Novel clostridial lineages recovered from metagenomes of a hot oil reservoir

Glenn D. Christman1, Rosa I. León-Zayas1,2, Rui Zhao1, Zarath M. Summers3 & Jennifer F. Biddle1✉

Oil reservoirs have been shown to house numerous microbial lineages that differ based on the in-situ pH, salinity and temperature of the subsurface environment. Lineages of Firmicutes, including Clostridiales, have been frequently detected in oil reservoirs, but are typically not considered impactful or relevant due to their spore-forming nature. Here we show, using metagenomics, a high temperature oil reservoir of marine salinity contains a microbial population that is predominantly from within the Order Clostridiales. These organisms form an oil-reservoir specific clade based on the phylogenies of both 16S rRNA genes and ribosomal proteins, which we propose to name *U* *Petromonas tenebris*, meaning they are single-celled organisms from dark rocks. Metagenome-assembled genomes (MAGs) of these *Petromonas* sp. were obtained and used to determine that these populations, while capable of spore-formation, were also likely replicating in situ in the reservoir. We compared these MAGs to closely related genomes and show that these subsurface Clostridiales differ, from the surface derived genomes, showing signatures of the ability to degrade plant-related compounds, whereas subsurface genomes only show the ability to process simple sugars. The estimation of in-situ replication from genomic data suggest that *U* *Petromonas tenebris* lineages are functional in-situ and may be specifically adapted to inhabit oil reservoirs.

Oil reservoirs are unique subsurface environments that can be hostile to life due to high temperatures, pressures, and salinity (reviewed in1). It is thought that temperatures in source rocks or reservoirs greater than 80 °C act to deter hydrocarbon degradation and sterilize microbial life2,3. As sterile crude oil and gases migrate from the hot source rocks and cool they accumulate and may be populated by existing subsurface communities in rock or flowing subterranean waters. Determining the indigenous microbial community structures in these reservoirs can be difficult because of the presence of pipeline and infrastructure contaminants, and the indigenous communities themselves can be altered in terms of member representation and activity by the injection of seawater or gases and chemical additives used for secondary recovery4.

Microbes can cause souring or consume significant quantities of lower molecular weight hydrocarbons, degrading the quality of the petroleum products. Given the potential for degradation of resources, there is interest in understanding the structure and source of these communities, and the metabolisms they are employing5. Clostridiales are frequently found in oil reservoir surveys4–7, including in systems with temperatures greater than 50 °C8,9. Many observed Clostridiales are spore formers, which may give them an advantage in persisting in these systems with high temperatures, pressures, and salinity. However, this ability also means that they are often overlooked in regards to their influence on the in-situ microbial community, as it is unknown if these organisms are active in the reservoir or present in a dormant, sporulated form. Some clostridial species have been implicated as reservoir souring culprits8, yet many reservoir descriptions lack any explanation of clostridial activity subsurface4–10. As such, their role in the reservoir microbial community remains poorly understood. Most Clostridiales employ fermentation as a metabolic strategy, however, many *Desulfotomaculum* species respire using sulfate reduction, and some *Desulfotomaculum* have been shown to directly utilize aromatic hydrocarbons (e.g., toluene, m-xylene, o-xylene) as carbon and energy sources15. Hydrocarbon enrichment cultures from oil reservoirs have stimulated clostridial lineages16, and clostridial hydrocarbon degradation genes have been observed in surficial oil-contaminated environments17. Previous metagenomic studies of oil reservoirs show clostridial

1School of Marine Science and Policy, University of Delaware, Lewes, DE, USA. 2Department of Biological Sciences, Willamette University, Salem, OR, USA. 3ExxonMobil Research and Engineering, Annandale, NJ, USA. ✉e-mail: jfbiddle@udel.edu
lineages containing genes initially classified as hydrocarbon degradation genes, but well-characterized anaerobic hydrocarbon-degrading genes, bssA and assA, were not detected. Instead, robust signatures of polysaccharide, peptide, and fatty acid degradation were seen as well as robust pathways of sugar fermentation. To our knowledge, few studies have assessed the portion of sporulated and vegetative cells of Clostridiales in oil reservoirs. In this study, we sampled production wells from Galveston 209, a mature oil field ~32 km south of Galveston Island, Texas. Well temperatures of this oil field range from 80 to 160° Celsius. Discovered in 1983, the field was extensively developed off two platforms (A & B) with a cumulative production over 23 MBO (million barrels of oil) and 300 GCF (gas cubic feet) out of stacked pay of Lower to Middle Miocene shallow marine sandy reservoirs (~2.1 to 4.6 km depth) with outer shelf mudstone seals, all deposited in a wave-dominated deltaic setting. Production data indicates that there is a strong water drive through well-connected, continuous reservoirs. Traps are in low-relief fault-dependent closes on basinward-dipping listric normal faults. In this study, microbial communities in the produced fluids (mixture of oil and water) of these hot wells were examined via metagenomics. Functional annotation and the generation of metagenome assembled genomes (MAGs) were used to examine potential community metabolisms. A novel clostridial lineage which we propose to name *Petromonas tenebris* was the dominant microbial signature in this subsurface environment, and *in-situ* replication of this organism can be detected via metagenomics.

**Methods**

**Collection of produced fluids.** Produced fluids were collected using sterile technique at the drilling rig offshore Texas. Fluids came from 4 distinct wells that had temperatures ranging from 88 to 102° C (Table 1). Produced fluids were filtered through a Sterivex filter until the filter clogged, which was seen between 300-400 mL of fluid. The Sterivex filter was frozen immediately and stored at −80 °C until use.

**DNA extraction and sequencing.** DNA was extracted using a modified version of the Qiagen PowerWater Sterivex filter extraction kit. DNA was checked versus a blank control extraction for bacterial PCR products, and once it was determined that DNA was amplifiable and contained a microbial signal, it was sent for metagenomic library preparation and sequencing via Illumina HiSeq at the University of Delaware Genomic Sequencing Facility. Raw sequences and MAGs for this project are deposited at NCBI under Bioproject PRJNA578106.

**Quality trim and assembly.** Raw Illumina reads were quality trimmed in CLCBio Workbench version 7.5.1 (Qiagen), with the following parameters: removal of low quality sequence (limit = 0.0016, but rounded to 0.002 by CLCBio, which represents a Phred score of 36 or better); removal of ambiguous nucleotides: No ambiguous nucleotides allowed; removal of terminal nucleotides: 2-12 nucleotides from either end to minimize sequencing errors and enriched 5mers; removal of sequences on length: minimum length 60 nucleotides. Whenever one read of a read pair was excluded due to the quality trim, the entire pair was excluded. Trimmed, paired reads were assembled using IDBA-UD version 1.1.1. with the following settings:–mink 40–maxk 120–step 20–min_contig 300. The resulting scaffolds were then used for further genome binning of each reservoir metagenome.

**Phylogeny.** The phylogeny of metagenome community members was determined using EMIRGE (Expectation-Maximization Iterative Reconstruction of Genes from the Environment) based on the reconstructed 16S rRNA gene sequences from unassembled data. A maximum likelihood phylogenetic tree of 16S

| Chemistry    | Reservoir Well |
|--------------|----------------|
| ppm, wt/vol  | A2 | B6 | B7 | B9 |
| Cl−          | 30034 | 28298 | 39183 | 39106 |
| NO3−         | <1  | <1  | <1  | <1  |
| PO4−         | <1  | <1  | <1  | <1  |
| SO4−         | 9   | 22  | 38  | 31  |
| Glycolate    | 7   | 25  | 26  | 19  |
| Formate      | 3   | 8   | 8   | 6   |
| Acetate      | 626 | 1041| 806 | 606 |
| Propionate   | 43  | 88  | 105 | 87  |
| Butyrate     | 9   | 18  | 14  | 11  |
| Vurate       | 3   | 6   | 4   | 4   |
| Bicarbonate  | 297 | 294 | 206 | 85  |
| Ca           | 512 | 578 | 1206| 1323 |
| K            | 151 | 120 | 216 | 199 |
| Mg           | 101 | 122 | 308 | 361 |
| Ba           | 11.4| 3.9 | 3.8 | 6.7 |
| Si           | 21.7| 18.9| 14.8| 16.3 |
| Temperature °C | 102 | 91  | 88  | 91  |

Table 1. Selected reservoir produced fluid (water) chemical concentrations and reservoir temperatures.
omes and MAGs. COG (Clusters of Orthologous Groups) data was obtained via a local copy of the RAMMCAP.

Metagenome assembly of individual samples were subjected to binning using MaxBin version 1.4.2 with the max iteration of 200. The taxonomical uniqueness of each resulting MAG was initially determined by Phylosift version 1.0.1 with the “lineage_wf” option. The VizBin program was then used to visually refine the MAGs to minimize outlier scaffolds. Close relative genomes of the clostridial MAGs were downloaded from NCBI (C. sporogenes was run with default settings.

Metagenome-assembled genomes (MAGs). Metagenome assembly of individual samples were subjected to binning using MaxBin version 1.4.2 with the max iteration of 200. The taxonomical uniqueness of each resulting MAG was initially determined by Phylosift version 1.0.1 with the “lineage_wf” option. The VizBin program was then used to visually refine the MAGs to minimize outlier scaffolds. Close relative genomes of the clostridial MAGs were downloaded from NCBI (C. sporogenes was run with default settings.

The presence or absence of functional genes in metabolic pathways of both cultured metagenome and MAGs were predicted using the BlastKOALA web service provided by the KEGG: Kyoto Encyclopedia of Genes and Genomes website (http://www.kegg.jp/blastkoala/), between July and October, 2017.

Carbohydrate enzymes. The Carbohydrate Active Enzymes (CAZY) annotation was performed via HMMER searches against the dbCAN release 4.0 HMM (hidden Markov model) database (downloaded from http://cys.bios.niu.edu/dbCAN/), and based on the CAZyDB released on March 17, 2015.

Estimation of growth. Growth signatures were created with the iRep program, which estimates the proportion of actively replicating cells by comparing the read recruitment to the origin of replication. The program was run with default settings.

Results
Metagenomes were generated using DNA extracted from four samples taken from produced fluids from an oil field located in the Gulf of Mexico, offshore Texas. The reservoirs in this formation are hot and salty, with temperatures ranging between 88-102 °C, salinity values exceeding 28% and sulfate values below seawater levels. Four separate wells access likely connected reservoir material in this system, and each well has somewhat similar but distinct geochemical conditions (Table 1).

We generated metagenomic sequencing data from each well, and the metagenome assembly resulted in 9,600–34,300 contigs varying in total length of 18–43 million basepairs (Table 2).

Based on 16S rRNA gene sequences reconstructed by EMIRGE, the most abundant organism in each metagenome was the unique lineage, which comprised from 69.8% to 96.7% of the microbial community (Fig. 1). Petrotoga, Geotoga, Euryarchaea and other members of Clostridiales were less abundant members of the community (Fig. 1). The closest relatives were uncultivated lineages found in other oil reservoirs (Fig. 2).

Phylogenetic analysis suggested that the nearest cultured relatives of the dominant 16S rRNA gene from the oil reservoirs were Caminicella sporogenes and Paramadeivibacter caminithermalis (formerly Clostridium

| Metagenome | A2 | B6 | B7 | B9 |
|------------|----|----|----|----|
| Total ng of DNA extracted | 77 | 58 | 172 | 34 |
| Basepairs DNA Sequenced (Gbp) | 9.70 | 11.84 | 10.67 | 9.60 |
| Basepairs after Quality Control (Gbp) | 4.37 | 4.89 | 4.92 | 3.90 |
| Basepairs in Assembly | 17,973,987 | 22,158,355 | 42,651,179 | 39,671,450 |
| Number of Contigs | 9,646 | 14,826 | 34,316 | 25,137 |
| Number of Genes | 16,297 | 20,414 | 40,930 | 37,663 |
| Unique 16S rRNA genes (EMIRGE) | 6 | 6 | 14 | 13 |

Table 2. Metagenome Statistics.

rRNA gene was inferred from these sequences using Mega version 5 using default parameters for alignment and tree construction with 500 bootstrap replicates.
caminthermale), both of which are moderately thermophilic and halophilic and were initially isolated from deep-sea hydrothermal vent systems (Fig. 2).

Metagenome assembled genomes (MAGs) of \textit{\textsuperscript{15}Petromonas tenebris} lineages were recovered. These high-quality MAGs were 99.2% to 100.0% complete with <5% contamination, as assessed based on single-copy genes by CheckM (Table 3). The estimated sizes of the genomes range between 2.7–3.0 Mb, which ranges in between the relatives of \textit{Ca. sporogenes} (2.5 Mb) and \textit{P. carminithalis} (4.1 Mb). The number of annotated genes of the MAGs ranged from 2,815–2,959 (Table 3).

Phylogenetic analysis of concatenated ribosomal protein sequences shows that the clostridial MAGs found in these reservoir samples are most closely related to \textit{Caminicella sporogenes} and \textit{Paramaledibacter caminithermales}, along with \textit{Maledibacter halophilus} (Fig. 3). These closely related organisms form a group of thermophilic and halophilic Clostridia within the Clostridiales\textsuperscript{38,39}. Both 16S rRNA gene and concatenated ribosomal protein sequences were similar for each of the reservoir MAGs suggesting that the similar organisms were present in all four reservoir samples (Figs. 2 and 3).

Average nucleotide identities (ANI) are 99-100% among the 4 MAGs, but the MAGs share only 74-75% ANI and 51-61% average amino acid identity (AAI) with the cultured genomes (Table 4).

As such, we provisionally name the organism \textit{\textsuperscript{15}Petromonas tenebris} as the MAGs fit the suggested metrics for establishment of a new genus and species\textsuperscript{40}. The name comes from “petra” (rock), “monas” (single celled organism) and “tenebris” (dark) since the phylogenetic trees show relatives of these organisms are all found in oil reservoirs, the \textsuperscript{15} indicates its uncultivated status.
The analysis of Clusters of Orthologous Groups (COGs) shared with either Ca. sporogenes and P. carminithermalis show that individual MAGs of UPetromonas tenebris have different levels of overlap with cultured relatives, with an average of 323 and 197 unique genes, respectively (Table 3). Genes present in the MAGs that are different versus Ca. sporogenes

| MAG or Genome source | A2     | B6     | B7     | B9     | Ca. sporogenes | P. carminithermalis |
|----------------------|--------|--------|--------|--------|----------------|---------------------|
| Size (bp)            | 2,946,899 | 3,033,216 | 2,689,941 | 2,893,964 | 2,473,920 | 4,060,620 |
| Completeness %       | 100.0% | 100.0% | 100.0% | 99.2% | — | — |
| Implied Size (bp)*   | 2,946,899 | 3,033,216 | 2,689,941 | 2,917,302 | 2,473,920 | 4,060,620 |
| Contamination %      | 4.14% | 1.65% | 1.42% | 1.65% | — | — |
| Strain Heterogeneity | 0.00 | 0.00 | 0.00 | 0.00 | — | — |
| Number of Contigs    | 121 | 85 | 146 | 140 | 43 | 155 |
| Number of Genes      | 2,874 | 2,959 | 2,815 | 2,844 | 2,422 | 3,378 |

Table 3. **Petromonas tenebris** MAG statistics and comparisons to cultured relative genomes. *Implied genome size equals MAG size divided by completeness.

The analysis of Clusters of Orthologous Groups (COGs) shared with either Ca. sporogenes and P. carminithermalis show that individual MAGs of **Petromonas tenebris** have different levels of overlap with cultured relatives, with an average of 323 and 197 unique genes, respectively (Table 3). Genes present in the MAGs that are different versus Ca. sporogenes

| versus Ca. sporogenes |
|-----------------------|
| Number of COGs        | 1,514 | 1,526 | 1,514 | 1,510 | 1,355 |
| COGs Unique to MAG    | 323   | 330   | 322   | 315   | —     |
| COGs Same as Genome   | 1,191 | 1,196 | 1,192 | 1,195 | 1,355 |

versus P. carminithermalis

| versus P. carminithermalis |
|----------------------------|
| Number of COGs            | 1,514 | 1,526 | 1,514 | 1,510 | 1,375 |
| COGs Unique to MAG        | 201   | 204   | 193   | 189   | —     |
| COGs Same as Genome       | 1,313 | 1,322 | 1,321 | 1,321 | 1,375 |

Figure 3. Phylogeny of **Petromonas tenebris** and related genomes based on the concatenated 16 ribosomal proteins: RpL2, 3, 4, 5, 6, 14, 15, 16, 18, 22, and 24, and RpS3, 8, 10, 17, and 19. The tree was reconstructed using the maximum likelihood algorithm with 100 iterations.
than those found in the isolate genomes include a number of oxidoreductases, including Fe-S oxidoreductase, citrate lyase, and CO dehydrogenase-CoA synthase subunits. There were also between 8-13 COGs unique to each MAG that did not appear in any other MAG or cultured genome (Fig. 4). These unique genes were primarily housekeeping genes except that \textit{UPetromonas tenebris} B7 had a unique COG1719, which is a predicted hydrocarbon binding protein.

The potential phylogenetic history of functional genes in the reservoir MAGs was examined to determine if they had any unusual evolutionary histories. Each gene was individually examined via BLAST and the taxonomy of the best hit was recorded. About 48% of the coding genes in the reservoir MAGs appear to have originated within the \textit{Caminicella}, \textit{Maledivibacter}, and \textit{Paramaledivibacter} genera, with the remainder coming from other Clostridiales or Firmicutes groups (Fig. 5).

Many metabolic pathways are shared between the reservoir MAGs and their cultured relatives. Shared core metabolisms include butyrate fermentation, sporulation, and common two component regulatory systems involving temperature, salt stress, chemotaxis, and flagella regulation. However, many differences between the MAGs and the closely related cultured organisms exist. For example, there are differences in Carbohydrate Active Enzymes (CAZY) between each of the cultured genomes and between the cultured genomes and the reservoir MAGs (Table 5). The reservoir MAGs have unique genes and transporters involved in the metabolism of sucrose.

|           | Ca. sporogenes | \textit{P. caminithermalis} | A2 | B6 | B7 | B9 |
|-----------|----------------|---------------------------|----|----|----|----|
| Average Amino acid identity AAI (1 way) | Ca. sporogenes | — | 51.5 | 56.6 | 56.3 | 57.1 | 57.1 |
|           | \textit{P. caminithermalis} | 61.1 (67.3) | — | 57.9 | 57.3 | 58.5 | 58.0 |
|           | A2 | 60.2 (69.4) | 51.8 (66.8) | — | 95.8 | 95.4 | 95.5 |
| B6 | 60.8 (69.4) | 52.0 (66.7) | 97.3 (99.7) | — | 96.2 | 96.7 |
| B7 | 60.4 (69.5) | 51.9 (67.1) | 93.6 (99.6) | 93.1 (99.6) | — | 94.3 |
| B9 | 60.8 (69.8) | 51.9 (67.0) | 94.8 (99) | 94.5 (99.6) | 95.3 (99.7) | — |

Table 4. ANI and AAI values of \textit{Petromonas tenebris} MAGs versus isolate genomes.
which are not present in the cultured genomes. Conversely, the cultured isolate genomes are capable of processing extracellular cellobiose and xylan/xylose, while the reservoir MAGs lack these genes.

Other key metabolic differences include a complete Wood–Ljungdahl CO₂ fixation pathway found in three of the four reservoir MAGs, but not in the other genomes (Table 6). COGs of the key enzyme, CO dehydrogenase/acyetyl CoA synthase, noted above, were not found in the cultured genomes. Certain sulfur metabolism genes are present in the reservoir MAGs but not in the cultured genomes. These include anaerobic sulfite reductase (ASR), adenosine 5′-phosphosulfate reductase (APR), and sulfite reductase (ferredoxin). ASR is typically part of the assimilatory sulfate reduction pathway, and APR is typically present in dissimilatory pathways, but neither of these pathways is complete in the MAGs. In addition, neither of the \textit{qmoABC} and \textit{dsrMKJOP} electron transport complex genes typically found in sulfate reducers are present in the MAGs. As a result, it does not appear that any of the reservoir MAGs and cultured genomes can perform dissimilatory sulfate reduction for energy conservation, but instead use sulfur compounds as electron sinks for fermentation.

Both \textit{Caminicella sporogenes} and \textit{Paramaledivibacter caminithermalis} contain a glycyl radical enzyme with the same active site as the gene annotated as a pyruvate formate lyase (\textit{pflD}, locus tag AF1449) found in \textit{Archaeoglobus} species, which may be used to anaerobically metabolize some hydrocarbons. However, the annotated pyruvate formate lyase genes found in the \textit{UPetromonas tenebris} MAGs contain a different active site that was found in more typical carbohydrate fermenters, leading to the conclusion that these are indeed typical pyruvate formate lyase genes used in the majority of anaerobic bacteria. We examined the metabolic profile of the MAGs via standard annotation by PROKKA and also via individual BLAST analysis for known anaerobic hydrocarbon degradation genes. We found no evidence for other anaerobic hydrocarbon metabolism pathways, except that the gamma subunit of acetophenone carboxylase, an enzyme in the ethylbenzene degradation pathway, was present in each of the MAGs. However, the gene encoding ethylbenzene dehydrogenase, the initial enzyme in this pathway, could not be found. As such, based on current genomes and annotations, \textit{UPetromonas tenebris} seems not to be capable of utilizing hydrocarbons. We do caveat that as seen in the \textit{Archaeoglobus} case, genes may exist for hydrocarbon degradation that are either unannotated or misannotated and full proof cannot be given until a culture is tested.

To determine if the \textit{UPetromonas tenebris} MAGs came from spores or active, vegetative cells we estimated the index of replication (iRep) of the MAGs in the four oil reservoirs. The resulting iRep indexes ranged from 1.35–1.40 for the reservoir MAGs. For reference, this iRep value is comparable to the median values seen in other environments, including soil (1.34) and human gut systems (1.37–1.42), and should be interpreted as a measurement indicative of growth. The value indicates 35–40% of these cells were replicating \textit{in situ} at time of sampling.

**Discussion**

The hot and salty oil reservoirs described here represent a challenging environment for microbial growth. Using metagenomic analysis, we found the dominant species in these reservoirs is related to the thermophilic \textit{Caminicellal/Paramaledivibacter} clades of thermophilic and halophilic Clostridiales, and form a distinct clade with other uncultured organisms found in high-temperature oil reservoirs. Due to these unique features,
we propose to name this lineage *UPetromonas tenebris* (Figs. 2 and 3). Related species have been detected in other oil well systems, including *Ca. sporogenes* in Oman at temperatures near 60 °C and formation water from the high temperature Ekofisk oil field in the North Sea. While these relatives do not grow at temperatures as high as observed in this environment, hyperthermophilic clostridial species have been documented from oil wells previously. However, no genomic information is available for comparison to these MAGs.

It is unlikely that these MAGs represent infrastructure contaminants or sporulated, inactive cells. First, 35–40% of these cells were in the process of replication. Additionally, the DNA was readily extractable and this species dominated the community, also suggesting it was not heavily sporulated. Considering the apparent lack of hydrocarbon consumption by these cells, it is plausible that clostridial spores present in the subsurface may have germinated as the oil seeped upward from hotter source rocks to a slightly cooler reservoir formation, where water was present. However, we interpret the low diversity of the community to be reflective of the challenging in-situ environment. We cannot refute the hypothesis that spores may be germinating en route to oil processing inside pipelines, however, we note that communities which show infrastructure influence are typically much more complex, reflecting the increase in electron acceptors and metals available within pipelines, as well as industrially introduced materials. As such, the entirety of the data presented suggests these are active in-situ.

Other organisms in this system in lower abundance (Fig. 1) include *Methanothermococcus* methanogens which use hydrogen and formate as electron donors, and *Desulfalls* species such as *D. gibsonae* and *D. geothermicum*, which utilize simple organic compounds, including some carbohydrates and/or fatty acids, and alcohols such as ethanol, propanol, and butanol, as electron donors and sulfur compounds as electron acceptors, producing carbon dioxide or acetate as end products. Also present is *Petrotoga mobilis*, which is a fermenter of a variety of carbohydrates including xylan. This system may be a syntrophic methanogenic system in which the *UPetromonas* and *Petrotoga* ferment complex organic compounds, with *Desulfalls* and *Methanothermococcus* scavenging the fermentation products. Based on the analysis of genomes, no alkane metabolizing partner has been detected, compared to other syntrophic methanogenesis systems that were explored via enrichment.

Table 5. Differences in Carbohydrate Active Enzymes (CAZY) between the *UPetromonas tenebris* MAGs and the *Caminicella sporogenes* and *Paramadevibacter caminithermalis* genomes. GH = Glycoside Hydrolases, CBM = Carbohydrate Binding Modules, CE = Carbohydrate Esterases, and GT = Glycosyltransferases.
Therefore, *U. Petromonas tenebris* represent one of the keystone members of the microbial community inhabiting this harsh environment.

Despite being retrieved from a similar geologic formation, the *U. Petromonas tenebris* MAGs show some distinctions between wells. In particular, they contained between 8-13 unique genes per genome (Fig. 4). They were clearly differentiated from their nearest cultivated neighbors by both ribosomal sequences (Figs. 2 and 3) and genome size and content (Tables 1 and 3, Fig. 4). The variations in genome all seem to be of clostridial origin (Fig. 5). Overall the MAGs in this reservoir system represent a single species, supported by phylogenies (Figs. 2 and 3) and ANI calculation (Table 2), despite the unique genome values mentioned previously.

The *U. Petromonas tenebris* MAGs are the most abundant organisms in this reservoir system, they show signatures of replication, and potential end products of their fermentative metabolism have built up in this system. Yet they lack potential hydrocarbon processing pyruvate formate lyase genes seen in the cultured relatives, and none of the other less abundant organisms have signatures of directly processing hydrocarbons either. Nevertheless, the *U. Petromonas tenebris* in this reservoir have become a dominant species in the environment, likely through fermentation, are growing in-situ and may be responsible for the increase of organic acids in the reservoirs. Due to their potential influence in reservoirs, Firmicutes and clostridial relatives should be regarded as key players in reservoir microbiomes, not only as sporulated, inactive cells. Furthermore, presence of these organisms in

| Metabolism                          | Caminicella sporogenes | Paramaledavibacter caminitermalis | A2 *P. tenebris* MAG | B6 *P. tenebris* MAG | B7 *P. tenebris* MAG | B9 *P. tenebris* MAG |
|-------------------------------------|------------------------|----------------------------------|---------------------|---------------------|---------------------|---------------------|
| Wood-Ljundahl Pathway               |                        |                                  | X                   | X                   | X                   | X                   |
| Sulfur metabolism                   |                        |                                  |                     |                     |                     |                     |
| Anaerobic Sulfite Reductase         |                        | X                                | X                   | X                   | X                   | X                   |
| Aden-5’phosphosulfate Reduct        |                        | X                                | X                   | X                   |                     |                     |
| Sulfite reductase (ferrodoxin)      |                        | X                                | X                   | X                   | X                   | X                   |
| Butanoate Metabolism                |                        |                                  |                     |                     |                     |                     |
| Pyruvate <→ Acetyl CoA              | X                      | X                                |                     |                     |                     |                     |
| Acetyl CoA <→ AcetoacetylCoA        | X                      | X                                |                     |                     |                     |                     |
| Acetyl CoA <→ Butanoate             | X                      | X                                |                     |                     |                     |                     |
| Acetyl CoA <→ Butanol               | X                      | X                                |                     |                     |                     |                     |
| Starch and Sucrose                  |                        |                                  |                     |                     |                     |                     |
| Extracellular Cellubiose            | X                      | X                                |                     |                     |                     |                     |
| Starch/Glycogen > Glucose           | X                      | X                                |                     |                     |                     |                     |
| Sucrose > Glucose                   | X                      | X                                |                     |                     |                     |                     |
| Membrane Transport                  |                        |                                  |                     |                     |                     |                     |
| Phosphonate                         | X                      | X                                |                     |                     |                     |                     |
| Liposaccharide                      | X                      | X                                |                     |                     |                     |                     |
| Glycine betaine/proline             | X                      | X                                |                     |                     |                     |                     |
| Cobalt                              | X                      | X                                |                     |                     |                     |                     |
| Nickel                              | X                      | X                                |                     |                     |                     |                     |
| Galactitol                          | X                      | X                                |                     |                     |                     |                     |
| Mannitol                            | X                      | X                                |                     |                     |                     |                     |
| Sorbitol                            | X                      | X                                |                     |                     |                     |                     |
| Sucrose                             | X                      | X                                |                     |                     |                     |                     |
| Amino Acid                          | X                      | X                                |                     |                     |                     |                     |
| Branched Chain Amino Acids          | X                      | X                                |                     |                     |                     |                     |
| Bacitracin/Lantibiotics              | X                      | X                                |                     |                     |                     |                     |
| Cellubiose/diacetylchitobiose       | X                      | X                                |                     |                     |                     |                     |
| Complete D-xylose                   | X                      | X                                |                     |                     |                     |                     |
| Galactose oligomer                  | X                      | X                                |                     |                     |                     |                     |
| N-acetyl-D-glucosamine              | X                      | X                                |                     |                     |                     |                     |
| Two-component systems               |                        |                                  |                     |                     |                     |                     |
| Temperature                         | X                      | X                                |                     |                     |                     |                     |
| Salt Stress                         | X                      | X                                |                     |                     |                     |                     |
| Chemotaxis                          | X                      | X                                |                     |                     |                     |                     |
| Flagella regulon genes              | X                      | X                                |                     |                     |                     |                     |
| Tricarboxylate transport            | X                      | X                                |                     |                     |                     |                     |
| Citrate formation genes             | X                      | X                                |                     |                     |                     |                     |

Table 6. Differences in metabolism between the MAGs and cultured isolate genomes.
produced fluids, drilling fluids, or naturally seeped fluids could be an indicator of the temperature of a connected subsurface reservoir, with applications to oil and gas exploration and development.

References

1. Pannekens, M., Kroll, L., Müller, H., Mbow, F. T. & Meckenstock, R. U. Oil reservoirs, an exceptional habitat for microorganisms. New Biotechnol. 49, 1–9 (2019).
2. Head, I. M., Jones, D. M. & Larer, S. R. Biological activity in the deep subsurface and the origin of heavy oil. Nature 426, 344–352 (2003).
3. Wilhelms, A. et al. Biodegradation of oil in uplifted basins prevented by deep-burial sterilization. Nature 411, 1034–1037 (2001).
4. Vigneron, A. et al. Succession in the petroleum reservoir microbiome through an oil field production lifecycle. ISME J. 11(9), 2141–2154 (2017).
5. Berdugo-Clavijo, C. & Gieg, L. M. Conversion of crude oil to methane by a microbial consortium enriched from oil reservoir production waters. Front. Microbiol. 5, 197, https://doi.org/10.3389/fmicb.2014.00197 (2014).
6. Fowler, S. J., Toth, C. R. A. & Gieg, L. M. Community structure in methanogenic enrichments provides insight into syntrophic interactions in hydrocarbon-impacted environments. Front. Microbiol. 7, 562 (2016).
7. Yousef, N., Elshahed, M. S. & McInerney, M. J. Microbial processes in oil fields: culprits, problems, and opportunities. Adv. Appl. Microbiol. 66, 141–251 (2009).
8. Kim, D. D. et al. Microbial community analyses of produced waters from high-temperature oil reservoirs reveal unexpected similarity between geographically distant oil reservoirs. Microb. Biotechnol. 11, 788–796 (2018).
9. Rosnes, I. T., Torsvik, T. & Lien, T. Spore-forming thermophilic sulfate-reducing bacteria isolated from North Sea oil field waters. Appl. Environ. Microbiol. 57(8), 2302–2307 (1991).
10. Gittel, A., Sorensen, K. B., Skovhus, T. L., Ingvorsen, K. & Schramm, A. Prokaryotic Community Structure and Sulfate Reducer Activity in Water from High-Temperature Oil Reservoirs with and without Nitrate Treatment. Appl. Environ. Microbiol. 75(22), 7086–7096 (2009).
11. 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(2), 204–212 (2012).
12. 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. Int. Biodeter. Biodegr. 81, 57–70 (2013).
13. Gao, P. et al. Spatial isolation and environmental factors drive distinct bacterial and archaeal communities in different types of petroleum reservoirs in China. Sci. Rep. 6, 20174, https://doi.org/10.1038/srep20174 (2016).
14. Hu, P. et al. Genome-Resolved Metagenomic Analysis Reveals Roles for Candidate Phyla and Other Microbial Community Members in Biogeochemical Transformations in Oil Reservoirs. MBIO 7(1), e01669–15 (2016).
15. Morasch, B., Schink, B., Tebbe, C. C. & Meckenstock, R. U. Degradation of α-xylene and m-xylene by a novel sulfate-reducer belonging to the genus Desulfotomaculum. Arch. Microbiol. 181, 407–417 (2014).
16. Tan, B. et al. Comparative analysis of metagenomes from three methanogenic hydrocarbon-degrading enrichment cultures with 41 environmental samples. The ISME J. 9(9), 2028–45 (2015).
17. Michas, A. et al. More than 2500 years of oil exposure shape sediment microbiomes with the potential for syntrophic degradation of hydrocarbons linked to methanogenesis. Microbiome 5, 118 (2017).
18. Regberg, A. B., Summers, Z. M., N’Guessan, A. L., Kirkpatrick, J. & D’Hondt S. Methods for Isolating Nucleic Acids from Samples. US Patent Application US20170342467 (2017).
19. Peng, Y., Leung, H. C., Yu, S. M. & Chin, F. Y. IDRA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28, 1420–1428 (2012).
20. Miller, C. S., Baker, B. J., Thomas, B. C., Singer, S. W. & Banfield, J. F. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. Genome Biol. 12, R44 (2011).
21. Tamura, K. et al. MEGAS: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol. Biol. Evol. 28, 2731–2739 (2011).
22. Wu, Y. W., Tang, Y. H., Tringe, S. G., Simmons, B. A. & Singer, S. W. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization maximization algorithm. Microbiome 2, 26 (2014).
23. Darling, A. E. et al. Phylodyn: phylogenetic analysis of genomes and metagenomes. PeerJ 2, e243, https://doi.org/10.7717/peerj.243 (2014).
24. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1053 (2015).
25. Lacroix, C. C. et al. ViSBin - an application for reference-independent visualization and human-annotated binning of metagenomic data. Microbiome 3, 1 (2015).
26. Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G. & Toth, I. K. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal. Methods 8, 12–24 (2016).
27. Eren, A. M. et al. Anvi’o: an advanced analysis and visualization platform for ‘omics data. PeerJ 3, e1319, https://doi.org/10.7717/peerj.1319 (2015).
28. Hug, L. A. et al. Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. Microbiome 1, 22 (2013).
29. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30(14), 3059–66 (2002).
30. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. 26(7), 1641–50 (2009).
31. Seemann, T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30(14), 2068–9 (2014).
32. Li, W. Analysis and comparison of very large metagenomes with fast clustering and functional annotation. BMC Bioinformatics 10, 359 (2009).
33. Kanesha, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. 428, 726–731 (2016).
34. Yin, Y. et al. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 40, W445–W451 (2012).
35. Cantarel, B. L. et al. The Carbohydrate-Active EnZymes database (CAzy): an expert resource for Glycomicrobiomes. Nucleic Acids Res. 37, D233–D238 (2009).
36. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, D490–D495 (2014).

Published online: 15 May 2020
37. Brown, C. T., Olm, M. R., Thomas, B. C. & Banfield, J. F. Measurement of bacterial replication rates in microbial communities. Nature Biotechnol. 34, 1256–1263 (2016).

38. Alain, K. et al. Caminicella sporogenes gen. nov., sp. nov., a novel thermophilic spore-forming bacterium isolated from an East-Pacific Rise hydrothermal vent. Int. J. Syst. Evol. Micr. 52, 1621–1628 (2002).

39. Li, G., Zheng, X., Liu, X., Zhang, X. & Shao, Z. Wukongibacter baidiensis gen. nov., sp. nov., an anaerobic bacterium isolated from hydrothermal sulfides, and proposal for the reclassification of the closely related Clostridium halophilum and Clostridium caminithermale within Maledrivibacter gen. nov. and Paramaledivibacter gen. nov., respectively. Int. J. Syst. Evol. Micr. 66, 4355–4361 (2016).

40. Konstantinidis, K. T., Rosselló-Móra, R. & Amann, R. Uncultivated microbes in need of their own taxonomy. ISME J. 11(11), 2399–2406 (2017).

41. Khelifi, N. et al. Anaerobic oxidation of long-chain n-alkanes by the hyperthermophilic sulfate-reducing archaeon, Archaeoglobus fulgidus. ISME J 8, 2153–2166 (2014).

42. Rabus, R. et al. Anaerobic microbial degradation of hydrocarbons: from enzymatic reactions to the environment. J. Mol. Microbiol. Biotechnol. 26, 5–28 (2016).

43. Heider, J., Spormann, A. M., Beller, H. R. & Widdel, F. Anaerobic bacterial metabolism of hydrocarbons. FEMS Microbiol. Reviews 22, 459–473.

44. Heider, J. & Schüle, K. Anaerobic Biodegradation of Hydrocarbons Including Methane in Rosenberg et al. (eds.), The Prokaryotes – Prokaryotic Physiology and Biochemistry, https://doi.org/10.1007/978-3-642-30141-4_80 (2013)

45. Al-Bahry, S. N. et al. Microbial Consortia in Oman Oil Fields: A Possible Use in Enhanced Oil Recovery. Int. Microbiol. Biotechnol. 23(1), 106–117 (2013).

46. Kaster, K. M., Bonaunet, K., Berland, H., Ketselaer-Ieltersen, G. & Brakstad, O. G. Characterisation of culture-independent and -dependent microbial communities in a high-temperature offshore chalk petroleum reservoir. Anton. Leeuw 96, 423–439 (2009).

47. Arora, P. & Ranade, D. R. Anaerobic Oxidation of Long-Chain N-Alkanes by the Hyperthermophilic Sulfate-Reducing Archaeon, Archaeoglobus Fulgidus. ISME J. 11(11), 2967–2973 (2017).

48. Whitman, W. B. Metanothermococcus gen. nov. Bergey’s Manual of Systematics of Archaea and Bacteria, Online. https://doi.org/10.1002/gbm00560 (2015).

49. Kuever, J. & Rainy F. A. Desulfotomaculum. Bergey’s Manual of Systematics of Archaea and Bacteria, Online. https://doi.org/10.1002/gbm00661 (2015).

50. Davey, M. E., MacGregor, B. J. & Stahl D. A. Petrotoga. Bergey’s Manual of Systematics of Archaea and Bacteria, Online. https://doi.org/10.1002/gbm00503 (2015).

51. Gray, N. D. et al. The quantitative significance of Syntrophaceae and syntrophic partnerships in methanogenic degradation of crude oil alkanes. Environ. Microbiol. 13(11), 2967–2975 (2011).

52. Summers, Z. M., N’Guessan, A. L., & Regberg, A. B. Methods to Determine Conditions of a Hydrocarbon Reservoir. US Patent Application US2018003690A1 (2017).

Acknowledgements
This work was funded by ExxonMobil Research and Engineering and ExxonMobil Upstream Research Company. R.L.Z. was funded by a postdoctoral fellowship from the NSF STC C-DEBI. R.Z. and J.F.B. were funded by the W.M. Keck Foundation. The authors acknowledge Clifford Walters for his technical support.

Author contributions
G.D.C. performed analyses and wrote the manuscript. R.L.Z., R.Z., Z.S. contributed data and analyses to the manuscript. J.F.B. directed the study and wrote the manuscript. All authors edited the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to J.F.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020