



Species Divergence vs. Functional Convergence Characterizes Crude Oil Microbial Community Assembly

Yong Nie^{1†}, Jie-Yu Zhao^{1†}, Yue-Qin Tang², Peng Guo¹, Yunfeng Yang³, Xiao-Lei Wu^{1,4*} and Fangqing Zhao^{5*}

¹ College of Engineering, Peking University, Beijing, China, ² College of Architecture and Environment, Sichuan University, Chengdu, China, ³ State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing, China, ⁴ Institute of Engineering (Baotou), College of Engineering, Peking University, Baotou, China, ⁵ Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing, China

OPEN ACCESS

Edited by:

William James Hickey,
University of Wisconsin-Madison,
USA

Reviewed by:

Sabine Kleinstüber,
Helmholtz Centre for Environmental
Research – UFZ, Germany
Chuanwu Xi,
University of Michigan, USA

*Correspondence:

Xiao-Lei Wu
xiaolei_wu@pku.edu.cn
Fangqing Zhao
zhfq@mail.biols.ac.cn

[†] These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Microbiotechnology, Ecotoxicology
and Bioremediation,
a section of the journal
Frontiers in Microbiology

Received: 22 January 2016

Accepted: 28 July 2016

Published: 12 August 2016

Citation:

Nie Y, Zhao J-Y, Tang Y-Q, Guo P,
Yang Y, Wu X-L and Zhao F (2016)
Species Divergence vs. Functional
Convergence Characterizes Crude Oil
Microbial Community Assembly.
Front. Microbiol. 7:1254.
doi: 10.3389/fmicb.2016.01254

Oil reservoirs exhibit extreme environmental conditions such as high salinity and high temperature. Insights into microbial community assemblages in oil reservoirs and their functional potentials are important for understanding biogeochemical cycles in the subterranean biosphere. In this study, we performed shotgun metagenomic sequencing of crude oil samples from two geographically distant oil reservoirs in China, and compared them with all the 948 available environmental metagenomes deposited in IMG database (until October 2013). Although the dominant bacteria and the proportion of hydrogenotrophic and acetoclastic methanogens were different among oil metagenomes, compared with the metagenomes from other environments, all the oil metagenomes contained higher abundances of genes involved in methanogenic hydrocarbon degradation and stress response systems. In addition, a “shape-sorting” manner was proposed for the assembly of microbial communities in oil reservoirs, with the oil reservoir acting as a function sorter to select microbes with special functions from its endemic pool of microorganisms. At the functional level, we found that environmental metagenomes were generally clustered according to their isolation environments but not their geographical locations, suggesting selective processes to be involved in the assembly of microbial communities based on functional gene composition.

Keywords: metagenome, microbial community, oil reservoirs, petroleum, methanogenesis, hydrocarbon degradation

INTRODUCTION

Oil reservoirs, which are located deep within the earth, exhibit extreme environmental conditions such as high pressure, high osmosis (due to high salinity or high hydrophobicity), and high temperature (Head et al., 2003). Culture-dependent and culture-independent analyses based on DNA/RNA fingerprints and high throughput 16S rRNA gene sequencing (Li et al., 2006, 2007, 2014; Dahle et al., 2008; Kaster et al., 2009; Pham et al., 2009; Zhang et al., 2012; Gao et al., 2016) have shown that oil reservoirs harbor various microbes. Because of the extreme environmental conditions in oil reservoirs, it is important to understand the functions and assemblies of microbial communities in this environment and differences between these microbial communities and those in other environments such as freshwater, seawater, and surface soils. To date, knowledge about functional genes of microbial communities in oil reservoirs is limited

despite recent studies on functional genes associated with sulfate reduction, ammonia oxidation, and methanogenesis by analyzing clone libraries (Li et al., 2011; Guan et al., 2013; Liu et al., 2015) and on metagenomes of offshore oil reservoirs of the Norwegian Sea by using shotgun sequencing (Kotlar et al., 2011; Lewin et al., 2014).

Despite the identification of various microbes in oil reservoirs, mechanisms underlying microbial assemblages in these environments remain unclear. Variation in environmental factors such as temperature and pH and biological interactions may influence the microbial communities. The structures of natural microbial communities can be explained by niche-based mechanisms, and differences in the abundance of taxa reflect the differences in the environmental factors. It is supported by numerous evidences that microbial communities change rapidly in response to changes in environmental factors. In contrast, microbial compositions are extremely different in the abundance of rare taxa, sometimes in the high abundant taxa despite similar geochemical conditions in some oil reservoirs (Ren et al., 2011; Kryachko et al., 2012; Tang et al., 2012). These variations in compositions of microbial communities among different oil reservoirs with similar geochemical conditions cannot be explained by environmental factors. It is possible that stochastic processes such as dispersal and colonization play significant roles in structuring microbial communities in these oil reservoirs with similar geochemical conditions. However, this stochastic process cannot explain how oil reservoirs with different geochemical properties show microbial communities with similar taxonomic and functional gene compositions.

Here, we analyzed the crude oil metagenomes from two oil reservoirs in Daqing and Qinghai oil fields in China. These metagenomes were compared with other two crude oil metagenomes from offshore oil reservoirs of Norwegian Sea (Kotlar et al., 2011; Lewin et al., 2014), and 948 environmental metagenomes which were all environmental metagenomes available in IMG database until October 2013, to examine how they differed within and among other environments, and how taxonomic and functional gene compositions were linked to oil reservoirs characteristic conditions.

MATERIALS AND METHODS

Sample Collection

Oil samples were collected from the wellhead of the high temperature Qinghai (QH) oilfield located in the Tibetan Plateau situated in the northwest of China and from the wellhead of the mesophilic Daqing oilfield (DQ) located in the northeast of China (Supplementary Figure S1). The distance between the two sampling sites is more than 2,500 km. Blocks of Qinghai and Daqing oilfield from where the oil samples were obtained have been subjected to water flooding for 15 and >30 years, respectively. Samples were collected directly from sampling valves located on the wellheads before the oil entered production facilities such as treatment station for separating water, oil, gas, and sand (Supplementary Figure S2). To avoid contamination,

sampling cans were washed, sealed, and sterilized by heating at 121°C for 20 min. The sampling valves were also washed and sterilized with 70% alcohol immediately before sampling. During sampling, an initial 20-L sample was discarded before collecting the oil samples in the sterilized cans. Once the cans were completely filled, they were sealed using screw caps, stored at <5°C, and sent to our laboratory within 2 days. Further analyses were performed immediately. Since the wellheads were directly connected to the working layers in reservoirs, the microbial community in the crude oil samples should reflect the microbial compositions in the reservoirs.

DNA Extraction, Sequencing, and Analysis

DNA was extracted from the samples as described previously (Kryachko et al., 2012) and the genomic libraries for Illumina sequencing were generated according to the Illumina paired-end library preparation protocol. The average size of the library fragments was 200 bp for both QH and DQ, while the average of the insert size was 190 bp for QH and DQ. The libraries were sequenced using HiSeq 2000 (Illumina, San Diego, CA, USA) with 2 × 100 bp paired-end sequencing.

For taxonomic analysis, low quality paired-end reads were removed using Illumina CASAVA pipeline with default parameters, and the remaining reads were compared against the NR database by using BLASTX to estimate functional and taxonomic composition and abundance as described before (Wang J. et al., 2013). MEGAN software (Huson et al., 2011) was used to parse the results of BLASTX and to estimate microbial composition using the lowest common ancestor (LCA) algorithm. Briefly, each read matching the sequence of any gene was placed onto the LCA node of those species in the NCBI taxonomy that are known to have that gene, and the taxonomic information of the LCA node was retrieved. To confirm the results, the 16S rRNA genes were also extracted from the reads based on hidden Markov Models (HMMs; Huang et al., 2009), and classified using the RDP classifier (Cole et al., 2009).

For metagenome assembly and gene prediction, after quality filtering of the raw reads, the remaining paired-end reads were used to build a *de novo* assembly by using SOAPdenovo (Luo et al., 2012) with default settings. Open reading frames (ORFs) of assembled contigs were predicted using MetaGeneMark program (Zhu et al., 2010). The predicted ORFs were compared against the NCBI NR database (downloaded from NCBI in September, 2012) of non-redundant protein sequences, the Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologous (KOs) database (Kanehisa et al., 2008), and Clusters of Orthologous Groups (COG) database (Tatusov et al., 2003) to assess the functional classification. For searching genes encoding alkane 1-monooxygenase (AlkB), cytochrome P450 CYP153 family of alkane hydroxylases (CYP153), long-chain alkane monooxygenase (LadA), flavin-binding monooxygenase (AlmA) involved in aerobic *n*-alkane metabolism, and hydrocarbon succinate synthases (AssA/BssA) involved in anaerobic hydrocarbon degradation in the two metagenomes, reference sequences were downloaded from GenBank (Supplementary Table S1). Genes encoding AlkB and CYP153 were annotated as described previously (Nie et al., 2014).

Two crude oil metagenomes NorOil1 (Kotlar et al., 2011) and NorOil2 (Lewin et al., 2014) from offshore oil reservoirs of Norwegian Sea were also analyzed using the same methods.

To identify special genes and pathways in the metagenomes of the oil reservoirs, all the published 1,233 metagenomic datasets from the IMG database (until October, 2013) were downloaded and compared. After removing metagenomes that were not environmental samples, or were without detailed information about the sampling sites, 948 metagenome datasets were selected for analysis. Both the oil metagenomes and all the reference metagenomes were clustered using Cluster 3.0 program (de Hoon et al., 2004) with average linkage method based on the relative abundance of KEGG genes after normalization according to *z*-score. To identify specific functions of adaptive importance in different samples, genes in different regions were affiliated to KEGG pathways. Next, abundance of genes in different pathways was normalized using the *z*-score and was clustered. Bray-Curtis dissimilarity was calculated using Vegan: community ecology package (version 1.8–6) in R statistical program 2.9.1 (R Development Core Team). Based on the KO abundance, principal component analysis (PCA) was performed to evaluate similarity among various metagenomes by using the Vegan package in the R statistical program.

The metagenomics data have been submitted to NCBI SRA and are accessible under the BioProject identifier PRJNA251580.

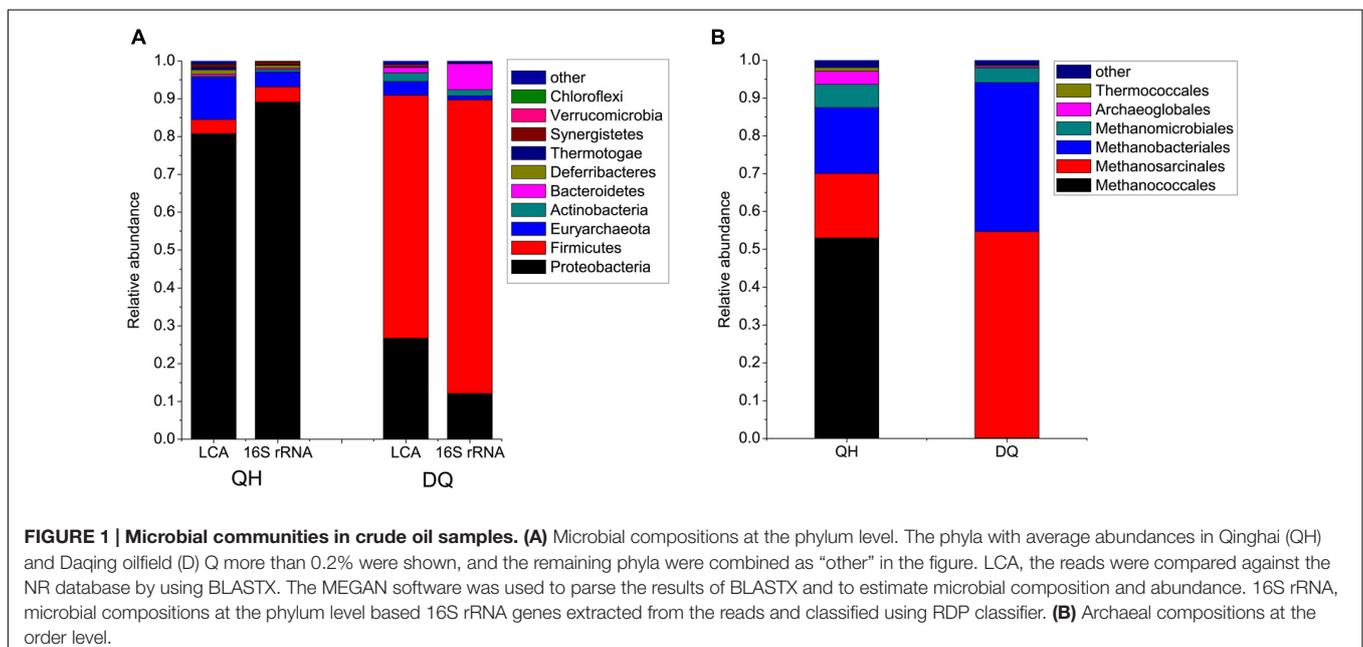
RESULTS

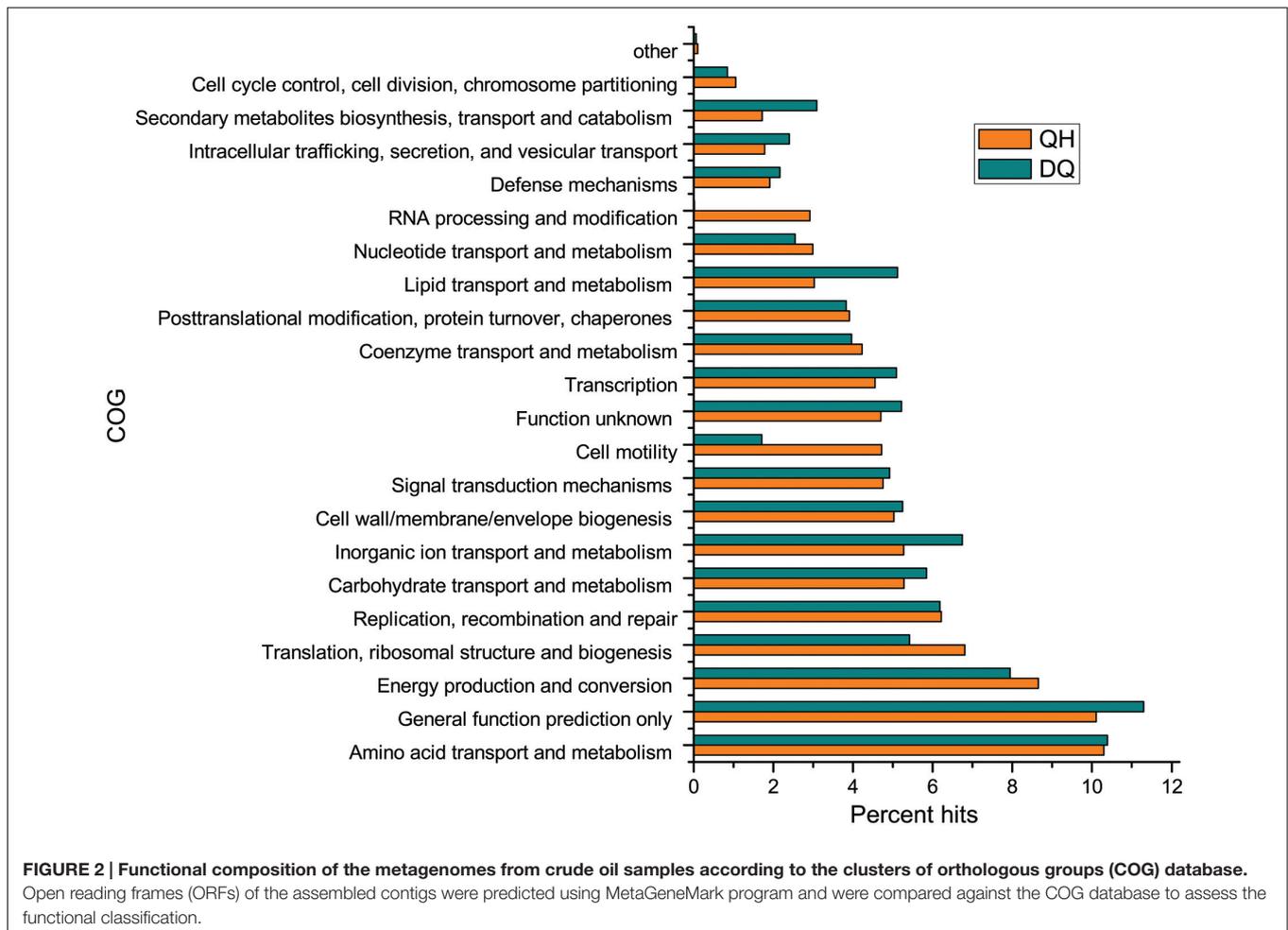
Microbial Community Structure in Oil Samples

After filtering low-quality reads, 68,937,779 and 68,362,793 read pairs were generated from the metagenomes of Qinghai (QH) and Daqing (DQ) oilfields, respectively. Approximately 72.2%

reads obtained from the QH metagenome were assigned to bacteria and archaea (Supplementary Figures S3 and S4); of these, approximately 93% reads were assigned to 34 known phyla (Figure 1; Supplementary Figure S4; Supplementary Table S2). Approximately 73.6% reads obtained from the DQ metagenome were assigned to bacteria and archaea (Supplementary Figure S4); of these, approximately 93% reads were assigned to 35 known phyla (Figure 1; Supplementary Figure S4; Supplementary Table S2). However, approximately 36.2% and 45.4% reads obtained from the QH and DQ metagenomes, respectively, that were assigned to known phyla could not be assigned to known genera (Supplementary Figure S4). The rarefaction analysis of reads based on the taxonomic information showed that a reasonable number of individual genomes were sampled, with the coverage over 99% both in QH and DQ at the genus level (Supplementary Figure S5).

The QH metagenome predominantly included bacteria belonging to *Proteobacteria*, which accounted for 80.8% of the total reads assigned to known phyla (Figure 1). Of the *Proteobacteria*, *Gammaproteobacteria* was the most dominant bacterial class, with genera such as *Pseudomonas* (54.4%), *Acinetobacter* (1%), and *Marinobacter* (1.3%) (Supplementary Figure S6). The remaining microbes mainly belonged to *Euryarchaeota* (11.1%), *Firmicutes* (3.8%), and *Deferribacteres* (1.2%). The QH metagenome included methanogenic archaea (47 genera) such as hydrogenotrophic *Methanococcus*, *Methanothermococcus*, *Methanobacterium*, *Methanoculleus*, and *Methanosarcina* that could produce methane from H₂, acetate or methyl compounds (Supplementary Figure S7; Supplementary Table S3). The DQ metagenome predominantly included bacteria belonging to *Firmicutes* (64.2%) and *Proteobacteria* (26.7%), namely, *Anoxybacillus* (37.2%), *Geobacillus* (2.8%), *Bacillus* (1.1%), *Pseudomonas* (6.4%), and *Brevundimonas* (0.9%) (Supplementary Figures S6 and S8). Microbes belonging to





Euryarchaeota (3.6%), *Actinobacteria* (2.2%), and *Bacteroidetes* (1.2%) were also detected in QH metagenome. The DQ metagenome included methanogenic archaea such as acetoclastic *Methanosaeta* (Supplementary Figure S7; Supplementary Table S3). *Poribacteria* and *Fibrobacteres* were only detected in the DQ metagenome, and *Nanoarchaeota* were only detected in the QH metagenome; however, abundances of these bacteria were very low in both the metagenomes (Supplementary Table S2).

To confirm the results, the 16S rRNA genes were also identified from metagenome reads and classified using the RDP classifier. A total of 3,210 and 1,792 16S rRNA gene fragments were extracted from the reads from QH and DQ metagenomes, respectively. The results revealed that the taxonomic distributions were similar to those from BLASTX-based classification (Figure 1).

Genes Related to Petroleum Degradation Pathways in Oil Reservoirs

In all, 184,871 and 305,582 ORFs were found in the QH and DQ metagenomes, respectively. Of these, 54.9 and 60.0% ORFs, respectively, were assigned to COG database (Figure 2) and 34.3 and 34.8% ORFs, respectively, were assigned to KOs

database (Supplementary Figure S9). Functional community profiles of the QH, DQ, NorOil1, and NorOil2 metagenomes were created based on KEGG annotations. KEGG annotations determined genes involved in complete hydrocarbon metabolism in both QH and DQ metagenomes; however, these genes showed remarkably different taxonomic affiliations. Both QH and DQ metagenomes contained high abundance of genes in lipid metabolism. For example, 129 and 395 genes for 3-oxoacyl-(acyl-carrier-protein) reductase (KO: K00059) were detected in the QH and DQ metagenomes, respectively. These genes were similar to sequences from *Pseudomonas* and *Marinobacter* in the QH metagenome and *Anoxybacillus* and *Pseudomonas* in the DQ metagenome. These data were consistent with the microbial taxonomic compositions of the two metagenomes. Both QH and DQ metagenomes contained glycosyltransferase genes essential for glycolipid and lipopeptide biosynthesis (Supplementary Tables S4 and S5). Moreover, both QH and DQ metagenomes contained high abundance of genes in lipid metabolism. For example, 129 and 395 genes for 3-oxoacyl-(acyl-carrier-protein) reductase (KO: K00059) were detected in the QH and DQ metagenomes, respectively. These genes were similar to sequences from *Pseudomonas* and *Marinobacter* in the QH metagenome and *Anoxybacillus* and *Pseudomonas* in the

DQ metagenome. These data were consistent with the microbial taxonomic compositions of the two metagenomes.

Alkane hydroxylases act in the first step of aerobic hydrocarbon degradation. Genes encoding alkane hydroxylase (EC: 1.14.15.3), alcohol dehydrogenase (EC: 1.1.1.1), and aldehyde dehydrogenase (EC: 1.2.1.3) were detected in both the metagenomes, indicating complete alkane degradation through β -oxidation pathway for fatty acid degradation. The QH and DQ metagenomes included ten and 46 genes encoding alkane 1-monooxygenase (AlkB; EC: 1.14.15.3), three and 66 genes encoding cytochrome P450 CYP153 family of alkane hydroxylases (EC: 1.14.15.3), 10 and 67 genes encoding long-chain alkane monooxygenase (LadA; Feng et al., 2007), and 36 and 149 genes encoding protein *n*-alkane metabolism A (AlmA; Throne-Holst et al., 2007), respectively (Supplementary Figures S10–S13). These results indicated that the DQ metagenome included more diverse genes encoding alkane hydroxylases than the QH metagenome. Genes encoding aerobic alkane hydroxylases were not detected in the NorOil metagenomes. Besides genes involved in aerobic alkane degradation, 16, 14, 18, and 18 genes encoding hydrocarbon succinate synthases (AssA/BssA) were present in the QH, DQ, NorOil1, and NorOil2 metagenomes, respectively. Proteins encoded by these genes activate anaerobic hydrocarbon degradation followed by benzoyl-CoA or fatty acid metabolism (Widdel and Rabus, 2001). Further, *assA/bssA* genes were detected with similarities to sequences from *Desulfotomaculum*, *Anaerofustis*, *Desulfatibacillum*, *Pelobacter*, *Bacteroides*, *Pseudomonas*, *Anaerobaculum*, *Spirochaeta*, and *Thermococcus* in the QH metagenome; *Desulfatibacillum*, *Desulfoglaeba*, *Desulfosporosinus*, *Desulfotomaculum*, *Anoxybacillus*, *Pseudomonas*, and *Sanguibacter* in the DQ metagenome; *Desulfovibrio*, *Pelobacter*, *Clostridium*, and *Thermovirga* in the NorOil1 metagenome; and *Desulfoglaeba*, *Desulfobacula*, *Geobacter*, *Azoarcus*, *Aromatoleum*, *Thauera*, and *Clostridia* in the NorOil2 metagenome (Supplementary Figure S14).

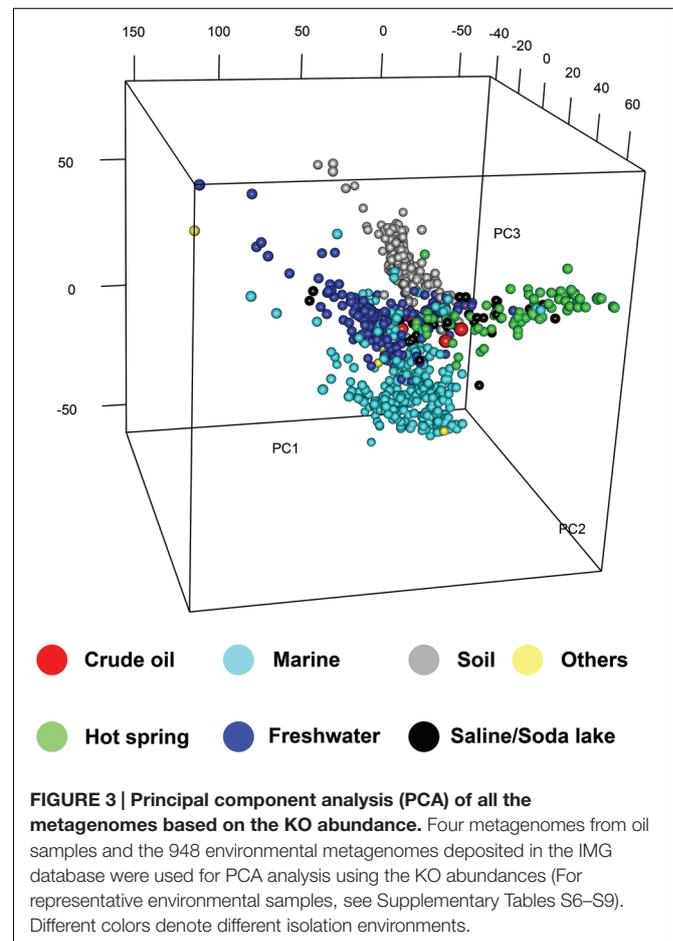
Bifunctional carbon monoxide dehydrogenase/acetyl-CoA synthase (CODH/ACS, EC: 2.3.1.-, 1.2.99.2, 1.2.7.4) is known as one of the key enzymes in Wood-Ljungdahl (W-L) pathway (Ragsdale, 2007). Here, we detected the genes encoding CODH/ACS in the QH and DQ metagenomes, to estimate the distribution of bacteria with W-L pathway and syntrophic acetate oxidation (SAO) potentials. In all, 79 and 60 reads matching the genes encoding CODH/ACS were detected in the QH and DQ metagenomes, respectively. Only 44 reads in the QH metagenome were assigned to classes *Clostridia*, *Deltaproteobacteria*, and *Synergistia* and 20 reads in the DQ metagenome were assigned to classes *Alphaproteobacteria*, *Deltaproteobacteria*, and unclassified *Poribacteria* (Supplementary Figure S15).

Clustering of Metagenomes of Oil Reservoirs

Taxonomic and functional compositions of QH, DQ, NorOil1, and NorOil2 metagenomes of oil reservoirs were compared. Bray–Curtis dissimilarity analysis indicated that the

metagenomes of the four oil reservoirs (QH, DQ, NorOil1, and NorOil2) showed 58.4% average dissimilarity in taxonomic compositions and 32.2% average dissimilarity in functional compositions (KEGG). These results showed that despite large differences in taxonomic compositions across metagenomes of oil reservoirs, they shared many genes, especially enriched or reduced genes with the same or similar functions, from different species that constituted the core functions of these environments.

Hierarchical clustering of KO abundances was performed for metagenomes from the four oil samples and the 948 environmental metagenomes deposited in the IMG database (until October 2013). PCA analysis was performed using the KO abundances to analyze the relationship among the functional compositions of metagenomes from different environments, which also classified all the metagenomes into four groups according to their isolation environments (Figure 3). Notably, all the environmental metagenomes did not cluster according to their geographical locations but clustered according to their isolation environments (Supplementary Figure S16A), which formed four major clusters (cluster I to cluster IV). Generally, metagenomes from soil, freshwater, and marine samples were clustered, respectively, and were separated from the metagenomes from oil reservoirs and other extreme



environments (Figure 3; Supplementary Figure S16A). In cluster I, the metagenomes from the QH, DQ, NorOil1, and NorOil2 reservoirs were clustered with metagenomes from other extreme environments (Supplementary Table S6). Cluster II included metagenomes mainly from terrestrial soil environments (Supplementary Table S7). Cluster III included metagenomes mainly from marine environments (Supplementary Table S8) and cluster IV included metagenomes from freshwater environments (Supplementary Table S9). These different metagenome clusters showed different enriched functional genes and were viewed as four “hot” gene pools, denoted as groups A–D (Supplementary Figure S16B). In these gene groups, Groups A, B, C, and D corresponded to Clusters I, II, III, and IV, respectively, in the heat map (Supplementary Figure S16A). Of these enriched genes, genes in group A were associated with 43 pathways, including pathways of carbon metabolism, methane metabolism, amino acid biosynthesis (Supplementary Figure S16B; Supplementary Tables S10 and S11). For example, 146 KOs in group A (5.1% of the total KOs in group A) were assigned to the carbon metabolism pathway, including genes encoding CODH/ACS involved in W-L pathway. In contrast, other genes in carbon fixation, such as RuBisCO in Calvin cycle and ATP-citrate lyase in the reductive tricarboxylic acid cycle (Atomi, 2002), were less abundant in group A and in the metagenomes of oil reservoirs. We also identified pathways in which genes in groups A and B were both highly distributed. These genes were mainly related to nitrogen metabolism, two-component systems (TCSs), ABC transporters, and bacterial chemotaxis (Supplementary Figure S16B; Supplementary Table S11).

Specific Functional Profiles of Metagenomes of Crude Oil Reservoirs

We closely examined the enriched gene pool in group A. Methanogenesis-related functions were among the enriched pathways. In all, 93 KOs of methane metabolism were detected, constituting both complete hydrogenotrophic (KEGG module M00567) and acetoclastic (M00357) methanogenic pathways. The QH and DQ metagenomes included 78 and 27 reads, respectively, matching methyl-coenzyme M reductase alpha subunit (*mcrA*), a key enzyme for methanogenesis and specific marker of methanogens (Luton et al., 2002; Supplementary Figure S17). *mcrA* genes similar to those from *Methanococcales* were dominant in the QH metagenome, accounting for 16.7% of the total *mcrA* genes. *Methanobacteriales* and *Methanosarcinales* similar *mcrA* genes were dominant in the DQ metagenome, accounting for 37.0 and 33.3% of the total *mcrA* genes, respectively. These results also showed that the DQ metagenome had more genes with similarities to sequences from *Methanosarcinales* and *Methanobacteriales*. Further, nitrogen fixation-related genes *nifDKH* (K02586, K02591, and K02588) were similar to the *nif* genes from methanogens such as *Methanococcales* (144/289 reads), *Methanobacteriales* (22/289 reads), and *Methanosarcinales* (19/289 reads) in the QH metagenome and *Methanobacteriales*

(13/90 reads), *Methanomicrobia* (5/90 reads) in the DQ metagenome (Supplementary Figure S18).

In all, 15 complete TCSs such as PhoR-PhoB, KinB-AlgB, YesM-YesN, and VicK-VicR (Supplementary Table S12); 19 complete ABC transporter systems (Supplementary Table S13) were detected in the enriched KOs obtained from the QH and DQ metagenomes. TCSs are important for microbes to adapt to the extreme environment in oil reservoirs, which is anaerobic and hydrophobic, lacks phosphates and amino acids, and contains hydrocarbons as carbon sources and solvents. Genes of TCSs were similar to sequences from *Methanococcales*, *Methanosarcinales*, *Clostridiales*, *Proteobacteria*, and *Deferribacteres* in the QH metagenome and *Anoxybacillus* in the DQ metagenome. The complete ABC transporter systems in the metagenomes of the oil reservoirs included membrane transport for glutamine and cysteine. However, transporters for L-arabinose, lactose, and oligogalacturonide and for multiple sugars were absent.

Although not enriched, many genes related to bacterial osmotic stress responses, including those encoding transcriptional regulators and proteins involved in compatible solute transport and synthesis, were present in the metagenomes of the oil reservoirs (Supplementary Table S14), which was similar to that observed in other extreme environments.

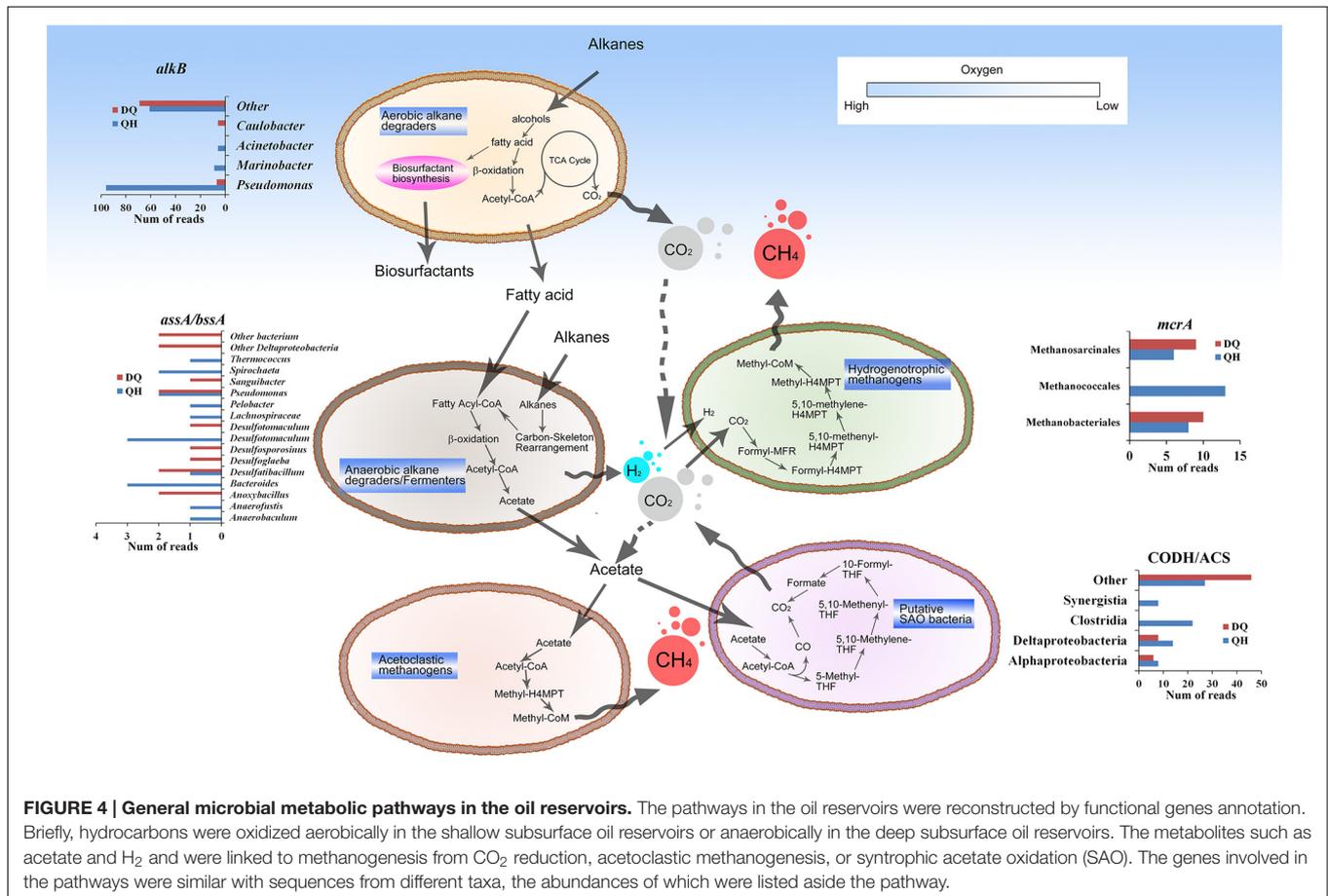
DISCUSSION

Although oil reservoirs exhibit extreme conditions to support life, accumulating evidence has indicated that oil reservoirs are complex ecosystems comprising numerous microbes (Li et al., 2007; Dahle et al., 2008; Kaster et al., 2009). To adapt to these ecosystems, microbes have developed compatible functions such as enriched systems for carbon metabolism, methane metabolism, stress response, sensing, and transport. Enrichment of genes involved in carbon metabolism pathways might be because of the special requirements of carbon metabolism in these harsh environments. For example, although petroleum constituents are hydrophobic and toxic, they are the only potential organic carbon source for microbes in oil reservoirs and contribute to the enrichment of genes involved in both aerobic and anaerobic alkane degradation pathways, methane metabolism, TCSs, and ABC transporter systems. Both QH and DQ metagenomes contained genes essential for glycolipid and lipopeptide biosynthesis (Supplementary Tables S4 and S5), which could help bacteria to access hydrophobic oil constituents (Wang X.B. et al., 2013). Enriched genes encoding AlkB, CYP153, LadA, and AlmA for aerobic alkane degradation were detected in the QH and DQ metagenomes but were absent in the NorOil1 or NorOil2 metagenomes. This might be because the DQ and QH reservoirs have been subjected to water flooding for more than 10 years. This may have provided small amount of oxygen through the injected water into the oil reservoirs and that may have stimulated aerobic alkane metabolism. Oxygen can also be provided by oxygen-bearing meteoric water, which results in the aerobic degradation of crude oil in shallow subsurface oil reservoirs

(Aitken et al., 2004). However, anaerobiosis is dominant in oil reservoirs (Head et al., 2003; Aitken et al., 2004; Dolfig et al., 2007; Jones et al., 2008). The metabolites such as organic acids generated during aerobic hydrocarbon degradation at the shallow subsurface could be further utilized in the deep surface of the oil reservoirs. Besides, anaerobic degradation of hydrocarbons in deep subsurface oil reservoirs are linked to methanogenesis via acetate or H₂/CO₂ (Figure 4; Dolfig et al., 2007). The acetate and H₂/CO₂ could be linked to several reactions and they might form a complex regulating network (Figure 4) which is influenced by factors such as H₂ and CO₂ concentration and temperature. SAO bacteria were thought using the reversal of W-L pathway to convert acetate to CO₂/H₂ that could be subsequently utilized by hydrogenotrophic methanogens (Hattori, 2008). Moreover, the enzyme activities in W-L pathway such as CO dehydrogenase, were detected in SAO bacteria (Hattori et al., 2005). The high abundance of hydrogenotrophic methanogens such as *Methanococcales*, *Methanobacteriales*, and *Methanomicrobiales* in the metagenomes of the oil reservoirs (Figure 1; Supplementary Figure S17), and the low concentration of detectable H₂ in oil reservoirs and microcosms of methanogenic hydrocarbon degradation (Mayumi et al., 2011) suggest a quick consumption of H₂ by hydrogenotrophic methanogens and low possibility of homoacetogenesis in oil reservoirs. The low concentration

of H₂ could also benefit SAO bacteria (Dolfig et al., 2007; Kotlar et al., 2011), which convert acetate to H₂ and CO₂. Moreover, at high temperatures, acetoclastic methanogenesis becomes thermodynamically less favorable (Dolfig et al., 2007). Thus, at high temperatures, hydrogenotrophic methanogenesis is dominant. Therefore, hydrogenotrophic methanogens such as *Methanothermococcus* were dominant in the NorOil1 (at 85°C) and NorOil2 (at 83°C) metagenomes (Kotlar et al., 2011; Lewin et al., 2014), and the proportion of *Methanococcus* was higher in the QH metagenome while that of *Methanosaeta* was higher in the DQ metagenome. These results were consistent with those reported by Pham et al (Pham et al., 2009) who showed that acetoclastic *Methanosaeta* were dominant in some mesophilic oil reservoirs.

In contrast to the abundance of hydrocarbons, absence of saccharides in oil reservoirs should contribute to the low abundance of genes associated with sugar metabolism in the metagenomes of oil reservoirs. Although genes such as RuBisCO in Calvin cycle, and ATP-citrate lyase in the reductive tricarboxylic acid cycle were detected in low abundances, the detection of genes in WL pathway suggested the existing of chemoautotrophic microbes in oil reservoirs. Besides metabolism of hydrocarbons, this pathway could also provide organic carbon sources for other microbes. Furthermore, due to the absence of amino acids in oil reservoirs, microbes in these environments



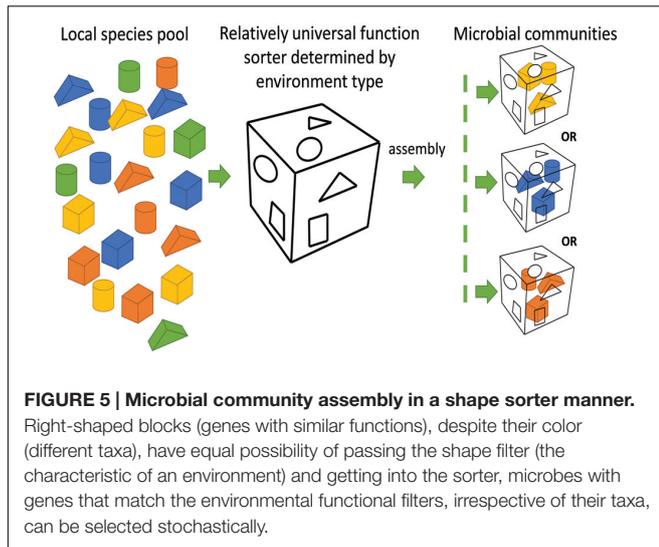
evolved effective systems to synthesize and transport amino acids for protein translation, as observed in group A (cluster I, Supplementary Figure S16A). Interestingly, *nif* was mainly enriched in methanogens in group A. It was not unusual because such genes have also been detected in some methane seep and consuming environments (Dekas et al., 2009; Miyazaki et al., 2009). Nitrogen-fixing archaea in these environments might use nitrogen fixation to overcome the scarcity of bioavailable nitrogen and may serve as the source of nitrogen for microbial communities (Dekas et al., 2009). Due to the low concentration of ammonia in oil reservoirs, it was suggested the important role of methanogens in the nitrogen cycle in oil reservoirs.

Besides hydrocarbon metabolism-related genes, microbes in oil reservoirs have developed special systems to adapt to crude oil habitats, to survive, to obtain nutrients and metabolize it, and to escape toxic compounds and high osmosis. For example, enriched VicR/VicK TCS, which mediates exopolysaccharide biosynthesis and biofilm formation (Fabret and Hoch, 1998; Winkler and Hoch, 2008), and LiaS/LiaR TCS, which provides resistance against organic solvents (Mascher et al., 2004), are beneficial for microbes to mitigate the impact of osmosis from petroleum hydrophobicity and to promote the survival of microbial communities. AtoS/AtoC can induce the bacterial growth in short-chain fatty acids, and positively regulate the levels of poly-(R)-3-hydroxybutyrate (PHB; Kyriakidis and Tiligada, 2009). PHB is a member of the polyhydroxyalkanoate (PHA) which serves as the intracellular carbon and energy storage compounds and enhances the survival of bacteria, usually when an essential nutrient such as nitrogen or phosphorus is limited in the presence of excess carbon source (Anderson and Dawes, 1990). These functions are important for microbes to survive in extreme environments of oil reservoirs, where the organic carbon sources, i.e., hydrocarbons are abundant, but the bioavailable nitrogen and phosphorus are limited.

High functional similarity observed among the metagenomes of the oil reservoirs was also observed among those of other environments. Functional profiles of all the 948 environmental metagenomes deposited in the IMG database (until October 2013) showed that metagenomes from soil, freshwater, marine, and extreme environments had their own special functions, irrespective of their geographical location (Figure 3; Supplementary Figure S16). This was consistent with the results of recent studies, which showed that microbes in similar environments selected similar functions to adapt to their surroundings (Nemergut et al., 2013). For example, higher availability of aromatic compounds such as aromatic carboxylic humic acids in soil (Van Trump et al., 2011) plays an important role in enriching genes involved in aerobic degradation of aromatic compounds (group D; Supplementary Figure S16B). In marine environments, high concentrations of nitrate and nitrite in euphotic and mesopelagic zones below the nitracline (Shi et al., 2010) and extreme oligotrophy in deep sea (Smedile et al., 2014) are important for enriching genes involved in nitrogen metabolism, bacterial chemotaxis, and flagellar assembly (group B; Supplementary Table S11). Therefore, it is not surprising that many marine bacteria are motile, with the

proportion of motile cells being as high as 20–60% (Stocker, 2012). Consequently, specific environmental factors to the four environments resulted in the selection of characteristic microbial functions and clustering of metagenomes from these environmental (Figure 3). Similarly, genes related to methane metabolism were enriched in the metagenomes of the oil reservoirs. This was consistent with the finding that methanogenic anaerobic degradation of crude oil is an important process in subsurface petroleum reservoirs, which distinguishes oil reservoirs from other environments (Jones et al., 2008).

Microbial compositions at the taxonomic level varied remarkably among oilfields. Microbial compositions of the metagenomes from offshore oil reservoirs of Norwegian Sea (NorOil1 and NorOil2; Kotlar et al., 2011; Lewin et al., 2014) which were located in the same geographical region but were physically separated, were remarkably different from the metagenomes of the QH and DQ reservoirs. *Delta/Epsilon-proteobacteria* and methanogens were abundant in the NorOil1 metagenome (Kotlar et al., 2011). However, archaea were the most dominant microbes in the NorOil2 metagenome (Lewin et al., 2014). In contrast, *Pseudomonas* and *Anoxybacillus* were dominant in the QH and DQ, respectively (Figure 1). Microbial compositions at the taxonomic level were remarkably different even between these oil reservoirs. However, functional compositions, i.e., metabolic pathways and functional genes, were common among geographically distant oil reservoirs and extreme environments such as hot spring and soda lakes. Bray–Curtis dissimilarity analysis also confirmed the results that despite large differences in taxonomic compositions across metagenomes of oil reservoirs, they shared many genes, especially enriched or reduced genes with the same or similar functions, from different species. These phenomena may be attributed to the mechanisms underlying the assembly of microbial communities. The popular niche theory suggests that the structure of a microbial community is shaped mainly by factors such as environmental conditions, competition, and niche differentiation. Environment and microbial interactions dominate this process (Ofiteru et al., 2010). In contrast, neutral theory proposes that species in environments are ecologically equivalent and that the structures of microbial communities are determined randomly (Woodcock et al., 2007). However, these two theories can only explain a part of our results. The niche theory explained the selection of similar microbial functions by similar environments (Figure 3) but could not explain the remarkable difference in taxonomic microbial compositions. In contrast, the neutral theory explained the variation in the distribution of microbes at the taxonomic level among different oil reservoirs, which was consistent with the stochastic model of community assembly. However, this theory could not explain the clustering of samples from four typical environments. Divergence of species and convergence of functional compositions observed in the metagenomes of the oil reservoirs suggested that the assembly and structure of bacterial communities may not be based on species but may be based on functional genes determined by the respective environments, which was consistent with the findings of a previous study (Burke et al., 2011).



In this aspect, the assembly of microbial communities in oil reservoirs was affected by the effects of the deterministic assembly processes in a “shape-sorting” manner based on the functional compositions. Like a shape sorter that contains different shape filters and that only allows blocks with the correct shape to pass (Figure 5), an environment with characteristic functional filters acts as a sorter of functions. This sorter, which is a characteristic of that environment type, is used to select microbes from nearby pools. Similar to the game in which right-shaped blocks, despite their color, have equal possibility of passing the shape filter and getting into the sorter, microbes with genes that match the environmental functional filters, irrespective of their taxa, can be selected stochastically (neutral theory). Next, these randomly selected successful microbes outcompete other microbes, colonize the environment, and occupy the niche. Therefore, different microbial species with similar properties or functions can randomly occupy the same niche in an ecosystem. Like oil reservoirs, soil, freshwater, marine, and extreme environments have their specific environmental factors and form four types of sorters with environment-specific functional filters for selecting genes.

Although the shape sorter equally allows the right-shaped blocks with different colors to pass through, colors with more number of blocks have higher possibilities to get into the shape sorter. Similarly, although the functional sorter is environment determined and is relatively universal within an environment type, microbes close to the sorter may be endemic because of their biogeographic patterns, especially at a larger scale (Martiny et al., 2006). Therefore, endemic differences in microbial composition should play a role in the assembly of microbial communities by offering endemic pools of diverse microbial candidates for selection. Thus, the shape-sorting system is comprehensively determined by the niche and neutral theories, local legacies of historical microbial events on a large scale, and differences among different microbial taxa. For example, crude oil has its own functional sorting system that selects specific functions from

different taxa. However, further research is needed to test this hypothesis.

In summary, we characterized microbial functional gene compositions in crude oil from two geographically distant oil reservoirs (QH and DQ) in China and found that these genes clustered with genes from extreme environments but not with genes from soil, freshwater, and marine environments. Although microbial compositions were taxonomically different between oil reservoirs with different geochemical conditions and between different wells from the same oilfield with similar geochemical conditions, functional gene compositions in the metagenomes of these oil reservoirs showed high similarities. The preserved common complete methanogenic hydrocarbon degradation pathways and enriched stress response genes were the key factors for distinguishing the metagenomes of oil reservoirs and extreme environments from those of soil, freshwater, and marine environments. These results suggested that microbial communities in oil reservoirs with consistent functional profiles are functionally selected by the niche theory but taxonomically selected by the neutral theory. Moreover, all the environmental metagenomes were clustered according to their isolation environments based on the hierarchical clustering of functional genes, suggesting that both selective and neutral processes were involved in the assembly of microbial communities in oil reservoirs.

AUTHOR CONTRIBUTIONS

YN and J-YZ contributed equally to this work. YN, X-LW, and FZ designed the research. YN, J-YZ, and Y-QT performed research. J-YZ re-analyzed the datasets, and revised the manuscript. PG and YY contributed analytic tools. YN, J-YZ, YY, X-LW, and FZ analyzed data. YN, J-YZ, YY, and X-LW wrote the paper.

FUNDING

This study was supported by the National Natural Science Foundation of China (31225001, 31300108), the National High Technology Research and Development Program of China (“863” Program: 2012AA02A703) and the National Basic Research Program of China (“973” Program, 2014CB846002).

ACKNOWLEDGMENTS

We thank Dr. Man Cai and Dr. Yan Li for their kindly help in DNA extraction. We thank Dr. Svein Vella and Dr. Anna Lewin for sharing the metagenome datasets of NorOil1 and NorOil2. We also thank Ms. Hui Fang for her help in bioinformatic analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.01254>

REFERENCES

- Aitken, C. M., Jones, D. M., and Larter, S. R. (2004). Anaerobic hydrocarbon biodegradation in deep subsurface oil reservoirs. *Nature* 431, 291–294. doi: 10.1038/nature02922
- Anderson, A. J., and Dawes, E. A. (1990). Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. *Microbiol. Rev.* 54, 450–472.
- Atomi, H. (2002). Microbial enzymes involved in carbon dioxide fixation. *J. Biosci. Bioeng.* 94, 497–505. doi: 10.1016/S1389-1723(02)80186-4
- Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S., and Thomas, T. (2011). Bacterial community assembly based on functional genes rather than species. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14288–14293. doi: 10.1073/pnas.1101591108
- Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., et al. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic. Acids Res.* 37(Suppl. 1), D141–D145. doi: 10.1093/nar/gkn879
- Dahle, H., Garshol, F., Madsen, M., and Birkeland, N. K. (2008). Microbial community structure analysis of produced water from a high-temperature North Sea oil-field. *Antonie Van Leeuwenhoek* 93, 37–49. doi: 10.1007/s10482-007-9177-z
- de Hoon, M. J., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. *Bioinformatics* 20, 1453–1454. doi: 10.1093/bioinformatics/bth078
- Dekas, A. E., Poretsky, R. S., and Orphan, V. J. (2009). Deep-sea archaea fix and share nitrogen in methane-consuming microbial consortia. *Science* 326, 422–426. doi: 10.1126/science.1178223
- Dolfing, J., Larter, S. R., and Head, I. M. (2007). Thermodynamic constraints on methanogenic crude oil biodegradation. *ISME J.* 2, 442–452. doi: 10.1038/ismej.2007.111
- Fabret, C., and Hoch, J. A. (1998). A two-component signal transduction system essential for growth of *Bacillus subtilis*: implications for anti-infective therapy. *J. Bacteriol.* 180, 6375–6383.
- Feng, L., Wang, W., Cheng, J., Ren, Y., Zhao, G., Gao, C., et al. (2007). Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5602–5607. doi: 10.1073/pnas.0609650104
- Gao, P. K., Tian, H. M., Wang, Y. S., Li, Y. S., Li, Y., Xie, J. X., et al. (2016). Spatial isolation and environmental factors drive distinct bacterial and archaeal communities in different types of petroleum reservoirs in China. *Sci. Rep.* 6:20174. doi: 10.1038/srep20174
- Guan, J., Xia, L. P., Wang, L. Y., Liu, J. F., Gu, J. D., and Mu, B. Z. (2013). Diversity and distribution of sulfate-reducing bacteria in four petroleum reservoirs detected by using 16S rRNA and dsrAB genes. *Int. Biodeterior. Biodegrad.* 76, 58–66. doi: 10.1016/j.ibiod.2012.06.021
- Hattori, S. (2008). Syntrophic acetate-oxidizing microbes in methanogenic environments. *Microbes Environ.* 23, 118–127. doi: 10.1264/jsme2.23.118
- Hattori, S., Galushko, A. S., Kamagata, Y., and Schink, B. (2005). Operation of the CO dehydrogenase/acetyl coenzyme A pathway in both acetate oxidation and acetate formation by the syntrophically acetate-oxidizing bacterium *Thermacetogenium phaeum*. *J. Bacteriol.* 187, 3471–3476. doi: 10.1128/JB.187.10.3471-3476.2005
- Head, I. M., Jones, D. M., and Larter, S. R. (2003). Biological activity in the deep subsurface and the origin of heavy oil. *Nature* 426, 344–352. doi: 10.1038/nature02134
- Huang, Y., Gilna, P., and Li, W. (2009). Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* 25, 1338–1340. doi: 10.1093/bioinformatics/btp161
- Huson, D. H., Mitra, S., Ruscchewey, H. J., Weber, N., and Schuster, S. C. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* 21, 1552–1560. doi: 10.1101/gr.120618.111
- Jones, D. M., Head, I. M., Gray, N. D., Adams, J. J., Rowan, A. K., Aitken, C. M., et al. (2008). Crude-oil biodegradation via methanogenesis in subsurface petroleum reservoirs. *Nature* 451, 176–180. doi: 10.1038/nature06484
- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic. Acids Res.* 36, D480–D484. doi: 10.1093/nar/gkm882
- Kaster, K. M., Bonaunet, K., Berland, H., Kjeilen-Eilertsen, G., and Brakstad, O. G. (2009). Characterisation of culture-independent and -dependent microbial communities in a high-temperature offshore chalk petroleum reservoir. *Antonie Van Leeuwenhoek* 96, 423–439. doi: 10.1007/s10482-009-9356-1
- Kotlar, H. K., Lewin, A., Johansen, J., Throne-Holst, M., Haverkamp, T., Markussen, S., et al. (2011). High coverage sequencing of DNA from microorganisms living in an oil reservoir 2.5 kilometres subsurface. *Environ. Microbiol. Rep.* 3, 674–681. doi: 10.1111/j.1758-2229.2011.00279.x
- Kryachko, Y., Dong, X., Sensen, C. W., and Voordouw, G. (2012). Compositions of microbial communities associated with oil and water in a mesothermic oil field. *Antonie Van Leeuwenhoek* 101, 493–506. doi: 10.1007/s10482-011-9658-y
- Kyriakidis, D. A., and Tiligada, E. (2009). Signal transduction and adaptive regulation through bacterial two-component systems: the *Escherichia coli* AtoSC paradigm. *Amino Acids* 37, 443–458. doi: 10.1007/s00726-009-0241-z
- Lewin, A., Johansen, J., Wentzel, A., Kotlar, H. K., Drablos, F., and Valla, S. (2014). The microbial communities in two apparently physically separated deep subsurface oil reservoirs show extensive DNA sequence similarities. *Environ. Microbiol.* 16, 545–558. doi: 10.1111/1462-2920.12181
- Li, G. Q., Gao, P. K., Wu, Y. Q., Tian, H. M., Dai, X. C., Wang, Y. S., et al. (2014). Microbial abundance and community composition influence production performance in a low-temperature petroleum reservoir. *Environ. Sci. Technol.* 48, 5336–5344. doi: 10.1021/es500239w
- Li, H., Mu, B. Z., Jiang, Y., and Gu, J. D. (2011). Production processes affected prokaryotic amoA Gene abundance and distribution in high-temperature petroleum reservoirs. *Geomicrobiol. J.* 28, 692–704. doi: 10.1080/01490451.2010.514026
- Li, H., Yang, S. Z., Mu, B. Z., Rong, Z. F., and Zhang, J. (2006). Molecular analysis of the bacterial community in a continental high-temperature and water-flooded petroleum reservoir. *FEMS Microbiol. Lett.* 257, 92–98. doi: 10.1111/j.1574-6968.2006.00149.x
- Li, H., Yang, S.-Z., Mu, B.-Z., Rong, Z.-F., and Zhang, J. (2007). Molecular phylogenetic diversity of the microbial community associated with a high-temperature petroleum reservoir at an offshore oilfield. *FEMS Microbiol. Ecol.* 60, 74–84. doi: 10.1111/j.1574-6941.2006.00266.x
- Liu, J. F., Sun, X. B., Yang, G. C., Mbadinga, S. M., Gu, J. D., and Mu, B. Z. (2015). Analysis of microbial communities in the oil reservoir subjected to CO₂-flooding by using functional genes as molecular biomarkers for microbial CO₂ sequestration. *Front. Microbiol.* 6:236. doi: 10.3389/fmicb.2015.00236
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1:18. doi: 10.1186/2047-217X-1-18
- Luton, P. E., Wayne, J. M., Sharp, R. J., and Riley, P. W. (2002). The mcrA gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogen populations in landfill. *Microbiology* 148, 3521–3530. doi: 10.1099/00221287-148-11-3521
- Martiny, J. B., Bohannan, B. J., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., et al. (2006). Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* 4, 102–112. doi: 10.1038/nrmicro1341
- Mascher, T., Zimmer, S. L., Smith, T. A., and Hellmann, J. D. (2004). Antibiotic-inducible promoter regulated by the cell envelope stress-sensing two-component system LiaRS of *Bacillus subtilis*. *Antimicrob. Agents Chemother.* 48, 2888–2896. doi: 10.1128/AAC.48.8.2888-2896.2004
- Mayumi, D., Mochimaru, H., Yoshioka, H., Sakata, S., Maeda, H., Miyagawa, Y., et al. (2011). Evidence for syntrophic acetate oxidation coupled to hydrogenotrophic methanogenesis in the high-temperature petroleum reservoir of Yabase oil field (Japan). *Environ. Microbiol.* 13, 1995–2006. doi: 10.1111/j.1462-2920.2010.02338.x
- Miyazaki, J., Higa, R., Toki, T., Ashi, J., Tsunogai, U., Nunoura, T., et al. (2009). Molecular characterization of potential nitrogen fixation by anaerobic methane-oxidizing archaea in the methane seep sediments at the number 8 Kumano Knoll in the Kumano Basin, offshore of Japan. *Appl. Environ. Microbiol.* 75, 7153–7162. doi: 10.1128/AEM.01184-09
- Nemergut, D. R., Schmidt, S. K., Fukami, T., O'neill, S. P., Bilinski, T. M., Stanish, L. F., et al. (2013). Patterns and processes of microbial community assembly. *Microbiol. Mol. Biol. Rev.* 77, 342–356. doi: 10.1128/MMBR.00051-12
- Nie, Y., Chi, C. Q., Fang, H., Liang, J. L., Lu, S. L., Lai, G. L., et al. (2014). Diverse alkane hydroxylase genes in microorganisms and environments. *Sci. Rep.* 4:4968. doi: 10.1038/srep04968
- Ofteru, I. D., Lunn, M., Curtis, T. P., Wells, G. F., Criddle, C. S., Francis, C. A., et al. (2010). Combined niche and neutral effects in a microbial wastewater

- treatment community. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15345–15350. doi: 10.1073/pnas.1000604107
- Pham, V. D., Hnatow, L. L., Zhang, S., Fallon, R. D., Jackson, S. C., Tomb, J. F., et al. (2009). Characterizing microbial diversity in production water from an Alaskan mesothermic petroleum reservoir with two independent molecular methods. *Environ. Microbiol.* 11, 176–187. doi: 10.1111/j.1462-2920.2008.01751.x
- Ragsdale, S. W. (2007). Nickel and the carbon cycle. *J. Inorg. Biochem.* 101, 1657–1666. doi: 10.1016/j.jinorgbio.2007.07.014
- Ren, H. Y., Zhang, X. J., Song, Z. Y., Rupert, W., Gao, G. J., Guo, S. X., et al. (2011). Comparison of microbial community compositions of injection and production well samples in a long-term water-flooded petroleum reservoir. *PLoS ONE* 6:e23258. doi: 10.1371/journal.pone.0023258
- Shi, Y., Tyson, G. W., Eppley, J. M., and Delong, E. F. (2010). Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J.* 5, 999–1013. doi: 10.1038/ismej.2010.189
- Smedile, F., Messina, E., La Cono, V., and Yakimov, M. M. (2014). Comparative analysis of deep-sea bacterioplankton OMICS revealed the occurrence of habitat-specific genomic attributes. *Mar. Genomics* 17, 1–8. doi: 10.1016/j.margen.2014.06.001
- Stocker, R. (2012). Marine microbes see a sea of gradients. *Science* 338, 628–633. doi: 10.1126/science.1208929
- Tang, Y. Q., Li, Y., Zhao, J. Y., Chi, C. Q., Huang, L. X., Dong, H. P., et al. (2012). Microbial communities in long-term, water-flooded petroleum reservoirs with different in situ temperatures in the Huabei Oilfield, China. *PLoS ONE* 7:e33535. doi: 10.1371/journal.pone.0033535
- Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. doi: 10.1186/1471-2105-4-41
- Throne-Holst, M., Wentzel, A., Ellingsen, T. E., Kotlar, H. K., and Zotchev, S. B. (2007). Identification of Novel Genes Involved in Long-Chain n-Alkane Degradation by *Acinetobacter* sp, Strain DSM 17874. *Appl. Environ. Microbiol.* 73, 3327–3332. doi: 10.1128/AEM.00064-07
- Van Trump, J. I., Wrighton, K. C., Thrash, J. C., Weber, K. A., Andersen, G. L., and Coates, J. D. (2011). Humic acid-oxidizing, nitrate-reducing bacteria in agricultural soils. *mBio* 2:e00044-11. doi: 10.1128/mBio.00044-11
- Wang, J., Qi, J., Zhao, H., He, S., Zhang, Y., and Zhao, F. (2013). Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease. *Sci. Rep.* 2013:3. doi: 10.1038/srep01843
- Wang, X. B., Nie, Y., Tang, Y. Q., Wu, G., and Wu, X. L. (2013). n-Alkane Chain length alters *Dietzia* sp. Strain DQ12-45-1b biosurfactant production and cell surface activity. *Appl. Environ. Microbiol.* 79, 400–402.
- Widdel, F., and Rabus, R. (2001). Anaerobic biodegradation of saturated and aromatic hydrocarbons. *Curr. Opin. Biotechnol.* 12, 259–276. doi: 10.1016/S0958-1669(00)00209-3
- Winkler, M. E., and Hoch, J. A. (2008). Essentiality, bypass, and targeting of the YycFG (VicRK) two-component regulatory system in gram-positive bacteria. *J. Bacteriol.* 190, 2645–2648. doi: 10.1128/JB.01682-07
- Woodcock, S., Van Der Gast, C. J., Bell, T., Lunn, M., Curtis, T. P., Head, I. M., et al. (2007). Neutral assembly of bacterial communities. *FEMS Microbiol. Ecol.* 62, 171–180. doi: 10.1111/j.1574-6941.2007.00379.x
- Zhang, F., She, Y. H., Chai, L. J., Banat, I. M., Zhang, X. T., Shu, F. C., et al. (2012). Microbial diversity in long-term water-flooded oil reservoirs with different in situ temperatures in China. *Sci. Rep.* 2:760. doi: 10.1038/srep00760
- Zhu, W., Lomsadze, A., and Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* 38:e132. doi: 10.1093/nar/gkq275

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Nie, Zhao, Tang, Guo, Yang, Wu and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.