



研究报告

油藏样本潜在核心功能微生物群的宏基因组挖掘

李志华¹ 隋维康¹ 向宝玉¹ 胡婧² 曹功泽² 吴晓灵² 林军章² 孙刚正²
张晓君^{*1} 张梦晖^{*1}

1 上海交通大学生命科学技术学院 微生物代谢国家重点实验室 上海 200240

2 中国石油化工股份有限公司胜利油田分公司石油工程技术研究院 山东 东营 257000

摘要:【背景】识别出具有驱油功能的内源性微生物群是当前设计高效微生物采油技术的迫切需求。【目的】通过宏基因组分析找出原位油藏微生物中具有产酸/产气功能的潜在核心微生物群。【方法】利用室内模拟体系,对原位油藏微生物进行有机营养(从有氧到无氧)的激活,运用生物信息、多元统计分析以及网络分析等方法,寻找潜在内源性核心功能微生物群。【结果】有机营养激活了原位油藏样本中以 *Bacillus licheniformis* 为核心的产酸菌群(包括 *Coprothermobacter proteolyticus*、*Marinobacter* spp.、*Anaerobaculum hydrogeniformans* 和 *Petrotoga mobilis*),该菌群具有以丙酮酸/乙酰辅酶 A 为原料发酵乳酸、乙酸以及甲酸的功能;激活了以 *Enterococcus faecium* 为核心的产气菌群(包括 *Shinella zoogloeoides*、*Paracoccus denitrificans*、*Paracoccus* spp. 和 *Enterobacter cloacae*),该菌群具备以(亚)硝酸盐、(亚)硫酸盐、石油烃为原料生产含氮/硫/碳等气体的能力。2 个核心菌群内部分菌同时具有产酸和产气通路基因,激活过程中 2 个菌群的丰度变化呈负相关关系。【结论】利用有机营养激活体系及宏基因组测序分析技术,筛选出了油藏样本中具备产酸、产气能力的核心功能菌群,为后续进一步的功能菌株研究提供了潜在靶标及功能指向。

关键词: 有机营养激活, 宏基因组, 核心功能菌群

Screening potential core functional microbiota of a crude oil sample through metagenomic analysis

LI Zhihua¹ SUI Weikang¹ XIANG Baoyu¹ HU Jing² CAO Gongze² WU Xiaoling²
LIN Junzhang² SUN Gangzheng² ZHANG Xiaojun^{*1} ZHANG Menghui^{*1}

1 State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

2 Research Institute of Petroleum Engineering Technology, Sinopec Shengli Oil Field Company, Dongying, Shandong 257000, China

Abstract: [Background] Understanding functional microbiota inside crude oil samples is crucial in microbiota design for enhancing oil recovery. [Objective] Identifying potential core functional microbiota from a crude oil sample that have ability to produce acid/gas through metagenomic analysis.

Foundation items: National Natural Science Foundation of China (31861133018, 32061133001)

***Corresponding authors:** E-mail: ZHANG Xiaojun: xjzhang68@sjtu.edu.cn; ZHANG Menghui: mhzhang@sjtu.edu.cn

Received: 10-11-2020; **Accepted:** 23-12-2020; **Published online:** 16-04-2021

基金项目: 国家自然科学基金(31861133018, 32061133001)

***通信作者:** E-mail: 张晓君: xjzhang68@sjtu.edu.cn; 张梦晖: mhzhang@sjtu.edu.cn

收稿日期: 2020-11-10; **接受日期:** 2020-12-23; **网络首发日期:** 2021-04-16

[Methods] The compositional and functional changes of the indoor activated crude oil samples were obtained with the metagenomic data. The potential endogenous core functional microbiota related to the production of gas and acid were obtained through bioinformatics, multivariate statistics and network analysis. **[Results]** The co-abundance analysis deduced that *Bacillus licheniformis*, along with *Coprothermobacter proteolyticus*, *Marinobacter* spp., *Anaerbaculum hydrogeniformis* and *Petrotoga mobilis* that clustered as a group, were the core functional bacteria relevant to the change of the pH value. Functional analysis of strains in this group through draft genome assembly indicated that the pathways for acid production, including for lactic acid, acetic acid and formic acid fermentation using pyruvate/acetyl coenzyme A, were enriched. *Enterococcus faecium* was considered as the core microbe in another group, which had *Shinella zoogloeoides*, *Paracoccus deniticans*, *Paracoccus* spp. and *Enterobacter cloacae*. Among these, *Enterobacter cloacae* was positively correlated with total gas production. The assembled bins in this group had the ability to produce nitrogen/sulfur/carbon using nitrate (nitrite)/sulfate (sulfite)/petroleum hydrocarbon. These two groups both had a small number of microbes that have dual ability for gas and acid production and the two groups were negatively correlated. **[Conclusion]** We had screened out core acid/gas producing microbiota in a crude oil sample through organic activation and metagenomic sequencing technology. These findings provided potential strain targets for further research.

Keywords: organic nutrient activation, metagenomics, core functional microbiota

石油作为一个国家重要的战略储备物资,其衍生的化工产品工业、民生、经济等领域扮演了重要的角色。目前全球剩余石油资源以稠油为主,占比 70%,约有 8 150 亿 t;稠油由于含有大量的长链烃且混杂胶质、沥青质、砂砾等,其粘度高(50–10 000 mPa·s)、流动性差,导致采收困难^[1]。气驱、水驱、化学驱等方法被用于提高石油的采收率^[2],但是这些方法对稠油的采收效率并不高。目前我国的稠油开采率仅占原油总产量的 16.2%^[2],亟须实现经济、高效的稠油开采。

微生物采油技术由于其绿色环保、低成本的特点,越来越受到关注。这种技术利用微生物的分解代谢能力,通过产酸、产气、产表面活性剂、产多糖和降解烃等多重方式来稀释稠油粘度,加强稠油的流动性,达到提高石油采收率的作用。在过去几十年里,微生物采油尝试了多种方式,包括:向油井投入某种能帮助产油的菌^[3]或者能促进微生物生长的营养物质^[4];利用合成生物学技术将一个菌改造成具备多种功能的“超级细菌”^[5–6];设计一组能够互养并且具备一定采油功能的“复配菌”注入油井^[7–9]等,以此来改性降粘,提高石油采收率。由于这些做法大都采用外源菌,而这些菌难以适应

油藏内部的复杂生态环境,其改善驱油效果常常不佳或者难以持久。从适应油藏极端环境的内源微生物中选取能形成互养关系并且具有多种不同驱油功能的微生物群,对其进行适当营养强化,在微生物采油应用中可能将具有更好的可行性以及更好的开采效率。

油藏内源性驱油微生物的选取首先需要了解油藏内微生物的组成及激活特点。本研究围绕一个典型性油藏样本,对如何寻找样本内的核心功能菌群进行了探索。由于微生物的产酸和产气功能是起到驱油作用的重要因素^[10],并且产酸/产气的观测较为方便且数据容易量化,本研究选取产酸/产气的微生物作为筛选的目标核心功能菌群,利用室内模拟油藏体系对油藏产出液样本进行了有机营养(从有氧到无氧)激活;在有氧培养末期、无氧培养的前期和后期激活时间点进行了产气、产酸的指标测量,并将对应的样本微生物进行了 HiSeq-X-Ten PE150×2 宏基因组测序。依据与油藏产出液原始样本测序数据的生物信息学与统计比较分析,以期发现油藏样本内微生物随激活时间点的组成演替以及功能变化特点;通过菌群丰度共变化分析,以期获得菌-菌间的相互关系以及这种相互关系形成的

聚类对产酸、产气的协同影响;最后依据发现的聚类功能菌群,通过对菌群内组装基因组草图的功能解析,以期进一步理清核心功能菌—功能基因—代谢通路的途径。

1 材料与方法

1.1 样本的采集和预处理

石油产出液原液样本来源于胜利油田南区一个油藏温度为 65 °C 的油田,从选定的油井井口取样。取样前使井口液体预先流出 1 min 以上,然后用无菌的采样桶接满井口产出液,密封后带回实验室。待样本静置分层后,将上层石油层和下层水相分离,备用。

1.2 主要试剂和仪器

胰蛋白胨、酵母提取物,赛默飞世尔科技公司;氯化铵、硝酸钠、磷酸氢二钠、磷酸二氢钾、可溶性淀粉、葡萄糖、氢氧化钠,国药集团化学试剂有限公司;刃天青,上海源叶生物科技有限公司。超净工作台,上海智城分析仪器制造有限公司;0.22 μm Millipore 滤膜,默克公司;pH 计, Mettler Toledo 公司;恒温摇床、离心机, Eppendorf 公司。

1.3 样本的有机营养激活

1.3.1 有机营养培养基的配制

有机营养培养基的配制采用游靖的配方^[11],由 2 种营养液混合而成。A 液:胰蛋白胨 5.00 g, 酵母提取物 5.00 g, NH_4Cl 5.35 g, NaNO_3 8.50 g, Na_2HPO_4 7.50 g, KH_2PO_4 2.80 g, 加入去离子水 200 mL, 于 1×10^5 Pa 灭菌 15 min。B 液:将 25.00 g 可溶性淀粉溶于 800 mL 去离子水中,于 100 °C 搅拌 30 min, 经 4 °C 放置过夜后取上清液,加入 25.00 g 葡萄糖,于 0.3×10^5 Pa 灭菌 15 min。A、B 液的混合在超净工作台中进行。用 0.22 μm Millipore 滤膜将 B 液过滤后,将滤液与 A 液混合,使用 1 mol/L NaOH 调节 pH 值至 7.0。

1.3.2 样本的有机营养激活

在无菌条件下,将 120 mL 产出液水相、3 g 左右石油以及 5 倍浓度 30 mL 有机培养基加入到

250 mL 培养瓶中;添加 2 滴 0.22 μm Millipore 滤膜过滤的氧气指示剂刃天青,胶塞密封后移入 55 °C、100 r/min 的恒温摇床进行培养;依据氧气含量状况以及游靖对油藏微生物功能菌的分离培养实验^[11],分别于有氧后期(A, 约培养 35 h, 快速生长期)、无氧前期(B, 培养 15 d, 菌浓平稳期)和无氧后期(C, 培养 47 d 后, 菌浓下降期)进行采样。实验另取 3 个原位油藏的平行样本(T01/T02/T03)作为对照。

1.3.3 样本总产气量及产酸的测量

每个采样时间点设置 6 个生物学重复。其中 3 个平行样本用于总产酸、产气的分析,另 3 个平行样本用于宏基因组测序。pH 值直接通过 pH 计测量。气体产生总量通过排水法测量,负压则认为其不具有产气功能。

1.4 DNA 的提取与宏基因组测序

1.4.1 样品的预处理

所有样本在培养之后于 4 °C、10 000 r/min 离心 15 min, 将收集到的上清液置于 50 mL 离心管,用提前在 4 °C 预冷的 PBS 重悬沉淀,重悬液在相同条件下离心 15 min, 再次取上清到 50 mL 离心管,沉淀保存在-80 °C,收集的上清以 0.22 μm 滤膜抽滤,保存滤膜在-80 °C。

1.4.2 DNA 的提取

DNA 的提取采用 CTAB 法^[12],分别对滤膜和沉淀进行提取,将提取后的 DNA 混合在一起,在 HiSeq-X-Ten PE150 \times 2 测序平台按测序服务商提供的标准流程进行宏基因组测序。

1.5 数据生物信息分析

1.5.1 测序数据的质控

原始测序数据的质控使用 PRINSEQ^[13]软件进行,先从序列的 5'端开始将第一个及其后的 Q 值小于 20 的碱基切除;再去除序列长度小于 60 bp 以及含模糊碱基“N”的序列。

1.5.2 基于 MetaPhlAn2 的组成分析

将上述质控后的高质量序列数据通过 MetaPhlAn2^[14]与其预设数据库(含 17 000 个细菌及古菌参考基因组)中的微生物标记(Marker)基

因比对, 获得每个样本内的鉴定物种分类以及丰度信息。

1.5.3 基于 MetaCyc 数据库的通路分析

使用 Bowtie2^[15]将物种特定的序列比对到 ChocoPhlAn 数据库, 进行快速的泛基因组功能注释; 使用 Diamond^[16]软件将未比对上的序列与 UniRef50 数据库进行比对, 识别其蛋白家族 (Protein Family)。经过 HUMAnN2 核心算法在 MetaCyc^[17]数据库中匹配代谢通路, 得到每个样本包含基因家族 (Gene Family) 的丰度、通路的表达量以及通路覆盖度的信息。以上步骤通过 HUMAnN2 流程^[18]完成。

1.5.4 基于宏基因组数据的基因组草图组装

将所有样本的高质量序列集合在一起, 使用 MEGAHIT^[19]默认参数进行共组装。使用 MetaBat2^[20]对组装出的片段 (Contigs) 进行装箱 (Binning), 设定最短片段为 2 500 bp。使用 CheckM^[21]软件 Lineage_wf 模块对装箱得到的 Bins 进行质控, 获得 70% 完整度以上的 Bins。使用 MetaBinG2^[22]对 Bins 进行分类鉴定到种水平。

1.5.5 基于基因组草图的代谢途径及子系统模块 (Subsystem) 分析

从 Kyoto Encyclopedia of Genes and Genomes (KEGG, <https://www.kegg.jp/>) 数据库、Rapid Annotation using Subsystem Technology (RAST, <https://rast.nmpdr.org/>) 数据库的通路及子系统模块中收集产乳酸/甲酸/乙酸、产含氮气体、含硫气体、含碳气体的通路、模块以及其中的基因。

通过 Prodigal^[23]预测基因组草图中的编码基因和蛋白文件, 并分别上传至 KEGG/RAST 网站进行基因组草图的功能注释。通过 RAST 注释识别基因组草图所包含的子系统模块, 筛选出样本覆盖到的产酸、产气功能模块及对应的基因家族, 利用 KEGG 重构产酸、产气相关的代谢途径。

1.6 统计分析

1.6.1 基于共丰度变化关系的菌群聚类

聚类计算在 Matlab[®] (R2019b, MathWorks) 环

境下使用其自带的程序进行。对于在 1.5.2 中获得的不同激活时间点的菌群分类及丰度矩阵, 计算每个菌-菌之间的 Spearman 关联系数及其 *P*-Value; 依据 Spearman 关联系数的相似度差异逐级进行菌-菌间的聚类, 从聚类树的最高点开始, 依次进行聚类间差异分析, 根据 $P < 0.05$ 的标准将菌群分为不同的聚类。

1.6.2 互作网络关系建立

将 1.6.1 中菌群两两之间的 Spearman 相关系数绝对值 > 0.5 并且同时满足 $P\text{-Value} < 0.05$ 的菌保留, 使用 CytoScape^[24]进行网络构建。

2 结果与分析

2.1 有机营养激活条件下油藏微生物的产酸、产气功能

用有机营养辅以一次性充入氧气对原位油藏样本进行激活, 激活体系模拟油藏原位状态历经有氧阶段和无氧阶段。体系内的总产气量在有氧后期以及无氧前期阶段逐渐增加, 而在无氧后期阶段总气体量减少 (图 1A), 说明产气过程主要发生在有氧阶段和无氧前期阶段。原位油藏样本的 pH 值为 7.84 ± 0.01 , 体系内 pH 值在有氧后期阶段降至 5.17 ± 0.05 , 在无氧阶段略有上升但变化不大, 说明产酸过程主要发生在有机营养有氧的阶段 (图 1B)。

2.2 营养激活条件下菌群的组成及演替

本研究中有 12 个样本的宏基因组平均测序量为 $8.03 \times 10^7 \pm 7.36 \times 10^6$ (均值 \pm 标准差) 条序列 (Reads), 模糊碱基率小于 0.002%, Q_{30} 碱基比例为 $96.71\% \pm 0.42\%$ 。使用 MetaPhlAn2 软件在原位油藏样本和激活样本中共识别到 48 个菌属 (种), 其中, 有 29 个识别到种, 19 个识别到属。原位油藏样本序列与数据库的比对率 (Mapping Rate) 为 59.48%, 有氧后期、无氧前期和无氧后期 3 个激活时间点的序列比对率分别为 81.02%、25.29% 和 2.13%, 大量的序列在无氧期无法从数据库获取比对信息。识别到的 48 个菌属 (种) 按照丰度共

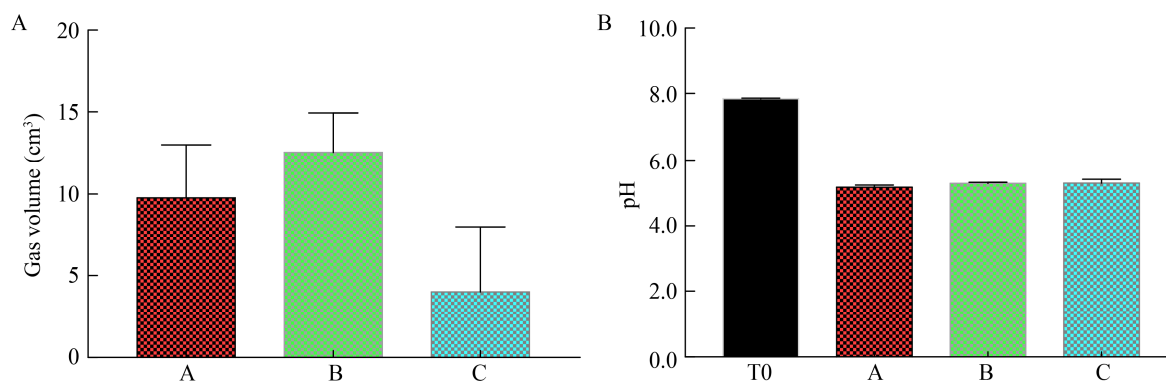


图1 营养激活条件下体系内总产气体积(A)和pH值(B)的变化

Figure 1 The changes of total gas volume (A) and pH value (B) under the organic activation

注: 误差线表示标准差。T0: 原位样本; A: 有氧后期; B: 无氧前期; C: 无氧后期

Note: The error bar represents the standard deviation. T0: *In-situ* sample; A: Aerobic later period; B: Anarobic earlier period; C: Anarobic later period

变化的 Spearman 关联, 可以划分为 6 个聚类 (Cluster) (图 2)。

在能够比对到数据库的菌中, 原位油藏样本中的优势菌为 *Pseudomonas stutzeri* (89.69%±0.40%) 和 *Marinobacter* spp. (8.26%±0.35%)。 *Cellvibrio* spp. 的丰度排第三, 占比为 1.36%±0.08%, 其余的物种总量仅占不到 1%。

有机营养的激活使原油优势菌属 *Marinobacter* spp. 在有氧后期阶段丰度显著下降至 0.09%±0.02% ($P<0.05$, t 检验), 并且大多数在原油中丰度次高的菌也无法存活, 如 *Cellvibrio* spp. 等的丰度降为 0。然而 *Bacillus licheniformis* 在有氧后期显著上升至 75.22%±3.94% ($P<0.05$), 成为新的优势菌种; *Enterococcus faecium* 的丰度在有氧后期显著上升 ($P<0.05$), 占比为 3.46%±0.67%。 *E. faecium* 在无氧前期升至 82.63%±13.82%, 其所在的聚类 (Cluster 5) 内的其他菌群的丰度均显著上升 ($P<0.05$); 与此同时丰度显著上升 ($P<0.05$) 的聚类菌群还有 Cluster 2 和 Cluster 4, 其中 Cluster 2 包含 *Escherichia coli*、*Thermus thermophilus* 等菌, Cluster 4 含常见的烃分解菌 *Nocardia farcinica* 等。 Cluster 6 的优势菌 *B. licheniformis* 及该聚类内的其他菌在无氧前期的丰度均显著下降 ($P<0.05$), 但在

无氧后期阶段再次上升, 其中 *B. licheniformis* 的丰度显著升高至 24.52%±32.19% ($P<0.05$), 仅次于 *P. stutzeri* (34.82%±24.13%)。然而 *E. faecium* 及其所在的聚类 (Cluster 5) 的菌群丰度在无氧后期则显著下降 ($P<0.05$)。原油优势菌种 *P. stutzeri* 虽然在有机营养激活后丰度显著下降 ($P<0.05$), 但其在激活过程中一直维持较高丰度, 在无氧后期阶段又重新成为优势菌。

根据菌群的共丰度变化, 使用 CytoScape 构筑菌群的互作关系网络。根据产酸与菌群的关联关系 (图 3) 来看, Cluster 6 的 *B. licheniformis* 与 pH 值的变化显著负相关。营养激活体系内产酸主要发生在有氧后期, *B. licheniformis* 作为有氧后期出现的优势菌, 其丰度占比高达 75.22%±3.94%, 而且其与 Cluster 6 内的关联菌均存在显著正相关关系 ($P<0.05$, Spearman 相关分析), 因此推测 Cluster 6 是以 *B. licheniformis* 为核心的产酸功能菌群。 Cluster 5 的 *Enterobacter cloacae* 与产气量显著正相关 ($P<0.05$), 该菌只与 Cluster 5 的 *E. faecium* 呈显著正相关关系 ($P<0.05$)。

E. faecium 在有氧后期丰度显著上升 ($P<0.05$), 在无氧前期升至 82.63%±13.82%, 而营养激活体系在有氧后期及无氧前期的产气量是逐步上升的, 因

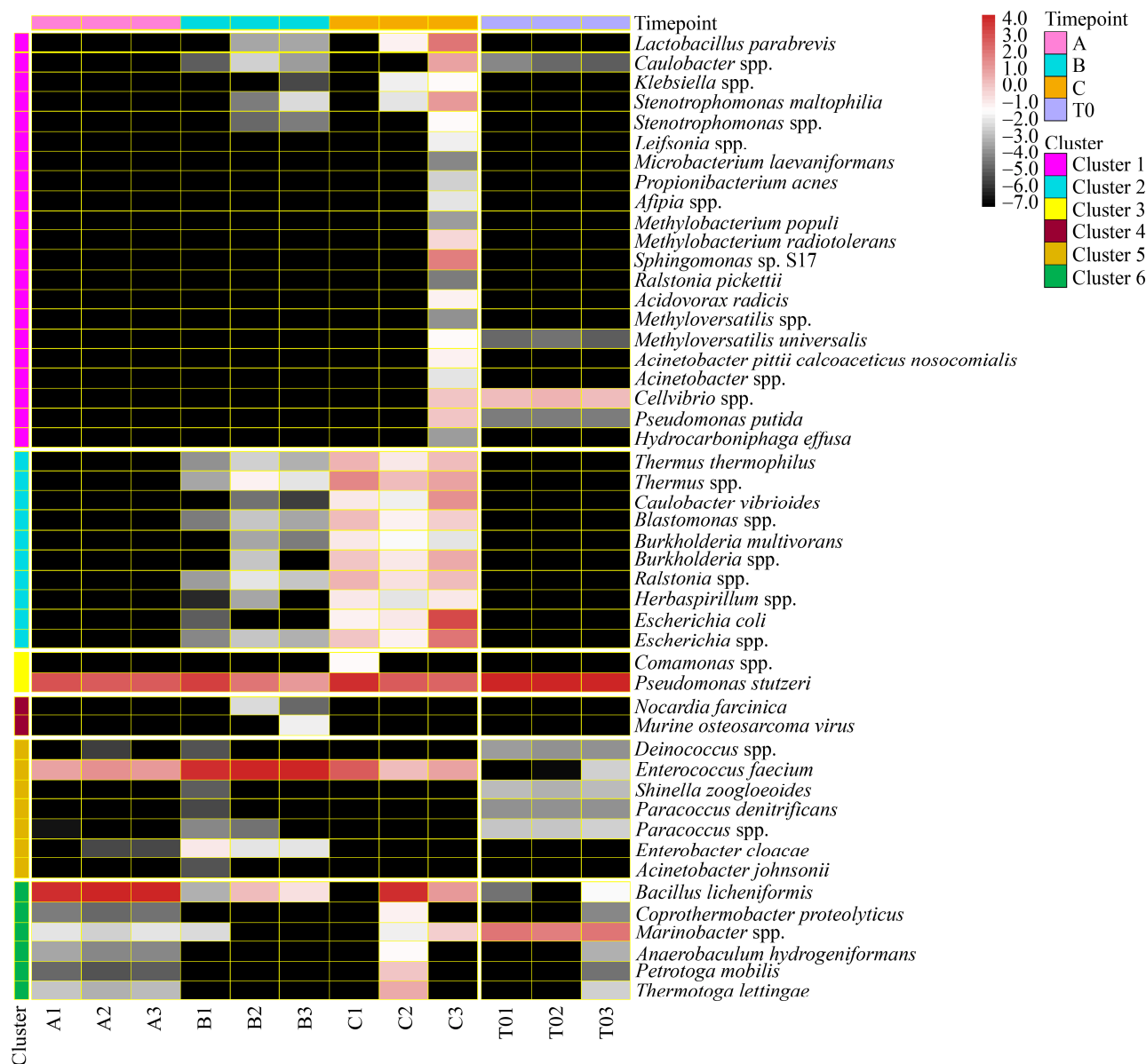


图 2 原位油藏样本及有机营养激活下不同时间点菌群的组成

Figure 2 The microbiota composition of the oil samples at different timepoints under the organic activation

注：菌群按照丰度变化的 Spearman 相关系数划分为 6 类，按聚类进行排序，热图中每个菌的相对丰度百分比经过 Log_{10} 处理

Note: The microbiota were classified into six clusters according to the pair-pair spearman correlations of the abundances, and ordered by the rank of clusters. The relative percentage of the abundance for each bacterium was Log_{10} -transformed

此推测 Cluster 5 是以 *E. faecium* 为核心的产气功能菌群。值得注意的是，*E. faecium* 还与 pH 值呈显著的正相关($P<0.05$)，其丰度变化与降低 pH 值正相关的 *B. licheniformis* 及其所属聚类中的所有菌群存在显著负相关关系($P<0.05$)。

2.3 油藏样本中产酸相关菌群的识别与功能鉴定
为了验证存在以 *B. licheniformis* 为核心的产酸功能菌群，使用 MEGAHIT 对样本宏基因组测序数据进行基因组草图的组装，并且使用 MetabinG2 识别出种属地位。将被识别为与 pH 值

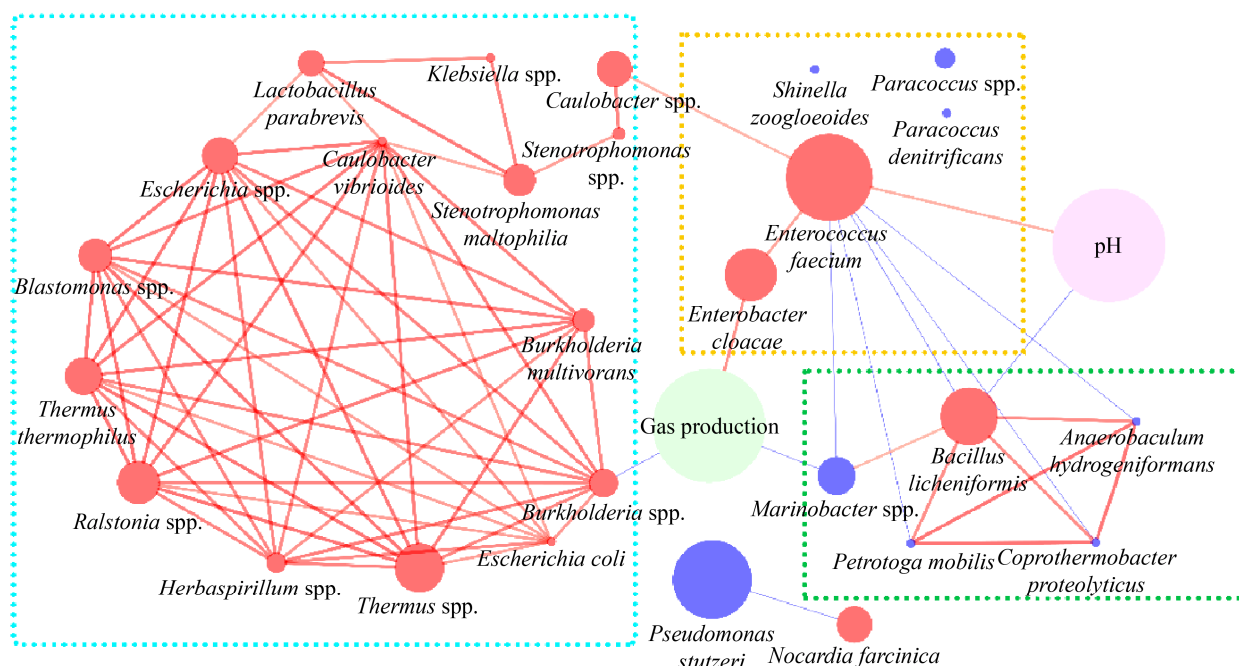


图3 有机营养激活下菌群间的丰度共变化网络以及与产酸、产气的关系

Figure 3 The co-abundance network of the microbiota and their relationships with the total gas volume and the pH value under the organic activation

注：节点为菌属(种)分类地位，其大小表征丰度，红色节点代表丰度增加，蓝色节点代表丰度降低；边表征菌属(种)间的相关性，红色表示正相关，蓝色表示负相关，其粗细表征相关系数(丰度排在前十的菌属作为节点保留)，框线颜色代表框线内菌群所属的聚类

Note: The node represents the taxonomy of microorganisms (abundance: red: up, blue: down), and its size represents the abundance. The edge represents the correlation between microbiota (red: positive, blue: negative), and its thickness represents the correlation coefficient (the top ten genera/species are reserved). The color of the frame represents the classified cluster

显著负相关的 *B. licheniformis* 所有菌株的基因组草图组合并作为该物种的基因组草图集合(泛基因组)，上传到 RAST 与 KEGG 网站进行功能分析。通过代谢途径以及子系统模块注释，识别到了发酵产乳酸(Fermentation: Lactate)以及发酵产甲/乙混合酸(Fermentations: Mixed Acid)模块，发现可以建立一条完整的由丙酮酸/乙酰辅酶 A 出发发酵产生乳酸/乙酸/甲酸的代谢途径(图 4)。因此，有完整产乳酸/混合酸子系统模块的 *B. licheniformis* 被认为在供能充足的情况下具有自主发酵混合酸的代谢能力。此外，在 *B. licheniformis* 同一聚类(Cluster 6)中的 *Coprothermobacter proteolyticus*、*Marinobacter* spp.、*Anaerobaculum hydrogeniformans*、*Petrotoga mobilis* 的组装基因组中均发现了产乳酸

或混合酸模块的基因富集，而在另外一个与产气功能正相关的聚类(Cluster 5)的 *E. faecium*、*Shinella zoogloeoides*、*Paracoccus denitrificans*、*Paracoccus* spp. 的组装基因组中也发现了少量产酸模块的富集。这 2 个聚类中的菌群都具有利用丙酮酸/乙酰辅酶 A 为营养的通路，提示 2 个聚类菌群之间可能在丙酮酸/乙酰辅酶 A 上存在营养竞争关系。

2.4 油藏样本中产气相关菌群的识别与功能鉴定

稠油油藏中含有最多的是以烷烃、芳香烃、各类不饱和和烃为主的碳氢化合物，此外还有氮、硫等元素^[25-26]，因此，识别产气菌群的重点在于寻找能产含碳气体、含氮气体及含硫气体的菌群。RAST 数据库具有硫代谢系统、氮代谢系统及呼吸代谢系统功能模块，在其中找到了许多油藏微生

Subsystem	Inorganic sulfur assimilation						Denitrifying reductase gene clusters				Ammonification	Formate hydrolase	Fermentations: mixed acid				Fermentations: lactate
Input compound	SO ₃ ⁻	APS	PAPS	APS	SO ₂ ⁻	SO ₂ ⁻	NO ₃ ⁻	NO ₂ ⁻	NO	N ₂ O	NO ₂ ⁻	Formate	Pyruvate	Acetyl CoA	Acetyl phosphate	Pyruvate	
Output compound	APS	PAPS	SO ₂ ⁻	SO ₂ ⁻	H ₂ S	NO ₂ ⁻	NO ₂ ⁻	NO	N ₂ O	N ₂	NH ₃	CO ₂	Formate	Acetyl phosphate	Acetate	Lactate	
Gene	PAPSS	PAPSS	Cys H	Apr AB	Cys JI	Nar GHI	Nir K	Nir S	Nos Z	Nir BD	Nir 6	FDH	PFL	PTA	ACK	LDH	
Taxonomy	SAT	Cys C			SIR	Nap AB						FDN					
<i>E. faecium</i>																	
<i>S. zoogloeoides</i>																	
<i>P. denitrificans</i>																	
<i>Paracoccus</i> spp.																	
<i>E. cloacae</i>																	
<i>B. licheniformis</i>																	
<i>C. proteolyticus</i>																	
<i>Marinobacter</i> spp.																	
<i>A. hydrogeniformans</i>																	
<i>P. mobilis</i>																	

Sulfur gas

Nitrogen gas

Carbon gas

Acidogenic module

图 4 产酸、产气相关潜在核心功能菌群及其对应的基因/功能通路/子系统模块
Figure 4 The potential core functional microbiota and the genes/pathways/subsystems they carried relevant to the production of gas and acid
注：菌名左侧颜色对应 Cluster 划分

Note: Taxonomy was classified according to the color of clusters

物能够利用油藏内营养产气的子系统模块及其通路和基因(与 KEGG 数据库关联),发现与产气量显著正相关的 *E. cloacae* 具有产含碳/氮/硫气体的所有完整的子系统模块(图 4)。*E. cloacae* 是样本菌群中唯一具有能够完整地将稠油中的硫酸盐及亚硫酸盐转化为硫化氢的无机硫同化(Inorganic Sulfur Assimilation)模块的菌。利用 KEGG 数据库重构出了 2 条完整的硫化氢代谢途径: $\text{SO}_3\text{—APS—PAPS—SO}_2\text{—H}_2\text{S}$ 途径(PAPS 途径)以及 $\text{SO}_3\text{—APS—SO}_2\text{—H}_2\text{S}$ 途径(APS 途径)。这 2 个代谢途径的第一步都是通过 *PAPSS/SAT/Cys ND* 将硫酸盐转化为 APS。PAPS 途径首先通过 *Cys C/PAPSS* 将 APS 转化为 PAPS, 然后通过 *Cys H* 催化 PAPS 转化为亚硫酸盐, 而 APS 途径直接通过 *Apr AB* 将 APS 转化为亚硫酸盐。此外, *E. cloacae* 中还具有将(亚)硝酸盐还原至氮气($\text{NO}_3\rightarrow\text{NO}_2\rightarrow\text{NO}\rightarrow\text{N}_2\text{O}\rightarrow\text{N}_2$)的反硝化还原酶基因簇(Denitrifying Reductase Gene Clusters)模块、将(亚)硝酸盐转化为氮气的(亚)硝酸盐氨化(Nitrate and Nitrite Ammonification)模块以及产二氧化碳的甲酸脱氢酶(Formate Dehydrogenase)模块, 并且该菌含有的基因能够重构出模块对应的完整的代谢途径。值得一提的是, Cluster 5 和 Cluster 6 这 2 个聚类丰度存在负相关关系, 而产酸菌群中的 *B. licheniformis* 和 *Marinobacter* spp. 与产气菌 *E. cloacae* 所在的聚类(Cluster 5)中的菌群都需要利用硝酸盐为营养, 提示 2 个聚类菌群之间可能在硝酸盐上存在营养竞争关系。

3 讨论与结论

原位油藏微生物由于处于极端环境,给准确分析其组成与功能带来很大的困难。本工作利用有机营养激活体系,结合宏基因组测序数据分析试图找出原位油藏微生物中具有产酸/产气功能的潜在核心微生物,为微生物采油的功能菌株研究提供靶标,有助于为制定更有效的微生物采油工艺提供理论指导。

本研究前期发现,由于目前石油微生物在公共数据库的功能信息相对匮乏,直接依据数据库进行油藏样本的宏基因组序列比对时有大量的序列无法被识别,只能找到非常少的驱油功能相关通路。然而直接利用基因组草图进行菌群组成与功能分析往往因为测序量及菌株组装效率的限制,只能拿到部分高质量基因组草图信息。

为了克服以上缺陷,本研究首先使用 MetaPhlAn2 对宏基因组进行数据库比对,尽可能拿到较多的微生物物种鉴别及相对丰度信息;再通过菌群丰度共变化对识别出的菌群进行聚类,获取菌群间的相互关联信息;并依据每个菌与产气、产酸的关系,推测出产气、产酸的功能菌群聚类;最后以聚类为单位,通过组装的基因组草图识别功能菌群的子系统功能模块并重构代谢途径,分析功能菌群的协同功能。基于这个分析策略,成功找到了油藏样本中以 *B. licheniformis* 为核心的产酸功能菌群、以 *E. cloacae* 为核心的产气功能菌群,以及它们在激活过程中的演替关系。

由于本研究的主要目的是探究如何寻找原位油藏潜在功能微生物的方法,因此没有对找出的产酸/产气菌进行驱油功能验证,只是对筛选出的功能微生物群进行了文献检索。在发现的产酸菌群中,有关 *B. licheniformis* 的报道较多,有研究显示该菌同时具有产乳酸和脂肪酸^[27-28]以及通过反硝化作用/氨化硝酸盐产含氮气体^[29]的功能,是重要的驱油菌种;*C. proteolyticus* 被发现具有发酵乙酸的功能^[30];Ruan 等证实 *Marinobacter* spp. 具有反硝化作用^[31],Kem 等证明其分解生物素时会释放脂肪酸^[32]。发现的产气菌群中,Ristuccia 等观察到了 *E. cloacae* 将糖发酵为酸以及还原(亚)硝酸盐产气的现象^[33],尽管未观察到硫化氢气体的产生,*E. cloacae* 被报道能利用培养基中的硫酸铵为营养源正常生长^[34],说明其是利用硫酸盐产硫化氢气体菌群的重要组成菌种;Bai 等发现 *S. zoogloeoides* 具有异养硝化及好氧反硝化作用^[35];Si 等发现

Paracoccus 中的 *P. denitrificans* 也具有反硝化作用^[36]。这些前人的大量工作为本研究结果的可信性提供了有力的佐证。

本研究还发现, 产酸菌群与产气菌群作为有机营养激活条件下的优势菌群, 在共同压制原位油藏微生物并且达到一定丰度后, 形成对硝酸盐、丙酮酸、乙酰辅酶 A 等营养物质的竞争, 并且产气功能菌群在无氧环境下处于弱势, 提示产酸与产气功能在菌群达到一定规模后可能会互相抑制, 通过对营养底物的调整可能有助于减轻它们的相互抑制。这些结果为微生物采油技术优化提供了有价值的新思路。

REFERENCES

- [1] Zhang XB, Hong S, Jiang H, Wang WQ. Research progress in heavy oil bio-degradation technology[J]. Contemporary Chemical Industry, 2016, 45(3): 617-621 (in Chinese)
张晓博, 洪帅, 姜晗, 王卫强. 微生物对稠油降解、降粘作用研究进展[J]. 当代化工, 2016, 45(3): 617-621
- [2] Liu XY, Zhao DX, Li YQ, Guo J, Qi CW. Progress of research on viscous crude production techniques[J]. Advances in Fine Petrochemicals, 2018, 19(1): 10-13 (in Chinese)
刘晓瑜, 赵德喜, 李元庆, 郭健, 齐从温. 稠油开采技术及研究进展[J]. 精细石油化工进展, 2018, 19(1): 10-13
- [3] Cao YB, Liu T, Li CF, Hu J, Ba Y, Wu XY. Efficiency and mechanism of oil displacement by a strain of *Geobacillus stearothermophilus*[J]. Chinese Journal of Applied & Environmental Biology, 2015, 21(6): 1060-1064 (in Chinese)
曹嫣嫔, 刘涛, 李彩凤, 胡婧, 巴燕, 吴昕宇. 一株嗜热脂肪地芽孢杆菌的驱油性能及机理[J]. 应用与环境生物学报, 2015, 21(6): 1060-1064
- [4] Du J, Luo T, Wang YX, Yang Y, Lu TJ. Study on the microbial remediation effect of oil-based drilling cuttings with different saw-wood powder[J]. Petrochemical Industry Technology, 2018, 25(1): 157-158 (in Chinese)
杜娇, 罗廷, 王雅璇, 杨越, 卢天娇. 不同锯木粉加量下油基钻井岩屑的微生物效果研究[J]. 石化技术, 2018, 25(1): 157-158
- [5] Nielsen AAK, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, Ross D, Densmore D, Voigt CA. Genetic circuit design automation[J]. Science, 2016, 352(6281): aac7341
- [6] Appleton E, Madsen C, Roehner N, Densmore D. Design automation in synthetic biology[J]. Cold Spring Harbor Perspectives in Biology, 2017, 9(4): a023978
- [7] Wang WQ, Ding YY, Wang GF, Li J. Analysis of oil viscosity reducing performance based on compound bacteria[J]. Special Oil & Gas Reservoirs, 2018, 25(2): 120-124 (in Chinese)
王卫强, 丁影影, 王国付, 李佳. 基于复配菌的原油降黏效果分析[J]. 特种油气藏, 2018, 25(2): 120-124
- [8] Wang WQ, Ding YY, Sun YH, Jiang WS, Wang GF. Analysis of the effect of compound bacteria on the microstructure of waxy crude oil[J]. Journal of Liaoning University of Petroleum & Chemical Technology, 2018, 38(5): 24-29 (in Chinese)
王卫强, 丁影影, 孙翊赫, 姜卫山, 王国付. 复配菌对含蜡原油微观结构影响分析[J]. 辽宁石油化工大学学报, 2018, 38(5): 24-29
- [9] Wang XT, Xiang LB, Zhang YX. Microorganism preparation and application evaluation on microbial enhanced high-pour point oil recovery in Liaohe Oilfield[J]. Lithologic Reservoirs, 2017, 29(5): 162-168 (in Chinese)
王小通, 向龙斌, 张艺馨. 辽河高凝油微生物采油菌剂研究及应用评价[J]. 岩性油气藏, 2017, 29(5): 162-168
- [10] Wang TY, Xiu JL, Huang LX, Bi YQ, Qi YB, Cong ZM, Yu L. Microbial enhanced oil recovery model in porous media[J]. Acta Petrolei Sinica, 2016, 37(1): 97-105 (in Chinese)
王天源, 修建龙, 黄立信, 毕永强, 齐义彬, 从拯民, 俞理. 多孔介质微生物提高原油采收率模型[J]. 石油学报, 2016, 37(1): 97-105
- [11] You J. The composition and distribution of microbial community and the detection of meor microbial groups in Huabei oil field[D]. Shanghai: Doctoral Dissertation of East China University of Science and Technology, 2018 (in Chinese)
游靖. 华北油田油藏微生物组成分布及采油功能微生物研究[D]. 上海: 华东理工大学博士学位论文, 2018
- [12] Del Sal G, Manfioletti G, Schneider C. The CTAB-DNA precipitation method: a common mini-scale preparation of template DNA from phagemids, phages or plasmids suitable for sequencing[J]. Biotechniques, 1989, 7(5): 514-520
- [13] Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets[J]. Bioinformatics, 2011, 27(6): 863-864
- [14] Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes[J]. Nature Methods, 2012, 9(8): 811-814
- [15] Langdon WB. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks[J]. Bio Data Mining, 2015, 8(1): 1
- [16] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND[J]. Nature Methods, 2015,

- 12(1): 59-60
- [17] Karp PD, Riley M, Paley SM, Pellegrini-Toole A. The MetaCyc database[J]. *Nucleic Acids Research*, 2002, 30(1): 59-61
- [18] Franzosa EA, McIver LJ, Rahnvard G, Thompson LR, Schirmer M, Weingart G, Lipson KS, Knight R, Caporaso JG, Segata N, et al. Species-level functional profiling of metagenomes and metatranscriptomes[J]. *Nature Methods*, 2018, 15(11): 962-968
- [19] Li DH, Liu CM, Luo RB, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph[J]. *Bioinformatics*, 2015, 31(10): 1674-1676
- [20] Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies[J]. *PeerJ*, 2019, 7: e7359
- [21] Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes[J]. *Genome Research*, 2015, 25(7): 1043-1055
- [22] Qiao YY, Jia B, Hu ZQ, Sun C, Xiang YJ, Wei CC. MetaBinG2: a fast and accurate metagenomic sequence classification system for samples with many unknown organisms[J]. *Biology Direct*, 2018, 13(1): 15
- [23] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification[J]. *BMC Bioinformatics*, 2010, 11: 119
- [24] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks[J]. *Genome Research*, 2003, 13(11): 2498-2504
- [25] Gray ND, Sherry A, Hubert C, Dolfing J, Head IM. Methanogenic degradation of petroleum hydrocarbons in subsurface environments: remediation, heavy oil formation, and energy recovery[J]. *Advances in Applied Microbiology*, 2010, 72: 137-161
- [26] Brown LR. Microbial enhanced oil recovery (MEOR)[J]. *Current Opinion in Microbiology*, 2010, 13(3): 316-320
- [27] Li C, Gai ZC, Wang K, Jin LP. Engineering *Bacillus licheniformis* as a thermophilic platform for the production of L-lactic acid from lignocellulose-derived sugars[J]. *Biotechnology for Biofuels*, 2017, 10: 235
- [28] Lopes C, Barbosa J, Maciel E, Da Costa E, Alves E, Ricardo F, Domingues P, Mendo S, Domingues RM. Decoding the fatty acid profile of *Bacillus licheniformis* I89 and its adaptation to different growth conditions to investigate possible biotechnological applications[J]. *Lipids*, 2019, 54(4): 245-253
- [29] Guo X, Liu BF, Gao LN, Zhou Y, Shan YY, Lü X. Isolation of nitrite-degrading strains from *Douchi* and their application to degrade high nitrite in *Jiangshui*[J]. *Journal of the Science of Food and Agriculture*, 2019, 99(1): 219-225
- [30] Tandishabo K, Nakamura K, Umetsu K, Takamizawa K. Distribution and role of *Coprothermobacter* spp. in anaerobic digesters[J]. *Journal of Bioscience and Bioengineering*, 2012, 114(5): 518-520
- [31] Ruan YJ, Kumar Awasthi M, Cai L, Lu HF, Xu XY, Li WB. Simultaneous aerobic denitrification and antibiotics degradation by strain *Marinobacter hydrocarbonoclasticus* RAD-2[J]. *Bioresource Technology*, 2020, 313: 123609
- [32] Kem MP, Naka H, Iinishi A, Haygood MG, Butler A. Fatty acid hydrolysis of acyl marinobactin siderophores by *Marinobacter* acylases[J]. *Biochemistry*, 2015, 54(3): 744-752
- [33] Ristuccia PA, Cunha BA. *Enterobacter*[J]. *Infection Control*, 1985, 6(3): 124-128
- [34] Moats WA, Kinner JA. Observations on brilliant green agar with H₂S indicator[J]. *Applied and Environmental Microbiology*, 1976, 31(3): 380-384
- [35] Bai YH, Sun QH, Zhao C, Wen DH, Tang XY. Aerobic degradation of pyridine by a new bacterial strain, *Shinella zoogloeoides* BC026[J]. *Journal of Industrial Microbiology and Biotechnology*, 2009, 36(11): 1391-1400
- [36] Si YY, Xu KH, Yu XY, Wang MF, Chen XH. Complete genome sequence of *Paracoccus denitrificans* ATCC 19367 and its denitrification characteristics[J]. *Canadian Journal of Microbiology*, 2019, 65(7): 486-495