

DNA 稳定同位素示踪与宏基因组单菌草图组装联用技术

DNA-Stable Isotope Probing and Metagenomics Binning

徐锐^{1,2}, 陶婉^{1,2}, 黄端仪^{1,2}, 苏平舟^{1,2}, 孙晓旭^{1,2}, 张苗苗^{1,2}, 孙蔚旻^{1,2,*}

¹华南土壤污染控制与修复国家地方联合工程研究中心, 广东省科学院生态环境与土壤研究所, 广州市, 广东省; ²广东省农业环境综合治理重点实验室, 广州市, 广东省

*通讯作者邮箱: wmsun@soil.gd.cn

摘要: 高通量测序技术的快速发展极大地促进了环境功能微生物基因组学的研究, 但传统的宏基因组学技术以样品的所有序列为研究对象, 原始测序数据量庞大 (Gb 级别), 存在较多的噪音干扰, 易导致微生物信息挖掘效率低下。本文介绍了 DNA 稳定同位素示踪与宏基因组单菌草图组装联用技术, 可将目标功能微生物的基因组进行有效标记和分离, 从而简化了宏基因组分析难度。具体包括: 首先通过含 ¹³C 稳定同位素的标记物作为唯一碳源与微生物共培养 (DNA-SIP), 可直接锚定参与代谢过程的功能菌群 (Neufeld *et al.*, 2007), 极大地降低了后续测序数据的噪音干扰, 提高生物信息挖掘效率。然后对锚定的 DNA 进行高深度的宏基因组测序, 再基于无参比对方法将短序列分箱 (Metagenomic Binning), 重构微生物在株水平的基因组草图 (Alneberg *et al.*, 2014; Sun *et al.*, 2018)。通过对非培养模式获得的功能微生物, 进行基因注释和代谢通路分析, 解析环境功能微生物的进化和代谢特性, 可有效发掘新型/未知功能物种的基因组学信息 (Quince *et al.*, 2017; Zhang *et al.*, 2020)。

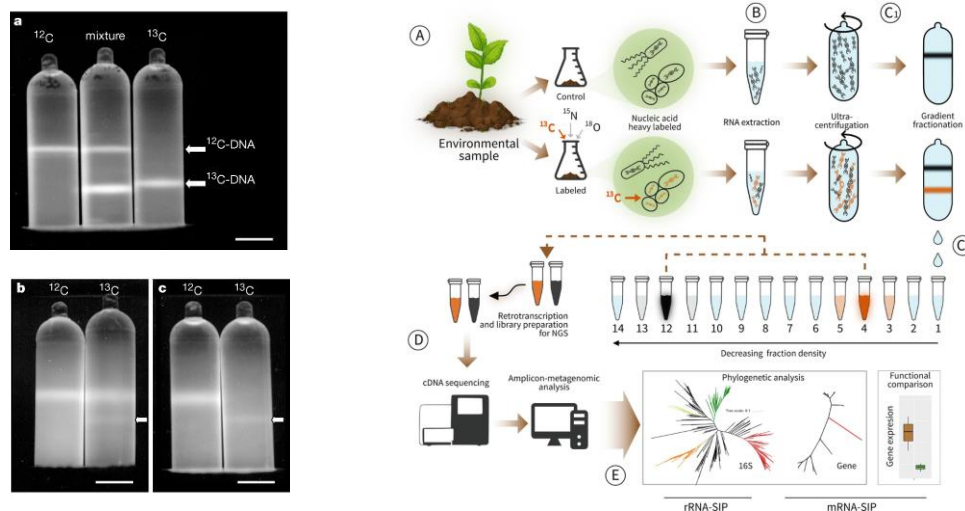
关键词: 稳定同位素示踪, 高通量测序, 宏基因组分箱, 基因组 DNA, 环境功能微生物

基本原理

1. 稳定同位素示踪原理

在碳同化培养体系中添加含 ¹³C 的标记物与微生物共培养。该过程中微生物如果可以利用标记物进行同化作用, 则可将 ¹³C 合并至自身 DNA, 从而完成对目标功能菌群的“标记”。然后收集样品总 DNA, 利用超速离心机分离, ¹³C-DNA 主要分布在底层 (重层), 未被标记的 ¹²C-DNA 则主要分布在上层 (轻层)。重点收集重层 DNA

进行下游的高通量测序及生物信息学分析。



2. 宏基因组分箱原理

宏基因组分箱 (Binning) 基于宏基因组原始测序序列, 可进一步将其中混合了不同微生物的序列或序列组装得到的 contigs 按物种进行分类。提取个体基因组草图(bins)可以实现宏基因组的单菌基因组水平分析。以往传统的单物种全基因组序列都是对目标菌群经过纯培养后, 方能进行全基因组 *de novo* 测序。但是环境中存在着大量的不可培养微生物, Binning 技术不依赖培养条件就可获得微生物的全基因组序列, 从而获得新物种的基因组序列和代谢功能等生物信息。

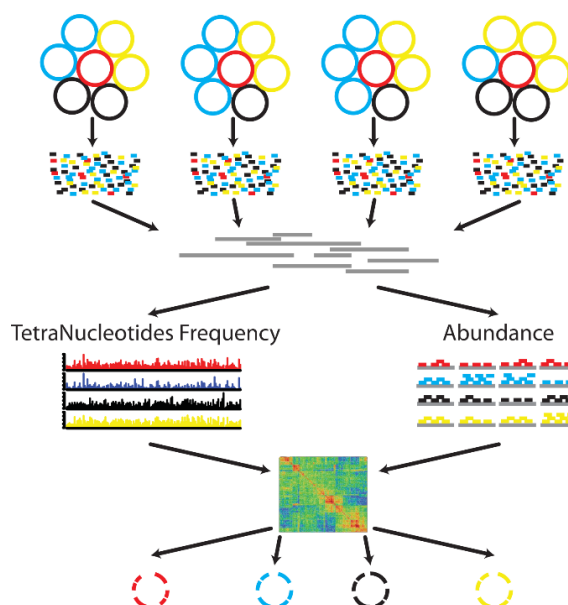


Figure from: MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis

51 材料与试剂

- 52 1. 普通 ^{12}C 乙酸 (北京百灵威科技有限公司, Lot: LE10Q200)
- 53 2. ^{13}C 乙酸 (Cambridge Isotope Laboratories, Inc, Lot: PR-29445)
- 54 3. DNA 提取试剂盒(QIAGEN GmbH, Lot: 163046382)
- 55 4. qPCR 试剂盒 (Takara, Lot: AJ91708A)
- 56 5. 氯化铵 (广州化学试剂厂)
- 57 6. 十二水合·磷酸氢二钠 (麦克林, Lot: C11160473)
- 58 7. 磷酸二氢钾 (广州化学试剂厂)
- 59 8. 氯化镁 (广东光华科技有限公司, Lot: 20190404)
- 60 9. 刃天青 (北京百灵威科技有限公司)
- 61 10. TE buffer (ZOMANBIO, Lot: 9BB08S)
- 62 11. 氯化铯 (BIOBOMEI, Lot: B23V1019)
- 63 12. 无水乙醇 (阿拉丁)
- 64 13. 核酸助沉剂 (ZOMANBIO)

65

66 仪器设备

- 67 1. 掌上离心机 (Scilogex)
- 68 2. 专用离心管 (Eppendorf Centrifuge 5430 R)
- 69 3. 超高速离心机 (Beckman Optima MAX-TL)
- 70 4. Qubit 定量 (Thermo Fisher Scientific)
- 71 5. qPCR 仪 (BIO-RAD CFX96 Real-Time System)
- 72 6. DNA 裂解仪 (MP Biomedicals FASTPREP-24)
- 73 7. SIP 分层泵 (Beckman)
- 74 8. 真空浓缩仪 (Eppendorf)
- 75 9. 振荡器 (Silentshake)
- 76 10. 折光率仪 (Atago)

77

78 软件和数据库

- 79 1. QIIME2 (扩增子分析主流软件)

系列教程: <https://docs.qiime2.org/2020.11/>

2. Megahit 组装软件

安装: `conda install -y megahit`

使用: `time megahit -t 30 \`

`-1 `ls 质控后序列路径/File_1.fq|tr '\n' ','|sed 's/,,$/'` \`

`-2 `ls 质控后序列路径/File _2.fq|tr '\n' ','|sed 's/,,$/'` \`

`-o YOUR PATH/megahit --k-min 21 --k-max 121 --k-step 10`

官方文档: 《MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph》Bioinformatics, 2015.

3. MetaWRAP 宏基因组分箱软件

安装: `conda install -c ursky metawrap -mg`

使用: `metawrap binning -o results/01_binning -t 30 -a YOUR PATH`

`TO/final.contigs.fa \`

`--metabat2 --maxbin2 --concoct YOUR PATH TO 指控后序列`

`/File_clean_*.fastq`

官方文档: 《MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis》Microbiome, 2018.

4. Kneaddata 质控软件

安装: `conda install -y kneaddata`

使用: `time kneaddata -i seq/sample_1.raw.fq.gz -i seq/ sample _2.raw.fq.gz \`

`-o temp/qc -v -t 30 --remove-intermediate-output \`

`--trimmomatic ~/miniconda2/envs/metagenome_env/share/trimmomatic/`

`--trimmomatic-options 'SLIDINGWINDOW:4:20 MINLEN:70' \`

`--bowtie2-options "--very-sensitive --dovetail" -db YOUR PATH TO/db/bowtie2'`

5. Prokka 细菌基因组注释

安装: `conda install -y prokka`

使用: `time prokka YOUR PATH TO/megahit/final.contigs.fa --outdir temp/prokka`

`\`

`--prefix mg --metagenome --kingdom Archaea,Bacteria,Mitochondria,Viruses \`

`--force --cpus 30`

官方文档：《Prokka: rapid prokaryotic genome annotation》Bioinformatics, 2015.

6. Salmon 基因定量工具

安装：conda install -y salmon

使用：time parallel -j 3 \

'salmon quant -i temp/salmon/index -l A -p 40 --meta \

-1 YOUR PATH TO 指控后序列/File_clean_1.fastq \

-2 YOUR PATH TO 指控后序列/File_clean_1.fastq \

-o temp/salmon/Sample_output.quant'

官方文档：《Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference》Nature Methods, 2017.

说明

1. 本文以最简单的乙酸降解功能菌分析流程进行介绍，读者可根据实际需要将标记物由 ^{13}C 乙酸替换成其他标记物，如 ^{13}C 乳酸、 ^{13}C 碳酸氢钠、 ^{13}C 甲苯等。
2. ^{13}C 标记物建议直接购买商用产品。
3. 本文描述的 ^{12}C 标记物即为普通试剂。
4. 以下实验操作中所用的枪头、离心管、西林瓶等耗材均需经过高压灭菌处理。

实验步骤

一、建立共培养微宇宙（以乙酸降解菌群为例）

1. 共培养实验设计分组如表 1 所示：

表 1.实验分组设计

组别	设定	目的	土量	培养液 (MSM)	乙酸	平行编号
A	新鲜土壤+乙酸 (^{13}C)	^{13}C 实验组	2 g	100 mL	10 mM	A-①/②/③
B	新鲜土壤+乙酸 (^{12}C)	^{12}C 对照组	2 g	100 mL	10 mM	B-①/②/③

134

135 2. 配制 1 L 矿物盐溶液 (MSM, 配方见表 2), 利用高压灭菌锅将 MSM、曝气所需针
136 头、西林瓶、胶塞灭菌, 然后冷却备用; MSM 当天配当天使用;

137 3. 称量 2 g 土壤于已灭菌的西林瓶中, 再加入 100 ml 的 MSM 和 100 ul 的 1 g/l 的
138 刃天青使用液;

139 4. 用氮气曝气脱氧 30 min 后用胶塞塞紧, 再用匹配的铝盖胶塞密封;

140 5. 根据不同分组, 利用注射器分别添加 ^{13}C -乙酸或 ^{12}C -乙酸至西林瓶内, 充分
141 摇匀;

142 6. 微宇宙放入恒温摇床, 在 30 °C 下进行共培养, 用于标记乙酸降解菌群;

143 7. 参与乙酸降解的菌群, 可将 ^{13}C 合并至自身 DNA, 从而完成标记。

144 二、提取 DNA

145 1. 待反应结束后, 破开微宇宙, 静置沉淀一段时间后倒掉部分上清液, 将剩余泥水
146 混合物小心转移至 50 ml 离心管后离心 15min (7500 x g);

147 2. 倒掉上清液, 仅保留湿土用于 DNA 的提取;

148 3. 建议使用 DNA 提取试剂盒提取样品的基因组 DNA;

149 4. 利用 Qubit 测定所提取的 DNA 浓度。

150 三、超速离心分层

151 1. 经 ^{13}C 标记后的 DNA 密度相对于未标记的 DNA 更大, 可由超速离心机进行分
152 离;

153 2. 用 TE buffer 和氯化铯粉末制备 CsCl 预配液 (折光率 $\text{RI} = 1.4032$) 和 CsCl 平衡溶
154 液 ($\text{RI} = 1.4020 \pm 0.0001$);

155 注: 每个样品需要大约 5 ml 的 CsCl 预配液, 根据所需样品配制足量的溶液, 如离
156 心 6 个样品, 制备 40 ml 即可; CsCl 平衡溶液所需体积较少, 制备 5 ml 左右即
157 可。

158 3. 添加 CsCl 预配液至 SIP 专用离心管刻线处, 示意图如下:

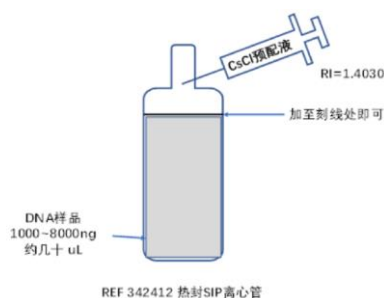


图 1. CsCl 预配液添加量示意图

4. 根据提取的 DNA 浓度，取总量 3000 ng 的 DNA 至 SIP 专用离心管中；
5. 离心管简单密封后涡旋震荡 20 s，混匀 CsCl 预配液与 DNA 样品；
6. 混匀后轻轻转动离心管，严格消除管内气泡；
7. 测定混合液的 RI，调节 RI 至 1.4020 ± 0.0001 。弱 RI 较低时可添加 CsCl 预配液增浓，较高时添加 TE buffer 稀释；
8. 注入 CsCl 平衡溶液至瓶口位置，完成封液；示意图如下：



图 2. CsCl 平衡溶液的封液示意图

9. 将 SIP 离心管两两配平 (两管间误差需小于 0.0010 g)，将离心管口热封后，对称放入转子中；配平时往较轻的离心管中添加 CsCl 平衡溶液，不可从管中取出，避免 DNA 样品的损失；
10. 超速离心条件：57000 x g，48 h，20 °C；离心不足将导致 DNA 分层效果不佳；
11. 离心完成后尽快取出，平稳取出转子，进行下一步分层收集。

四、分层接管

1. 每一管超速离心样品预计分为 24 层；将提前灭菌的离心管写上对应编号；
2. 打开分层取样泵，每管超离样品分 24 层，每层取样 20 s，设定流速为 595 ul/min；

注：实验前可提前进行预实验，观察取样情况，调整合适的流速。

3. 小心剪去离心管瓶颈，注意瓶口要剪平；装入收集台并压紧，上提针头，扎破瓶底；
注：使用前可利用清水测试管路气密性，确保气密性良好再上样。
4. 开启取样泵，借用计时器手动接管，观察出液管路，以管路中连续液柱的第一滴液体滴出开始计算，每 20 s 手动移动至下一分层管，最后一滴可快速靠壁防止洒漏；
5. 依次收集各个样品的 1-24 层 DNA (fraction, F1-F24, 由重至轻)；
6. 不同超离管样品采集的间隙需使用无菌水清洗管路，防止相互污染；
7. 测定各层样品的折光率 RI 值，并将 RI 值转换成浮力密度 (Buoyant Density; BD)。
注：若分层效果较好，各层 RI 应呈现等差递减的趋势。
 $BD \text{ 值} = RI * 10.8601 - 13.4974$ 。
8. 各层样品收集与 RI 测定示意图，建议以 RI=1.4002 为界，选取上下 10 层进行下游分析。

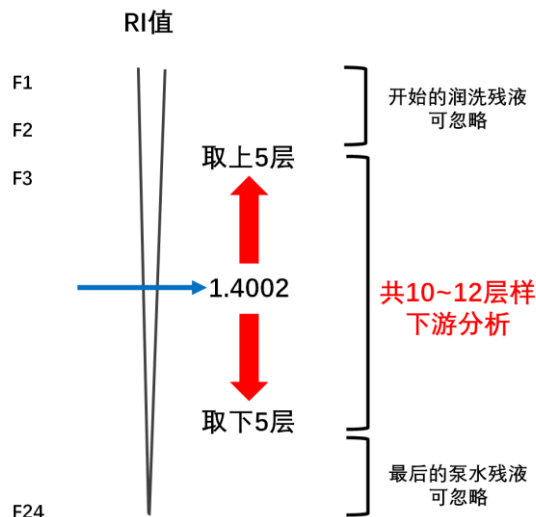


图 3. 各层样品收集示意图

五、纯化回收 DNA

1. 配制纯化溶液 (配方见表 3)，根据接管所得样品数量配制总纯化溶液；
注：可以多预算 5 个样品量的体积，防止分加纯化溶液过程中损失。

2. 往各层收集样品中加入 1106 ul 纯化溶液，置于 4 °C 冰箱过夜反应 (帮助核酸沉淀)；

注：由于每管的第 1、2 层和第 23、24 层误差较大故可不进行纯化回收，仅回收剩余 20 层。

3. 各分层管子离心 30 min (11000 x g, 4 °C)，离心完后缓慢倾斜离心管小心倒掉管中液体；

4. 将离心管放入真空浓缩仪内进行干燥 (D-AL 模式，45 °C，20 min)；直至离心管内已无肉眼可见水珠；

5. 干燥完毕后，加 30 ul ddH₂O 或 TE buffer，震荡混匀，用掌上离心机离心 5 s，放入 -20 °C 冰箱内长期保存；

注：若短时间内马上进行下游分子生物学分析可先放在 4 °C 冰箱中保存。

6. 纯化后再次测定各层样品的浮力密度 BD 值。

六、荧光定量 qPCR 分析

1. 利用 qPCR 方法对各层 DNA 量进行定量测定，以判断标记、分层效果；

2. 选取纯化回收之前所得的浮力密度 BD 值在 1.69 - 1.78 之间的分层 DNA 进行测试；

3. qPCR 检测的目标基因可以是 16S rRNA (乙酸为例)，或其他功能基因；

4. 将 qPCR 所需试剂和选取的 DNA 样品提前放入 4 °C 冰箱解冻、离心摇匀；

5. 按照 qPCR kit 说明书配置 20 ul 反应溶液；

6. qPCR 模板包括各层样品、标注曲线、阴性对照，且均需三个实验平行；

qPCR 反应程序根据目标基因进行设定，本文使用 16S rRNA 检测程序如下：

Step1: 95 °C，15min (预变性)

Step2: 94 °C，15 s；55 °C，30s ；72 °C，30 s；44 个循环 (扩增)

Step3: 72 °C 7 min

Step4: 起始温度 94 °C，终止温度 94 °C，温度变化速率为 0.5 °C (溶解曲线)

Step5: 读板

记录 qPCR 结果，换算得到各层 DNA 样品的基因浓度，绘制渐变曲线。

七、16S rRNA 测序与生物信息学分析

1. 参考 qPCR 结果, 若 ^{13}C -乙酸 ^{12}C -乙酸两组样品错峰明显, 表明标记效果较好, 可继续对各层 DNA 样品进行扩增子测序;
2. 为减少测序成本, 可根据 qPCR 结果选取各峰尖、及左右各一层, 共 3 个样品测序。
3. 使用 Earth Microbiome Project 推荐的引物对基因组 DNA 的 16S rRNA 片段进行高通量测序 (Illumina MiSeq PE250); 主要分析细菌群落组成;
4. 下机数据主要使用 QIIME2 分析, 包括序列除杂、过滤、拼接、去噪、注释等主要过程;
5. 根据物种丰度统计表, 绘制各样品不同分层下的物种变化图。

八、宏基因组测序与生物信息学分析

1. 经过 ^{13}C 标记的重层 DNA 大大简化了乙酸降解功能菌群的生物信息, 从而减少冗余数据的干扰, 极大地降低了生物信息分析难度;
2. 重点针对重层 DNA 进行宏基因组测序 (Illumina HiSeq PE250), 测序数据量建议不低于 6 G/样品;
3. 下机数据主要使用 Kneaddata 进行质控、除杂、过滤;
4. 利用 Megahit 软件对优化序列进行拼接;
5. 使用 MetaWRAP 进行分箱 (包括 metaBAT2、CONCOCT、MaxBin2 三种主流方法), 将短序列逐一组装成单菌基因组草图 (Bins), 建议拼装完整度 > 70 %, 污染率 < 10 %;
6. 使用 CheckM 评估组装得到的 Bins 的质量, 如基因组大小、完整性、杂合度等主要参数;
7. 使用 Salmon 对各个 Bins 进行基因定量;
8. 使用 Prokka 软件预测目标 Bins 的基因结构;
9. 利用 Diamond 软件作 KEGG/ COG/eggNOG 基因功能注释, 分别获得 KO 和 EC 号后, 进一步对功能基因作富集分析和通路分析;
10. 重点关注与乙酸降解密切相关的功能基因或代谢途径。

结果与分析

一、共培养标记后的 qPCR 检测结果

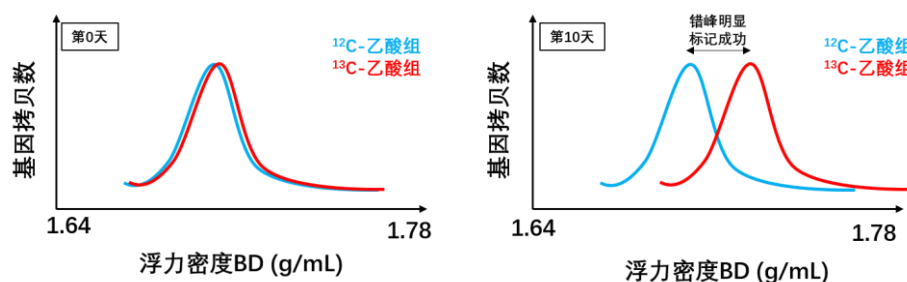


图 4. 定量 qPCR 用于指示 DNA-SIP 标记效率的示意图

图 4 所示：实验开始阶段（第 0 天），乙酸降解功能菌的基因组 DNA 尚未得到 ^{13}C 标记，故实验组与对照组的各层中基因拷贝数浓度较为一致。随着标记时长逐渐增加（第 10 天），乙酸降解功能菌的基因组 DNA 被 ^{13}C 标记，从而“变重”，故其各层中的基因拷贝数将向浮力密度更大的“重层”迁移，表现为与 ^{12}C -乙酸对照组明显错峰，即为标记成功。

二、16S rRNA 测序结果

如何挑选测序样品：

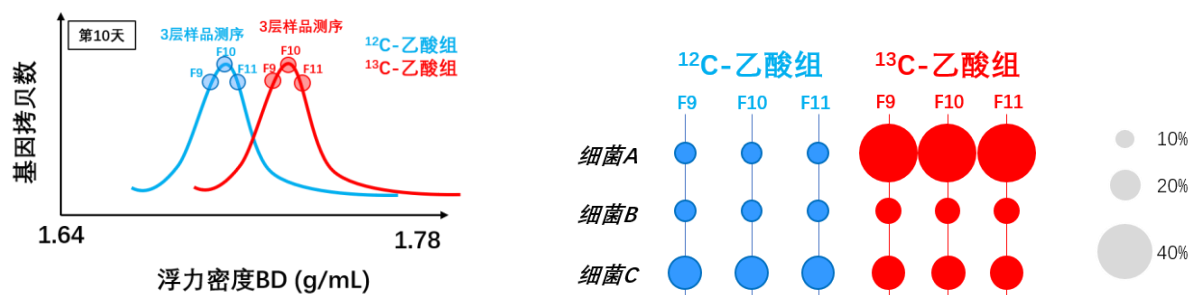


图 5. 标记样品的选择与高通量测序

图 5 所示：重点选取每个离心样品的各峰尖层（F10）、及左右两层（F9+F11）进行测序可减少测序成本。经过物种丰度统计后，对比发现细菌 A 在 ^{13}C 实验组中富集，而在 ^{12}C 对照组中丰度较低，说明细菌 A 是潜在的乙酸降解菌群。而细菌 B 和细菌 C 在两组中的丰度差异不大，不是潜在的乙酸降解菌群。说明：本文仅设置了一个对照组进行简要说明，读者可根据实际情况设置更多对照组，排除假阳性标记结果。

三、宏基因组分箱结果

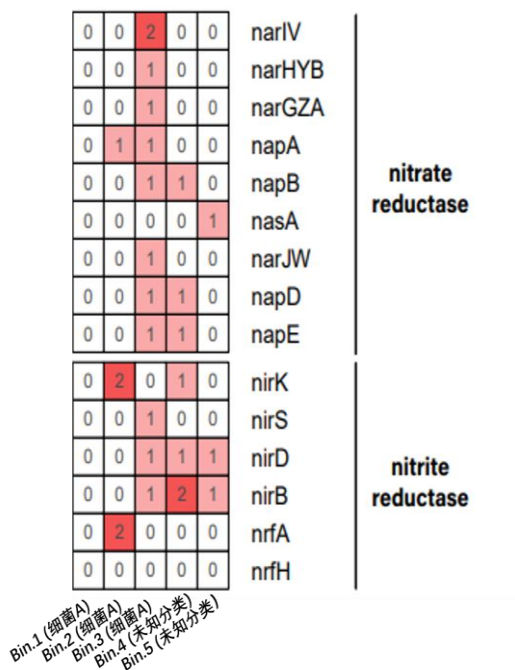


图 6. 宏基因组 Bins 的物种注释与功能基因注释结果

图 6 所示：重点针对“重层”DNA 样品进行宏基因组测序，并采用 MetaWRAP 分箱组装后一共获得了 5 个高质量的基因组草图 (Bins)。基于 KEGG、COG 等功能基因注释后，发现这些 Bins 含有特定的功能基因，可用于解释潜在的代谢机制。其中 Bin1、2、3 可注释至细菌 A，与 16S 扩增子结果较为吻合。同时还发现了两种新型细菌 (Bin4、5)也可能是潜在的功能细菌，可进一步构建系统发育树探究其物种分类及潜在代谢能力。

溶液配方

1. 微宇宙培养的 MSM 溶液配方 (以 1 L 计，用去离子水制备)

表 1. MSM 溶液配方

试剂	添加量
KH ₂ PO ₄	1.5 g
Na ₂ HPO ₄ ·12H ₂ O	10.55 g

NH ₄ Cl	0.3 g	
MgCl ₂	0.1 g	
混合维生素	1 mL	}
微量元素	1 mL	

2. 离心样品的纯化溶液配方

表 2. 纯化溶液配方

试剂	体积 (×每样)
无菌水	800 ul
分子级无水乙醇	300 ul
核酸助沉剂	6 ul

失败经验

1. 共培养实验需要通过预实验探索合适的标记时长，标记过久容易导致交叉污染 (Cross Feeding)，即被标记的功能细菌残体被其他非功能菌所捕食，导致非功能菌被错误标记而出现假阳性结果。

2. 判断标记效果可利用荧光、16S rRNA-qPCR、功能基因-qPCR 等方法综合判定。

致谢

感谢广东省科学院实施创新驱动发展能力建设专项 (2020GDASYL-20200103086)，国家自然科学基金资助项目 (42007357)，广东省基础与应用基础研究基金 (2019A1515110351)，中国博士后科学基金资助项目 (2019M662825 和 2020T130127)，广州市科技计划项目资助 (202002030271 和 202002020072)，广东省科学院实施创新驱动发展能力建设专项 (2019GDASYL-0103053)，广东省“珠江人才计划”项目 (2017BT01Z176 和 2017GC010570) 对本工作的大力支持。

参考文献

1. Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F. and Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*. 11(11): 1144-1146.<https://doi.org/10.1038/nmeth.3103>
2. Neufeld, J. D., Vohra, J., Dumont, M. G., Lueders, T., Manefield, M., Friedrich, M. W. and Murrell, J. C. (2007). DNA stable-isotope probing. *Nature Protocols*. 2: 860.<https://doi.org/10.1038/nprot.2007.109>
<https://www.nature.com/articles/nprot.2007.109.pdf>
3. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*. 35(9): 833.<https://www.nature.com/articles/nbt.3935>
4. Sun, W., Xiao, E., Häggblom, M., Krumins, V., Dong, Y., Sun, X., Li, F., Wang, Q., Li, B. and Yan, B. (2018). Bacterial survival strategies in an alkaline tailing site and the physiological mechanisms of dominant phylotypes as revealed by metagenomic analyses. *Environmental Science & Technology* 52(22): 13370-13380.<https://doi.org/10.1021/acs.est.8b03853>
5. Zhang, M., Li, Z., Häggblom, M. M., Young, L. Y., He, Z., Li, F., Xu, R., Sun, X. and Sun, W. (2020). Characterization of nitrate-dependent As(III)-oxidizing communities in arsenic-contaminated soil and investigation of their metabolic potentials by the combination of DNA-SIP and metagenomics. *Environmental Science & Technology*.<https://doi.org/10.1021/acs.est.0c01601>