

基于高通量测序平台的组学研究 及其应用简介

刘伟强

Weiqiang.liu@majorbio.com

18516656530

- 一、高通量测序基础知识
- 二、微生物群落的研究
- 三、微生物基因组的研究
- 四、RNA的研究

一、高通量测序基础知识

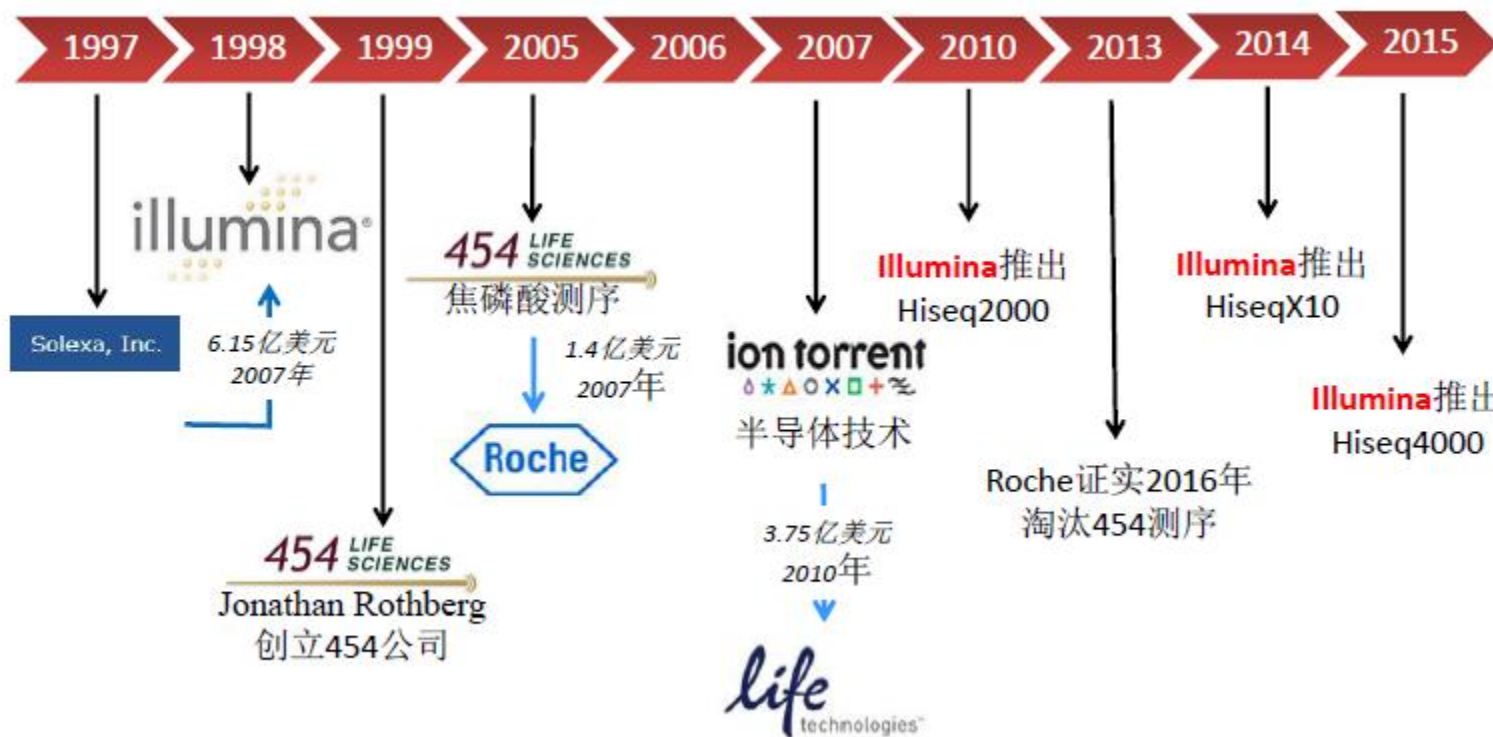


高通量测序简介

高通量测序技术 又称“下一代”测序技术（“Next-generation” sequencing technology, NGS），以能一次并行对几十万到几百万条DNA分子进行序列测定和一般读长较短等为标志。

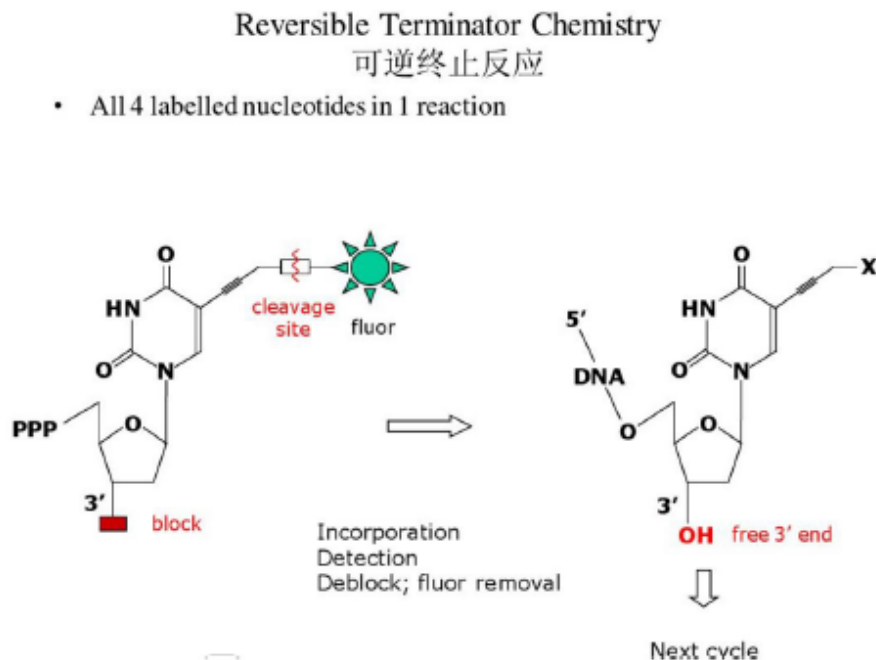
特点：费用低、时间短、通量高、可同时处理成百上千个样本。

主流测序仪公司： Illumina



illumina测序原理

可逆阻断技术：利用3'端的阻断基团，确保一次只能合成一个碱基，再利用相应的激光激发荧光集团，捕获激发光，从而读取碱基信息。

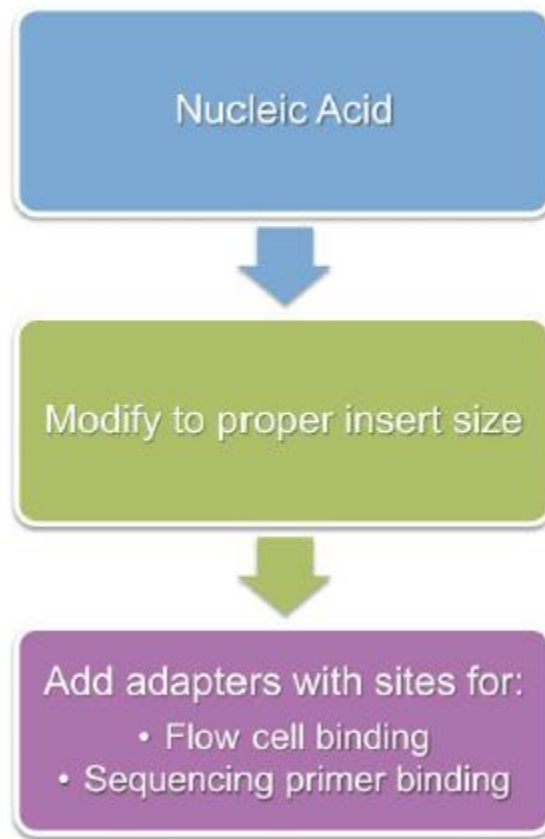


文库制备

Library Preparation



illumina文库概念:
将DNA/RNA样本处理成合适的片段大小，并在两端连接上特定illumina接头的DNA片段即可称为illumina文库。



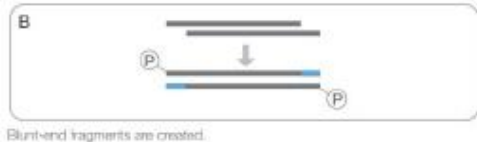
文库制备

TruSeq Nano DNA Library Prep Kit

DNA片段化



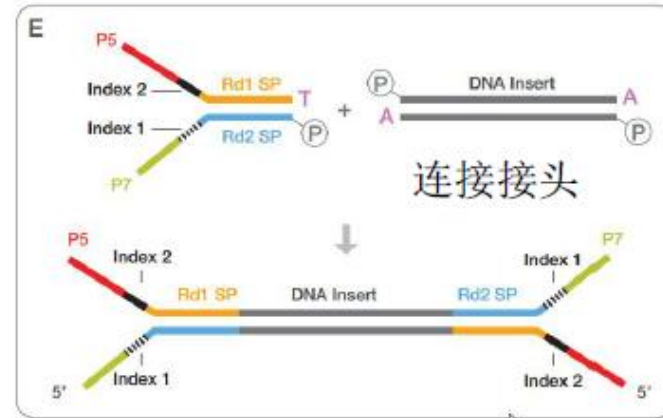
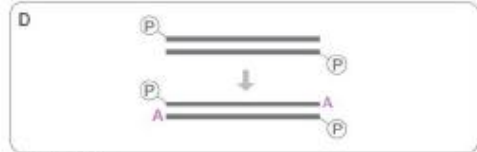
末端修复



片段筛选



加A

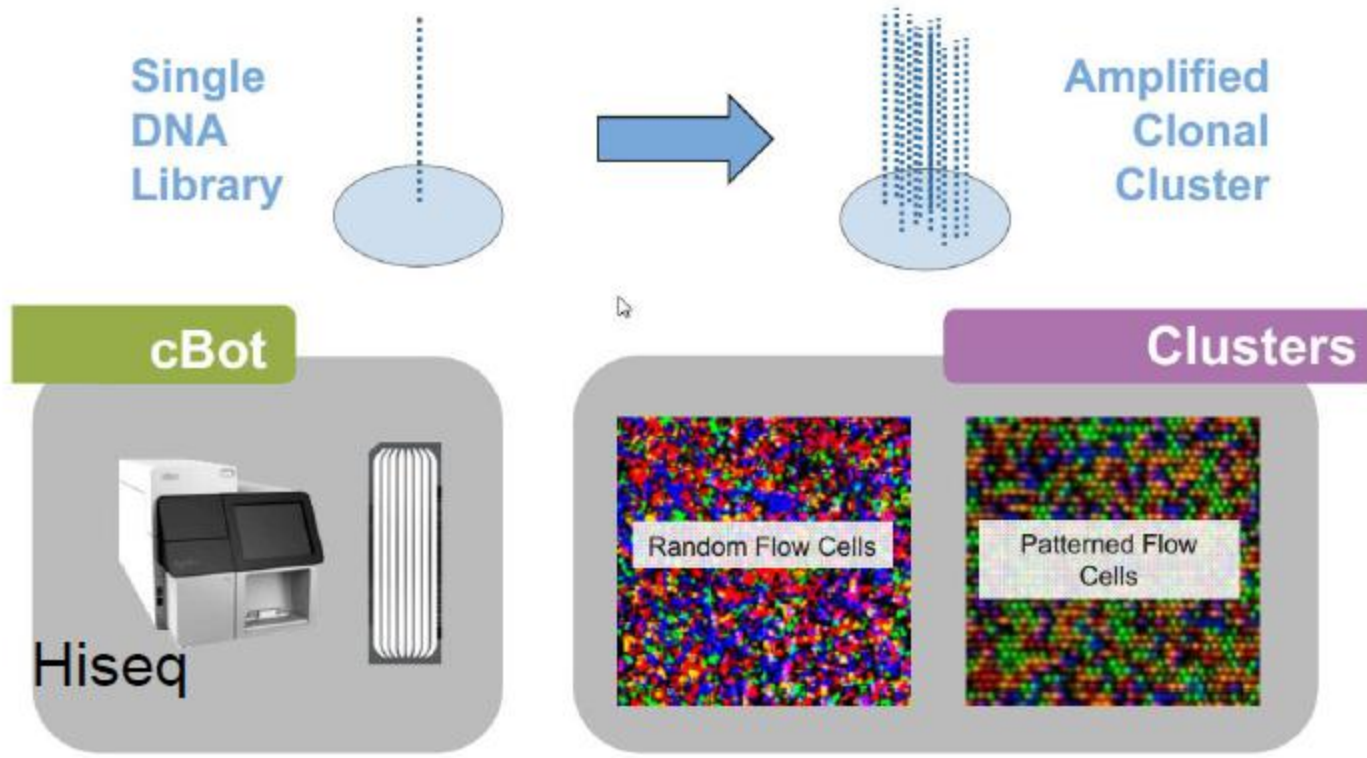


↓ PCR扩增



桥式PCR，形成cluster

Cluster Generation Instrumentation



三代测序平台

- 2009年Pacific Biosciences (PacBio) 公司推出了单分子实时合成测序SMRT (single molecule real-time) 技术, 该方法基于边合成边测序的思想, 采用四色荧光标记的dNTP和零级波导 (zero-mode waveguides, ZMW) 的纳米芯片对单个DNA分子进行测序。
- SMRT技术最大的优势是读长超长, PacBio RS II测序平台最大读长可超过60 kb, 平均读长约8.5 kb, 是目前所有商品化测序仪中读长最长的, 这对基因组拼接, 尤其是含有较多串联重复序列的基因组拼接是一个福音。

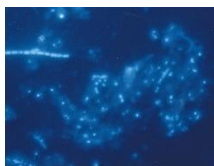
DNA 产品	实验技术
各物种基因组扫描图	PE文库
动植物基因组测序	PE文库+MP文库
细菌完成图	PACBIO文库
全基因组甲基化测序	bisulfite 文库
动植物简化基因组	GBS/RAD
外显子测序	PE文库+目标捕获
目标区域基因组测序	PE文库+目标捕获
	多重PCR
单细胞基因组测序	全基因组扩增+PE文库
宏基因组测序	PE文库
微生物多样性	META文库
微生物16S全长测序	PCR+PACBIO文库
功能基因组测序	META文库

RNA 产品	实验技术
真核转录组测序	polyA捕获+转录组文库
原核转录组测序	rRNA去除+转录组文库
链特异转录组	rRNA去除+链特异转录组文库
small RNA测序	small RNA文库
circular RNA测序	rRNA去除+lineRNA去除+链特异转录组文库
	rRNA去除+链特异转录组文库
全长转录组测序	SMART cDNA+PACBIO文库
单细胞转录组测序	SMART cDNA+PE文库
宏转录组测序	rRNA去除+转录组文库

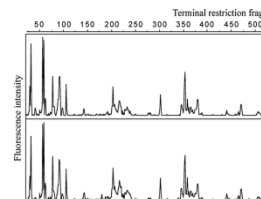
二、微生物群落的研究

Laura Alonso-Saez, *et al.*,
FEMS Microbiology Ecology
(2007), 60,98-112

FISH

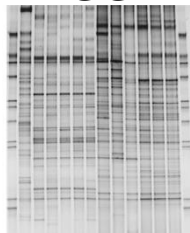


T-RFLP



W T Liu, *et al.*, AEM,
(1997), 63,4516-4522

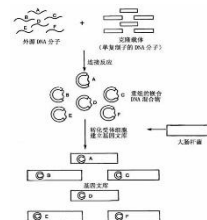
DGGE



分子生物学检测手段

Gerard Muyzer, Kornelia Smalla,
Antonie van Leeuwenhoek
(1998), 73,127-141

Clone library



Todd Z. DeSantis *et al.*, Microbial Ecology
(2007), 53,371-383

微生物群落组成及丰度

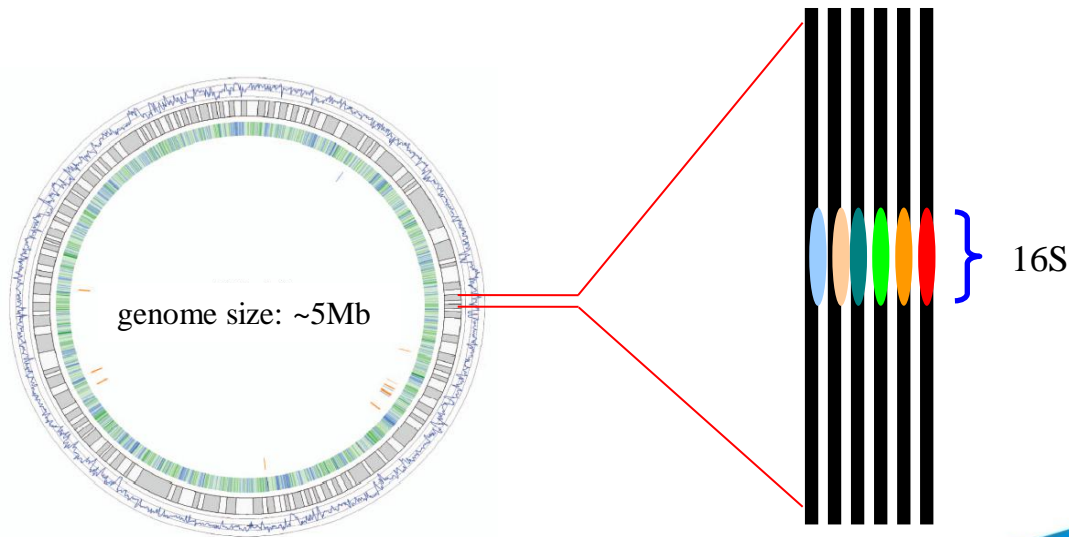


环境样品
(土壤、水体、滤膜、粪便等)

高通量测序

上海美吉生物医药科技有限公司
Shanghai Majorbio Bio-Pharm Technology Co., Ltd.

1991年Pace首次提出环境基因组学的概念，基于核糖体rRNA等基因的扩增子测序来反映系统中的微生物多样性的组成（如16S、18S、ITS等）。



微生物多样性一般用于研究特定生境微生物组成，不同生境微生物群落结构比较，环境因子/临床指标与群落结构相关分析等；核心是微生物群落结构。

微生物群落功能？

宏基因组定义：生物环境中**全部微小生物的基因组**。

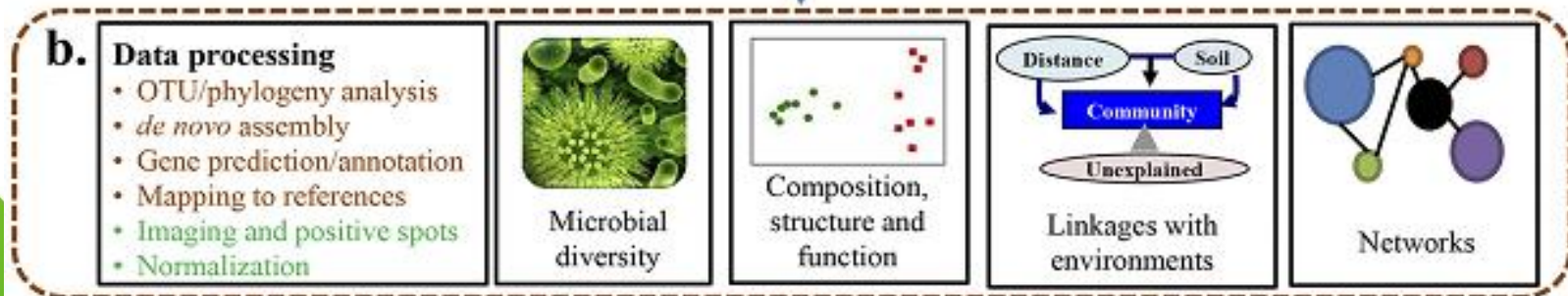
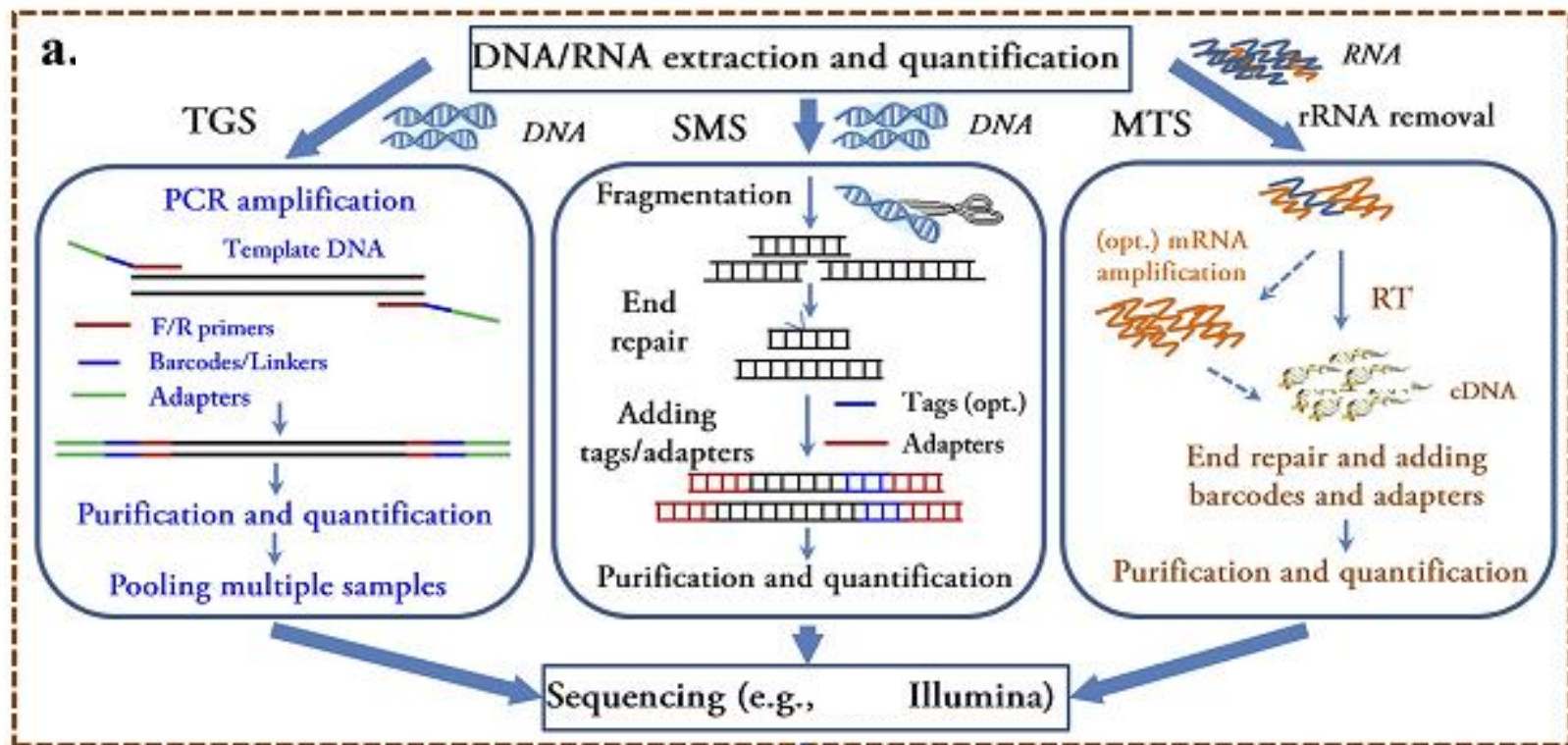
宏基因组学：以研究生境中微生物的种群结构、进化关系、功能活性、相互协作关系及与环境之间的关系为目的的微生物研究方法。

研究手段：二代测序（高通量测序）

多样性

宏基因组

宏转录组



多样性

1、物种注释与评估

2、物种组成分析

3、样本比较分析

4、物种差异分析

5、相关性分析

6、进化分析

宏基因组

1、物种与功能注释

2、物种与功能组成分析

3、样本比较分析

4、物种与功能差异分析

5、相关性分析

6、进化分析

宏转录组

1、物种与功能注释

2、差异表达基因分析

3、物种与功能组成分析

4、样本比较分析

5、物种与功能差异分析

6、相关性分析

7、进化分析

- **物种分类学注释：**将基因集序列与参考数据库比对，并根据分类学谱系系统得到该序列的物种分类地位，从而分析得到整个基因集的物种信息；
- **COG功能注释：**将基因集序列与COG参考数据库比对，获得基因集的同源聚类蛋白群的相应功能注释信息；
- **KEGG功能注释：**将基因集序列与KEGG的基因组数据库比对，获得基因集在代谢、酶、反应和功能模块方面的注释信息。
- **个性化功能注释：**将基因集序列与特殊数据库比对，获得基因集特殊的功能信息，如：CAZy、ARDB、VFDB。

三、微生物基因组的研究

基因组 *denovo* 测序



微生物基因组

小基因组

真菌类

古生菌

真细菌

蓝藻等

支原体等

质粒

病毒

线粒体

叶绿体

基因组 *denovo* 测序



原核基因组

小基因组

扫描图测序

精细图测序

完成图测序

扫描图测序

完成图测序

三个基础概念:



存在Gap;
基因组覆盖度大于95%;
基因覆盖度大于98%;
单碱基错误率小于十万分之一。

存在Gap;
基因组覆盖度大于98%;
基因覆盖度大于99%;
单碱基错误率小于十万分之一。

无Gap;
获得全长序列和全部基因序列;
单碱基错误率小于十万分之一。

三个基础概念：



简单基因组已不再
提精细图的概念

扫描图测序

精细图测序

完成图测序

使用Illumina PE
文库进行高通量
测序。

使用Illumina PE
文库加上MP文科
进行高通量测序。

使用PacBio三代
测序仪测序获得
全长组装结果。



A

质控检测

B

片段化和建库

C

上机测序

d

数据分析

纯度
检测

NanoDrop分析检测核酸浓度，分析OD260/OD280以及OD260/OD230的吸光值比例，判定纯度

完整性
检测

利用琼脂糖凝胶电泳检测DNA样品的片段大小和范围

总量
检测

利用Qubit/BS-380精准检测双链DNA的含量

富集
检测

检测磁珠对特定长度片段的吸附能力



A

质控检测

B

片段化和建库

C

上机测序

d

数据分析

基因组
片段化

根据基因组完整性情况和测序需要，利用 G-tubes 方法将基因组 DNA 处理成 8-20 kb 的片段应用于 PacBio 测序，利用 Covaris M220 将基因组 DNA 处理成 400 bp 左右的短片段用于 Illumina 测序。

末端
修复

PacBio 测序末端补平，片段两端分别连接环状单链得到一个套马环结构；Illumina 测序末端加 A 补平。

片段
筛选

根据测序需要，筛选特定长度的片段用于上机测序。



A

质控检测

B

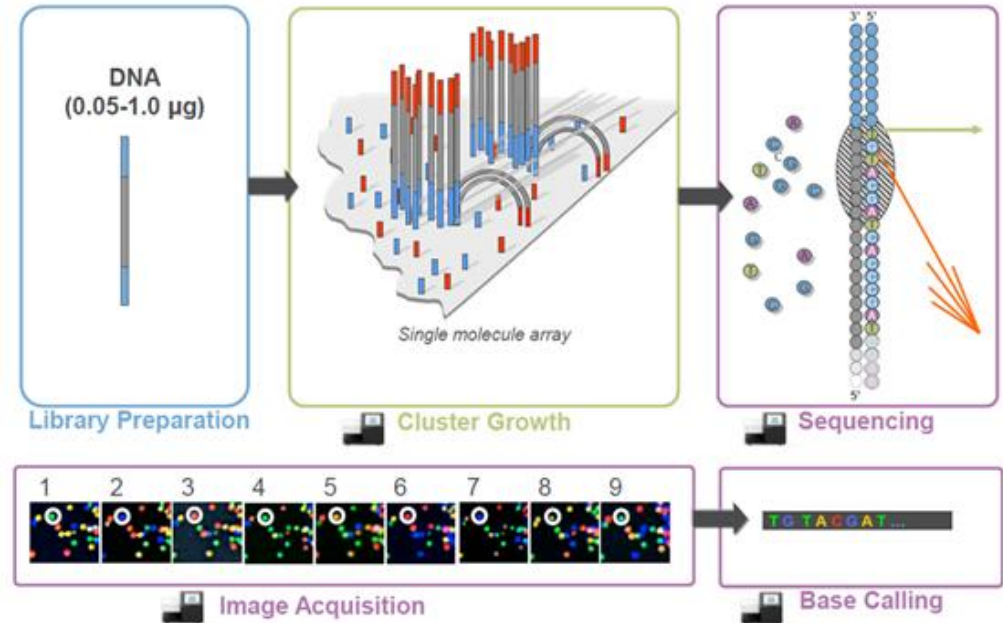
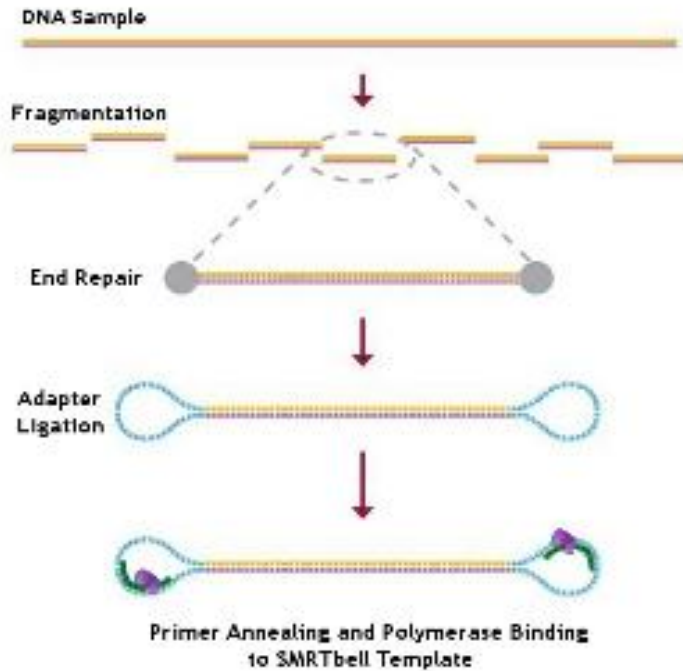
片段化和建库

C

上机测序

d

数据分析



PacBio三代测序

Illumina二代测序

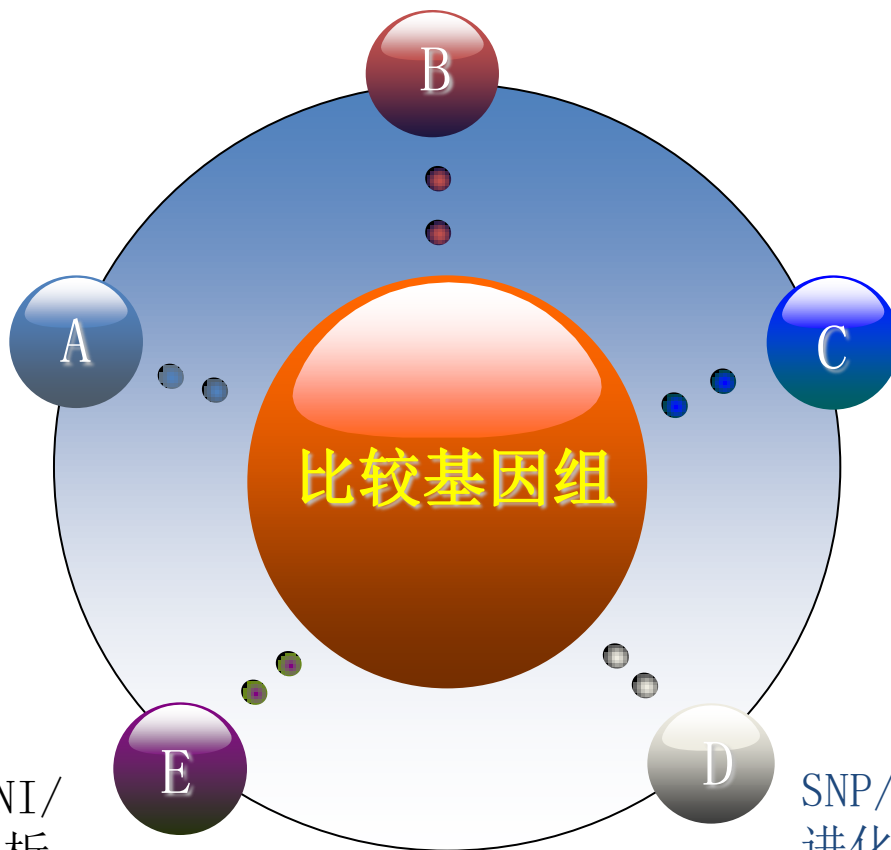

A
质控检测
B
片段化和建库
C
上机测序
d
数据分析

标准分析	高级分析I	高级分析II	个性化分析I	个性化分析II	个性化分析III
原始数据统计与处理	GI 预测分析	同源基因分析	耐药基因注释	SNP、InDel 检测及注释	其他个性定制
基因组组装	CRISPR 分析	基于全基因组的进化树分析	毒力基因注释	SV 检测和注释	
基因预测	前噬菌体序列分析	共线性分析	碳水化合物相关酶数据库注释		
重复序列分析	圈图绘制	泛基因组分析	转座单元预测与分析		
tRNA&rRNA预测	假基因预测	ANI 分析	分泌蛋白预测分析(信号肽和跨膜螺旋)		
Nr/Swiss-prot 注释			基因簇比较作图分析		
COG/GO/KEGG 注释			数据上传 NCBI 数据库		

物种进化树分析、
基因进化树分析



同源基因分析、
泛基因组分析



共线性分析、
基因水平转移分析

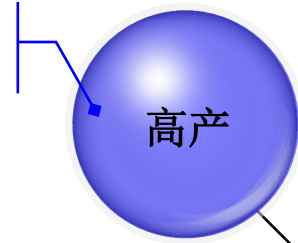
ANI/TNI/
AAI分析

SNP/Indel、SV分析
进化速率分析 (dN/dS)

个性化分析



1. 微进化分析;
差异关联分析;
操纵子挖掘



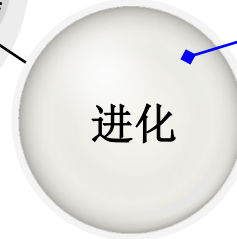
高产

4. 次级代谢产物
基因簇分析;
代谢途径分析



代谢

5. 结构变异分析;
基因水平转移;
群体进化分析



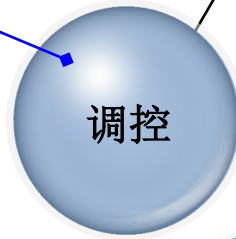
进化

2. 耐受基因、代谢
途径分析;
基因簇作图;



高耐受性

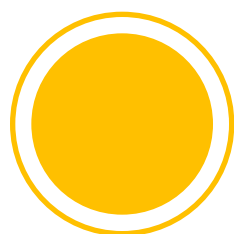
3. 群体感应系统分析;
组成型/诱导性表达分析



调控

您提出要求,
我们提供解决方案

上海美吉生物医药科技有限公司
Shanghai Majorbio Bio-Pharm Technology Co., Ltd.



经典案例分享

乳酸杆菌扫描图测序



中文题目：
通过213株菌比较基因组学和属的相关性分析挖掘乳酸杆菌潜在的生物功能信息

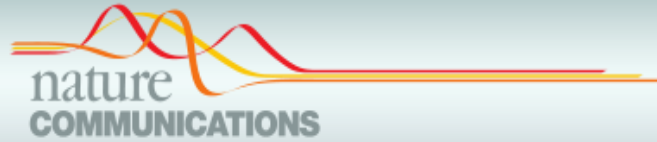
发表时间：
2015.9

IF: 11.47

合作单位：
内蒙古农业大学

测序平台：
Illumina HiSeq2000

样品：
213株乳酸杆菌



ARTICLE

Received 24 Oct 2014 | Accepted 11 Aug 2015 | Published 29 Sep 2015

DOI: 10.1038/ncomms9322

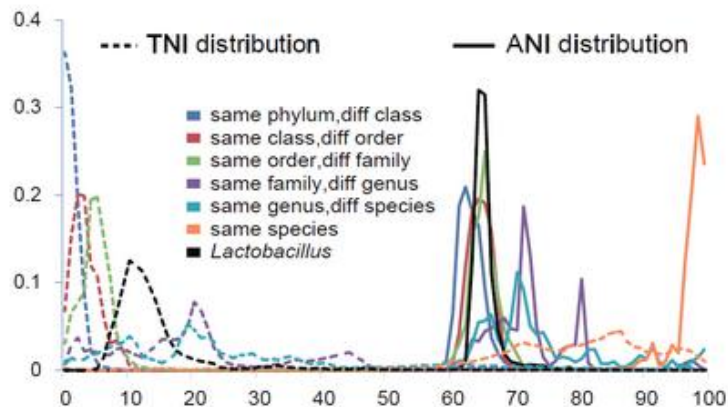
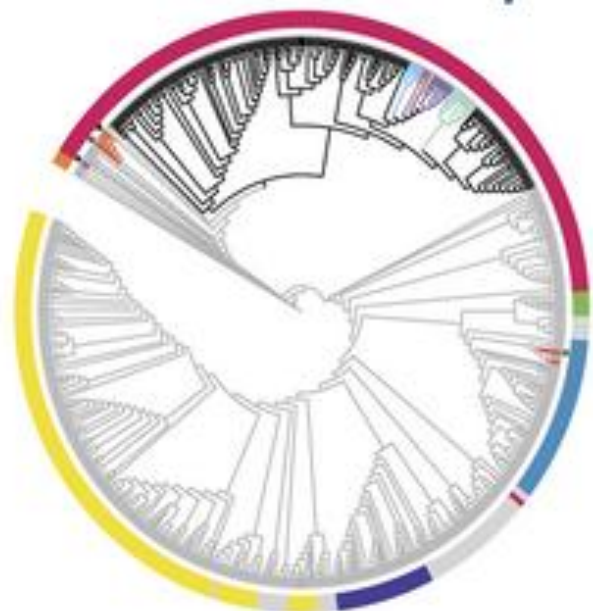
OPEN

Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera

Zhihong Sun^{1,*}, Hugh M.B Harris^{2,*}, Angela McCann^{2,*}, Chenyi Guo^{3,*}, Silvia Argimón^{4,*}, Wenyi Zhang^{1,*}, Xianwei Yang³, Ian B. Jeffery², Jakki C. Cooney⁵, Todd F. Kagawa⁵, Wenjun Liu¹, Yuqin Song¹, Elisa Salvetti⁶, Agnieszka Wrobel², Pia Rasinkangas⁷, Julian Parkhill⁸, Mary C. Rea⁹, Orla O'Sullivan⁹, Jarmo Ritari⁷, François P. Douillard⁷, R. Paul Ross⁹, Ruifu Yang³, Alexandra E. Briner¹⁰, Giovanna E. Felis⁶, Willem M. de Vos^{7,11}, Rodolphe Barrangou¹⁰, Todd R. Klaenhammer¹⁰, Page W. Caufield⁴, Yujun Cui³, Heping Zhang¹ & Paul W. O'Toole²

Lactobacilli are a diverse group of species that occupy diverse nutrient-rich niches associated with humans, animals, plants and food. They are used widely in biotechnology and food preservation, and are being explored as therapeutics. Exploiting lactobacilli has been complicated by metabolic diversity, unclear species identity and uncertain relationships between them and other commercially important lactic acid bacteria. The capacity for biotransformations catalysed by lactobacilli is an untapped biotechnology resource. Here we report the genome sequences of 213 *Lactobacillus* strains and associated genera, and their

进化分析研究



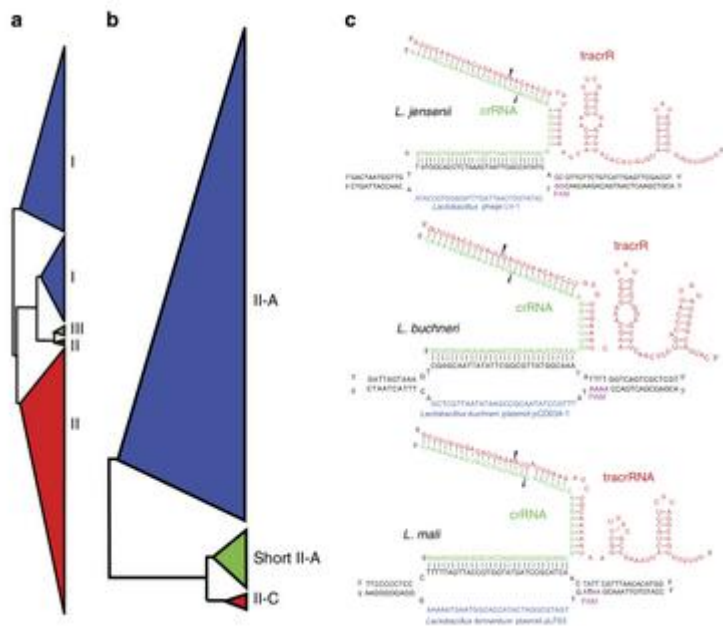
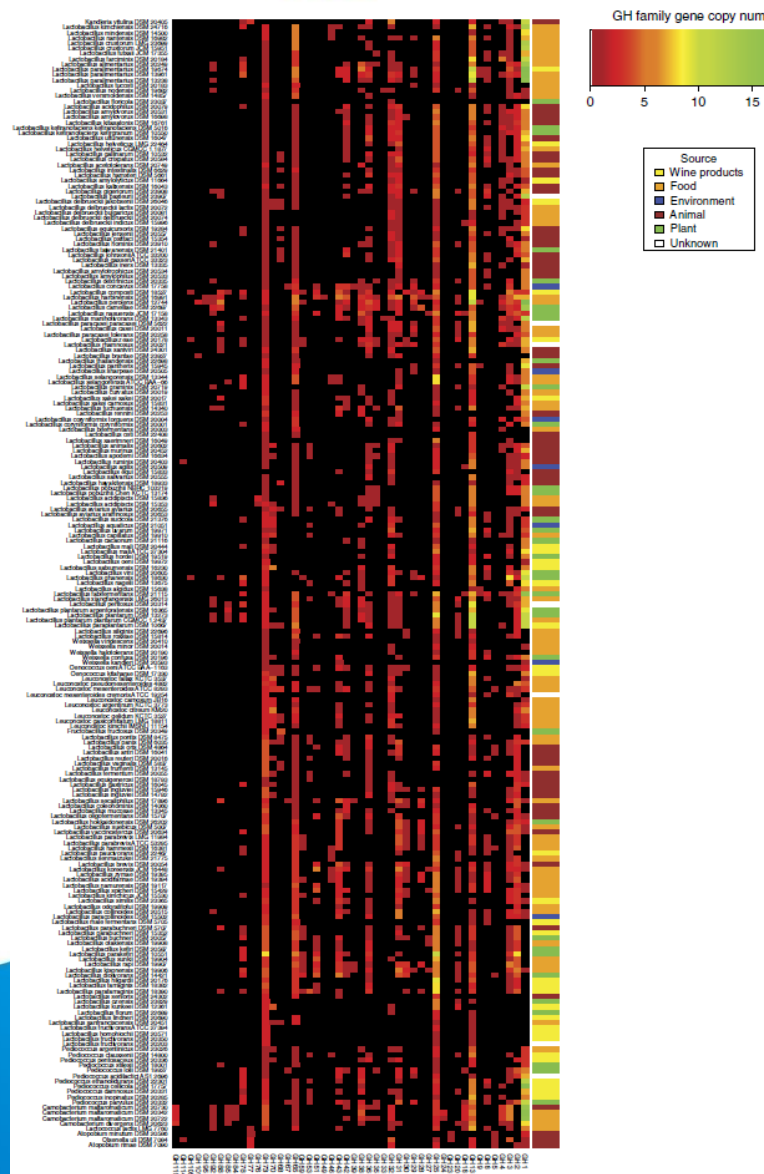
Lactobacillus的ANI和TNI分布频率与传统的菌种分类比较

- 选择来自26个门452个属的菌株和Lactobacillus属模式菌株，以它们**同有的16个基因**的氨基酸序列构建系统发育树
- 结果表明**乳杆菌属为一个并系类群**，是由从一个特定的共同祖先不断进化分化而成，即兼性异型发酵乳杆菌为了适应环境逐渐分化形成严格同型和异型发酵乳杆菌。

代谢分析研究

碳水化合物和糖苷水解酶在所研究的213个菌株中的分布

挖掘乳杆菌细胞表面互作因子相关基因
CRISPR-Cas系统和可动的遗传因子



根际克雷白氏杆菌完成图分析



- 合作单位：
中国科学院南京土壤研究所
- 发表时间：2016年5月
- IF：5.228
- 测序类型：细菌基因组完成图
- 研究对象：根际克雷白氏杆菌
- 研究目的：
基于全基因组测序分析细菌高浓度盐耐受系统、较宽的pH适应性。

SCIENTIFIC REPORTS

OPEN

Whole genome analysis of halotolerant and alkalotolerant plant growth-promoting rhizobacterium *Klebsiella* sp. D5A

Received: 22 January 2016

Accepted: 09 May 2016

Published: 24 May 2016

Wuxing Liu¹, Qingling Wang¹, Jinyu Hou¹, Chen Tu², Yongming Luo² & Peter Christie¹

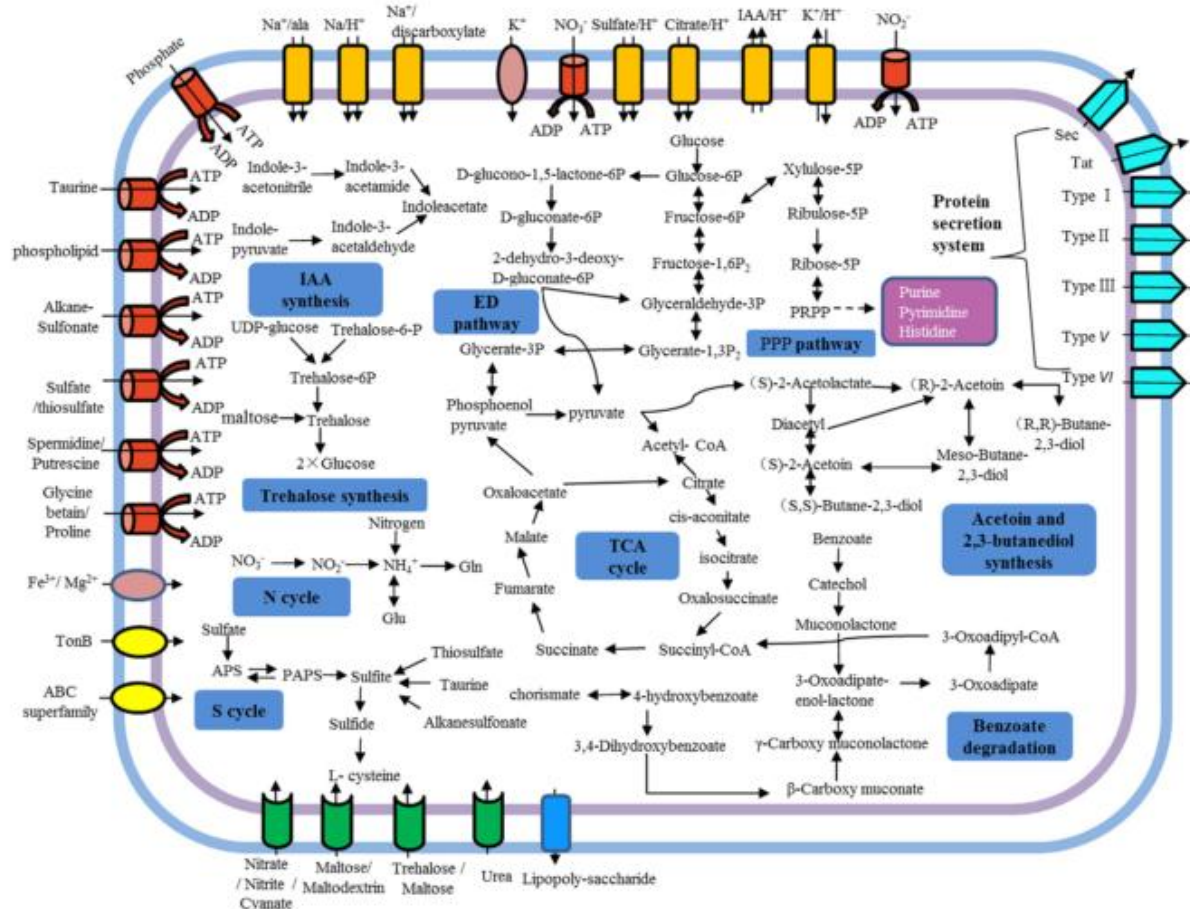
This research undertook the systematic analysis of the *Klebsiella* sp. D5A genome and identification of genes that contribute to plant growth-promoting (PGP) traits, especially genes related to salt tolerance and wide pH adaptability. The genome sequence of isolate D5A was obtained using an Illumina HiSeq 2000 sequencing system with average coverages of 174.7× and 200.1× using the paired-end and mate-pair sequencing, respectively. Predicted and annotated gene sequences were analyzed for similarity with the Kyoto Encyclopedia of Genes and Genomes (KEGG) enzyme database followed by assignment of each gene into the KEGG pathway charts. The results show that the *Klebsiella* sp. D5A genome has a total of 5,540,009 bp with 57.15% G + C content. PGP conferring genes such as indole-3-acetic acid (IAA) biosynthesis, phosphate solubilization, siderophore production, acetoin and 2,3-butanediol synthesis, and N₂ fixation were determined. Moreover, genes putatively responsible for resistance to high salinity including glycine-betaine synthesis, trehalose synthesis and a number of osmoregulation receptors and transport systems were also observed in the D5A genome together with numerous genes that contribute to pH homeostasis. These genes reveal the genetic adaptation of D5A to versatile environmental conditions and the effectiveness of the isolate to serve as a plant growth stimulator.

文章的关键数据点



- Klebsiella sp. D5A的一般基因组特征
- 促进植物生长相关的基因
- 固氮基因
- 耐盐性基因
- 如何降解芳香化合物
- 中心代谢途径
- 分泌系统

关键代谢网络分析研究



Schematic overview of metabolic pathways and transport systems of *Klebsiella sp. D5A*

一株嗜冷乳杆菌的完成图测序



- 发表时间：2015年5月
- IF：5.228
- 测序类型：
细菌基因组完成图
- 研究对象：
一种嗜冷乳酸菌
- 研究目的：
基于全基因组测序研究乳酸菌的嗜冷性和基因组的特征。
- 测序平台：三代PacBio

Tanizawa et al. *BMC Genomics* (2015) 16:240
DOI 10.1186/s12864-015-1435-2



RESEARCH ARTICLE

Open Access

Complete genome sequence and analysis of *Lactobacillus hokkaidonensis* LOOC260^T, a psychrotrophic lactic acid bacterium isolated from silage

Yasuhiro Tanizawa^{1,2}, Masanori Tohno^{3*}, Eli Kaminuma², Yasukazu Nakamura² and Masanori Arita^{2,4}

Abstract

Background: *Lactobacillus hokkaidonensis* is an obligate heterofermentative lactic acid bacterium, which is isolated from Timothy grass silage in Hokkaido, a subarctic region of Japan. This bacterium is expected to be useful as a silage starter culture in cold regions because of its remarkable psychrotolerance; it can grow at temperatures as low as 4°C. To elucidate its genetic background, particularly in relation to the source of psychrotolerance, we constructed the complete genome sequence of *L. hokkaidonensis* LOOC260^T using PacBio single-molecule real-time sequencing technology.

Results: The genome of LOOC260^T comprises one circular chromosome (2.28 Mbp) and two circular plasmids: pLOOC260-1 (81.6 kbp) and pLOOC260-2 (41.0 kbp). We identified diverse mobile genetic elements, such as prophages, integrated and conjugative elements, and conjugative plasmids, which may reflect adaptation to plant-associated niches. Comparative genome analysis also detected unique genomic features, such as genes involved in pentose assimilation and NADPH generation.

Conclusions: This is the first complete genome in the *L. vaccinostercus* group, which is poorly characterized, so the genomic information obtained in this study provides insight into the genetics and evolution of this group. We also found several factors that may contribute to the ability of *L. hokkaidonensis* to grow at cold temperatures. The results of this study will facilitate further investigation for the cold-tolerance mechanism of *L. hokkaidonensis*.

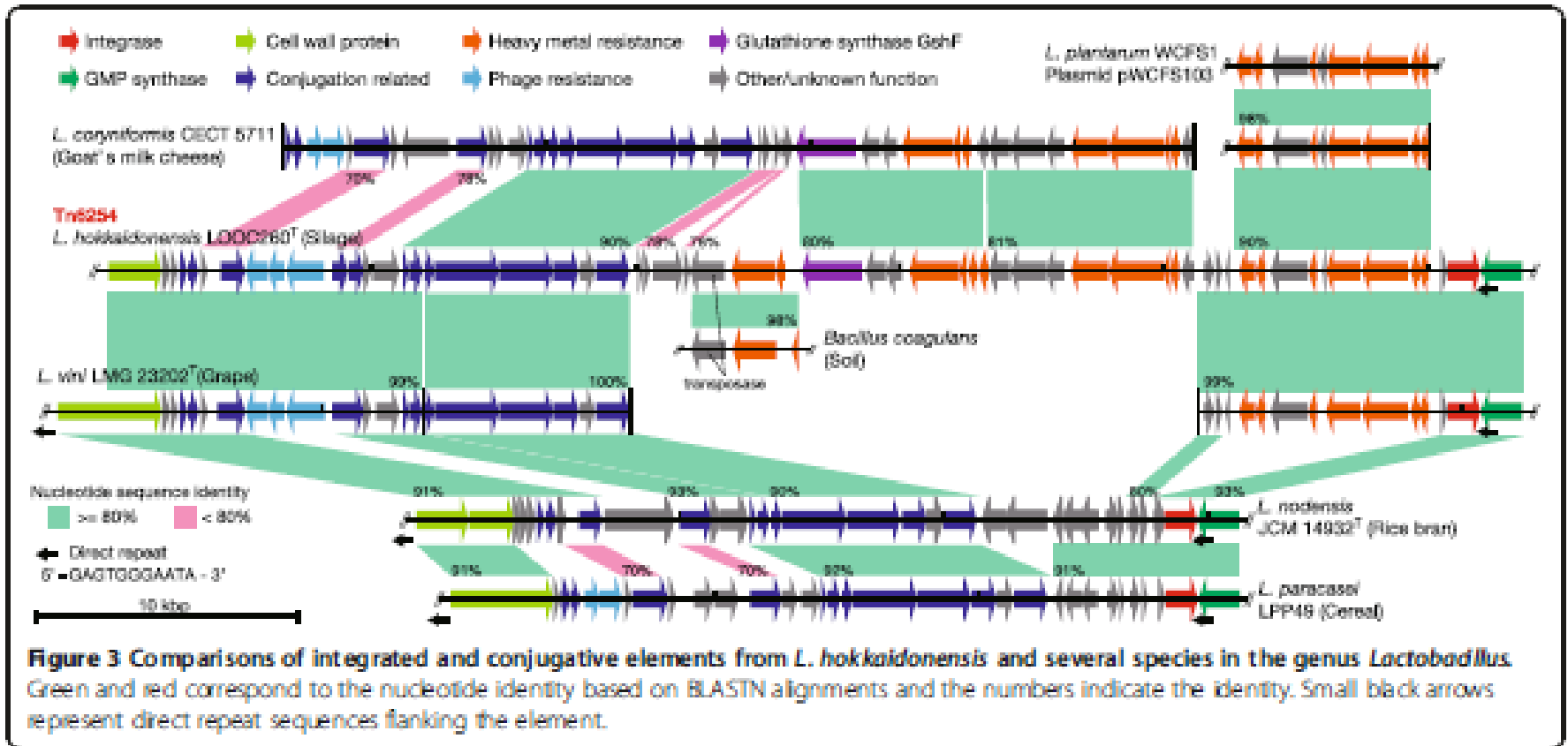
Keywords: Lactic acid bacteria, Silage fermentation, Cold adaptation, Pentose metabolism, Mobile genetic element, Integrative and conjugative element

文章的关键数据点

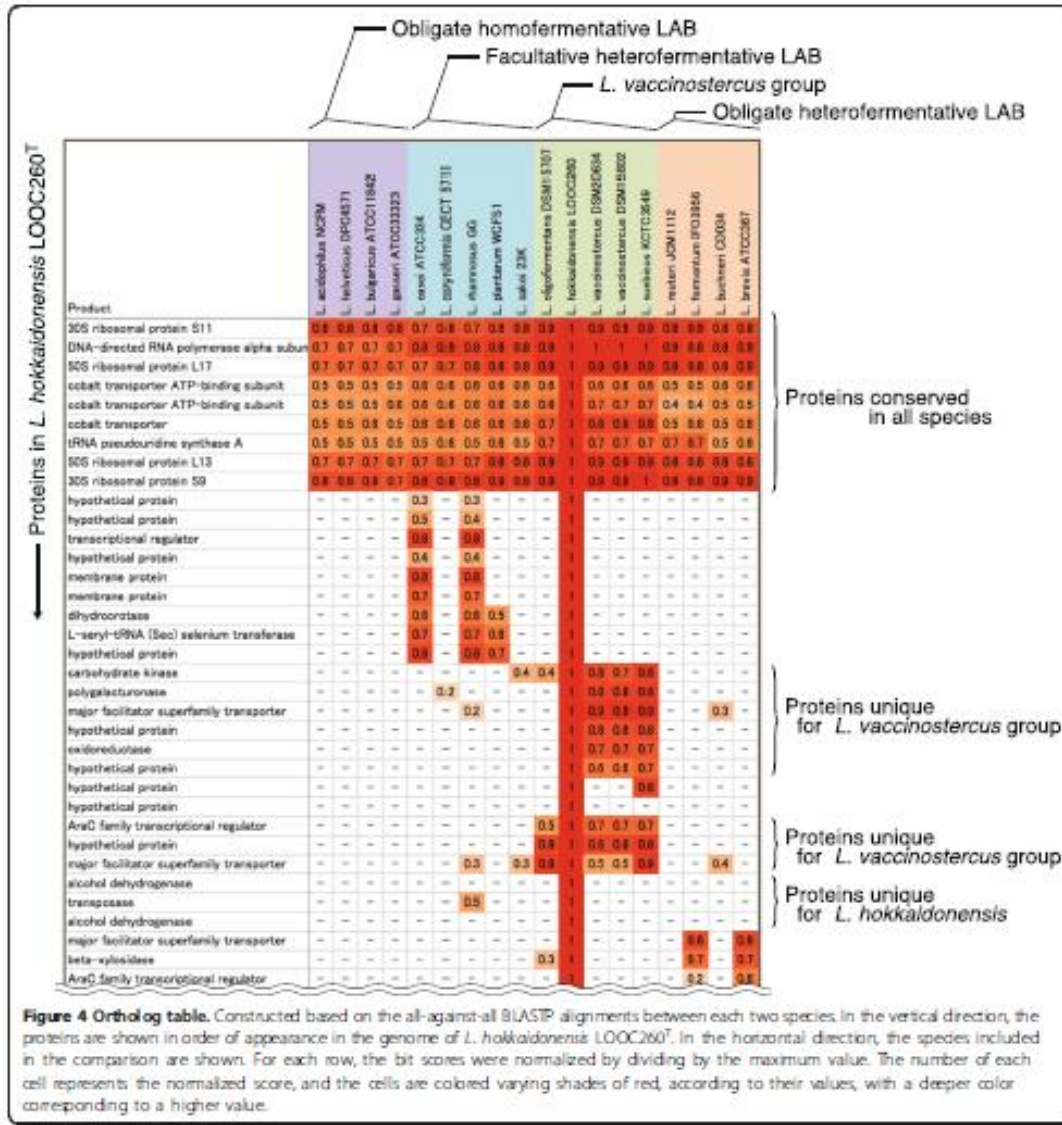


- 分析了基因组的基本序列和基因特征
- 找到5个移动原件，包括前噬菌体、整合和接合元件以及内源质粒等；
- 通过比较基因组找到了基因组的一些特性信息；
- 分析了一些代谢系统，如戊糖同化作用相关基因分析和NADPH生成相关基因分析；
- 冷适应相关基因分析。

整合和接合移动元件



基因组中特性基因的比较分析



冷适应相关基因的挖掘



- 锁定脂肪酸代谢途径，存在于质粒中，且与其他菌属的染色体中相关基因组成和排序相似；
- 找到了cold shock protein A (CspA)；
- 推测与某些代谢产物如甜菜碱、肉毒碱；
- 计划用转录组测序的方法挖掘更多相关基因。

精细化分析



- 精细化分析的意义：
 - 更精细的数据可发现更多基因组细节方面的规律；
 - 精细化的分析指导后续的科学研究；发表更高档次的文章，扩大科研成果的影响力。

相似的方法

创意性的思路

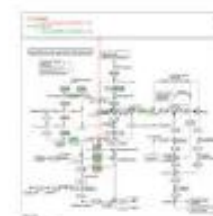
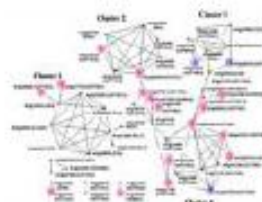
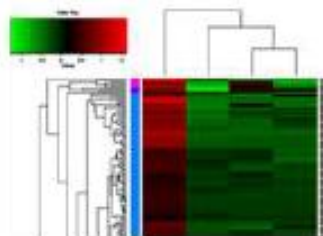
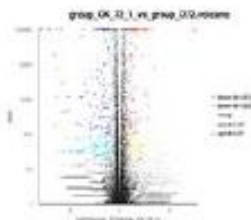
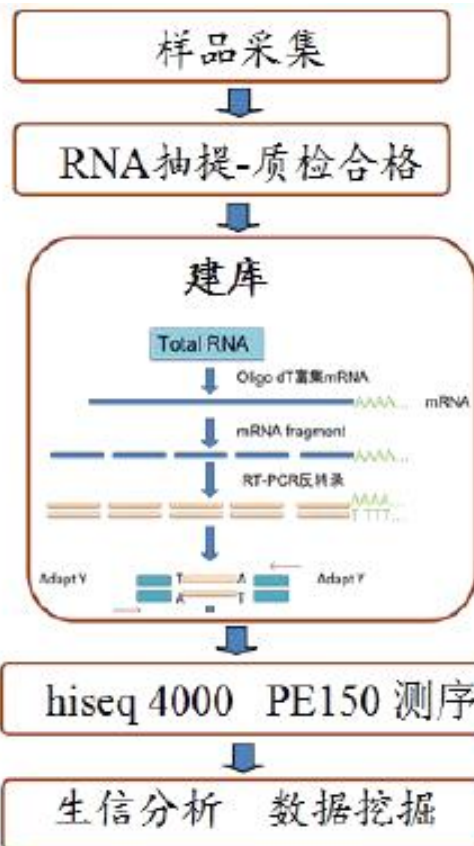


高质量的科研成果

四、RNA的研究

基于二代
测序转录
组的优势

- ⌘ 大规模功能基因的发现 (*de novo*)
- ⌘ 低丰度及新转录本的发现
- ⌘ 发现特定组织或者条件下的可变剪切或编辑
- ⌘ small RNA及lncRNA的鉴定
- ⌘ 基因表达水平研究及差异分析
- ⌘ SNP、SSR检测与分析



转录组项目建库流程

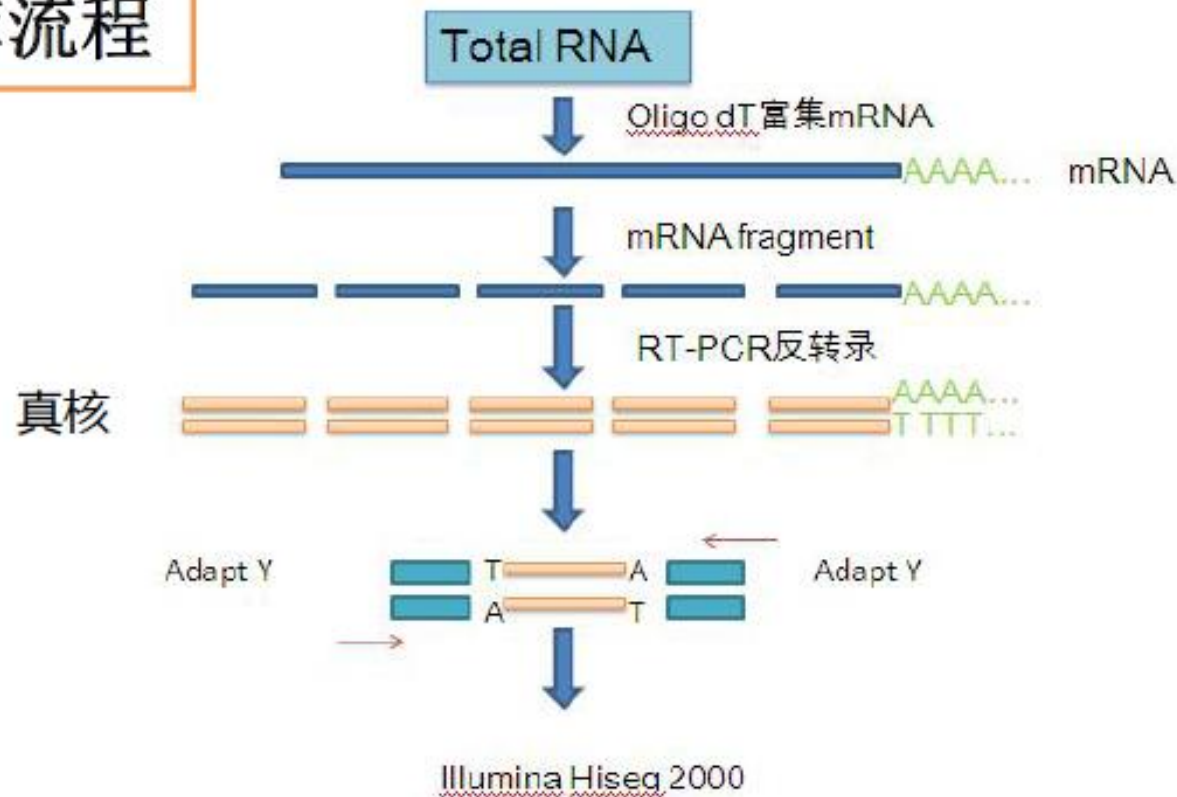
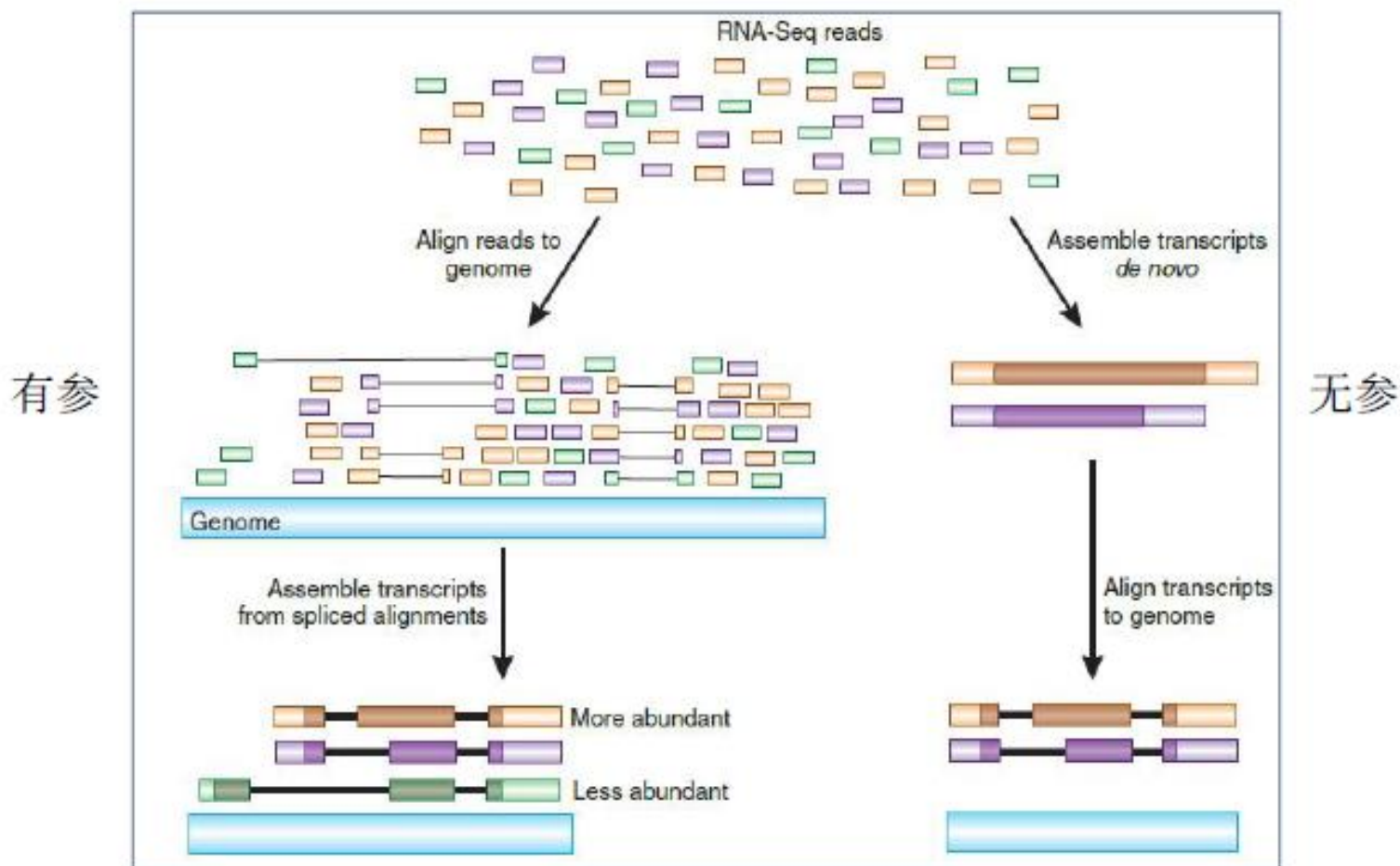


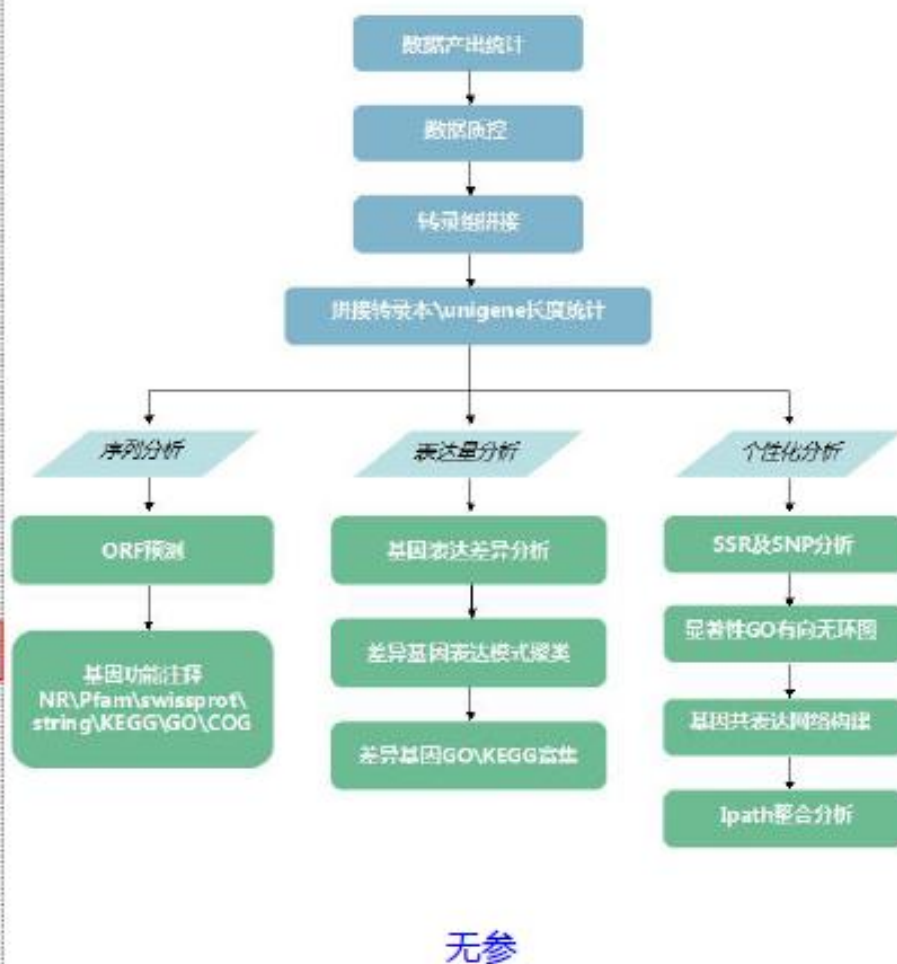
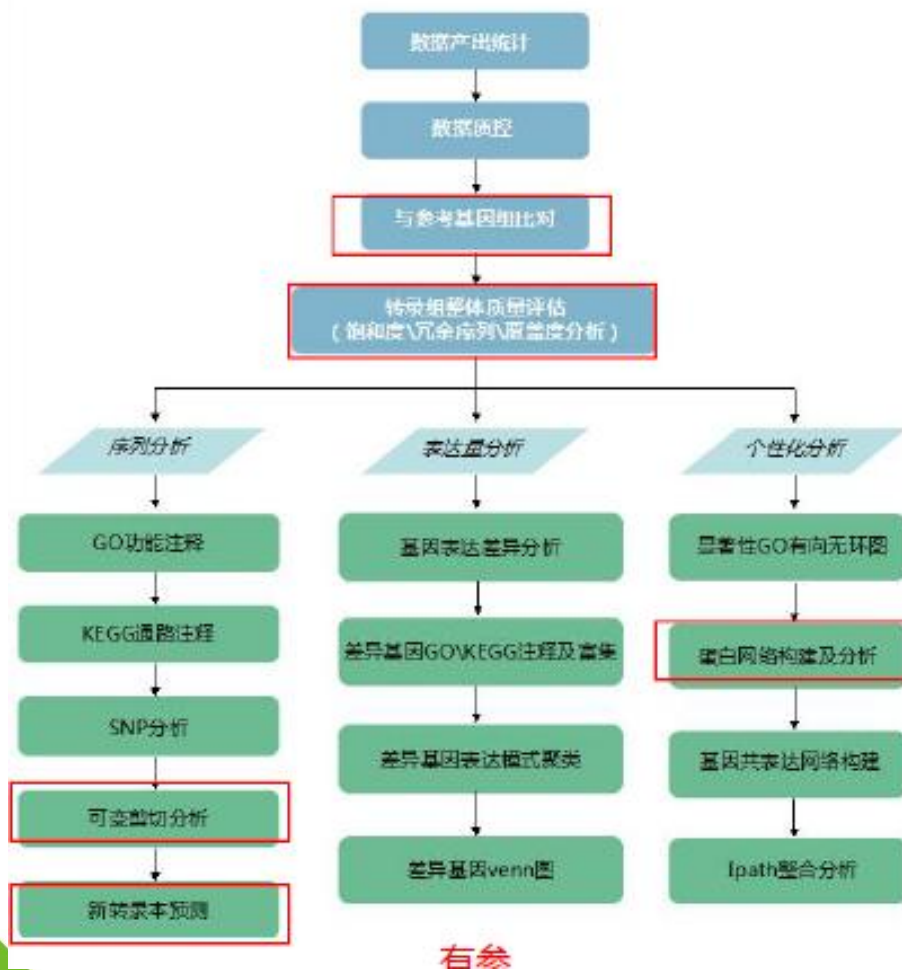
图 2-1 原核转录组实验流程

转录组-有参与无参



✓有参和无参分析对比

红方框为有参转录组特有分析部分



上海美吉生物医药科技有限公司

Shanghai Majorbio Bio-Pharm Technology Co., Ltd.

互作转录组

通过转录组测序，比较寄生菌与寄主或者病原菌与宿主在互作状态下的转录表达，得到病原菌相关的分子响应机制或寄主中响应的通路变化，揭示宿主抗病机理或者寄主致病分子机制

生理适应和分子响应

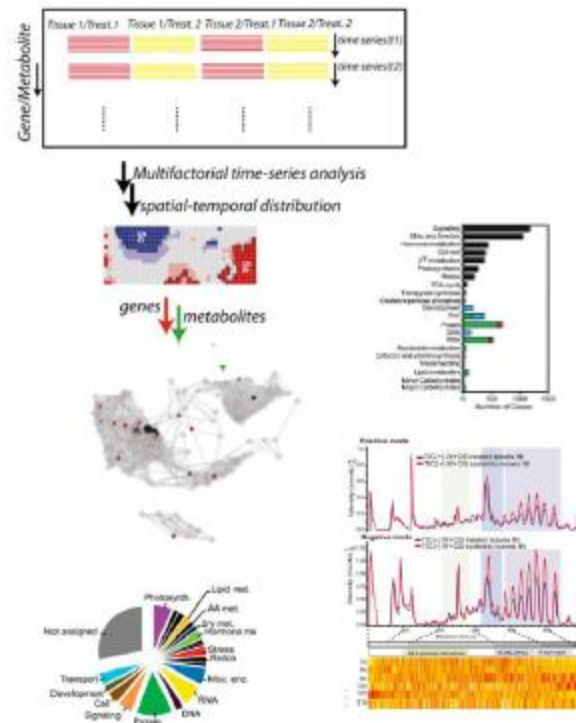
捕食与防御

寄生

共生过程

抗病机理

病虫害



单细胞测序

2013年，单细胞测序技术（single-cell sequencing）荣膺《自然-方法》年度技术。

一个细胞里的DNA或RNA仅仅处在皮克（picograms）级的水平，这么少的量远远达不到现有测序仪的最低上样需求。因此必须先对单细胞内的微量核酸分子进行扩增，而且必须保证尽可能少地出现技术误差，以便开展后续的测序及其他研究。

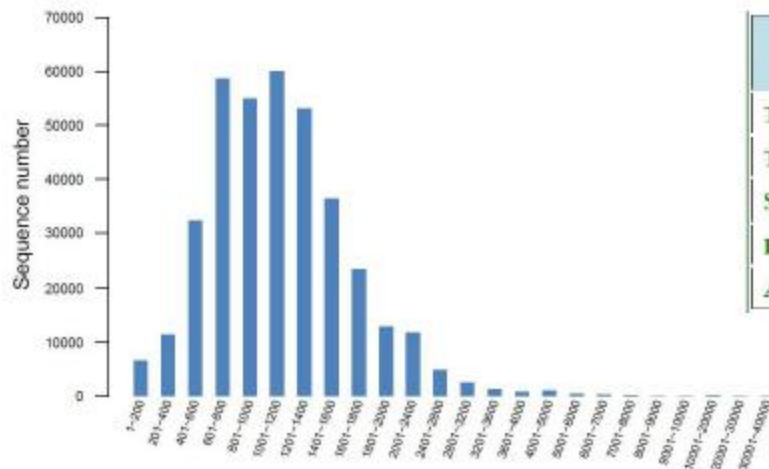
单细胞转录组应用

单细胞测序技术有助于我们剖析细胞的异质性。比较罕见的细胞、胚胎植入前遗传学诊断、异质性的样本、与遗传嵌合或突变相关的表型、不能人工培养的微生物，这些都是单细胞测序技术能够一展所长的研究领域。

其他产品介绍 -- Pacbio全长转录组 iso-seq



Sequence length distribution



	reads of insert(CCS)	full-length reads (non-chimeric)	non-full-length reads
Total sequence num	166306	63482	73626
Total sequence base	247441744	64407844	167875318
Smallestseq	10	300	300
Largest seq	40009	18487	40009
Average length	1487.9	1014.6	2280.1

感谢您的欣赏



美吉生物
Majorbio

地址/Add: 上海市浦东新区国际医学园区康新公路3399号3号楼

电话/Tel: 021-51875086

服务热线: 400-660-1216

网址/Web: www.majorbio.com

传真/Fax: 021-51875086-8002