Research Article

Analysis of Bacterial Diversity in Different Heavy Oil Wells of a Reservoir in South Oman with Alkaline pH

Biji Shibulal 1,1, Saif N. Al-Bahry 1,1, Yahya M. Al-Wahaibi 2,2, Abdulkadir E. Elshafie 1,1, Ali S. Al-Bemani 2,2 and Sanket J. Joshi 1,3

1Department of Biology, College of Science, Sultan Qaboos University, Muscat, Oman
2Department of Petroleum and Chemical Engineering, College of Engineering, Sultan Qaboos University, Muscat, Oman
3Central Analytical and Applied Research Unit, College of Science, Sultan Qaboos University, Muscat, Oman

Correspondence should be addressed to Saif N. Al-Bahry; snbahry@squ.edu.om

Received 29 October 2017; Accepted 15 February 2018; Published 21 March 2018

Copyright © 2018 Biji Shibulal et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The identification of potential hydrocarbon utilizing bacteria is an essential requirement in microbial enhanced oil recovery (MEOR). Molecular approaches like proteomic and genomic characterization of the isolates are replacing the traditional method of identification with systemic classification. Genotypic profiling of the isolates includes fingerprint or pattern-based technique and sequence-based technique. Understanding community structure and dynamics is essential for studying diversity profiles and is challenging in the case of microbial analysis. The present study aims to understand the bacterial community composition from different heavy oil contaminated soil samples collected from geographically related oil well areas in Oman and to identify spore-forming hydrocarbon utilizing cultivable bacteria. V4 region of 16S rDNA gene was the target for Ion PGM™. A total of 825081 raw sequences were obtained from Ion torrent from all the 10 soil samples. The species richness and evenness were found to be moderate in all the samples with four main phyla, Firmicutes, Bacteroidetes, Proteobacteria, and Actinobacteria, the most abundant being Firmicutes. *Bacillus* sp. ubiquitously dominated in all samples followed by *Paenibacillus*, which was followed by *Brevibacillus*, *Planococcus*, and *Flavobacterium*. Principal Coordinate Analysis (PCoA) and UPGMA dendrogram clustered the 10 soil samples into four main groups. Weighted UniFrac significance test determined that there was significant difference in the communities present in soil samples examined. It can be concluded that the microbial community was different in all the 10 soil samples with *Bacillus* and *Paenibacillus* sp. as predominating genus. The 16S rDNA sequencing of cultivable spore-forming bacteria identified the hydrocarbon utilizing bacteria as *Bacillus* and *Paenibacillus* sp. and the nucleotide sequences were submitted to NCBI GenBank under accession numbers KP119097–KP119115. *Bacillus* and *Paenibacillus* sp., which were relatively abundant in the oil fields, can be recommended to be chosen as candidates for hydrocarbon utilization study.

1. Introduction

Oil production has been experiencing decline in many parts of the world due to oilfield’s maturity and an example of such includes the major oilfields in the North Sea [1]. Another major concern is the increasing energy demands due to global population growth and the difficulty in discovering new oilfields as an alternative to the exploited oil fields. Therefore, there is an urge to find out alternative technologies to increase oil recovery from existing oilfields around the world. It is a fact that fossil fuels will still remain the key source of energy, regardless of the gross investments in other energy sources such as biofuels, solar energy, and wind energy. This fact is highlighted by the current global energy production from fossil fuels which currently stand at about 80–90% with oil and gas representing about 60% [2]. During oil production, primary oil recovery can account for 30–40% oil productions, while additional 15–25% can be recovered by secondary methods such as water injection leaving behind about 35–55% of oil as residual oil in the reservoirs [3]. This residual oil is usually the target of many enhanced oil recovery technologies and it amounts to about 2–4 trillion barrels [4] or about 67%
2. Materials and Methods

2.1. Sample Collection and Preparation. The heavy crude oil contaminated subsurface soil was collected from different heavy oil wells in Southern Oman as described previously [11]. Total ten soil samples were collected aseptically in sterile sampling bags and stored at 4°C for further studies. The samples were kindly provided by a local oil company.

2.2. Genomic DNA Isolation. The genomic DNA (gDNA) was isolated from cultivable spore-forming bacteria by boiling suspended soil samples in 10 ml distilled water at 90°C for 30 min to kill all the vegetative cells and enrichment of the bacteria in Bushnell-Haas media (BH media) containing 1% heavy crude oil. The flasks inoculated with the boiled sample supernatant were incubated at 40°C for 24 h and plated on fresh BH agar plates to obtain pure cultures. The gDNA from the isolates and the soil samples was isolated using PowerSoil® DNA Isolation Kit (Mo Bio Laboratories Inc.). The nucleic acid concentration and purity were measured by Thermo Fischer NanoDrop™ 2000/2000 c spectrophotometer.

2.3. Identification of Cultivable Spore-Forming Isolates Using 16S rDNA Sequencing. Cultivable spore-forming bacterial isolates were identified by 16S rDNA sequencing using 27F and 1492R universal primers as described previously [11], where briefly the genomic DNA was extracted using PowerSoil DNA Isolation Kit (Mo Bio Laboratories Inc.), and the amplification was performed using T100™ thermal cycler (Bio-Rad, USA). The PCR products were purified using QIAquick PCR purification kit (QIAGen). The BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems™) was used for de novo sequencing. The sequencing was done using 3130 XL Genetic Analyzer (Applied Biosystem, Hitachi) at Central Analytical and Applied Research Unit (CAARU), Sultan Qaboos University, and was submitted to NCBI GenBank USA. The dendrogram was constructed using maximum likelihood (ML) methods, respectively, using PHYLIP, the Phylogeny Inference Package program. The ML program uses a Hidden Markov Model (HMM) method of inferring different rates of evolution at different sites [12].

2.4. NGS Analysis of Microbial Community in Soil Samples. The metagenomics analysis of all the 10 soil samples was done following the instruction manual of the Ion PGM System. The total gDNA extracted from the soil samples was amplified using 16S Primer Sets targeting hypervariable V4 region: 515F (5’-GTGCCAGCMGCGGTTAA-3’) and 806R (5’-GGACTACVSGGTATCTAAT-3’). The PCR procedure was with an initial denaturation at 95°C for 10 min, 25 cycles of denaturation for 30 sec at 95°C, annealing at 58°C, extension at 72°C for 20 sec, followed by holding at 72°C for 7 min and a final hold at 4°C for oo. The amplified product was purified using Agencourt® AMPure® XP Reagent and 70% ethanol. The DNA input for library preparation was calculated using Agilent® 2100 Bioanalyzer® and the library was prepared using Ion Plus Fragment Library Kit, following the user instructions. The pooled short amplicons were end-repaired using 5x End Repair Buffer and End Repair Enzyme and purified using Agencourt AMPure XP Reagent and 70% ethanol. Barcoded libraries were prepared using Ion Xpress™ Barcode Adapters 1–16 Kit. The adapters are ligated using DNA Ligase and the nicks were repaired using Nick Repair Polymerase. The DNA template for Ion PGM System was prepared using the Ion PGM Template OT2 400 Kit and the Ion OneTouch™ 2 System following the instructions in Ion PGM Template OT2 400 Kit User Guide. And the library was sequenced using the Ion Personal Genome Machine® (PGM™) System and the Ion PGM Sequencing 400 Kit.

2.5. Data Analysis. Primary data analysis was performed with Torrent Suite™ Software v4.0 with automated secondary analysis using Ion Reporter™ Software v4.0. Further analysis was done using a variety of computer packages including Past3, XLstat, NCSS 2007, “R” and NCSS 2010. Alpha diversity analysis was conducted using QIIME pipeline (version 1.8.0) (QIIME, 2016). Significance reported for any analysis is defined as p < 0.05.

Short sequences < 200 bp were removed after depleting primers and barcodes; sequences with ambiguous base calls were also removed. Sequences with homopolymer runs
Shannon diversity, related directly to species richness ($S$), found in species $i$. Monte Carlo simulations (weighted UniFrac significance test) were done to test the significant difference in OTUs in the samples in the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) tree generated by UniFrac analysis [25], followed by Bonferroni correction [26] to reduce type I error rate.

3. Results

3.1. Sample Collection and Preparation. The soil moisture was determined to be 0.03–0.08 m$^3$/m$^3$, measured using EM50 Digital data logger, and pH was measured by Jenway 3505 pH meter as 8.5. The viscosity of the heavy crude oil from the well head, which contaminated the soil, was estimated as 4.5°FAPI using RheolabQC Rotation Viscometer. The eTPH of the soil samples ranged within 3.2–4.8%.

3.2. Genomic DNA Isolation and Identification of Cultivable Spore-Forming Isolates Using 16S rDNA Sequencing. The gDNA from all the isolates (Figure 1(a)) and the soil samples (Figure 1(b)) was extracted using PowerSoil DNA Isolation Kit and was successfully amplified using universal primers for bacteria, 27F and 1492R. The amplified product was 1400 bp (Figure 2).

The purified PCR product was sequenced using 3130 XL Genetic Analyzer and the sequences were submitted to NCBI under the accession numbers KP119097–KP119115 (Table I). The sequences were aligned with the closely related species sequences from NCBI nucleotide blast. The branches with bootstrap values above 70% are reliable. The dendrogram was made using ML (maximum likelihood) method, which assumes every single site of the multiple sequence alignment independently with bootstrapping of 1000 replicates. The ln likelihood for the cladogram was $-6866.09722$. The closer the value of likelihood to zero, the better is the dendrogram (Figure 3). Among the listed isolates in Table I, Paenibacillus ehimensis BSI showed very good potential for applications in MEOR and heavy crude oil biodegradation [II].

3.3. NGS Analysis of Microbial Community in Soil Samples. The gDNA was amplified using specific primers targeting the V4 region of the 16S rRNA bacterial gene using
Figure 2: 16S rDNA amplified product electrophoresed on a 2% agarose gel and stained with ethidium bromide. First lane contains the 10000 bp molecular weight markers from 300 bp to 10200 bp in length. Other lanes are PCR products amplified using 27F and 1492R primers.

Table 1: NCBI GenBank accession numbers for the identified isolates.

| Accession No. | Identification                        |
|---------------|--------------------------------------|
| KPI19097      | Bacillus subtilis strain BG1          |
| KPI19098      | Bacillus licheniformis strain BG2     |
| KPI19099      | Bacillus foraminis strain BG3         |
| KPI19100      | Bacillus firmus strain BG4            |
| KPI19101      | Bacillus halodurans strain BG5        |
| KPI19102      | Bacillus subtilis strain BG7          |
| KPI19103      | Bacillus halodurans strain BG8        |
| KPI19104      | Bacillus licheniformis strain BG9     |
| KPI19105      | Bacillus licheniformis strain BG10    |
| KPI19106      | Paenibacillus ehimensis strain BS1    |
| KPI19107      | Bacillus subtilis strain BS2          |
| KPI19108      | Bacillus subtilis strain BS3          |
| KPI19109      | Bacillus subtilis strain BS4          |
| KPI19110      | Bacillus licheniformis strain BS5     |
| KPI19111      | Bacillus licheniformis strain BS6     |
| KPI19112      | Bacillus subtilis strain BS7          |
| KPI19113      | Bacillus licheniformis strain BS8     |
| KPI19114      | Bacillus licheniformis strain BS9     |
| KPI19115      | Bacillus licheniformis strain BS10    |

Table 2: Shannon index and Simpson index.

| Samples | Shannon index ($H'$) | Simpson index ($D$) |
|---------|----------------------|---------------------|
| (1)     | 2.67                 | 0.89                |
| (2)     | 1.81                 | 0.73                |
| (3)     | 2.26                 | 0.81                |
| (4)     | 2.61                 | 0.86                |
| (5)     | 2.69                 | 0.90                |
| (6)     | 2.08                 | 0.79                |
| (7)     | 2.42                 | 0.85                |
| (8)     | 2.03                 | 0.71                |
| (9)     | 2.78                 | 0.90                |
| (10)    | 2.37                 | 0.84                |

The α-diversity measures of the samples were done in QIIME pipeline and Past3. Shannon index ($H'$) and Simpson index ($D$) were calculated for all the samples (Table 2), which revealed that the species richness and relative abundance were present in all the samples, soil sample 2 being with the least diversity.

Rarefaction curves with chao1 and Shannon diversity were plotted with the average number of OTUs at each interval against the size of the subsample [27] (Figure 4). It was found that, for all the 10 soil samples, the curve reached a plateau at approximately 5000 sequences indicating that sequencing depth was sufficient to capture the full scope of microbial diversity.

Taxonomic comparison of the OTUs identified four main phyla, Firmicutes, Bacteroidetes, Proteobacteria, and Actinobacteria, the most abundant being Firmicutes followed by Bacteroidetes (Figure 5). Further comparison of genus belonging to these phyla showed that *Bacillus* was the most abundant genus followed by *Paenibacillus*, which was

515F (5'-GTGCCAGCMCGCGCCGGA-3') and 806R (5'-GACTACVCVGGGTATCTAAT-3') for Ion Torrent PGM analysis. A total of 825081 raw sequences were obtained from Ion torrent from all the 10 soil samples. After quality filtering 758305 reads were obtained. Clustering with UCLUST provided 755423 OTUs. The average reads per sample was 75342. The OTUs were rarefied to 30000 sequences and jackknifed at 25000 sequences.
followed by *Brevibacillus, Planococcus, Flavobacterium*, and so forth (Figure 6).

The result was confirmed by heatmap analysis in which the highest relative abundance of *Bacillus, Paenibacillus,* and *Brevibacillus* was found in all the 10 soil samples and the soil samples with more similar microbial populations were mathematically clustered closer together. The genera (consortium) were used for clustering. Thus, the samples with more similar consortium of genera cluster closer together with the length of connecting lines (top of heatmap) (Figure 7) related to the similarity; shorter lines between two samples indicate closely matched microbial consortium. The heatmap represents the relative percentages of each genus. The predominant genera are represented along the right y-axis.

SHE analysis was performed to evaluate whether species proportion was similar in all of the 10 soil samples to assess the microbial diversity. The community structure was determined as a log-normal distribution, that is, a few species with high or low abundance and many with intermediate abundance [28] (Figure 8).

The resemblance between the bacterial communities, β-diversity, was measured using UniFrac analysis, which provided a tree-based (Figure 9(a)) Principal Coordinate Analysis (PCoA) graph (Figure 9(b)), which grouped the soil samples to 4 main clusters. The eigenvalues for PC1, PC2, and PC3 were 0.014, 0.002, and 0.001, respectively, and accounted for 74%, 12%, and 6% of the total variance. Both the two approaches revealed the same pattern.

**Figure 3:** Dendrogram of the isolates using maximum likelihood method (the green dots represent the isolates). The length of each tree segment in the dendrogram represents in units of expected nucleotide substitutions per sites.
Figure 4: Rarefaction curves showing Shannon diversity index and Chao1 at various sequencing depths. The curve reached a plateau at approximately 5000 sequences.

Figure 5: Relative abundance (%) of bacteria phyla. Firmicutes were found to be the most abundant phyla in all the soil samples.

UniFrac $p$ values were based on comparisons to 1,000 randomized trees. The $p$ values are significant only if they were $<0.05$ (Table 3).

4. Discussion

Ion PGM™ was used to delineate the bacterial community structure of 10 soil samples contaminated with heavy crude oil collected from near oil wells used in the study. The hypervariable V4 region of 16S rDNA sequence was the target region. Stringent quality sequence curation of a total of 825081 raw reads obtained resulted in 84% reduction of initial reads. There are reports stating 50–80% filtering of initial reads [29–31].

Figure 6: Relative abundance (%) of bacterial genera. *Bacillus* was the most abundant genus followed by *Paenibacillus* and *Brevibacillus*.
The relative abundance of each phyla varied among the 10 soil samples, the predominant phyla observed to be Firmicutes, Bacteroidetes, Proteobacteria, and Actinobacteria, the most abundant being Firmicutes followed by Bacteroidetes.
The possible reason for abundance of Firmicutes (*Bacillus* sp.) could be their ability to form endospores to resist adverse conditions of the oil fields, as in desert habitats of Oman [32–35]. Also, because of their metabolic and physiologic adaptability and ability to produce enzyme inhibitors and antibiotics, Firmicutes are considered to be better competitors in natural environment [36]. The moisture content of the soil samples was too low, about 0.03–0.08 m³/m³, which can be another possible reason. The heatmap analysis of genera showed that *Bacillus* had the highest relative abundance followed by *Paenibacillus* and *Brevibacillus* in all the 10 soil samples.

Measures of α-diversity such as Shannon index and Simpson index showed that there was diversity of species OTUs within the community in all the 10 soil samples. Multiple rarefaction curves assembled from each sample’s Shannon diversity index reached a plateau at approximately 5000 sequences suggesting that the sequencing depth was sufficient to capture the full scope of microbial diversity [37]. SHE analysis revealed the constant proportion of species in all 10 samples.

β-Diversity measured using UniFrac analysis provided a tree-based PCoA (Principal Coordinate Analysis) graph. The two approaches revealed the same pattern of clustering; the Monte Carlo simulation test with Bonferroni corrections revealed that there was significant difference in the communities present in the soil samples examined [24].

The microbial community in soils is determined by the physicochemical parameters such as moisture content, temperature, salinity, and pH [38, 39]. The extreme temperature conditions along with low moisture content (0.03–0.08 m³/m³) and a slightly alkaline pH will have an impact on the diversity of bacterial community in the soil. Similar results were reported earlier for Tibetan plateau [40]. The presence of heavy crude oil can be another limiting factor for the microbial community.

The identification of cultivable spore-forming isolates by 16S rDNA sequencing from the soil samples resulted in *Paenibacillus* and *Bacillus* sp. The relative abundance of *Bacillus* sp. in the microbial community in heavy crude oil sludge was reported in Saudi Arabia, Oman, and Nigeria. [41–43]. *Bacillus* and *Paenibacillus* sp. which were relatively abundant in the oil fields can be recommended to be chosen as candidates for hydrocarbon utilization study. One of our isolates, *Paenibacillus ehimensis* BSI, showed maximum growth in presence of heavy oil and biotransformed it to lighter aliphatic and aromatic compounds demonstrating its potential in EOR and environmental bioremediation under aerobic and anaerobic reservoir conditions [11].

**Conflicts of Interest**

The authors declare that they have no conflicts of interest.

**References**

[1] K. Aleklett, M. Höök, K. Jakobsson, M. Lardelli, S. Snowden, and B. Söderbergh, “The Peak of the Oil Age - Analyzing the world oil production Reference Scenario in World Energy Outlook 2008,” *Energy Policy*, vol. 38, no. 3, pp. 1398–1414, 2010.

[2] W. Graus, M. Rogieri, P. Jaworski, L. Alberio, and E. Worrell, “The promise of carbon capture and storage: Evaluating the capture-readiness of new EU fossil fuel power plants,” *Climate Policy*, vol. 11, no. 1, pp. 789–812, 2011.

[3] R. Cosse and M. Koch, “Basics of reservoir engineering,” *Pure and Applied Geophys.*, vol. 140, pp. 739–741, 1993.

[4] C. Hall, P. Thorakam, J. Hallock, C. Cleveland, and M. Jefferson, “Hydrocarbons and the evolution of human culture,” *Nature*, vol. 426, no. 6964, pp. 318–322, 2003.

[5] R. S. Bryant, A. K. Stepp, K. M. Bertus, T. E. Burchfield, and M. Dennis, “Microbial-Enhanced Waterflooding Field Pilots,” *Developments in Petroleum Science*, vol. 39, no. C, pp. 289–306, 1993.

[6] E. Sunde, J. Beeder, R. Nilsen, and T. Torsvik, “Aerobic microbial enhanced oil recovery for offshore use,” in *SPE/DOE Enhanced Oil Recovery Symposium*, pp. 22–24, Society of Petroleum Engineers SPE-24204-MS, Tulsa, Oklahoma, 1992.

[7] S. J. Geetha, I. M. Banat, and S. J. Joshi, “Biosurfactants: Production and potential applications in microbial enhanced oil
[42] I. Allamin, U. Ijah, H. Ismail, and M. Riskuwa, “Occurrence of hydrocarbon degrading bacteria in soil in Kukawa, Borno State,” *International Journal of Environment*, vol. 3, no. 2, 2014.

[43] A. Al-Sayegh, Y. Al-Wahaibi, S. Al-Bahry, A. Elshafie, A. Al-Bemani, and S. Joshi, “Microbial enhanced heavy crude oil recovery through biodegradation using bacterial isolates from an Omani oil field,” *Microbial Cell Factories*, vol. 14, no. 1, article no. 141, 2015.