Thermodynamically Favorable Reactions Shape the Archaeal Community Affecting Bacterial Community Assembly in Oil Reservoirs

Xiao-Lei Wu (✉️ xiaolei_wu@pku.edu.cn)
Peking University

Jie-Yu Zhao
Peking University

Bing Hu
Beijing Institute of Technology College of Chemistry and Chemical Engineering

Yan Li
Jiangsu University

Yue-Qin Tang
Sichuan University

Yiming Jiang
Helmholtz Zentrum Munchen Institute of Bioinformatics and Systems Biology

Chang-Qiao Chi
Peking University College of Engineering

Yong Nie
Peking University College of Engineering

Research

Keywords: Microbial community assembly, oil reservoirs, thermodynamic constraint, core archaea, microbiome

DOI: https://doi.org/10.21203/rs.3.rs-41490/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background: Microbial communities exist everywhere on the earth, and play essential roles in biogeochemical cycling in all ecosystems. Understanding microbial community assembly mechanisms could improve our ability to manage microbial ecosystems for industrial, pharmaceutical, and agricultural applications. Previous studies have shown that microbial communities are shaped by deterministic or stochastic processes, but how these two processes influence the microbial community together is rarely investigated, especially in deep terrestrial ecology.

Methods: Here, the microbial compositions in the production waters collected from water injection wells and oil production wells across eight oil reservoirs throughout northern China were determined using high-throughput 454 pyrosequencing of 16S rRNA genes, and analyzed by proportional distribution analysis and null model analysis.

Results: In the study, a ‘core’ microbiota consisting of three bacterial genera and eight archaeal genera were found to be existent in all production water samples. Canonical correlation analysis reflected that these core archaea were significantly influenced by the abiotic factors of temperature and reservoir depth, while the core bacteria were affected by the combined impact of the core archaea and environmental factors. Considering that two of the core archaeal genera, acetoclastic methanogens and hydrogenotrophic methanogens, were enriched in low- and high-temperature oil reservoirs, respectively, it was proposed that the archaeal communities in oil reservoirs were characterized by thermodynamic constraints.

Conclusion: Together, our study indicates that microbial community structures in wells of oil reservoirs are originally determined by the thermodynamic conditions, through which the core archaeal communities are shaped directly followed by the deterministic recruiting of core bacterial genera, and then the stochastically selection of some other microbial members from local environments.

Introduction

Microorganisms are present in nearly all habitats on the earth and play fundamental roles in global biogeochemical cycles, thus shaping the entire environment of the planet [1]. Community assembly mechanisms are essential for understanding the ecological functions of microorganisms in biogeochemical cycling in Earth’s ecosystems [2–4]. Both deterministic and stochastic processes play important roles in community assembly [5, 6]. The deterministic processes involve niche-based mechanisms, including microbial interactions and environmental filtering, while the stochastic processes include random births, deaths, colonization, extinction, and speciation [7]. However, how these two processes shape the microbial community together is still unclear.

Several recent studies have reported that the taxonomic compositions of microbial communities are inconsistent within a given biome on a global scale, but the compositions of protein-coding genes were more consistent across similar biomes. [8–10]. The profiles of oxidoreductase genes better elucidated the
associations of the microbial communities with their respective biomes, which suggested that redox processes separate microbial communities into their corresponding biomes [11].

The functional microbial structures of communities are shaped by physical-chemical factors [11] and can be explained by mechanistic physics-based models [12]. The general laws of thermodynamics and stoichiometry play important roles in the assembly of ecosystems through chemical reactions catalyzed by enzymes of organisms [12]. For example, acetoclastic methanogenic archaea was predominant in the anaerobic microbial communities under high CO$_2$ pressure, because acetoclastic methanogenesis performed by acetoclastic methanogens was thermodynamically easier than hydrogenotrophic methanogenesis, which was catalyzed by hydrogenotrophic methanogens, under this condition [13]. However, how thermodynamic laws drive the assembly of microbial communities remains unknown.

Modeling the intrinsic mechanism of microbial community assembly in complex natural ecosystems such as soils and oceans, with their multilevel, multifactor, and multivariable characteristics, is a considerable challenge. The use of petroleum reservoirs as the objective to learn about the microbial community assembly mechanism seems to be potential. Firstly, microbial ecosystems in petroleum reservoirs are seldomly influenced by exogenous abiotic and biotic variables due to their space independence, making them one important type of natural ecosystems on the earth. Secondly, the carbon sources available to microorganisms in petroleum reservoirs were limited to hydrocarbons with different molecular weights, making the microbial members in the ecosystems either oil degraders or companions of oil degraders, such as the ones with the function of the methanogenic crude oil degradation [14, 15]. Dolfing et al. reported that the methanogenic degradation of crude oil hydrocarbons was an important biogeochemical process in petroleum reservoirs, which was roughly divided into three stages with relatively clear thermodynamics parameters: hydrocarbon oxidation, syntrophic acetate oxidation, and methanogenesis [16, 17]. The microbiota compositions in petroleum reservoirs have been investigated in Asia, Europe, Africa, North America, and South America [18–21], but most previous studies have focused on small amounts of petroleum reservoirs with special properties. The lack of a comprehensive understanding of the microbiota in different petroleum reservoirs interferes with the determination of the pattern of microbiota composition.

On the basis of these earlier insights, we hypothesize in our study that the assembly of the microbial community is shaped by thermodynamically favorable reactions. Here, we sampled production and injection water samples from different petroleum reservoirs with various geochemical conditions in China and analyzed their microbial compositions using high-throughput 454 pyrosequencing of 16S rRNA genes. Our results showed that the functional archaeal distribution was in accord with the predicted thermodynamic constraints, indicating that the operation of an intrinsic mechanism likely leads toward a thermodynamically favored state in the assembly of the microbial community.

**Results**
Spatial variations in physical and geochemical conditions in petroleum reservoirs

In this study, a total of 50 water samples were collected for microbial community detection from eight injection wells and 42 production wells in blocks of eight oilfield in northern China, including XJT (T), XJQ (Q), HBM (M), HBB (B), DQN (N), DGD (D), SLZ (Z), and LHA (A). The locations of the eight oilfields and the geochemical conditions in each water-flooded petroleum reservoir at the sampling time are shown in Fig. 1.

Principal component analysis (PCA) based on environmental parameters showed that geochemical conditions in petroleum reservoirs varied along with spatial variations (Fig. S1). For instance, the geochemical conditions of XJT and XJQ in Northwest China were significantly different from those of DGD and SLZ in North China. Furthermore, Spearman’s rank correlation test showed that temperature and depth exhibited a significantly positive correlation \( (\rho = 0.851, P = 3.444 \times 10^{-15}) \), while TOC and pH showed a significantly negative correlation \( (\rho = -0.807, P = 9.8 \times 10^{-9}) \) (Table S1).

The composition of the microbial communities varies geographically

Pyrosequencing analysis of the microbial communities in the 50 water samples identified 246,497 and 144,194 optimized bacterial and archaeal sequences, respectively. These sequences were arranged into 6,174 bacterial and 1,727 archaeal OTUs. Good’s coverage indices for most samples were found to be over 90% (Table S2). Remarkably, the bacterial communities exhibited higher alpha diversity than the archaeal communities, as indicated by the Shannon-Weiner and Simpson indices \( (P = 1.391 \times 10^{-6} \) and \( P = 4.514 \times 10^{-6} \), Table S2).

Taxonomically, the microbial community structures at the phylum level were similar in different petroleum reservoirs. As shown in Fig S2, Proteobacteria (45.41–100%) and Euryarchaeota (69.93–100%) were the most abundant bacterial and archaeal phyla, respectively, across all samples, and Bacteroidetes, Actinobacteria, Firmicutes, Caldiserica, Crenarchaeota, and Thaumarchaeota were detected in most of the petroleum reservoirs. However, predominance at the class level varied geographically. Epsilonproteobacteria and Methanomicrobia were dominant in XJT, DQN, and several HBM production wells, while Gammaproteobacteria and Methanobacteria were dominant in XJQ, DGD, and HBB, and Alphaproteobacteria was only predominant in LHA (Fig. S2).

Core archaea as the fundamental participants in the microbial communities of petroleum reservoirs

Based on their frequency to be identified in the samples of one oilfield, the microbes in each of the eight oil reservoir were grouped into block-level core (OTUs or genera present in > 70% of the samples in one block), block-level variable (OTUs or genera present in at least two but < 70% of the samples in one block),
and block-level unique (OTUs or genera only present in a unique sample) microbes (Table S3 and Fig S3).

As shown in Table S4, the block-level core and variable microbes showed greater relative abundances than the block-level unique microbes in every oil reservoir. The average proportions of the block-level core and variable microbes at the OUT level and the genus level for all blocks were over 72.19% and 88.26%, respectively. As summarized in Fig. 2, some block-level core bacterial genera, including *Pseudomonas*, *Acinetobacter*, and *Arcobacter*, and some block-level archaeal genera, including *Archaeoglobus, Methanobacterium, Methanothermobacter*, unclassified genera affiliated with *Methanobacteriaceae, Methanoculleus, Methanolinea*, and *Methanosaeta*, were existent in all of the eight blocks, and they were classed as the ‘core microbial genera’ in the study. The analysis of the relative abundances of core bacterial and archaeal genera in all blocks showed that *Pseudomonas, unclassified Methanobacteriaceae, and Methanothermobacter* were dominant in XJQ, DGD, and HBB blocks, where the temperature at the sampling sites was moderate-to-high. In contrast, *Arcobacter, Methanobacterium, Methanosaeta, Methanoculleus*, and *Methanolinea* were predominant in XJT, HBM, and DQN blocks, where the temperature was low-to-moderate (Fig. S3). These results indicated that environmental variables should have tight relationship with the core microbial communities in the oilfield blocks.

**Abiotic and biotic factors gradually and consistently shape the microbiota**

To evaluate the interactions between the oilfield-derived microbial communities and the abiotic and biotic factors, we conducted CCA analysis. During the analysis, the core bacterial and archaeal genera were setup as the key biotic factors considering of their high relative abundances in all samples. As shown in Table S5, temperature, depth, TOC, pH, AH, and ASP significantly influenced the microbial community structures (*P* < 0.05), four core archaeal genera (*Archaeoglobus, Methanothermobacter, unclassified Methanobacteriaceae, and Methanosaeta*) significantly influenced the bacterial microbial community structures (*P* < 0.05), while no core bacterial genera showed a significant influence on the archaeal communities.

When we compared the explanation degrees of abiotic and biotic variables to the bacterial and archaeal communities using both total and constrained variance, it was found that only 38.28% of the bacterial community was explained by abiotic factors, while 58.28% of the archaeal community was explained by abiotic factors. Moreover, the core archaeal genera explained 31.96% of the bacterial community, while the core bacterial genera explained only 16.94% of the archaeal community. The integration of abiotic and biotic influences showed that 67.45% of the bacterial communities were influenced by the integrated factors, which was much higher than the explanation degrees of either abiotic or biotic variables. In contrast to the bacterial communities, the explanation degrees of the integrated factors to the archaeal communities were similar with those of the individual abiotic variables (Fig. S4). These results indicated that archaeal community structures should be mainly determined by the abiotic variables in petroleum reservoirs, while the bacterial communities were influenced by the integrated abiotic variables and the core archaeal genera.
To analyze association between the abiotic or biotic variables and the abundances of microbial members in the water samples, Spearman’s correlation coefficients were calculated and visualized in Fig. 3. It was found that abiotic factors, especially temperature and depth, showed close connections with the surrounding core archaea and provided favorable conditions for the growth of various microbes, indicating that the abiotic factors of temperature and depth and the biotic factors of core archaeal microbiota were the key nodes in the microbial communities of oil reservoirs. In the four blocks with low-to-moderate temperatures (XJT, XJQ, HBM, and DQN), the core archaeal genera with relatively high abundances, including *Methanosaeta*, *Methanolinea*, *Methanoculleus*, and *Methanobacterium*, were negatively correlated with temperature and depth. In the four blocks with moderate-to-high temperatures (DGD, HBB, SLZ, and LHA), the core archaea being predominant in the core microbiota, including *Methanothermobacter* and unclassified *Methanobacteriaceae*, showed positively correlated with temperature and depth. For the remaining microbes in the correlation networks, they were directly or indirectly connected with temperature, depth, and the core archaea. For example, in XJT and HBM, the variable bacterial genera of *Geobacter*, *Smithella*, *Arcobacter*, *Hydrogenophaga*, *Azonexus*, *Thauera*, and *Erysipelothrix* showed significantly positive correlations with the core archaeal genera of *Methanoculleus*, *Methanolinea*, *Methanobacterium*, and *Methanosaeta*; In DGD, SLZ and LHA, the core bacterial genus of *Pseudomonas* and the variable bacterial genera of *Brevundimonas*, *Sphingomonas*, and *Sphingobium* showed significantly positive correlations with the presence of *Methanothermobacter*. These findings directed our exploration of the mechanism responsible for microbial community assembly in petroleum reservoirs.

**Deterministic processes influence microbial community assembly in petroleum reservoirs**

We analyzed the microbial community structure dissimilarities of water samples derived from different blocks using Bray-Curtis and Jaccard distances. As shown in Table S6, the compositions of the microbiota in the same block were more similar than those identified in different blocks, indicating that different blocks contained microbes that were specifically adapted to those blocks. We next conducted a null model analysis to assess whether the microbiota assembly was characterized by a deterministic or stochastic process. The permutational analysis of multivariate dispersions (PERMDISP) showed that the observed bacterial and archaeal β-diversities in each of the eight blocks were significantly different from the null random expectations ($P < 0.05$), suggesting that microbial community assemblage in the sampled petroleum reservoirs should be in a deterministic manner (Table S7). To measure the importance of deterministic processes for the assembly of archaeal and bacterial communities in petroleum reservoirs, we then analyzed the proportions of the deterministic processes to the archaeal and bacterial communities, respectively. As shown in Fig. 4, the archaeal communities exhibited higher proportions of deterministic processes (97.08 ± 6.70%) than the bacterial communities (78.75 ± 18.45%). The results indicated that abiotic factors exert a greater influence on the archaeal communities of petroleum reservoirs than the bacterial communities, which was in accordance with the CCA analysis results in Fig. S4.
Thermodynamic constraints on core archaeal microbiota

To determine the underlying mechanism by which the abiotic factors, such as temperature and depth, influence the archaeal communities and then the bacterial communities in the sampled petroleum reservoirs, the thermodynamic constraints of methanogenic hydrocarbon degradation within the Archaea domain under various temperatures were calculated (Fig. 5). Theoretically, when the temperature was below 140 °C, both hydrocarbon oxidation and acetate oxidation were endergonic and the ΔG values for each reaction decreased with increasing temperatures, regardless of the presence of saturated hydrocarbons (C_{16}H_{34} and C_{100}H_{202}) or aromatic hydrocarbons (benzene C_{6}H_{6}). When the temperature was higher than 60 °C, acetate oxidation was more and more favorable than acetate production in the incomplete hydrocarbon oxidation reactions, indicating that relative potential of acetate consumption to produce H_{2} and CO_{2} increased with the increased temperature when the biosystem was exposed to hydrocarbons. When the temperature was higher than 40 °C, the ΔG values of acetoclastic methanogenesis increased rapidly with increasing temperature, whereas hydrogenotrophic methanogenesis was less sensitive to temperature increase. In general, acetoclastic methanogenesis was inferred to be more favorable at low temperature, and hydrogenotrophic methanogenesis and syntrophic acetate oxidation were more favorable at high temperature. This inference was in agreement with the fact that the relative abundances of hydrogenotrophic methanogens (HMAs) in our 50 samples of the eight blocks, including the archaeal genera of *Methanothermobacter*, unclassified *Methanobacteriaceae*, *Methanolinea*, and *Methanobacterium*, exhibited a significant positive correlation with temperature, whereas the relative abundances of acetoclastic methanogen (AMA) *Methanosaeta* and hydrogenotrophic methanogens requiring acetate (HAMA) *Methanoculleus* exhibited a significantly negative correlation with temperature (Fig. 6). Thus, it was considered that the abiotic factors determined the archaeal community structures in petroleum reservoirs by the thermodynamic constraint-derived recruitment of the core archaeal strains whose metabolism were thermodynamically favorable under special conditions.

Discussion

Abiotic factors drive core archaeal microbiota in petroleum reservoirs

In this study, we found that the microbiota of Chinese petroleum reservoirs was shaped by deterministic factors, including geochemical conditions (temperature, depth, TOC, pH, AH, and ASP) and species interactions. Our results showed that 58.28% of the archaeal community was influenced by abiotic factors. This is a much higher percentage than that of the bacterial community influenced by abiotic factors. Here, we show that abiotic factors, particularly temperature and depth, mainly influence the core archaea by regulating the metabolic steps required for the methanogenetic degradation of hydrocarbons, a key process to maintain the suitable, optimal, and sustainable living habitats in the subsurface ecosystems [14]. Among the core archaea, *Methanosaeta*, *Methanolinea*, *Methanoculleus*, and
Methanobacterium showed a relatively high abundance at low-to-moderate temperatures and were negatively correlated with temperature and depth. This result is in accordance with previous findings. It was reported that AMA Methanoseta was present in production wells in Canada and China at low temperatures [20, 28]. Some Methanoseta harundinacea strains could utilize acetate exclusively for growth and methane production, and the cells proliferated when the temperature was at 25–45 °C with the optimum temperature as 34–37 °C [29]. HAMA Methanoculleus has been detected in the Enermark Oilfield, Canada [30], and the optimum growth temperature of Methanoculleus chikugoensis strain MG62(T) was 25–30 °C [31]. HMA Methanolinea has been detected in production wells in China at low temperatures [20], and the isolated Methanolinea mesophila strain TNR$^T$ could grow at 20–40 °C, with the optimum growth temperature of 37 °C [32]. HMA Methanobacterium has been detected in production wells in petroleum reservoirs in Canada at low temperatures [30], and the isolated Methanobacterium strain Mic5c12(T) and Mic6c05(T) exhibited the optimum growth temperatures of 35 °C and 40 °C, respectively [33]. Conversely, the abundance of the core archaea of Methanothermobacter and unclassified Methanobacteriaceae, which was relatively high at moderate-to-high temperatures, was positively correlated with temperature and depth. It is well known that Methanothermobacter survives in high-temperature petroleum reservoirs [20, 21]. Methanothermobacter tenebrarum strain RMAS(T), a hydrogenotrophic and thermophilic methanogen, could proliferate at a temperature range of 45–80 °C (optimum: 70 °C) [34].

Microbial correlations contribute to complex microbial networks

In this study, the null model analysis showed that microbial community assembly was shaped by deterministic factors, including geochemical conditions and species interactions. Abiotic factors and core archaea sustain the network of microbiota in petroleum reservoirs.

In petroleum reservoirs, microbial members may form syntrophic relationships, in which organisms positively, neutrally, or negatively affect their partners, either by direct contact or by the cross-feeding of metabolites, resulting in mutualistic, commensal, or parasitical interactions [35, 36]. Previous studies have shown that Pseudomonas, Brevundimonas, Sphingomonas, Sphingobium, Hydrogenophaga, and Thauera, which are hydrocarbon-oxidizing microorganisms, initiate hydrocarbon degradation in petroleum reservoirs [37]. After the initial oxidation, subsequent products such as alcohols, aldehydes, and fatty acids are excreted into the surrounding environment, affecting other microbes. Subsequently, fermentative bacteria (such as Geobacter and Smithella) take up and reduce compounds such as amino acids, sugars, long-chain fatty acids, lactate, butyrate, and propionate for the production of H$_2$, CO$_2$, and acetate [37]. Methanogens (such as Methanobacterium, Methanothermobacter, Methanolinea, and Methanoseta) then take up H$_2$, CO$_2$, acetate, and other small organic acids, resulting in the production of CH$_4$ [38]. Second, numerous microbial species may play similar roles in a single petroleum reservoir because crude oil is an important source of abundant carbon compounds. Our results showed that Thauera, Azonexus, Zoogloea, and Hydrogenophaga, affiliated with Betaproteobacteria, were positively correlated with Methanoseta and Methanolinea and coexisted in the HBM and DQN blocks.
Sphingomonas, Novosphingobium, Brevundimonas, and Rhizobium, affiliated with Alphaproteobacteria, were positively correlated with Methanothermobacter and coexisted in the DGD, SLZ, and LHA blocks. The existence of diverse microbes with similar functions may increase the biodiversity of the microbiota. Based on the diversity-stability theory, high-diversity systems tend to be robust and resilient [39, 40].

The dynamic pattern of the ecological relationships between different microbes includes symbiosis, competition, and parasitism [41]. Microbes show a trade-off between growth, reproduction, restoration, and resistance under conditions of limited availability of energy sources. Furthermore, microbes compete and cooperate to maintain growth and obtain energy. Our results indicate that abiotic factors, particularly temperature, depth, and the presence of the core archaea, lay the foundation for the microbial network. Furthermore, abundant microbial interactions enrich the microbial network and increase the stability and elasticity of the energy flow in petroleum reservoirs.

**Feasibility and challenge of thermodynamic constraints on microbial community assembly**

Our results of null model analysis indicated that the microbiota of the petroleum reservoirs was shaped by deterministic processes. Abiotic factors, particularly temperature and depth, shaped the core archaeal microbiota, which are among the most important participants in the metabolic steps of the methanogenic degradation of petroleum hydrocarbons. Abiotic factors and the core archaeal microbiota then lay the foundation for the microbial food web.

A petroleum reservoir is a chemosynthetically driven ecosystem characterized by a relatively simple carbon source and stable geochemical conditions compared with those in the soil and ocean [4, 42]. The methanogenic degradation of crude oil hydrocarbons shows relatively clear thermodynamic constraints [16, 17]. In our study, the change in the Gibbs free energy, based on thermodynamic reactions, indicated that HMAs preferred high-temperature petroleum reservoirs, whereas AMAs and HAMA preferred low-temperature petroleum reservoirs. This result was consistent with the abundances of the functional archaea determined across our samples using 16S rRNA sequencing (Fig. 5, Fig. 6). This result is in agreement with previous findings showing that HMAs are transformed to AMAs under high CO$_2$ concentrations, which is more thermodynamically favorable [13]. Our results suggest that microbial community assembly is thermodynamically driven.

The elucidation of the intrinsic assembly mechanism is a central goal in ecology. However, complex multifactor, multivariate, and multilevel ecosystems mask this underlying mechanism. The cryosphere of the Earth is a simple ecosystem that includes sea ice, glacial, and subglacial habitats, with a reduced content of organic matter inhabited by diverse organisms. In sea ice and supraglacial habitats, sunlight penetrates the ice. Photoautotrophy, mostly contributed by sea-ice diatoms, serves as the basis for complex food webs, whereas in subglacial habitats, chemoautotrophy is observed [43, 44]. The overwhelming majority of microbes in these environments are difficult to culture and model. Therefore, we encountered obstacles in trying to determine the energy flow of these microbes. For these reasons,
assessing microbial community assembly based on thermodynamic constraints remains a key challenge in microbial ecology.

**Conclusion**

To date, little is known about the underlying mechanism of microbial community assembly. In this study, the bacterial and archaeal communities in eight Chinese petroleum reservoirs were comprehensively investigated under heterogeneous geochemical conditions to identify the intrinsic mechanism of microbial community assembly. In the petroleum reservoirs, abiotic factors (temperature, depth, TOC, pH, AH, and ASP) significantly influenced the microbiota. Abiotic factors influenced the archaeal communities more than the bacterial communities. The combination of the core archaea and abiotic factors, especially temperature and depth, lay the foundation for the microbial community structures in Chinese petroleum reservoirs, indicating that deterministic processes shape the microbial community. The distribution of the core archaea was predicted according to the thermodynamic constraints controlled by the abiotic factors, indicating that microbial community assembly tended toward a thermodynamically favored state. These findings will lead to new horizons for exploring the assembly mechanisms of various ecological habitats.

**Materials And Methods**

**Sampling and analysis of the physico-geochemical characteristics of the water samples**

From 2009 to 2010, a total of 50 water samples collected from eight injection wells and 42 production wells of eight water-flooding Chinese oil reservoirs were collected. The eight sampling blocks were XJT (T) and XJQ (Q) in Xinjiang Province, HBM (M) and HBB (B) in Inner Mongolia Province, DQN (N) in Heilongjiang Province, LHA (A) in Jilin Province, DGD (D) in Tianjin Province and SLZ (Z) in Shandong Province. Physical characteristics of the petroleum reservoirs, such as permeability (PER), porosity (POR), temperature, and depth, were recorded. The distribution of the sampling sites and detailed information on the samples are shown in Fig. 1 and Table S2. All samples were immediately transported to the laboratory, centrifuged at 500 × g to discard the upper-phase oil and the precipitates, and then stored at -80 °C for further analysis. Upon arrival, the chemical indices of floating crude oil in the production mixture, including sulfur (S), nitrogen (NI) and the contents of four subfractions (wt/wt) (saturated hydrocarbons (SH), aromatic hydrocarbons (AH), nonhydrocarbons (NH), and asphaltene (ASP)) were determined. The chemical characteristics of the injection water and production water, such as the pH, total organic carbon (TOC), and total salinity degree (TSD), which included anions (Cl⁻, NO₃⁻, and SO₄²⁻) and cations (Na⁺, K⁺, Mg²⁺, and Ca²⁺), were also analyzed. All analyses were performed according to the methods of Tang et al. [20].

**Microbe collection and DNA extraction**
For the microbiota composition analysis, microbes from the 50 water samples were collected by filtering approximately 200–250 ml samples through 0.22-µm hydrophilic membrane filters (Ф47 mm, Millipore, USA). DNA was extracted using the FastDNA® Spin Kit for Soil (MP Biomedicals, Cleveland, USA) according to the manufacturer’s instructions and was quantified using a spectrophotometer (Pharmaspec UV-1700, Shimadzu, Kyoto, Japan).

**Sequencing library construction and pyrosequencing**

For bacterial community analysis, the ~ 192 bp V3 region of the 16S rRNA gene was amplified with a bacterial universal primer pair 341F/533R [22]. For archaeal community analysis, nested PCR was performed to amplify the V3-V6 region (approximately 709 bp). The first primer pair was 109F/1386R, followed by 339F/1048R [23]. The primers 341F, 533R, and 339F were linked with 454 Life Sciences adaptor sequences, a unique 10-bp error-correcting Golay barcode, and a ‘T’ linker sequence that was inserted between the barcode and primer, while primer 1048R only harbored the adapter sequence. The primers, pipetting conditions, and thermal conditions for bacterial and archaeal PCR amplification used in this study are listed in Tables S2, S8, and S9. Three biological replicates were performed for each experiment. All amplicons were visualized using 2% (w/v) agarose gels (TaKaRa, Japan) and purified using a DNA gel extraction kit (BioTek, China). Purified amplicons of the same sample were mixed in equimolar ratios, followed by sequencing on the 454 Life Sciences Genome Sequencer FLX Titanium platform (Shanghai Majorbio Bio-Pharm Technology Co., Ltd. and TEDA Institute of Biological Sciences and Biotechnology, Nankai University, China).

**Processing of pyrosequencing data and statistics**

For bacterial community analysis, valid sequences were processed using Mothur 1.10.2. The sequences were optimized for the preferences of both the forward and reverse primers, allowing up to two mismatches, with perfect primer and barcode sequence matching, no ambiguous base calls, a 100-bp minimum sequence length, and more than 30 sequences. For archaeal community analysis, raw sequences were processed using QIIME version 1.7.0 to obtain optimized sequences, which removed ≤ 200 bp and ≥ 1000 bp sequences, ambiguous bases exceeding 6 bp, a missing qual score or a mean qual score below 25, max homopolymer runs exceeding 6 bp, no mismatches between primers, and uncorrected barcodes. All optimized sequences were compared by using BLASTN with a reference database based on the SILVA database (version 106) and then clustered into operational taxonomic units (OTUs). OTUs were defined according to a farthest neighbor Jukes-Cantor distance of 0.03 and assigned taxonomically using the Ribosomal Database Project (RDP) Naive Bayes classifier. The Shannon-Wiener, Simpson, Chao1, and Good’s coverage indices were estimated to determine the diversity of the bacterial and archaeal communities. The sff files were deposited at DDBJ (http://www.ddbj.nig.ac.jp) under the accession numbers listed in Table S2.

Statistical analyses were performed using R software (v.3.4.2, R Project for Statistical Computing). To analyze the heterogeneity of the geochemical conditions, including PER, POR, temperature, depth, TSD, TOC, pH, S, NI, SH, AH, NH, and ASP, in different petroleum reservoirs, principal component analysis (PCA)
was performed on a geochemical data matrix using the `pca2d` function in the `pca3d` package of R. The `cor` and `cor.test` functions in the stats package of R were used to calculate and test Spearman's rank correlations between geochemical conditions. A Spearman's correlation coefficient (\( \rho \)) of \( \geq 0.4 \) and \( P < 0.05 \) were considered statistically significant [24].

Based on the frequency of the occurrence of OTUs or genera, the microbes in a single block were divided into three groups: (I) OTUs or genera present in >70% of production water samples in one block were grouped as core OTUs or genera; (II) OTUs or genera present in at least two but <70% of production water samples in one block were grouped as variable OTUs or genera; and (III) OTUs or genera present only in one production water sample were grouped as unique OTUs or genera [25]. Using a similar calculation method, the core, variable, and unique microbes in all blocks were summarized. A Venn diagram based on OTU levels was used to compare the core and variable microbes in the injection and production water samples.

The abiotic and biotic influences on the microbiota were tested using the `envfit` function in the `vegan` package of R based on canonical correlation analysis (CCA) [26]. \( P < 0.05 \) based on 999 permutations indicated the significance of geochemical conditions. To more specifically identify the relationships between geochemical conditions and among core and variable microbes in different blocks, the \( \rho \) and \( P \) values of Spearman's rank correlations were calculated. All possible Spearman's rank correlations (\( |\rho| \geq 0.4, P < 0.05 \)) were considered valid influencing events, and the topology of the network was described in Cytoscape 3.2.1. The nodes in the reconstructed network represent the geochemical conditions and core-variable genera, whereas the lines correspond to a strong and significant correlation between nodes. Path lines document only the positive and negative correlations, regardless of connectivity (\( \rho \)). Node size roughly describes the relative abundance of core microbes.

\( \beta \)-diversity (Bray-Curtis dissimilarity and Jaccard's dissimilarity) represents compositional variations among communities across various spatial scales and provides insights into mechanisms of community assembly [7, 27]. We calculated the dissimilarity (1-similarity) of the microbiota based on OTU levels using the `vegdist` function in the `vegan` package of R. In the null mode analysis, microbiota assembly via a stochastic process, in which the microbial composition is not regulated by abiotic and biotic factors, was set as the null model. Permutational analysis of multivariate dispersions (PERMDISP) was used to discern whether differences in similarity in the same block were different from the null expectation. The proportions of deterministic and stochastic processes provided a quantitative estimation of the extent of niche-based deterministic selection in shaping the community composition and structure [7]. The analysis of the correlations between the microbiota and geochemical conditions was conducted with a multiple regression model using the `lm` function in the stats package of R [4].

**Evaluating the thermodynamics of methanogenic hydrocarbon degradation routes**
The degradation of either saturated hydrocarbons (C_{16}H_{34} and C_{100}H_{202}) or aromatic hydrocarbons (C_{6}H_{6}) is divided into the following three steps.

Step 1 (hydrocarbon oxidation) has three types of reactions:

1. Complete oxidation: \(4C_xH_y + 8xH_2O \rightarrow 4xCO_2 + (8x + 2y)H_2\)
2. Incomplete oxidation to produce acetate: \(4C_xH_y + (4x - y)H_2O + yCO_2 \rightarrow (2x + y/2)CH_3COOH\)
3. Incomplete oxidation to produce acetate and H\(_2\): \(4C_xH_y + 4xH_2O \rightarrow 2xCH_3COOH + 2yH_2\)

Step 2 (syntrophic acetate oxidation):

\((2x + y/2)CH_3COOH + (4x + y)H_2O \rightarrow (4x + y)CO_2 + (8x + 2y)H_2\)

Step 3 (methanogenesis) has two types of reactions:

1. Hydrogenotrophic methanogenesis: \((8x + 2y)H_2 + (2x + y/2)CO_2 \rightarrow (2x + y/2)CH_4 + (4x + y)H_2O\)
2. Acetoclastic methanogenesis: \((2x + y/2)CH_3COOH \rightarrow (2x + y/2)CO_2 + (2x + y/2)CH_4\)

The Gibbs free energy of each chemical reaction was calculated according to the study by Dolfin et al., and temperature corrections for \(\Delta G^0\) were performed using the Gibbs-Helmholtz equation:

\[
\Delta G_{Tact}^0 = \Delta G_{Tref}^0 \cdot \left(\frac{T_{act}}{T_{ref}}\right) + \Delta H_{Tref}^0 \cdot \frac{T_{ref} - T_{act}}{T_{ref}}, \text{ with the temperature in Kelvin; } T_{ref} = 298.15 \text{ K} \ [16, 17].
\]

**Declarations**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Availability of data and material**

The sff files of the sequence data were deposited at DDBJ under the accession numbers listed in Table S2 (https://ddbj.nig.ac.jp/DRASearch/).

**Competing Interests**
The authors declare that they have no conflict of interest.

**Funding**

The study was supported by the National Key R&D Program of China (2018YFA0902100 and 2018YFA0902103 to XLW) and the National Natural Science Foundation of China (31770120 to XLW).

**Authors’ contributions**

XW and YN conceived of the study, and participated in its design and helped to polish the manuscript. JZ, BH, YL, and CC performed the study, including sample collection, DNA extraction, sequencing data analysis, and modeling strategy development. JZ and BH wrote the manuscript. YT and YJ assisted with the statistical analysis and polished the manuscript. All the authors read and approved the final manuscript.

**Acknowledgements**

We like to thank Dr. Yingying Yao at School of Human Settlements and Civil Engineering, Xi’an Jiaotong University for illustration preparation and analysis.

**References**


Figures
Figure 1

Geographical distribution of eight oil reservoir blocks in China according to ArcGIS. Fifty water samples from 8 injection wells and 42 production wells were obtained from eight Chinese oilfields (XJT, XJQ, HBM, DQN, DGD, HBB, SLZ, and LHA). The names of the samples taken from the injection and production wells, as well as the physical and geochemical conditions of samples from each oilfield are listed under the map. *

* PER, Permeability; POR, Porosity; WC, Water content; TSD, Total salinity degree; TOC, Total organic carbon; S, Sulfur; N, Nitrogen; SH, Saturated hydrocarbon; AH, Aromatic hydrocarbon; NH, Non-hydrocarbon; SAP, Asphaltene.
Figure 2

Core microbial communities in petroleum reservoirs. The core bacteria are indicated in pink, variable bacteria in green, and unique bacteria in white (black font), while archaea are indicated in red font in different petroleum reservoirs.
**Figure 3**

Network of interactions based on Spearman’s rank correlation test (P < 0.05, |ρ| ≥ 0.4), among the core-variable genera in each block and under the indicated geochemical conditions. The nodes represent the geochemical conditions (green), core archaea (red), variable archaea (pink), core bacteria (blue), and variable bacteria (lavender). Only core bacteria and core archaea were present in relative abundance. The path lines correspond to a strong and significant correlation between nodes (P < 0.05), shown by positive (red solid line) and negative (blue dashed line) correlations, regardless of connectivity (ρ).

**Figure 4**

Proportions of deterministic and stochastic processes of bacterial and archaeal communities in eight blocks.
Figure 5

Effect of temperature on the changes in Gibbs free energy in hydrocarbon oxidation, acetate oxidation, and methanogenesis. Hydrocarbon oxidation is classified as complete oxidation to H2 and CO2 (black line), incomplete oxidation to CH3COOH alone (red line), and incomplete oxidation to CH3COOH and H2 (green line). The blue line indicates acetate oxidation. Methanogenesis is classified into hydrogenotrophic
methanogenesis (yellow line) and acetotrophic methanogenesis (magenta line). Thermodynamic calculations for different hydrocarbons, such as (a) C16H34, (b) C100H202, and (c) benzene, were evaluated under a temperature range of 0–160°C. Vertical dark lines indicate different temperatures (20 and 80°C) for different reactions and hydrocarbons.

Figure 6

Relationships between functional methanogens (HMAs, AMAs, and HAMA) and temperature analyzed using a multiple regression model. HMAs, Methanothermobacter, other Methanobacteriaceae, Methanolinea, and Methanobacterium; AMAs, Methanosaeta; HAMA, Methanoculleus.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Zhaoetal.SupplementaryTables.xlsx