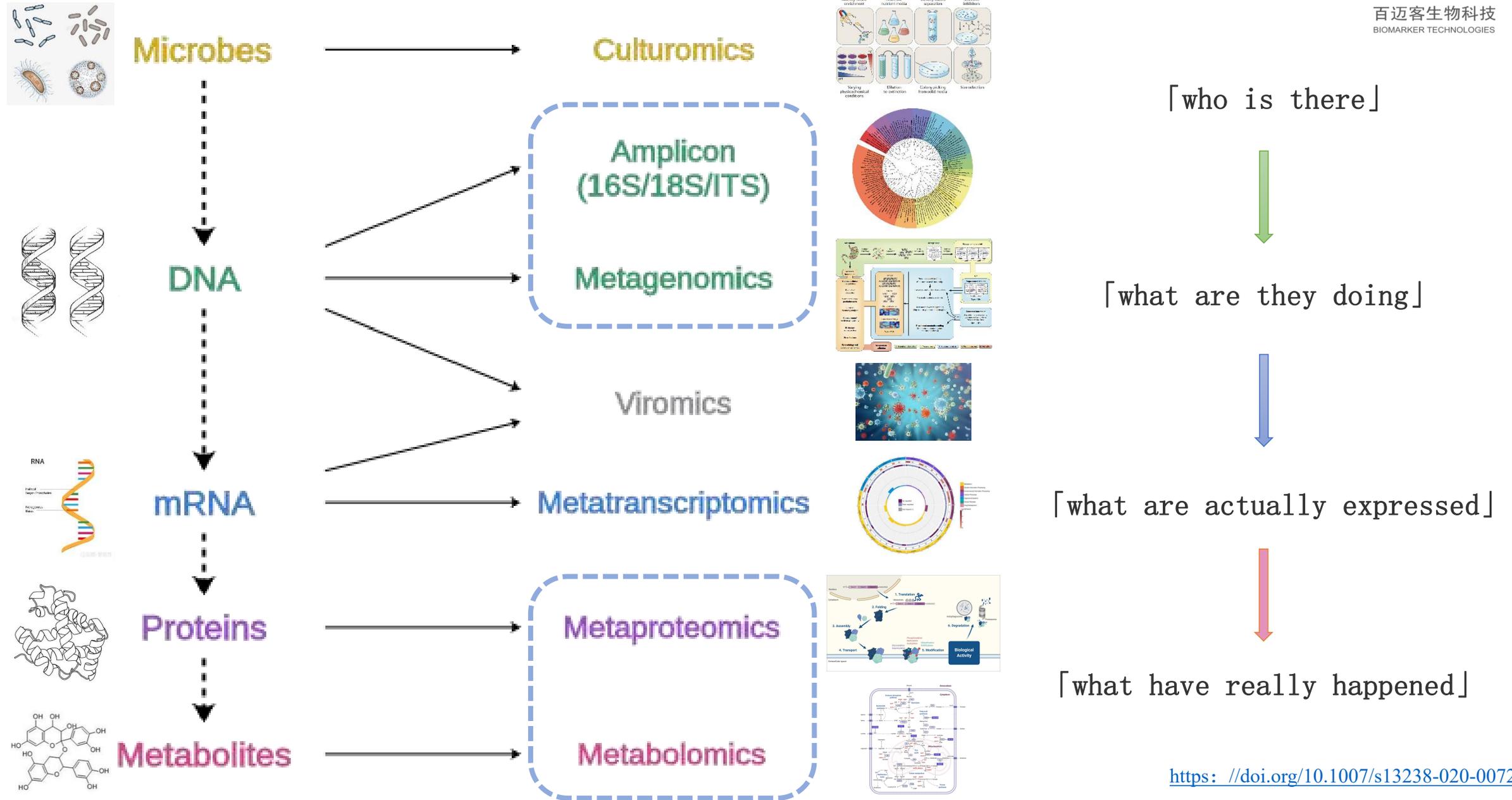
DOI: 10.1016/j.cell.2022.06.040

微生物组学的三个概念



- a. Microbiota (微生物群)：通过**16S rRNA** 基因对环境中的微生物分类鉴定。
- b. Metagenome (宏基因组)：微生物群的**基因和基因组**，突出了种群的遗传潜力。
- c. Microbiome (微生物组)：微生物群的基因和基因组，以及微生物群和宿主环境的**产物**。

微生物组学研究方法



微生物组研究类型



微生物

环境微生物

分离培养

单菌基因组

细菌基因组

真菌基因组

微生物多样性

宏基因组

物种组成

物种多样性

物种和功能组成

物种和功能多样性

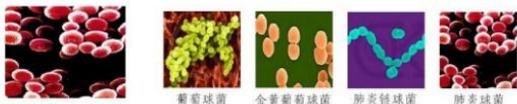
完整基因组结构

基因组特有功能

完整基因组结构

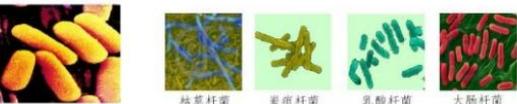
基因组特有功能

球菌



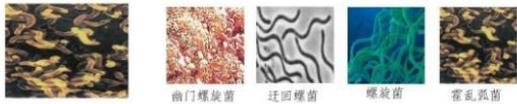
葡萄球菌 金黄色葡萄球菌 肺炎链球菌 肺炎球菌

杆菌



枯草杆菌 炭疽杆菌 乳酸杆菌 大肠杆菌

螺旋菌



幽门螺旋菌 迂回螺菌 螺旋菌 霍乱弧菌



2

环境微生物

环境微生物研究对象

微生物无处不在

人类：口腔、皮肤、粪便、肠道...

动物：瘤胃、肠道...

植物：菌根、内生菌...

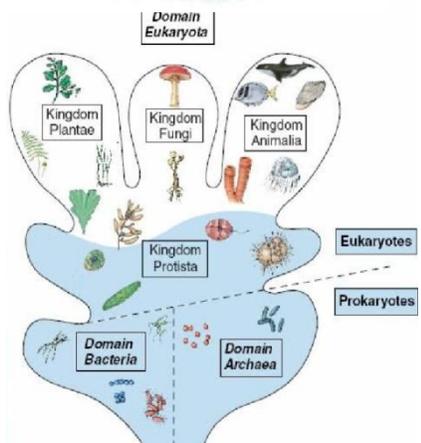
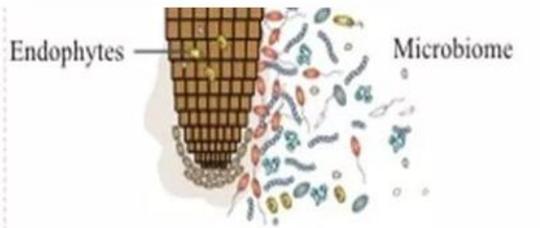
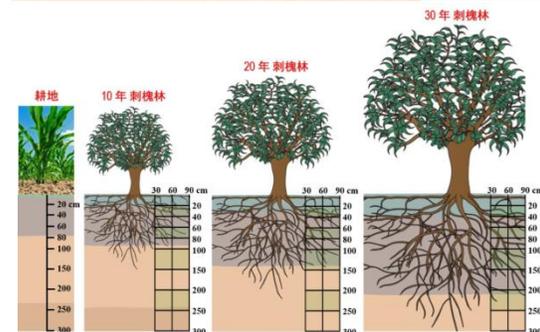
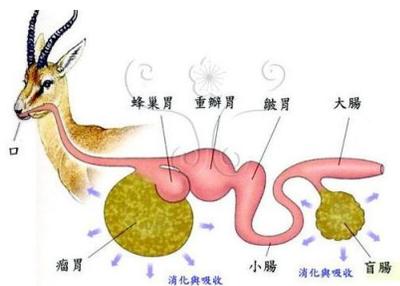
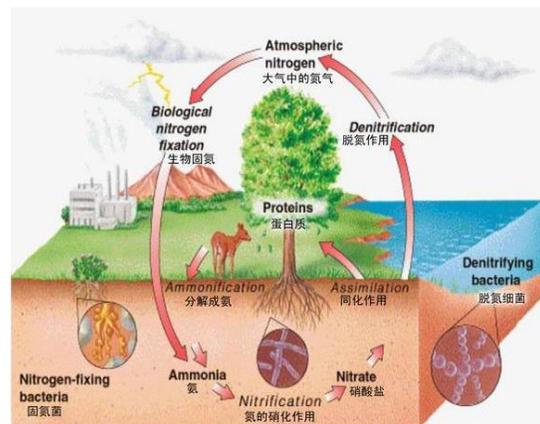
土壤：草原、矿地、冰川、冻土...

水体：海水、河水、湖水、污水...

工业：发酵液、酒曲...

昆虫：蚊子、蚜虫...

环境：矿井、火山、空气...

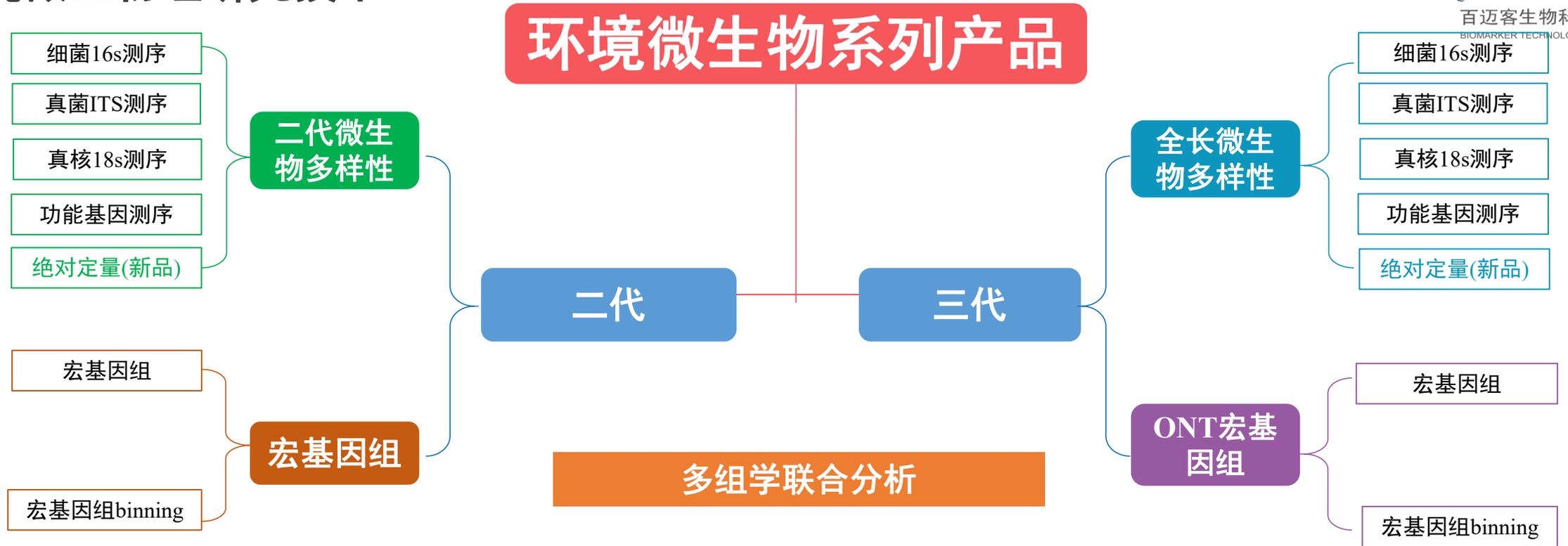


环境微生物组样本类型及客户群体



样本类型			客户群体
土壤	常见类型	植物根系土壤、受污染土壤、极地土壤、高原土壤等	植物所、地理所等
	根系分类	非根际微生物、根际微生物、根表面微生物、根系微生物	
	取样分组依据	地域、土壤类型、植物品种、深度、耕作方式、追肥方式、污染程度	
淤泥	常见类型	海岸带淤泥、污水池淤泥、沼气池淤泥、发酵淤泥、生物电池淤泥等	水产所、环境学院
	取样分组依据	地域、发酵时间、污染程度	
水体	常见类型	海水、冰川、污水、养殖水、注水油藏水系等	水产所、环境学院
	取样分组依据	深度、经纬度、水污染程度、污水处理时间、养殖水等	
粪便	常见类型	人、鼠、动物等	消化科、动物所
	取样分组依据	不同患病程度、给药组处理组健康组、不同饮食处理、不同饲用处理	
肠道	常见类型	人鼠肠粘膜，鼠肠道内容物、胃液、瘤胃液、动物肠道内容物、鱼虾蟹、昆虫	消化科、动物所、水产所、昆虫所
	取样分组依据	不同患病程度、给药健康干预组、不同饮食、不同饲用处理、不同肠道部位	
口腔	常见类型	唾液、痰液、牙菌斑、牙垢、舌苔	口腔科、口腔医院
	取样分组依据	健康患病、不同病程	
拭子	常见类型	皮肤、阴道分泌物等	皮肤科、妇产科
器官	常见类型	人鼠肝脏、肺泡灌洗液等	内科
食品	常见类型	面团、鱼罐头、红肠、	食品学院

环境微生物组研究技术



Illumina测序平台



Water G2-XS QToF



核磁共振仪 Agilent 600mz



PacBio(左) , Nanopore测序平台



微生物多样性研究原理

原理：以扩增序列变体为分子标记研究环境样品中微生物系统分类、物种构成。

□ 二代微生物多样性（16S/ITS/18S/功能基因）

通过PE 250 测序及双端拼接，得到特征Feature(OTUs、ASVs)注释表，快速获得环境样本的菌群结构，分辨率达“属水平”。

□ 全长微生物多样性（16S/ITS/18S 全长）

通过三代PacBio HiFi全长测序，单碱基准确性达99%，无需拼接，分辨率可达“种水平”，更真实还原菌群结构。

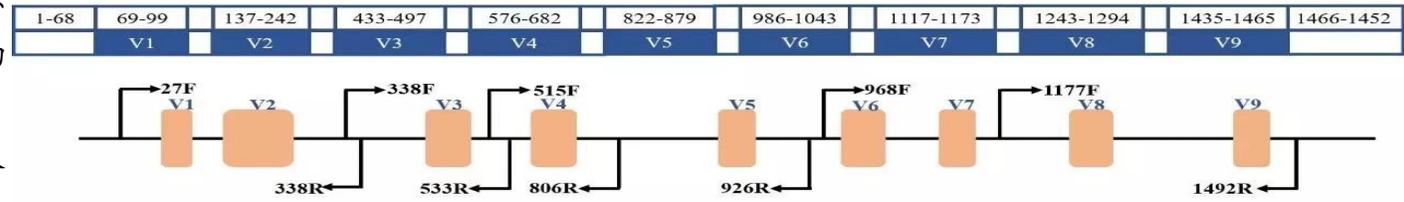
	一代测序	二代测序	三代测序（PacBio测序）
测序读长	——	常规用PE250	15-30 Kb
信号类型	——	荧光信号	荧光信号
优点	准确度高	通量高，准确性相对高， 成本低	CCS纠错，准确性高于99%， 通量高，读长长 种水平注释率
缺点	通量低，需分离培养	读长相对较短，注释到属 水平	成本高

选择合适的测序区域

➤ 细菌多样性

细菌核糖体RNA有三种类型：5S、16S和23S。16S rRNA普遍存在于原核细胞中，且含量较高、拷贝数较多（占细菌RNA总量的80%以上），便于获取模板，功能同源性强，遗传信息量适中，适于作为细菌多样性分析的标准。

16S rDNA编码原核生物核糖体小亚基rRNA的DNA序列，在结构上分为10个保守区和9个可变区，保守区反映生物物种间的亲缘关系，可变区反映物种间的差异。



➤ 真菌多样性

ITS是编码真核生物核糖体rRNA的DNA内转录间隔区序列，由于ITS区属于中度保守的区域，在绝大多数真核生物中表现出极为广泛的序列多态性。

同时，ITS的保守型表现为种内相对一致，种间差异较明显，能够反映出种属间，甚至菌株间的差异。利用它可研究种及种以下的分类阶元。

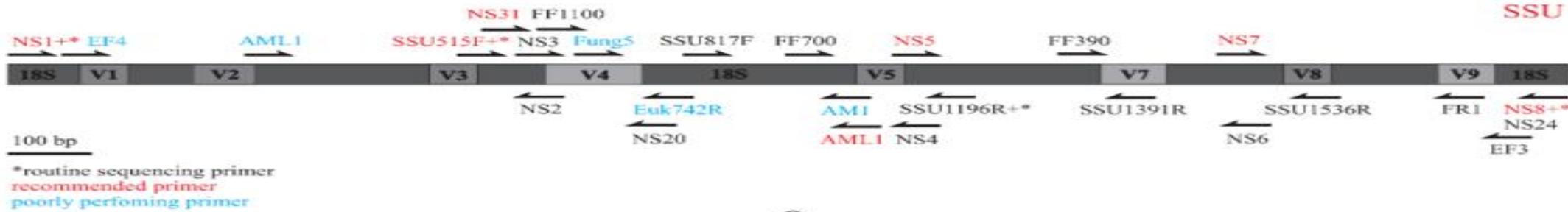


选择合适的测序区域

真核18S多样性

真核微生物中也有三类核糖体RNA，包括5.8S、18S和28S。

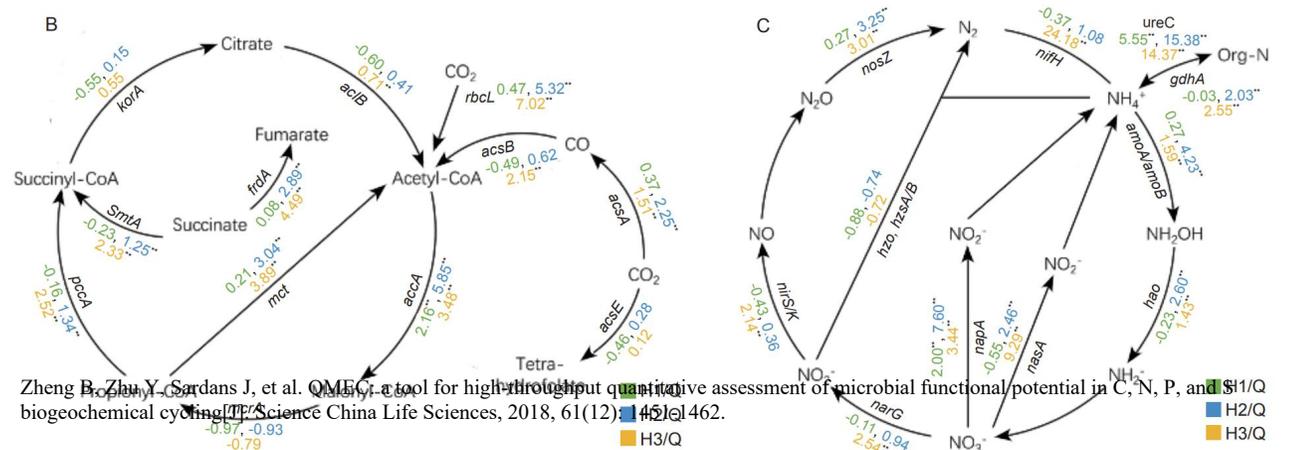
18S rDNA 编码真核生物核糖体小亚基 rRNA的DNA序列，在结构上也有9个可变区（没有V6区），保守区域反映了生物物种间的亲缘关系，而可变区则能体现物种间的差异，适用于作种级及以上的分类标准。



功能基因多样性

每种功能微生物在分类学上可能有很大不同，但却具有相类似的基因使其能够发挥同样的功能，功能基因测序可有效研究特定环境中的功能微生物物种信息

- ❑ 氮循环功能基因: *nifH*, *amoA*, *amoB*, *nirS*, *nirK*, *nosZ*
- ❑ 碳循环功能基因: *pmoA*, *mcrA*
- ❑ 丛枝菌根: AMF
- ❑ 特殊DNA宏条形码: 植物鉴定, 动物鉴定, 鱼类鉴定等



植物内生菌

内生菌：指那些在其生活史的一定阶段或者全部阶段生活于健康植物的各组织和器官内部的微生物。植物内生菌的研究主要集中在药用植物、农作物及特殊生境植物。

研究发现，植物内生菌广泛存在，几乎存在于植物的所有组织中，不仅存在于植物的根、茎、叶、花、果实、胚、种子中，在植物的根瘤中也分离到了内生菌。



属名	拉丁学名	属名	拉丁学名	属名	拉丁学名
食酸菌属	<i>Acidovorax</i>	放线菌属	<i>Actinomyces</i>	产碱菌属	<i>Alcaligenes</i>
气单胞菌属	<i>Aeromonas</i>	土壤杆菌属	<i>Agrobacterium</i>	交替单胞菌属	<i>Alteromonas</i>
节杆菌属	<i>Arthrobacter</i>	固氮菌属	<i>Azotobacter</i>	伯克氏菌属	<i>Burkholderia</i>
芽生杆菌属	<i>Blastobacter</i>	芽单胞菌属	<i>Blastomonas</i>	芽孢杆菌属	<i>Bacillus</i>
棒杆菌属	<i>Corynebacterium</i>	肠杆菌属	<i>Enterobacter</i>	根瘤菌属	<i>Rhizobium</i>
假单胞菌属	<i>Pseudomonas</i>	黄单胞菌属	<i>Xanthomonas</i>	木杆菌属	<i>Xyella</i>

内生细菌、内生真菌项目注意事项

植物宿主：叶绿体、线粒体 线下去宿主后线上分析

内生细菌测序数据量：平均5W tags，至少4W tags（该数据量为去除宿主前的数据量）

内生真菌测序数据量：平均30W tags（该数据量为去除宿主前的数据量，不保证注释结果）**特殊引物内生真菌测序数据量可按常规操作**

项目经验反馈：根的内生真菌注释结果相对于叶和茎会好些

全长微生物多样性研究

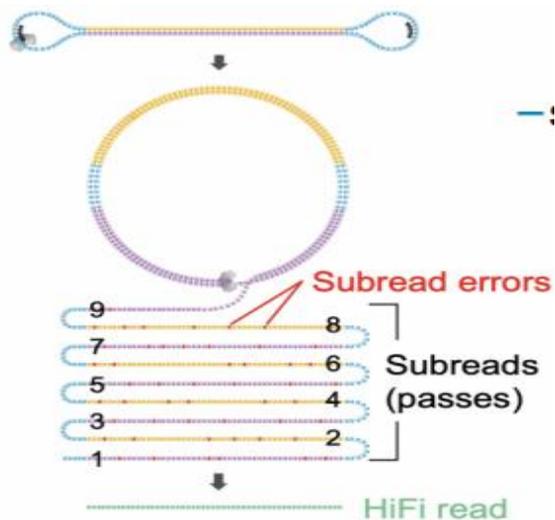


百迈客生物科技

BIOMARKER TECHNOLOGIES

三代微生物多样性是基于 PacBio 测序平台，利用单分子实时测序（SMRT Cell）的方法，基于HIFI模式对对单一片段进行多轮测序的方式来提升准确性，由于Pacbio的原始错误为随机错误，可通过获取CCS进行自身纠正，来提升数据的准确性。

目前 PacBio Sequel II 测序平台平均酶读长可以达到 70 Kb，当测序酶读长达到 8Kb时，就可以满足一条1.5Kb的16S全长纠错5次，minPasses ≥ 5，据官方数据，同一片段测序 5 次后，单一Read的准确性至少Q20（99%），同一片段测序10次后，单一Read的准确性可达Q30（99.9%）；

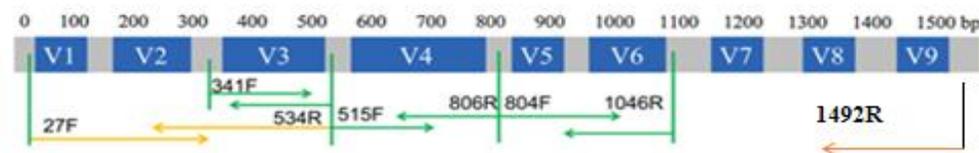


16S全长

27F_(16S-F) 5'- AGRGTTTGATYNTGGCTCAG-3'

1492R_(16S-R) 5'-TASGGHTACCTTGTTASGACTT-3'

产物长度：1.5K

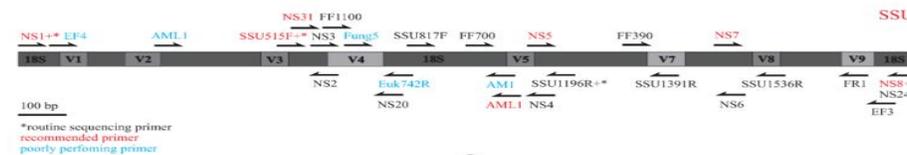


18S全长

Euk-A_(18S-F) 5'-AACCTGGTTGATCCTGCCAGT -3'

Euk-B_(18S-R) 5'-GATCCTTCTGCAGGTTACCTAC-3'

产物长度：1.8K

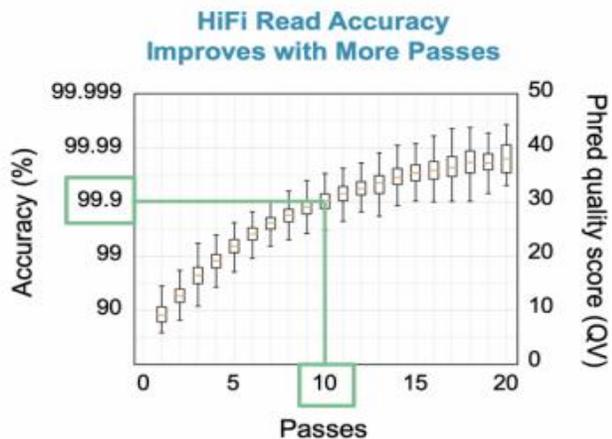


ITS全长

ITS1F 5'-CTTGGTCATTTAGAGGAAGTAA-3'

ITS4 5'-TCCTCCGCTTATTGATATGC-3'

产物长度：0.6-0.7K



为什么要选择全长微生物多样性研究?

为什么要选择全长测序?

二代测序由于读长的限制通常只能**针对部分可变区**进行测序，只提供科或属级别的分类，并且由于不同物种可变区的变异程度不一致会对环境样品中的物种组成和鉴定结果带来一定的偏差（过高/过低的估计群落中的微生物多样性情况）；三代测序由于其测序读长的优势能够对全部可变区进行测序，可以在属和种水平上提供更好的分类鉴定。

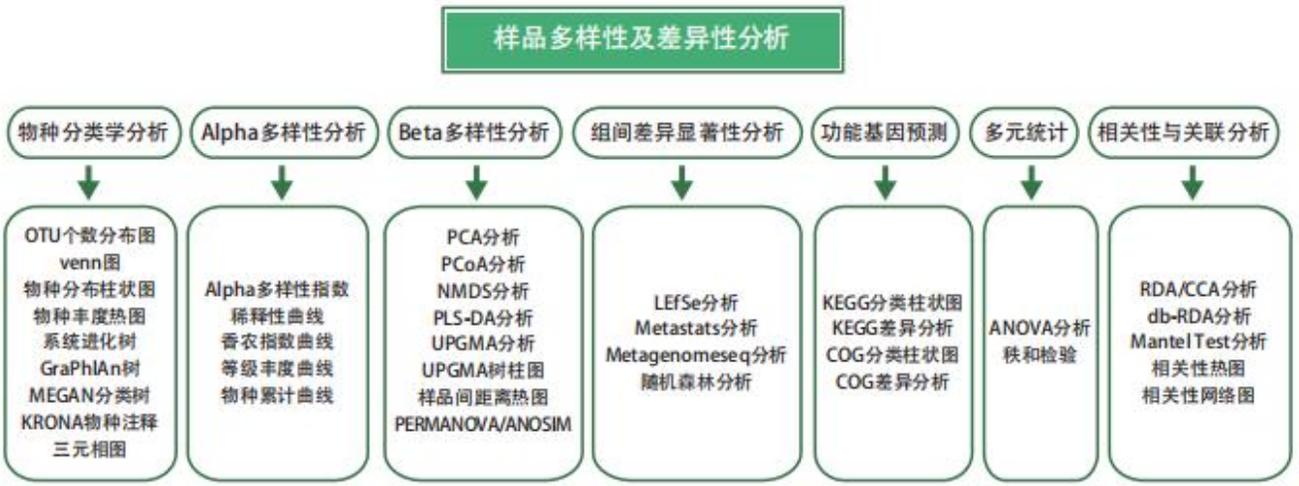
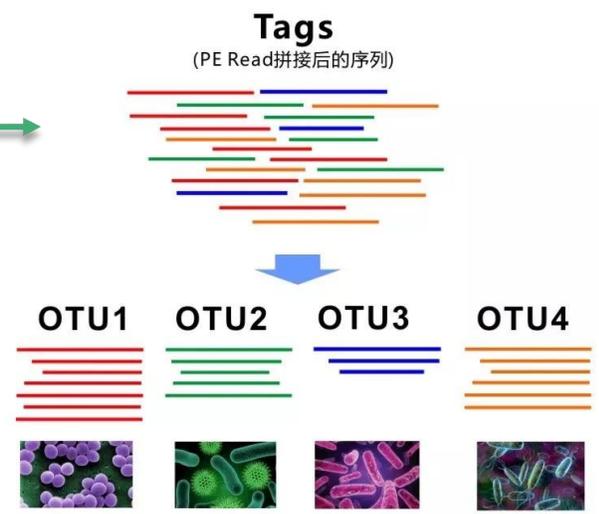
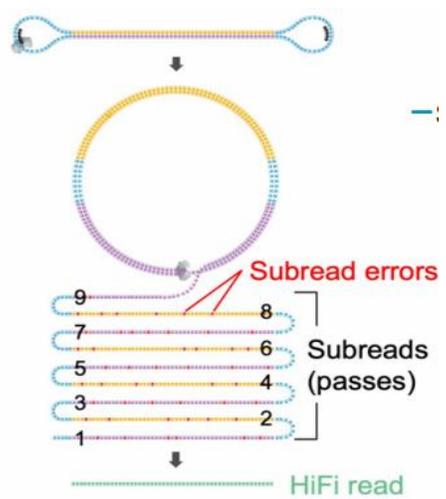
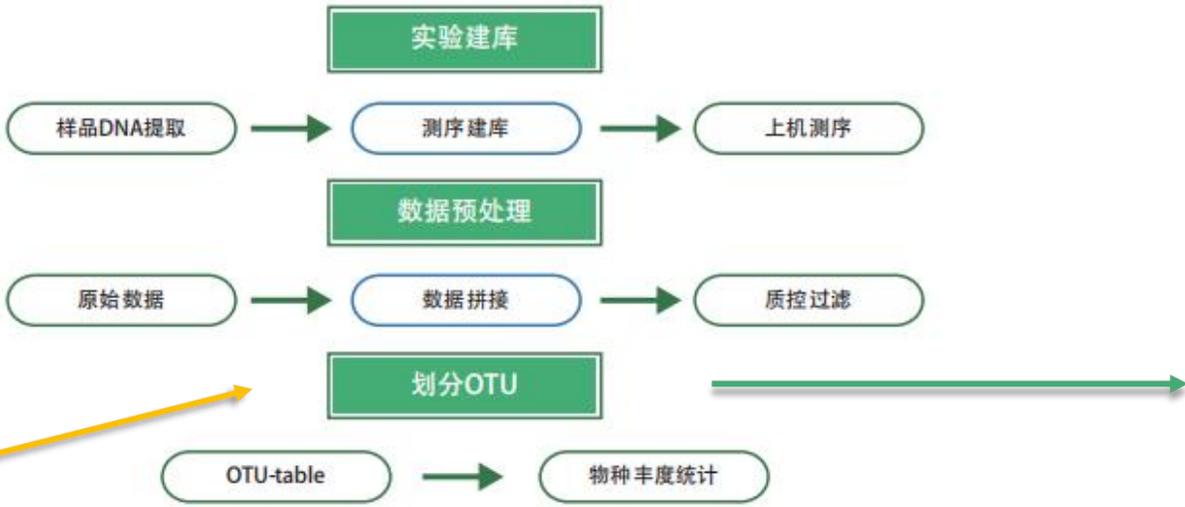
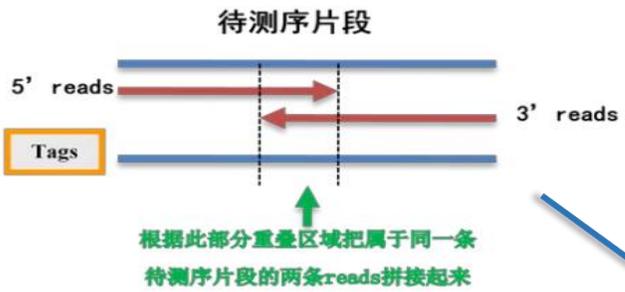
V4区无法识别不同种属



V4区变异程度较高，过高的估计了群落多样性

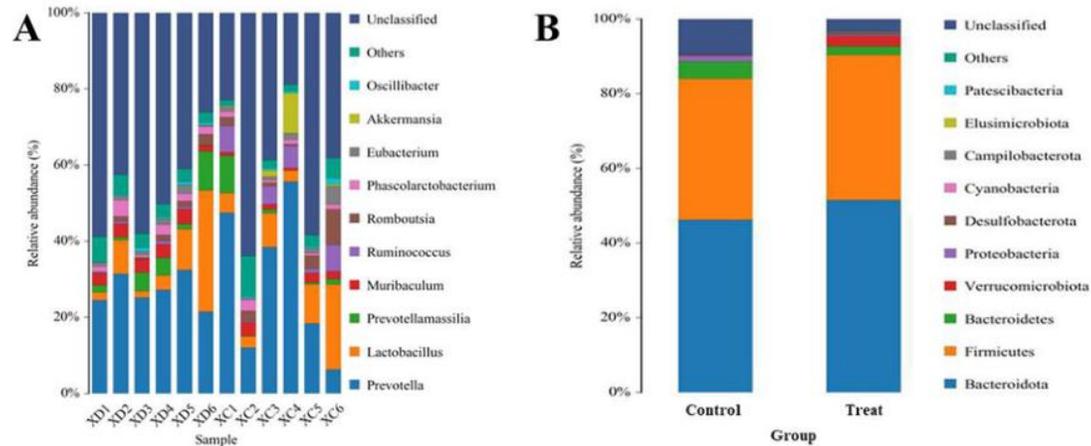


微生物多样性研究技术路线

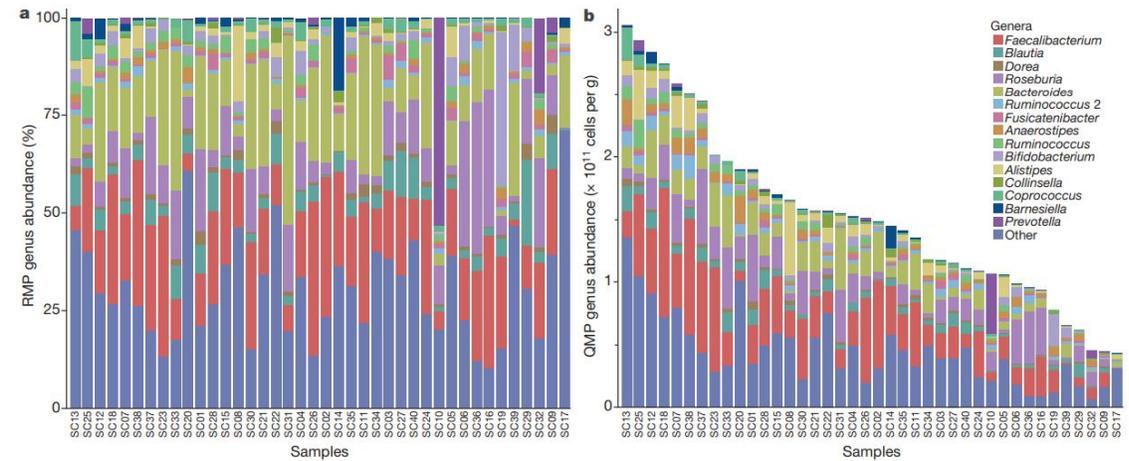


常规微生物多样性测序局限性

- 获取样品中物种的相对丰度，从而进行样品间、组间进行差异及相关性分析等，不能反映样品中物种真实的绝对丰度和真实细菌载量情况。
- 传统QPRC微生物定量通过标准品构建标准曲线，样本需单独扩增，扩增效率存在差异从而造成定量的不准确性。

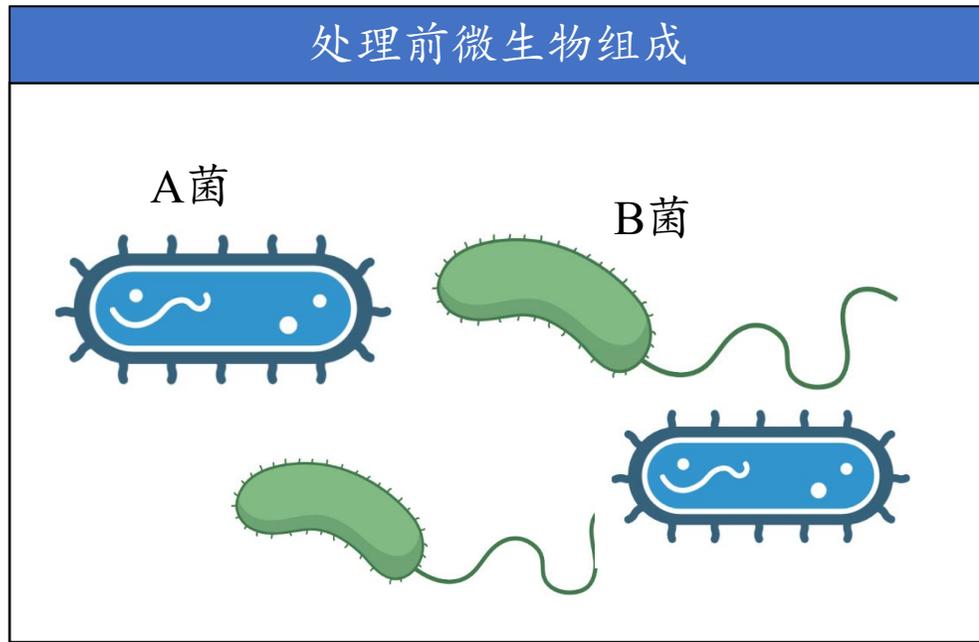


X. Dong et al. *Science of the Total Environment*, 2022

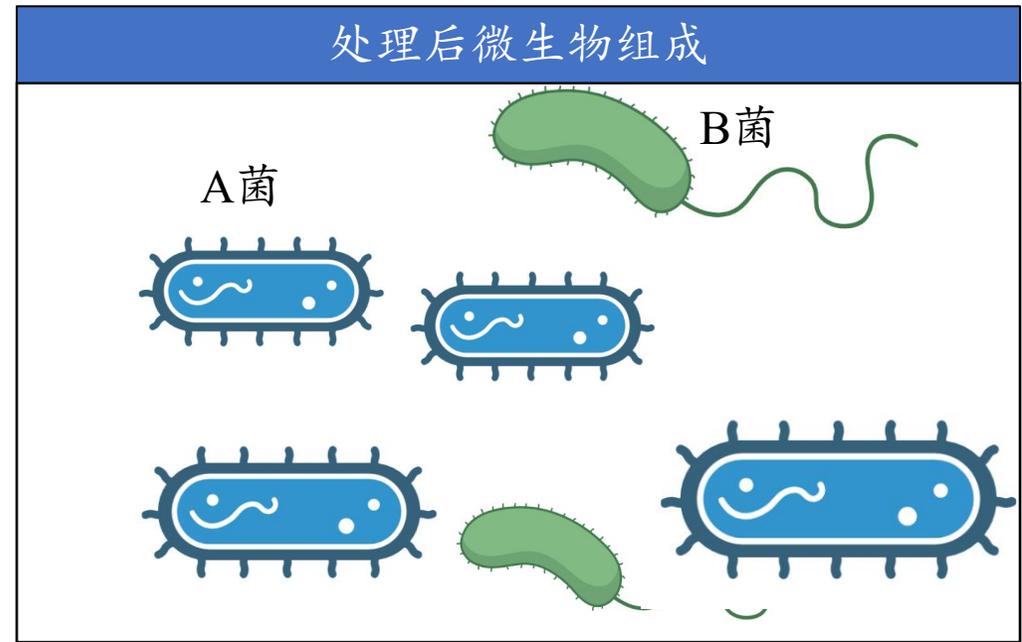


Vandeputte D et al. *Nature*, 2017

相对定量无法反映物种真实含量



相对丰度占比: A: 50% B: 50%
绝对定量cell数: A: 2 B: 2



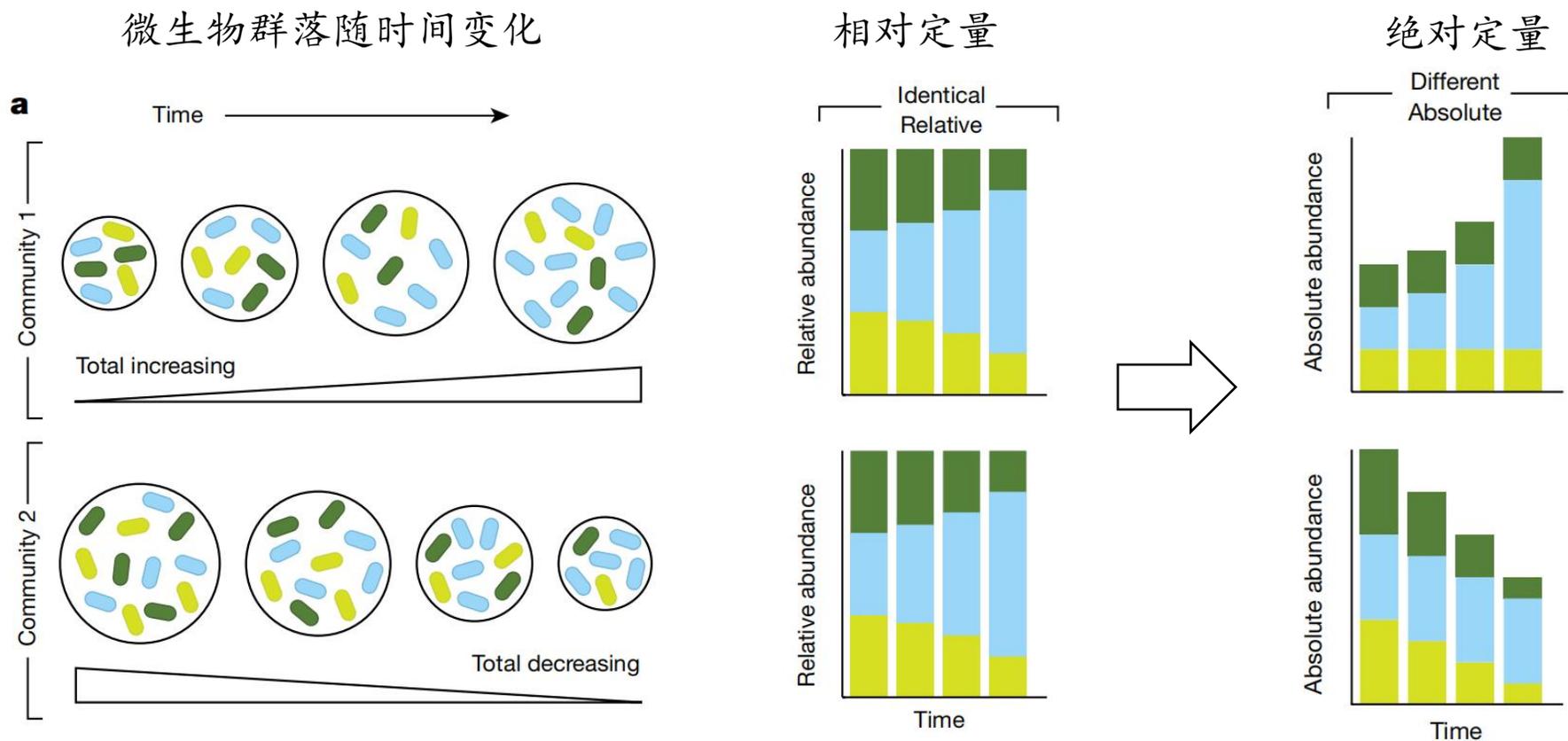
相对丰度占比: A: 66.7% B: 33.3%
绝对定量cell数: A: 4 B: 2

相对定量结果为: A菌丰度增加, B菌丰度降低, 而真实情况B菌并未减少, A菌增加

绝对定量一般是指获得样本中物种的绝对丰度, 真实的物种含量 (细胞数/16s rRNA基因拷贝数)。

Cells or 16s rRNA gene copies拷贝数

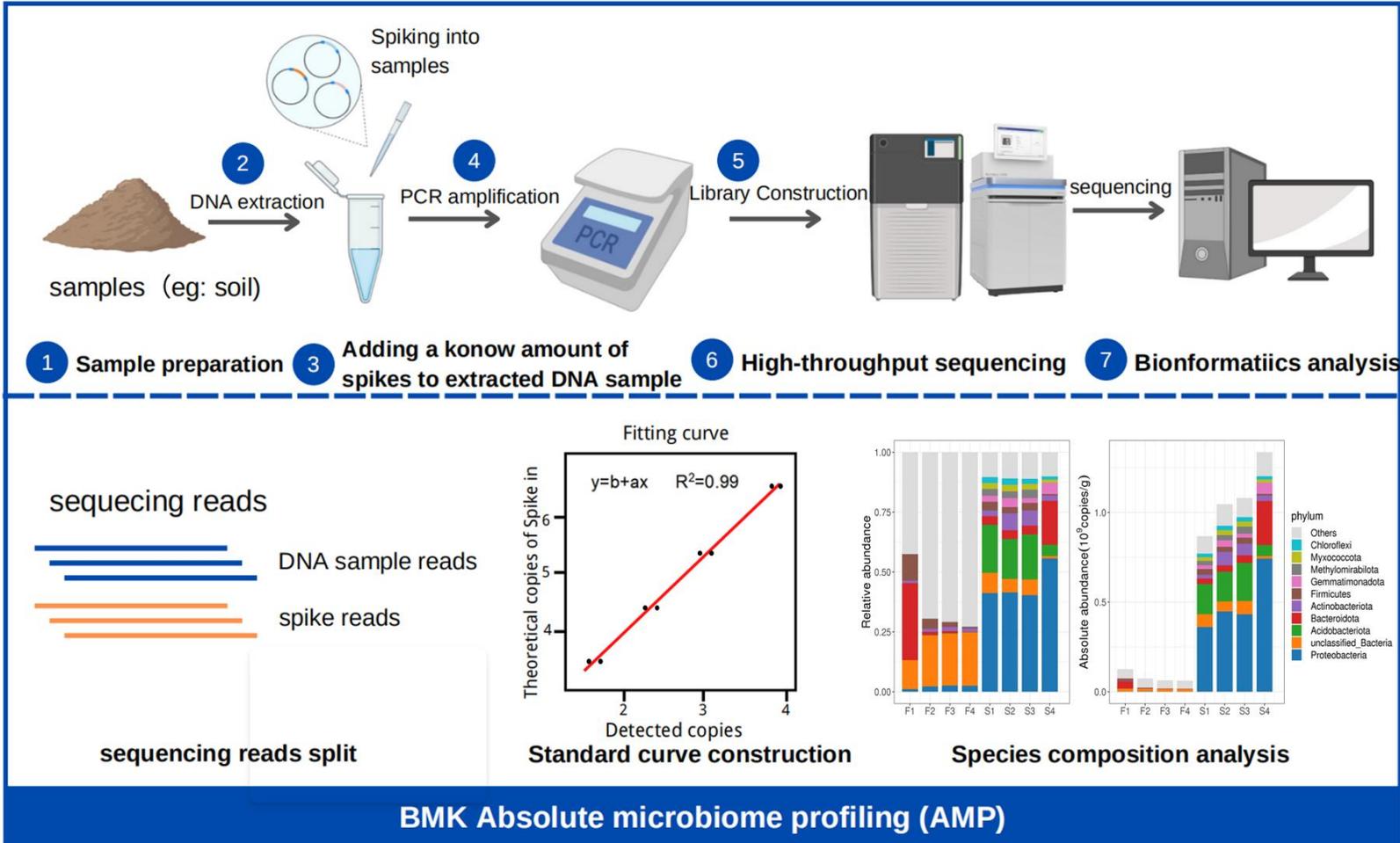
微生物绝对定量测序解决传统微生物组研究局限



两个群落，一个**随着时间的推移**逐渐繁盛，另一个逐渐消亡，但相对定量显示为相同的结论
相对丰度数据的组成性质因此可能掩盖群落动态，无法真实还原群落的繁盛和消亡。

绝对定量能够更真实还原微生物的**群落动态变化**

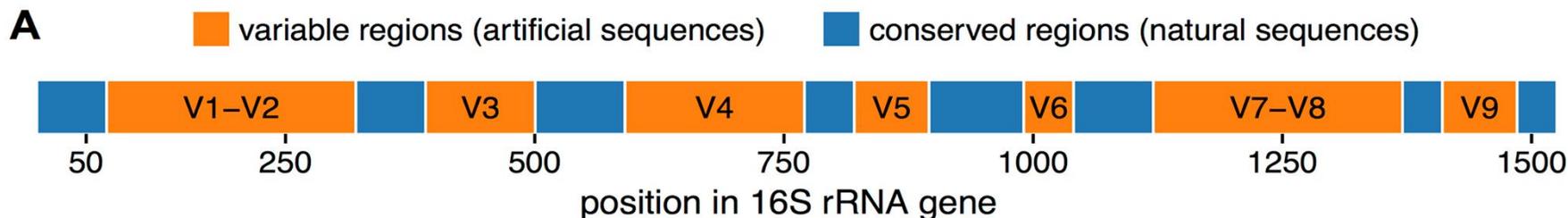
百迈客微生物绝对定量技术原理



- 运用高通量测序技术和spike-in内标法实现对细菌的绝对定量。
- 通过在样品中加入已知拷贝数的spike-in内参序列，与样品同时进行扩增、建库测序，然后通过内参的理论拷贝数与测序tags数，绘制标准曲线，结合标准曲线及样品的组织、核酸、扩增用量等计算得到OTUs/ASVs的绝对拷贝数表。
- 能够一次获得所有细菌的绝对定量结果，包括总丰度和每种菌的丰度。

Spike-in 内参序列设计原则

什么是 Spike-in ?



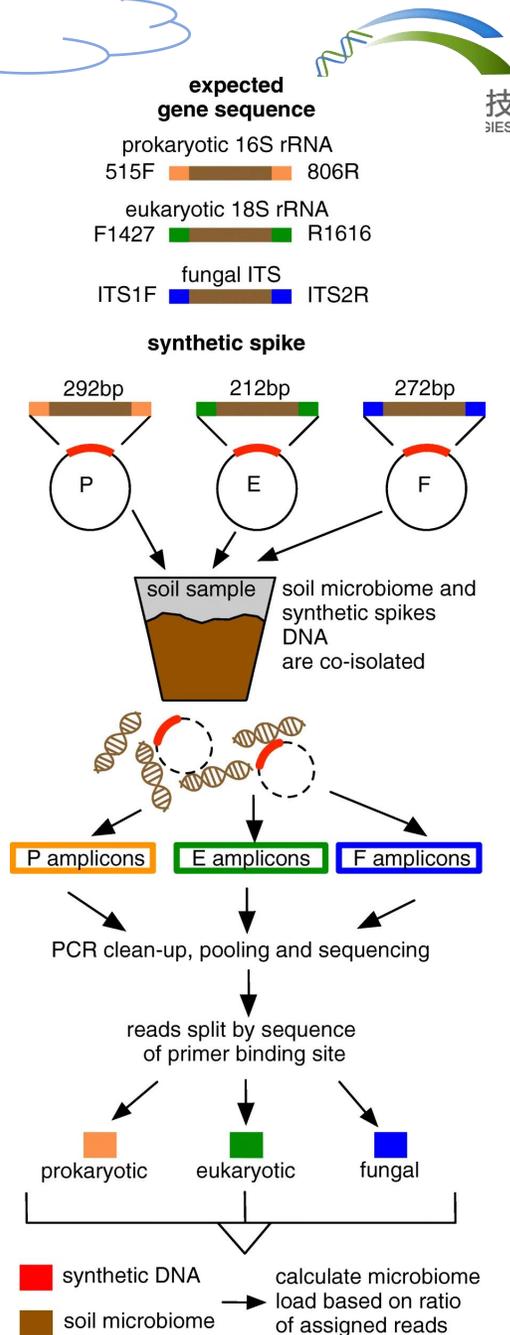
- spike-in是人工合成的一段核酸片段(Synthetic spike-in), 由与选定的天然 16S rRNA 基因相同的保守区和人工可变区组成。
- Spike in 序列与样本中的微生物基因组序列缺乏一致性。

简单理解就是它在样本中是独一无二的, 不会影响到其他序列, 易于区分, 后续分析可以根据已知序列信息进行提取序列。目前我们是以80%以下同源性作为设计原则。

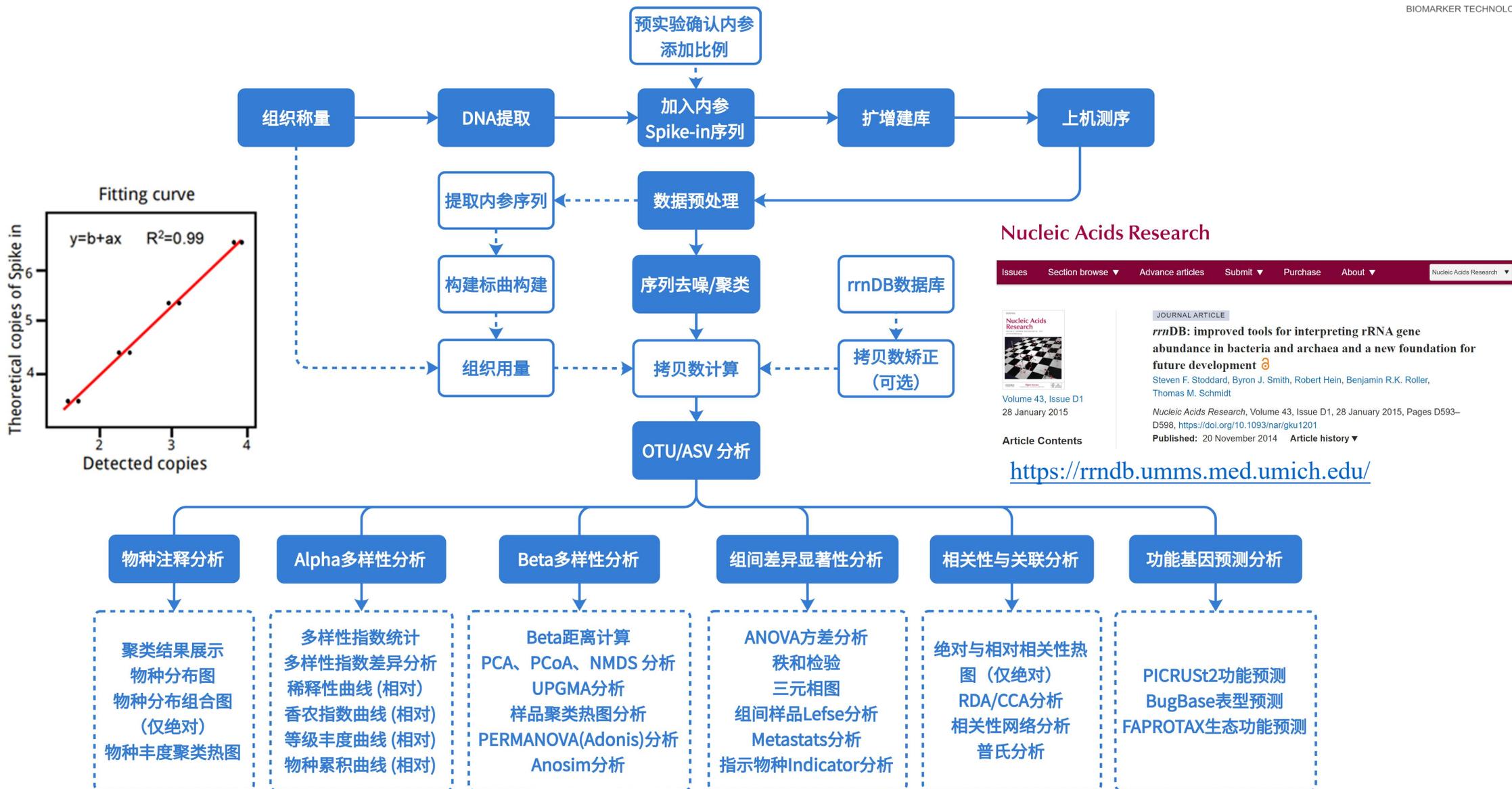
spike-in的设计有三个关键因素:

- ① 具有共同的基因引物结合位点(primer binding sites, PBSs);
- ② 与待测靶标序列相似的长度和GC含量;
- ③ 容易获得、易于处理;

将合成的序列克隆到质粒载体中, 将其转化到大肠杆菌中进行扩繁得到大量的 spike-in 序列。



百迈客微生物绝对定量技术路线



rrnDB数据库拷贝数校正

Nucleic Acids Research



Issues Section browse ▼ Advance articles Submit ▼ Purchase About ▼ All Nucleic Acids Resear ▼

<https://rrnodb.umms.med.umich.edu/>



Volume 43, Issue D1
28 January 2015

rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development 

Steven F. Stoddard, Byron J. Smith, Robert Hein, Benjamin R.K. Roller, Thomas M. Schmidt

Nucleic Acids Research, Volume 43, Issue D1, 28 January 2015, Pages D593–D598, <https://doi.org/10.1093/nar/gku1201>

Published: 20 November 2014 [Article history ▼](#)

Last update: rrnDB
version 5.8
June 23, 2022

- 16S rRNA基因在一个物种中会有多拷贝的现象存在，在PCR后会有多个扩增产物，比如物种A只有1个16S基因，物种B有2个16S基因，在原始群落中，二者丰度相同，在经过相同循环次数的PCR，理论上最终测序得到的reads中，物种B的reads数会是物种A的2倍；
- 而我们通常使用 reads 数表征某个OTU的丰度，尽管在原始群落中物种A和B相同，但是由于拷贝数的差异，所以定量的结果不能正确的反映在群落中二者的丰度；这对于后续的 alpha 和 beta 多样性的分析都会有影响；
- rrnDB数据库中收录了16S基因为多拷贝的物种；经过校正后的数据更接近微生物的细胞数，生信分析结果更**真实可靠**。

利用rrnDB数据库进行拷贝校正，使拷贝数更接近细胞数

Stoddard S.F, Smith B.J., Hein R., Roller B.R.K. and Schmidt T.M. (2015) rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Research* 2014

高

超高性价比

- 1个样本单次测序，同时获得相对定量和绝对定量结果
- OTU聚类和ASV去噪可选
- 独有特色分析

高准确度

rrnDB数据库进行拷贝数校正
16s拷贝数更接近真实细菌细胞数

准

种

种水平定量

绝对定量结合Pacbio三代测序实现物种水平定量
Silva138等多种数据库任选

新发现

技术成熟，2016年提出，两篇Nature
可广泛应用于现有研究领域
从绝对定量角度阐述问题

新

METHODOLOGY

Adjusting microbiome profiles for differences in microbial load by spike-in bacteria

Frank Stammler^{1,2*}, Joachim Gläser³, Andreas Hiergeist³, Ernst Holler³, Daniela Weber³, André Gessner⁴ and Rainer Spang¹

Abstract

Background: Next-generation 16S ribosomal RNA gene sequencing is widely used to determine the composition of the mammalian gut microbiomes. However, in the absence of a reference alterations in absolute abundance of specific operational taxonomic units if microbial load varies. Here we suggest the spiking of exogenous bacteria into crude specimens to quantify bacterial abundances. We use the 16S rRNA read counts of the spike-in bacteria to adjust the endogenous bacteria for changes in total microbial loads. Using a series of dilutions of pooled rice containing defined amounts of the spike-in bacteria *Salibacter ruber*, *Rhizobium radiolium*, *acidiphilum*, we demonstrate that spike-in-based calibration to microbial loads allows accurate absolute endogenous bacteria abundances. Applied to stool specimens of patients undergoing transplantation, we were able to determine changes in both relative and absolute abundance especially the genus *Enterococcus*, in response to antibiotic treatment and radio-chemotherapy. **Conclusion:** Exogenous spike-in bacteria in gut microbiome studies enable estimation of absolute abundances, providing novel insights into the structure and the dynamics of intestinal microbial communities. **Keywords:** Microbial load, Spike-in bacteria, 16S rRNA gene sequencing, Standardization, Microbiome communities, Community analysis

Background

The human intestinal tract is populated by an ecological community of microorganisms, the gut microbiome. This complex community plays an important role in health and disease [1–7] and varies greatly among individuals. Next generation sequencing of the 16S rRNA gene allows profiling of the bacterial and archaeal components of the gut microbiome at unprecedented precision and depth [8–10]. Computational tools such as QIIME [11] and mothur [12] cluster reads into operational taxonomic units (OTUs) [13], which may then be joined into taxonomic groups at the genus, family, order, class, and phylum level.

Current studies focus on the portions of OTUs [14, 15], as may contribute 5% to microbiome B corresponding to a rare as it is the total number of of in A is four times larger than for twice as many bacteria as for the actual ratio to 2:1.

Antibiotic treatment, diet, microbial load and composition [16] observed that the genus *Enterococcus* in stool patients undergoing allogeneic ASCT) can increase from 1% to ASCT) up to 94% after ASCT relative shift to *Enterococcus* increased risk of acute graft

* Correspondence: Frank.Stammler@klinikum-regensburg.de
¹Chair of Statistical Bioinformatics, University of Regensburg, Am Bischofshof 9, 93055 Regensburg, Germany
²Institute of Clinical Microbiology and Hygiene, University Medical Centre, Franz Josef Strauß Allee 11, 93053 Regensburg, Germany
Full list of author information is available at the end of the article

© 2016 The Author(s). Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Open Access

LETTER

2017-Nature

Quantitative microbiome profiling link community variation to microbial load

Doris Vandeputte^{1,2,3*}, Gunter Kathagen^{1,2,3}, Kevin D'hoel^{2,3,4}, Sara Vieira-Silva^{2,3,4}, Mireia Valle Jun Wang^{2,3}, Raai V. Tito^{2,3}, Lindsey De Commer^{2,3}, Youssef Darzi^{2,3}, Séverine Vermeire^{2,3}, Gwen F

Current sequencing-based analyses of faecal microbiota quantify microbial taxa and metabolic pathways as fractions of the sample sequencing library generated by each analysis^{1,2}. Although these relative approaches permit detection of disease-associated microbial variation, they are limited in their ability to reveal the interplay between microbiota and host health^{3,4}. Comparative analyses of relative microbiome data cannot provide information about the extent or directionality of changes in taxa abundance or metabolic potential^{5,6}. If microbial load varies substantially between samples, relative profiling will hamper attempts to link microbiome features to quantitative data such as physiological parameters or metabolite concentrations^{7,8}. Saliently, relative approaches ignore the possibility that altered overall microbiota abundance itself could be a key identifier of a disease-associated ecosystem configuration⁹. To enable genuine characterization of host-microbiota interactions, microbiome research must exchange ratios for counts^{10,11}. Here we build a workflow for the quantitative microbiome profiling of faecal material, through parallelization of amplicon sequencing and flow cytometric enumeration of microbial cells. We observe up to tenfold differences in the microbial loads of healthy individuals and relate this variation to enterotype differentiation. We show how microbial abundances underpin both microbiome variation between individuals and covariation with host phenotype. Quantitative profiling bypasses compositionality effects in the reconstruction of gut microbiota interaction networks and reveals that the taxonomic trade-off between *Bacteroides* and *Prevotella* is an artefact of relative microbiome analyses. Finally, we identify microbial load as a key driver of observed microbiota alterations in a cohort of patients with Crohn's disease¹², here associated with a low-cell-count *Bacteroides* enterotype (as defined through relative profiling)^{13,14}.

First, we collected a set of 40 fresh faecal samples (the study cohort), which were processed within one hour of acquisition. We compiled an accompanying set of basic matching metadata, with an emphasis on anthropometrics and stool characteristics (Supplementary Table 1). Given expected dietary effect sizes¹⁵ and cohort limitations, participants were not requested to keep food records. Sample analysis was aligned with Flemish Gut Flora Project (FGFP) protocols¹⁶. Metadata exploration reaffirmed the previously reported association of stool consistency (Bristol Stool Scale (BSS) score¹⁷) with moisture¹⁸ (Spearman's $\rho = 0.45$, $P = 5.2 \times 10^{-3}$, Supplementary Table 2). Microbiome analysis of frozen faecal aliquots characterized the sample set as within the bounds of FGFP community space, distributed over four enterotypes that were identified on the basis of Dirichlet multinomial mixtures (DMM) (Fig. 1a). Stool moisture and donor age were identified as the microbiome covariates that displayed the largest non-redundant effect size, jointly explaining 9.3% of inter-individual microbiota variation (stepwise distance-based redundancy analysis) (BRED; Fig. 1b).

¹KU Leuven - University of Leuven, Department of Microbiology and Immunology, Rega Institute, Herestraat 49, B-3000 Leuven, Belgium; ²Rega Institute, Herestraat 49, B-3000 Leuven, Belgium; ³Research Group of Microbiology, Department of Biomedical Sciences, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium; ⁴Center for Gastrointestinal Disorders (GARGO), KU Leuven, B-3000 Leuven, Belgium; ⁵These authors contributed equally to this work; ⁶These authors jointly supervised this work.

doi:10.1038/nature24460

2019-Nature Microbiology

LETTERS

<https://doi.org/10.1038/s41564-019-0483-9>

nature
microbiology

Quantitative microbiome profiling disentangles inflammation- and bile duct obstruction-associated microbiota alterations across PSC/IBD diagnosis

Sara Vieira-Silva^{1,2,3*}, João Sabino^{3,7}, Mireia Valles-Colomer^{1,2,3}, Gwen Falony^{1,2,3}, Gunter Clara Caenepeel⁴, Isabelle Cleyneux⁴, Schalk van der Merwe^{3,5,6}, Séverine Vermeire^{3,8,9}, Jeroen Raes^{3,10,9*}

Recent work has highlighted the importance of confounder control in microbiome association studies^{1,2}. For instance, multiple pathologies previously linked to gut ecosystem dysbiosis display concomitant changes in stool consistency³, a major covariate of microbiome variation^{4,5}. In those cases, observed microbiota alterations could largely reflect variation in faecal water content. Moreover, stool moisture variation has been linked to fluctuations in faecal microbial load, inducing artefacts in relative abundance profile analyses^{6,7}. Hence, the identification of associations between the gut microbiota and specific disease manifestations in pathologies with complex aetiologies requires a deconfounded, quantitative assessment of microbiome variation. Here, we revisit a disease association microbiome data set comprising 106 patients with primary sclerosing cholangitis (PSC) and/or inflammatory bowel disease^{8,9}. Assessing quantitative taxon abundances¹⁰, we study microbiome alterations beyond symptomatic stool moisture variation. We observe an increased prevalence of a low cell count *Bacteroides* 2 enterotype across the pathologies studied, with microbial loads correlating inversely with intestinal and systemic inflammation markers. Quantitative analyses allow us to differentiate between taxa associated with either intestinal inflammation severity (*Fusobacterium*) or cholangitis/biliary obstruction (*Enterococcus*) among previously suggested PSC marker genera. We identify and validate a near-exclusion pattern between the inflammation-associated *Fusobacterium* and *Veillonella* genera, with *Fusobacterium* detection being restricted to Crohn's disease and patients with PSC-Crohn's disease. Overall, through absolute quantification and confounder control, we single out clear-cut microbiome markers associated with pathophysiological manifestations and disease diagnosis.

¹Laboratory of Molecular Bacteriology, Department of Microbiology and Immunology, Rega Institute, KU Leuven, Leuven, Belgium; ²Center for Gastrointestinal Disorders (GARGO), KU Leuven, Leuven, Belgium; ³Translational Research Center for Gastrointestinal Disorders (TRGID), KU Leuven, Leuven, Belgium; ⁴Human Genetics, KU Leuven, Leuven, Belgium; ⁵Department of Gastroenterology and Hepatology, KU Leuven, Leuven, Belgium; ⁶Laboratory of Microbiology, KU Leuven, Leuven, Belgium; ⁷These authors contributed equally; ⁸Sara Vieira-Silva, João Sabino, Mireia Valles-Colomer, Gwen Falony, Jeroen Raes jointly supervised this work; ⁹Séverine Vermeire, Jeroen Raes, *email: severine.vermeire@kuleuven.be, jeroen.raes@kuleuven.be

2021-Nature

Article

Multi-kingdom ecological drivers of microbiota assembly in preterm infants

Chitong Rao^{1*}, Katharine Z. Coyte^{2,3,4,5}, Wayne Baintner⁶, Raif S. Gheh⁷, Camilla R. Martin⁸ & Seth Rakoff-Nahoum^{1,4,7,9,10}

Received: 18 March 2020

Accepted: 14 January 2021

Published online: 24 February 2021

Check for updates

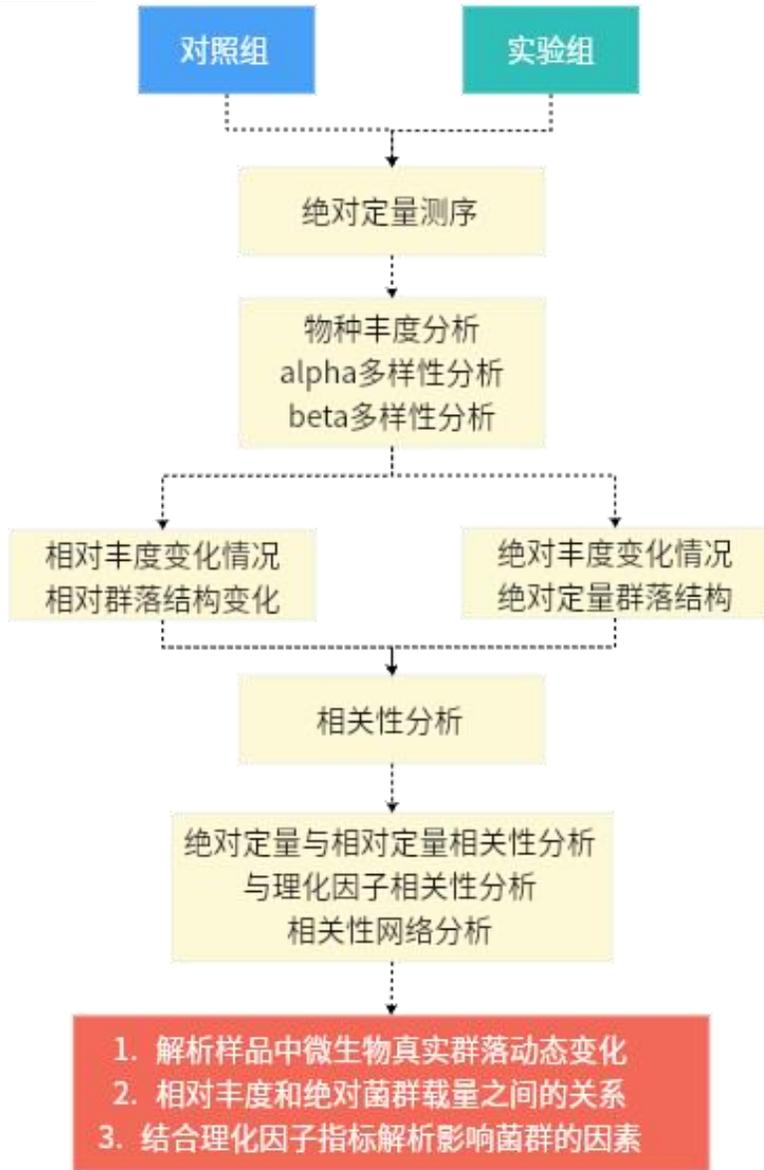
The gut microbiota of preterm infants develops predictably^{1,2}, with pioneer species colonizing the gut after birth, followed by an ordered succession of microorganisms. The gut microbiota is vital to the health of preterm infants^{3,4}, but the forces that shape these predictable dynamics of microbiome assembly are unknown. The environment, the host and interactions between microorganisms all potentially shape the dynamics of the microbiota, but in such a complex ecosystem, identifying the specific role of any individual factor is challenging^{5,6}. Here we use multi-kingdom absolute abundance quantification, ecological modelling and experimental validation to address this challenge. We quantify the absolute dynamics of bacteria, fungi and archaea in a longitudinal cohort of 178 preterm infants. We uncover microbial blooms and extinctions, and show that there is an inverse correlation between bacterial and fungal loads in the infant gut. We infer computationally and demonstrate experimentally in vitro and in vivo that predictable assembly dynamics may be driven by directed, context-dependent interactions between specific microorganisms. Mirroring the dynamics of macroscopic ecosystems^{7,8}, a late-arriving member of the microbiome, *Keleisia*, exploits the pioneer microorganism, *Staphylococcus*, to gain a foothold within the gut. Notably, we find that interactions between different kingdoms can influence assembly, with a single fungal species—*Candida albicans*—inhibiting multiple dominant genera of gut bacteria. Our work reveals the centrality of simple micro-microbe interactions in shaping host-associated microbiota, which is critical both for our understanding of microbial ecology and for targeted microbiota interventions.

Humans are colonized by vast communities of microorganisms, particularly within the gastrointestinal tract, that have key roles in the health of the host^{1,2}. Infants are generally born uncolonized by microorganisms, and their gut microbiomes gradually assemble after birth³. Notably, this developmental process occurs in a predictable manner, with specific bacterial taxa establishing in the gut at distinct points in infant life^{4,5}. The early-life microbiota is critical to infant health, and microbiota composition is linked to a range of diseases, morbidity and mortality, particularly within preterm infants^{6,7,8}. Yet despite its importance, we do not understand what drives the patterned progression of the infant gut community^{9,10}. Gestational age, delivery mode, host epithelial and immune ontogeny, diet, antibiotics and the interactions between individual microorganisms may each influence microbiota composition^{11–13,14}. But with such complexity, the effect of any individual factor on the development of the microbiota has remained unclear, and disentangling how and why microbial communities change over time remains a major challenge—both for the human microbiota and for host-associated and environmental microbiomes more broadly.

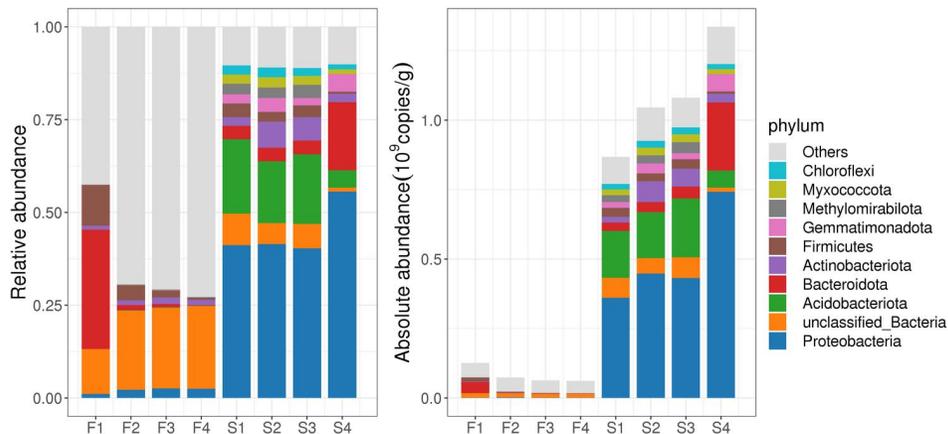
Our ability to identify drivers of microbiota development has been hampered by the complexity of microbial ecosystems and also by fundamental limitations in how we quantify the composition of microbial communities¹⁵. First, although next-generation sequencing (NGS) has provided a comprehensive map of bacterial diversity within the human gut^{16,17}, we still know little about the other microorganisms—such as fungi and archaea—that colonize the infant microbiota^{18,19}, which constrains our ability to identify inter-kingdom interactions that drive ecosystem dynamics²⁰. Second, NGS data typically chart only the relative abundances of taxa, providing the proportions of different microorganisms within a community²¹, but not the absolute abundance, delivery mode, host epithelial and immune ontogeny, diet, antibiotics and the interactions between individual microorganisms may each influence microbiota composition^{11–13,14}. But with such complexity, the effect of any individual factor on the development of the microbiota has remained unclear, and disentangling how and why microbial communities change over time remains a major challenge—both for the human microbiota and for host-associated and environmental microbiomes more broadly.

¹Division of Infectious Diseases, Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA; ²Division of Evolution and Genomic Sciences, Institute of Biological Sciences, University of Manchester, Manchester, UK; ³Division of Immunology, Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA; ⁴Department of Immunology and Division of Translational Research, Beth Israel Deaconess Medical Center, Boston, MA, USA; ⁵Division of Gastroenterology, Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA; ⁶Department of Microbiology, Harvard Medical School, Boston, MA, USA; ⁷Institut für MFI and Harvard, Cambridge, MA, USA; ⁸These authors contributed equally; ⁹Chitong Rao, Katharine Z. Coyte, *email: katharine.coyte@manchester.ac.uk, seth.rakoff-nahoum@childrens.harvard.edu

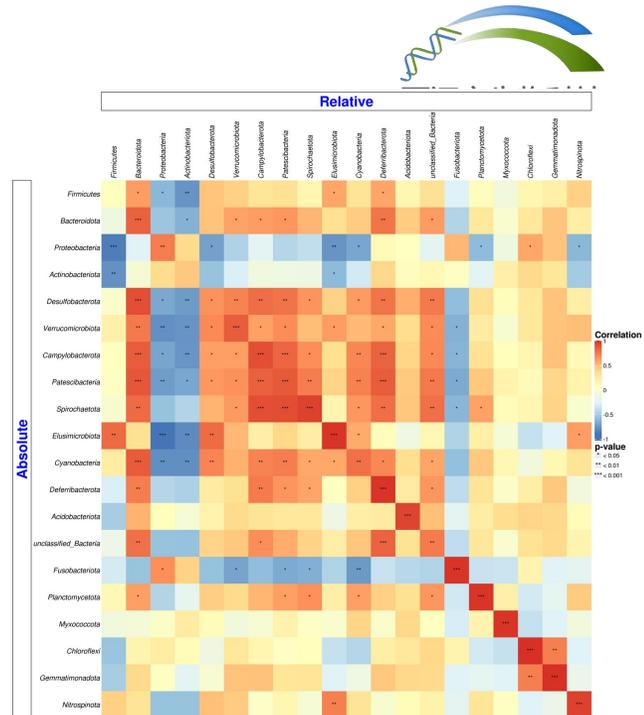
绝对定量研究思路总结



相对/绝对共同解析群落结构



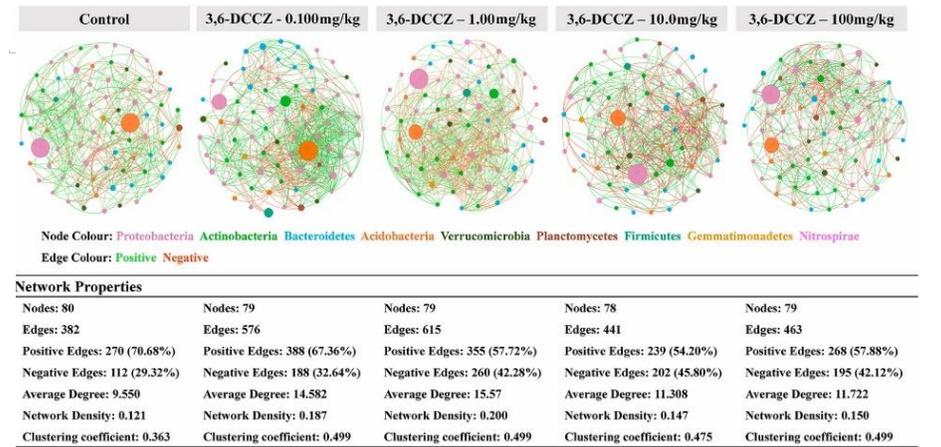
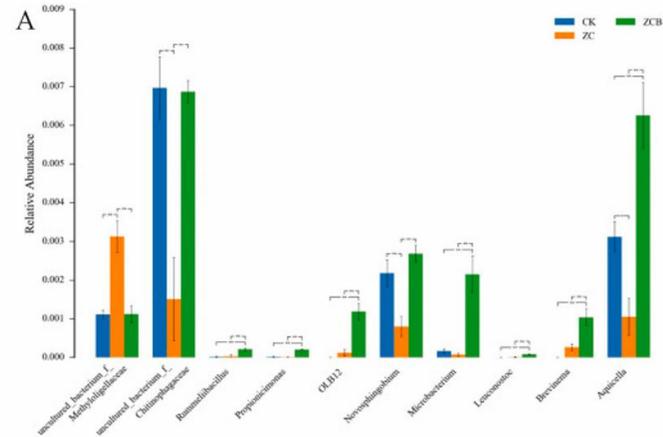
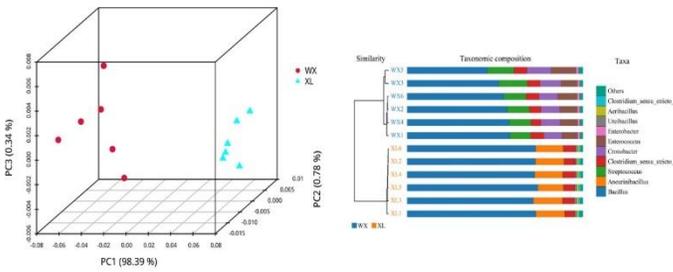
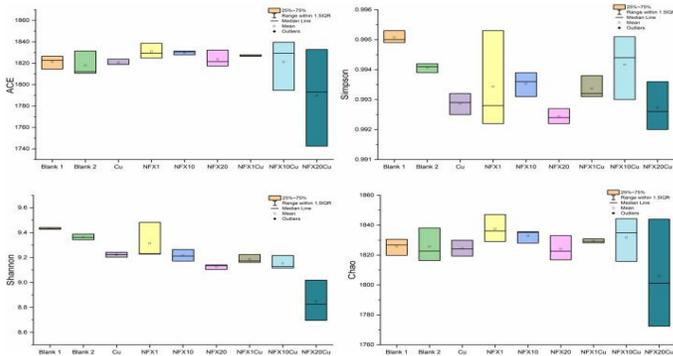
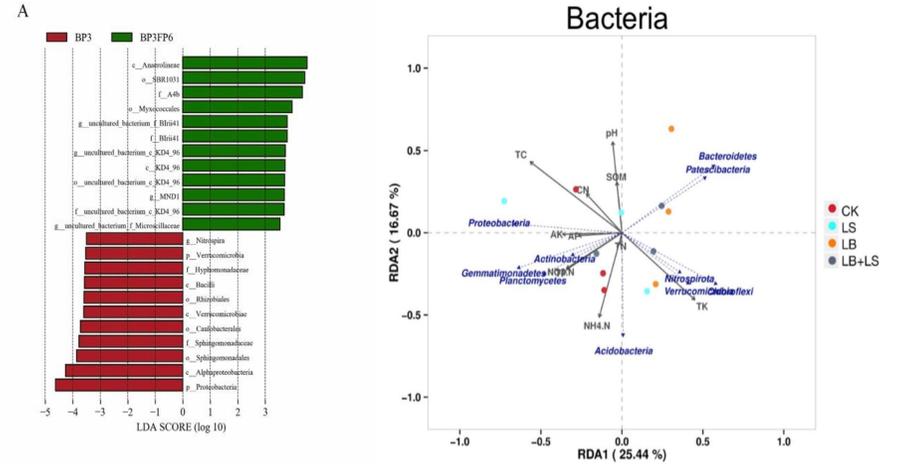
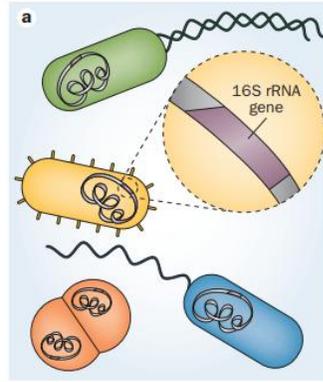
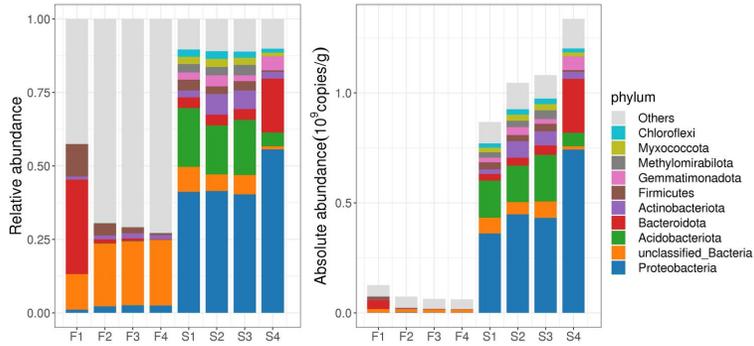
相对丰度（左）与绝对丰度（右）物种分布图



绝对定量与相对定量相关性热图

- ❑ 微生物群落总含量变化，真实还原微生物群落动态变化
- ❑ 优势菌绝对丰度变化情况，优势菌响应环境变化机制
- ❑ 从相对丰度，绝对丰度两个维度进行组间差异分析
- ❑ 相对丰度，绝对丰度相关性分析，研究变化趋势

微生物多样性研究结果



Q1: 微生物绝对定量目前可扩增的区域有哪些?

Q2: 微生物绝对定量目前可不与实验沟通直接送样的样本类型有哪些?

扩增区域与样本类型

产品类型	支持区域	数据量	分析内容	周期
细菌绝对定量 (三代)	16S全长 (V1-V9)	1W CCS	标准曲线绘制 常规分析+绝对定量	样本量: 0-50 30天 提取+建库测序+分析
细菌绝对定量 (二代)	16S V3-V4	5W tags	标准曲线绘制 常规分析+绝对定量	样本量: 0-50 30天 提取+建库测序+分析

组织样类型	送样量
土壤	1~2g
粪便	0.5~2g
粪便(大鼠、小鼠)	3~5粒
备注: 上述组织送样量均为三次提取的量	

1. 样本需要称量g重, 所以无法称重的样本需要注意, 提前咨询。
2. 水体类样本由于无法称重, 所以需要客户告知所收集水体的体积, 后期需要换算到环境样本中的拷贝数信息。

宏基因组研究方法

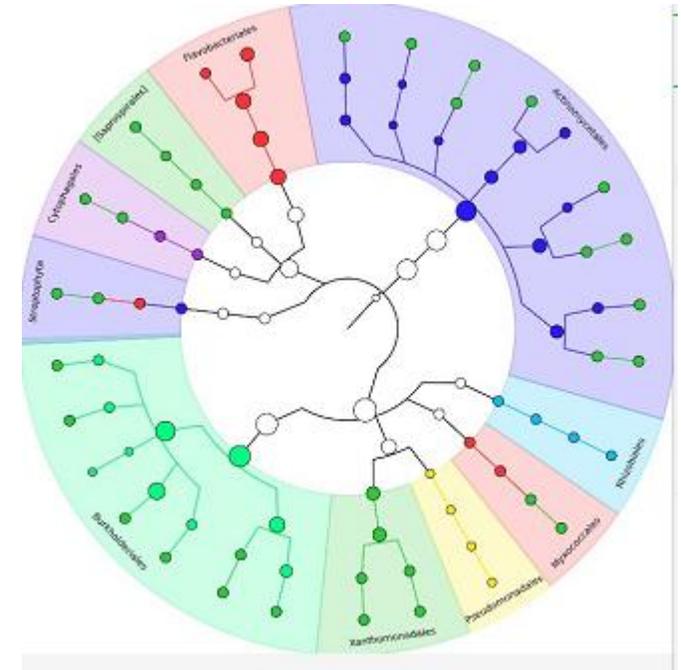
宏基因组 (Metagenome): (也称微生物环境基因组 Microbial Environmental Genome, 或元基因组) 是由 Handelsman 等 1998 年提出的, 其定义为 “the genomes of the total microbiota found in nature”, 即**环境中全部微小生物遗传物质的总和**。它包含了可培养的和不可培养的微生物的基因, 目前主要指环境样品中的**细菌和真菌**等微生物基因组总和。

□ 二代宏基因组

通过二代测序, 同时得到细菌、真菌种群结构与功能信息, 分辨率超“种水平”。深入了解微生物功能与环境互作机制。

□ ONT宏基因组

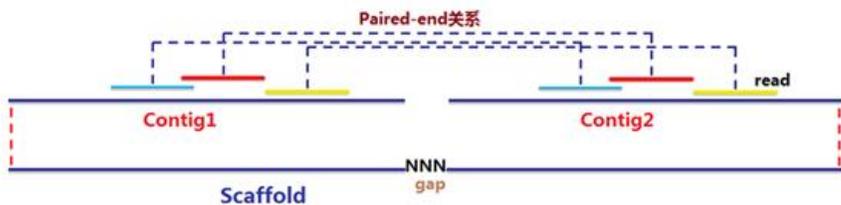
通过二代和三代ONT测序, 跨越复杂基因结构, 组装更高质量宏基因组, 使得基因结构更完整, 物种及功能基因预测精准度更高。



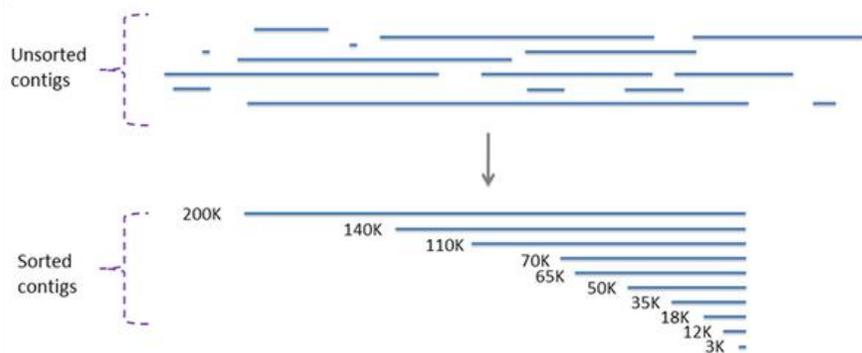
宏基因组研究技术路线

技术路线

基因组组装 Scaffold、Contig



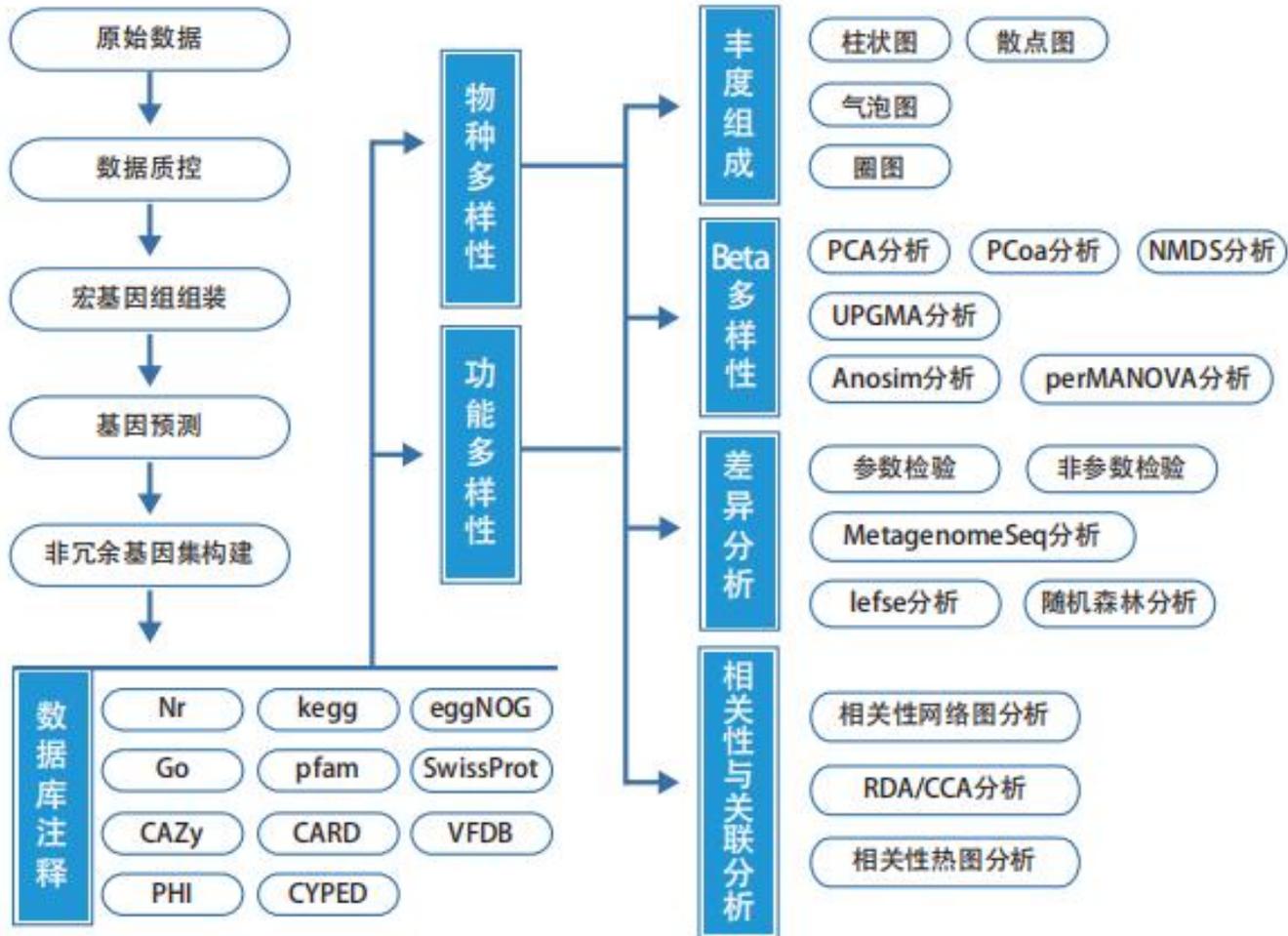
ScaffoldN50 ContigN50



Total contig length= 200K + 140K + 110K + 70K + 65K + 50K + 35K + 18K + 12K + 3K= 703K

50% total contig length= 703K x 50%= 351.5K

∵ 200K+140K + 110K > 351.5K ∴ N50= 110K



原始数据

数据质控

宏基因组组装

基因预测

非冗余基因集构建

物种多样性

功能多样性

数据库注释

Nr

kegg

eggNOG

Go

pfam

SwissProt

CAZy

CARD

VFDB

PHI

CYPED

丰度组成

Beta多样性

差异分析

相关性 & 关联分析

柱状图

散点图

气泡图

圈图

PCA分析

PCoA分析

NMDS分析

UPGMA分析

Anosim分析

perMANOVA分析

参数检验

非参数检验

MetagenomeSeq分析

lefse分析

随机森林分析

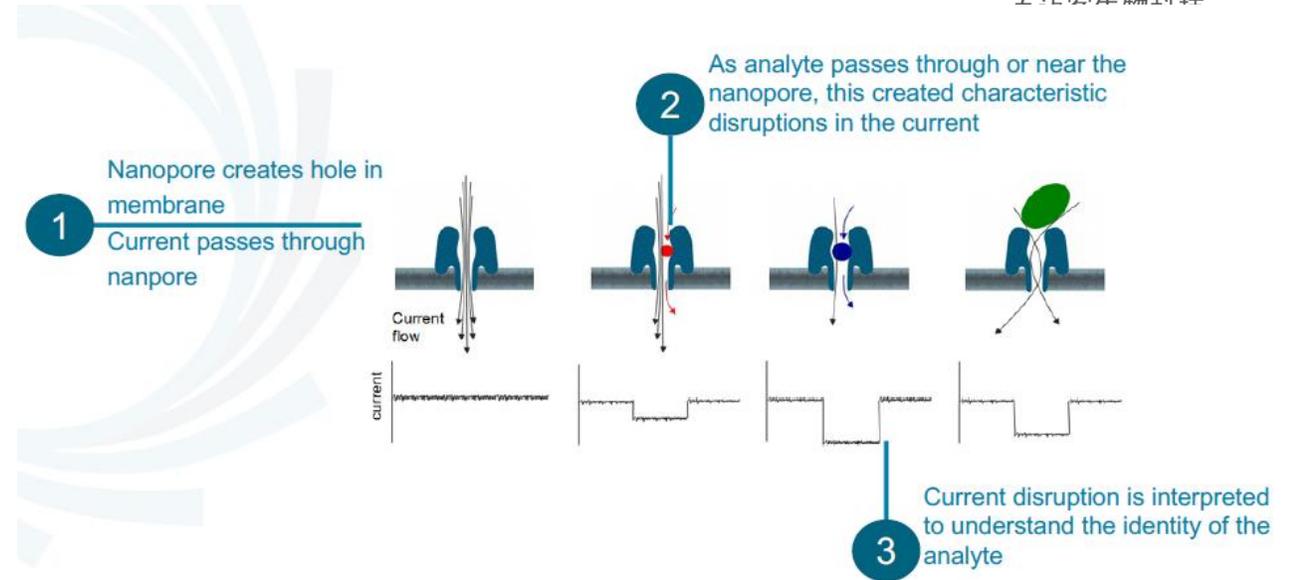
相关性网络图分析

RDA/CCA分析

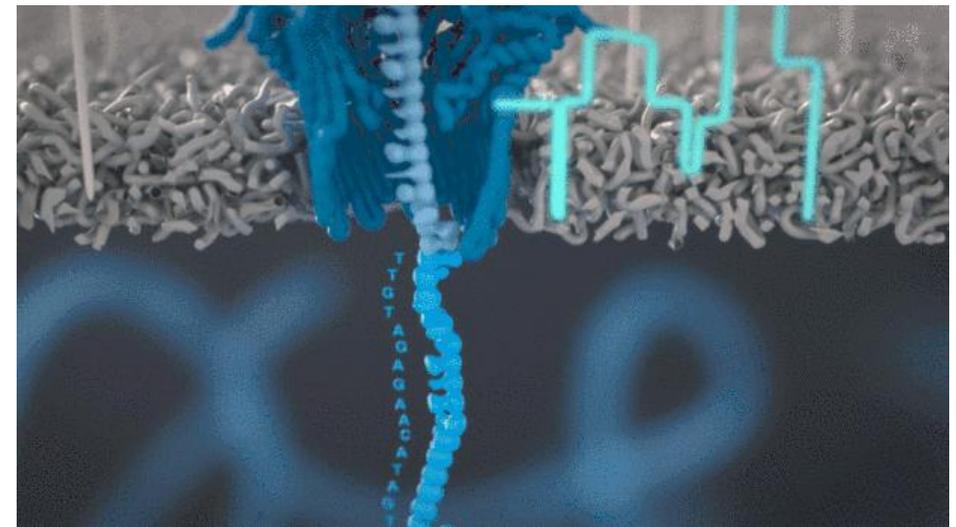
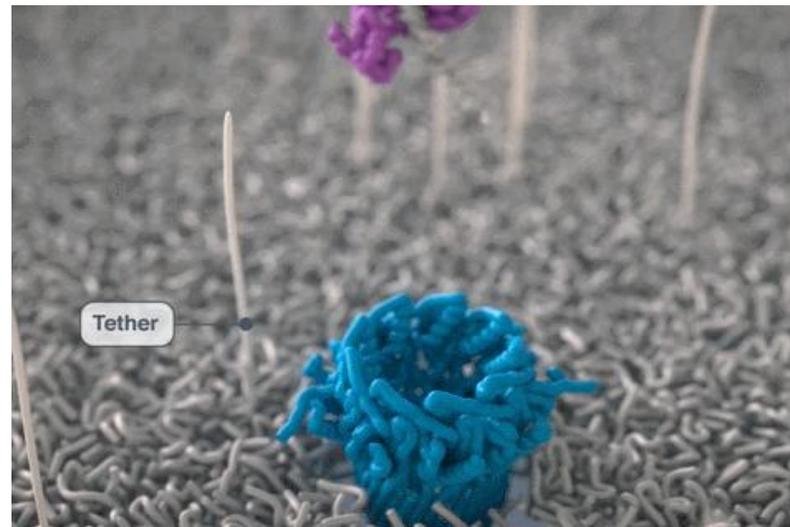
相关性热图分析

Nanopore测序原理和优势

- 长读长：Nanopore 技术的测序平均读长能够达到几十到上百 Kb，最长读长能达到2 Mb以上级别；
- 跨越重复区：长读长可跨越重复序列区域，更有利于基因组组装；
- 低成本：相比其他三代测序技术，ONT 测序样本处理极其简单，无需DNA聚合酶、连接酶和dNTPs，测序价格低；
- 无PCR偏好：不进行PCR扩增，避免二代测序中PCR扩增可能引入的错误。



PromethION



ONT宏基因组优势



环节	比较点	NGS宏基因组	ONT宏基因组
测序	测序原理	边扩增合成边测序	无需扩增，边解链边测序
	测序读长	PE150	平均10Kb，最长M级别，可跨越重复区域
	测序单价	低	中
	序列准确度	Q20>85%	众数准确度98%左右
组装	组装原理	复杂的De Bruijn Graph algorithm	简单的Overlap-Layout-Consensus algorithm
	组装N50	~ 1 Kbp	~100 K bp (100倍)
	最长组装长度	几百Kbp	Mbp 单菌基因组级别
	数据矫正	无需矫正	需二代序列辅助矫正
注释	物种注释	读长短，注释准确度低	读长更长，信息量更多，注释更全面准确
	基因预测	基因碎片化，长度短，数量冗余	基因更完整，长度长，数量更契合真实
	蛋白注释	-	由于基因更长更完整，注释更准确
深度挖掘	Binning	草图很碎(几百条contigs)，完整度低	(准)完整细菌基因组(几十~1条环形)
	抗性基因簇	碎片化的抗性基因(簇)	相对完整的抗性基因(簇)
	病毒序列等	碎片化的病毒序列	相对完整的病毒序列

Nanopore宏基因组技术路线与实测数据

Nanopore宏基因组组装指标较二代组装提升两个数量级，最长contig可达MB级别

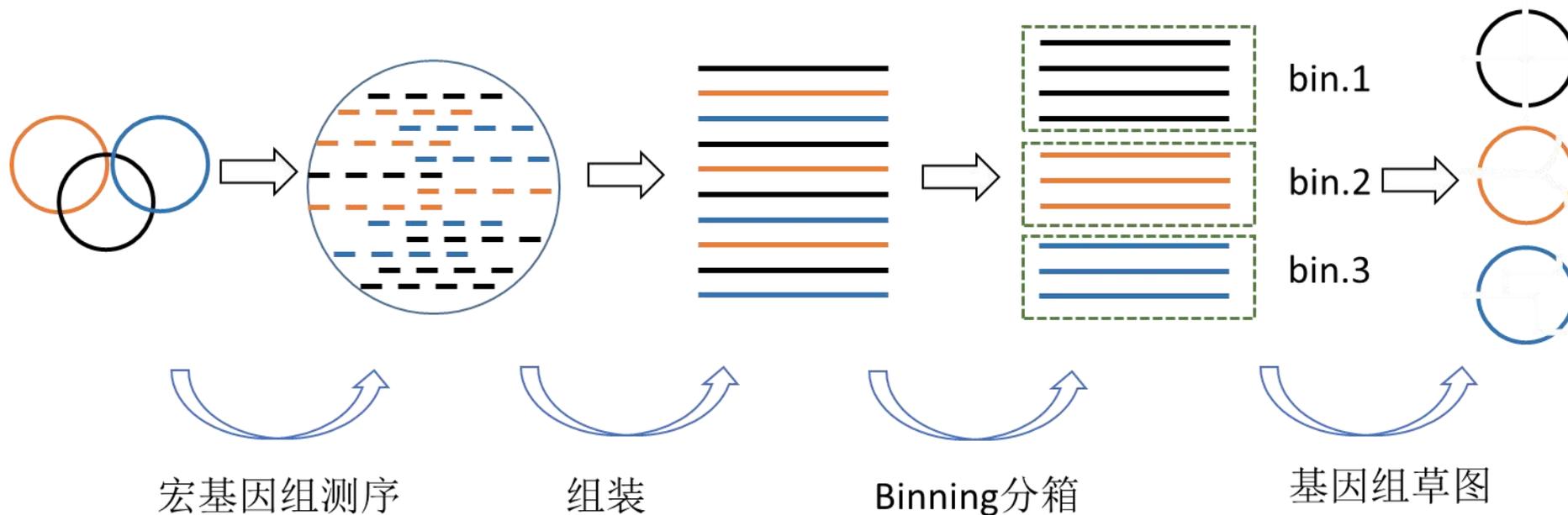
Sample ID	TotalNum	CtgNum	CtgLen	CtgN50	CtgMax	GC(%)
某水体	1237	1234	103581918	102994	655097	46.91
某粪便	1781	1779	145624839	99853	2311559	46.35
某肠道	926	926	89245723	126756	1684291	47.84
某水体	2970	2970	189527073	76148	1534036	47.72
某土壤	3535	3532	199161026	67022	2089238	47.68

宏基因组高级分析-宏基因组binning

Binning:

对序列进行聚类、分箱，根据基因组特征以及组装信息等将属于不同基因组的序列分离开来的过程。

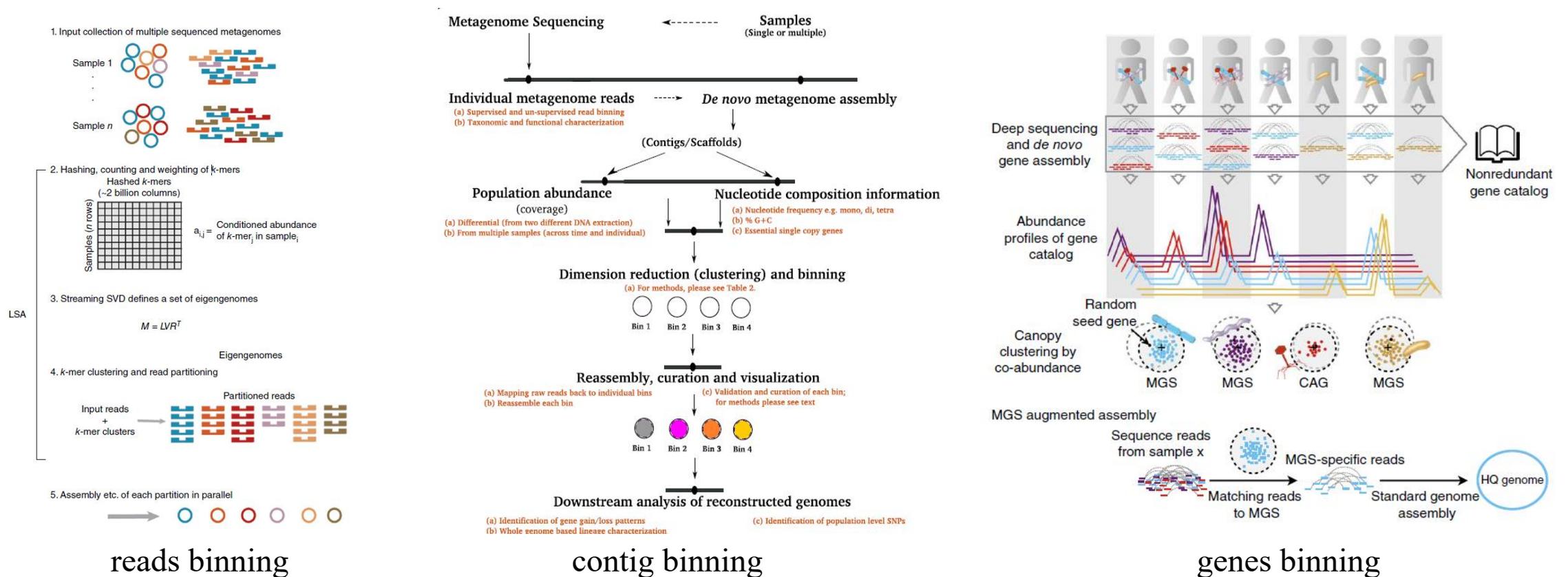
通过binning得到的bins，很可能是实验室无法纯培养的未知微生物，从而实现从环境微生物群落中构建菌株水平(或更高级别分类单元)的草图基因组。



宏基因组高级分析-宏基因组binning

根据基于聚类的序列类型的不同，分为 reads binning, **contig binning** 和 genes binning。基于 reads binning 的优势是可以聚类出宏基因组中丰度非常低的物种。基于 genes binning 目前已发表的宏基因组关联分析 (MWAS) 和多组学联合分析文章中，尤其是疾病的 MWAS 研究中基本都用 genes binning。

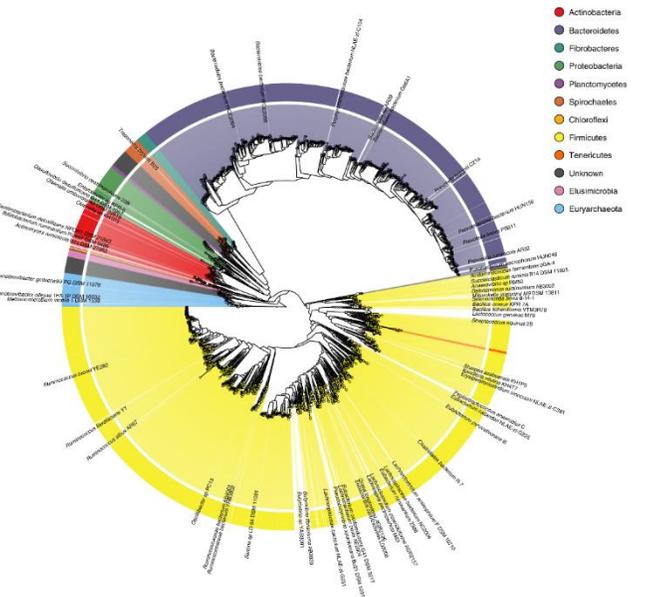
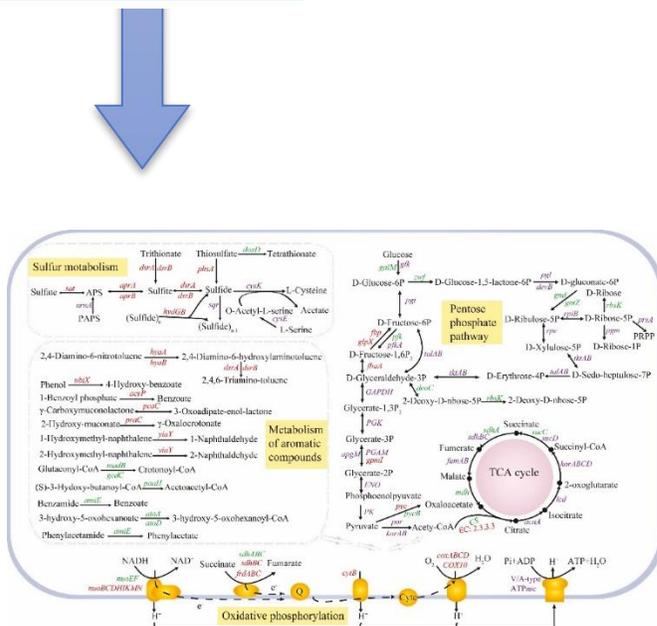
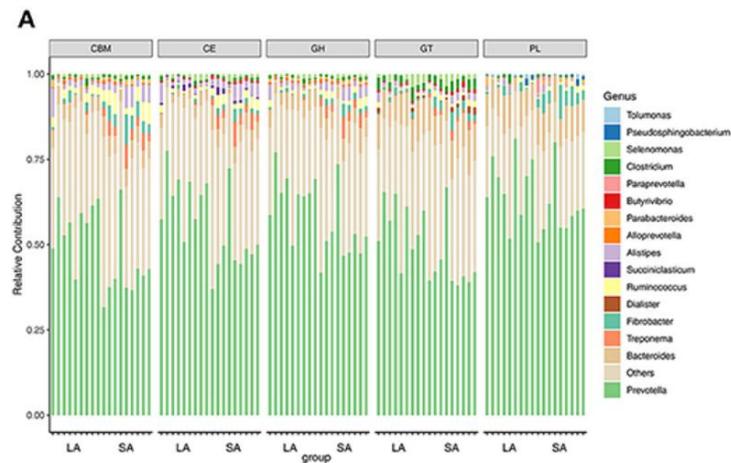
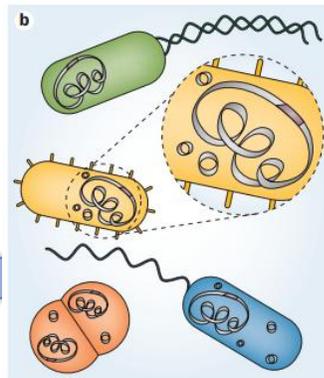
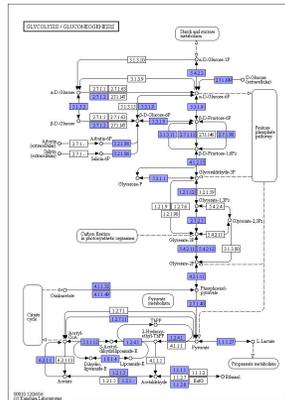
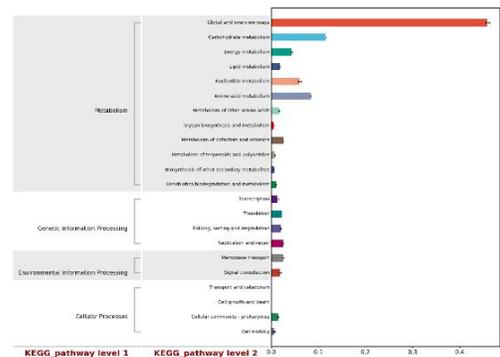
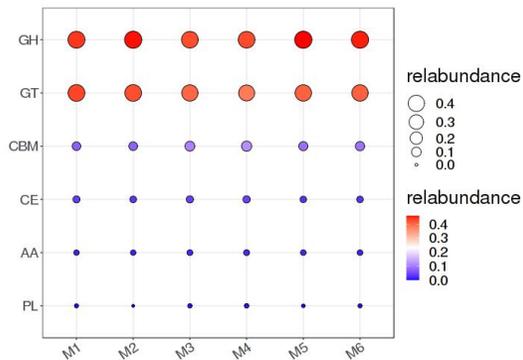
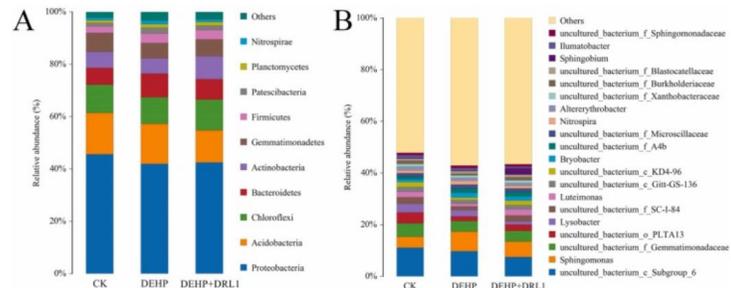
由于核酸组成和物种丰度变化模式在越长的序列中越显著和稳定，目前越多的研究发现，基于 contig binning 效果可能更好，现在的很多软件也大都是基于 contig binning。



宏基因组研究



环境中全部微生物遗传物质DNA的总和，同时得到微生物群落结构组成与功能信息，分辨率超“种水平”。深入了解微生物功能与环境互作机制。

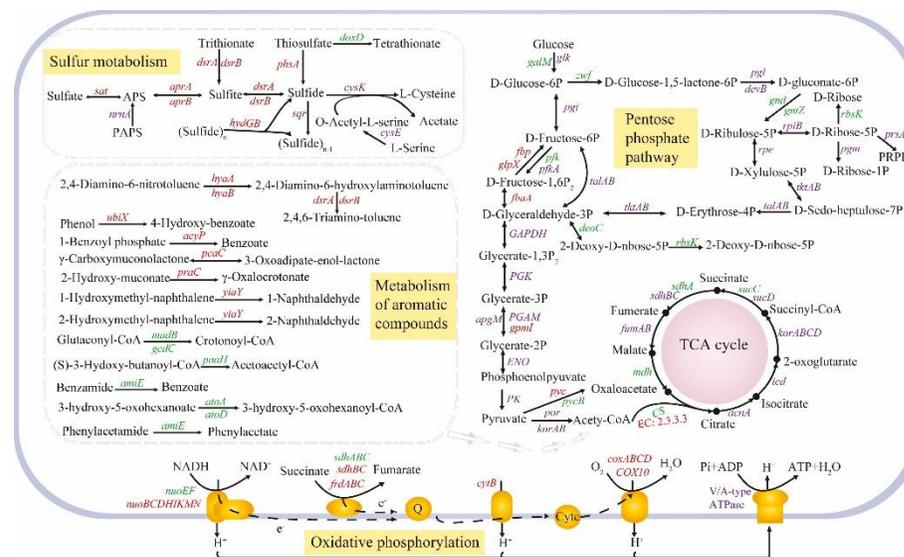
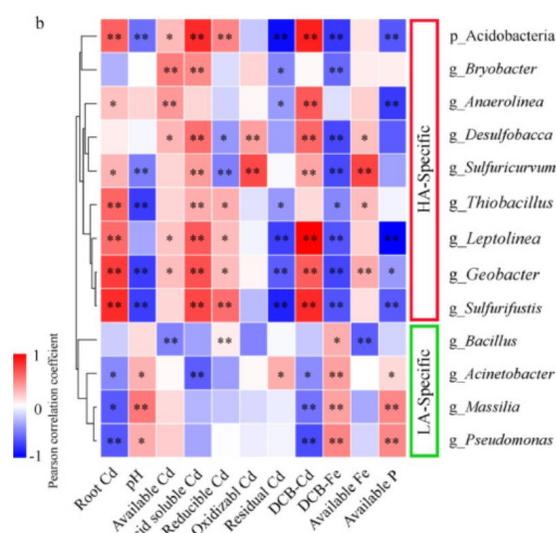
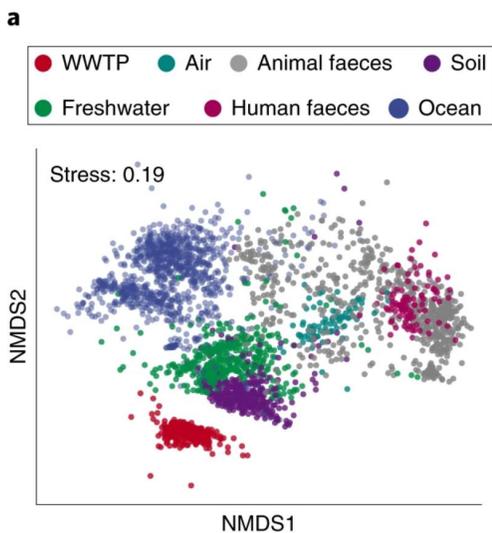
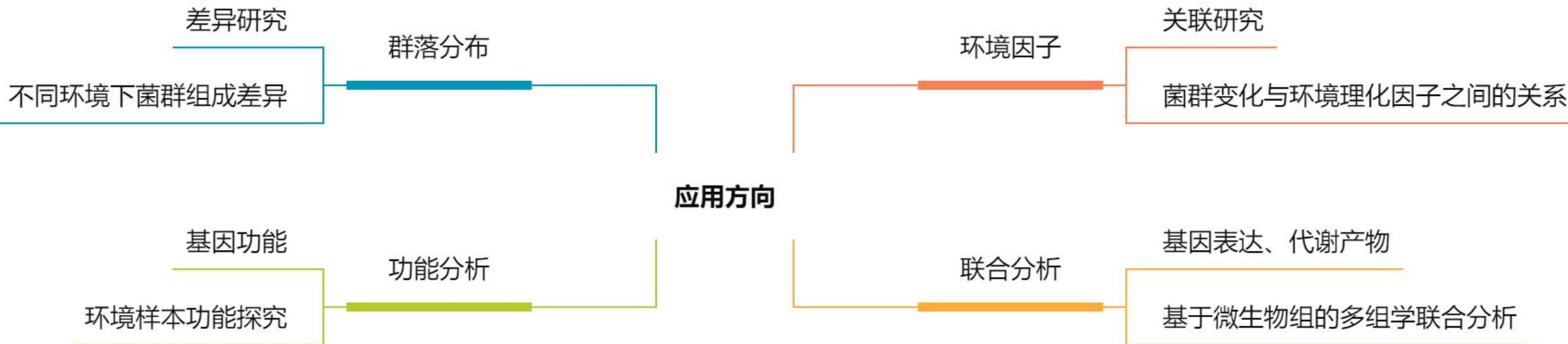


环境微生物组学研究技术比较



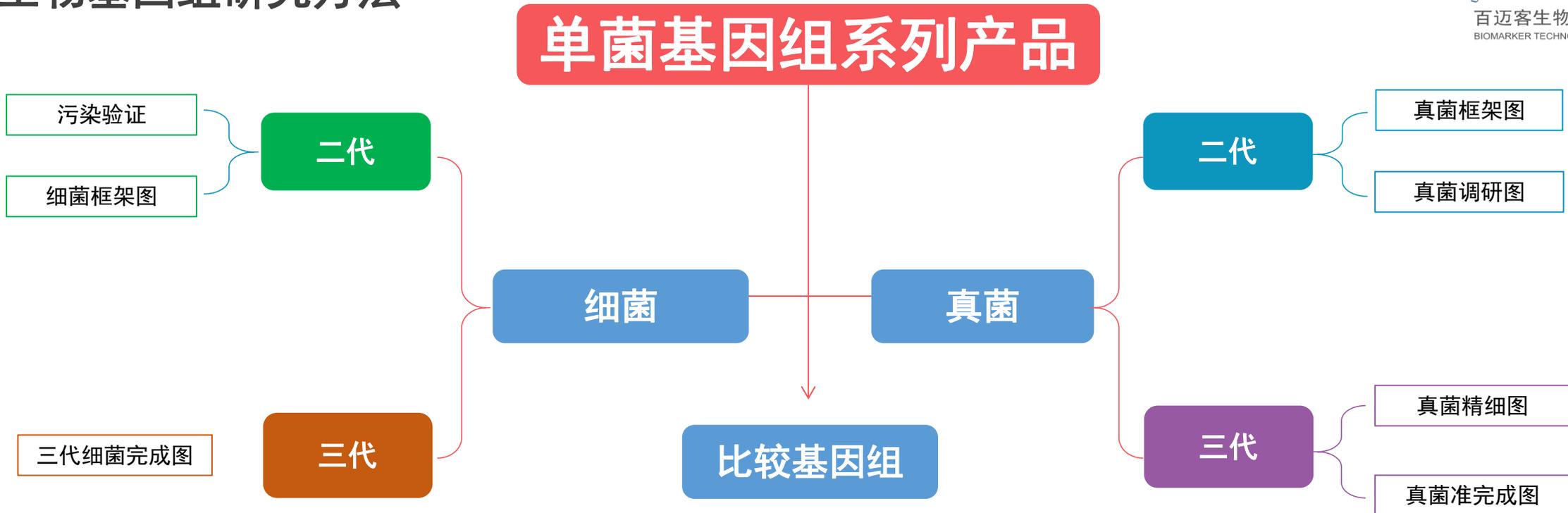
产品类型	研究内容	优势	技术限制
二代微生物多样性	物种种类 物种相对丰度 Alpha多样性 Beta多样性 RDA/CCA分析	成本低，分析简单，周期短	PE250/PE300 测序区域少，V3+V4等 “属水平” 引物设计和可变区选择的偏好影响
三代微生物多样性	组间显著性差异分析 功能基因预测	“种水平” “种水平”平均注释率 $\geq 60\%$ 测序区域全 V1--V9	成本高
二代宏基因组	物种种类 物种相对丰度 功能基因种类 功能基因相对丰度	环境样本中所有微生物基因组 分辨率超“种水平” 可组装获得细菌基因组信息 可进行基因家族挖掘	对测序深度要求高 分析准确度依赖组装质量
三代宏基因组	Alpha多样性 Beta多样性 RDA/CCA分析 组间显著性差异分析 Binning高级分析	环境样本中所有微生物基因组 无PCR扩增的偏好性 测序读长长，可跨越复杂的重复序列 组装指标较二代提升两个数量级 更容易组装出细菌基因组	成本高 单碱基准确率 需要二代数据纠错 对测序数据量要求高

环境微生物组学应用方向



3

单菌基因组



Illumina 测序平台



PacBio (左), Nanopore测序平台

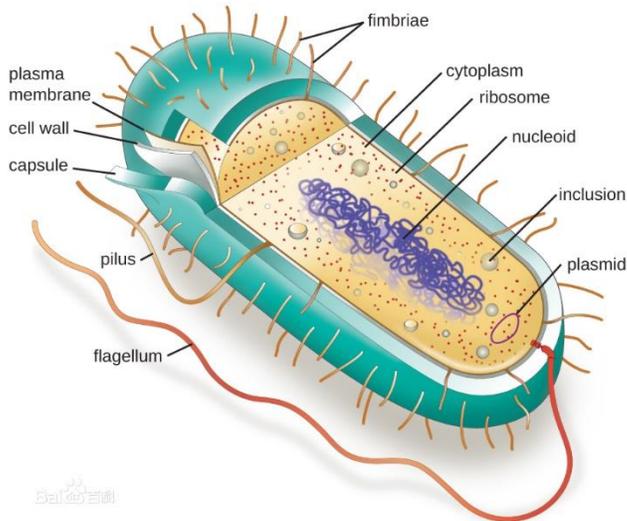
细菌基因组

细菌：属于**原核生物**，无核膜、DNA裸露，分**真细菌**和**古细菌**两大类的微生物。是在自然界分布最广、个体数量最多的有机体，是大自然物质循环的主要参与者。

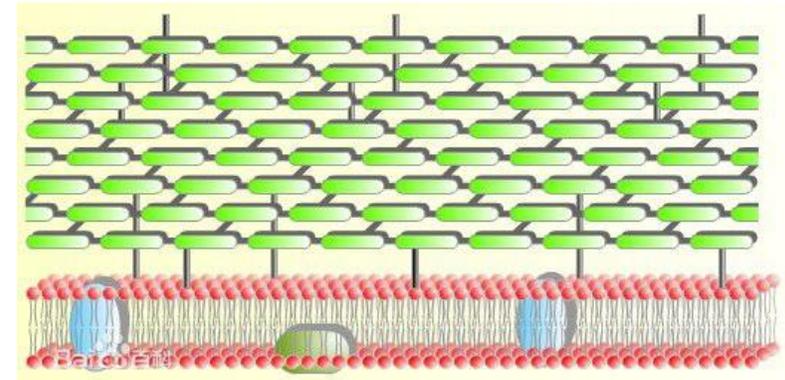
基因组大小：一般在**0.16-13Mb**，大部分在**5M**左右。放线菌 **8M**左右；

编码序列：占总长度的90%，基因平均大小为 1kb；

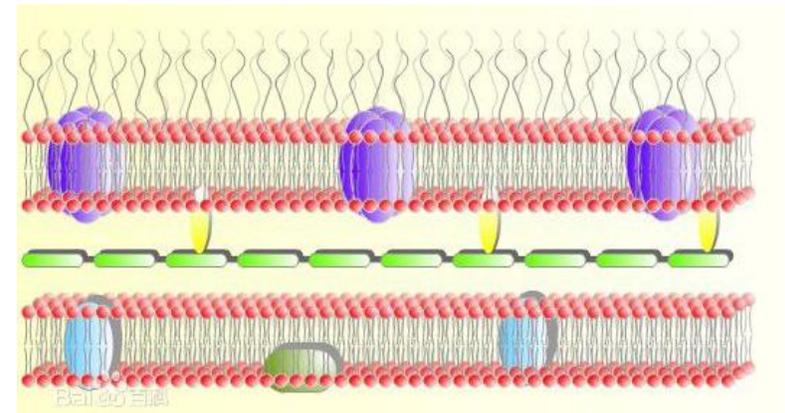
结构特点：多为一条环状闭合双链DNA，无内含子；



一个典型的原核生物基因的结构



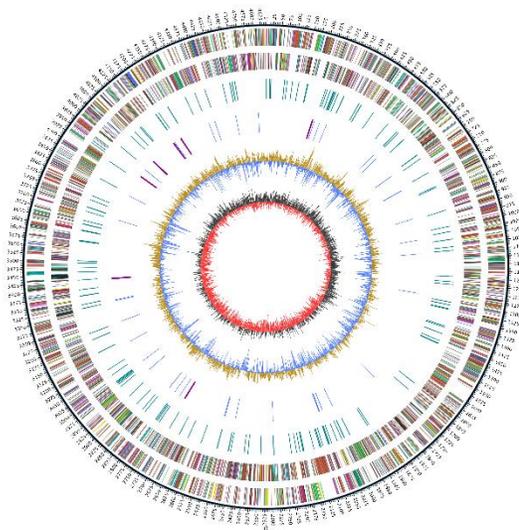
G+细胞壁示意图



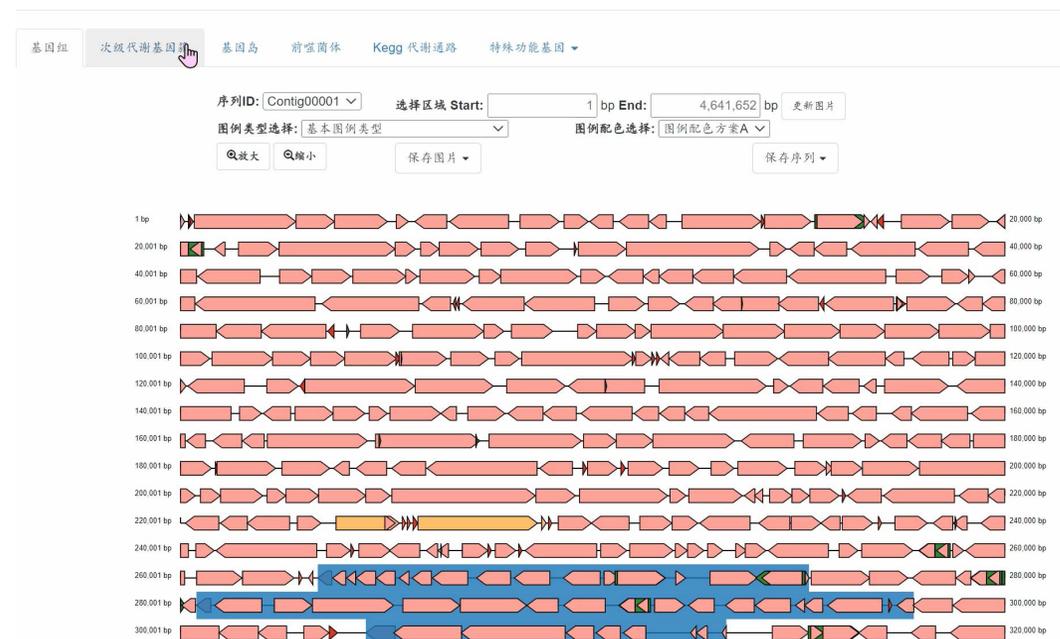
G-细胞壁示意图

细菌基因组产品类型

产品类型	测序平台	总测序深度	项目周期	承诺指标
细菌基因组框架图	Illumina	≥100 X	35天	/
ONT细菌基因组完成图	Illumina + ONT	二代 ≥ 100 X ONT ≥ 100 X	35天	0gap
PB细菌基因组完成图	Illumina + PacBio	二代 ≥ 100 X PB HiFi ≥ 30 X	35天	0gap

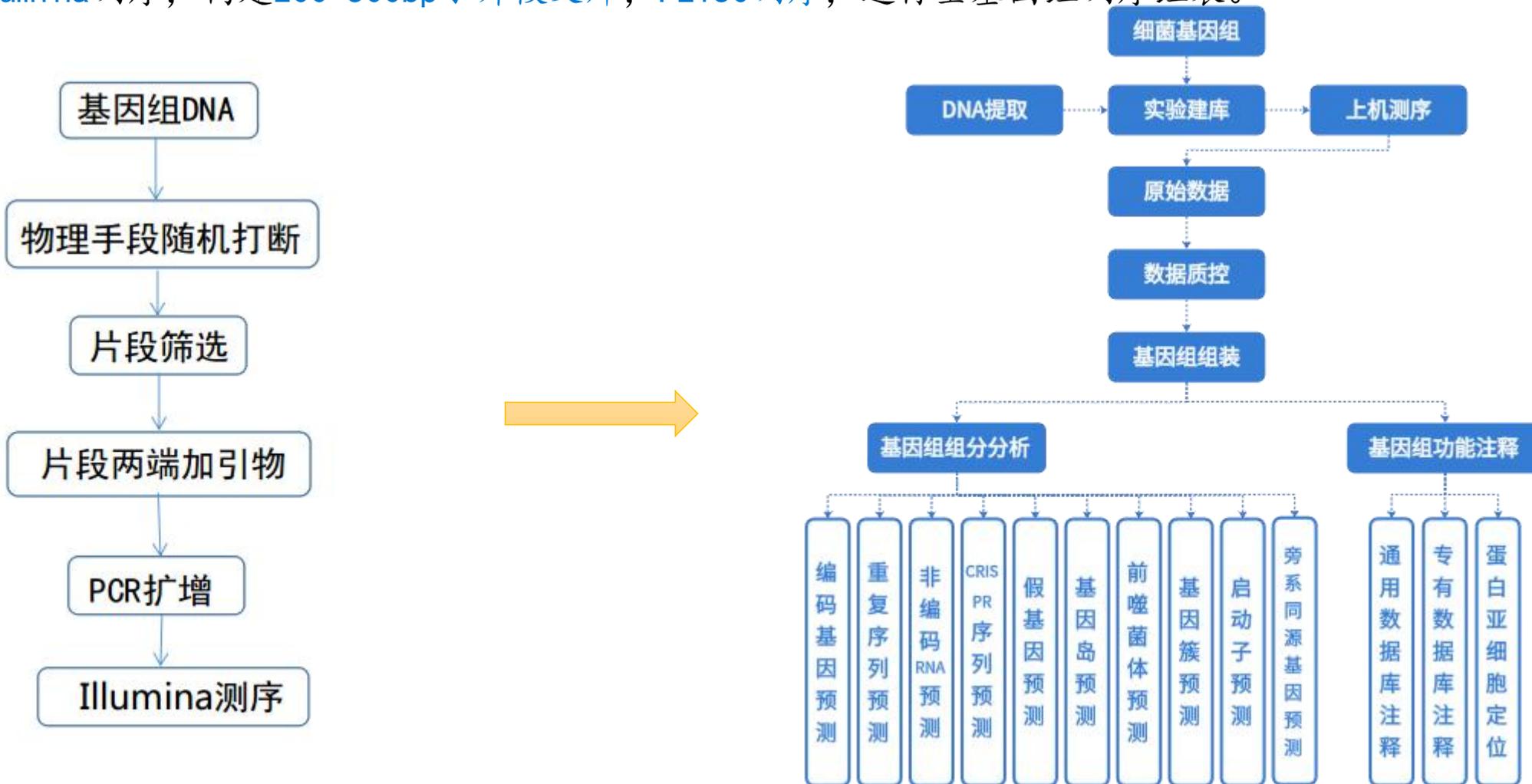


- A: RNA processing and modification(3)
- B: Chromatin structure and dynamics(0)
- C: Energy production and conversion(305)
- D: Cell cycle control, cell division, chromosome partitioning(40)
- E: Amino acid transport and metabolism(344)
- F: Nucleotide transport and metabolism(91)
- G: Carbohydrate transport and metabolism(343)
- H: Coenzyme transport and metabolism(116)
- I: Lipid transport and metabolism(91)
- J: Translation, ribosomal structure and biogenesis(176)
- K: Transcription(302)
- L: Replication, recombination and repair(251)
- M: Cell wall/membrane/envelope biogenesis(267)
- N: Cell motility(65)
- O: Posttranslational modification, protein turnover, chaperones(148)
- P: Inorganic ion transport and metabolism(243)
- Q: Secondary metabolites biosynthesis, transport and catabolism(48)
- R: General function prediction only(333)
- S: Function unknown(661)
- T: Signal transduction mechanisms(134)
- U: Intracellular trafficking, secretion, and vesicular transport(56)
- V: Defense mechanisms(45)
- W: Extracellular structures(0)
- Y: Nuclear structure(0)
- Z: Cytoskeleton(0)
- Not Cog annotated(257)



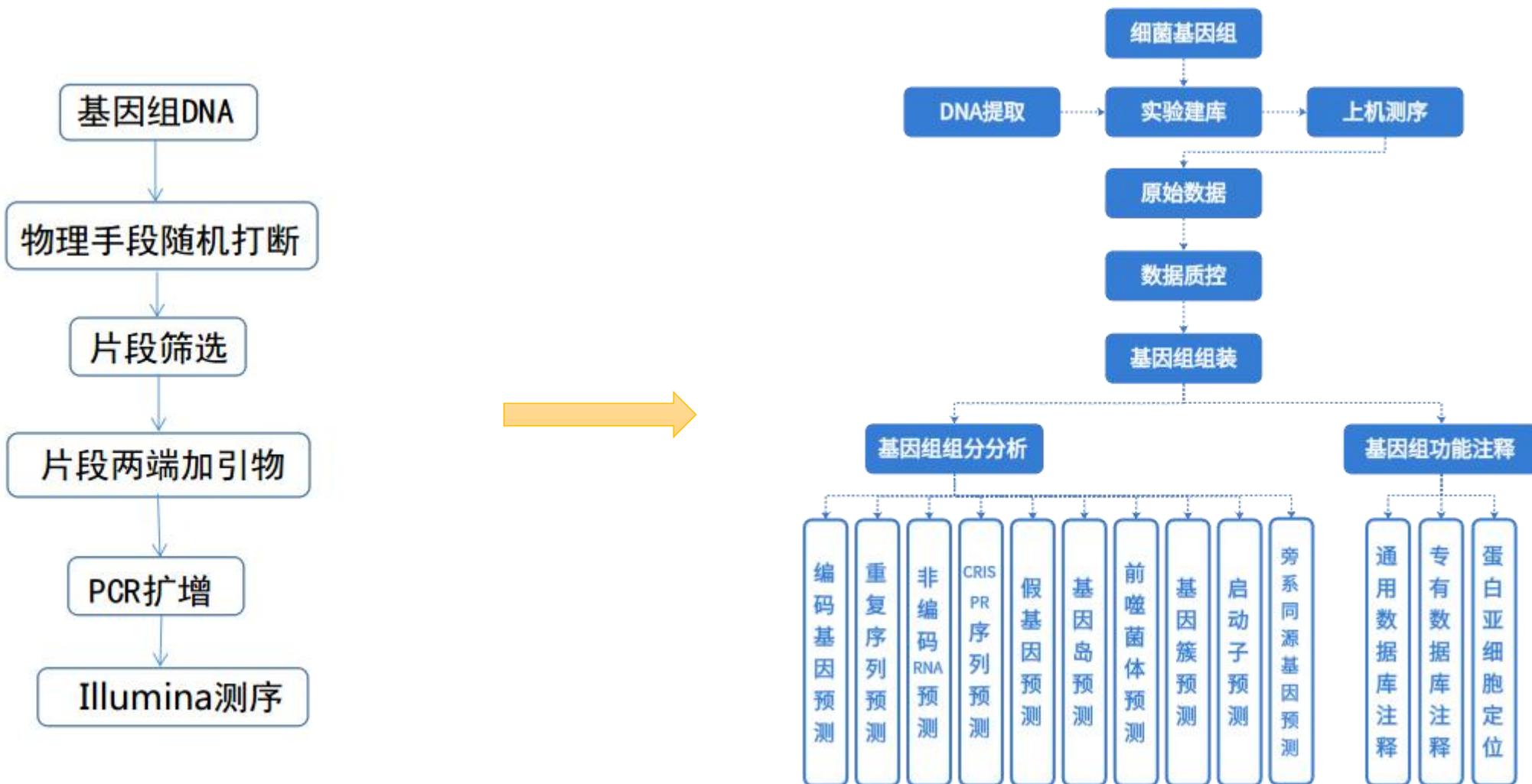
细菌基因组框架图

ILLUMINA测序，构建200-500bp小片段文库，PE150测序，进行全基因组测序组装。



细菌基因组框架图

Illumina测序，构建200–500bp小片段文库，PE150测序，进行全基因组测序组装。

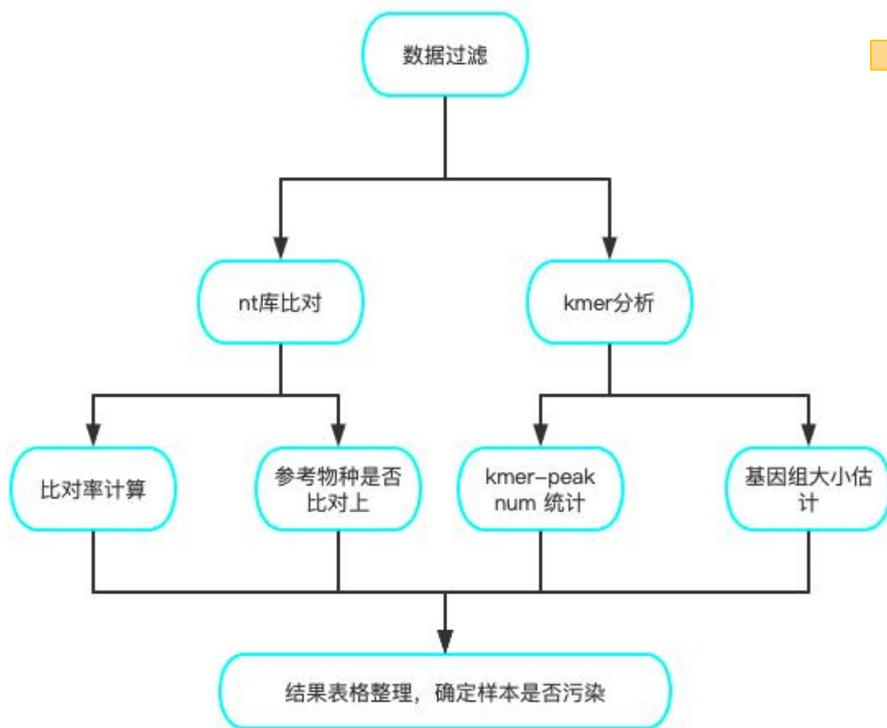


细菌完成图

污染验证

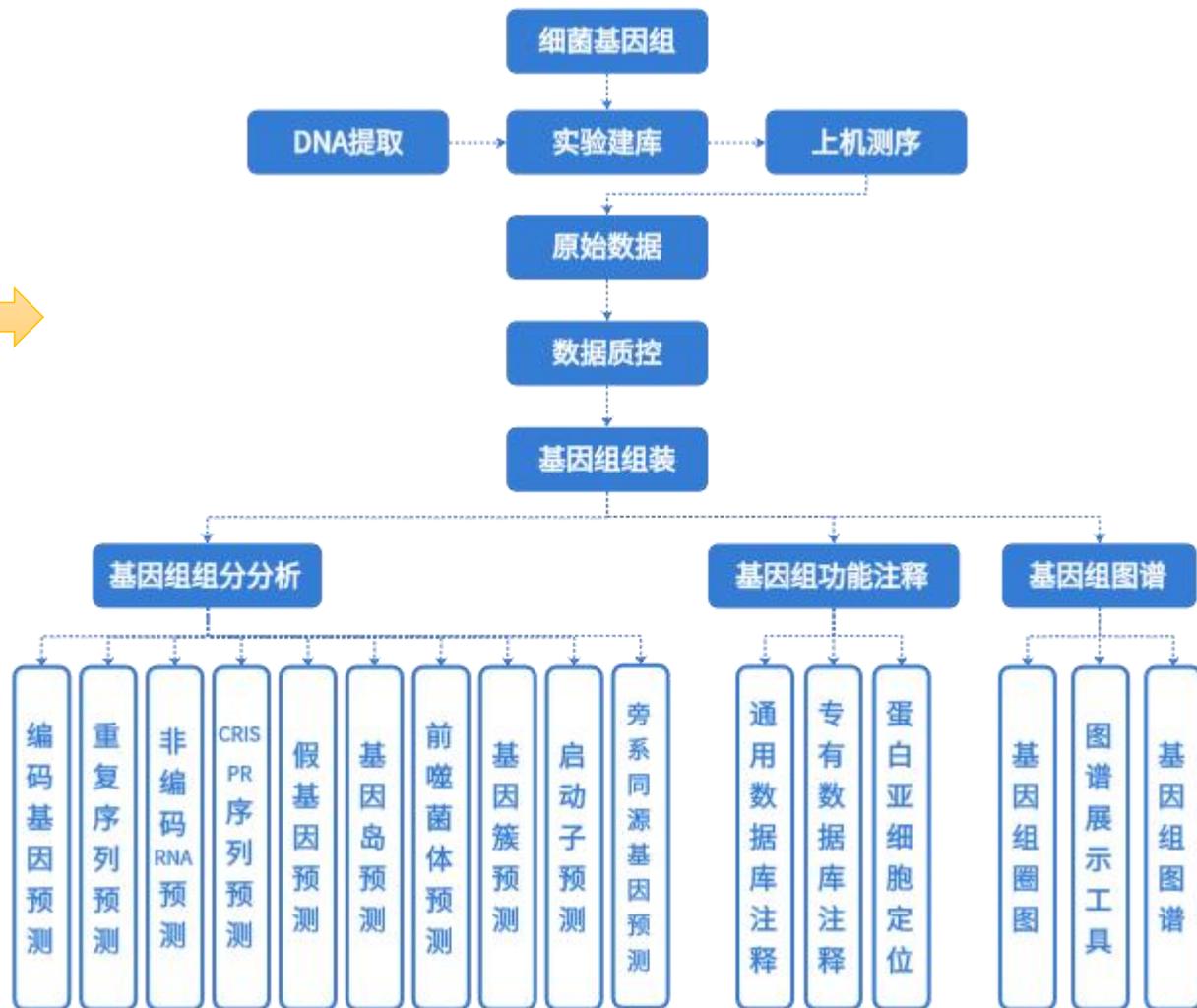
二代 20天

目的：通过进行nt库比对和kmer分析，来**鉴定测序数据有无污染，是不是目标菌**。用于三代数据辅助纠错。



完成图分析

三代 35天



细菌框架图与完成图

框架图

表格1 基因组组装结果统计

SampleID	Scaffold Length (bp)	Scaffold Number	Scaffold N50 (bp)	Scaffold N90 (bp)	Contig Length (bp)	Contig
██████	2339536	43	125570	36146	2339152	44

表格1 基因组组装结果统计

SampleID	Scaffold Length (bp)	Scaffold Number	Scaffold N50 (bp)	Scaffold N90 (bp)	Contig Length (bp)	Contig
██████	2611712	45	111560	36190	2611424	46

表格1 基因组组装结果统计

SampleID	Scaffold Length (bp)	Scaffold Number	Scaffold N50 (bp)	Scaffold N90 (bp)	Contig Length (bp)	Contig
██████	2847366	31	177497	57654	2846952	32

完成图

表格3 基因组组装结果统计

SampleID	Scaffold Length (bp)	Scaffold Number	Scaffold N50 (bp)	Scaffold N90 (bp)	Contig Length (bp)	Contig
██████	2829752	1	2829752	2829752	2829752	1

表格3 基因组组装结果统计

SampleID	Scaffold Length (bp)	Scaffold Number	Scaffold N50 (bp)	Scaffold N90 (bp)	Contig Length (bp)	Contig
██████	4230437	1	4230437	4230437	4230437	1

表格3 基因组组装结果统计

SampleID	Scaffold Length (bp)	Scaffold Number	Scaffold N50 (bp)	Scaffold N90 (bp)	Contig Length (bp)	Contig
██████	2168965	1	2168965	2168965	2168965	1

真菌基因组

真菌基因组主要指各种真菌，包括各种酵母，霉菌，食用真菌等。一般在**10M-150M**，大部分是**50M**左右。复杂真菌有多核，杂合等现象。



大型真菌与单细胞真菌

真菌基因组产品类型

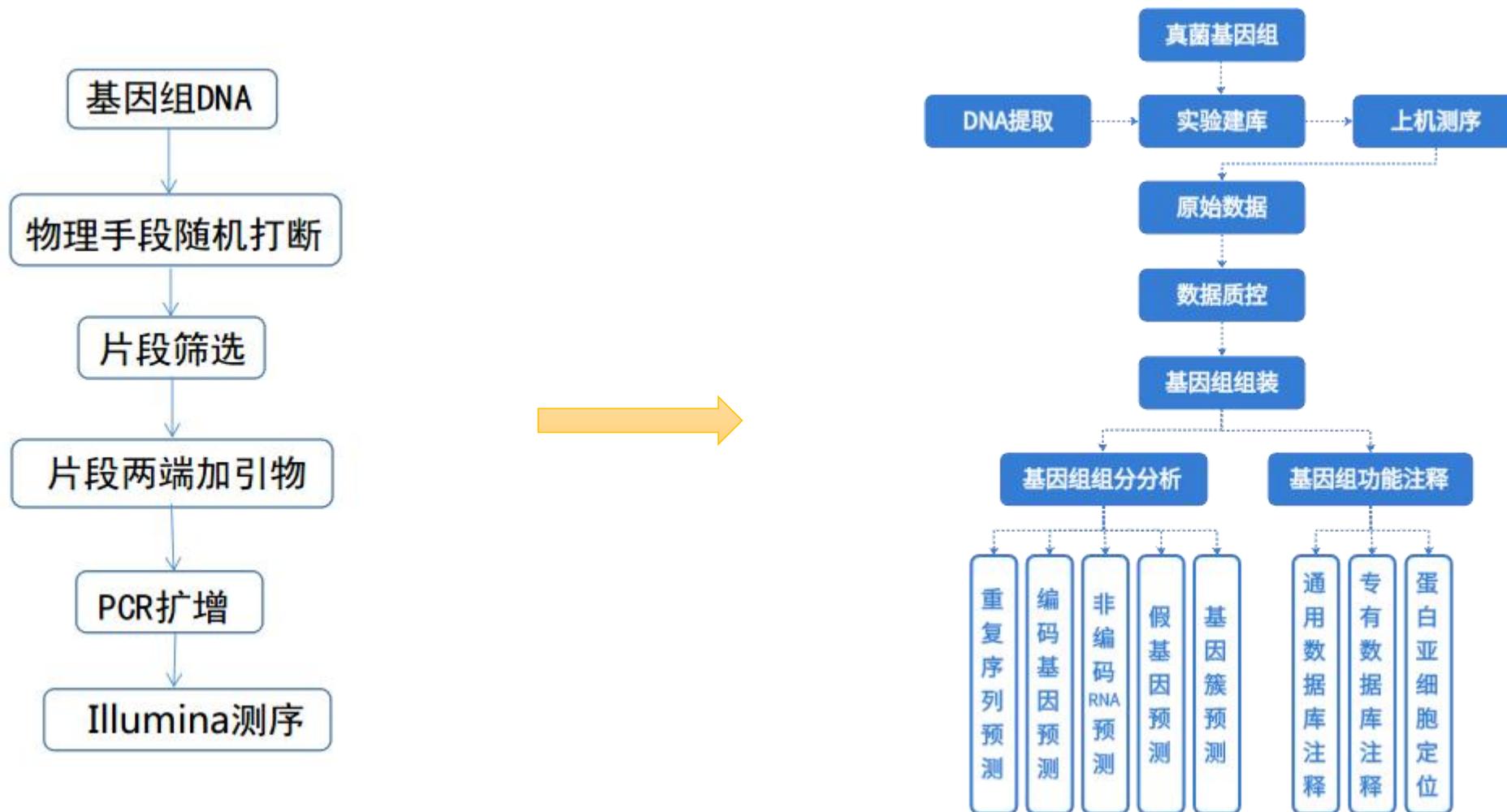


产品类型	测序平台	总测序深度	项目周期	承诺指标
真菌基因组框架图	Illumina	$\geq 100 X$	30天	初步组装，快速获得真菌基因组草图序列，和功能信息
真菌基因组调研图	Illumina	$\geq 100 X$	30天	基本调研，确定基因组大小，杂合度，重复序列等
真菌基因组精细图	Illumina ONT/Pacbio	二代 $\geq 50 X$ ONT $\geq 100 X$ / PB HiFi $\geq 30 X$	45天	基因组小于100M contig N50 $\geq 2Mb$ (单核) contig N50 $\geq 500kb$ (多核)
真菌基因组准完成图	Illumina ONT/Pacbio HiC	精细图+100 X HiC	90天	染色体挂载率90%

目前我们公司完成的Hi-C项目：白灵菇、木霉菌、酸腐菌、酵母、炭疽菌

真菌框架图

Illumina测序，构建200–500bp小片段文库，PE150测序，进行全基因组测序组装。

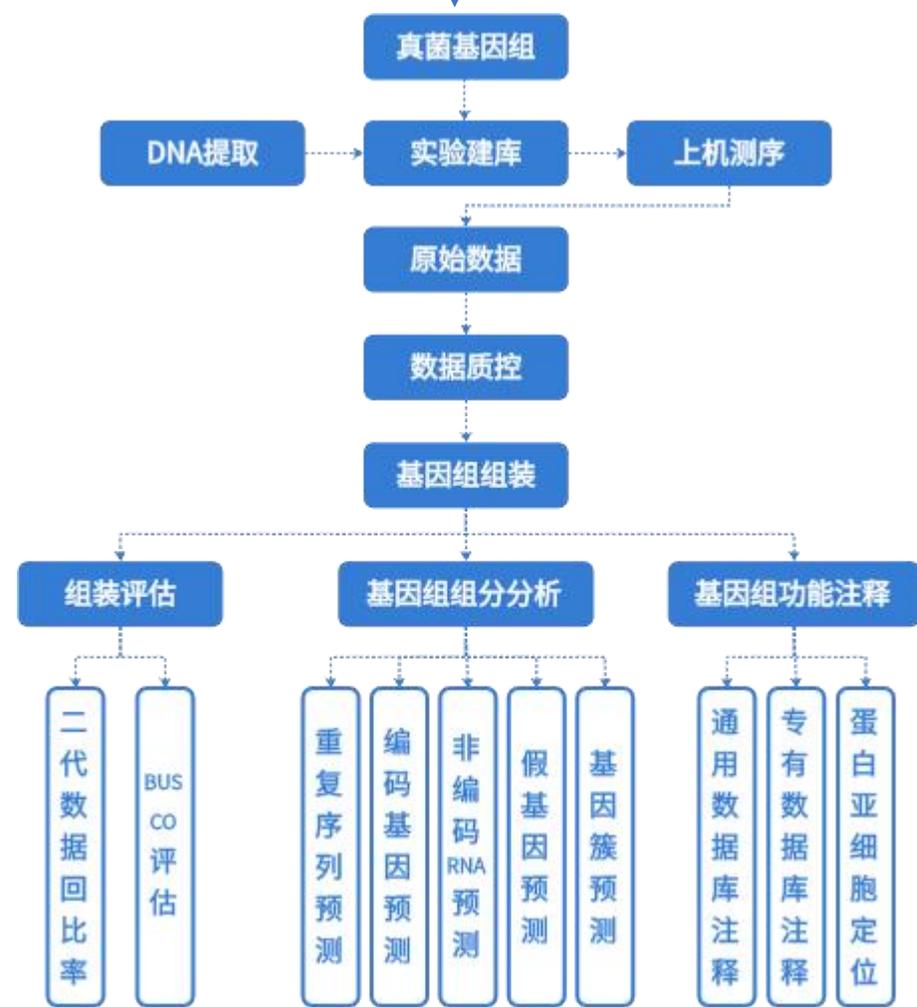
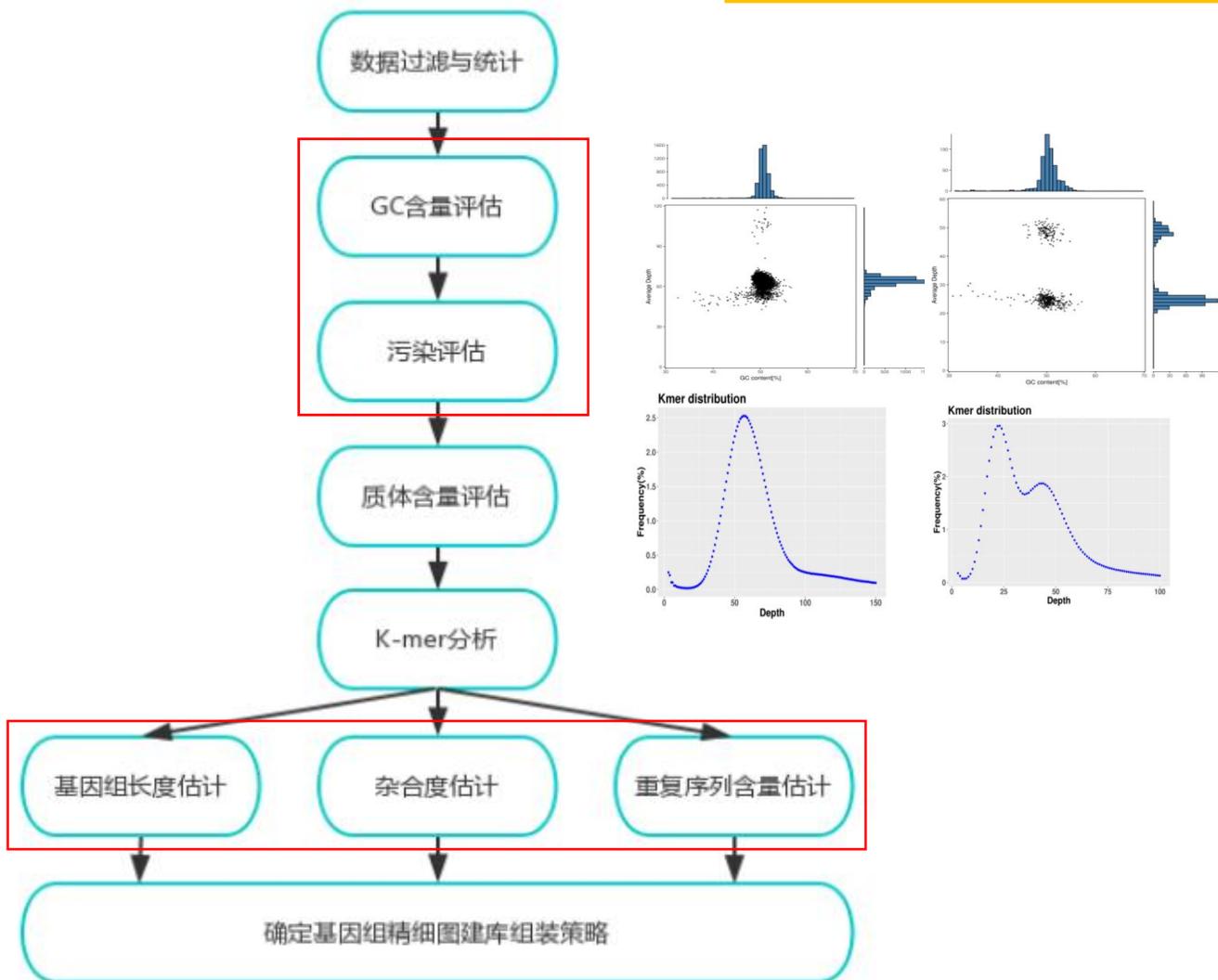


真菌精细图

调研图-二代

确定真菌基因组大小、杂合度和重复序列，用于三代辅助纠错
多核样本建议单核化

精细图-三代

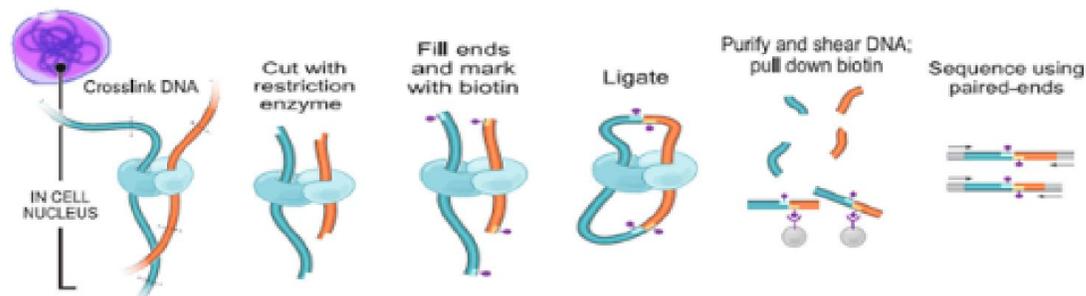


真菌准完成图

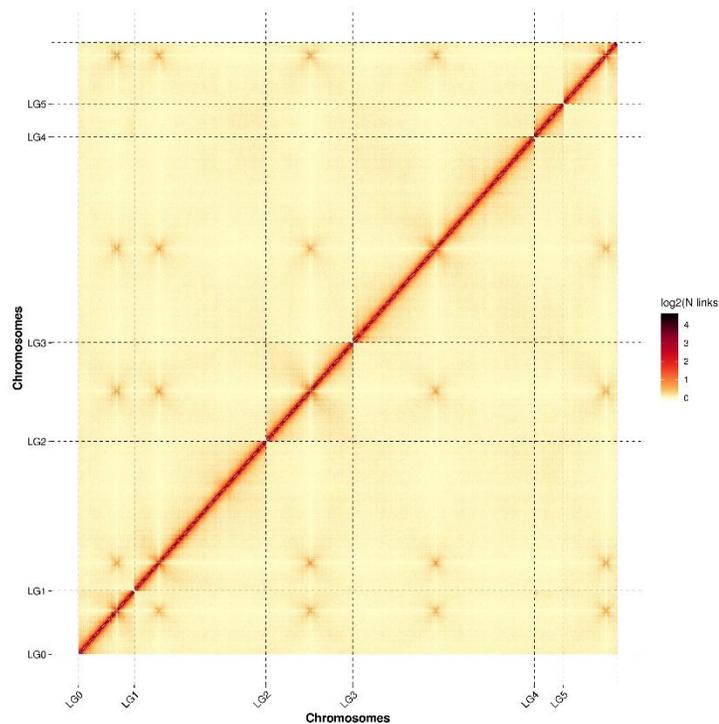
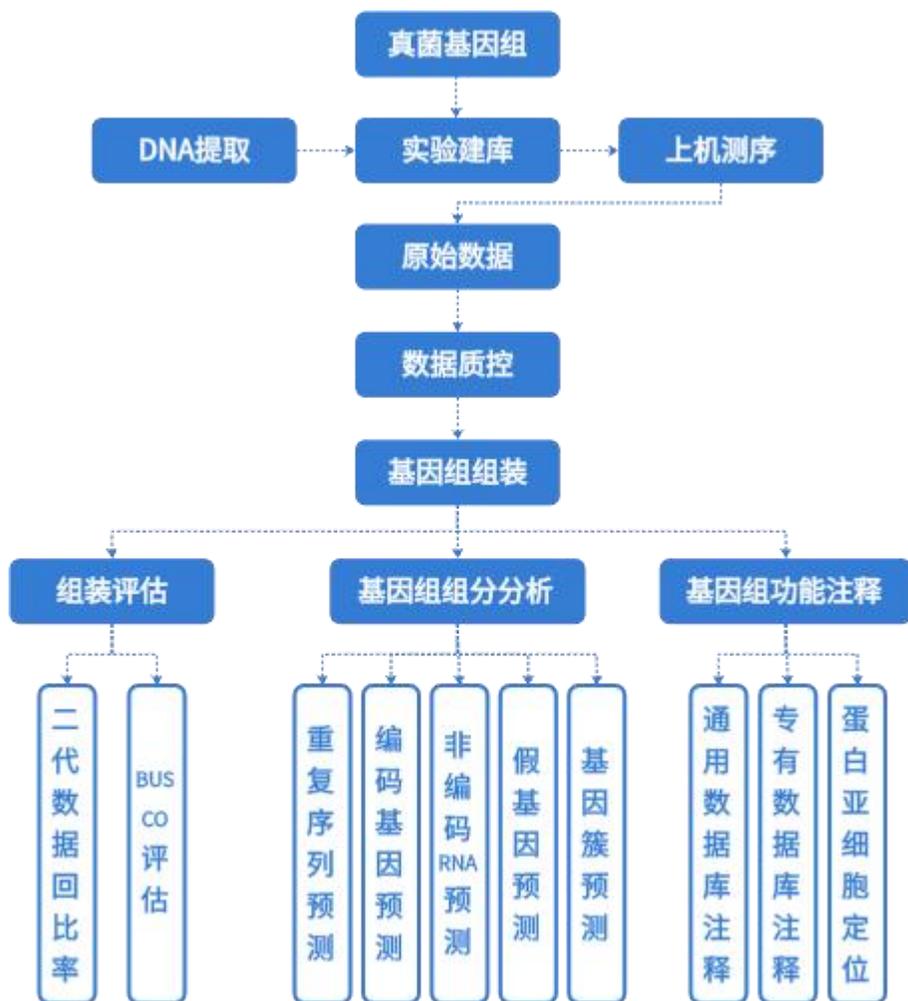
精细图-三代

Hi-C技术, 将contig定位到染色体上

准完成图



A: 细胞交联 B: 内切酶酶切 C: 末端修复 D: 环化 E: DNA 纯化和捕获 F: Illumina Pair-End 测序



真菌框架图与精细图

框架图

表1 基因组组装结果统计

Scaffold Length (bp)	Scaffold Number	Scaffold N50 (bp)	Scaffold N90 (bp)	Contig Length (bp)	Contig Number	Contig N50 (bp)	Contig N90 (bp)	GC content (%)	Gaps Number
30,578,270	326	394,877	78,686	30,576,667	591	344,746	62,910	41.23	265

表1 基因组组装结果统计

Contig Length (bp)	Contig Number	Contig N50 (bp)	Contig N90 (bp)	GC content (%)	Gaps Number
37,235,197	47	3,461,147	1,161,397	49.21	0

表2 比对结果统计

精细图

Library	Mapped(%)	Properly mapped(%)	Coverage(%)	Depth(X)
350bp	99.81	98.29	99.95	119.92

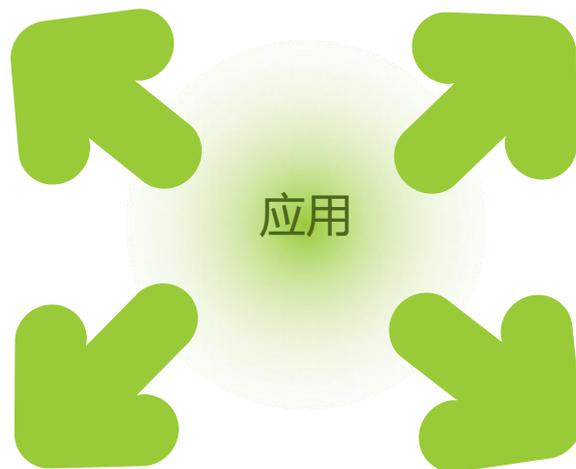
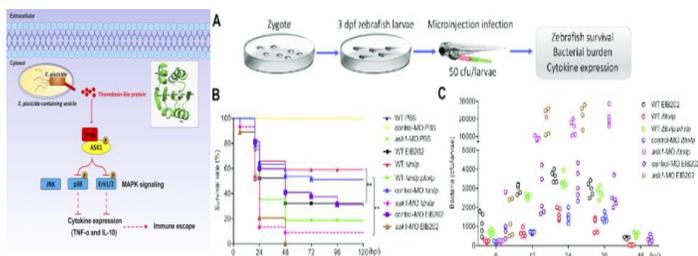
表3 BUSCO评估统计

Complete BUSCOs(C)	Complete and single-copy BUSCOs(S)	Complete and duplicated BUSCOs(D)	Fragmented BUSCOs(F)	Missing BUSCOs(M)	Total Lineage BUSCOs
272 (93.79%)	272 (93.79%)	0 (0.00%)	10 (3.45%)	8 (2.76%)	290

微生物基因组应用方向

病原菌研究

基因岛、毒力因子分析



耐药性研究

CARD抗生素抗性基因注释

分泌蛋白研究

发酵工程研究

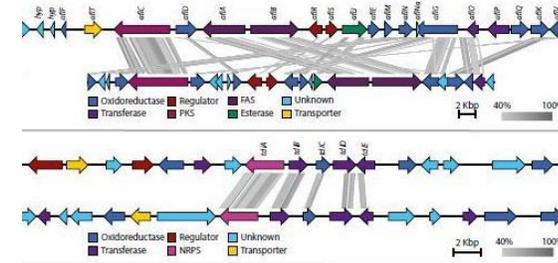
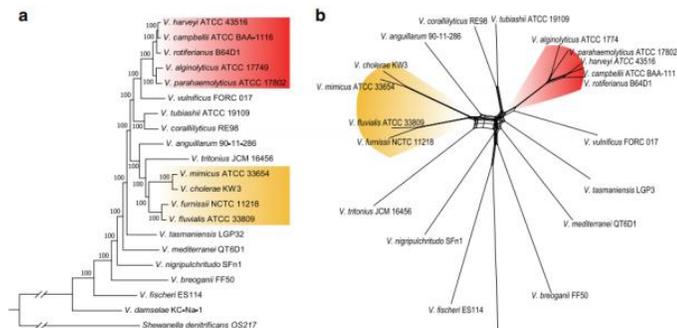
代谢通路研究、次级代谢产物

环境适应性研究

碳水化合物酶注释

跨膜蛋白研究

比较基因组学分析



Thanks for your attention!

百迈客基因 为世界创造新的可能