Molecular analysis of the bacterial community in a continental high-temperature and water-flooded petroleum reservoir

Hui Li¹, Shi-Zhong Yang¹, Bo-Zhong Mu¹, Zhao-Feng Rong² & Jie Zhang²

¹Institute of Applied Chemistry and ²Institute of Bioengineering, East China University of Science and Technology, Shanghai, China

Correspondence: Bo-Zhong Mu, Institute of Applied Chemistry and 2Institute of Bioengineering, East China University of Science and Technology, Shanghai, China. Tel.: +86 21 64252063; fax: +86 21 64252485; e-mail: bzmu@ecust.edu.cn

Received 8 November 2005; revised 4 January 2006; accepted 11 January 2006.
First published online 15 February 2006.
doi:10.1111/j.1574-6968.2006.00149.x
Editor: J.C. Murrell

Keywords
petroleum reservoir; 16S rRNA gene diversity; restriction fragment length polymorphism; bacterial community; clone library.

Abstract
Water from a continental high-temperature, long-term water-flooded petroleum reservoir in Huabei Oilfield in China was analysed for its bacterial community and diversity. The bacteria were characterized by their 16S rRNA genes. A 16S rRNA gene clone library was constructed from the community DNA, and using restriction fragment length polymorphism analysis, 337 randomly selected clones were clustered with 74 operational taxonomic units. Sequencing and phylogenetic analyses showed that the screened clones were affiliated with Gammaproteobacteria (85.7%), Thermotogales (6.8%), Epsilonproteobacteria (2.4%), low-G+C Gram-positive (2.1%), high-G+C Gram-positive, Betaproteobacteria and Nitrosira (each < 1.0%). Thermophilic bacteria were found in the high-temperature water from the flooded petroleum reservoir, as well as mesophilic bacteria such as Pseudomonas-like clones. The mesophilic bacteria were probably introduced into the reservoir as it was being exploited. This work provides significant information on the structure of bacterial communities in high-temperature, long-term water-flooded petroleum reservoirs.

Introduction
The environment of deep subsurface petroleum reservoirs is generally characterized by a high temperature (up to 180 °C), high pressure (up to 40 MPa), high salinity (up to 20 g L⁻¹ total dissolved solids) and by anaerobic systems with multiphase fluids of oil, gas and water. Recently, much attention has been paid to the microorganisms which inhabit this environment due to their possible application in industrial processes. Over the last decade, a body of convergent observations has highlighted the great diversity of indigenous microorganisms in subsurface petroleum reservoirs, as well as in the exogenous microorganisms introduced to reservoirs in drilling operations and from water flooding in oil exploitation (Stetter et al., 1993). These various aerobic and anaerobic groups of microorganism are able to degrade oil hydrocarbons and to synthesize such oil-releasing agents as fatty acids, alcohols, polysaccharides and biosurfactants. The technology known as microbial enhanced oil recovery has been applied by either the injection of selected microorganisms with their nutrients into the target petroleum reservoir, or simply the injection of microbial nutrients into the target reservoir to stimulate those indigenous microorganisms that have an oil-releasing capacity. Laboratory investigations and field trials, in close collaboration with the petroleum industry, have indicated that the in situ biophysical and biochemical activities of these microorganisms, as well as their viability, has contributed to an enhancement of oil recovery. To understand these activities, a knowledge of the structure of microbial communities in subsurface petroleum reservoirs is of great importance.

Using traditional culture-dependent approaches many different species of microorganisms have been found in subsurface petroleum reservoirs, including sulfate reducers (Nilsen et al., 1996; Rueter et al., 1994), sulfidogens (L’Haridon et al., 1995; Magot et al., 2000), hydrocarbon-oxidizing bacteria (Nazina et al., 2001), fermentative microorganisms (Grassia et al., 1996; Van Hamme et al., 2003), methanogens (Nilsen & Torsvik, 1996), manganese and iron reducers (Rees et al., 1995; Greene et al., 1997; Slobodkin et al., 1999), and acetogens (Davydova-Charakhch’yan et al., 1993). Nevertheless, our knowledge of the microbial diversity and community structure of these particular subsurface ecosystems is still limited, since many of the microorganisms in these ecosystems are ‘non-culturable’. An application of various molecular techniques has allowed a more complete characterization of them, and evidence has increasingly
indicated that culture-independent techniques, in particular the analysis of retrieved 16S rRNA genes, are effective in characterizing complex microbial assemblages in environmental samples (Amann et al., 1995). Molecular methods based on reverse sample genome probing, dot-blot DNA hybridization with functional gene probes and 16S rDNA analysis have been applied in identifying sulfate-reducing bacterial populations inhabiting a low-temperature water-flooded of western Canadian harboring reservoir (Voor- douw et al., 1992, 1996). Both 16S rRNA gene phylogenetic analysis and enrichment culture techniques have recently been used to characterize thermophilic microbial assemblages in the Miocene Monterey formation, a prominent high-temperature, oil-bearing formation in California (Orphan et al., 2000). In contrast, a parallel measurement using culture-based enrichments, 16S rRNA gene sequence analysis and oligonucleotide matrix array hybridization methods was carried out to investigate the microbial groups encompassing key genera of thermophilic bacteria and archaea of a continental high-temperature oil reservoir in Western Siberia, Russia (Bonch-Osmolovskaya et al., 2003). A comparable study of the microbial community in a low-temperature, low-salinity and biodegraded petroleum reservoir from a Western Canadian Sedimentary Basin has been reported, which employed a multidisciplinary approach including chemical and geochemical examinations, biodegradation studies, and culture-based and 16S rRNA gene analyses (Grabowski et al., 2005).

In this paper we used an rRNA approach – the cloning and sequencing of 16S rRNA gene fragments – to analyze bacterial communities and their proportions in a continental high-temperature, water-flooded petroleum reservoir in the J-12 Unit at Huabei Oilfield, Hebei Province, China. The J-12 Unit has been in primary production since March 1987. Flooding of the reservoir began in May 1988 and from that time on, it has been continuously flooded by the recycling of production water from the reservoir.

Materials and methods

Sample collection and DNA extraction

The samples of production water were directly collected in sterile steel screw-cap bottles from a sampling valve at the pipeline of the well head in May 2005. The bottles, which were completely filled with the production water (an oil/water mixture), were sealed from contamination and immediately taken to the laboratory where they were stored at 4°C before concentration. The in-situ temperature and pressure of the target reservoir were about 75°C and 18 MPa, respectively. The salinity of the groundwater in oil-bearing formations at a depth of 1500–1700 m in the reservoir was 16 622 mg L⁻¹.

The oil in the water samples was removed by heating the samples to 70°C for 15 min, followed by phase separation in a 2 L sterile separation funnel. The microbial biomass was collected from approximately 1 L of the water phase by centrifugation at 15 000 g, at 4°C. Total community DNA was extracted from the cell pellets using a lysozyme/proteinase K/sodium dodecyl sulfate (SDS) treatment, followed by standard phenol/chloroform extractions (Murray et al., 1998). Nucleic acids were purified with a DNA purification kit (V-gene, China).

16S rRNA gene library construction and RFLP analysis

Bacterial 16S rRNA genes were amplified from total community DNA by PCR using the combination of a bacterial primer 27f (5'-AGA GTT TGA TCC TGG CTC AG-3') and universal primer 1492r (5'-GGT TAC CTT GTT ACG ACT T-3') (Lane, 1991). The PCR was performed with a ‘reconditioning PCR’ program (Thompson et al., 2000) and the amplified products were purified (Sambrook & David, 2001) and cloned with a T-vector (Takara, Japan).

Insert DNA was reamplified with pUC universal primers and subjected to restriction fragment length polymorphism (RFLP) by separate enzymatic digestion (Giselle et al., 1994) with HinfI and HaeIII restriction endonucleases. Clones which showed an identical RFLP pattern were clustered as an operational taxonomic unit.

Sequencing and phylogenetic analysis

One to three representative clones from each unique OTU were selected for sequencing. Insert DNA was sequenced on an automated ABI 377 sequencer (Dye-Terminator Cycle Sequencing Ready Reaction FS Kit; PE Applied Biosystems) using M13 universal primers. Chimeras were checked with the Chimera-Check program (Cole et al., 2003) from the Ribosomal Database Project II (http://rdp.cme.msu.edu/). The sequences were initially submitted to GenBank at the NCBI (http://www.ncbi.nlm.nih.gov), using the BLAST network service (Altschul et al., 1997) and FASTA program (version 3) (Pearson & Lipman, 1988) to determine their closest phylogenetic relatives. Sequences that differed by less than 3% were considered to belong to the same phylotype (Stackebrandt & Goebel, 1994), and each phylotype was represented by a sequence type. Sequences were aligned to their nearest neighbor using Clustal X software (Thompson et al., 1997). Phylogenetic trees were constructed based on the Kimura two-parameter model (Kimura, 1980) and neighbor-joining algorithm (Saitou & Nei, 1987) using the Phylip package (Felsenstein, 2005). Bootstrap analysis with 1000 replicates was applied to assign confidence levels to the nodes in the trees.
Nucleotide sequence accession numbers

The GenBank accession numbers for the rDNA sequences are as follows: HBO1–HBO23, DQ201206–DQ201228; HBO25–HBO33, DQ201229–DQ201237; HBO35–HBO45, DQ201238–DQ201248; HBO46, DQ240873; HBO47, DQ201249; HBO48, DQ223058; HBO49–HBO61, DQ201250–DQ201262; HBO62–HBO68, DQ223044–DQ223050; HBO69, DQ223052; HBO70, DQ223051; and HBO71–HBO75, DQ223053–DQ223057.

Results

A total of 337 clones were characterized by PCR as having the correct insert DNA. The RFLP analysis showed that all of the clones fell into 74 OTUs, with four predominant OTUs accounting for 73.6% of the gene library. The remaining 70 OTUs were presented at low levels, of which 57 OTUs were represented by a single clone. A comparative analysis of the retrieved sequences showed that all the clones were clustered within the domain Bacteria. The phylogenetic analysis indicated that most sequence types had relatively high levels of similarity with their closest counterparts in the public databases (Table 1). Our phylogenetic analysis placed the 69 HBO (Huabei Oilfield) sequence types into the following groups of the domain Bacteria: Gammaproteobacteria, Epsilonproteobacteria, low-G+C Gram-positive, high-G+C Gram-positive, Betaproteobacteria, and Nitrospira.

Gammaproteobacteria

A total of 289 clones, represented by 48 sequence types and accounting for 85.7% of the gene library, fell into the Gammaproteobacteria group (Table 1, Fig. 1). Of which 272 clones, represented by 40 sequence types and accounting for 80.7% of the gene library, were related (> 92.8% identity) to Pseudomonas bacteria, and 16 clones represented by seven sequence types and accounting for 4.5% were related (> 99% identity) to Serratia bacteria.

The most abundant sequence type in this group, as well as in the gene library, HBO45, displaying 176 clones and accounting for 52.3% of the library, was closely related (100% identity) to Pseudomonas sp. PDB (Fig. 1), a chlo-rate-respiring bacterium retrieved from the Pennsylvania State University wastewater treatment plant (Logan et al., 2001). The second most abundant sequence type, HBO46, including 52 clones and accounting for 15.4%, was closely related to Pseudomonas sp. BWDY-9 (DQ200852) with 97.9% similarity, a cultivated species isolated from Yellow River estuary. Representing three clones, the sequence type HBO50 was highly related (98.4% identity) to the uncultured bacterium O1 (AY770933), an clone from a production well of the Dagang oilfield in China.

Epsilonproteobacteria

Seven sequence types, representing eight clones and accounting for 2.4% of the gene library, were clustered within the Epsilonproteobacteria group. The similarities of these phylotypes to previously determined rRNA gene sequences were greater than 93%. Six of clones were related to uncultivated species and the remaining two were related to cultivated bacterial clones of the genera Wolinella and Campylobacter (Table 1).

The sequence types HBO3, HBO38 and HBO43, each representing one clone, were closely related (98% identity) to the uncultured bacterial clones PL-5B5, PL-28B12 and PL-14B3, respectively (Table 2, Fig. 2), which had previously been recovered from the production waters of a non-water-flooded low-temperature and low-salinity petroleum reservoir in Canada (Elizaveta et al., 2003). The sequence type, HBO75, only one clone, was identical (100% identity) to the uncultured bacterium clone CCSD DF2030 B20, a clone which had been recovered from ultra-high-pressure rocks and drilling fluids from the Chinese Continental Scientific Drilling Project (Zhang et al., 2005).

Low-G+C Gram-positive

Seven clones, represented by five sequence types and accounting for 2.1% of the gene library, were clustered within low-G+C Gram-positive bacteria (Table 1, Fig. 2). Three clones, represented by HBO65, which was the most abundant sequence type in this group, were closely related (97.6% identity) to Thermoanaerobacter sp. SL9 (AY216597), a thermophilic anaerobic bacterial clone recovered from an oil reservoir in France. The HBO68 sequence type, displaying one clone, was 96% similar to the rRNA gene of Thermoanaerobacter sp. MET-G (AY800104), a novel species of the genus Thermoanaerobacter isolated from an oilfield in France. No cultivated bacterial members were

Table 1. Distribution of sequence types from the bacterial 16S rRNA gene library

<table>
<thead>
<tr>
<th>Putative division</th>
<th>No. of RFLP patterns</th>
<th>No. of sequence types</th>
<th>Similarity (%)</th>
<th>No. of clones (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-G+C Gram-positive</td>
<td>1</td>
<td>1</td>
<td>98.5</td>
<td>3 (&lt; 1)</td>
</tr>
<tr>
<td>Low-G+C Gram-positive</td>
<td>5</td>
<td>5</td>
<td>93.0–97.6</td>
<td>7 (2.1)</td>
</tr>
<tr>
<td>Nitrospira</td>
<td>1</td>
<td>1</td>
<td>93.6</td>
<td>1 (&lt; 1)</td>
</tr>
<tr>
<td>Thermotogales</td>
<td>3</td>
<td>3</td>
<td>92.8–96.1</td>
<td>24 (6.8)</td>
</tr>
<tr>
<td>Betaproteobacteria</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>1 (&lt; 1)</td>
</tr>
<tr>
<td>Epsilonproteobacteria</td>
<td>7</td>
<td>7</td>
<td>93.8–100</td>
<td>8 (2.4)</td>
</tr>
<tr>
<td>Gammaproteobacteria</td>
<td>55</td>
<td>50</td>
<td>70.9–100</td>
<td>290 (85.7)</td>
</tr>
</tbody>
</table>

*Percentage of 16S rRNA gene similarity to its closest relative.
^Proportion of clones in the library.
found for HBO28. It represented one clone that was closely related (96.6% identity) to the uncultured *Thermovenabulum* sp. B8-otu12 (DQ097677), an anaerobic microorganism found in the high-temperature regions of the Dagang oilfield in China.

**Thermotogales**

A total of 23 clones, represented by two sequence types and accounting for 6.8% of the gene library, clustered with the *Thermotogales* group (Table 1, Fig. 2). Sequence type HBO69, including 20 clones and holding 5.9%, exhibited a 93.6% similarity to *Thermotogaes hypogea* SEBR 7054, a thermophilic, strictly anaerobic and rod-shaped bacterium isolated from an African oil-producing well (Fardeau et al., 1997).

**Others**

Three clones displayed by the single sequence type, HBO70, clustered with the high-G+C Gram-positive group (Table 1). They were closely related (98.5% identity) to *Mycobacterium massiliense* CCUG 48898 (Fig. 2), a strain isolated from a pure culture of the sputum and bronchoalveolar fluid of a hospital patient (Adékambi et al., 2004). One clone, represented by HBO60, was grouped with the *Nitrospira* (Table 1). Its phylotype was related (93.6% identity) to *Thermodesulfovibrio* sp. TGL-LS1 (AB021302) (Fig. 2), a Gram-negative thermophilic sulfate-reducing bacterial clone recovered from thermophilic anaerobic wastewater treatment processes. One clone, represented by the single sequence type HBO61, clustered with the *Betaproteobacteria* (Table 1). Its rDNA phylotype was 100% identical to a previously determined rDNA sequence from *Burkholderia* sp. SAPII (AF052387) (Fig. 1), a bacterial clone isolated from the hepatopancreatic symbionts in the freshwater isopod, *Asellus aquaticus*.

**Discussion**

Representation of the domain Bacteria included 68 sequence types distributing to seven bacterial divisions. Some microorganisms in this study are identical to those previously
reported from other oilfields, while others are different (Orphan et al., 2000; Bonch-Osmolovskaya et al., 2003; Grabowski et al., 2005). The differences are probably due to the highly heterogeneous geological and physical conditions of the petroleum reservoir studied, which may also select for physiologically diverse assemblages of microorganisms.

Nine types clustering within known thermophilic groups included the low-G+C Gram-positive Thermotogales and Nitrospira, which have previously been isolated from a number of high-temperature petroleum reservoirs worldwide (Orphan et al., 2000; Bonch-Osmolovskaya et al., 2003), suggesting that these thermophiles may be a common component of geothermally heated specialized subsurface environments and probably play a role in the trophic web of these ecosystems. In addition, eight sequence types were associated with clones previously screened from seven

---

**Table 2. Closest relatives isolated from the oilfield**

<table>
<thead>
<tr>
<th>Type (accession no.)</th>
<th>Clone no.</th>
<th>Phylogenetically closest related organism (accession no.)</th>
<th>Similarity (%)</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBO68 (DQ223050)</td>
<td>1</td>
<td>Thermoanaerobacter sp. MET-G (AY800104)</td>
<td>96</td>
<td>Oilfield in France</td>
</tr>
<tr>
<td>HBO65 (DQ223047)</td>
<td>3</td>
<td>Thermoanaerobacter subterraneus SL9 (AY216597)</td>
<td>97.6</td>
<td>Oilfield in France</td>
</tr>
<tr>
<td>HBO69 (DQ223052)</td>
<td>20</td>
<td>Thermotoga hypogea (T) SEBR 7054 (U89768)</td>
<td>93.6</td>
<td>Oilfield in Africa</td>
</tr>
<tr>
<td>HBO43 (DQ201246)</td>
<td>1</td>
<td>Uncultured bacterium clone PL-14B3 (AY570598)</td>
<td>98</td>
<td>Oilfield in Canada</td>
</tr>
<tr>
<td>HBO38 (DQ201241)</td>
<td>1</td>
<td>Uncultured bacterium clone PL-28B12 (AY570614)</td>
<td>98</td>
<td>Oilfield in Canada</td>
</tr>
<tr>
<td>HBO3 (DQ20120)</td>
<td>1</td>
<td>Uncultured bacterium clone PL-5S5 (AY570559)</td>
<td>98</td>
<td>Oilfield in Canada</td>
</tr>
<tr>
<td>HBO50 (DQ201215)</td>
<td>3</td>
<td>Uncultured bacterium O1 (AY770933)</td>
<td>98.4</td>
<td>Oilfield in China</td>
</tr>
<tr>
<td>HBO28 (DQ201232)</td>
<td>1</td>
<td>Thermovenabulum sp.</td>
<td>96.6</td>
<td>High-temperature oil reservoir in China</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metagenomic clone B8-otu12 (DQ097677)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Percentage of 16S rRNA gene similarity to its closest relative.

---

**Fig. 2.** Phyllogenetic tree of the Thermotogales, low-G+C Gram-positive, high-G+C Gram-positive and Nitrospira 16S rRNA gene phylotypes of HBO sequence types (shown in bold) and closely related sequences from the GenBank database. Putative divisions are listed on the right. Bootstrap values (1000 replicates) of ≥50% are reported as percentages. The topology shown was that obtained with the neighbor-joining method. The scale bar represents the number of changes per nucleotide position. Thermococcus litoralis (Z70252) and Methanothermobacter thermoautotrophicum (X05482) were used as the outgroups.
oilfields across the world. The wide spread of these eight types within different petroleum reservoirs indicates that they may be an indigenous bacteria to petroleum reservoirs and may have a significant impact on oil reservoir geochemistry.

Many sequence types were closely related to the Pseudomonas-like clones found in other oilfields (Orphan et al., 2000). These phylotypes are not likely to be derived from thermophilic microorganisms, and may be representative of mesophilic microorganisms, although this petroleum system is characterized by a high-temperature. The water flooding of our petroleum reservoir has been continuous for 17 years, which implies that it is an open system. The injection water was recycled from the water produced from the reservoir and was not sterile during this operation; therefore a number of microorganisms originally present in the surface environment may have been introduced into the reservoir along with the re-injected water. Some bacterial cells will not lyse in the formation water and some microorganisms may remain in the cooler portions of the reservoir, such as in the bottoms of injection wells, or along the wall of production wells (Orphan et al., 2000).

Molecular analysis is a useful tool in any investigation of the microbial ecosystems of oil pools, since the species involved in this environment cannot be easily isolated using conventional methods. However, a previous study indicated that the intensive use of molecular techniques would involve in this environment cannot be easily isolated using conventional methods. However, a previous study indicated that the intensive use of molecular techniques would introduce biases (Suzuki & Giovannoni, 1996; von Wintzen-gerode et al., 1997). Any deficiency in the DNA extraction step may result in members of the Gram-positive bacteria accounting for a smaller fraction, which is probably the reason for the small fraction of high-G+C Gram-positive, low-G+C Gram-positive and Thermotogales groups in our study. In addition, a number of mesophilic microorganisms such as the Pseudomonas-like clones in the production water did not seem to be consistent with the high temperature and extreme conditions present, which implies that our current knowledge of the microbial diversity in such an ecosystem is limited and further study is still necessary. Nevertheless, an analysis of the rRNA gene sequences we retrieved reveals an insight into the bacterial community of high-temperature, long-term water-flooded petroleum reservoirs, and our results will contribute to the promotion of applications in microbially-enhanced oil recovery.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (50374038, 50574040), the Ministry of Education of China (20030251002, 03071) and the Shanghai Municipal Science and Technology Commission (045407017).

References


Nazina TN, Tourova TP, Poltaraus AB, Novikova EV, Grigoriyan AA, Ivanova AE, Petrovyanka VV, Osipov GA, Belyaev SS & Ivanov MV (2001) Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermooleivorans*, *Bacillus kaustophilus*, *Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stea*rothermophilus*, *G. thermocatenulatus*, *G. thermooleivorans*, *G. kaustophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans*. *Int J Syst Evol Microbiol* **51**: 433–446.


