Universal Indicators for Oil and Gas Prospecting Based on Bacterial Communities Shaped by Light-Hydrocarbon Microseepage in China

Chunping Deng¹, Xuejian Yu¹, Jinshui Yang¹, Baozhen Li¹, Weilin Sun², and Hongli Yuan*¹

¹State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing 100193, P.R. China
²National Research Center for Geoanalysis, Beijing 100037, P.R. China

Introduction

Hydrocarbon seepage is prevalent in the carbon cycle on Earth [13]. A large body of studies have addressed the microbial community structures and diversity of functional genes at hydrocarbon macroseeps (active seeps with large concentrations of migrated hydrocarbons) [56], such as oil-spill areas and oil-contaminated soil, seawater, and sediment, using high-throughput sequencing of 16S rRNA genes, clone libraries of the pmoA, alkB, and nah genes [53, 57, 58], and GeoChip [25, 40, 48]. These reports indicate that high concentrations of hydrocarbons have significant effects on the abundance and diversity of microbial and functional populations in these environments.

Driven by the pressure of subterranean oil and gas reservoirs, low-molecular-weight hydrocarbons, such as methane, ethane, propane, and butane, can vertically penetrate faults and fractures in the reservoirs and migrate upward to the near-surface soils [8], and be utilized by indigenous hydrocarbon-oxidizing microorganisms as potential carbon and energy sources [37]. Thus, the anomalous enrichment of these bacteria caused by long-term and continuous light-hydrocarbon microseeps (passive seeps with low concentrations of migrated hydrocarbons)
could be used as an indicator for petroleum prospecting [38, 49]. Studies of the ecological characteristics of bacteria in these ecosystems could be of great importance for the improvement of petroleum prospecting technology. Until now, most studies on the microbial community structures in these environments have been based on culture-independent approaches, usually with short-chain alkanes (C1–C6) as the sole carbon sources [37, 38], with the results showing higher numbers of light-hydrocarbon oxidizers in hydrocarbon prospective areas than non-prospective areas. However, the vast majority of bacteria are viable but uncultivable in the laboratory. The development of culture-independent molecular biotechnology, especially next-generation high-throughput sequencing, which is known to be superior for detecting rarer bacterial populations at unprecedented depths [5, 51], has greatly facilitated our knowledge of microbial community structures. Nevertheless, only a few studies of microbial communities in hydrocarbon microseepage ecosystems have been conducted with culture-independent methods, such as the clone library approach [32, 60] and denaturing gradient gel electrophoresis (DGGE) analysis [55]. To date, knowledge of the patterns of microbial communities in these ecosystems is still lacking.

The overall ecological characteristics of bacteria are influenced by geographic location and various environmental factors, such as soil type, vegetation, pH, and nutrients [7, 24]. Thus, oil and gas fields with different geographic locations and significant heterogeneity could comprise distinct microbial community structures, and only co-occurring taxa enriched in different oil and gas fields could be useful as “universal indicators” for microbial prospecting of oil and gas reservoirs. However, existing studies have been limited to only one individual oil or gas field, such as the Ban 876 gas and oil field [60], Beihanzhuang oil field in China [55], and a sedimentary basin in Brazil [32]. Thus, it is difficult to find shared taxa enriched in different fields based on the previous studies. Research on the microbial compositions in subsurface soil among different oil and gas fields is much needed to determine universal microbes.

The general method for determining microbial community structures is sampling, high-throughput sequencing, and data analysis [3, 31]. Laboratory simulations are important for the verification of microorganisms with special functions, such as hydrocarbon degradation [16, 41]; however, samples obtained from light-hydrocarbon microseepage ecosystems have been rarely addressed. Despite the fact that methane is the most abundant gas hydrocarbon in petroleum reservoirs, the presence of C2–C6 n-alkane-oxidizing bacteria seems to be more indicative of microseepage from subsurface deposits [46]. Among them, butane originates only from gas and oil fields. Thus, the existence of anomalously high densities of butane-oxidizing bacteria in soil could be an indicator of subterranean petroleum or gas deposits [63]. Only a few studies have been conducted on the anaerobic degradation of butane by marine sulfate-reducing bacteria at marine seeps [20-22]. The community structure of butane-utilizing bacteria above terrestrial oil and gas fields is quite poorly understood.

To address the above issues, we conducted a study to determine the response of microbial communities to long-term light-hydrocarbon microseepage at two typical oil fields and one gas field in China with relatively distinct climates, using high-throughput sequencing and laboratory-simulated incubation with n-butane. The aims of this study were to (i) determine the bacterial community structures in the oil and gas fields; (ii) determine the impact of geographic location and environmental factors on the bacterial community structure; and (iii) propose universal microbial taxa that are enriched by light hydrocarbons at different oil and gas fields compared with the background. Our results provide comprehensive information about the microbial community structures at long-term light-hydrocarbon microseeps, and suggest a highly potential basis for the microbial prospecting of oil and gas.

Materials and Methods

Sample Collection

Soil samples were collected from two typical oil fields and one gas field at different geographic locations in China with the help of the National Research Center for Geoanalysis, Chinese Academy of Geological Sciences: Jianghan (JH) oil field in central China, Shengli (SL) oil field in north China [26], and Puguang (PG) gas field in southwest China (Fig. 1). The three regions have relatively different climates. JH has a subtropical humid monsoon climate, SL has a warm temperate continental semihumid monsoon climate [25], and PG has a subtropical monsoon climate. Soil samples named OJH, OSL, and OPG were taken adjacent to crude oil or gas pumping wells from the JH, SL, and PG sites, respectively. Corresponding background samples named NJH, NSL, and NPG were acquired from the same sites but far away from the known oil and gas fields. Details of the samples are given in Table S1. All the soil samples were collected in five replicates (200 g each) from a depth of about 50 cm to avoid anthropogenic disturbance, mixed thoroughly, and packaged in sterile bags and stored at 4°C and ~20°C for incubation experiments and DNA extraction, respectively.

Gas samples from the subsurface soil (about 80 cm below the ground) were collected using a 615C penetration type sampler equipped with a hand vacuum pump (Eijkelkamp, the Netherlands) at each sampling site. About 12 ml of gas was pumped into each
downward serum bottle (15 ml) prefilled with saturated saltwater. An air sample above each site was also collected. The bottles were sealed with a butyl rubber stopper and kept bottom-up to prevent gas leakage.

**Geochemical Analysis**

All liquid and solid chemicals were of analytical reagent grade and purchased from Sinopharm Chemical Reagent Co., Ltd (China). Mixed standard gas (4.55% methane, 0.25% ethane, 0.054% propane, 0.0537% n-butane, 0.151% iso-butane, and 94.941% nitrogen) was purchased from Beiwen Gas Manufacturer (China) with ≥99.99% purity. The pH values of the samples were determined using a 1:2.5 (w/v) soil:deionized water suspension and twin pH B-212 (Horiba, Japan). The concentrations of total nitrogen (TN), available nitrogen (NO$_3^-$-N and NH$_4^+$-N), total organic carbon (TOC), and total phosphorus (TP) were determined according to a Chinese handbook [12], and the concentrations of water-soluble salts (K$^+$, Ca$^{2+}$, and Mg$^{2+}$) were determined at the Institute of Plant Nutrition and Resources, Beijing Academy of Agriculture and Forestry Sciences.

The compositions of short-chain alkanes in the gas samples (1 ml) were measured using a gas chromatography–flame ionization detection (Finnigan TraceGC Ultra; Germany) according to the method described previously [63], with slight modifications. Specifically, the gas chromatograph was run at a column temperature of 35°C for 4 min and then the temperature was increased to 160°C (20°C/min) and held for 3 min. The inlet and detector temperatures were 200°C and 300°C, respectively. Room air was used as a blank control. Total solvent extractable matter (TSEM) was prepared from 100 g of soil according to the Ultrasonic-Soxhlet extraction gravimetric method [19].

**Butane Incubation Microcosms**

The three OJH samples and two NJH samples collected from Jianghan oil field were chosen for subsequent incubation. For one set of samples, n-butane was used as the sole carbon source and another set of samples were used as controls (room air). For each set, 10 g of fresh, homogenized, and 2-mm-sieved soil was added to a 50 ml serum bottle. The bottles were tightly sealed with butyl rubber stoppers and aluminum crimp caps, and then injected with 6 ml of n-butane or 6 ml of air with a gas-tight syringe (SGE, Australia). Each set was cultured in the dark at room temperature. The concentration of n-butane in the headspace of each microcosm was measured every week (7, 14, 21, and 26 days) by gas chromatography–flame ionization detection, as previously described, to make sure there was sufficient n-butane. Destructive sampling was also performed at each time point and soils were stored at -20°C until DNA extraction.

**DNA Extraction**

Metagenomic DNA was extracted from the soil samples using an E.Z.N.A. Soil DNA Kit (Omega, USA), according to the manufacturer’s instructions. Two or three replicates of each soil sample were individually prepared for subsequent analysis.

**16S rRNA Amplification and DGGE Analysis of Incubated Samples**

The variable V3 region of the bacterial 16S rRNA gene was amplified using primers 2 and 3 [33]. Each PCR mixture contained 2.5 µl of 10× Taq buffer, 1.5 µl of 25 mM MgCl$_2$, 0.5 µl of 2.5 mM dNTP mixture (TaKaRa Co., Shiga, Japan), 0.5 µl of 10 µM each primer, 0.25 µl of 5 U/µl Taq DNA polymerase (Fermentas, Waltham, MA, USA), and approximately 100 ng of genomic DNA.
as the template. Deionized water was added to a total volume of 25 µl. The PCR program included an initial denaturation step at 95°C for 5 min, 19 touchdown cycles of 95°C for 1 min, annealing temperature decreased every cycle by 0.5°C (from 65°C to 55.5°C) for 1 min, and extension at 72°C for 3 min, followed by 5 cycles consisting of 95°C for 1 min, 65°C for 1 min, and 72°C for 3 min, and a final extension step of 10 min at 72°C. The PCR products were concentrated and purified using Wizard SV Gel and the PCR Clean-Up System (Promega, USA) following the manufacturer's instructions.

An approximately 250 ng aliquot of each PCR product was separated on 6% (w/v) denatured polyacrylamide gel using a Dcode System apparatus (Bio-Rad, USA) with a 45–70% denaturing gradient (100% denaturant corresponded to 7 M urea and 40% deionized formamide). DGGE was performed in 0.5× Tris-acetate-EDTA (TAE) buffer at a constant voltage of 50V and a temperature of 60°C for 15 h. The DNA bands were stained with SYBR green I (Amresco, USA) for 30min and photographed through a UV gel documentation system (Bio-Rad, USA).

Illumina MiSeq Sequencing of Bacterial 16S rRNA Genes and Data Analysis

The diversity and composition of the bacterial communities in each of the samples were determined using a protocol described previously [4]. PCR amplifications were conducted using the 515f/806r primer set specific to the V4 region of the 16S rRNA gene [28]. Sequencing was conducted on an Illumina MiSeq platform by Novogene (China).

Sequencing reads were assigned to each sample according to the 6bp unique barcode of each sample, and the barcodes and primers were trimmed after that. Pairs of reads from the original DNA fragments were merged using FLASH [10]. Raw tags were filtered by Quantitative Insights Into Microbial Ecology (QIIME) quality filters with default settings [2, 4]. Chimeric sequences were removed according to previous publications [11, 17]. The sequencing data were analyzed with the QIIME software package [32] and UPARSE pipeline [10], in addition to custom Perl scripts to analyze alpha diversity and beta diversity. Sequences with ≥ 97% similarity were assigned to the same operational taxonomic units (OTUs). The OTU table was rarified to eliminate the effect of sequencing depth on the indices. The small percentage of archaeal sequences was removed [61].

Principal component analysis (PCA) of the geochemical data and the relative contributions of soil geographic location and geochemical parameters to the variations in the bacterial communities in samples (variation partitioning analysis, VPA) were conducted using the vegan package ver. 2.2-1 in the R computing environment [23, 34]. The significance test was carried out by Monte Carlo permutation (999 times). The Pearson’s correlation coefficients of the diversity of the bacterial communities with soil geochemical parameters and the Duncan tests for statistical significance were calculated using SPSS 18.0 software. A non-metric multidimensional scaling analysis (NMDS) of bacterial composition based on the Bray-Curtis distance matrix was performed with PAST ver. 3.0 [47]. The characterization of bacterial features differentiating the samples obtained from oil and gas fields and background areas was carried out using the linear discriminant analysis (LDA) effect size (LEfSe) method, a useful tool that emphasizes both biological relevance and statistical significance [45]. The Kruskal-Wallis rank sum test was used with a significance alpha value of 0.05 to detect features with significantly different abundance among classes, and the threshold on the logarithmic LDA score for discriminative features was 2.5. A more strict strategy for multiclass analysis was set in this study. The relative abundance of each taxon was standardized by subtracting the mean relative abundance in objective samples and dividing the difference by its standard deviation. After that, the hierarchical cluster analysis of these samples was performed using the R package of pheatmap ver. 1.0.8 [36] based on the normalized matrix.

Nucleotide Sequence Accession Number

The sequences obtained by Illumina MiSeq sequencing were deposited in the National Center for Biotechnology Information Sequence Read Archive (SRA) under the accession number SRP063715.

Results

Geochemical Analysis of Soil Samples

The geochemical parameters of the soil samples are shown in Table S1. The concentration of TSEM among samples ranged from 20 to 541.6 µg/g. Methane was present at all the sampling sites in Jianghan (JH) with a concentration of 259.55–424.20 ppm in OJH (oil field) and 291.80–372.91 ppm in NJH (background) samples (Table S2). Notably, n-butane and iso-butane were only detected in the OJH2, OJH3, and OJH5 samples obtained next to the oil and gas wells. A relatively low concentration of methane (1.65–25.2 ppb) was detected in the Puguang (PG) and Shengli (SL) samples. Thus, the subsequent analysis did not include information on light hydrocarbons.

PCA of the geochemical parameters in all samples revealed that samples were distinctly separated based on geographic location, especially for samples collected from PG and JH (Fig. 2). Moreover, the key environmental factors were distinct among the different oil and gas fields. Specifically, JH samples (OJH and NJH) contained relatively high-level nutrient factors (TOC, TN, TP, and C/N) and moisture, and PG samples (OPG and NPG) were mainly positively affected by altitude and NO3-N. The TSEM, Mg2+ and Ca2+ contents, pH value, and location were positively correlated with the OSL samples, whereas weak correlations were found between NSL samples and the...
For all the initial samples, a total of 2,198,466 clean reads were obtained after rigorous quality control through Illumina MiSeq sequencing, and were affiliated with 4,554 OTUs on average at 97% sequence identity. The α-diversity indices are shown in Table 1. Samples OJH and OPG revealed significantly higher richness and diversity than samples NJH and NPG, respectively. The phylogenetic diversity index exhibited the same trend. In order to compare the effect of environmental factors on the bacterial diversity between the next-to-well samples and background samples, Pearson’s correlation analysis was performed. The result showed that the diversity of the background samples was significantly influenced by almost all the environmental factors detected in our study (Table 2). Unlike the background samples, the diversity of next-to-well samples was significantly influenced by the major nutrient factors (TOC, TN, NH₄⁺-N, and TP), whereas the geographic location (latitude, longitude, and altitude) and pH values showed a nonsignificant effect. Notably, the bacterial diversity of the next-to-well and background samples did not show a significant correlation with TSEM concentration.

**Comparison of Bacterial Assemblages and the Influence of Environmental Factors**

The NMDS analysis of the bacterial compositions among the samples showed that samples OJH2, OJH3, and OJH5 were separated from NJH6 and NJH7 by NMDS2 (Fig. 3). The bacterial compositions of pristine samples NPG10, NPG11, and NPG13 were remarkably different from samples OPG5 and OPG7. Samples OSL5 and OSL7 were considerably dissimilar to samples NSL1, NSL2, and NSL4. These results indicated that the bacterial communities of samples from the two oil fields and one gas field were notably different from those in the corresponding background samples. Surprisingly, the detected bacterial compositions in samples OPG5 and OPG7 were relatively similar to the OSL samples, implying that samples collected

**Table 1. Alpha-diversity indices of samples in the present study.**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Chao1 richness</th>
<th>OTU numbers</th>
<th>Shannon index</th>
<th>Phylogenetic diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>OJH</td>
<td>10,661.35 (986.99)</td>
<td>4,972.33 (475.00)</td>
<td>9.23 (0.83)</td>
<td>380.54 (30.62)</td>
</tr>
<tr>
<td>NJH</td>
<td>8,915.04 (1,197.43)</td>
<td>4,255.25 (641.97)</td>
<td>8.52 (1.39)</td>
<td>327.31 (39.37)</td>
</tr>
<tr>
<td>OPG</td>
<td>8,940.66 (1,246.87)</td>
<td>4,481.67 (625.03)</td>
<td>9.32 (1.07)</td>
<td>332.24 (33.56)</td>
</tr>
<tr>
<td>NPG</td>
<td>7,600.76 (472.73)</td>
<td>3,845.33 (173.31)</td>
<td>8.29 (0.24)</td>
<td>299.00 (11.27)</td>
</tr>
<tr>
<td>OSL</td>
<td>9,460.00 (376.87)</td>
<td>4,655.50 (219.62)</td>
<td>9.98 (0.34)</td>
<td>341.34 (12.58)</td>
</tr>
<tr>
<td>NSL</td>
<td>10,160.72 (418.42)</td>
<td>5,105.25 (149.12)</td>
<td>10.41 (0.07)</td>
<td>364.98 (6.96)</td>
</tr>
</tbody>
</table>

Standard deviations are in parentheses.

Sample names with the letter O are samples obtained adjacent to oil or gas pumping wells, and those with the letter N are background samples. OJH represents the mean diversity of OJH2, OJH3, and OJH5. NJH represents the mean diversity of NJH6 and NJH7. OSL represents the mean diversity of NSL1, NSL2, and NSL4. NSL represents the mean diversity of OSL5, OSL6, and OSL7. OPG represents the mean diversity of OPG5, OPG6, and OPG7. NPG represents the mean diversity of NPG10, NPG11, and NPG13. Values with different lowercase letters in the same column are significantly (p < 0.05) different from each other, according to Duncan’s test.
from distant regions with light-hydrocarbon microseepage could have similar bacterial compositions, which suggests the possibility of the presence of universal indicators for oil and gas prospecting.

To quantify the contributions of geographic location and soil environmental variables to the bacterial community variation in the OS samples, VPA was carried out. The result revealed that a total of 82.93% of the variation was explained by the detected environmental parameters. Geographic location and geochemical parameters were able to independently explain 7.05% and 36.23% of the total variation, respectively. Interactions between the two components showed more influence (39.65%) than the individual components did, implying a strong correlation between them that should not be ignored. Only 17.07% of the total variation could not be explained.

### Bacterial Composition and Discriminating Taxa

The relative abundance of the top 14 phyla in all the samples is shown in Fig. 4A. Samples obtained from different areas had similar dominant phyla, but varied in their relative abundance. Proteobacteria was the most abundant phylum among all the samples, ranging from 22.81% to 41.83%. Notably, Firmicutes was the second dominant phylum among JH and PG samples, ranging from 6.71% to 28.57% and represented mostly by bacilli. Acidobacteria, Actinobacteria, Chloroflexi, Bacteroidetes, and Gemmatimonadetes were prevalent among these samples as well. Other phyla, such as Planctomycetes, Nitrospirae, Verrucomicrobia, WS3, Cyanobacteria, Chlorobi, and Armatimonadetes, were found to account for only small proportions in the samples. The bacterial communities at the phylum level varied substantially between the next-to-well and background samples at each site, and were even more distinct among the three sites (Fig. 4B).

**Table 2. Statistical analysis of the relationship between richness and diversity of bacterial communities and soil physicochemical parameters based on Pearson’s correlation coefficients.**

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>OS Community index</th>
<th>NS Community index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Richness</td>
<td>Shannon</td>
</tr>
<tr>
<td>Latitude</td>
<td>-0.320</td>
<td>-0.143</td>
</tr>
<tr>
<td>Longitude</td>
<td>-0.140</td>
<td>-0.313</td>
</tr>
<tr>
<td>Altitude</td>
<td>-0.030</td>
<td>0.368</td>
</tr>
<tr>
<td>K⁺</td>
<td>-0.434*</td>
<td>-0.372</td>
</tr>
<tr>
<td>Ca²⁺</td>
<td>-0.323</td>
<td>-0.130</td>
</tr>
<tr>
<td>Mg²⁺</td>
<td>-0.271</td>
<td>-0.088</td>
</tr>
<tr>
<td>Moisture</td>
<td>0.530**</td>
<td>0.008</td>
</tr>
<tr>
<td>pH</td>
<td>-0.052</td>
<td>0.003</td>
</tr>
<tr>
<td>TOC</td>
<td>0.418*</td>
<td>-0.030</td>
</tr>
<tr>
<td>TN</td>
<td>-0.050</td>
<td>-0.594**</td>
</tr>
<tr>
<td>C/N</td>
<td>0.465*</td>
<td>0.195</td>
</tr>
<tr>
<td>NH₄⁻-N</td>
<td>0.504*</td>
<td>-0.034</td>
</tr>
<tr>
<td>NO₃⁻-N</td>
<td>-0.201</td>
<td>0.095</td>
</tr>
<tr>
<td>TP</td>
<td>0.172</td>
<td>-0.443*</td>
</tr>
<tr>
<td>TSEM</td>
<td>-0.315</td>
<td>-0.222</td>
</tr>
</tbody>
</table>

*Observed operational taxonomic unit numbers. OS represents samples collected adjacent to crude oil or gas pumping wells, including OJH, OSL, and OPG; NS refers to samples obtained from the corresponding background area (NJH, NSL, and NPG).

**Fig. 3.** Analysis of bacterial community compositions using non-metric multidimensional scaling (NMDS) analysis based on the Bray-Curtis distance matrix.

The dominant taxa in all samples were analyzed using LEfSe to identify specific phylotypes as biomarkers between next-to-well and background samples. The cladogram showed that Bacteroidetes was the discriminating phylum, with a significantly higher abundance in next-to-well samples compared with background samples (Fig. 5 and Table S3; \(p < 0.05\)), whereas the phyla Acidobacteria and Verrucomicrobia were much less in proportion. Moreover, the class Acidimicrobia within the phylum Actinobacteria, classes Cytophaga, Flavobacteria, and Sphingobacteria within the phylum Bacteroidetes, and class Anaerolineae within the phylum Chloroflexi were significantly enriched in next-to-well samples. At the order level, next-to-well samples had a remarkably higher abundance of Cytophagales.
Flavobacteriales, Desulfuromonadales, Alteromonadales, Oceanospirillales, Pseudomonadales, and Thiotrichales, whereas background samples favored the growth of orders Acidobacterales, Solibacterales, Actinomycetales, Bacteroidales,
and Thermogemmatisporales. The abundant families Cytophagaceae, Mycobacteriaceae, Pirellulaceae, Geobacteraceae, Campylobacteraceae, Alteromonadaceae, Ectothiorhodospiraceae, and Piscirickettsiaceae occupied a notably larger proportion in next-to-well samples. It is worth mentioning that LEfSe highlighted several genera that were significantly more abundant in next-to-well samples compared with background samples, such as *Mycobacterium*, *Flavobacterium*, *Geobacter*, *Pseudomonas*, *Arenimonas*, and *Lysobacter*, suggesting their potential application as distinguishing biomarkers of underlying oil and gas deposits. In contrast, the relative abundance of the genera *Nocardiooides* and *Enterococcus* was significantly less in next-to-well than in background samples.

**Response of Key Taxa under Laboratory Simulation with n-Butane**

To confirm the response of taxa enriched in next-to-well samples to light hydrocarbons, laboratory-simulated incubation with n-butane was carried out. Samples collected from JH were selected as representatives based on the relatively high concentration of light hydrocarbons detected in these sites. To determine the appropriate incubation time, DGGE analysis was conducted on samples OJH5 and NJH7 as representatives to uncover the bacterial community structure dynamics during the cultivation. The DGGE fingerprints showed remarkably different microbial community structures between treatments with or without n-butane after incubation for 14 days and an even longer time (e.g. 21 and 26 days; see Fig. S1). Considering that long-term incubation might raise the possibility of cross-feeding to a large extent [16], samples treated for 14 and 21 days with n-butane or an equal volume of air were chosen for Illumina MiSeq sequencing to investigate the potential bacteria able to utilize n-butane.

The hierarchical clustering map at the genus level showed that the microbial community structure of samples obtained from the oil field were distinct from those from pristine soil, even after the n-butane stimulation (Fig. 6). n-Butane greatly favored the growth of *Azoarcus* (the relative abundance increased from 0.42% to 1.16%), *Hydrogenophaga* (from 0.08% to 0.24%), *Mycobacterium* (from 0.51% to 0.74%), *Pseudomonas* (from 0.58% to 1.02%), *Rhodococcus* (from 0.04% to 0.15%), and *Rubrivivax* (from 0.17% to 0.32%) in OJH samples in 14 or 21 days, indicating an immediate response of these genera to butane stimulation. For the NJH samples, the relative abundance of *Methylibium*, *Hydrogenophaga*, *Nocardiooides*, *Sphingomonas*, *Pseudonocardia*, *Polaromonas*, *Nevskia*, and *Rubrivivax* was notably increased after incubation with n-butane as compared with the treatment with air. However, the relative abundance of *Nocardia*, *Lactococcus*, *Bacillus*, *Steroidobacter*, *Acinetobacter*, and *Carnobacterium* increased considerably even without n-butane.

**Discussion**

**Bacterial Community Structures in Light-Hydrocarbon Microseepage Ecosystems and Their Correlation with Environmental Factors**

A large body of studies have addressed the influence of soil pH on the diversity and richness of bacterial communities [14, 62]; however, no significant correlation was found
between pH values and the diversity of next-to-well samples in this study. The diversity of next-to-well samples was found to be significantly correlated with the major nutrient factors (TOC, TN, and TP), whereas the diversity of background samples was affected by almost all the detected environmental variables. The notable distinction between the next-to-well and background samples was probably because of the continuous migration of light hydrocarbons from subsurface petroleum reservoirs in the next-to-well samples. Light hydrocarbons can be utilized as a carbon source by hydrocarbon-oxidizing bacteria [37], which might result in the change of TOC. As nitrogen and phosphorus are the key nutrient elements for microbes [9, 44], the utilization of extra light hydrocarbons in next-to-well samples might be mainly affected by these nutrient factors.

Although the diversity of OSL was lower than NSL, an unexpected outcome in our work was the significantly higher diversity in next-to-well samples compared with background samples. Oil contamination has been reported to decrease the diversity of microbes and functional genes in previous studies [1, 25, 35, 57]. This is probably because relatively high concentrations of oil are highly toxic, mutagenic, and/or carcinogenic to microorganisms [42]. In contrast, microbial communities at long-term light-hydrocarbon microseeps might have acclimatized to trace and continuous light hydrocarbons through horizontal gene transfer [27]. In addition, there was no significant correlation between the bacterial diversity and the concentration of TSEM. Although this could be attributed to the much lower extent of oil contamination in our samples compared with previous reports [25, 57], this further illustrated the relatively different eco-environments of light-hydrocarbon microseepage ecosystems in the present study and oil-contaminated ecosystems. Wu et al. [55] suggested that light-hydrocarbon microseepage could change the composition of bacterial communities, but showed no significant influence on bacteria in the Beihanzhuang oil field using the DGGE method. However, Man et al. [29] found relatively higher α-diversity at oil and gas fields compared with the control area by PCR-DGGE, which showed a similar trend with our study. Our results showed notably different trends in bacterial diversity stimulated by relatively long-term light-hydrocarbon microseepage compared with previous reports in short-term oil contamination, suggesting that bacterial diversity could be a new indicator for microbial prospecting of oil and gas. Considering that scant information about the bacterial diversity at light-hydrocarbon microseeps compared with corresponding background samples is available, further work is urgently required.

Understanding the factors that influence microbial community structures could be of great help to unravel the patterns of microbial distribution, which is crucial in microbial ecology. In the present study, PCA suggested that samples collected from different oil and gas fields comprised remarkably different geochemical properties. Surprisingly, NMDS analysis indicated that samples obtained from distant fields could have similar bacterial community structures, such as samples OPG and OSL. Moreover, there was no significant correlation between the diversity of the next-to-well samples and the geographic location. Furthermore, VPA showed that only 7.05% of the bacterial community in next-to-well samples was explained by geographic location independently, whereas it explained 33.5% in five oil-contaminated fields in China [25]. These results imply the possibility of the presence of “core microbes” stimulated by long-term light-hydrocarbon microseepage, even at distant geographic locations with great heterogeneities, which could be used as potential “universal indicators” for subsurface oil and gas reservoirs. The bacterial community structures between samples OSL and NSL were not significantly different compared with other samples, which might be because of the anthropogenic disturbances at this area with developed industry. Previous study indicated that anthropogenic disturbance has great effect on the biosphere and global biogeochemical processes, which substantially altered the community structures and their ecological functions [6]. However, further research is still needed.

Proposed “Universal Indicators” in Oil and Gas Fields

The few reports on microbial communities in light-hydrocarbon microseepage ecosystems have illustrated relatively different results at different phylogenetic levels. Based on the 16S rRNA gene clone libraries, Chloroflexi and Gemmatimonadetes were found to be dominant in the Ban 876 gas and oil field in China [60] and Actinobacteria, Proteobacteria, and Acidobacteria were the most dominant phyla, whereas Gemmatimonadetes, Bacteroidetes, Chloroflexi, Cyanobacteria, and Firmicutes were the least numerous phyla in petroliferous soil from a sedimentary basin in Brazil [32]. Nocardioides, Aciditerrimonas, sulfate-reducing bacteria, and Chloroflexi were proposed to be novel indicators for microbial prospecting of oil and gas in the Beihanzhuang oil field using the DGGE method [55]. Methylocystaceae might act as a potential indicator for an unexploited gas resource, and Methyllophaga and Alcanivorax for oil using PCR-DGGE [29].

work was limited by methods with relatively low resolution, and only one oil or gas field was considered in each study, there is still insufficient information to determine universal indicators in different oil and gas fields for oil and gas prospecting.

In this study, LEfSe analysis indicated that the phylum Bacteroidetes was significantly enriched in next-to-well samples compared with background samples, which is relatively consistent with previous reports, suggesting the possibility of Bacteroidetes as an indicator. However, hierarchical clustering of dominant phyla indicated that samples were grouped based on geographic location (Fig. 4B). As a high phylogenetic level, one phylum comprises many taxa, which could be affected by many environmental variables [7, 18]; this might decrease their reliability as a universal phylum among different fields.

At a relatively low phylogenetic level, the genera *Mycobacterium* and *Pseudomonas* were found to be significantly enriched in next-to-well samples through the LEfSe method. Interestingly, both genera notably bloomed in OJH samples after n-butane incubation. *Mycobacterium* and *Pseudomonas* have been reported to be capable of degrading short-chain hydrocarbons [46], as well as other alkanes [50], and have also been detected in other light-hydrocarbon microseepage ecosystems, such as a sedimentary basin in Brazil [32]. In addition, primers were designed based on several strains of *Mycobacterium* and *Pseudomonas* for the detection of propane and butane-oxidizing microorganisms (e.g., Cano et al., 2013. USA Patent). Our results, in combination with previous reports, further underpin their possibility as “universal indicators” for subsurface oil and gas reservoirs, despite the dramatically different environmental characteristics among these study sites. Still, more extensive samples should be studied to make these findings more confirmative. Moreover, *Lysobacter*, *Geobacter*, and *Arenimonas* were found to be more abundant in the oil field samples, implying that they may take part in hydrocarbon metabolism directly or indirectly and their potential application as biomarkers. Further work based on DNA-SIP (stable-isotope probing) or RNA-SIP is necessary to further confirm their hydrocarbon-oxidizing ability and explore more hydrocarbon-degrading bacteria.

*Nocardioides* and *Rhodococcus* are also capable of utilizing light hydrocarbons [39]. In this study, although *Rhodococcus* was not significantly enriched in next-to-well samples, the relative abundance indeed increased notably after n-butane incubation for 21 days, indicating that *Rhodococcus* could be proposed as an assistant indicator. However, the relative abundance of *Nocardioides* increased considerably even without n-butane, as well as *Lactococcus*, *Bacillus*, *Steroidobacter*, *Acinetobacter*, and *Carnobacterium*, indicating that these genera have abilities other than butane degradation. *Nocardioides* has been reported to be able to degrade short-chain hydrocarbons [46] and has been proposed as a new indicator [55]. However, our results showed a much lower abundance of *Nocardioides* in next-to-well samples than in background samples, which might be due to their metabolic versatility [59]. Therefore, they should not be adopted as reliable indicators for petroleum deposits.

Contemporary environmental disturbances and historical contingencies are considered to be the two main factors shaping microbial communities [30]. Previous reports have suggested that the relative influence of historical contingencies and environmental factors on bacterial communities is scale dependent [15, 43, 54]. At the present scale (about 1,000 km), VPA showed that the bacterial communities in next-to-well samples were mainly explained by environmental factors other than geographic location independently, which is relatively consistent with Wu et al. [54], implying the potential application of our findings for microbial prospecting of oil and gas deposits at a regional spatial scale (about 1,000 km). Still, more extensive samples at a much larger scale should be studied.

One of the final goals of investigating microbial populations and distribution patterns in long-term light-hydrocarbon microseepage environments is to reduce the drilling risks in petroleum exploration. For the first time, our results have illustrated remarkably higher diversity at light-hydrocarbon microseeps compared with the background area at two of the three sites, based on high-throughput sequencing. Owing to the notable enrichment of *Mycobacterium* and *Pseudomonas* in next-to-well samples and their substantial increase in abundance after laboratory simulation with n-butane, these genera are proposed as potential universal indicators for the microbial prospecting of oil and gas reservoirs. Further work on the ecological characteristics of microbial communities in light-hydrocarbon microseeps among different areas at much broader scales will be necessary. These results, integrated with geological, geophysical, and geochemical evidence, could be invaluable in achieving higher success in forecasting the existence of oil and gas deposits.

**Acknowledgments**

This work was financially supported by the National Natural Science Foundation of China (No. 31270533). The authors declare that they have no conflict of interest. This
article does not contain any studies with human participants or animals performed by any of the authors.

References


58. Yang YY, Wang J, Liao JQ, Xie SG, Huang Y. 2015. Abundance and diversity of soil petroleum hydrocarbon-degrading...