

# SCIENTIFIC REPORTS



OPEN

## Crude oil as a microbial seed bank with unexpected functional potentials

Man Cai<sup>1,\*</sup>, Yong Nie<sup>1,\*</sup>, Chang-Qiao Chi<sup>1</sup>, Yue-Qin Tang<sup>2</sup>, Yan Li<sup>1</sup>, Xing-Biao Wang<sup>1</sup>, Ze-Shen Liu<sup>1</sup>, Yunfeng Yang<sup>3</sup>, Jizhong Zhou<sup>3</sup> & Xiao-Lei Wu<sup>1</sup>

Received: 24 June 2015

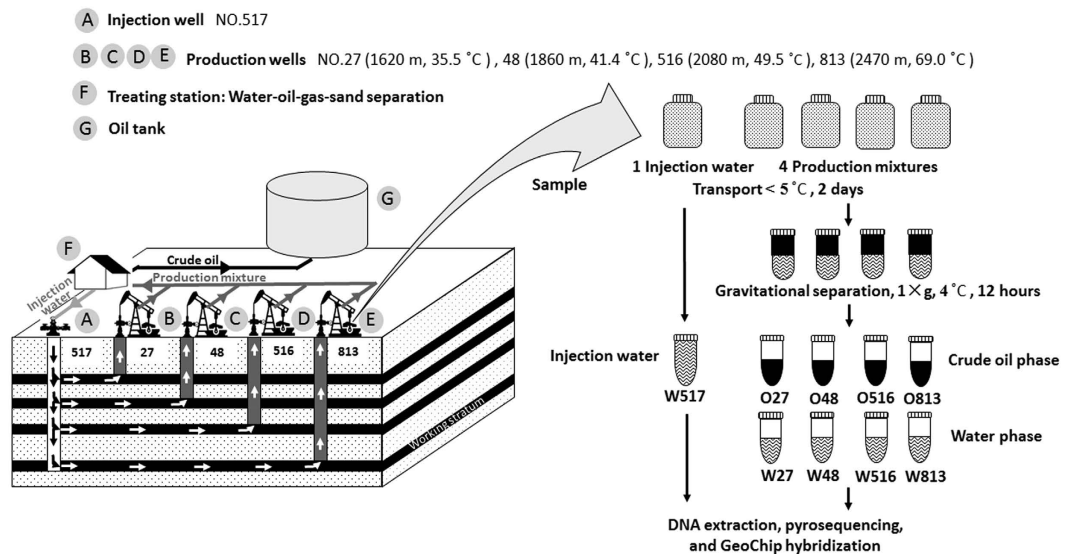
Accepted: 01 October 2015

Published: 03 November 2015

It was widely believed that oil is a harsh habitat for microbes because of its high toxicity and hydrophobicity. However, accumulating evidence has revealed the presence of live microbes in crude oil. Therefore, it's of value to conduct an in-depth investigation on microbial communities in crude oil. To this end, microorganisms in oil and water phases were collected from four oil-well production mixtures in Qinghai Oilfield, China, and analyzed for their taxonomic and functional compositions via pyrosequencing and GeoChip, respectively. Hierarchical clustering of 16S rRNA gene sequences and functional genes clearly separated crude oil and water phases, suggestive of distinct taxonomic and functional gene compositions between crude oil and water phases. Unexpectedly, *Pseudomonas* dominated oil phase where diverse functional gene groups were identified, which significantly differed from those in the corresponding water phases. Meanwhile, most functional genes were significantly more abundant in oil phase, which was consistent with their important roles in facilitating survival of their host organisms in crude oil. These findings provide strong evidence that crude oil could be a "seed bank" of functional microorganisms with rich functional potentials. This offers novel insights for industrial applications of microbial-enhanced oil recovery and bioremediation of petroleum-polluted environments.

Oil is modern industry's lifeblood, but it is also a source of environmental pollution worldwide. In the oil recovery industry, water flooding, i.e., pressurized water being pumped into oil-bearing rock strata to push oil out of the reservoirs, is widely used<sup>1</sup>. However, water flooding has its production limit at ~30% oil recovery, leaving a significant amount of oil untapped<sup>2</sup>. Microbial enhanced oil recovery (MEOR) is a cost-effective technique to recover the residual oil<sup>3</sup>. Generally, microorganisms or nutrients are injected into the oil reservoirs to propagate microorganisms in deep subsurface, which resulted in higher microbial activities that improve oil fluidity and recovery, as field tests in Romashkinskoe Oilfield<sup>4</sup> and Dagang Oilfield<sup>5</sup>. Therefore, understanding the microbial processes in deep oil reservoirs is important for the development of MEOR. Because it is extremely difficult to aseptically sample the oil-bearing rock cores, the injection water and production mixtures are often sampled and analyzed using both culture-dependent and -independent methods to assess the microbial community of an oil reservoir. For example, analysis of the 16S rRNA gene fingerprints using a culture-independent clone library<sup>6–8</sup>, denaturing gradient gel electrophoresis (DGGE)<sup>9,10</sup>, terminal restriction fragment polymorphism (T-RFLP)<sup>11</sup>, and pyrosequencing<sup>1,12</sup> revealed that the major bacteria in oilfields worldwide included *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, and *Planctomycetes*, though their relative abundances vary among oilfields<sup>1,13</sup>. In addition, significant differences in bacterial composition have been detected between injected water and production mixtures in most oilfields. For example, only 2.86% of bacterial operational taxonomic

<sup>1</sup>Peking University, College of Engineering, Beijing, 100871, China. <sup>2</sup>Sichuan University, College of Architecture and Environment, Chengdu, 610065, China. <sup>3</sup>Tsinghua University, School of Environment, State Key Joint Laboratory of Environment Simulation and Pollution Control, Beijing, 100084, China. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to X.-L.W. (email: xiaolei\_wu@pku.edu.cn)

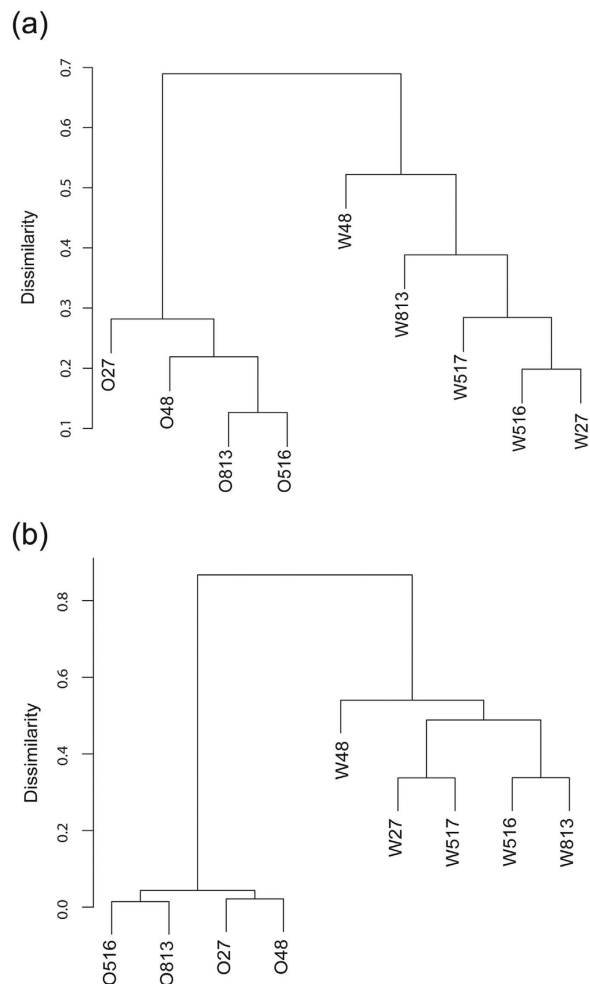


**Figure 1. Outline of sampling and sample management.** The model of four oil-containing rock strata with different depths and proceeding of oil extraction in Qinghai Oilfield, as well as the sample management in lab after sampled from oilfield. All drawings were drew by Man Cai.

units (OTUs) and 15.15% of archaeal OTUs were shared between injection water and the production mixture in the Gudao Oilfield, suggesting that injected water had only a minor influence on the microbial community indigenous to the oil containing strata<sup>10,11</sup>. Although diversities of several functional genes were investigated in some oilfields, such as the methyl coenzyme M reductase genes (*mcr*)<sup>14</sup> and dissimilatory sulfite reductase genes (*dsr*)<sup>15</sup>, community-scale functional gene diversity has not been assessed thoroughly. Therefore, the overall functional potential of microbes from these oilfields remains unclear.

It is widely believed that oil is not a preferred habitat for microbes because of its potential toxicity and high hydrophobicity<sup>16</sup>. However, recent studies have shown that microorganisms including *Acinetobacter*, *Propionibacterium*, *Sphingobium*, *Bacillales*, *Burkholderia*, and *Brevundimonas* are present in crude oil phase<sup>17,18</sup>. In a recent study, we observed alive microbial cells in crude oil, which could be stimulated to bloom by introduction of other bacteria<sup>19</sup>. In addition, different microbial communities from the oil and water phases of the same oil production mixture were detected in a high-temperature oilfield in Japan<sup>20</sup> and a mesothermic oilfield in Canada<sup>12</sup>. Hydrogen- and/or hydrocarbon-utilizing bacteria and archaea preferentially attached to the crude oil phase, whereas *Deltaproteobacteria* were the main microbial group identified in the water phase<sup>12</sup>. The detection of these microorganisms in crude oil is changing our understanding of microorganisms, as well as crude oil itself. However, it is questionable whether highly hydrophobic oil accommodates microorganisms ubiquitously, and whether functional potentials of the microorganisms residing in crude oil are distinct from those in corresponding water phase. Moreover, because crude oil microenvironments in oil reservoirs are thought to be hardly disturbed by the surface water<sup>21</sup>, it is of much interest whether the microbes residing in the subterranean oil phase evolve in a way different from those in the water phase or other environments. In addition, it is typical to detect several oil-bearing strata with different depth in an oilfield. Do these different crude oils accommodate different functional potentials due to variations in, pressure and temperature? If the answers to these questions are yes, it would be of great interest to examine whether crude oil itself is a reservoir of rare species that can potentially expand rapidly under suitable conditions, which is defined as a “seed bank” - a reservoir of dormant individuals that can expand rapidly when conditions are right<sup>22,23</sup>. This reservoir of functional microorganisms might play important roles in the natural oil attenuation process during the long history of oil reservoir formation. Moreover, such an investigation could be important for developing technologies such as MEOR and bioremediation of oil-polluted environments.

Here, we used high-throughput pyrosequencing and a microarray-based metagenomic tool GeoChip 4.0<sup>24</sup> to investigate the microbial communities between oil and water phases which were sampled from four oil-containing rock strata with different depths (temperatures) and groundwater salinities at the Qinghai Oilfield (Fig. 1). Unexpectedly, highly diverse functional groups in microorganisms associated with the oil phase were detected, and significant differences between the crude oil- and water-related microbial communities were observed. Furthermore, microbial communities in oil phases from different rock strata were taxonomically similar, but their functional potentials differed significantly. Because of the diverse functional potentials, crude oil itself could be a “seed bank” of functional microorganisms for elucidating crude oil-related natural microbial processes and developing new oil recovery technologies.



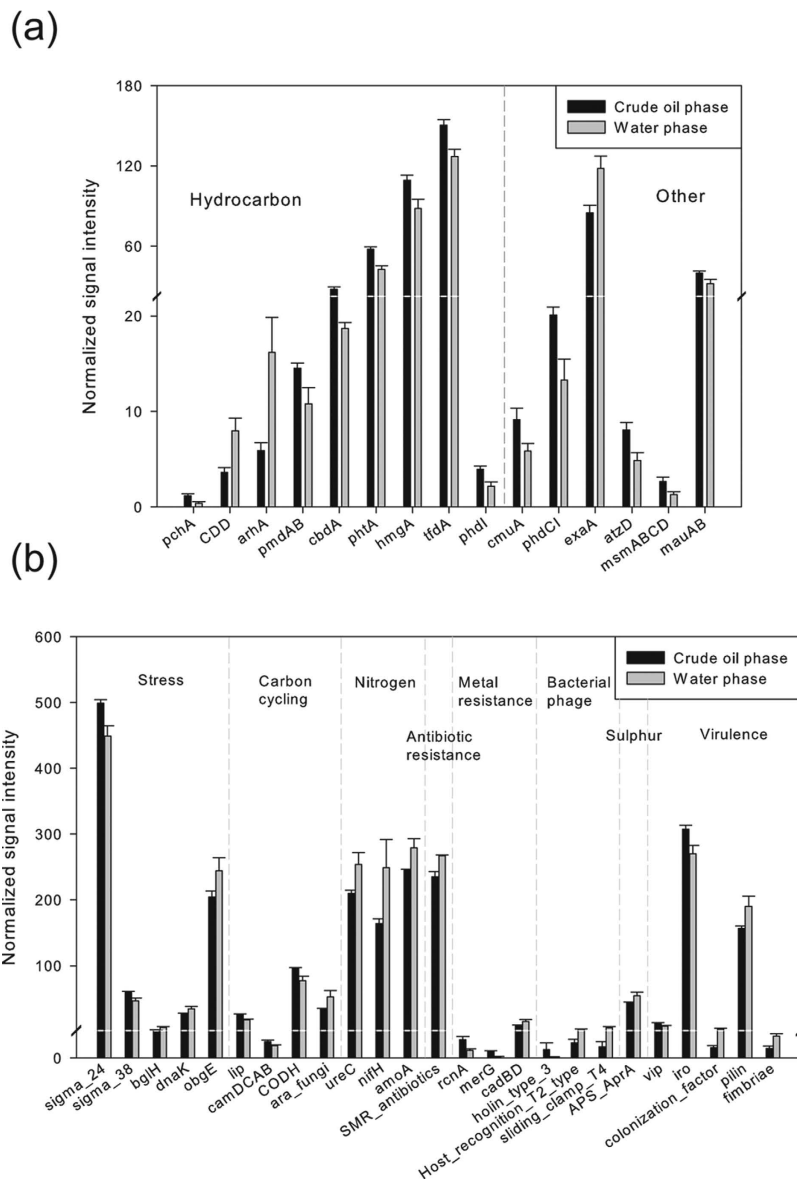
**Figure 2. Hierarchical clustering analysis of functional genes and microbial communities.** Hierarchical clustering analysis of all samples based on relative signal abundances of functional genes in GeoChip hybridization (a) and on the abundance of OTUs obtained by pyrosequencing (b).

## Results

**Geochemical characteristics of samples.** Chemical analyses revealed that concentrations of the total dissolved solids (TDS) of the water (salinity) from the oilfield blocks ranged from 33 to 98 g·L<sup>-1</sup> (Table S1). TDS differed substantially between water samples from different layers of oil-containing rock in oil reservoir (oil strata) where the oil was extracted (Fig. 1). Because the injection water was the same amongst the reservoirs, so the differences of TDS concentrations were reservoir-attributable changes, suggesting that the process of water flooding for 10 years had a minor influence on ground water in different oil strata, which was in consistent with previous results<sup>10,25</sup>. As the oil-containing stratum temperature/well depth increased, weight ratios of aromatic hydrocarbon (ARH) decreased and aliphatic hydrocarbons (ALH) increased (Table S2).

**Functional distribution of microbial communities.** GeoChip 4.0 contains approximately 82,000 probes for targeting >141,995 coding sequences from 410 functional gene families related to microbial carbon (C), nitrogen (N), sulphur (S), and phosphorus (P) cycling, energy metabolism, antibiotic resistance, metal resistance/reduction, organic remediation, stress responses, bacteriophage and virulence<sup>26</sup>. GeoChip 4.0 hybridization experiments led to detection of total of 35,017 genes from 397 gene families in the nine samples. Hierarchical clustering analysis based on functional gene abundances showed that the samples clustered into two groups. One group contained all four samples from oil phase, and another group contained all four samples from water phase, which suggested that microbial functional diversity was different between the oil and water phases (Fig. 2a).

*Functional gene distribution differences between the oil and water phases.* Unexpectedly, alpha diversities of detected functional genes, as examined by the Shannon-Weiner index, were significantly ( $P < 0.05$ ) higher in the oil phase ( $9.621 \pm 0.074$ , mean  $\pm$  SD) than those in water phase ( $9.438 \pm 0.255$ )



**Figure 3. Significant genes in different living activities between the crude oil and water phase samples.** The normalized signal intensity of the significant ( $P < 0.05$ ) gene in the organic degradation category (a) and stress regulation, carbon cycling, nitrogen and sulfur metabolism, antibiotic resistance, metal resistance, bacteriophage, and virulence categories (b) between the crude oil phase and water phase samples. The signal intensity for the gene was the average of the total signal intensity from all four samples of the same phase.

from the same production well (Table S3). Beta diversities of functional genes between oil phases ( $0.261 \pm 0.045$ ) were lower than those between water phases ( $0.361 \pm 0.077$ ), and between oil phases and water phases ( $0.360 \pm 0.096$ ) (Table S4). The significant difference between the oil and water phases in functional composition was demonstrated by three non-parametric multivariate statistical tests (MRPP, ANOSIM and Adonis; Table S5).

Organic compound degradation genes were highly abundant, with significantly ( $P < 0.05$ ) more abundance in crude oil than in the water phase (Fig. 3a). For example, the average total abundance of *phtA*, *hmgA*, and *pmdAB* genes encoding the first enzymes in the polycyclic aromatic hydrocarbon catabolic pathway that converts hydrocarbons into aromatic carboxylic acids were significantly ( $P < 0.05$ ) higher in the crude oil phase (Fig. 3a). The carbon cycle genes *lip*, *camDCAB* and *CODH*, which encode manganese peroxidase for lignin degradation, camphor monooxygenase, and carbon monoxide oxidase, respectively, were also detected at a significantly ( $P < 0.05$ ) higher abundance in the oil phase (Fig. 3b). By contrast, some genes such as the alkane 1-monoxygenase gene (*alkB*), which encodes a key enzyme in aerobic metabolism of aliphatic hydrocarbons, showed similar abundance in the water and oil phases.

Of the stress response genes, a group of 711  $\sigma^{24}$  genes and 93  $\sigma^{38}$  genes was found to be significantly ( $P < 0.05$ ) more abundant in crude oil than the 685  $\sigma^{24}$  genes and 75  $\sigma^{38}$  genes detected in the water phase (Fig. 3b). Conversely, three other stress-related genes, *bglH* (encoding a beta-glucosidase), *dnaK* (encoding a heat shock protein), and *obgE* (encoding a GTPase), had significantly ( $P < 0.05$ ) higher abundance in water than in the crude oil phase (Fig. 3b). Furthermore, genes involved in nitrogen and sulfur metabolism (including *ureC*, *nifH*, *amoA*, and *APS\_AprA*, which encode urease for ammonification, nitrogenase for nitrogen fixation, ammonia monooxygenase for ammonia oxidation, and adenylylsulfate reductase for sulfate reduction, respectively), showed significantly ( $P < 0.05$ ) higher abundance in water than in the oil phase (Fig. 3b). A total of 38 bacteriophage genes were detected, of which the abundances of three genes varied significantly ( $P < 0.05$ ) between the crude oil and water phases: *holin\_type\_3* (higher abundance in the oil phase), *host\_recognition\_T2\_type* (higher abundance in the water phase), and *sliding\_clamp\_T4* (higher abundance in the water phase), which were related to bacterial lysis, host recognition, and replication, respectively (Fig. 3b).

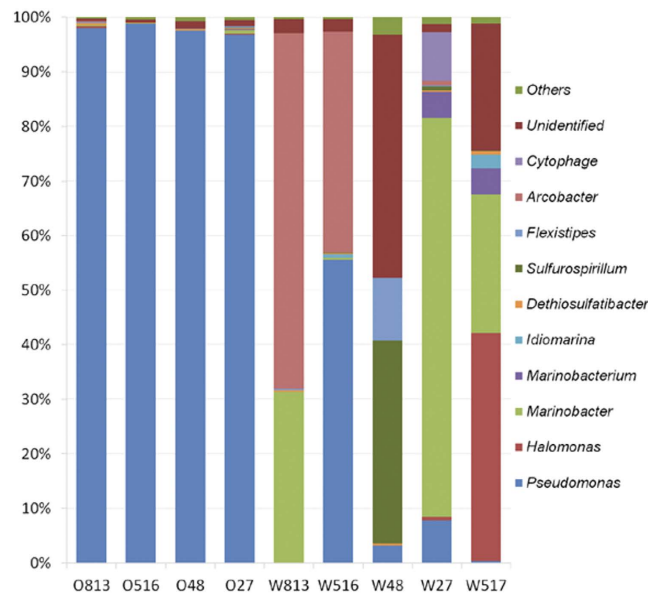
**Differences in functional gene distribution among oil samples.** As the temperature and depth of oil-containing stratum decreased (Fig. 1), the relative abundance of genes related to the degradation of aromatic compounds, carbon cycling, and metabolism of other organic compounds increased concomitantly (by 2.2%, 3.1%, and 15.3%, respectively) (Table S6). For example, the abundance of *nahA*, *HBH* and *pobA* genes involved in the aerobic degradation of benzoate, a common central intermediate in the metabolism of aromatic compounds, increased. The abundance of *hmgA*, which encodes phthalate dioxygenase, also increased (Table S7). In contrast, the increase/decrease of genes abundances was not found among the five water samples as the temperature of oil-containing stratum decreased (Table S6).

Similarly, the relative abundance of genes associated with the stress response, antibiotic resistance, and sulfur metabolism decreased (1.6%, 3.1%, and 1.6%, respectively) with temperature/depth (Table S6). For example, 133 genes, including two *dnaK* genes, which encode the molecular chaperone HSP70 and up-regulated upon heat shock<sup>27</sup>, increased with temperature (Table S8), whereas the abundance of 75 genes, including *hrcA*, which encodes a heat-inducible transcriptional repressor of bacterial heat shock genes<sup>28</sup>, decreased (Table S9).

**Taxonomic distribution of the bacterial community.** A total of 41,792 valid sequences were retrieved from pyrosequencing analysis, including 4,613–6,612 sequences from each oil-phase sample that were assigned to 152–186 bacterial OTUs (>97%), and 2,837–4,154 sequences from each water-phase sample that were assigned to 138–246 bacterial OTUs (Table S10). The coverage values and Chao1 indices (219–452) showed sequencing saturation, thus the number of sequences was adequate to represent the bacterial community in each sample. The Shannon-Wiener indices indicated that bacterial communities were less diverse ( $1.605 \pm 0.093$ ) in oil phases than in water phases ( $3.246 \pm 0.321$ , mean  $\pm$  SD) (Table S10). Hierarchical clustering analysis of OTU abundance clearly separated the samples into two clusters of crude oil and water, suggesting that these microbial communities were markedly different (Fig. 2b). In addition, OTUs common among the injection, oil, and water phases were limited (Fig. S1). Furthermore, MRPP, ANOSIM and Adonis analyses showed the significant difference between the oil and water phases in taxonomic composition, in accordance with the results based on functional gene composition (Table S5).

A total of nine phyla were retrieved from the 41,792 sequences, of which 361 sequences could not be assigned to any known bacterial phyla. *Proteobacteria* was the most abundant phylum in all the oil phases, ranging from 97.86% to 99.02%. By contrast, *Firmicutes* and *Bacteroidetes* accounted for 9.20% and 17.31% of the total sequences in the W27 and W517 (injection) water samples, respectively. *Chloroflexi* was detected in only W516 (0.03%), whereas *Deferribacteres* (11.60%) and an abundant, unidentified bacteria (10.86%) were present in W48 (Table S11). Specifically, *Gammaproteobacteria* dominated the crude oil phase and both *Gammaproteobacteria* and *Epsilonproteobacteria* were generally dominant in the water phases, but with varying abundances (Fig. 4, Table S12).

At the genus level, 50 genera were retrieved from all the samples. Among these genera, 33 and 31 known genera were retrieved from the oil samples and the water samples, respectively. *Pseudomonas* was abundant in the four oil samples, accounting for 96.84–98.87% of the total sequences in each sample (Fig. 4, Table S12). However, it was present at much lower levels in the water samples with a relative abundance of 0.00–7.86%, except for 55.57% in W516. *Acinetobacter* and *Dethiosulfatibacter* were also common in oil phases, but at a very low abundance (Fig. 4, Table S12). Finally, *Marinobacter*, *Halomonas*, and *Flexistipes* were detected in three of the oil samples. The bacteria in water samples were more diverse, without a common dominant representative. The bacteria common to the five water samples were *Sulfurospirillum* (0.03–37.19%) and *Dethiosulfatibacter* (0.16–0.63%) (Fig. 4, Table S12). The major bacterial genera in W813, W516, W48, and W27 were *Arcobacter* (65.13%) and *Marinobacter* (32.16%); *Pseudomonas* (55.57%) and *Arcobacter* (40.50%); *Sulfurospirillum* (37.19%) and *Flexistipes* (11.56%); and *Marinobacter* (73.12%), *Cytophage* (8.92%), *Pseudomonas* (7.86%), and *Marinobacterium* (4.74%), respectively. In the injection water (W517), *Halomonas* (41.86%), *Marinobacter* (25.35%), *Marinobacterium* (4.84%), and *Idiomarina* (2.50%) were the major bacterial genera, accounting for 74.55% of the total number sequences.



**Figure 4. Relative abundance of major bacterial genera detected in each samples.** The relative abundance based on OTU numbers of different genera by pyrosequencing from each crude oil and water phases sample.

Although *Pseudomonas* was detected in both the water and oil phases, detected species were distinct between the water and oil phases (Fig. 5), i.e., those from the oil phase clustered with *P. oleovorans* and *P. alcaliphila*, while those from the water phase clustered with *P. aeruginosa*, *P. balearica*, and *P. citronellolis*.

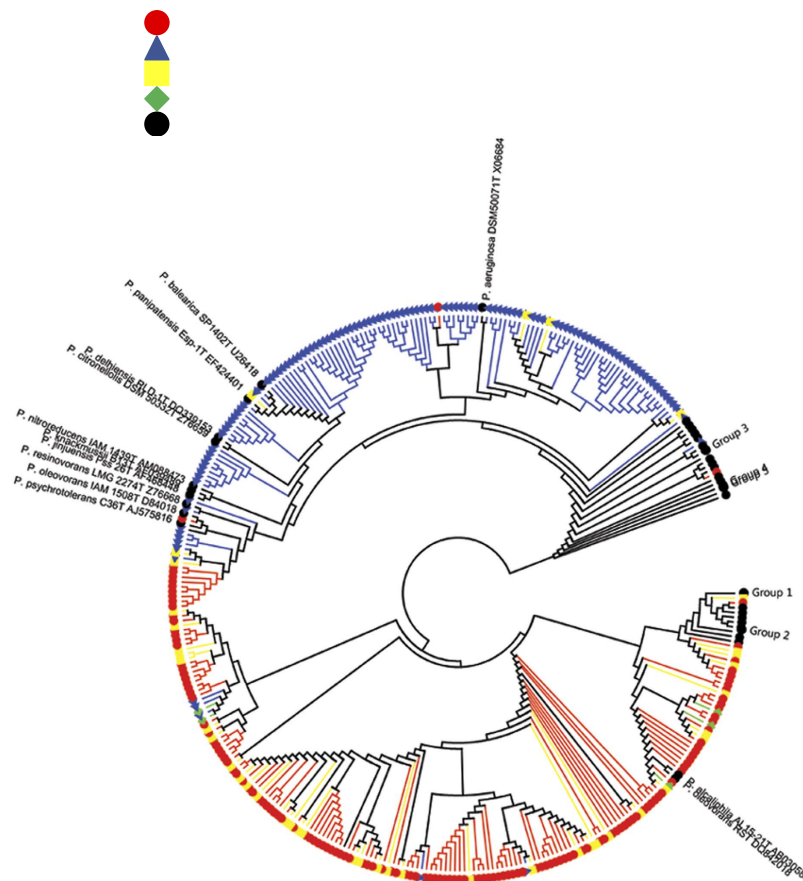
**Association between microbial community and geochemical characteristics.** The links between environmental conditions and microbial taxonomic structures and functional gene compositions were evaluated using canonical correspondence analysis (CCA) (Fig. 6a,b). The key geochemical parameters for both water and crude oil phases included TDS, pH, ALH, ARH, polar fraction with heteroatoms nitrogen, sulfur, and oxygen (NSO), and oil-containing stratum temperature (T), which were selected after various pre-evaluations. Similar to the results of hierarchical clustering using the functional gene abundances or OTU numbers, all of the oil phases were clustered and well separated from the water samples. In addition, the four water phases were separated into different quadrants.

For taxonomic composition, the first axis positively correlated with T and ALH, whereas the second axis positively correlated with ARH and negatively correlated with pH and NSO (Fig. 6a). For functional gene composition, the first axis negatively correlated with ARH, whereas the second axis positively correlated with ALH and T, and negatively correlated with NSO and pH (Fig. 6b).

## Discussion

Despite the marked significance in understanding microorganisms mediated processes in an oilfield, microorganisms in crude oil, their functional profiles and their differences between oil and water phases remain unclear. To our knowledge, this is the first comprehensive study to investigate and compare microbial communities in crude oil and water phases at both the taxonomic and functional gene levels. Our results revealed that microorganisms in crude oils were taxonomically narrow but extremely diverse in functional potentials (Table S3 and S10, and Figs 3 and 5).

Although the microbial communities in the oil phase were dominated by *Pseudomonas*, sequencing results showed that there existed a rare biosphere, which accounted for less than 3.16% of the total OTUs in the oil samples (Table S12). These OTUs spanned from 33 known genera belonging to phyla *Proteobacteria*, *Firmicutes*, *Deferribacteres*, and *Bacteroidetes*. In contrast, although the bacterial communities in the water phase are more diverse than those in the oil phase, the OTUs in the water phase could be only assigned into 31 known genera, which was even less than those in the oil phase. The genera with high abundances such as *Arcobacter*, *Marinobacter* and *Sulfurospirillum* in the water phase could also be found in the oil phase. Moreover, dominant taxa in the water phases were different, which might be caused by the different temperatures, depths and pH, suggesting a rapid growth of these taxa in the water phase under the appropriate environmental conditions. The overlap of taxa also indicated the dispersal of microorganisms between the oil and water phase, although the direction of dispersal seemed to be unclear. However, at least for *Pseudomonas*, most of the shared OTUs by both oil and water phases were embedded in the OTUs clusters of oil phase (Fig. 5), which suggested that the dispersal of these OTUs was most likely from the oil phase to the water phase. *Pseudomonas*, usually a mesophilic microorganisms, has been detected more and more from high temperature oilfields<sup>11,29–31</sup>, and in our

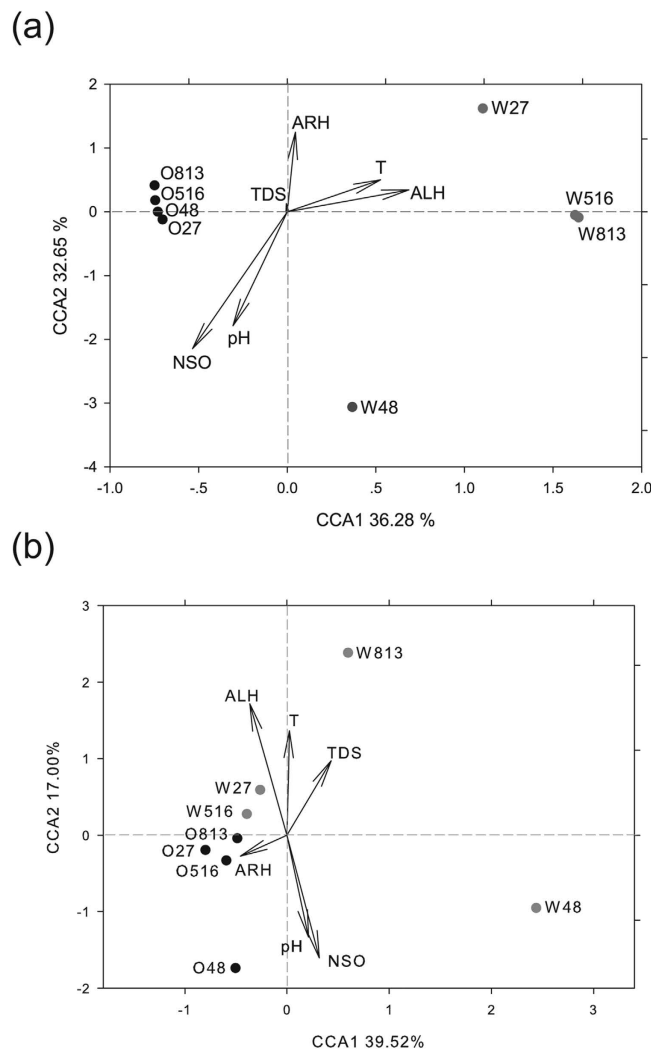


**Figure 5. Phylogenetic relationships of *Pseudomonas* in crude oil and water phases.** Circle topology tree of the phylogenetic relationships between the 308 bacterial 16S rRNA gene sequences obtained from pyrosequencing OTUs in this study and 125 valid *Pseudomonas* species strains from RDP. Bar, 1nt substitution per 100 nt. ● OTUs only from crude oil phases; ▲ OTUs only from oil production water phases; ■ OTUs both from oil and production water phases; ◆ OTUs both from injection water and oil and production water phases; no OTUs only from injection water; ● valid *Pseudomonas* species (all the names of *Pseudomonas* species strains were list in Table S13).

previous investigation, we also detected *Pseudomonas* from autoclaved crude oil<sup>19</sup>. It may be related to the known ability of these organisms to form biofilms<sup>32,33</sup> thus giving them a competitive advantage in high-temperature anhydrous environments. Recent study also suggested that microorganisms in oil lived in minuscule water droplets which provided closed microenvironments to protect them<sup>21</sup>. These results supported that the crude oil could act as a microbial “seed bank” and influence the microbial communities in oil reservoirs. As the sequencing depth is expected to increase due to rapidly expanding sequencing capacity, it is very likely that more taxa will be found in the oil phase, in accordance with relatively high abundance of functional genes found in the oil phase. Because GeoChip is sensitive in detecting genes of low abundance<sup>34</sup>, extremely diverse gene functions were detected, suggesting that crude oil harbors a microbial seed bank with rich functional potential.

Several studies have demonstrated the taxonomic differences between the oil and water phases<sup>12,20</sup>. In this study, both GeoChip and pyrosequencing analyses revealed a significant ( $P < 0.05$ ) difference in microbial communities between the crude oil and water phases, even though they were from the same oil production mixture (Figs 2 and 4; Tables S3 and S4). However, little was known why there were significant taxonomic differences between the oil and water phases. This question could also be explained as what conditions in the water phase promoted the growth or changed the relative abundance of taxa in the communities, if the crude oil served as a “seed bank”. It was possible that different microbial communities were caused by dispersal or growth of rare and dominant taxa from the microbial “seed bank”, which changed the abundance of taxa rather than the presence or absence of different taxa. So different environmental filtering might be important in shaping the communities.

The microorganisms in oil and water phases displayed different correlations with the environmental conditions. CCA analysis revealed that the microbial functional community was closely related to the crude oil components (ARH, ALH, and NSO), the stratum temperature, and pH (Fig. 6b). It is well established that crude oil contains more than 3,000 organic compounds<sup>35</sup>, which provides a wide variety



**Figure 6. Canonical correspondence analysis (CCA) of significant geochemical variables and microbial communities from all production well samples. (a)** CCA of significant geochemical variables (arrows) and relative abundances of bacterial OTUs (symbols). **(b)** CCA of significant geochemical variables (arrows) and relative abundance of functional genes from GeoChip hybridization signals (symbols). Variables: aliphatic hydrocarbons (ALH), aromatic hydrocarbons (ARH), polar fraction with heteroatoms nitrogen, sulfur, and oxygen (NSO) from crude oil, total dissolved solid (TDS) of production water, pH value, and bottom hole temperature of production wells (T).

of niches for microbial communities and allows for gene drift among microbial groups<sup>36</sup>. However, high toxicity and hydrophobicity impose strong natural selection on these microorganisms<sup>37,38</sup>, resulting in highly uneven microbial communities that could serve as a microbial seed bank. Remarkably, our previous results showed that *Pseudomonas*, *Bacillus*, and *Alcaligenes* survived autoclaving in crude oil and could be stimulated by *Dietzia* and *Acinetobacter* strains<sup>19</sup>. *Pseudomonas* and *Bacillus* eventually outcompeted other bacteria and played the most important roles in crude oil degradation.

The oil phase contained a higher abundance of genes involved in aromatic degradation and carbon cycling compared to the water phase. Crude oil contains more carbon sources such as alkanes and aromatic hydrocarbons than the water phase, though these substrates are hydrophobic<sup>12</sup>. Of the bacteria detected by pyrosequencing, *Pseudomonas* has a known ability to degrade oil aromatic compounds using the enzymes encoded by *phtA*, *hmgA*, and *pmdAB*<sup>39–41</sup>. Moreover, microorganisms could produce surface-active materials and changing their cell surface hydrophobic activities to access hydrophobic oil constituents. Some *Pseudomonas* spp., e.g., *P. aeruginosa*, can produce biosurfactants such as rhamnolipids<sup>42</sup> that help access crude oil. This allowed *Pseudomonas* to attach to crude oil and utilize the aromatic carbon sources, explaining the significantly ( $P < 0.05$ ) higher abundance of aromatic degradation-related genes in the oil phase (Fig. 3a). In contrast to the oil phase, the water phase contained more oxygenated and sulfonated molecules from biodegradation of crude oil, such as carboxylic acids<sup>21</sup>. As the results, it was unsurprising that we found many microorganism that were not oil degraders in the water phase.



The different environmental conditions could also explain the differences among the water samples from four wells. The water phase samples from the Qinghai Oilfield had high TDS contents ranging from 33–98 g·L<sup>-1</sup>, which promotes the growth of halophilic organisms such as *Halomonas* and *Marinobacter*<sup>43,44</sup>. Consistently, non-halophilic *Pseudomonas* was the most abundant microbial group in W516, which had the lowest salinity/TDS ratio (33.2 g·L<sup>-1</sup>). However, significantly ( $P < 0.05$ ) more genes related to stress response, e.g.,  $\sigma^{24}$  and  $\sigma^{28}$ , were found in the oil phase (Fig. 3b), which cannot be attributed to salinity. Rather, high hydrophobicity of crude oil causes high osmotic pressure, which might lead to more stress response genes in oil phase.

It was interesting that more abundant host recognition- and replication-related bacteriophage genes, such as *host\_recognition\_T2\_type* and *sliding\_clamp\_T4*, were detected in the water phase than those in the oil phase (Fig. 3b). Because the water samples were filtered with 0.22- $\mu$ m membranes, it could be imagined that free phages could hardly remain in the samples. The detection of these phage-related genes might be possibly from the microbial genomes, which might be the traits of the phage infection events during the evolution. Recent study suggested that microorganisms in oil lived in minuscule water droplets, and the high salinity and water-stable isotopes of the droplets indicated a deep subsurface origin<sup>21</sup>, raising a possibility that water droplets and their microbial inhabitants might originate from the ancient seawater or brines. The water droplets provided closed environments and protected the migration of microorganisms from the outside intrusions such as water injection to the oil embedded water droplets. In contrast, the production water was more prone to dispersal and migration and could be influenced by the injection water or the surface water<sup>45</sup>, as well as by the oil phase in the oil-water transition zone. The more closed microenvironments in the oil phase might protect the inner microbes from phage infections, which could explain the higher abundance of bacteriophage genes in the water phase. This discovery was also consistent with the important role of phage in phage-driven microbial life cycles under the condition of oil-polluted water treatment<sup>46</sup>. However, further investigations are still needed to clarify the hypotheses.

In summary, despite the wide belief that oil is a harsh habitat, here we show that oil harbors unexpectedly rich microbial functional potentials. This is the first study to demonstrate that crude oil could be a “seed bank” of functional microorganisms enriched for hydrocarbon degradation-related capacities. Therefore, a microbial community related to crude oil should be considered when evaluating the natural crude oil attenuation process, as well as when developing and applying technologies for bioremediation and MEOR.

## Methods

**Site and sampling.** We collected the injection water and production mixture samples from Yue-II block of the Qinghai Oilfield, which is located in the Tibetan Plateau, northwestern China. This block had been subjected to water flooding by the local oil company since 1999. In this work, we selected four oil wells (numbers 813, 516, 48, and 27) which were connected to four different oil-containing rock strata with different depth, temperature, and salinity (Table S1 and S2), as well as their common injection well (number 517) for sampling (Fig. 1). The distances between the four production wells and the injection well ranged from 0.2 to 0.5 km. For each oil well or injection well, we collected production mixtures or the injection water from the sampling valve on the well head into sterile bottles, and filled the bottles to exclude oxygen. Then we sealed the bottles with sterile screw caps and sent them immediately back to the laboratory within 48 h below 5 °C. After receiving, we settled all the bottles filled with production mixtures with gravitational precipitation<sup>12</sup> for 12 h at 4 °C to facilitate separation of the oil/water phases. Subsequently, we collected the production water layer and oil layer from each bottle as four water phase samples (W813, W516, W48, and W27, corresponding to the four wells), and oil phase samples (O813, O516, O48, and O27), respectively. The injection water sample (W517) was subjected to direct analysis.

**Chemical analysis.** Four sub-fractions comprising aliphatic hydrocarbons (ALH), aromatic hydrocarbons (ARH), a polar fraction with nitrogen-, sulfur- and oxygen-containing compounds (NSO), and asphaltenes (ASP) were separated from the four crude oil samples using liquid-solid chromatography<sup>47</sup>. We weighed them to calculate the weight ratios of the fractions, then measured contents of the different elements in the crude oil using a Vario EL III elemental analyzer (Elementar Analysensysteme GmbH, Germany) for carbon and hydrogen, and a multi EA3100 elemental analyzer (Analytik Jena AG, Germany) for sulfur and nitrogen. After filtration through 0.22  $\mu$ m cellulose ester membrane filters, we measured the concentrations of Na<sup>+</sup>, K<sup>+</sup>, Mg<sup>2+</sup>, Ca<sup>2+</sup>, Cl<sup>-</sup>, and SO<sub>4</sub><sup>2-</sup> in the water samples using cation and anion columns (IC-C4 and IC-A3[s], Shimadzu, Japan) in PIA-1000 ion analyzer (Shimadzu, Japan), according to the manufacturer's instructions. Then total dissolved solids (TDS) was calculated as sum of concentrations of all the tested ions.

**DNA extraction, GeoChip hybridization, and pyrosequencing.** After separation, we collected the microbial cells of the injection water and the four water phase samples by filtering them through 0.22- $\mu$ m cellulose ester membrane filters. Meanwhile, we collected the microbial cells from the four oil-phase samples as described previously<sup>18</sup> with several modifications. Briefly, we mixed 15 mL of the oil with 15 mL of sterile isooctane (2, 2, 4-trimethylpentane) and settled the mixture overnight at 4 °C. Then we vortexed (1000 rpm) the mixture for 5 min and centrifuged at 5,000  $\times$  g for 1 h at 4 °C. After resuspension of the

precipitants in 20 mL isoctane, centrifuged at  $5,000 \times g$  three times, and lyophilized. We extracted and purified microbial genomic DNA from each of these portions as previously described<sup>48,49</sup>. Because the samples were kept at low temperature for less than 12 hours (overnight), the nutrients in the samples, such as nitrogen and phosphorus, were low, and the hydrocarbons were not easily degradable as carbon sources at such low temperature and short time, we don't think that there were major changes in the microbial communities by isoctane addition and subsequent steps. DNA quality and quantity were measured using NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA) and with PicoGreen using a FLUOstar Optima (BMG Labtech, Jena, Germany), respectively<sup>26</sup>. For microbial functional diversity study, we used GeoChip 4.0 to dissect the microbial community functional gene composition as described elsewhere<sup>50</sup>. We labeled DNA samples with the fluorescent dye Cy-3 using a random priming method and purified using the QIA quick purification kit (Qiagen, Valencia, CA, USA). And then dried DNA in a SpeedVac (ThermoSavant, Milford, MA, USA) at 45 °C for 45 min. The hybridization was carried out at 42 °C for 16 h on a MAUI hybridization station (BioMicro, Salt Lake City, UT, USA). After purification, we scanned GeoChip microarrays using a NimbleGen MS200 scanner (Roche, Madison, WI, USA) at 633 nm using a laser power and photomultiplier tube gain of 100% and 75%, respectively<sup>26</sup>. For microbial composition study, we used purified DNA to amplify the bacterial hypervariable V3 region of the 16S rRNA gene (at *Escherichia coli* 16S rRNA gene position 341–534 bp) to construct a community library using tag pyrosequencing (for more details, please see Supplementary Methods).

**Statistical analysis.** For GeoChip analysis, we quantified and processed signal intensities using the data analysis pipeline as previously described<sup>27,49,51</sup>. After removing the poor quality data with ratio of signal to noise lower than 2.0, we normalized the signal intensity of each probe as described previously<sup>49</sup>. These normalized data were relative functional gene abundances or relative signal intensities which were then used for further analysis. We examined beta diversity to analyze the dissimilarity of two samples using the quantitative Bray-Curtis index<sup>13,52</sup>. To examine whether there is a significant difference of functional gene composition between oil phase and water phase, we used three different complementary non-parametric analyses for multivariate data: multi-response permutation procedure (MRPP) (McCune, B. *et al.* Analysis of ecological communities. MJM Software Design: Gleneden Beach, OR, 2002), analysis of similarities (ANOSIM)<sup>53</sup>, and permutational multivariate analysis of variance using distance matrices (Adonis)<sup>54</sup>. Euclidean distance was used to calculate the distance matrix for MRPP, ANOSIM and Adonis analyses. All three methods are based on dissimilarities among samples and their rank order to calculate test statistics. The Monte Carlo permutation was used to test the significance of statistics<sup>55</sup>. All three procedures (MRPP, ANOSIM and Adonis) were performed using the Vegan package v 2.0–4 (Oksanen, J. *et al.* Package Vegan: Community ecology package, version 2.0–4, 2012) in R v. 2.14.0 (RDC Team, 2011).

For pyrosequencing analysis, we removed the low-quality reads shorter than 150 base pairs, with mismatched primer and barcode and with more than one ambiguous nucleotide within correct barcodes or primers. Then we processed and analyzed the data as previously described<sup>56</sup>, and compared them with the Bacterial SILVA database (SILVA version 106, 2011, <http://www.arb-silva.de/documentation/release-106/>). The OTU was defined by 97% similarity<sup>56</sup>. Venn diagram was used to reflect the bacterial community relationship among production waters, crude oils and injection water. To reveal the diversity and distribution of *Pseudomonas* in oil field samples, we constructed phylogenetic trees based on partial 16S rRNA gene sequences of representative *Pseudomonas* OTUs from this study and *Pseudomonas* type species from RDP (Ribosomal Database Project, 1998, <http://rdp.cme.msu.edu/>) using the MEGA program, version 5.0<sup>57</sup>.

We calculated the diversity indices including richness, Shannon–Weiner index and Reciprocal of Simpson's index on the basis of OTUs or functional genes. To understand association between microbial community and geochemical characteristics, we performed canonical correspondence analysis (CCA) to examine whether the geochemical characteristics impacted the microbial communities based on the abundance of OTUs or functional genes, using the Vegan package v 2.0–4 in R v. 2.14.0 (RDC Team, 2011), respectively. To examine overall patterns of variation among the samples, we did hierarchical cluster analysis of functional gene abundance (i.e. normalized signal intensity from GeoChip analysis) or OTU numbers using the Pearson distance method and the complete linkage clustering algorithm with CLUSTER (Gene Cluster 3.0, 2002, <http://rana.stanford.edu/>), and visualized the results in TREEVIEW (TreeView software, 2002, <http://rana.stanford.edu/>)<sup>51</sup>.

## References

1. Lenchi, N. *et al.* Diversity of microbial communities in production and injection waters of Algerian Oilfields revealed by 16S rRNA gene amplicon 454 pyrosequencing. *PLoS One* **8**, e66588 (2013).
2. Singer, M. E. & Finnerty, W. R. in *Petroleum Microbiology*, (ed. Atlas, R. M.) 1–59 (Collier MacMillan, 1984).
3. Amani, H., Sarrafzadeh, M., Haghghi, M. & Mehrnia, M. Comparative study of biosurfactant producing bacteria in MEOR applications. *J. Petrol. Sci. Eng.* **75**, 209–214 (2010).
4. Nazina, T. N. *et al.* Results of the trial of the microbiological method for the enhancement of oil recovery at the carbonate collector of the Romashkinskoe oil field: biogeochemical and productional characteristics. *Microbiology* **68**, 222–226 (1999).
5. Liu, J. *et al.* The field pilot of microbial enhanced oil recovery in a high temperature petroleum reservoir. *J. Petrol. Sci. Eng.* **48**, 265–271 (2005).

6. Grabowski, A., Nercessian, O., Fayolle, F., Blanchet, D. & Jeanthon, C. Microbial diversity in production waters of a low-temperature biodegraded oil reservoir. *FEMS Microbiol. Ecol.* **54**, 427–443 (2005).
7. Dahle, H., Garshol, F., Madsen, M. & Birkeland, N. K. Microbial community structure analysis of produced water from a high-temperature North Sea oil-field. *Anton. Leeuw. Int. J. G.* **93**, 37–49 (2008).
8. Sette, L. D. *et al.* Oliveira. Analysis of the composition of bacterial communities in oil reservoirs from a southern offshore Brazilian basin. *Anton. Leeuw. Int. J. G.* **91**, 253–266 (2007).
9. Kraan, G. M., Bruining, G., Lomans, B. P., Van Loosdrecht, M. C. & Muyzer, G. Microbial diversity of an oil-water processing site and its associated Oil field: the possible role of microorganisms as information carriers from oil-associated environments. *FEMS Microbiol. Ecol.* **71**, 428–443 (2010).
10. Ren, H. Y. *et al.* Comparison of microbial community compositions of injection and production well samples in a long-term water-flooded petroleum reservoir. *PLoS One* **6**, e23258 (2011).
11. Tang, Y. Q. *et al.* Microbial communities in long-term, water-flooded petroleum reservoirs with different *in situ* temperatures in the Huabei Oilfield, China. *PLoS One* **7**, e33535 (2012).
12. Kryachko, Y., Dong, X., Sensen, C. W. & Voordouw, G. Compositions of microbial communities associated with oil and water in a mesothermic oil field. *Anton. Leeuw. Int. J. G.* **101**, 493–506 (2012).
13. Youssef, N. H., Couger, M. B. & Elshahed, M. S. Fine-scale bacterial beta diversity within a complex ecosystem (Zodlone Spring, OK, USA): the role of the rare biosphere. *PLoS One* **5**, e12414 (2010).
14. Gieg, L. M., Davidova, I. A., Duncan, K. E. & Suflita, J. M. Methanogenesis, sulfate reduction and crude oil biodegradation in hot Alaskan oilfields. *Environ. Microbiol.* **12**, 3074–3086 (2010).
15. Guan, J. *et al.* Mu. Functional genes (*dsr*) approach reveals similar sulphidogenic prokaryotes diversity but different structure in saline waters from corroding high temperature petroleum reservoirs. *Appl. Microbiol. Biotechnol.* **98**, 1871–1882 (2013).
16. Atlas, R. M. & Atlas, M. C. Biodegradation of oil and bioremediation of oil spills. *Curr. Opin. Biotechnol.* **2**, 440–443 (1991).
17. Yoshida, N. *et al.* Bacterial communities in petroleum oil in stockpile. *J. Biosci. Bioeng.* **99**, 143–149 (2005).
18. Yamane, K. *et al.* Diversity and similarity of microbial communities in petroleum crude oils produced in Asia. *Biosci. Biotechnol. Biochem.* **72**, 2831–2839 (2008).
19. Gong, X. C. *et al.* Bacteria in crude oil survived autoclaving and stimulated differentially by exogenous bacteria. *PLoS One* **7**, e40842 (2012).
20. Kobayashi, H. *et al.* Phylogenetic diversity of microbial communities associated with the crude-oil, large- insoluble-particle and formation-water components of the reservoir fluid from a non-flooded high-temperature petroleum reservoir. *J. Biosci. Bioeng.* **113**, 204–210 (2012).
21. Meckenstock, R. U. *et al.* Water droplets in oil are microhabitats for microbial life. *Science* **345**, 673–676 (2014).
22. Lennon, J. T. & Jones, S. E. Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat. Rev. Microbiol.* **9**, 119–130 (2011).
23. Gibbons, S. M. *et al.* Evidence for a persistent microbial seed bank throughout the global ocean. *Proc. Natl. Acad. Sci. USA.* **110**, 4651–4655 (2013).
24. Zhou, J. *et al.* Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. *Proc. Natl. Acad. Sci. USA.* **111**, E836–E845 (2014).
25. Pham, V. D. *et al.* Characterizing microbial diversity in production water from an Alaskan mesothermic petroleum reservoir with two independent molecular methods. *Environ. Microbiol.* **11**, 176–87 (2009).
26. Tu, Q. *et al.* GeoChip 4: a functional gene-array-based high-throughput environmental technology for microbial community analysis. *Mol. Ecol.* **14**, 914–928 (2014).
27. Crapoulet, N., Barbry, P., Raoult, D. & Renesto, P. Global transcriptome analysis of *Tropheryma whipplei* response to temperature stresses. *J. Bacteriol.* **188**, 5228–5239 (2006).
28. Narberhaus, F. Negative regulation of bacterial heat shock genes. *Mol. Microbiol.* **31**, 1–8 (1999).
29. Li, H., Yang, S. Z., Mu, B. Z., Rong, Z. F. & Zhang, J. Molecular phylogenetic diversity of the microbial community associated with a high-temperature petroleum reservoir at an offshore oilfield. *FEMS Microbiol. Ecol.* **60**, 74–84 (2007).
30. Orphan, V. J., Taylor, L. T., Hafenbradl, D. & Delong, E. F. Culture-dependent and culture-independent characterization of microbial assemblages associated with high-temperature petroleum reservoirs. *Appl. Environ. Microbiol.* **66**, 700–711 (2000).
31. Korenblum, E., Souza, D. B., Penna, M. & Seldin, L. Molecular analysis of the bacterial communities in crude oil samples from two Brazilian offshore petroleum platforms. *Int. J. Microbiol.* 156537 (2012).
32. Klausen, M., Gjermansen, M., Kreft, J. U. & Tolker-Nieslen, T. Dynamics of development and dispersal in sessile microbial communities: examples from *Pseudomonas aeruginosa* and *Pseudomonas putida* model biofilms. *FEMS Microbiol. Lett.* **261**, 1–11 (2006).
33. Li, M. Y., Zhang, J., Lu, P., Xu, J. L. & Li, S. P. Evaluation of biological characteristics of bacteria contributing to biofilm formation. *Pedosphere* **19**, 554–561 (2009).
34. He, Z., *et al.* GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *ISME J.* **4**, 1167–1179 (2010).
35. Qian, K. *et al.* Resolution and identification of elemental compositions for more than 3000 crude acids in heavy petroleum by negative-ion microelectrospray high-field fourier transform ion cyclotron resonance mass spectrometry. *Energ. Fuel.* **15**, 1505–1511 (2001).
36. Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C. & Martiny, J. B. H. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat. Rev. Micro.* **10**, 497–506 (2012).
37. Van Hamme, J. D. Singh, A. & Ward, O. P. Recent advances in petroleum microbiology. *Microbiol. Mol. Biol. Rev.* **67**, 503–549 (2003).
38. Liu, Y. *et al.* Cell hydrophobicity is a triggering force of biogranulation. *Enzyme. Microb. Tech.* **34**, 371–379 (2004).
39. Nomura, Y., Nakagawa, M., Ogawa, N., Harashima, S. & Oshima, Y. Genes in PHT plasmid encoding the initial degradation pathway of phthalate in *Pseudomonas putida*. *J. Ferment. Bioeng.* **74**, 333–344 (1992).
40. Mathee, K. *et al.* Dynamics of *Pseudomonas aeruginosa* genome evolution. *Proc. Natl. Acad. Sci. USA* **105**, 3100–3105 (2008).
41. Yun, S. H., Yun, C. Y. & Kim, S. I. Characterization of protocatechuate 4, 5-dioxygenase induced from p-hydroxybenzoate-cultured *Pseudomonas* sp. K82. *J. Microbiol.* **42**, 152–155 (2004).
42. Gunther, N. W., Nuñez, A., Fett, W. & Solaiman, D. K. Y. Production of rhamnolipids by *Pseudomonas chlororaphis*, a nonpathogenic bacterium. *Appl. Environ. Microbiol.* **71**, 2288–2293 (2005).
43. Vreeland, R. H., Litchfield, C. D., Martin, E. L. & Elliot, E. *Halomonas elongata*, a new genus and species of extremely salt-tolerant bacteria. *Int. J. Syst. Bacteriol.* **30**, 485–495 (1980).
44. Liebgott, P. P., Casalot, L., Paillard, S., Lorquin, J. & Labat, M. *Marinobacter vinifirmus* sp. nov., a moderately halophilic bacterium isolated from a wine-barrel-decalcification wastewater. *Int. J. Syst. Evol. Microbiol.* **56**, 2511–2516 (2006).
45. Aitken, C. M., Jones, D. M. & Larter, S. R. Anaerobic hydrocarbon biodegradation in deep subsurface oil reservoirs. *Nature* **431**, 291–294 (2004).
46. Rosenberg, E. *et al.* The phage-driven microbial loop in petroleum bioremediation. *Microb. Biotechnol.* **3**, 467–472 (2010).

47. Chaillan, F. *et al.* Identification and biodegradation potential of tropical aerobic hydrocarbon-degrading microorganisms. *Res. Biomicro.* **155**, 587–595 (2004).
48. Zhou, J., Bruns, M. A. & Tiedje, J. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**, 316–322 (1996).
49. Yang, Y. *et al.* The microbial gene diversity along an elevation gradient of the Tibetan grassland. *ISME J.* **8**, 430–440 (2014).
50. Yang, Y. *et al.* Responses of the functional structure of soil microbial community to livestock grazing in the Tibetan alpine grassland. *Glob. Change. Biology* **19**, 637–648 (2013).
51. Lu, Z. *et al.* Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. *ISME J.* **6**, 451–460 (2012).
52. Bray, J. R. & Curtis, J. T. An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* **27**, 325–349 (1957).
53. Clarke, K. R. Nonparametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* **18**, 117–143 (1993).
54. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral. Ecol.* **26**, 32–46 (2001).
55. Zhou, J. *et al.* Reproducibility and quantitation of amplicon sequencing-based detection. *ISME J.* **5**, 1303–1313 (2011).
56. Gu, S. *et al.* Bacterial community mapping of the mouse gastrointestinal tract. *PLoS One* **8**, e74957 (2013).
57. Tamura, K. *et al.* MEGA 5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).

## Acknowledgements

The study was supported by National Natural Science Foundation of China (31225001) and National High Technology Research and Development Program (2012AA02A703, 2014AA021505).

## Author Contributions

C.M., Y.Q.T. and X.L.W. designed research; C.M. performed research; X.B.W., N.Y., Z.S.L. and Y.L. contributed new reagents/analytic tools; C.M., C.Q.C., Y.Y., J.Z. and X.L.W. analyzed data; C.M., N.Y., Y.Y. and X.L.W. wrote the paper. All authors reviewed the manuscript.

## Additional Information

**Accession codes:** The GeoChip data of all samples have been deposited in GEO repository of NCBI under the accession number GSE55293, and the pyrosequencing sequences of all samples have been deposited in NCBI Sequencing Read Archive under the accession number SRP037706.

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Cai, M. *et al.* Crude oil as a microbial seed bank with unexpected functional potentials. *Sci. Rep.* **5**, 16057; doi: 10.1038/srep16057 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>