Article

Microbiome assessment of global oil reservoirs reveals site-specific hydrocarbon degradation functional profiles

Kelly J. Hidalgo [1,2], *, Isabel N. Sierra-Garcia [3], German Zafra [4] and Valéria Maia de Oliveira [1]

[1] Microbial Resources Division, Research Center for Chemistry, Biology and Agriculture (CPQBA), University of Campinas – UNICAMP, Paulínia, Brazil, Av. Alexandre Cazellato, 999, ZIP 13148-218; khidalgo@javeriana.edu.co

[2] Graduate Program in Genetics and Molecular Biology, Institute of Biology, University of Campinas (UNICAMP), Rua Monteiro Lobato 255, Cidade Universitária, Campinas, SP. ZIP 13083-862, Brazil

[3] Biology Department & CESAM, University of Aveiro, Aveiro, Portugal, Campus de Santiago, Avenida João Jacinto de Magalhães, 3810-193 Aveiro, Portugal; inatalia.sierra@gmail.com

[4] Grupo de Investigación en Bioquímica y Microbiología (GIBIM), Escuela de Microbiología, Universidad Industrial de Santander, Bucaramanga, Colombia; geralzaf@uis.edu.co

* Correspondence: khidalgo@javeriana.edu.co; Tel.: +55 19981721510

Abstract: Microorganisms inhabiting subsurface petroleum reservoirs are key players in biochemical transformations. The interactions of microbial communities in these environments are highly complex and still poorly understood. This work aimed to assess publicly available metagenomes from oil reservoirs and implement a robust pipeline of genome-resolved metagenomics to decipher metabolic and taxonomic profiles of petroleum reservoirs worldwide. Analysis of 301,2 Gb of metagenomic information derived from heavily flooded petroleum reservoirs in China and Alaska to non-flooded petroleum reservoirs in Brazil enabled us to reconstruct 148 MAGs of high and medium quality. At the phylum level, 74% of MAGs belonged to bacteria and 26% to archaea. The profiles of these MAGs were related to the physicochemical parameters and recovery management applied. The analysis of the potential functional core in the reservoirs showed that the microbiota was specialized for each site, with 31.7% of the total KEGG orthologies annotated as functions (1,690 genes) common to all oil fields, while 18% of the functions were site-specific, i.e., present only in one of the oil fields. The oil reservoirs with lower level of intervention were the most similar to the potential functional core, while the oil fields with longer history of water injection had greater variation in functional profile. These results show how key microorganisms and their functions respond to the distinct physicochemical parameters and interventions of the oil field operations such as water injection and expand the knowledge of biogeochemical transformations in these ecosystems.

Keywords: petroleum; metagenomics; core microbiota, functional potential, oil fields

1. Introduction

Oil reservoirs are unique subsurface environments often located deep in the earth, characterized by a complex matrix of hydrocarbons, low oxygen levels, and a long history of separation from the surface [1]. Currently, while a plethora of bacteria and archaea have been detected in petroleum reservoirs, some patterns in microbial taxonomic composition have been identified that correlate to specific environmental factors such as temperature and depth, with little role for the geographic distance between petroleum reservoirs worldwide [2,3]. Shallow and low-temperature reservoirs show an abundance of Epsilonbacteria and Deltaproteobacteria members, while deeper high-temperature reservoirs are enriched in Clostridiales and Thermotogales. Despite these specific taxa that distinguish oil reservoirs according to their attributes, a core microbiome composition across the world has been identified, showing that the bacterial classes Gammaproteobacteria,

Clostridia and Bacteroidia and the archaeal class Methanomicrobia are ubiquitous in petroleum reservoirs [4].

There have been several studies using metagenomics aiming to provide insights into the functional potential of microorganisms in oil field environments [5-7]. Pioneering studies comparing metagenomes from oil reservoirs with metagenomes from hydrocarbon-rich or non-hydrocarbon-rich environments reported higher abundances of genes related to anaerobic hydrocarbon metabolism and methanogenesis in samples from oil reservoirs [5,8]. However, the metabolic capability of individual organisms and intra-community interactions remains largely unresolved due to a lack of enough genome information of biochemical transformations in oil reservoirs. Emerging approaches to recover metagenome-assembled genomes (MAGs) from metagenomic data [8-10] allow the recovery of microbial genomes and access the functional potential of keystone taxa, including the uncultivated ones, and their interactions.

At present, several metagenomic datasets derived from oil reservoirs located in China [2,11,12], Denmark [13], Alaska [6] and Brazil [3] have been added to public repositories. Most of these metagenomic sequence data were analyzed as isolated niche studies, however no analysis has been carried out so far on the universal patterns or core functions of microorganisms in such deep environments. The understanding of the functional role and specificity of microorganisms inhabiting the subsurface biosphere requires the integration of the metagenomic data using novel tools.

Here, we aimed to integrate the publicly available metagenomic data from oil reservoirs in order to deeply characterize community composition and metabolic profiles in petroleum reservoirs worldwide. The questions that guided our work were: *i) Is there a potential functional core shared by microorganisms in oil reservoirs worldwide? ii) Are there metabolic processes that are specific to particular petroleum reservoirs? iii) Are pathways for hydrocarbon degradation in oil reservoirs different from the ones in other environments?* The answers to these questions are important for the comprehension of metabolic capabilities, biogeochemical cycles and geological resources in deep subsurface environments.

2. Materials and Methods

2.1. Data Acquisition

Seventeen datasets derived from publicly available shotgun metagenomes associated to oil fields were downloaded and used in this study. These data correspond to 17 samples distributed in six studies [1-3,6,11,12]. Metagenome data from oil reservoirs with different recovery management and temperatures located in different regions were included in order to investigate the influence of geographic location, temperature and anthropogenic interventions on functional and taxonomic profiles of the indigenous microbial communities through the recovery of petroleum-associated Metagenome Assembled Genomes (MAGs). All sequence data were obtained from the National Center for Biotechnology Information (NCBI). The sample accession number, location, type of sample and related references is listed in the Supplementary Table S1.

2.1.1. Description of datasets

The geographic distribution of the metagenomes analyzed and the amount of data in Gigabytes from each site were plot in Figure 1. Five oil fields from China comprising 10 metagenomes and almost 69 Gb of information were analyzed [1,2,11,12]. Three formations in Alaska with 6 metagenomes and 216 Gb [6] were analyzed. Finally, one metagenome with 16 Gb of data from a Brazilian oil field was also included in the meta-analysis [3].
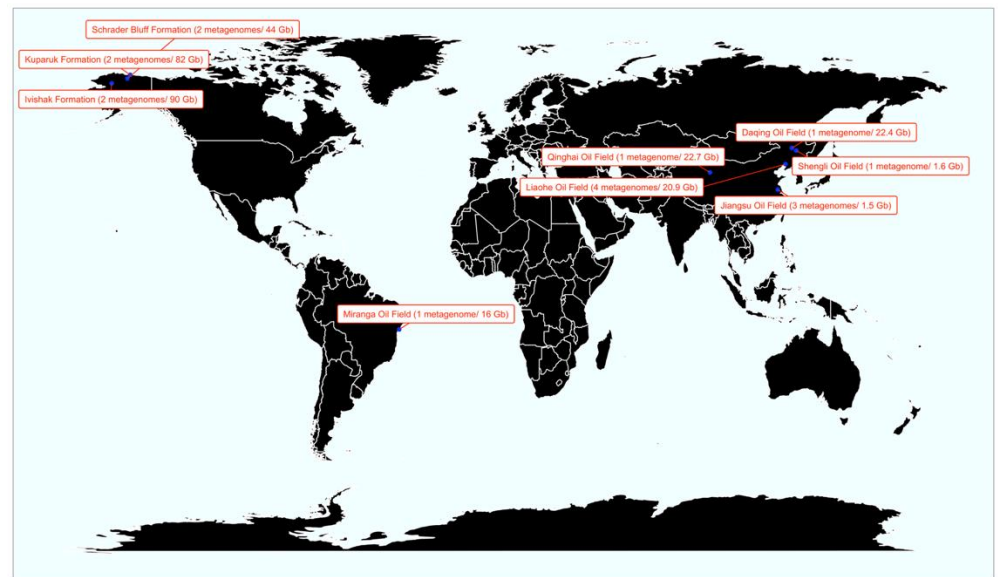
Figure 1. Geographic distribution of metagenomes. The number of metagenomes and amount of information in gigabytes from each site is in parenthesis.

All downloaded metagenome datasets were submitted to the binning pipeline approach established in this work (Figure 2) in order to recover MAGs. The study performed by Liu and colleagues [11] described the metabolic potential and activity of the microbial communities obtained from the Jiangsu Oil Reservoir (China). They sampled produced water from three different wells, named W2-71, W9-18 and W15-5 and sequenced in Illumina MiSeq platform (2 x 75 bp). Hu and collaborators (2016) used metagenomic shotgun sequencing of six samples from two North Slope oil fields in Alaska to compare the microbial community across the range of physical and chemical conditions and to predict metabolic roles. The produced water samples were obtained from different depths and temperatures, with or without hydrogen sulfide production (souring), and with or without impact by sea water injection. Samples SB1 and SB2 were from Schrader Bluff formation, with temperature range 24 – 27°C and depths from 1,200 to 1,400 m. Samples K2 and K3 were from the Kuparuk formation, with temperature range between 47 and 70°C and depths between 1,785 and 2,150 m. And finally, samples I1 and I2 were collected from the Ivishak Formation, the hottest and deepest reservoir of the study, with temperature range between 80 and 83°C and depths of 2,750 to 3,100 m. In the study of Wang et al. (2019), the community structure and metabolic potential of oil-water mixture samples collected from the water-flooded Shen84 oil reservoir (Liaohe oil field, in China) was investigated. Wells 6111_O and 6111_W showed medium permeability, medium oil saturation and active edge water, well AJ-5 showed high permeability, high soil saturation and intensive anthropogenic perturbation, and well 71551 showed extremely low permeability and low oil saturation. Song and colleagues (2019) studied a water-flooded petroleum reservoir at the Shengli oil field, located in China. They sampled produced fluids from the well head of the production well to assess the taxonomic and functional information of the subsurface microbial community. In the study performed by Nie et al. (2016), crude oil metagenomes from two oil reservoirs in Daqing and Qinghai oil fields in China were analyzed. The oil samples were collected from the well head of the high temperature Qinghai (QH) oil field and from the well head of the mesophilic Daqing oil field (DQ), in order to compare the microbiome taxonomic and functional profiles. Finally, Sierra-Garcia and collaborators (2020) sequenced the metagenome from a crude oil sample collected from a non-water flooded and high-temperature Brazilian oil reservoir located at Miranga oil field in Brazil (BA-1) in order to assess the functional genetic potential of the microbial community. Due to the differences on the stage of recovery management and temperatures among the wells, five groups of samples were defined: i) without water injection

and high temperature; ii) with water injection and high temperature; iii) sea water injection and high temperature; iv) water injection and mesophilic temperatures; and v) recycled water injection.

Figure 2. Computational pipeline for obtaining MAGs.

2.2. Computational analysis

2.2.1. Quality control and filtering

In this study, we used metagenomic sequencing datasets previously obtained from oil reservoirs worldwide to recover metagenome-assembled genomes (MAGs). Sequence datasets were downloaded from databases and processed according to Figure 2. Briefly, quality control of sequencing reads was performed using FASTQ v.0.11.5 [14]. Adapter presence was detected using BBMap v38.90 [15]. Trimmomatic v. 0.36 [16] was used to remove the adapters and to trim reads with quality lower than 30 (Phred score) and length smaller than 100 bp. For some metagenomes with low quality and reads lengths of 100 bp, Phred score 20 and 75 bp minimum length were used as the thresholds.

2.2.2. Metagenome Assembly and binning

Mash software was used to calculate the pairwise distances between the datasets by employing the MinHash technique [17] (Figure 2), in order to know which samples could be co-assembled. Metagenome datasets with distance values below 0.1 were co-assembled. Both the trimmed read pairs and the reads left unpaired after quality filtering were used for metagenomic assembly. Assembly and co-assembly modes were performed using Megahit v.1.1.2 [18] with the following parameters: `--k-min 27 –k-max 147 –k-step 10 –min-contig-len 500`. Quality of the generated assemblies was assessed using MetaQUAST v5.0.2 [19]. The coverage profile of the contigs were obtained by using Bowtie2 v2.3.5.1 [20] read alignment program to map the reads of each sample to each assembly. SAM files generated were translated to BAM files, then sorted and indexed using SAMtools v1.9 [21].

Oil reservoir-associated MAGs were generated per assembly using five different bin-ning tools (Figure 2) with default options: MetaBAT2 v2.15 [22], MaxBin v2.2.7 [23], CON-COCT v1.1.0 [24], BinSanity v.0.5.3 [25] and Vamb v.3.0.2 [26], which use a combination of sequence composition and coverage information from the coverage profile. DAS Tool v1.1.2 [27] with option `--score_threshold 0.0` was used to combine MAGs produced by the five tools from each assembly to generate a non-redundant MAGs set. Resulting bins were analyzed for quality, contamination and completeness using CheckM v1.1.3 [28], which also uses HMMER v.3.2.1 [29] as a third-party software to compare the obtained bin against the single-copy gene profile of the genome of the affiliated taxon and to quantify how many genes are in the bin (completeness) and how many are in single-copy (contam-ination). Magpurify2 (https://github.com/apcamargo/MAGpurify) was used to the refine-ment of the obtained MAGs in order to remove putative contaminant contigs within each MAG (Figure 2).

MAGs quality was defined according to the MIMAG standard [Completeness – (5*Contamination)] [30]. MAGs with quality score > 50 were used in downstream anal-yses. High-quality MAGs are defined as completeness >90% and contamination <5% and medium-quality MAGs are defined as completeness >50% and contamination <10. The mapping rate was estimated using Bowtie2 v1.3.0 [20] to map the reads of each meta-genome to the MAGs originated from it; the proportion of reads that mapped with at least 95% identity was obtained with CoverM v0.6.1 (https://github.com/wwood/CoverM) in genome mode.

2.3. Taxonomic assignment and phylogeny

Taxonomic affiliation of MAGs was carried out using Genome Taxonomy Database and the associated GTDB-Tk toolkit v1.1.0 [31-33] and the refseq Release 89. GTDB-tk, which also uses pplacer as a third-party software [34], uses average nucleotide identity (ANI) and genome topology to find the closest genomic relative in its database of > 140,000 public prokaryote genomes. A phylogenetic approach was used to place the obtained MAGs in the tree of Bacteria and Archaea. PhylophlAn v3.0 [35] was used to place the genomes in a high-resolution phylogeny, and iTOL [36] was used for the visualization using the newick file as input.

2.4. Functional Assignment

Genes encoded in MAGs were predicted using Prodigal v.2.6.3 [37] and annotated using sequence aligner Diamond v0.9.30 [38] based on KEGG database (Kyoto Encyclope-dia of Genes and Genomes [39]. After annotation, genes found were filtered based on a list of specific genes related with hydrocarbon metabolism (Supplementary Table S2). The completeness of each pathway was calculated based on the fraction of genes present in KEGG Modules.

2.5. Environment-specific orthogroups

Proteins annotated in the oil reservoir MAGs were compared to related organisms in order to assess functional novelty related to oil reservoir environments. First, MAGs with taxonomic affiliation at least to family level were identified and grouped. Then, protein sequences from species of the same families of the oil reservoir MAG set were recovered from GTDB (release 89). Only families with at least four genomes were used for this anal-ysis (including the MAGs generated in this study). All-versus-all pairwise comparisons between proteins in each family were performed using Orthofinder v2.5.2 [40], clustering the proteins into orthogroups (orthologous gene cluster), and generating count matrices with the number of genes within each orthogroup per genome. To identify orthogroups that are exclusive to the oil reservoir MAGs (environment-specific orthogroups), tspex

v0.6.2 [41] was used to calculate the specificity measure (SPM) of each orthogroup from the matrices. Only orthogroups with at least three genes were used. If any orthogroup was classified as exclusive to the oil reservoir MAGs, the percentage of enrichment in all orthogroups in each family was calculated comparing with the representative genome from the GTDB. The orthogroups with at least 60% of enrichment were selected and functionally annotated using Diamond v. [38] against KEGG [42] and EggNogg [43] databases. Orthogroups with ambiguous functions (different annotations within the orthogroup) were ignored.

2.6. Functional core analysis

Annotation of a functional core shared among all oil reservoir datasets analyzed was performed based on KEGG databases. Gene annotations from each oil field were pooled and compared in order to assess both exclusive and common KEGG categories among sites. The percentage of orthologies in each group was calculated. The relative abundance of the functional core among all sites and of the exclusive functions based on Level 2 categories from KEGG orthologies were assessed.

3. Results & Discussion

3.1. Co-assembling statistics

After filtering low-quality reads, metagenomic datasets were processed by grouping metagenomes by read-level similarity (Minhash distances), in order to determine which of the metagenome sets could be co-assembled. The resultant distance matrix was plot in a clustering heatmap (Supplementary Figure S1). Datasets with distances below 0.1 were co-assembled. Four co-assemblies and six individual assemblies were performed. In all cases the co-assemblies corresponded to samples of the same study. Supplementary Table S3 summarizes the statistics of the assemblies.

3.2. Binning

The individual assemblies and co-assemblies were binned to MAGs. Merging of bins obtained using the pipeline described herein (Figure 2) yielded 176 MAGs. Quality evaluation results indicated that 74 of MAGs had estimated completeness higher than 50% and contamination less than 10% and were classified as medium quality, while 74 MAGs had completeness higher than 90% and contamination less than 5% and were classified as high-quality MAG, according to the MiMAG standard [30] (Supplementary Figure S2). MAGs with low-quality were not used for the further analyses.

3.3. Taxonomic affiliation and phylogeny

For the taxonomic assignment of MAGs, a phylogenetic tree was reconstructed using the 148 MAGs obtained in this study (Figure 3). One hundred and ten MAGs belonged to the Bacteria domain while 38 belonged to Archaea. The MAGs were grouped in 25 bacterial phyla and four archaeal phyla. The tree showed three large clusters representing the phyla Firmicutes (21 MAGs), Proteobacteria (19 MAGs), and Halobacterota (24 MAGs) (Figure 3). Smaller clades represented the Campylobacterota, Deferribacterota, Nitrospira, Desulfobacterota, Thermotoga, Euryarchaeota, among others. The most representative family was Methanotrichaceae with nine MAGs, followed by Pseudomonadaceae with seven MAGs, Thermovirgaceae with six MAGs and Archaeoglobaceae with four MAGs. At genus level, MAGs were assigned to 94 different genera. The most represented ones were the methanogenic archaeum Methanothrix, with five MAGs, and the proteobacterium Pseudomonas, also with five MAGs, followed by Thermodesulfobacterium (4) and Thermodesulfovibrio (3). The taxonomic assignment of all MAGs is detailed in the Supplementary Table S4.
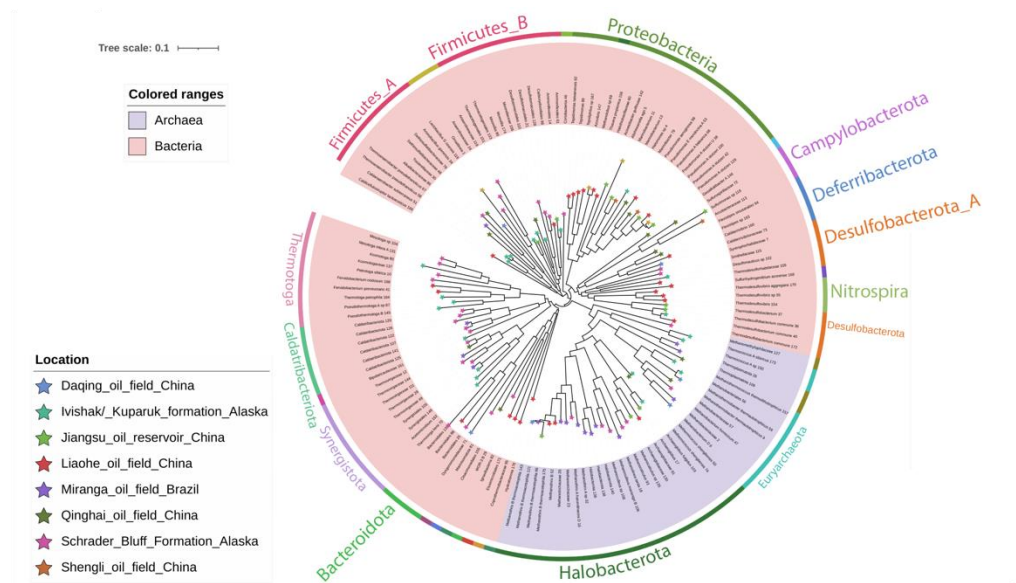
Figure 3. Phylogenetic tree of the 148 MAGs recovered from the oil reservoirs. The tree is decorated with colored backgrounds corresponding to domains, the most represented phyla are labelled in color in the outermost ring. The star shapes in the tips of the branches indicate the location of origin of the MAGs (see legend).

The first group of samples (without water injection plus high temperature) was composed only by BA-1 sample from Miranga Oil Field, the solely well that had not been water-flooded among the samples studied here. This well may reflect, to some degree, the indigenous thermophilic microbial communities in oil reservoirs before flooding operations. Sixteen MAGs were recovered from this sample, of which 11 MAGs were distributed in three archaeal phyla and five MAGs in four bacterial phyla (Figure 3). The Archaeal phylum Halobacterota was represented by 50% of the MAGs. The bacterial MAGs were assigned to the phyla Synergistota, Firmicutes, Patescibacteria and Caldatribacteriota. At the genus level, 13 out of 16 MAGs could not be assigned to any taxa, suggesting that the reservoir harbours several unknown taxa. Two MAGs, including one affiliated to the Microgenomatia class and other affiliated to the Tissierellaceae family, were found exclusively in this reservoir. The class Microgenomatia belongs to the Patescibacteria, recently proposed as phylum [33]. The majority of taxa from this phylum have been proposed through Metagenome-Assembled Genome approach of difficult access-environments, such as groundwater, deep sea sediments and deep subsurface [44-49]. In addition, three of the Microgenomatia genomes were found in activated sludge from a petrochemical plant [50], and one MAG belonging to Patescibacteria was found in crude oil metagenome from an oil field in Wietze, Germany [51]. Microbes from this phylum are characterized by small cell and genome sizes [52], containing only essential genes and lacking many metabolic functions like amino acid or nucleotide biosynthesis [44]. Due to these special characteristics, a parasitic lifestyle has been suggested [8,53]. Two MAGs from the Synergistales order were found in BA-1 sample from Miranga Oil Field. One of them was affiliated to the Thermovirgaceae family. Members of this family were found in a hydrocarbon reservoir that was submitted to CO2 enhanced oil recovery flood 40 years ago [54]. In terms of the archaeal community, this oil reservoir metagenome was characterized by the predominant presence of members of Halobacterota (8) phylum, followed by Thermoplasmota (2) and Euryarchaeota (1). MAGs from Thermoplasmota phylum could not be affiliated to higher taxonomic ranks. MAG2 was affiliated to the Methanobacteriaceae family. Members of this family are known to reduce H2 and CO2 to produce methane [55]. In a study across 22 oil reservoirs, Methanobacteria were found in all of them [56]. Zhao and collaborators analyzed eight oil reservoirs from northern China and found that the

core microbiota was composed by three bacterial and eight archaeal genera, including the genus Methanobacterium from the Methanobacteriaceae family [57]. In this study, one MAG from family Methanoregulaceae was recovered and affiliated to the species Methanolinea tarda. Members of Methanolinea are hydrogenotrophic archaea [58] and are also found as part of the core in some studies [56,57,59]. Four MAGs were affiliated to the Methanotrichaceae family, two of them belonging to Methanothrix genus. Finally, three MAGs were assigned to Halobacterota phylum, without any classification at higher ranks. However, they were placed into the Methanotrichales order clade (Figure. 3 and Supplementary table S4). The family Methanotrichaceae was proposed as a new family to replace Methanosaetaceae. Methanothrix is the only genus in this family, with three species, M. harundinacea, M. thermoacetophila and M. soehgenii [60]. This genus produces methane mainly from acetate, but it is also able to reduce CO_2 via direct interspecies electron transfer (DIET) with Geobacter spp. [61,62]. Members of Methanothrix are more frequently dominant in reservoirs from medium to high temperature reservoirs [56].

   In the second group of samples (water injection plus high temperature), only the dataset from Qinghai oil field (QH) was included. This oil field has been submitted to waterflooding for 15 years. Binning using this metagenome yielded four archaeal MAGs (Supplementary Table S4), which were assigned to Halobacterota (Methermicoccus shengliensis), Euryarchaeota (Methanothermococcus thermolithotrophicus and Methanobacteriaceae member) and to Nanoarchaeota (member of Woesearchaeales order). The last one is an ultra-small hyperthermophilic obligate ectosymbiont with a reduced genome that lives on the surface of several archaeal hosts of the Crenarchaeota phylum [63]. Members of Nanoarchaeota phylum have already been found in hot springs in Yellowstone National Park (United States) under high temperature (~90oC) [64]. They are also associated to volcanoes rich in hydrocarbons [64], deep-sea hydrothermal vents and deep lakes [63-73]. Methermicoccus shengliensis is a methylotrophic methanogen. It has been isolated from production water samples from a high temperature oil reservoir in Japan [74] and from Shengli Oil Field, where the in-situ temperatures range from 75 to 80oC [75]. Methanothermococcus genus is an hydrogenotrophic and methylotrophic methanogen [56]. Gao and colleagues investigated the spatial distribution of microbial communities and their drivers in 20 water-flooded oil reservoirs and two that had not been flooded, in China, and observed that Methanothermococcus was found more frequently in reservoirs with high salinity [56]. A study in the Xinjiang Luliang long-term water-flooded oil reservoir detected Methanothermococcus. thermolithotrophicus by mcrA sequencing [76]. With regards to the Bacteria domain, 12 MAGs were recovered from QH dataset, which belonged to eight phyla, as follows: Campylobacterota (1), Cloacimonadota (1), Deferribacterota (2), Desulfuromonadota (1), Firmicutes (3), Patescibacteria (1), Proteobacteria (2) and Synergistota (1). Two genomes of Flexistipes (Deferribacterota phylum) were found. There is evidence that species from this genus, as F. sinusarabici (MAG94), have genes that encode for electrically conductive pili (e-pili), similar to Geobacter species, that is responsible for the DIET process [77]. Geoalkalibacter (MAG146), belonging to Proteobacteria, is a strictly anaerobic bacterium, able to reduce Fe(III). A strain of this genus was isolated from a produced water in the Redwash oil field in Utah, USA, with 52°C of in situ temperature [78]. This genus was also found in a reservoir that had not been waterflooded [56]. Species of the genera Pseudomonas, Marinobacter, Sulfurospirillum, among others, are universally detected in reservoirs with a wide range of temperature, and most of them are hydrocarbon-degraders, surfactant producers, and nitrate or sulphate reducers [56]. Many of these microorganisms are aerobic. Due to the low redox potential in the oil reservoirs, anaerobic and facultative microorganisms are abundant. However, a lot of aerobic microbes are also detected in this environment, and water injection water may be closely related to that phenomenon [76].

   In the third group of samples (sea water injection plus high temperature) were clustered Kuparuk (K2 and K3) and Ivishak formations (I1 and I2). The binning process of these four metagenomes allowed the recovery of 24 MAGs. Some different physico-chemical characteristics were found among these samples; first the temperatures in the Ivishak

formation were between 80-83°C and in the Kuparuk were from 47 to 70°C, further the well I2 has the longest history of sea water injection among all samples, and that is reflected by the presence of the Desulfonauticus spp. (Desulfobacterota phylum), sulfate-reducing bacteria present in the deep-sea [79,80]. On the other hand, I1 well has only recently been injected with sea water. Only two MAGs were recovered from I1 sample. One of them was assigned to the thermophilic fermenting anaerobe Thermoanaerobacter pseudethanolicus (Firmicutes phylum), that has been isolated from oil reservoirs. This species is able to respire thiosulfate contributing to souring [81] and, depending on the ecological conditions, may act in syntrophy with acetoclastic methanogens [82]. The second MAG was assigned to Halomonas, Gammaproteobacteria class, a known hydrocarbon degrader [83]. MAGs found in the Kuparuk wells were different probably because the temperatures were lower, and the use of sea water injection is more recent than in the other site. These were classified as Methermicoccus shengliensis and the hyperthermophilic sulfide producing Archaeoglobus fulgidus [59,79], belonging to the Halobacterota phylum, and as the CO_2/H_2-reducing methanogen Methanothermobacter thermautotrophicus and the sulfide-producing Thermococcus sibiricus, members of the Euryarchaeota phylum. The last two are frequently found in oil thermophilic oil reservoirs [54,56,57,76,79]. Seven genomes belonging to Firmicutes phylum were recovered: Thermoanaerobacter pseudethanolicus (2), Caldanaerobacter subterraneus (1), Thermoacetogeniales (1), Ammonifexales (1) and Moorellia (2), which are known to be syntrophic and fermentative bacteria [75,84-86]. The family Thermotogaceae was represented with three genomes, belonging to two different thermophilic genera, Thermotoga and Pseudothermotoga. These microbes are indigenous to petroleum reservoirs and more commonly abundant in non-water flooded oil fields [87]. They are fermentative bacteria and have been detected in several high temperature oil reservoirs [75,81,88,89]. Lastly, one MAG affiliated to Thermodesulfobacterium commune, Desulfobacterota phylum, was found. This bacterium is able to incompletely oxidize fatty acids to produce acetate [90].

In the fourth group of samples (water injection plus mesophilic temperatures) were clustered Liaohe, Jiangsu, Shengli and Daqing Oil Field datasets. In this cluster, 52 MAGs were recovered, of which only 10 belonged to Archaea domain. As expected, this group of wells was characterized by the presence of many Proteobacteria members, especially the well-known hydrocarbon-degraders and surfactant producers like the facultative anaerobe organotrophs Pseudomonas (P. stutzeri, P. aureginosa, P. balearica), Acinetobacter, Marinobacterium and Tepidimonas, and the fermenters and nitrate reducers Azovibrio and Thauera. In many investigations, these bacteria have been detected in most of the oil reservoirs studied [56] and some authors have proposed these genera as part of the core microbiota of oil reservoirs [57,91,92]. However, they can be more abundant in water-flooded reservoirs [86]. Two MAGs belonging to Firmicutes phylum were recovered, Lactobacillus ozensis and Anoxybacillus gonensis. Species of the fermentative genus Lactobacillus have already been detected in oil samples, although its role in oil reservoirs still remains unknown [93]. A member of the Anoxybacillus genus was isolated from production water of a high temperature oil reservoir and showed ability to reduce nitrate and to control sulfate reducers [94]. One MAG assigned to the thermophilic sulfate reducing genus Thermodesulfovibrio (Nitrospira phylum) was reconstructed. Members of this genus were previously detected in oil field environments [95]. Archaeal MAGs were mainly assigned to Methanotrichaceae and Archaeogloboceae families and one MAG was assigned to the genus Methanolobus, a methylotrophic methanogen belonging to the Methanosarcinaceae family [96,97]. This methanogen was predominant in low and medium temperature oil reservoirs [56].

Finally, in the fifth group of samples (recycled water injection), SB1 and SB2 from the Schrader formation were included. Assembly enabled the recovery of 40 MAGs, of which seven were assigned to the Archaea domain, represented by 5 families, Methanobacteriaceae, Methanocorpusculaceae, Methanocullaceae, Methanomethylophilaceae and Methanotrichaceae. The hydrogenotrophic and mesopholic Methanocalculus archaeal genus was found in reservoirs without anthropogenic perturbation and was dominant in

medium to high temperature reservoirs [56]. Methanoculleus is a methanogen able to re-duce CO2 in syntrophy with bacteria such as Syntrophus and Marinobacter [79]. This ge-nus was proposed as part of the core microbiota across several oil reservoirs in China [57] and was found as dominant in low to medium temperature reservoirs [56]. A study per-formed by Liang and collaborators proposed a cooperation between Methanoculleus and Anaerolineaceae members for the n-alkanes degradation [98]. Three MAGs affiliated to Anaerolinaceae family were found in these samples, and other three MAGs belonging to the Kosmotogaceae family and one to the Petrotogaceae family, both from the Thermoto-gota phylum, were recovered. Members from these two last families have been found in oil field environments [98,99]. Lastly, three MAGs classified as sulphate-reducing bacteria from the Desulfotomaculales order were found. Desulfotomaculum has already been pro-posed as hydrocarbon degrading bacteria because of high similarity with gene sequences coding for benzyl succinate synthase (bssA), responsible for the activation of alkyl-substi-tuted aromatic hydrocarbons [59].

3.4. Metabolic potential

Similarly to the taxonomic analysis, annotation of the potential metabolism of MAGs revealed the presence of a microbial functional core across all oil fields and the occurrence of exclusive (or site-specific) functions, depending on the recovery management and in situ temperatures (Figure 4, Supplementary Table S2). Without any anthropogenic inter-vention, the indigenous microbial communities in oil reservoirs are initially dominated by slow growing anaerobes such as methanogens and some Clostridia members that are adapted to living on high concentration of hydrocarbons and high temperatures, among other extreme conditions [13]. This was observed in the first group, represented by the sample from Miranga oil field, that was mainly governed by methanogenic archaea, and as consequence, the potential metabolism prevailing was mainly methanogenesis-related (Figure 4, taxa names in purple / modules 12 - 16). In the second group, originated from high-temperature water flooded oil reservoir (Figure 4, taxa names in dark green), it was possible to observe diversification in both taxonomic and metabolic levels. In addition to the methanogenesis metabolism (modules 12 - 16), alkane-degradation (module 4) and dissimilatory nitrate reduction (module 17) were also present. Depending on the water source, the injection can deliver oxidants and different electron acceptors, thus promoting fast growing microbes as members from Deferribacterota, Proteobacteria and Bacteroi-dota phyla, with thermodynamically more favorable metabolism as nitrate or sulfate re-duction [13]. In the third cluster, comprising Kuparuk and Ivishak oil fields, sea water was mainly used for the secondary recovery. The use of sea water can introduce sulfate to the reservoirs, promoting sulfate reduction in the community and the consequent production of H2S [100]. Desulfonauticus sp. was probably introduced in the oil reservoir by sea wa-ter, and the functional analysis of this MAG showed the capability to reduce sulfate with the respective module almost complete (Figure 4, taxa names green / module 18). In this comparative study, the fourth group can be considered as the "greatest level of interven-tion" amongst all. These were mesophilic water-flooded oil fields. Here, a great shift in the functional profile of the microbial community was observed (Figure 4, taxa names red, light green and brown). Modules of hydrocarbon aerobic degradation (Figure 4, 3 - 8) are found in several MAGs. This was expected since samples were dominated by the fast-growing facultative anaerobes and opportunistic members of Marinobacter, Marinobac-terium, and other well-known hydrocarbon degraders as Pseudomonas and Tepidimonas [3,13,56]. Specially in the Liaohe and Jiangsu oil fields, many genomes with metabolic po-tential to reduce nitrate and sulfate were recovered. This is in accordance with the fact that recycled treated water was used for secondary recovery in the Schrader oil field. From all MAGs obtained in this study, only a few from this group showed metabolic potential to degrade aromatic hydrocarbons anaerobically. These MAGs were associated to Desul-fotomaculales and Syntrophorhabdaceae. As it was discussed above, members of Desul-fotomaculum had already been associated with fumarate activation mechanism in alkyl-substituted aromatic hydrocarbons as toluene and ethylbenzene [59]. The family

Syntrophorhabdaceae embraces a single mesophilic genus Syntrophorhabdus which is capable of oxidizing phenol p-cresol, 4-hydroxybenzoate, isophthalate and benzoate in association with an H(2)-scavenging partner (e.g., Methanospirillum hungatei or Desulfovibrio sp.) [101,102]. Here, MAGs belonging to Syntrophorhabdaceae and Desulfotomaculales showed several genes from the module of anaerobic Benzoyl-CoA degradation (module 11), a universal intermediate formed during degradation of aromatic compounds [103]

The metabolic analysis carried out in this study also provided evidence that anaerobic hydrocarbon degradation via fumarate addition (module 10) is a rare metabolism detected in oil reservoirs. The latter results are expected once mechanisms alternative to the archetypical fumarate addition reactions (e.g., anaerobic hydrocarbon carboxylation or hydroxylation) are present in oil reservoirs [3,104], which are not described in any KEEG category or module yet and therefore could not be detected in this work. However, genes for further activated hydrocarbon transformations were frequently detected, such as module 9, phenol anaerobic degradation, which was found in 61 MAGs out of 148, suggesting that the organisms represented by these MAGs mediate the downstream degradation of aromatic compounds via anaerobic benzoate degradation. On the other hand, the potential for aerobic hydrocarbon degradation was showed to occur in oil wells submitted to water flooded and in oil reservoirs with mesophilic temperatures. Last but not least, catechol meta-cleavage and all methanogenesis modules were found across all oil fields, suggesting that these metabolisms can be ubiquitous in oil reservoirs.
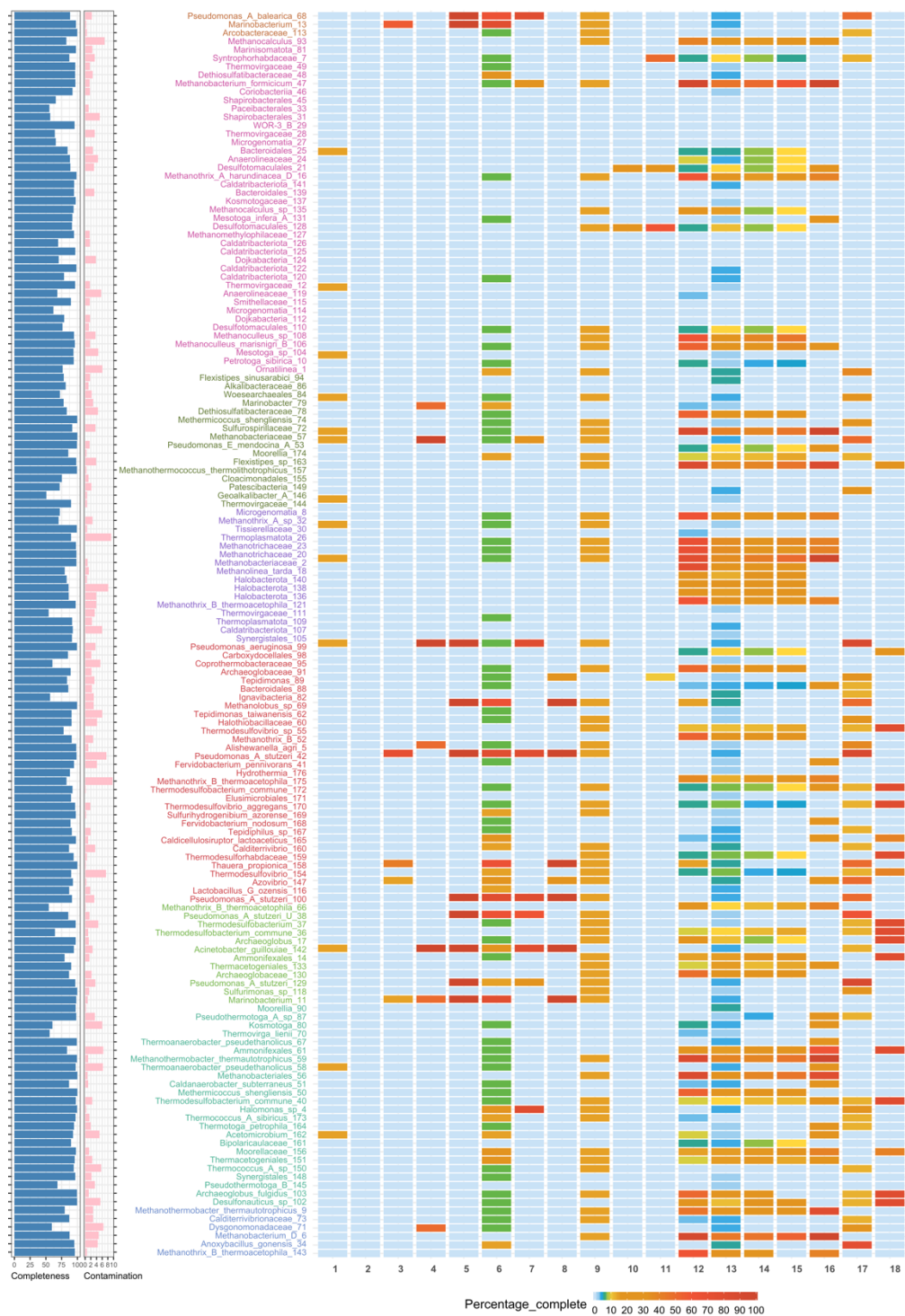
Figure 4. Heat map showing KEGG Module completeness in each MAG. Numbers below the bars: 1 - Ring Cleavage via Baeyer-Villiger oxidation; 2 - Naphthalene degradation; 3 - Biphenyl degradation; 4 - Alkane-degradation; 5 - Benzoate degradation; 6 - Catechol meta-cleavage; 7 - Catechol ortho-cleavage; 8 - Benzene degradation, aerobic; 9 - phenol degradation, anaerobic; 10 - Toluene degradation, anaerobic; 11 - Benzoyl-CoA degradation, anaerobic; 12 - Methanogenesis, CO_2; 13 - Methanogenesis, acetate; 14 - Methanogenesis, methylamines; 15 - Methanogenesis, methanol; 16 - Coenzyme M biosynthesis; 17 - Dissimilatory nitrate reduction; 18 - Dissimilatory sulfate reduction. Left bar plots represent the completeness (blue bars) and contamination (pink bars). The highest taxonomic affiliation was used. Color taxa names represent the oil field (see key color figure 3).

The module of catechol meta-cleavage degradation was found in 52% of the MAGs (78 MAGs). Even with a module completeness around 10%, it was a relevant result as this metabolic pathway is aerobic and was found in all MAGs belonging to Archaea. A bar plot with the frequency of each gene from catechol meta-cleavage and its phylum affiliation was constructed (Figure 5). The gene praC, that encodes the 4-oxalocrotonate tautomerase (4-OT), showed a higher frequency and was present in MAGs affiliated to 14 phyla, mainly in Proteobacteria, Halobacterota and Euryarchaeota (Figure 5). This gene was found in 15 archaeal MAGs, which were submitted to PATRIC platform [105] in order to search for evidence if it is part of an operon. As it can observed in the MAG comparison, this gene was not found in an operon region (Supplementary Figure S3). In three MAGs, this gene was annotated as 4-oxalocrotonate tautomerase-like (4-OT) enzyme (MAG17, MAG103 and MAG106). In MAG103 (Figure S3 a) and MAG17 (Figure S3 b), assigned to Archaeoglobus genus, the upstream genes found encode unknown proteins, while the downstream genes encode chorismate synthase and Acyl-CoA dehydrogenase. In MAG106 (Figure S3 c), assigned to Methanoculleus marisnigri, a gene that encodes an haloacid dehalogenase-like hydrolase was found upstream of praC gene. 4-OT enzyme is part of a group known as promiscuous enzymes, since they have multiple low-level catalytic activities in addition to their primary physiological function [106,107]. It is already known that many organisms through Bacteria and Archaea domains harbor 4-OT enzyme homologues which have unidentified physiological functions. Many of these organisms are not aromatic hydrocarbon degraders, suggesting that these homologues might have other functions [106].
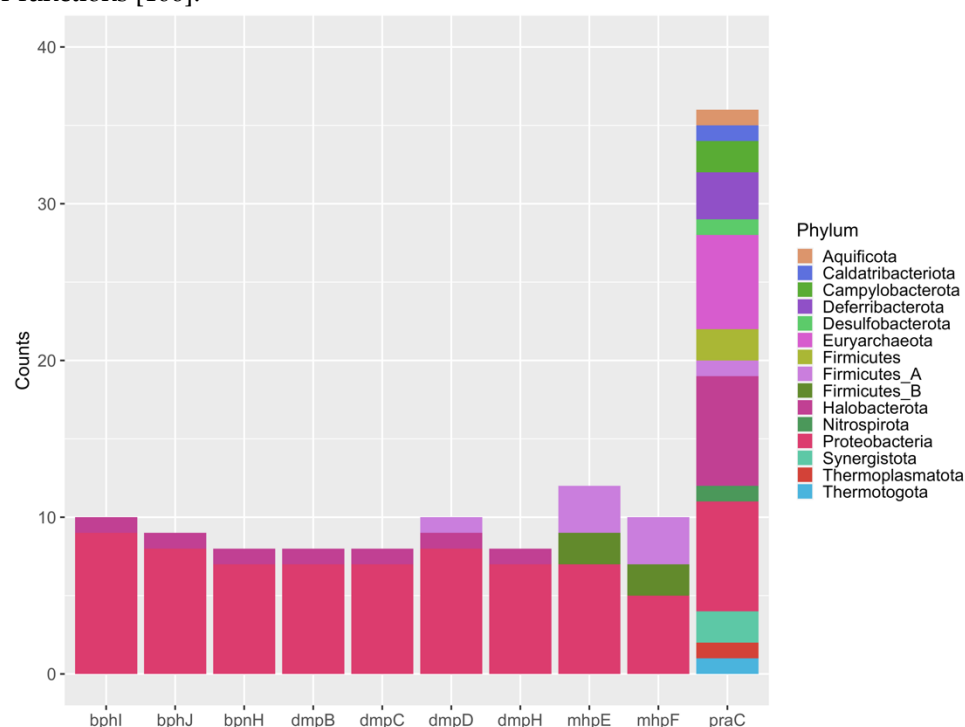

Figure 5. Frequencies of the genes from the Catechol meta-cleavage module across the phyla detected in oils reservoirs under study. Colors represent the phylum annotation.

Several recent studies of the methyl-coenzyme M reductase complex (Mcr) in MAGs have shown that divergent *mcr*-like genes are involved in methane/alkane metabolism. Here, we recovered the *mcr*A genes from MAGs assigned to archaea for further phylogenetic analysis. A phylogenetic tree was constructed using *mcr*A sequences recovered from 11 MAGs and approximately 280 curated *mcr*A sequences recovered from PhyMet² [108]. Interestingly, two MAGs (MAG 47 and MAG 59) each contained two different *mcr*A sequences, and two archaeoglobaceae-assigned MAGs contained *mcr*A sequences. In addition to the mcrA sequences found, four *mcr*-like sequences belonging to NM1 (*Candidatus*

Methanoliparia) and NM3 new lineages were included [109]. According to the phylogenetic analysis (Figure 6), MAG 16 assigned to *Methanothrix* clustered with NM3, Methanopyrales and Methanomassiliicoccales. NM3 is suggested to perform methyl-dependent hydrogenotrophic methanogenesis with the potential of using methanol and methanethiol [109]. The two MAGs assigned to Archaeoglobacea were related with NM1 lineage. NM1 belongs to Ca. *Methanoliparia* which has been reported to be capable of both methane and short chain alkane metabolisms [109,110]. These results indicate that archaeal MAGs identified in this work could also be involved with hydrocarbon degradation.
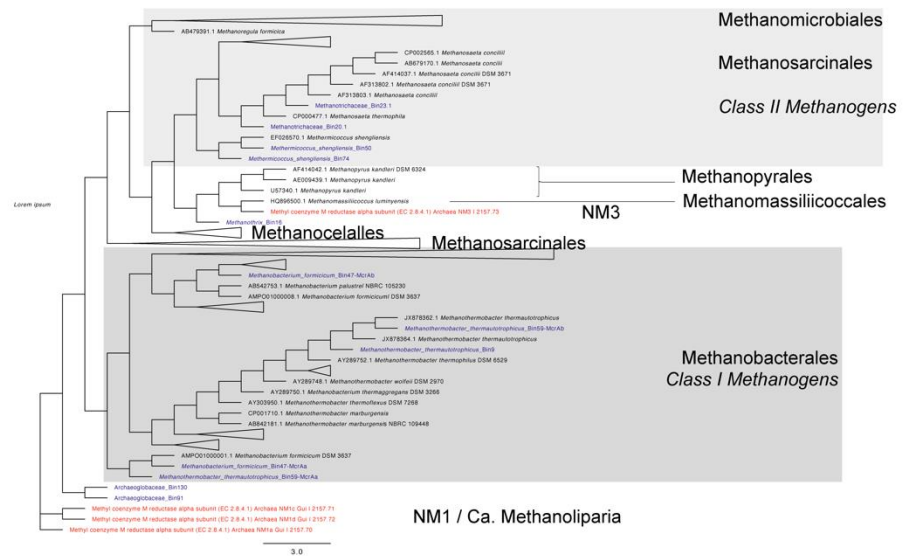
Figure 6. Phylogenetic tree of mcrA gene sequences recovered from the archaeal MAGs. Sequences of mcrA extracted from MAGs are shown in blue and Ca. Methanoliparia (NM1) and NM3 lineages are shown in red. The initial tags for mcrA sequences refer to accession numbers.

### 3.4. Functional core analysis

In order to investigate if there is a potential functional core shared by all oil reservoirs under analysis, KEGG annotations of the MAGs from each oil field were pooled and the KEGG orthologies common to all were assessed. At the same time the exclusive functions of each oil field were also evaluated. A total of 25,563 ORFs were annotated, distributed in 5,332 different KEGG orthologies. Functions that were present in at least 8 out of 9 oil fields were called as potential functional core (1,690 genes), which represented 31.7% of total functional annotation (Figure 7A). Site-specific genes (orthologies that were found in one oil field) accounted for 18.8% (1,007 KEGG functions) of total annotated KEGG orthologies (Figure 7A). The percentage of specific functions of each site was calculated (Figure 7A). Liaohe oil field showed the most distinct metabolic potential compared to the "functional core", comprising 41.9% of exclusive functions, followed by Daqing oil field with 17.37%, and Kuparuk and Jiangsu with 10.62% and 8.64% of exclusive genes, respectively. Shengli and Miranga oil fields had the lowest percentage of exclusive functional potential, and thus showed the most similar potential metabolism compared to the "core", with 1.39%% and 1.09% of the exclusive KEGG orthologies, respectively. These results suggest a possible correlation between the anthropogenic intervention level and the percentage of exclusive functions. For example, Miranga was the oil field without any intervention, and it had the lower percentage of exclusive functions, while Lioahe with the highest level of perturbations (water injection and mesophilic temperatures) had the highest percentage of exclusive genes (more different from the core).

The main metabolisms encompassed by the potential functional core were assessed (Figure 7B). The functional core comprised almost all metabolic categories. The most abundant core metabolisms were carbohydrate and energy metabolisms, the latter

including methane, nitrogen and sulfur metabolisms. All methanogenesis pathways were found in the functional core, with hydrogenotrophic methanogenesis as the most represented one. In addition, assimilatory sulfate reduction and nitrogen fixation were also part of the functional core. Xenobiotics biodegradation and metabolism was less represented in the functional core, with five different functions, four from benzoate degradation (genes praC, pcaC, mhpE and bsdB) and one from nitrotoluene degradation (gene hyaBC). Genes mhpE and praC belong to catechol meta-cleavage pathway, and as it was discussed above, gene praC encodes to a 4-OT, a promiscuous enzyme. Genes pcaC and bsdB are also part of an aerobic metabolism. These results suggest that the aerobic hydrocarbon degradation metabolism is likely an important potential energy strategy for the microbial community in oil fields. However, methanogenesis seems to be the most relevant metabolism in this extreme environment, being omnipresent independently of the well management practice employed.

Within the exclusive orthologies of each oil field, Ivishak, Schrader and Shengli oil fields showed functional specialization in xenobiotics biodegradation (Figure 7C). For example, in Shengli oil field more than 50% of specific genes were from xenobiotics degradation category, and these genes were different from the ones occurring in the other oil fields. These results suggest that in each site microorganisms are specialized in different biodegradation pathways.
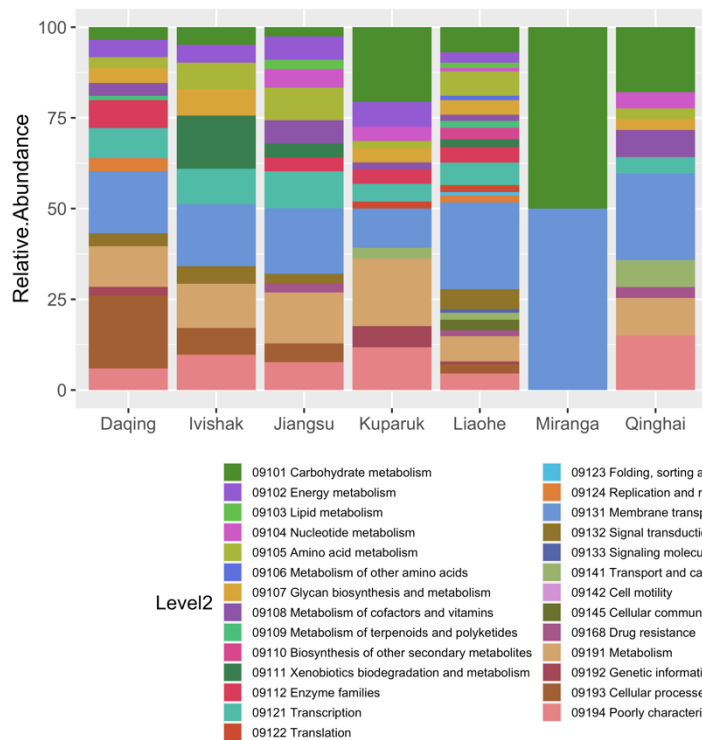
On the other hand, genes for sulfur oxidizing (sox genes) were exclusive in Liaohe oil field. This is consistent with the use of nitrate injection in this reservoir, a promising oil souring control strategy that promotes sulfur oxidation [13,111]. Despite the beneficial role, intermediates from the sulfide oxidation are potentially corrosive [112-114].
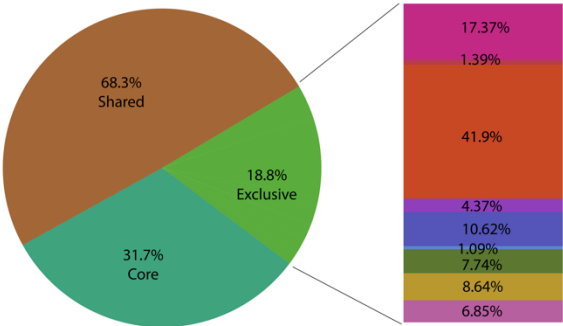
(a)
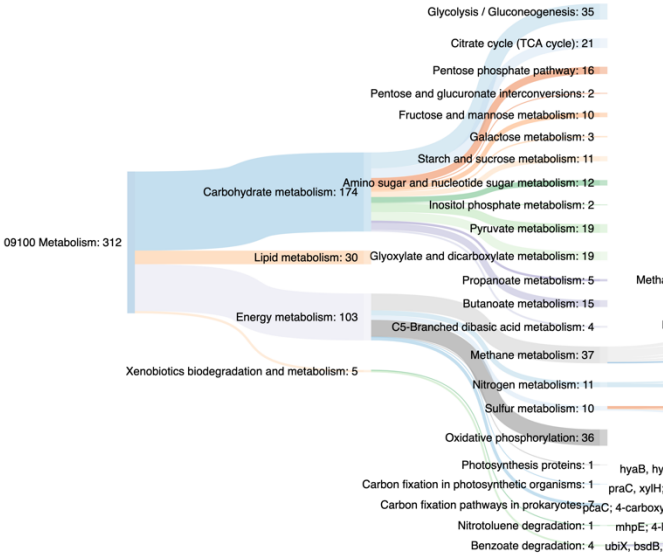
(b)

(c)

Figure 7. (a) Pie chart with percentages of specific KEGG orthologies from each oil field.(b) Main core metabolisms. (c) Exclusive orthologies at level 2 KEGG categories.

3.4.1. Environment-specific orthogroups

To explore the functional novelty in the microbiomes of oil reservoirs, a taxonomy informed approach was used by comparing 67 out of 148 MAGs obtained to related genomes at the family level and identifying orthogroups (cluster of orthologous genes) that are exclusive to the oil reservoir genomes at the family level. Among the 46 families that were analyzed, no unique orthogroups (out of 58,769) were found in the oil field datasets. Due to this result, tspex specificity measure (SPM) output was used to identify enriched orthogroups in the MAGs obtained in this study compared to the representative genomes from GTDB. No orthogroups with at least 60% of enrichment were found (Supplementary Table S5). With the thresholds and conditions of this analysis, no evidence of environment-specific or gene enrichment at family level was found amongst the microbiomes of oil fields under study, possibly due to the fact that the families of these environments are already highly adapted to these environments. Maybe at higher taxonomic levels it would be possible to find exclusive orthogroups related to oil fields.

4. Conclusions

Oil reservoirs are very complex niches that can be disturbed by intervention practices for oil recovery (water injection, use of biocides, etc). Understanding the indigenous and "disturbed" microbial community profiles and their functional potential is essential for evaluation of the efficiency and consequences of each treatment applied to the oil reservoir. Herein, information about the microorganisms and their functions in onshore oil reservoirs under different management practices and in situ temperatures across distinct geographic regions has been provided. The potential functional core shared by all oil reservoirs studied comprised mainly basic cellular functions, related to energy and replication. All methanogenesis pathway metabolisms and some genes related to aerobic

hydrocarbon degradation were part of the core. On the other hand, anaerobic degradation was not relevant, since it was not shared by microbial communities of all sites. In contrast, analysis of site-specific functions (orthologies that were found only in one oil field) revealed functional specialization in xenobiotics biodegradation in each oil field. Such metabolic specialization is likely driven by the selective pressure imposed by the distinct anthropogenic intervention practices and in situ temperatures. In order to improve oil recovery, water-flooding is effective, but in the long term it may have undesirable consequences as souring, corrosion and increased hydrocarbon degradation. Knowledge on the taxonomic and metabolic shifts in the oil field microbiome related to anthropogenic intervention is important to design future rationale and efficient oil recovery and toxic pollutant mitigation measures.

References

1. Wang, X.; Li, X.; Yu, L.; Huang, L.; Xiu, J.; Lin, W.; Zhang, Y. Characterizing the microbiome in petroleum reservoir flooded by different water sources. *Science of The Total Environment* **2019**, *653*, 872-885, doi:https://doi.org/10.1016/j.scitotenv.2018.10.410.

2. Nie, Y.; Zhao, J.-Y.; Tang, Y.-Q.; Guo, P.; Yang, Y.; Wu, X.-L.; Zhao, F. Species Divergence vs. Functional Convergence Characterizes Crude Oil Microbial Community Assembly. *Frontiers in Microbiology* **2016**, *7*, doi:10.3389/fmicb.2016.01254.

3. Sierra-Garcia, I.N.; Belgini, D.R.B.; Torres-Ballesteros, A.; Paez-Espino, D.; Capilla, R.; Santos Neto, E.V.; Gray, N.; de Oliveira, V.M. In depth metagenomic analysis in contrasting oil wells reveals syntrophic bacterial and archaeal associations for oil biodegradation in petroleum reservoirs. *Science of The Total Environment* **2020**, *715*, 136646, doi:https://doi.org/10.1016/j.scitotenv.2020.136646.

4.   Sierra-Garcia, I.N.; Dellagnezze, B.M.; Santos, V.P.; Chaves B, M.R.; Capilla, R.; Santos Neto, E.V.; Gray, N.; Oliveira, V.M. Microbial diversity in degraded and non-degraded petroleum samples and comparison across oil reservoirs at local and global scales. *Extremophiles* **2017**, *21*, 211-229, doi:10.1007/s00792-016-0897-8.

5.   An, D.; Caffrey, S.M.; Soh, J.; Agrawal, A.; Brown, D.; Budwill, K.; Dong, X.; Dunfield, P.F.; Foght, J.; Gieg, L.M.; et al. Metagenomics of Hydrocarbon Resource Environments Indicates Aerobic Taxa and Genes to be Unexpectedly Common. *Environmental Science & Technology* **2013**, *47*, 10708-10717, doi:10.1021/es4020184.

6.   Hu, P.; Tom, L.; Singh, A.; Thomas, B.C.; Baker, B.J.; Piceno, Y.M.; Andersen, G.L.; Banfield, J.F. Genome-Resolved Metagenomic Analysis Reveals Roles for Candidate Phyla and Other Microbial Community Members in Biogeochemical Transformations in Oil Reservoirs. *mBio* **2016**, *7*, e01669-01615, doi:10.1128/mBio.01669-15.

7.   Kotlar, H.K.; Lewin, A.; Johansen, J.; Throne-Holst, M.; Haverkamp, T.; Markussen, S.; Winnberg, A.; Ringrose, P.; Aakvik, T.; Ryeng, E. High coverage sequencing of DNA from microorganisms living in an oil reservoir 2.5 kilometres subsurface. *Environmental microbiology reports* **2011**, *3*, 674-681.

8.   Hug, L.A.; Baker, B.J.; Anantharaman, K.; Brown, C.T.; Probst, A.J.; Castelle, C.J.; Butterfield, C.N.; Hernsdorf, A.W.; Amano, Y.; Ise, K. A new view of the tree of life. *Nature microbiology* **2016**, *1*, 1-6.

9.   Parks, D.H.; Rinke, C.; Chuvochina, M.; Chaumeil, P.-A.; Woodcroft, B.J.; Evans, P.N.; Hugenholtz, P.; Tyson, G.W. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology* **2017**, *2*, 1533-1542, doi:10.1038/s41564-017-0012-7.

10.   Tyson, G.W.; Chapman, J.; Hugenholtz, P.; Allen, E.E.; Ram, R.J.; Richardson, P.M.; Solovyev, V.V.; Rubin, E.M.; Rokhsar, D.S.; Banfield, J.F. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **2004**, *428*, 37-43.

11.   Liu, Y.-F.; Galzerani, D.D.; Mbadinga, S.M.; Zaramela, L.S.; Gu, J.-D.; Mu, B.-Z.; Zengler, K. Metabolic capability and in situ activity of microorganisms in an oil reservoir. *Microbiome* **2018**, *6*, 5, doi:10.1186/s40168-017-0392-1.

12.   Song, Z.; Chen, S.; Zhao, F.; Zhu, W. Whole metagenome of injected and produced fluids reveal the heterogenetic characteristics of the microbial community in a water-flooded oil reservoir. *Journal of Petroleum Science and Engineering* **2019**, *176*, 1198-1207, doi:https://doi.org/10.1016/j.petrol.2019.02.008.

13.   Vigneron, A.; Alsop, E.B.; Lomans, B.P.; Kyrpides, N.C.; Head, I.M.; Tsesmetzis, N. Succession in the petroleum reservoir microbiome through an oil field production lifecycle. *The ISME Journal* **2017**, *11*, 2141-2154, doi:10.1038/ismej.2017.78.

14.   Andrews, S. FastQC: a quality control tool for high throughput sequence data. **2010**, doi:https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 07 September 2020.

15.   Bushnell, B.; Rood, J.; Singer, E. BBMerge – Accurate paired shotgun read merging via overlap. *PLOS ONE* **2017**, *12*, e0185056, doi:10.1371/journal.pone.0185056.

16.   Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, doi:https://doi.org/10.1093/bioinformatics/btu170.

17.   Ondov, B.D.; Treangen, T.J.; Melsted, P.; Mallonee, A.B.; Bergman, N.H.; Koren, S.; Phillippy, A.M. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology* **2016**, *17*, 132, doi:10.1186/s13059-016-0997-x.

18.   Li, D.; Liu, C.-M.; Luo, R.; Sadakane, K.; Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **2015**, *31*, 1674-1676.

19.   Mikheenko, A.; Saveliev, V.; Gurevich, A. MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics* **2016**, *32*, doi:10.1093/bioinformatics/btv697.

20.   Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **2012**, *9*, 357.

21.   Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078-2079.

22.   Kang, D.D.; Froula, J.; Egan, R.; Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **2015**, *3*, e1165.

23.   Wu, Y.-W.; Simmons, B.A.; Singer, S.W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **2015**, *32*, 605-607.

24.   Alneberg, J.; Bjarnason, B.S.; de Bruijn, I.; Schirmer, M.; Quick, J.; Ijaz, U.Z.; Lahti, L.; Loman, N.J.; Andersson, A.F.; Quince, C. Binning metagenomic contigs by coverage and composition. *Nature Methods* **2014**, *11*, 1144-1146, doi:10.1038/nmeth.3103.

25.   Graham, E.D.; Heidelberg, J.F.; Tully, B.J. BinSanity: unsupervised clustering of environmental microbial assemblies using coverage and affinity propagation. *PeerJ* **2017**, *5*, e3035.

26.   Nissen, J.N.; Johansen, J.; Allesøe, R.L.; Sønderby, C.K.; Armenteros, J.J.A.; Grønbech, C.H.; Jensen, L.J.; Nielsen, H.B.; Petersen, T.N.; Winther, O.; et al. Improved metagenome binning and assembly using deep variational autoencoders. *Nature Biotechnology* **2021**, *39*, 555-560, doi:10.1038/s41587-020-00777-4.

27.   Sieber, C.M.K.; Probst, A.J.; Sharrar, A.; Thomas, B.C.; Hess, M.; Tringe, S.G.; Banfield, J.F. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology* **2018**, *3*, 836-843, doi:10.1038/s41564-018-0171-1.

28.   Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* **2015**, gr. 186072.186114.

29.   Wheeler, T.J.; Eddy, S.R. nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **2013**, *29*, 2487-2489.

30.   Bowers, R.M.; Kyrpides, N.C.; Stepanauskas, R.; Harmon-Smith, M.; Doud, D.; Reddy, T.B.K.; Schulz, F.; Jarett, J.; Rivers, A.R.; Eloe-Fadrosh, E.A.; et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology* **2017**, *35*, 725-731, doi:10.1038/nbt.3893.

31.   Chaumeil, P.-A.; Mussig, A.J.; Hugenholtz, P.; Parks, D.H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **2019**, *36*, 1925-1927, doi:10.1093/bioinformatics/btz848.

32.   Parks, D.H.; Chuvochina, M.; Chaumeil, P.-A.; Rinke, C.; Mussig, A.J.; Hugenholtz, P. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nature Biotechnology* **2020**, *38*, 1079-1086, doi:10.1038/s41587-020-0501-8.

33.   Parks, D.H.; Chuvochina, M.; Waite, D.W.; Rinke, C.; Skarshewski, A.; Chaumeil, P.-A.; Hugenholtz, P. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature biotechnology* **2018**, *36*, 996-1004.

34.   Matsen, F.A.; Kodner, R.B.; Armbrust, E.V. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC bioinformatics* **2010**, *11*, 1-16.

35.   Segata, N.; Börnigen, D.; Morgan, X.C.; Huttenhower, C. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature communications* **2013**, *4*, 1-11.

36.   Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic acids research* **2019**, *47*, W256-W259.

37.   Hyatt, D.; Chen, G.-L.; LoCascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **2010**, *11*, 119, doi:DOI: 10.1186/1471-2105-11-119.

38.   Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nature methods* **2014**, *12*, 59.

39.   Kanehisa, M.; Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **2000**, *28*, 27-30.

40.   Emms, D.M.; Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology* **2019**, *20*, 1-14.

41.   Camargo, A.P.; Vasconcelos, A.A.; Fiamenghi, M.B.; Pereira, G.A.; Carazzolle, M.F. tspex: a tissue-specificity calculator for gene expression data. **2020**.

42.   Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **2017**, *45*, D353-d361, doi:10.1093/nar/gkw1092.

43.   Jensen, L.J.; Julien, P.; Kuhn, M.; von Mering, C.; Muller, J.; Doerks, T.; Bork, P. eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic acids research* **2007**, *36*, D250-D254.

44.   Brown, C.T.; Hug, L.A.; Thomas, B.C.; Sharon, I.; Castelle, C.J.; Singh, A.; Wilkins, M.J.; Wrighton, K.C.; Williams, K.H.; Banfield, J.F. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **2015**, *523*, 208-211.

45.   Frey, B.; Rime, T.; Phillips, M.; Stierli, B.; Hajdas, I.; Widmer, F.; Hartmann, M. Microbial diversity in European alpine permafrost and active layers. *FEMS microbiology ecology* **2016**, *92*, fiw018.

46.   Herrmann, M.; Wegner, C.-E.; Taubert, M.; Geesink, P.; Lehmann, K.; Yan, L.; Lehmann, R.; Totsche, K.U.; Küsel, K. Predominance of Cand. Patescibacteria in Groundwater Is Caused by Their Preferential Mobilization From Soils and Flourishing Under Oligotrophic Conditions. *Frontiers in Microbiology* **2019**, *10*, doi:10.3389/fmicb.2019.01407.

47.   Hubalek, V.; Wu, X.; Eiler, A.; Buck, M.; Heim, C.; Dopson, M.; Bertilsson, S.; Ionescu, D. Connectivity to the surface determines diversity patterns in subsurface aquifers of the Fennoscandian shield. *The ISME journal* **2016**, *10*, 2447-2458.

48.   León-Zayas, R.; Peoples, L.; Biddle, J.F.; Podell, S.; Novotny, M.; Cameron, J.; Lasken, R.S.; Bartlett, D.H. The metabolic potential of the single cell genomes obtained from the Challenger Deep, Mariana Trench within the candidate superphylum Parcubacteria (OD1). *Environmental microbiology* **2017**, *19*, 2769-2784.

49.   Luef, B.; Frischkorn, K.R.; Wrighton, K.C.; Holman, H.-Y.N.; Birarda, G.; Thomas, B.C.; Singh, A.; Williams, K.H.; Siegerist, C.E.; Tringe, S.G. Diverse uncultivated ultra-small bacterial cells in groundwater. *Nature communications* **2015**, *6*, 1-8.

50.   Antunes, T.C.; Marconatto, L.; Borges, L.G.d.A.; Giongo, A.; Sand, S.T.V.D. Analysis of microbial community biodiversity in activated sludge from a petrochemical plant. *Revista Ambiente & Água* **2021**, *16*.

51.   Eze, M.O.; Lütgert, S.A.; Neubauer, H.; Balouri, A.; Kraft, A.A.; Sieven, A.; Daniel, R.; Wemheuer, B.; Putonti, C. Metagenome Assembly and Metagenome-Assembled Genome Sequences from a Historical Oil Field Located in Wietze, Germany. *Microbiology Resource Announcements* **2020**, *9*, e00333-00320, doi:doi:10.1128/MRA.00333-20.

52.   Tian, R.; Ning, D.; He, Z.; Zhang, P.; Spencer, S.J.; Gao, S.; Shi, W.; Wu, L.; Zhang, Y.; Yang, Y.; et al. Small and mighty: adaptation of superphylum Patescibacteria to groundwater environment drives their genome simplicity. *Microbiome* **2020**, *8*, 51, doi:10.1186/s40168-020-00825-w.

53.   Castelle, C.J.; Brown, C.T.; Anantharaman, K.; Probst, A.J.; Huang, R.H.; Banfield, J.F. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nature Reviews Microbiology* **2018**, *16*, 629-645.

54.   Shelton, J.L.; Andrews, R.S.; Akob, D.M.; DeVera, C.A.; Mumford, A.; McCray, J.E.; McIntosh, J.C. Microbial community composition of a hydrocarbon reservoir 40 years after a CO2 enhanced oil recovery flood. *FEMS Microbiology Ecology* **2018**, *94*, doi:10.1093/femsec/fiy153.

55.   Alsouleman, K.; Linke, B.; Klang, J.; Klocke, M.; Krakat, N.; Theuerl, S. Reorganisation of a mesophilic biogas microbiome as response to a stepwise increase of ammonium nitrogen induced by poultry manure supply. *Bioresource technology* **2016**, *208*, 200-204.

56.   Gao, P.; Tian, H.; Wang, Y.; Li, Y.; Li, Y.; Xie, J.; Zeng, B.; Zhou, J.; Li, G.; Ma, T. Spatial isolation and environmental factors drive distinct bacterial and archaeal communities in different types of petroleum reservoirs in China. *Scientific Reports* **2016**, *6*, 20174, doi:10.1038/srep20174.

57.   Zhao, J.-Y.; Hu, B.; Dolfing, J.; Li, Y.; Tang, Y.-Q.; Jiang, Y.; Chi, C.-Q.; Xing, J.; Nie, Y.; Wu, X.-L. Thermodynamically favorable reactions shape the archaeal community affecting bacterial community assembly in oil reservoirs. *Science of The Total Environment* **2021**, *781*, 146506, doi:https://doi.org/10.1016/j.scitotenv.2021.146506.

58.   Wang, H.-Z.; Gou, M.; Yi, Y.; Xia, Z.-Y.; Tang, Y.-Q. Identification of novel potential acetate-oxidizing bacteria in an acetate-fed methanogenic chemostat based on DNA stable isotope probing. *The Journal of general and applied microbiology* **2018**, *64*, 221-231.

59.   Sengupta, K.; Pal, S. A review on microbial diversity and genetic markers involved in methanogenic degradation of hydrocarbons: futuristic prospects of biofuel recovery from contaminated regions. *Environmental Science and Pollution Research* **2021**, doi:10.1007/s11356-021-13666-3.

60.   Oren, A. The family Methanotrichaceae. *The Prokaryotes* **2014**, 298-306.

61.   Akinyemi, T.S.; Shao, N.; Whitman, W.B. Methanothrix. In *Bergey's Manual of Systematics of Archaea and Bacteria*; pp. 1-12.

62.   Holmes, D.E.; Shrestha, P.M.; Walker, D.J.F.; Dang, Y.; Nevin, K.P.; Woodard, T.L.; Lovley, D.R.; Schloss, P.D. Metatranscriptomic Evidence for Direct Interspecies Electron Transfer between Geobacter and Methanothrix Species in Methanogenic Rice Paddy Soils. *Applied and Environmental Microbiology* **2017**, *83*, e00223-00217, doi:doi:10.1128/AEM.00223-17.

63.   St. John, E.; Flores, G.E.; Meneghin, J.; Reysenbach, A.-L. Deep-sea hydrothermal vent metagenome-assembled genomes provide insight into the phylum Nanoarchaeota. *Environmental Microbiology Reports* **2019**, *11*, 262-270, doi:https://doi.org/10.1111/1758-2229.12740.

64.   Munson-McGee, J.H.; Field, E.K.; Bateson, M.; Rooney, C.; Stepanauskas, R.; Young, M.J.; Wommack, K.E. Nanoarchaeota, Their Sulfolobales Host, and Nanoarchaeota Virus Distribution across Yellowstone National Park Hot Springs. *Applied and Environmental Microbiology* **2015**, *81*, 7860-7868, doi:doi:10.1128/AEM.01539-15.

65.   Carrier, V.; Svenning, M.M.; Gründger, F.; Niemann, H.; Dessandier, P.-A.; Panieri, G.; Kalenitchenko, D. The Impact of Methane on Microbial Communities at Marine Arctic Gas Hydrate Bearing Sediment. *Frontiers in Microbiology* **2020**, *11*, doi:10.3389/fmicb.2020.01932.

66.   Casanueva, A.; Galada, N.; Baker, G.C.; Grant, W.D.; Heaphy, S.; Jones, B.; Yanhe, M.; Ventosa, A.; Blamey, J.; Cowan, D.A. Nanoarchaeal 16S rRNA gene sequences are widely dispersed in hyperthermophilic and mesophilic halophilic environments. *Extremophiles* **2008**, *12*, 651-656.

67.   Chernyh, N.A.; Mardanov, A.V.; Gumerov, V.M.; Miroshnichenko, M.L.; Lebedinsky, A.V.; Merkel, A.Y.; Crowe, D.; Pimenov, N.V.; Rusanov, I.I.; Ravin, N.V. Microbial life in Bourlyashchy, the hottest thermal pool of Uzon Caldera, Kamchatka. *Extremophiles* **2015**, *19*, 1157-1171.

68.   Clingenpeel, S.; Kan, J.; Macur, R.; Woyke, T.; Lovalvo, D.; Varley, J.; Inskeep, W.P.; Nealson, K.; McDermott, T. Yellowstone lake nanoarchaeota. *Frontiers in microbiology* **2013**, *4*, 274.

69.   Flores, G.; Shakya, M.; Meneghin, J.; Yang, Z.; Seewald, J.; Geoff Wheat, C.; Podar, M.; Reysenbach, A.L. Inter-field variability in the microbial communities of hydrothermal vent deposits from a back-arc basin. *Geobiology* **2012**, *10*, 333-346.

70.   Flores, G.E.; Campbell, J.H.; Kirshtein, J.D.; Meneghin, J.; Podar, M.; Steinberg, J.I.; Seewald, J.S.; Tivey, M.K.; Voytek, M.A.; Yang, Z.K. Microbial community structure of hydrothermal deposits from geochemically different vent fields along the Mid-Atlantic Ridge. *Environmental microbiology* **2011**, *13*, 2158-2171.

71.   Hohn, M.J.; Hedlund, B.P.; Huber, H. Detection of 16S rDNA sequences representing the novel phylum "Nanoarchaeota": indication for a wide distribution in high temperature biotopes. *Systematic and applied microbiology* **2002**, *25*, 551-554.

72.   McCliment, E.A.; Voglesonger, K.M.; O'Day, P.A.; Dunn, E.E.; Holloway, J.R.; Cary, S.C. Colonization of nascent, deep-sea hydrothermal vents by a novel Archaeal and Nanoarchaeal assemblage. *Environmental Microbiology* **2006**, *8*, 114-125.

73.   Zemskaya, T.I.; Cabello-Yeves, P.J.; Pavlova, O.N.; Rodriguez-Valera, F. Microorganisms of Lake Baikal—the deepest and most ancient lake on Earth. *Appl Microbiol Biotechnol* **2020**, *104*, 6079-6090.

74.   Mayumi, D.; Mochimaru, H.; Tamaki, H.; Yamamoto, K.; Yoshioka, H.; Suzuki, Y.; Kamagata, Y.; Sakata, S. Methane production from coal by a single methanogen. *Science* **2016**, *354*, 222-225, doi:10.1126/science.aaf8821.

75.   Cheng, L.; Ding, C.; Li, Q.; He, Q.; Dai, L.-r.; Zhang, H. DNA-SIP Reveals That Syntrophaceae Play an Important Role in Methanogenic Hexadecane Degradation. *PLOS ONE* **2013**, *8*, e66784, doi:10.1371/journal.pone.0066784.

76.   Gao, P.; Tian, H.; Li, G.; Sun, H.; Ma, T. Microbial diversity and abundance in the Xinjiang Luliang long-term water-flooding petroleum reservoir. *MicrobiologyOpen* **2015**, *4*, 332-342, doi:https://doi.org/10.1002/mbo3.241.

77.   Walker, D.J.F.; Adhikari, R.Y.; Holmes, D.E.; Ward, J.E.; Woodard, T.L.; Nevin, K.P.; Lovley, D.R. Electrically conductive pili from pilin genes of phylogenetically diverse microorganisms. *The ISME Journal* **2018**, *12*, 48-58, doi:10.1038/ismej.2017.141.

78.   Greene, A.C.; Patel, B.K.C.; Yacob, S. Geoalkalibacter subterraneus sp. nov., an anaerobic Fe(III)- and Mn(IV)-reducing bacterium from a petroleum reservoir, and emended descriptions of the family Desulfuromonadaceae and the genus

Geoalkalibacter. *International Journal of Systematic and Evolutionary Microbiology* **2009**, *59*, 781-785, doi:https://doi.org/10.1099/ijs.0.001537-0.

79.   Piceno, Y.M.; Reid, F.C.; Tom, L.M.; Conrad, M.E.; Bill, M.; Hubbard, C.G.; Fouke, B.W.; Graff, C.J.; Han, J.; Stringfellow, W.T.; et al. Temperature and injection water source influence microbial community structure in four Alaskan North Slope hydrocarbon reservoirs. *Frontiers in Microbiology* **2014**, *5*, doi:10.3389/fmicb.2014.00409.

80.   Mayilraj, S.; Kaksonen, A.H.; Cord-Ruwisch, R.; Schumann, P.; Spröer, C.; Tindall, B.J.; Spring, S. Desulfonauticus autotrophicus sp. nov., a novel thermophilic sulfate-reducing bacterium isolated from oil-production water and emended description of the genus Desulfonauticus. *Extremophiles* **2009**, *13*, 247-255, doi:10.1007/s00792-008-0212-4.

81.   Orphan, V.; Taylor, L.; Hafenbradl, D.; Delong, E. Culture-dependent and culture-independent characterization of microbial assemblages associated with high-temperature petroleum reservoirs. *Applied and environmental microbiology* **2000**, *66*, 700-711.

82.   Shestakova, N.; Ivoilov, V.; Tourova, T.; Belyaev, S.; Poltaraus, A.; Nazina, T. Application of clone libraries: syntrophic acetate degradation to methane in a high-temperature petroleum reservoir: culture-based and 16S rRNA genes characterization. *Applied Microbiology and Molecular Biology in Oil Field Systems* **2011**, 45-53.

83.   Mnif, S.; Chamkha, M.; Sayadi, S. Isolation and characterization of Halomonas sp. strain C2SS100, a hydrocarbon-degrading bacterium under hypersaline conditions. *Journal of applied microbiology* **2009**, *107*, 785-794.

84.   Cheng, L.; He, Q.; Ding, C.; Dai, L.-r.; Li, Q.; Zhang, H. Novel bacterial groups dominate in a thermophilic methanogenic hexadecane-degrading consortium. *FEMS Microbiology Ecology* **2013**, *85*, 568-577, doi:10.1111/1574-6941.12141.

85.   Mbadinga, S.M.; Li, K.-P.; Zhou, L.; Wang, L.-Y.; Yang, S.-Z.; Liu, J.-F.; Gu, J.-D.; Mu, B.-Z. Analysis of alkane-dependent methanogenic community derived from production water of a high-temperature petroleum reservoir. *Appl Microbiol Biotechnol* **2012**, *96*, 531-542.

86.   Zhou, L.; Li, K.-P.; Mbadinga, S.M.; Yang, S.-Z.; Gu, J.-D.; Mu, B.-Z. Analyses of n-alkanes degrading community dynamics of a high-temperature methanogenic consortium enriched from production water of a petroleum reservoir by a combination of molecular techniques. *Ecotoxicology* **2012**, *21*, 1680-1691.

87.   Pournia, M.; Bahador, N.; Tabatabaei, M.; Azarbayjani, R.; Salekdeh, G.H. Microbial Diversity of Non-flooded High Temperature Petroleum Reservoir in South of Iran. *Biological Journal of Microorganism* **2020**, *8*.

88.   Dahle, H.; Garshol, F.; Madsen, M.; Birkeland, N.-K. Microbial community structure analysis of produced water from a high-temperature North Sea oil-field. *Antonie van Leeuwenhoek* **2008**, *93*, 37-49, doi:10.1007/s10482-007-9177-z.

89.   Magot, M.; Ollivier, B.; Patel, B.K.C. Microbiology of petroleum reservoirs. *Antonie van Leeuwenhoek* **2000**, *77*, 103-116, doi:10.1023/A:1002434330514.

90.   Henry, E.A.; Devereux, R.; Maki, J.S.; Gilmour, C.C.; Woese, C.R.; Mandelco, L.; Schauder, R.; Remsen, C.C.; Mitchell, R. Characterization of a new thermophilic sulfate-reducing bacterium. *Archives of Microbiology* **1994**, *161*, 62-69, doi:10.1007/BF00248894.

91.   Meslé, M.; Dromart, G.; Oger, P. Microbial methanogenesis in subsurface oil and coal. *Research in microbiology* **2013**, *164*, 959-972.

92.   Nazina, T.N.; Shestakova, N.M.; Semenova, E.M.; Korshunova, A.V.; Kostrukova, N.K.; Tourova, T.P.; Min, L.; Feng, Q.; Poltaraus, A.B. Diversity of Metabolically Active Bacteria in Water-Flooded High-Temperature Heavy Oil Reservoir. *Frontiers in Microbiology* **2017**, *8*, doi:10.3389/fmicb.2017.00707.

93.   Silva, T.R.; Verde, L.C.L.; Santos Neto, E.V.; Oliveira, V.M. Diversity analyses of microbial communities in petroleum samples from Brazilian oil fields. *International Biodeterioration & Biodegradation* **2013**, *81*, 14, doi:doi:10.1016/j.ibiod.2012.05.005.

94.   Liu, J.-F.; Wu, W.-L.; Yao, F.; Wang, B.; Zhang, B.-L.; Mbadinga, S.M.; Gu, J.-D.; Mu, B.-Z. A thermophilic nitrate-reducing bacterium isolated from production water of a high temperature oil reservoir and its inhibition on sulfate-reducing bacteria. *Applied Environmental Biotechnology* **2016**, *2*, 35-42.

95.   Tüccar, T.; Ilhan-Sungur, E.; Abbas, B.; Muyzer, G. Coexistence of sulfate reducers with the other oil bacterial groups in Diyarbakır oil fields. *Anaerobe* **2019**, *59*, 19-31, doi:https://doi.org/10.1016/j.anaerobe.2019.04.008.

96.   Nazaries, L.; Murrell, J.C.; Millard, P.; Baggs, L.; Singh, B.K. Methane, microbes and models: fundamental understanding of the soil methane cycle for future predictions. *Environmental microbiology* **2013**, *15*, 2395-2417.

97.   Penger, J.; Conrad, R.; Blaser, M. Stable carbon isotope fractionation by methylotrophic methanogenic archaea. *Applied and environmental microbiology* **2012**, *78*, 7596-7602.

98.   Santos, J.C.d.; Lopes, D.R.G.; Da Silva, J.D.; De Oliveira, M.D.; Dias, R.S.; Lima, H.S.; De Sousa, M.P.; De Paula, S.O.; Silva, C.C.d. Diversity of sulfate-reducing prokaryotes in petroleum production water and oil samples. *International Biodeterioration & Biodegradation* **2020**, *151*, 104966, doi:https://doi.org/10.1016/j.ibiod.2020.104966.

99.   Summers, Z.M.; Belahbib, H.; Pradel, N.; Bartoli, M.; Mishra, P.; Tamburini, C.; Dolla, A.; Ollivier, B.; Armougom, F. A novel Thermotoga strain TFO isolated from a Californian petroleum reservoir phylogenetically related to Thermotoga petrophila and T. naphthophila, two thermophilic anaerobic isolates from a Japanese reservoir: Taxonomic and genomic considerations. *Systematic and Applied Microbiology* **2020**, *43*, 126132, doi:https://doi.org/10.1016/j.syapm.2020.126132.

100.   Gieg, L. Microbial communities in oil shales, biodegraded and heavy oil reservoirs, and bitumen deposits. *Microbial Communities Utilizing Hydrocarbons and Lipids: Members, Metagenomics and Ecophysiology; Springer International Publishing: Cham, Switzerland* **2018**, 1-21.

101.   Qiu, Y.-L.; Hanada, S.; Ohashi, A.; Harada, H.; Kamagata, Y.; Sekiguchi, Y. <i>Syntrophorhabdus aromaticivorans</i> gen. nov., sp. nov., the First Cultured Anaerobe Capable of Degrading Phenol to Acetate in Obligate Syntrophic Associations with a Hydrogenotrophic Methanogen. *Applied and Environmental Microbiology* **2008**, *74*, 2051-2058, doi:doi:10.1128/AEM.02378-07.

102.   Kuever, J. The Family Syntrophorhabdaceae. In *The Prokaryotes: Deltaproteobacteria and Epsilonproteobacteria*, Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2014; pp. 301-303.

103.   Dong, X.; Greening, C.; Rattray, J.E.; Chakraborty, A.; Chuvochina, M.; Mayumi, D.; Dolfing, J.; Li, C.; Brooks, J.M.; Bernard, B.B.; et al. Metabolic potential of uncultured bacteria and archaea associated with petroleum seepage in deep-sea sediments. *Nature Communications* **2019**, *10*, 1816, doi:10.1038/s41467-019-09747-0.

104.   Aitken, C.M.; Jones, D.M.; Maguire, M.J.; Gray, N.D.; Sherry, A.; Bowler, B.F.J.; Ditchfield, A.K.; Larter, S.R.; Head, I.M. Evidence that crude oil alkane activation proceeds by different mechanisms under sulfate-reducing and methanogenic conditions. *Geochimica et Cosmochimica Acta* **2013**, *109*, 162-174, doi:https://doi.org/10.1016/j.gca.2013.01.031.

105.   Davis, J.J.; Wattam, A.R.; Aziz, R.K.; Brettin, T.; Butler, R.; Butler, R.M.; Chlenski, P.; Conrad, N.; Dickerman, A.; Dietrich, E.M. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic acids research* **2020**, *48*, D606-D612.

106.   Wang, S.C. *Studies of bacterial catabolic enzymes: implications for the evolution of enzymes and metabolic pathways*; The University of Texas at Austin: 2003.

107.   Davidson, R.; Baas, B.-J.; Akiva, E.; Holliday, G.L.; Polacco, B.J.; LeVieux, J.A.; Pullara, C.R.; Zhang, Y.J.; Whitman, C.P.; Babbitt, P.C. A global view of structure–function relationships in the tautomerase superfamily. *Journal of Biological Chemistry* **2018**, *293*, 2342-2357, doi:https://doi.org/10.1074/jbc.M117.815340.

108.   Michał, B.; Gagat, P.; Jabłoński, S.; Chilimoniuk, J.; Gaworski, M.; Mackiewicz, P.; Marcin, Ł. PhyMet2: a database and toolkit for phylogenetic and metabolic analyses of methanogens. *Environmental Microbiology Reports* **2018**, *10*, 378-382, doi:https://doi.org/10.1111/1758-2229.12648.

109.   Borrel, G.; Adam, P.S.; McKay, L.J.; Chen, L.-X.; Sierra-García, I.N.; Sieber, C.M.K.; Letourneur, Q.; Ghozlane, A.; Andersen, G.L.; Li, W.-J.; et al. Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea. *Nature Microbiology* **2019**, *4*, 603-613, doi:10.1038/s41564-019-0363-3.

110.   Laso-Pérez, R.; Hahn, C.; Vliet, D.M.v.; Tegetmeyer, H.E.; Schubotz, F.; Smit, N.T.; Pape, T.; Sahling, H.; Bohrmann, G.; Boetius, A.; et al. Anaerobic Degradation of Non-Methane Alkanes by “<i>Candidatus</i> Methanoliparia” in Hydrocarbon Seeps of the Gulf of Mexico. *mBio* **2019**, *10*, e01814-01819, doi:doi:10.1128/mBio.01814-19.

111.   Lahme, S.; Callbeck, C.M.; Eland, L.E.; Wipat, A.; Enning, D.; Head, I.M.; Hubert, C.R.J. Comparison of sulfide-oxidizing Sulfurimonas strains reveals a new mode of thiosulfate formation in subsurface environments. *Environmental Microbiology* **2020**, *22*, 1784-1800, doi:<u>https://doi.org/10.1111/1462-2920.14894</u>.

112.   Dahle, H.; Roalkvam, I.; Thorseth, I.H.; Pedersen, R.B.; Steen, I.H. The versatile in situ gene expression of an E psilonproteobacteria-dominated biofilm from a hydrothermal chimney. *Environmental microbiology reports* **2013**, *5*, 282-290.

113.   Hensen, D.; Sperling, D.; Trüper, H.G.; Brune, D.C.; Dahl, C. Thiosulphate oxidation in the phototrophic sulphur bacterium Allochromatium vinosum. *Molecular microbiology* **2006**, *62*, 794-810.

114.   Zumft, W.; Kroneck, P.; Robert, K. Advances in Microbial Physiology, Vol. 52. *by RK Poole, Elsevier Ltd* **2007**, 107-227.