

## 研究报告



## 16S rRNA 基因高变区 V4 和 V3–V4 及测序深度对油藏细菌菌群分析的影响

刘明艳 马嘉晗 李瑜 刘文霞 秦鸿娟 高配科\*

曲阜师范大学生命科学学院 山东 曲阜 273165

**摘要:**【背景】16S rRNA 基因测序是当前研究微生物群落组成及其分布的重要手段。【目的】揭示 16S rRNA 基因高变区 V4 (515–806)和 V3–V4 (338–806)及测序深度(1–2 万条和 10 万条)对油藏微生物细菌群落组成和多样性分析的影响。【方法】所用油水样细菌 16S rRNA 基因拷贝数为  $(6.51 \pm 0.56) \times 10^8/\text{L}$ , 16S rRNA 基因 V4 区测序使用 Illumina MiSeq PE250 测序平台, V3–V4 区测序使用 MiSeq PE300 测序平台。【结果】测序深度达到 1–2 万条时, V4 和 V3–V4 区测序文库覆盖率均达到 99.6%以上, 且具有较好的可重复性; V4 区测序深度为 1–2 万和 10 万时, 菌群  $\alpha$  多样性指数受测序深度影响不显著; 与 V4 区测序相比, 同样测序深度(1–2 万)下, V3–V4 区测序获得的菌群  $\alpha$  多样性指数有所降低。V4 测序 1–2 万与 10 万获得的菌群中几乎未出现显著性差异微生物类群; 同样测序深度(1–2 万)下, V4 与 V3–V4 测序相比, 优势微生物类群 *Epsilonproteobacteria* (51.37%:64.23%)和 *Deltaproteobacteria* (17.96%:11.40%)相对丰度表现出显著差异。【结论】测序深度达到一定水平, 增加测序深度会一定程度上影响菌群  $\alpha$  多样性指数, 对菌群  $\beta$  多样性分析的影响十分有限; 同一测序深度下, V4 区与 V3–V4 区测序获得的细菌菌群  $\alpha$  多样性指数明显不同, 部分优势微生物类群的相对丰度值之间具有显著性差异。鉴于测序读长的提升和测序成本降低, 与 V4 区测序相比, V3–V4 区测序在更低的测序深度下文库覆盖率更高, 可提供更多用于反映物种亲缘关系的 16S rRNA 碱基信息, 本文认为 V3–V4 区测序可作为当下菌群分析的首选区域。

关键词: 油藏, 细菌, 16S rRNA, V4, V3–V4, MiSeq 测序

## Influence of 16S rRNA gene V4 and V3–V4 sequencing and sequencing depth on unraveling bacterial communities inhabiting oil reservoirs

LIU Ming-Yan MA Jia-Han LI Yu LIU Wen-Xia QIN Hong-Juan GAO Pei-Ke\*

College of Life Sciences, Qufu Normal University, Qufu, Shandong 273165, China

**Abstract:** [Background] 16S rRNA gene sequencing is an important method to investigate microbial communities inhabiting diverse environments. [Objective] This study aimed to reveal the influences of

**Foundation items:** National Natural Science Foundation of China (31500414); National Undergraduate Innovation and Entrepreneurship Training Program of China (201710446078)

\*Corresponding author: Tel: 86-537-4456415; E-mail: gpkyl-001@163.com

Received: 28-03-2019; Accepted: 15-08-2019; Published online: 30-09-2019

基金项目: 国家自然科学基金(31500414); 国家级大学生创新创业训练计划项目(201710446078)

\*通信作者: Tel: 0537-4456415; E-mail: gpkyl-001@163.com

收稿日期: 2019-03-28; 接受日期: 2019-08-15; 网络首发日期: 2019-09-30

16S rRNA gene V4 and V3-V4 sequencing with sequencing depth of 10 000–20 000 and 100 000 on analyzing bacterial communities inhabiting oil reservoirs. **[Methods]** The number of the bacterial 16S rRNA gene copies of the sample used here was  $(6.51 \pm 0.56) \times 10^8/\text{L}$ . The bacterial 16S rRNA gene amplicons were sequenced on MiSeq PE250 and PE300 sequencing platforms, respectively. **[Results]** The Good's coverage of each sequencing library reached up to 99.6%, when the sequencing depth reached 10 000. The data obtained by V4 and V3-V4 sequencing were repeatable. The  $\alpha$  diversity indices of the communities obtained by V4 sequencing with depth of 10 000–20 000 and 100 000 did not show significantly differences. Furthermore, there were few microbial populations that showed significant differences. Compared with the communities obtained by V4 sequencing, lower  $\alpha$  diversity indices were observed in the communities obtained by V3-V4 sequencing with a same sequencing depth, and the relative abundances of *Epsilonproteobacteria* (51.37%:64.23%) and *Deltaproteobacteria* (17.96%:11.40%) had significant differences. **[Conclusion]** The results indicated that increasing sequencing depth had a more distinct impact on community  $\alpha$  diversity analysis than community  $\beta$  diversity analysis for a given sample, differential  $\alpha$  diversity indices and some dominant microbial populations with significantly differential relative abundances were observed between the communities obtained by V4 and V3-V4 sequencing. In view of the increase of sequencing length and lower cost of MiSeq sequencing, higher sequencing coverage and more 16S rRNA bases, V3-V4 sequencing is suggested as the preferred method at the current stage.

**Keywords:** Oil reservoirs, Bacteria, 16S rRNA, V4, V3-V4, MiSeq sequencing

微生物在碳循环、氮循环、磷循环和硫循环等生物地球化学过程、营养盐再生、环境污染物降解、能源勘探开发和能量传递等过程中发挥着重要作用<sup>[1-2]</sup>。特定环境微生物的群落组成、多样性及其生态功能是微生物生态学研究的核心问题。微生物群落结构决定了其生态功能,研究微生物群落有助于人类更好地驾驭这些微生物净化环境、勘探开发矿藏、生产合成医药等。

微生物群落分析已从最初的分离培养发展到分子水平<sup>[3-4]</sup>,分子检测技术主要包括荧光原位杂交<sup>[5-6]</sup>、克隆文库<sup>[7]</sup>、基因芯片<sup>[8-9]</sup>、基于高通量测序的 16S rRNA 基因序列分析<sup>[10-12]</sup>、宏基因组<sup>[13]</sup>和宏转录组<sup>[14]</sup>等,其中 16S rRNA 基因测序在研究微生物群落组成及其分布过程中得到广泛应用。

油藏孕育着丰富的微生物类群,这些微生物及其代谢产物能够一定程度上改善油藏环境和原油物性,提高原油采收。油藏微生物群落分析是石油微生物学的一个重要研究方向,对微生物采油技术的应用具有重要指导意义。454 焦磷酸测序开创了 16S rRNA 基因高通量测序的先河,2014 年前后

454 焦磷酸测序在油藏微生物群落解析方面发挥了重要作用<sup>[12,15]</sup>,但要达到较高测序深度,不可避免地要增加测序成本。MiSeq 测序技术的出现使得同样测序成本情况下测序深度大大提高。现阶段,基于 MiSeq 测序平台的 16S rRNA 基因 V4 区测序被广泛用于各环境(含油藏)微生物群落研究<sup>[10,16-17]</sup>。基于 V3-V4 区测序可提供更丰富的序列信息,同样测序深度下测序成本略有提高,该测序区域在油藏环境的应用鲜有报道。鉴于此,本文基于 MiSeq 测序平台探究测序区域(V4 区和 V3-V4 区)及测序深度对油藏细菌菌群分析的影响。

油藏是一个高压缺氧的极端环境,微生物数量远远低于地表土壤和湖泊等环境,基于待分析油样微生物 16S rRNA 基因拷贝数,设置 V4 区和 V3-V4 区 MiSeq 测序所需深度。分析内容包括:(1) 不同测序区域和深度对油藏细菌菌群多样性指数分析的影响;(2) 不同测序区域和深度对菌群组成分析的影响;(3) 不同测序区域和深度下相对丰度之间具有显著性差异的微生物类群分析。本文研究结果可为基于 16S rRNA 基因高通量测序研究油藏微生物群落提供一定参考。

## 1 材料与方法

### 1.1 主要试剂和仪器

AxyPrep™ Genomic DNA Miniprep Kit, 康宁生命科学(吴江)有限公司; FastStart Universal SYBR Green Master (ROX)试剂盒, 上海罗氏制药有限公司; 大肠杆菌 DH5 $\alpha$  和 pEASY-T 载体, 北京全式金生物技术有限公司。NanoDrop 分光光度计, 赛默飞世尔科技公司; PCR 仪, Bio-Rad 公司; MiSeq 测序仪, Illumina 公司。

### 1.2 微生物菌体收集与总 DNA 提取

实验所用油水样取自新疆克拉玛依油田六中区克下组采油井 T6890 产出液。该油藏区块已经历近 30 年水驱, 截至取样时采油井综合含水 92.5%。油藏平均深度 480 m, 地层温度 22.6 °C, 地层压力 7.2 MPa。菌体收集过程为: 将 2 L 油水样置于 60 °C 下静置 30 min, 利用无菌分液漏斗进行油水相分离; 水相经 0.22  $\mu$ m 孔径滤膜过滤收集菌体; 油相加入 3 倍体积异辛烷, 混合均匀后 12 000 r/min 离心 20 min 收集菌体, 同时将离心后的油相经 0.22  $\mu$ m 孔径滤膜过滤收集菌体; 将油水相所得菌体混合用于基因组提取。基因组提取采用玻璃珠研磨和 Axygen 基因组提取试剂盒相结合的方法<sup>[15]</sup>: 将 1 mL TE 缓冲液(Tris 80 mmol/L, EDTA 40 mmol/L, pH 8.0)加入上述菌体, 加入 0.2 g 0.1 mm 玻璃珠 2 500 r/min 振荡 1 min, 此过程重复 3 次, 期间置于冰上冷却 1 min。后续基因组提取过程用 Axygen 基因组提取试剂盒完成。

### 1.3 16S rRNA 基因数量定量分析

以 16S rRNA 基因通用引物 8F (5'-AGAGTTT GAT(CT)(AC)TGGCTC-3')和 338R (5'-GCTGCCT CCGTAGGAGT-3')为定量 PCR 引物, 检测样品基因组中 16S rRNA 基因数量<sup>[18]</sup>。回收纯化 16S rRNA 基因扩增子, 将目的基因片段连接入 pEASY-T 载体, 进而克隆转化到大肠杆菌 DH5 $\alpha$  中(具体操作按照试剂盒说明书进行)。选取阳性克隆于液体肉汤培养基中 37 °C、160 r/min 培养过夜, 收集菌体提取质粒。用 NanoDrop 分光光度计检测

所提质粒 DNA 浓度, 计算质粒数量的公式: 样品浓度(ng/ $\mu$ L)= $OD_{260} \times 50$  ng/ $\mu$ L $\times$ 稀释倍数; 样品摩尔质量=碱基对数 $\times$ 660 g/mol; 样品基因数量(copies/ $\mu$ L)= $6.02 \times 10^{23} \times$ 待测样品浓度(ng/ $\mu$ L) $\times$ 10<sup>-9</sup>/样品摩尔质量; 核酸最高吸收波长为 260 nm; 1 OD 相当于 50  $\mu$ g/mL 的 dsDNA。将质粒进行 10 倍比梯度稀释至 10<sup>2</sup> copies/ $\mu$ L, 用于标准曲线绘制。荧光定量采用 SYBR Green I 染料法, PCR 反应体系: Fast Start Universal SYBR Green Master 10  $\mu$ L, 引物 8F 和 338R 各 6 pmol/L, 模板 DNA 1  $\mu$ L, ddH<sub>2</sub>O 补足至 20  $\mu$ L。PCR 反应条件: 95 °C 10 min; 95 °C 20 s, 60 °C 1 min 采集荧光信号, 40 个循环; 以每循环 0.5 °C 由 60 °C 上升至 95 °C, 每循环保留 10 s, 获得溶解曲线。根据标准曲线计算待测样品中 16S rRNA 基因数。

### 1.4 16S rRNA 基因 MiSeq 测序及序列处理

16S rRNA 基因 V4 区测序使用 Illumina MiSeq PE250 测序平台(单端最大读长 250 bp), 16S rRNA 基因 V3-V4 区测序使用 MiSeq PE300 测序平台(单端最大读长 300 bp)。V4 区 PCR 扩增所用引物为 515f (5'-GTGCCAGCMGCCGCGGTAA-3')和 806r (5'-GGACTACHVGGGTWTCTAAT-3'), 扩增子长度约 290 bp<sup>[10]</sup>; V3-V4 区 PCR 扩增所用引物为 338f (5'-ACTCCTACGGGAGGCAGCAG-3')和 806r (5'-GGACTACHVGGGTWTCTAAT-3'), 扩增子长度约 460 bp。PCR 反应体系(20  $\mu$ L): 5 $\times$ TransStart® FastPfu Buffer 4  $\mu$ L, dNTPs (2.5 mmol/L) 2  $\mu$ L, TransStart® FastPfu DNA Polymerase (2.5 U/ $\mu$ L) 0.5  $\mu$ L, 正、反向引物(10  $\mu$ mol/L)各 0.5  $\mu$ L, 样品 DNA 2  $\mu$ L, ddH<sub>2</sub>O 10.5  $\mu$ L。PCR 反应条件: 94 °C 4 min; 94 °C 30 s, 50 °C 30 s, 72 °C 30 s, 30 个循环; 72 °C 5 min。测序由上海美吉生物医药科技有限公司完成。

测序数据处理通过 Quantitative insights into microbial ecology (QIIME)软件完成。PE reads 首先根据 Overlap 关系进行拼接, 对低质量的数据进行过滤; 根据 Barcode 序列将序列按照样品来源进行分拣, 对每个样品的序列数目进行统计; 在 97%

相似度水平上, 通过 Usearch 方法对序列进行操作分类单元(operational taxonomic units, OTU)划分, 在聚类过程中去除嵌合体, 得到 OTU 的代表序列, 并生产 OTU 表格; 采用 RDP classifier 贝叶斯算法对 97%相似水平的 OTU 代表序列进行分类学分析; 基于 OTU 计算物种多样性指数, 如菌群丰度指数 Ace、菌群多样性指数 Shannon 和 Simpson、测序深度指数 Good's coverage。

1.5 统计分析

采用相关性分析、方差分析和 Duncan 检验、基于 Bray-Curtis 算法的非度量多维尺度分析(non-metric multidimensional scaling, NMDS)和 ANOSIM 分析对各测序方法得到的微生物群落组成异同进行比较分析。差异性微生物类群分析基于 Welch's *t* test, 所用软件为 Statistical analysis of metagenomic profiles (STAMP 2.1.3)<sup>[19]</sup>。

2 结果与分析

2.1 采油井产出液微生物群落 16S rRNA 基因数

荧光定量 PCR 结果表明, 阴性对照组扩增无信号, 扩增效率达到 90%, 标准曲线 *R* 达到 0.99, 经计算本文所用油水样中微生物 16S rRNA 基因数量为(6.51±0.56)×10<sup>8</sup>/L。

2.2 16S rRNA 基因不同测序区域和深度测序概况及多样性指数

所用样品 V4 区测序 1-2 万条、V4 区测序 10 万条和 V3-V4 测序 1-2 万条, 其文库覆盖率均达到 99.6%以上; 同一测序区域(V4), 随测序深度增加, OTU 数量明显增加(表 1); 同 V4 区测序相比, V3-V4 测序获得的 OTU 数量明显较少(表 1)。多样性指数分析表明, V4 区测序 1-2 万和 10 万时, 菌群 Shannon 和 Simpson 指数受测序深度影响不显著; 与 V4 区测序相比, V3-V4 区测序 Shannon 指数偏低, Simpson 指数偏高。以上数据表明, 对于同一测序区域, 即使文库覆盖率已达到 99.6% (如 V4 区测序, 表 1), 增加测序深度仍会有相当数量 OTU 被检测到; 与 V4 区测序相比, 在相近测序深度下(V3-V4 区测序条带数略少), V3-V4 区测序检测到的 OTU 数量明显减少, 菌群 Shannon 多样性指数明显降低, 但 Simpson 多样性指数和文库覆盖率更高。

2.3 不同测序区域和测序深度获得菌群的物种组成一致性

在纲水平上, V4 区测序 1-2 万、V4 区测序 10 万和 V3-V4 区测序 1-2 万 3 种测序方式获得的菌群中共有微生物类群所占百分比分别为 98.15%、

表 1 16S rRNA 基因测序概况

Table 1 Overview of the 16S rRNA gene sequencing

测序方法 Sequencing method	条数 Reads	Coverage	OTUs	Shannon	Simpson
V4 区测序 10 万条	130 750 (4 858)	0.999 6	440 (18)*	3.085 (0.262)a	0.161 (0.025)b
V4 with sequencing depth of 100 000					
V4 区测序 1-2 万条	22 492 (5 494)	0.996 2	340 (23)**	3.153 (0.130)a	0.150 (0.017)b
V4 with sequencing depth of 10 000-20 000					
V3-V4 区 1-2 万条	17 276 (6 635)	0.999 6	145 (4)	2.533 (0.050)b	0.236 (0.012)a
V3-V4 with sequencing depth of 10 000-20 000					

注: 以上数据均为平均值, 括号中为标准偏差; V4 区测序 10 万条 2 个平行, V4 区和 V3-V4 区测序 1-2 万条各 3 个平行; \*表示含 16 个古菌 OTU (相对丰度为 0.45%), \*\*表示含 15 个古菌 OTU (相对丰度为 1.2%); Duncan 多重比较结果的字母标记参考文献[20], 小写字母表示  $\alpha<0.05$  水平。

Note: The above data are averages of the parallel samples, with standard deviation in brackets; V4 with sequencing depth of 100 000 included two parallel samples, V4 and V3-V4 with sequencing depth of 10 000-20 000 included three parallel samples, respectively; \* indicated that 16 archaeal OTUs were detected, \*\* indicated that 15 archaeal OTUs were detected; Letter markers for multiple comparison result refer to references [20], the lower-case letters represent  $\alpha<0.05$ .

98.99%和 99.5% (图 1A)。在属水平上, 3 种测序方式获得的菌群中共有微生物类群在 V4 区测序 1–2 万中占 94.06%, 在 V4 区测序 10 万中占 94.61%, 在 V3–V4 区测序 1–2 万中占 98.13% (图 1B)。在属水平上对上述 3 种测序方式获得的菌群进行 Pearson 线性相关性分析表明, 3 种测序方式获得的菌群中共有微生物类群具有极为相近的相对丰度(图 2)。以上数据表明, 3 种测序方式获得的菌

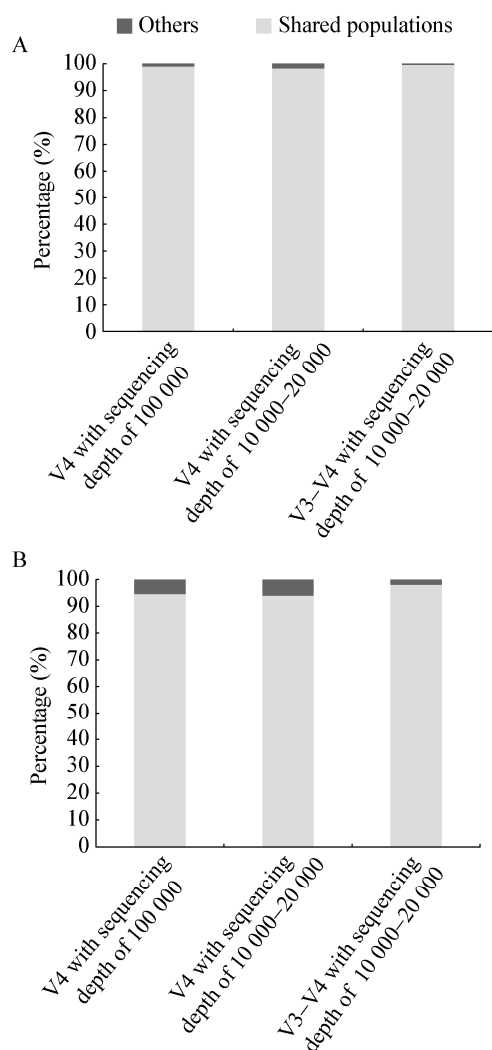


图 1 V4 和 V3–V4 区不同测序深度获得菌群中共有微生物类群和特有微生物类群在各菌群纲水平(A)和属水平(B)中所占百分比

Figure 1 The abundances of different and shared microbial populations at class (A) and genus (B) level in the microbial communities obtained by 16S rRNA gene V4 and V3–V4 sequencing

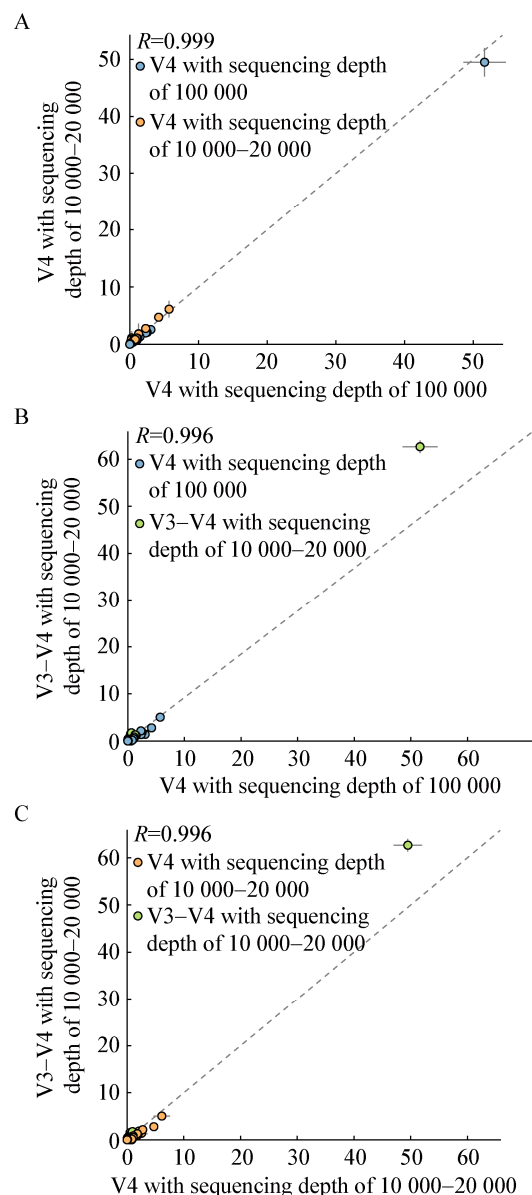


图 2 V4 和 V3–V4 区不同测序深度获得菌群的 Pearson 线性相关性分析

Figure 2 The Pearson analysis for the microbial populations obtained by 16S rRNA gene V4 and V3–V4 sequencing

群中优势微生物类群(纲和属水平)及其相对丰度具有较高的一致性。

菌群  $\beta$  多样性分析表明, 同一测序体系重复测序获得的菌群仍具有一定差异, 但整体上优于不同测序体系获得的菌群, 在 NMDS 图中表现为具有更近的距离(图 3)。相同测序区域(V4)不同测

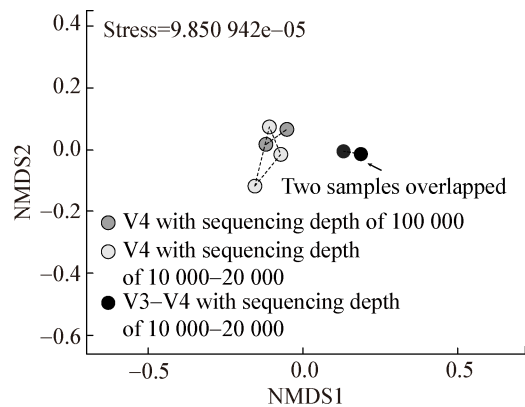


图 3 V4 和 V3-V4 区不同测序深度获得菌群的 NMDS 分析

Figure 3 The NMDS analysis for the microbial communities obtained by 16S rRNA V4 and V3-V4 sequencing

序深度(1-2 万和 10 万条)在 NMDS 图中并未明显分开,不同测序区域(V4 和 V3-V4 区)相同测序深度(1-2 万条)获得的菌群在 NMDS 图中明显分开。即便如此,V4 和 V3-V4 区测序获得的菌群间并未表现出显著性差异(ANOSIM 分析,  $P=0.1$ )。

2.4 不同测序区域和深度获得的菌群中相对丰度存在显著差异的微生物类群

如表 2 所示,两两显著性差异分析表明,在纲水平上,V4 区测序 10 万与 V4 区测序 1-2 万两种测序深度下几乎不存在相对丰度具有显著性差异的微生物类群(图 4A);V4 区测序 10 万与 V3-V4 区测序 1-2 万两种测序下优势微生物类群 *Deltaproteobacteria*、*Gammaproteobacteria* 和

*Betaproteobacteria* 相对丰度间存在显著性差异(图 4B,  $P<0.05$ );V4 区测序 1-2 万与 V3-V4 区测序 1-2 万两种测序下优势微生物类群 *Epsilonproteobacteria*、*Deltaproteobacteria*、*Alphaproteobacteria* 和 *Deferribacteres* 相对丰度间存在显著性差异(图 4C,  $P<0.05$ )。在属水平上 V4 区测序 10 万与 V4 区测序 1-2 万两种测序深度下不存在相对丰度具有显著性差异的优势微生物类群(图 5A);V4 区测序 10 万与 V3-V4 区测序 1-2 万两种测序下 *Desulfobotulus*、*Arcobacter*、*Sulfurospirillum*、*Desulfomicrobium*、*Desulfococcus*、*Geobacter*、*Desulfotignum* 和 *Desulfocapsa* 等微生物类群相对丰度间存在显著性差异(图 5B,  $P<0.05$ ),但这些微生物类群在群落中占比在 5%以下;V4 区测序 1-2 万与 V3-V4 区测序 1-2 万两种测序下相对丰度间存在显著差异的微生物类群数量增多(图 5C,  $P<0.05$ ),其中包含优势微生物 *Sulfurovum* (V4 区 49.49%, V3-V4 区 62.67%),其他具有显著性差异的微生物类群在群落中占比在 5%以下。

3 讨论与结论

鉴于 Illumina MiSeq 高通量测序平台读长及测序成本等问题,基于引物 515f 和 806r 扩增的 V4 区测序被广泛使用,是发起于 2010 年的地球微生物组计划(Earth Microbiome Project, <http://www.earthmicrobiome.org/>)建议使用的测序引物及区域。引物

表 2 相对丰度存在显著差异的优势微生物类群在 V4 和 V3-V4 区测序获得菌群中的分布

Table 2 The relative abundances of differential microbial populations in the microbial communities obtained by 16S rRNA V4 and V3-V4 sequencing (%)

纲水平	V4 区测序 10 万	V4 区测序 1-2 万	V3-V4 区测序 1-2 万
Class level	V4 with sequencing depth of 100 000	V4 with sequencing depth of 10 000-20 000	V3-V4 with sequencing depth of 10 000-20 000
<i>Epsilonproteobacteria</i>	53.96	51.37	64.23
<i>Deltaproteobacteria</i>	17.99	17.96	11.40
<i>Gammaproteobacteria</i>	2.75	2.31	1.43
<i>Deferribacteres</i>	1.37	1.90	0.99
<i>Sphingobacteriia</i>	1.16	1.22	0.69
<i>Betaproteobacteria</i>	0.83	0.66	0.46
<i>Alphaproteobacteria</i>	0.12	0.12	0.05

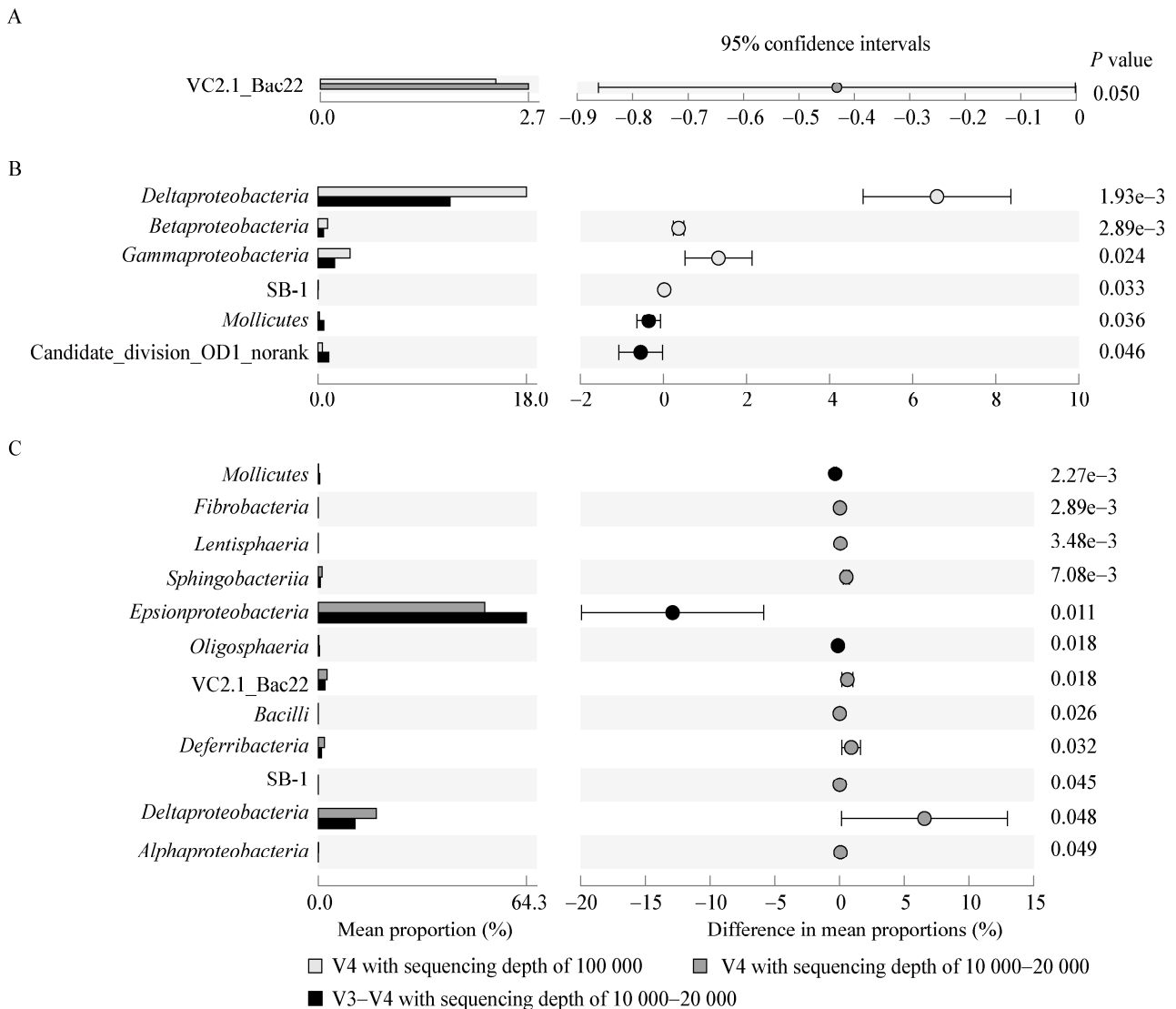


图 4 V4 和 V3–V4 区不同测序深度获得菌群中相对丰度具有显著差异的微生物类群(纲水平)

Figure 4 The microbial populations that showed significant difference between the communities obtained by 16S rRNA gene V4 and V3–V4 sequencing (at the class level)

515f 和 806r 由 Caporaso 等设计,对微生物覆盖度高,同时也可以检测到部分古菌类群<sup>[10]</sup>。然而 Parada 等发现该引物对会对 *Gammaproteobacteria* 造成高估,因此对其进行了优化<sup>[21]</sup>。这一现象在本文中也有所体现,如同一测序深度下,V4 区测序获得的菌群中 *Gammaproteobacteria* 的相对丰度是 2.31%,高于 V3–V4 区测序获得的菌群中 *Gammaproteobacteria* 相对丰度是 1.43%。基于引物 338f 和 806r 对 V3–V4 区进行扩增,可获

得约 460 bp 扩增子。更长的序列信息意味着可用于反映种属亲缘关系的位点数增多,能够更准确地进行物种分类分析。与此同时,对测序平台的读长要求也相应提高,测序成本相应增加,如相对 V4 区测序(约 290 bp),单端测序 150 bp 的读长即可满足分析需求,而 V3–V4 区单端测序读长至少要达到 250 bp。现阶段 Illumina 测序平台读长已经可以达到 400 bp,测序成本也在不断降低。



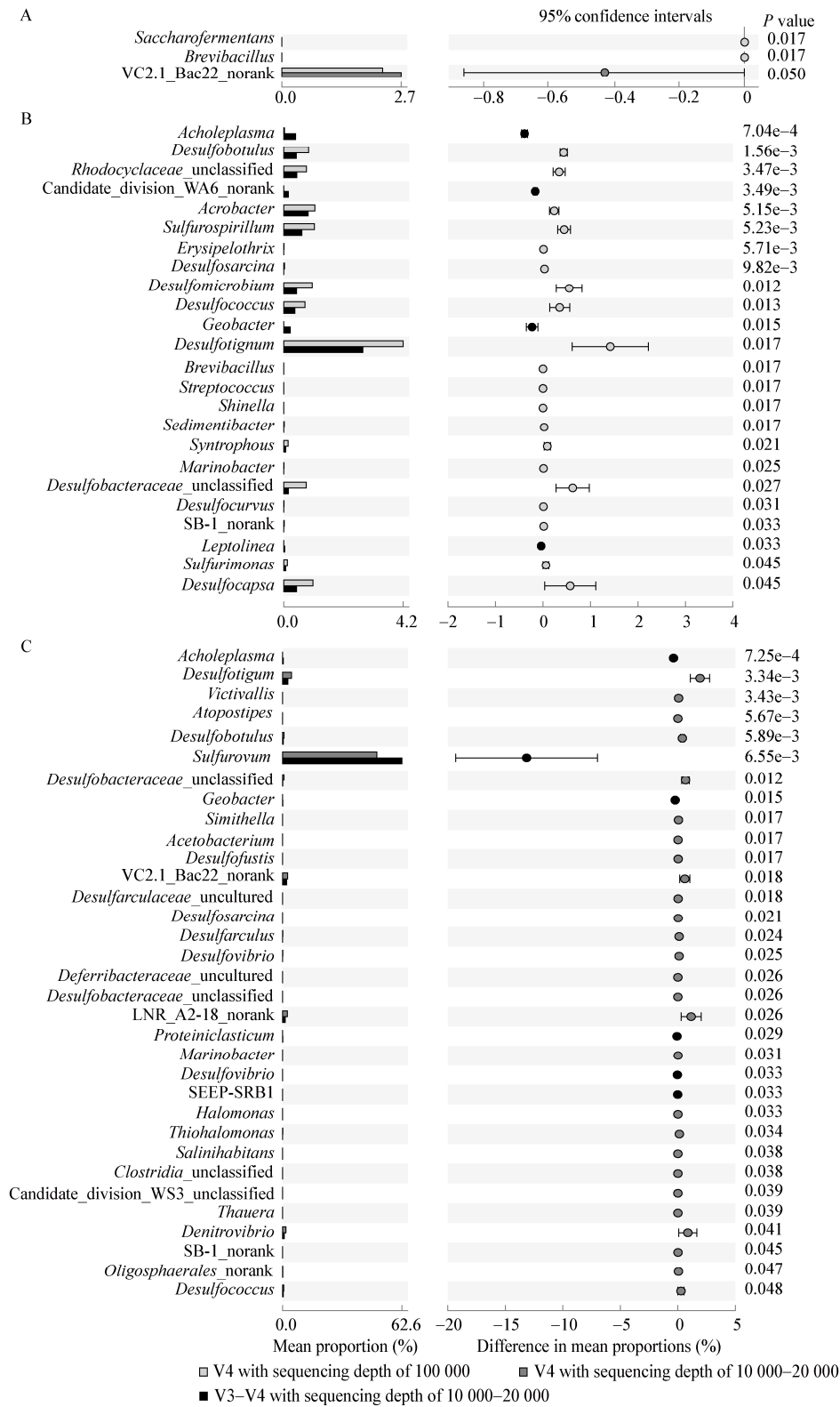


图 5 V4 和 V3-V4 区不同测序深度获得菌群中相对丰度具有显著差异微生物类群(属水平)  
Figure 5 The microbial populations that showed significant differences between the communities obtained by 16S rRNA gene V4 and V3-V4 sequencing (at the genus level)



16S rRNA 基因测序区域和测序深度对菌群  $\alpha$  多样性指数分析具有明显影响。就本文所用样本, V4 区测序深度由 1–2 万条增加至 10 万条, 文库覆盖率仅由 99.62% 增加至 99.96%, 但被检测到的 OTU 数量由 340 个(含 15 个古菌 OTU) 增加到 440 个(含 16 个古菌 OTU)。多出来的 OTU 大多仅出现一次, 这表明增加测序深度可以检测到样本中极低丰度微生物类群。由于测序平台、序列质控、序列拼接及 OTU 划分等过程完全相同, 因此测序深度增加是导致 OTU 数量大幅度增加的根本原因。由于共同检测到的微生物类群在菌群中所占比重高(在属水平上, 本文涉及到的 3 种测序方式获得的菌群中共有微生物类群在 V4 区测序 1–2 万中占 94.06%, 在 V4 区测序 10 万中占 94.61%), 差异微生物类群在菌群中所占比重极低, 测序深度对菌群  $\beta$  多样性并未表现出显著影响。在测序深度略高的情况下, V4 区测序获得了 340 个 OTU (含 15 个古菌 OTU), 明显高于 V3–V4 区测序的 145 个 OTU; V4 区测序 Shannon 多样性指数显著高于 V3–V4 区测序, 但 Simpson 多样性指数和文库覆盖率却明显低于 V3–V4 区测序。上述现象表明, V4 区测序可能会导致对菌群的  $\alpha$  多样性指数高估。值得关注的是, V4 区测序 10 万计算出的  $\alpha$  多样性指数比 V4 区测序 1–2 万更接近于 V3–V4 区测序 1–2 万。

无论是纲水平还是属水平, V4 区测序 10 万与 1–2 万两种测序深度下几乎不存在相对丰度具有显著性差异的微生物类群, 这应该与此两种测序深度下文库覆盖率分别达到 99.96% 和 99.62% 相关。此外, 由于测序平台、测序引物、数据处理方法等完全相同, 可以排除系统误差。因此, 只要达到相应的测序深度, 获得的菌群数据可以有效反映整个菌群组成。V3–V4 区测序与 V4 区测序相比, 同样测序深度下, 相对丰度具有显著性差异的微生物类群数量增多, 其中包含优势微生物类群 *Epsilonproteobacteria* (51.37%:64.23%) 和 *Deltaproteobacteria* (17.96%:11.40%)。

本文数据反映了 16S rRNA 基因不同测序区域和测序深度对同一测序样本菌群分析的影响: 对于特定样品, 测序深度达到一定水平(送样测序前检测样本菌群基因组 16S rRNA 拷贝数, 根据拷贝数确定测序深度), 不同测序深度对菌群组成分析的影响十分有限, 但增加测序深度会一定程度影响菌群  $\alpha$  多样性指数; V4 区与 V3–V4 区测序获得的菌群  $\alpha$  多样性指数明显不同, 且存在相对丰度具有显著差异的优势微生物类群。鉴于测序读长的提升和测序成本降低, 以及在更低的测序深度下 V3–V4 区测序文库覆盖率更高, 而且可提供更多可用于反映物种亲缘关系的 16S rRNA 碱基信息, V3–V4 区测序可作为当下菌群分析的首选区域。此外, 不同领域微生物群落分析在满足测序深度需要的前提下均应采用统一的测序平台、测序引物和数据处理方法, 从而有助于不同研究者对不同采样位点产生的实验数据进行 Meta 分析, 在更大的时空尺度上揭示特定环境(如油藏)微生物群落组成特征及演替规律。

本文实验设计之初考虑到 V4 区测序扩增子长度约 290 bp, V3–V4 区测序扩增子长度约 460 bp, 因此, V4 区测序由 MiSeq PE250 测序平台(单端测序长度可达 250 bp)完成, V3–V4 区测序由 MiSeq PE300 测序平台(单端测序长度可达 300 bp)完成。尽管 MiSeq PE250 和 MiSeq PE300 测序准确度都较高, 但依旧无法排除因测序平台不同可能对菌群分析造成的影响。

## REFERENCES

- [1] Youssef N, Elshahed MS, McInerney MJ. Microbial processes in oil fields: culprits, problems, and opportunities[J]. *Advances in Applied Microbiology*, 2009, 66: 141–251
- [2] Zheng P. *Environmental Microbiology*[M]. 2nd ed. Hangzhou: Zhejiang University Press, 2012 (in Chinese)  
郑平. 环境微生物学[M]. 2 版. 杭州: 浙江大学出版社, 2012
- [3] Li H, Mu BZ. Recent advances in molecular microbial ecology of petroleum rese[J]. *Microbiology China*, 2008, 35(5): 803–808 (in Chinese)  
李辉, 牟伯中. 油藏微生物多样性的分子生态学研究进

- 展[J]. 微生物学通报, 2008, 35(5): 803-808
- [4] Shendure J, Ji H. Next-generation DNA sequencing[J]. Nature Biotechnology, 2008, 26(10): 1135-1145
- [5] Raap AK. Advances in fluorescence *in situ* hybridization[J]. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 1998, 400(1/2): 287-298
- [6] Hu Q, Qi HY, Zhang HX. Fluorescence *in situ* hybridization (FISH) and its applications in microbial ecology[J]. Acta Ecologica Sinica, 2004, 24(5): 1048-1054 (in Chinese)  
呼庆, 齐鸿雁, 张洪勋. 荧光原位杂交技术及其在微生物生态学中的应用[J]. 生态学报, 2004, 24(5): 1048-1054
- [7] Zhao LX, Ma T, Gao PK, et al. Characterization of microbial diversity and community in water flooding oil reservoirs in China[J]. World Journal of Microbiology and Biotechnology, 2012, 28(10): 3039-3052
- [8] He ZL, Deng Y, van Nostrand JD, et al. GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity[J]. The ISME Journal, 2010, 4(9): 1167-1179
- [9] Sagaram US, DeAngelis KM, Trivedi P, et al. Bacterial diversity analysis of Huanglongbing pathogen-infected citrus, using PhyloChip arrays and 16S rRNA gene clone library sequencing[J]. Applied and Environmental Microbiology, 2009, 75(6): 1566-1574
- [10] Caporaso JG, Lauber CL, Walters WA, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample[J]. Proceedings of the National Academy of Sciences of the United States of America, 2011, 108 (Suppl 1): 4516-4522
- [11] Xiao M, Sun SS, Zhang ZZ, et al. Analysis of bacterial diversity in two oil blocks from two low-permeability reservoirs with high salinities[J]. Scientific Reports, 2016, 6: 19600
- [12] Lenchi N, Inceoglu Ö, Kebbouche-Gana S, et al. Diversity of microbial communities in production and injection waters of Algerian oilfields revealed by 16S rRNA gene amplicon 454 pyrosequencing[J]. PLoS One, 2013, 8(6): e66588
- [13] An DS, Caffrey SM, Soh J, et al. Metagenomics of hydrocarbon resource environments indicates aerobic taxa and genes to be unexpectedly common[J]. Environmental Science & Technology, 2013, 47(18): 10708-10717
- [14] Liu YF, Galzerani DD, Mbadinga SM, et al. Metabolic capability and *in situ* activity of microorganisms in an oil reservoir[J]. Microbiome, 2018, 6: 5
- [15] Li GQ, Gao PK, Wu YQ, et al. Microbial abundance and community composition influence production performance in a low-temperature petroleum reservoir[J]. Environmental Science & Technology, 2014, 48(9): 5336-5344
- [16] Song YT, Hu J, Wu XL, et al. Stability of microbial community structure in reservoir water samples at room temperature[J]. Chinese Journal of Applied & Environmental Biology, 2017, 23(3): 495-501 (in Chinese)  
宋永亭, 胡婧, 吴晓玲, 等. 室温条件下油藏采出液微生物群落结构稳定性[J]. 应用与环境生物学报, 2017, 23(3): 495-501
- [17] Hu J, Sun GZ, Liu T, et al. Spatial distribution of indigenous microbial communities under simulated reservoir conditions[J]. Microbiology China, 2019, 46(5): 1008-1017 (in Chinese)  
胡婧, 孙刚正, 刘涛, 等. 模拟油藏条件下内源微生物群落空间分布规律[J]. 微生物学通报, 2019, 46(5): 1008-1017
- [18] Gittel A, Sørensen KB, Skovhus TL, et al. Prokaryotic community structure and sulfate reducer activity in water from high-temperature oil reservoirs with and without nitrate treatment[J]. Applied and Environmental Microbiology, 2009, 75(22): 7086-7096
- [19] Parks DH, Tyson GW, Hugenholtz P, et al. STAMP: statistical analysis of taxonomic and functional profiles[J]. Bioinformatics, 2014, 30(21): 3123-3124
- [20] Qiu NW, Liu Q, Liu H. Multiple comparisons of biological experimental data in statistical analysis[J]. Journal of Biomathematics, 2015, 30(3): 535-541 (in Chinese)  
邱念伟, 刘倩, 刘慧. 生物学实验数据统计分析中的多重比较法[J]. 生物数学学报, 2015, 30(3): 535-541
- [21] Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples[J]. Environmental Microbiology, 2016, 18(5): 1403-1414