microLife, 2025, **6**, uqae027
DOI: 10.1093/femsml/uqae027
Advance access publication date: 23 January 2025
Research Article

Metagenomic analysis of pristine oil sheds new light on the global distribution of microbial genetic repertoire in hydrocarbon-associated ecosystems

Julia Plewka [1,2], Armando Alibrandi[3], Till L. V. Bornemann[1], Sarah P. Esser[1], Tom L. Stach[1,4], Katharina Sures [1], Jannis Becker[1], Cristina Moraru[1], André Soares[1], Rolando di Primio[5], Jens Kallmeyer[3], Alexander J. Probst [1,2,4,6,*]

[1]Environmental Metagenomics, Research Center One Health Ruhr of the University Alliance Ruhr, Faculty of Chemistry, University of Duisburg-Essen, 45141 Essen, Germany
[2]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Cyclotron Road, Berkeley, CA 94720, United States of America
[3]GFZ German Research Centre for Geoscience, Telegrafenberg, 14473 Potsdam, Germany
[4]Centre of Water and Environmental Research (ZWU), University of Duisburg-Essen, 45141 Essen, Germany
[5]Aker BP, NO-1324 Lysaker Norway
[6]Centre of Medical Biotechnology (ZMB), University of Duisburg-Essen, 45141 Essen, Germany
*Corresponding author. Environmental Metagenomics, Research Center One Health Ruhr of the University Alliance Ruhr, Faculty of Chemistry, University of Duisburg-Essen, 45141 Essen, Germany. E-mail: alexander.probst@uni-due.de
Editor: [Axel Brakhage]

Abstract

Oil reservoirs are society's primary source of hydrocarbons. While microbial communities in industrially exploited oil reservoirs have been investigated in the past, pristine microbial communities in untapped oil reservoirs are little explored, as are distribution patterns of respective genetic signatures. Here, we show that a pristine oil sample contains a complex community consisting of bacteria and fungi for the degradation of hydrocarbons. We identified microorganisms and their pathways for the degradation of methane, *n*-alkanes, mono-aromatic, and polycyclic aromatic compounds in a metagenome retrieved from biodegraded petroleum encountered in a subsurface reservoir in the Barents Sea. Capitalizing on marker genes from metagenomes and public data mining, we compared the prokaryotes, putative viruses, and putative plasmids of the sampled site to those from 10 other hydrocarbon-associated sites, revealing a shared network of species and genetic elements across the globe. To test for the potential dispersal of the microbes and predicted elements via seawater, we compared our findings to the Tara Ocean dataset, resulting in a broad distribution of prokaryotic and viral signatures. Although frequently shared between hydrocarbon-associated sites, putative plasmids, however, showed little coverage in the Tara Oceans dataset, suggesting an undiscovered mode of transfer between hydrocarbon-affected ecosystems. Based on our analyses, genetic information is globally shared between oil reservoirs and hydrocarbon-associated sites, and we propose that currents and other physical occurrences within the ocean along with deep aquifers are major distributors of prokaryotes and viruses into these subsurface ecosystems.

Keywords: metagenomics; oil microbiome; subsurface; viruses; dispersal; plasmids

Introduction

Oil reservoirs have been of great interest due to the still-increasing global energy demand and climate change. Microorganisms within these reservoirs must withstand a range of different temperatures, salinity, high pressure, and low water activity, with temperature having a strong influence on the community structure (Pannekens et al. 2019). For instance, the greatest diversity of microbial communities was discovered in oil reservoirs with moderate temperatures compared to those with high temperatures (Vigneron et al. 2017). The combination of abiotic factors in reservoirs is apparently heavily associated with microbial community compositions, as in some instances, they explain at least 64% of the variance of the microbial community structure (Gao et al. 2016).

Denitrifiers, sulfate-reducing bacteria, and methanogens from different bacterial and archaeal taxa are commonly found in oil reservoirs and heavily oil-polluted sites (Lin et al. 2014, Cai et al. 2015, Seitz et al. 2019, Semenova et al. 2020). They utilize petroleum hydrocarbons alongside sulfur and nitrogen compounds as energy and carbon sources (Reyes-Avila et al. 2013, Heider et al. 2016). Regarding Eukarya, fungi have not yet been described in oil reservoirs but were found in oil-polluted environments (Asemoloye et al. 2020, Liu et al. 2023). Cocultivation of fungi and bacteria resulted in increased biodegradation since fungi can decrease the toxicity of the hydrocarbons (Atakpa et al. 2022).

Understanding the oil microbiome has revealed great potential for application in bioremediation (Xu et al. 2018, Benedek et al. 2021) and enhanced oil extraction mechanisms (Youssef et al. 2009). Most metagenomic analyses of oil reservoirs were conducted after fluid injections (also termed water flooding) or from production water samples (Li et al. 2013, Hu et al. 2016, Wang et al. 2019, Yang et al. 2019). Reservoir exploitation can lead to drastic changes in reservoir conditions, water flooding decreases temper-

Received 21 January 2024; **revised** 23 October 2024; **accepted** 22 January 25

ature and salinity while dissolved oxygen increases, which has a huge impact on microbial diversity and on community composition (Gao et al. 2016, Bian et al. 2023). Comparisons of microbial communities within water-flooded oil reservoirs found high similarities on genus level despite large geological distances (Gao et al. 2016, Karthikeyan et al. 2019). It remains unknown if these similarities also apply to pristine reservoirs, which are thought to have little to no contact with surrounding environments (Pannekens et al. 2019). By contrast, two studies suggest that specific microorganisms (e.g. *Thermotoga*) can potentially infiltrate hydrocarbon reservoirs from the surrounding seawater and that production water has a substantial influence on the pristine microbial community composition over time (Nesbø et al. 2015, Vigneron et al. 2017). However, the mode of distribution of mobile genetic elements (MGEs; e.g. prospective DNA sequences of plasmids and viruses), remains little explored for hydrocarbon-affected sites and/or oil reservoirs. MGEs which usually have higher evolutionary pressure and thus mutate faster (Sobecky and Hazen 2009, Sanjuán et al. 2010), might be better suited for distribution studies in globally distinct sites.

Here, we investigated the microbial community of a pristine oil reservoir Well 7219/12–1, which had a temperature of 53°C and was drilled in the Filicudi prospect in the Barents Sea. The petroleum in the reservoir was sampled using subsurface tools (MDT chamber), ensuring the acquisition of a pristine oil sample. This oil sample is herein called "Filicudi" based on the name of the petroleum prospect. The reservoir was encountered at roughly 1600 m below the sea surface (with an additional water column of 300 m). After DNA extraction, sequencing, and metagenome assembly, we reconstructed 17 metagenome-assembled genomes (MAGs), including 15 bacterial genomes of 12 different species-level lineages, and two fungal genomes. Various hydrocarbon-degrading pathways were found within the MAGs, indicating the potential for biodegradation in the reservoir. The oil reservoir also harbored 58 distinct viruses and 39 plasmids partially carrying auxiliary/metabolic genes for different metal tolerances. Sequence comparisons of the microbial taxonomic units, viruses, and plasmids of the reservoir to 10 other hydrocarbon-associated sites, including water-flooded oil reservoirs, and to the Tara Ocean dataset (Pesant et al. 2015) confirm the pristine state of our samples and revealed global patterns of shared microorganisms and MGEs in hydrocarbon-associated sites.

Methods

Sample preparation, DNA extraction, and metagenome sequencing

Three oil samples in biological duplicates originating from the Filicudi oil reservoir (Barents Sea 72° 12' 44.21" N; 19° 46' 38.32" E) were provided by AkerBP (Lysaker, Norway). The reservoir had a temperature of 53°C, and the oil had an API gravity of 17. The samples were placed in sterile PTFE (polytetrafluoroethylene)-capped bottles and filled to capacity to prevent air contamination. Upon arrival at GFZ Potsdam, the samples were cooled at 4°C. They were opened to transfer their content in reaction tubes exclusively in an anaerobic glove box to prevent aerobic growth during long-term storage.

Sample processing within the reaction tubes was carried out in a fumehood to avoid exposure to volatile hydrocarbons from the oils and the solvents used in the DNA extraction.

DNA was extracted using the isooctane method described by Alibrandi et al. (2023). This involved mixing 25 ml of crude oil with an equal volume of isooctane (2,2,4-trimethylpentane), followed by centrifugation at 5000 *g* (Sigma 6–16KS, Sigma-Aldrich, St. Louis, USA). After centrifugation, the supernatant was discarded, and the pellet was added to the bead tubes of the DNA extraction kit (DNeasy PowerSoil Pro Kit, Qiagen, Hilden, Germany) (catalog number 47014). To each bead tube, 40 µl of 10% sodium dodecyl sulfate was added, followed by incubation at 65°C for 10 min. Subsequently, we followed the manufacturer's protocol for the DNA extraction kit. Due to the small volume of crude oil available and the high amount of DNA required for metagenomic analysis, the duplicates were merged for sequencing (Eurofins Genomics Europe Sequencing GmbH, Constance, Germany), and three metagenomes were obtained. Samples were sequenced at Eurofins Genomics Europe Sequencing GmbH with an Illumina NovaSeq sequencer and paired end-run type. The sequencing parameters included a read length of 2 × 150 bp, with a guaranteed yield of 10 million read pairs (20 million reads) per package (± 3%).

Metagenomic analysis

Raw reads from shotgun metagenomic sequencing and those downloaded from NCBI were treated similarly in the downstream analyses. If not otherwise stated, software was used with default parameters. Illumina artifacts and adapters were removed from raw reads with BBDuk (version 37.09) (Brian Bushnell 2014). Reads were quality-controlled and trimmed with sickle (version 1.33) (Joshi and Fass 2011). Quality-controlled reads were assembled using the MetaviralSPAdes pipeline (version 3.14.0) (Antipov et al. 2020) to recover prospective DNA sequences of viruses and plasmids. Assembled reads were filtered from the original set of quality-controlled reads. Unassembled reads, i.e. reads not matching the recovered putative viruses and plasmids, were separately assembled with the MetaSPAdes pipeline (version 3.14.0) (Nurk et al. 2017) to assemble the microbial community (including viruses and plasmids not assembled with MetaviralSPAdes). SPAdes creates scaffolds from contigs that have known gaps based on overlapping reads (Nurk et al. 2017). Scaffold outputs from the assemblies were pulled together. Scaffolds with lengths below 1000 base pairs were removed. All following analyses were performed on the combined length-filtered scaffold output from the assemblies. Prodigal (version 2.6.3) (Hyatt et al. 2010) in meta mode was used to predict genes and translate these into amino acid sequences. Protein sequences of the genes were annotated using diamond (version 2.0.5) and blastp (version 2.9.0+) (Buchfink et al. 2015) against the UniRef100 database (downloaded on 23.06.2021) (The UniProt Consortium 2019). The dominating taxonomy assignment of UniRef100 protein annotations was used to classify unbinned scaffolds as described for the uBin software (Bornemann et al. 2023).

Nonpareil3 (version 3.303) (Rodriguez-R et al. 2018) was used to estimate covered sequence diversity.

Abundance estimation based on the *rpS3* marker gene in microorganisms

Community composition was reconstructed based on the *ribosomal protein S3* (*rpS3*) marker gene. The advantages of the *rpS3* gene compared to other conventionally used markers, such as the *16S rRNA* gene, were described by Sharon et al. (2015). Marker genes were identified with species-specific hidden Markov models (HMMs) and compared against the UniRef100 database (downloaded on 23.06.2021) (The UniProt Consortium 2019). Coverage of scaffolds with *rpS3* gene sequences was calculated by mapping the quality-controlled reads using bowtie2 (version 2.3.5.1)

(Langmead and Salzberg 2012) against all scaffolds from the meta(viral)SPAdes assemblies in sensitive mode. The coverage was used to approximate relative abundance.

Binning and taxonomic classification of prokaryotes

ABAWACA (version 1.0.0) (Brown et al. 2015), Maxbin2 (version 2.2.7) (Wu et al. 2016), and CONCOCT (version 1.1.0) (Alneberg et al. 2014) were used to bin the samples from the oil reservoir using default parameters. DASTool (version 1.1.6) (Sieber et al. 2018) aggregated the recovered bins to the best set of bins. The selection was manually curated with uBin (version 0.9.14) (Bornemann et al. 2023) to minimize contamination and maximize completeness. Only MAGs with contamination below 5% and completeness above 65% were included, according to checkM2 (version 1.0.1) (Chklovski et al. 2023).

Prokaryotic MAGs were taxonomically annotated and placed in a tree with GTDB-Tk2 (version 2.1.0) (Chaumeil et al. 2022) and GTDB (version 207). DRAM (version 1.4.6) (Shaffer et al. 2020) generated metabolic profiles of said MAGs. Hydrocarbon-degrading genes were screened with the CANT-HYD hmm database (downloaded on 22.06.2023) (Khot et al. 2022).

Gene prediction, gene annotation, and binning of eukaryotes

EukRep (version 0.6.7) (West et al. 2018) with default parameters was used to identify eukaryotic scaffolds in the metagenome. The eukaryotic origin identified scaffolds were binned with CONCOCT (version 1.1.0) (Alneberg et al. 2014), and only bins larger than 2.5 Mbp were kept. Gene prediction and genome annotation were done with the MOSGA2 pipeline (version 2.1.5) (Martin et al. 2021), including BRAKER (version 3.0.3) (Brůna et al. 2021) and completeness calculation with BUSCO (version 5.3.2) (Simão et al. 2015). Predicted genes were annotated against the UniRef100 database (downloaded on 23.06.2021) (The UniProt Consortium 2019). The folding structure of potential oil-degrading genes in the *Pichiacaea* bin was predicted with AlphaFold (version 2.3.2) (Jumper et al. 2021) and annotated the output with FoldSeek (Van Kempen et al. 2023) against the protein folding structures of the UniRef database (The UniProt Consortium 2019) presented in Table S14. The average nucleotide identity between the eukaryotic MAG "bin 8" and reference genome WGS accession AAFM01 from the NCBI database was calculated with fastANI (version 1.34–2) (Jain et al. 2018). The other eukaryotic MAG (bin 19) was taxonomically placed using the UFCG pipeline (version 1.0.5) (Kim et al. 2023). The tree was constructed with IQ-TREE2 (version 2.1.4-beta) (Minh et al. 2020), using ModelFinder (Karthikeyan et al. 2019) and pruned in iTOLs (version 5) (Letunic and Bork 2021).

Detection of viruses and plasmids

Putative viral scaffolds in the metagenomes were identified using three different tools: VIBRANT (version 1.2.1) (Kieft et al. 2019); VirSorter2 (version 2.2.4) (Guo et al. 2021) in sensitive mode; and DeepVirFinder (version 1.0) (Ren et al. 2020) with a threshold of 0.7. Identified scaffolds were further examined and cleaned from host contamination using CheckV (version 1.0.1) (Nayfach et al. 2020). Only scaffolds with predicted completeness ≥25% and no warnings by CheckV (version 1.0.1) were used for further analyses. Taxonomy was predicted with geNomad (version 1.5.0) (Camargo et al. 2023) annotation mode.

SCAPP (version 1.0) (Pellow et al. 2021) was used to reconstruct putative plasmids from the assemblies. The results from SCAPP were filtered not to include scaffolds with genes closely related to those from viral genomes (by matching the genes to the UniRef100 database (downloaded on 23.06.2021) (The UniProt Consortium 2019). Only putative plasmids with a minimum length of 2000 bp were included in the downstream analysis. The plasmid taxonomic units were annotated with COPLA (version 1.0, with PlasmidFinder database from 31.07.2019) (Redondo-Salvo et al. 2021). Auxiliary/metabolic genes were annotated with the UniRef100 database (downloaded on 23.06.2021) (The UniProt Consortium 2019) and CANT-HYD (downloaded on 22.06.2023) (Khot et al. 2022).

Mapping of metagenomic reads to *rpS3* gene sequences, viruses, and plasmids for detection

The collected MGEs and gene sequences from all ten sites were clustered with VIRIDIC (version 1.0) (Moraru et al. 2020) (viruses) on species level (identity threshold 95%) and MeShClust (version 2.0) (James et al. 2018) (plasmids, expanded *rpS3* gene sequences; identity threshold of 95%), respectively. For each cluster, the longest sequence was used as the cluster representative and filtered out. Quality-controlled Illumina reads were mapped to the cluster representative sequences with bowtie2 (version 2.3.5.1) (Langmead and Salzberg 2012) in sensitive mode. The mapping was filtered to only include reads that matched the target sequence with maximum 5% mismatching nucleotide bases. The breadth (herein defined as the percentage of positions of a target sequence covered by at least one read) of the mappings was calculated with genomecov from bedtools (version 2.3.1.1) (Quinlan and Hall 2010). The breadth threshold for the *rpS3* sequence mapping was set to ≥99% and for predicted viruses and plasmids to ≥95%; putative viruses were additionally investigated with a breadth filter of ≥75%. This was done to ensure that the measured coverage values are not the result of many reads only aligning to a small portion of the target sequence.

Results and discussion

Bacteria and fungi potentially mediate hydrocarbon degradation from a pristine oil reservoir

The genetic diversity of microorganisms within the pristine Filicudi oil reservoir sample was analysed using metagenome sequencing of three individual oil samples. Due to the complexity of handling such samples for DNA extraction (Alibrandi et al. 2023), only one sample resulted in a high sequencing depth metagenome covering more than 90% of the biome's diversity, as estimated by Nonpareil3 (Rodriguez-R et al. 2018) (Figure S1). To investigate the diversity of microorganisms in the pristine oil sample, we used the high-sequencing depth metagenome and phylogenetically analysed *rpS3* marker gene sequences (as described by Sharon et al. 2015) from the assembled metagenome, resulting in 26 distinct taxonomic units (Fig. 1). For 14 of these taxonomic units, we recovered high-quality MAGs (15 bacterial MAGs; 2 fungal MAGs), which we subsequently used to determine their capacity for hydrocarbon degradation (information on the contamination and completeness of the MAGs are provided in Tables S3 and S4). Based on relative abundances, the reservoir oil sample predominantly harbored organisms from the *Sphingomonas* genus (over 50% of population share), a group of well-characterized microorganisms previously found in petroleum-contaminated soils and carrying the ability for aromatic hydrocarbon degradation (Zhou et al. 2016). The only *Sphingomonas* MAG recovered in our samples encoded

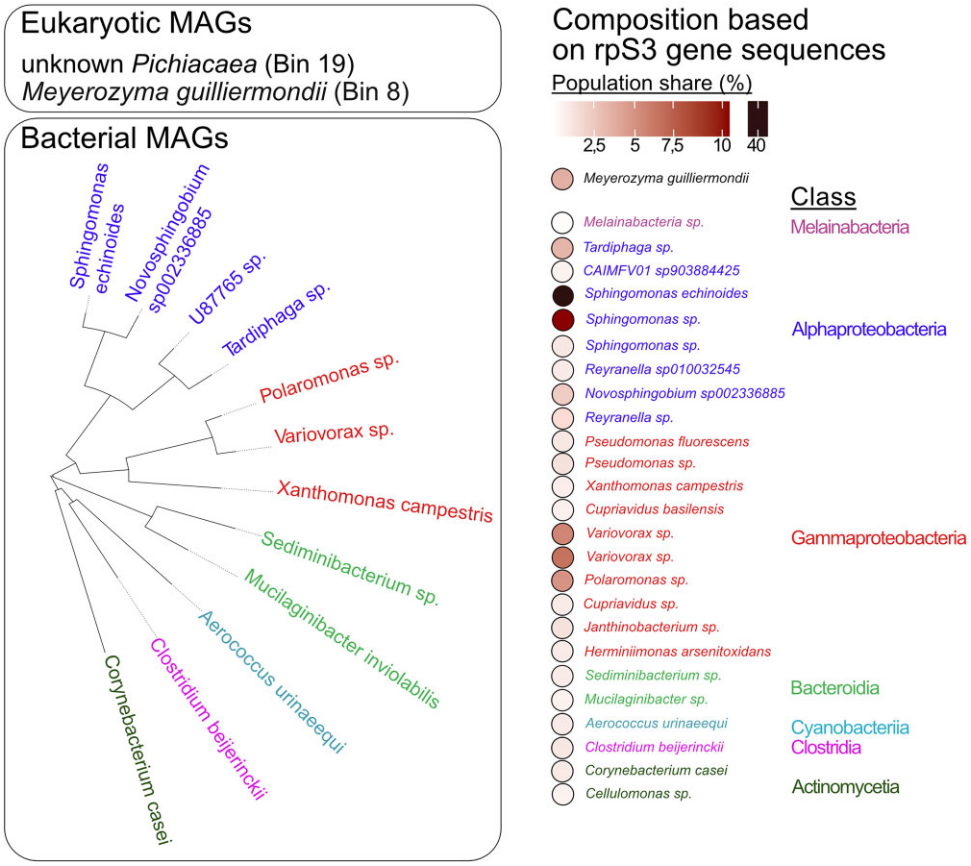

Figure 1. Microbial community composition in the Filicudi oil reservoir from the high-sequencing depth metagenome. For information about taxonomic assignment of eukaryotic MAGs please see the section "Methods" and Figure S2. The phylogenetic tree of bacterial MAGs was constructed with GTDB-Tk2 (Chaumeil et al. 2022). The relative abundance of extended marker gene *rpS3* sequences was used to estimate the community composition. Bacterial taxa are colored by classes.

for the ability to degrade alkanes, toluene, and protocatechuate (Fig. 2) [based on matches against the UniRef100 database (The UniProt Consortium 2019), DRAM (Shaffer et al. 2020), and CANT-Hyd HMMs (Khot et al. 2022)]. Genes for subunits of *benzylsuccinate synthases*, a key enzyme in anaerobic toluene degradation (Acosta-González et al. 2013) were observed in 12 bacterial MAGs from different genera (further database IDs of the mentioned genes can be found in Table S7). Some of these MAGs belonged to the genera *Variovorax* and *Polaromonas*, which were found to be highly abundant in the Filicudi sample (11.3% and 4.7%, respectively).

Additionally, *alkane monooxygenases* (*alkB*) and *alkane hydroxylases* (*CYP153*) were detected in four bacterial MAGs and one fungal MAG (Fig. 2). Regarding the degradation of gaseous hydrocarbons, a MAG classified as *Sediminibacterium* and a scaffold classified as *Mucilaginibacter* (not binned) encode for *pmoA-like methane monooxygenases*, highlighting the organisms' potential to utilize methane as an energy and potential carbon source. Other prominent hydrocarbon-degrading pathways identified in bacterial MAGs include (tetra-) phthalate degradation to pyruvate with (homo-) protocatechuate as intermediate and catechol, xylene, toluene, or benzoate degradation (Fig. 2).

Besides bacteria, we recovered two fungal MAGs from the oil reservoir. One (bin 8) is closely related to *Meyerozyma guilliermondii*, and the other (bin 19) belongs to the family *Pichiacaea*. The *M. guilliermondii* MAG (bin 8) was 10.3 Mbp in size and 88.23% complete based on BUSCO analyses [C:88.23% (S:87.84, D:0.39%), F:4.31%, M:7.45% (Table S3)]. Its genome revealed a mean amino acid identity of 99.7% with another *M. guilliermondii* species from an ex-type culture (NCBI WGS accession AAFM01); additionally, more than 90% of its genes aligned to those from other *M. guilliermondii* in the UniRef100 database. Hydrocarbon-degrading *Meyerozyma* were previously found in oil sludge samples (Liu et al. 2023). The *Pichiacaea* MAG (bin 19) [BUSCO results: C:91.38% (S:90.2%, D:1.18%), F:1.96%, M:6.67% (Table S3)] was found to carry different oil-degrading genes relevant to alkane (*alkB*) and terpene degradation (Fig. 2). Based on its phylogenetic relationship (Figure S2), it is related to the genus *Ogataea* and an unclassified *Candida boidinii* species. To summarize, the recovered pristine community of 17 taxonomic units (15 bacteria and 2 fungi) encoded a highly complex concert of pathways for hydrocarbon degradation of simple and complex hydrocarbons, summarized in Fig. 2. Further metabolic features beyond hydrocarbon degradation are summarized in the Supplementary Results.

Microbial diversity at *Filicudi* is substantially different from other reservoirs and hydrocarbon-affected sites

For comparing the microbial community of our pristine oil sample with those of other hydrocarbon-associated sites (e.g. oil reservoirs, oil-polluted water samples, drilling sites for hydrocarbon exploration, and so on; Tables S1 and S2), we compared the *rpS3* gene sequences (including 1000 bp upstream and downstream of the gene), with those of 35 publicly available metagenomes from 10 sites (Fig. 3) via stringent read mapping (99% nucleotide sequence

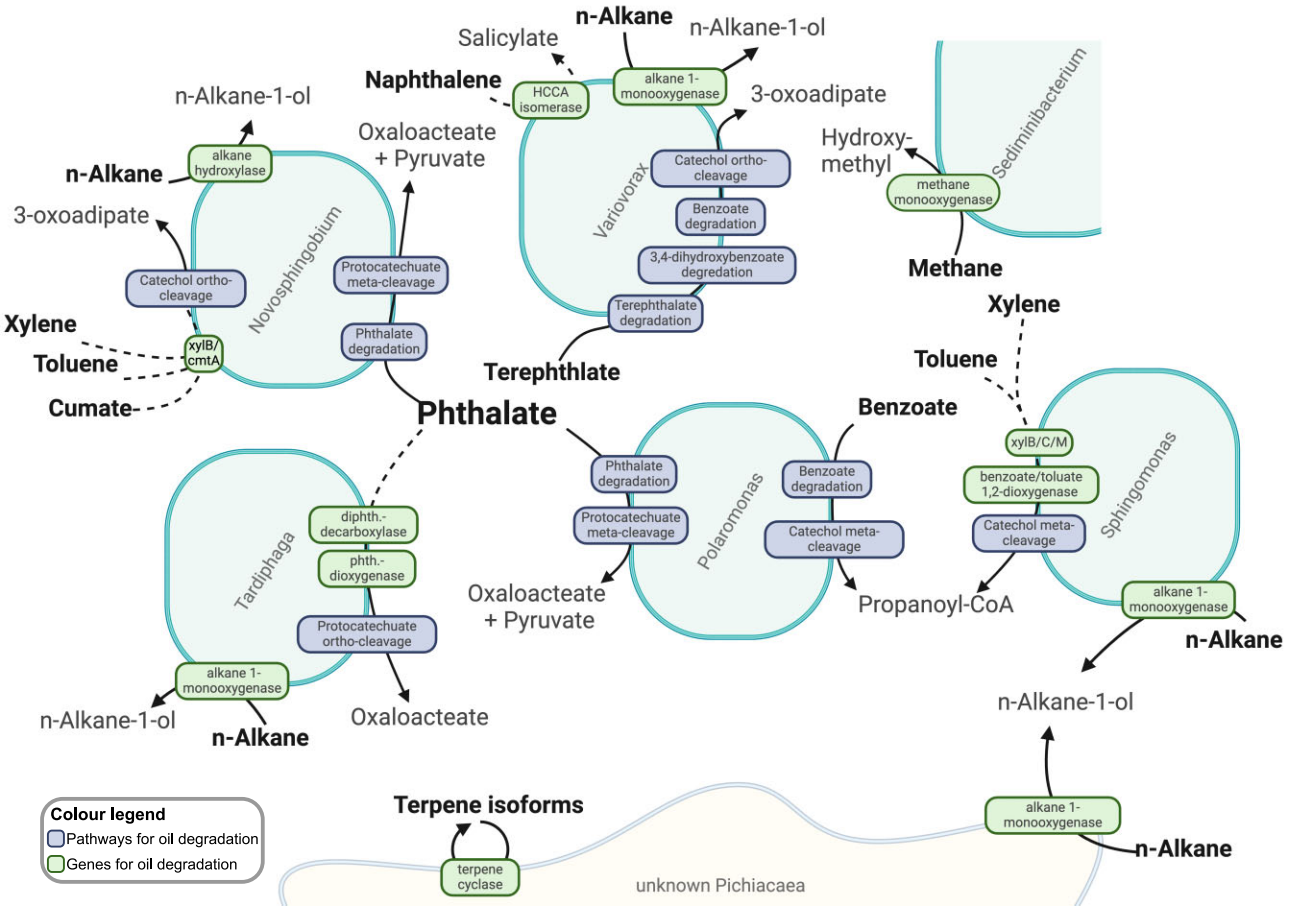

Figure 2. Genetically encoded oil degradation potential of MAGs in the Filicudi oil. Pathways of oil degradation and the predicted encoded genes in a subset of the MAGs from the Filicudi oil reservoir sample. Dashed lines indicate incomplete pathways. The position of pathways and genes does not resemble cellular location. Metabolites and enzymes do not provide information on intra-/extracellular localization thereof. This figure was illustrated with BioRender.com.

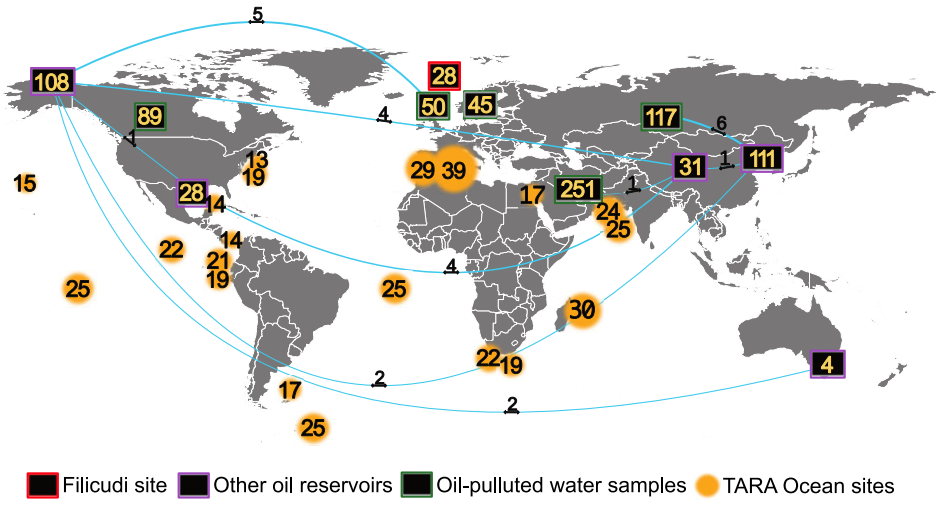

Figure 3. Global distribution of extended *rpS3* gene sequences across hydrocarbon-associated sites. Line thickness along with number on top of the line indicate the number of shared *rpS3* sequences (*rpS3* gene sequence + 1000 bp flanking regions). *RpS3* sequences were filtered for a minimum mapping breadth (br) of 99%. The count of *rpS3* sequences found in each location is represented by boxes. The size of the spheres, along with the number in the sphere, represents the number of *rpS3* sequences from oil-associated sites that were found in the TARA Ocean database.

similarity). In some instances, we identified highly similar *rpS3* gene sequences between the other 10 hydrocarbon-associated sites, indicating that similar species occur in the geologically distinct locations (26 from 1086 *rpS3* gene sequences; 2.39%). The mean coverage of *rpS3* sequences was between 6.9 and 667.8 (median 21.5) based on read mapping when called present in metagenomes of sites from which they were not originally assembled (Table S8). However, we found no shared species between previously released metagenomes and the metagenomes from the Filicudi site, likely rendering the herein-described sample pristine and harboring an indigenous population (see Fig. 1). To track the origin of the potential microbes identified in the 10 publicly available hydrocarbon-impacted sites, which were not pristine, we stringently mapped the reads from the Tara Ocean dataset (Pesant et al. 2015) to the recovered *rpS3* gene sequences, resulting in multiple shared species across the globe (Fig. 3). For instance, oil-contaminated marine samples contained 215 *rpS3* gene sequences that were also covered by reads from the Tara Ocean datasets (mean depth of coverage between 1.9 and 3916.5; median 20). By contrast, we found only two *rpS3* gene sequences from water-polluted oil reservoirs to be also present in ocean microbiomes, none of which belonged to Filicudi, (mean depth of coverage between 4.5 and 55.1; median 13.15; for further depth of coverage values please see Table S9). Based on 19.98% (217 from 1086) of *rpS3* gene sequences from hydrocarbon-affected sites being covered by metagenomic reads from the Tara Ocean dataset, we posit that these marine microorganisms are adapted to global hydrocarbon exploration sites and oil contamination; they also seem to be introduced to oil reservoirs that are not contamination-controlled if seawater is used as a drilling fluid.

Viruses and plasmids demonstrate contrasting distribution patterns across 35 metagenomes from hydrocarbon sites, suggesting a global dispersal of MGEs

To investigate information encoded on MGEs, we analysed putative plasmids and viruses reconstructed from the Filicudi site, discovering 39 putative plasmids and 58 putative viruses. Of these, 98% and 90%, respectively, were novel compared to previously described MGEs in public databases [based on annotation with COPLA (Redondo-Salvo et al. 2021) and geNomad (Camargo et al. 2023)]. This highlights the heavily underexplored genomic repertoire residing in hydrocarbon-rich environments.

Regarding the putative plasmids, their length varied between 2 and 230 kb, and all 39 of them carried different metabolic genes, including an *alkanesulfonate monooxygenase* that has a protective function during hexadecane sulfonate production (Park et al. 2020). A substantial fraction of the predicted plasmids (16 out of 39, 41%) also carried genes relevant to metal tolerance and resistance, specifically mercury, copper, arsenate, and tellurite (Fig. 4A), which are generally known to be typical contaminants in oil reservoirs. We conclude that predicted plasmids can encode crucial metabolic genes of microorganisms in hydrocarbon-associated sites and might be suitable targets for modifying hydrocarbon-degrading microorganisms in biotechnology. Concerning viral scaffolds, most of them were classified as dsDNA viruses from the *Caudoviricetes* class (48), while a small subset were ssDNA viruses from *Malgrandaviricetes* (5), *Arfiviricetes* (4), and one belonging to *Faserviricetes* (Fig. 4B). Given the selective sequencing of only double-stranded DNA viruses using Illumina technology, we can conclude that the ssDNA viruses detected in the metagenome were actively replicating and thus appeared as double-stranded prior to DNA extraction, indicating a potential activity of ssDNA viruses in oil reservoirs.

In contrast to the isolated population of microorganisms in the pristine Filicudi sample, mapping of the abovementioned 35 public metagenomes from hydrocarbon-associated sites to the recovered putative plasmidome and virome from the Filicudi sample revealed a shared mobilome between hydrocarbon-impacted sites. To further complement our study, we predicted the previously undescribed mobilome of the 10 hydrocarbon-associated sites, resulting in 1738 putative viruses and 386 putative plasmids after clustering using VIRIDIC (Moraru et al. 2020) at 95% identity (Table S6). Since usually only a minor fraction of viruses get assembled in metagenomes, we then used the clustered set of putative viruses and plasmids as databases for read mapping (breadth ≥75%; for details, see the section "Methods"), resulting in 63 shared putative viruses, 18% of which were also found as shared in the initial VIRIDIC clustering. Based on these mapping results, we identified three similar putative viral sequences in a highly water-contaminated oil reservoir located in the Shengli region (China) (Wang et al. 2019) and in the Filicudi sample. We also found evidence of 16 highly similar viruses between oil-associated sites, the ocean, and geologically distinct oil reservoirs [Fig. 5A; breadth ≥95%; mean coverage depth between 1.5 and 6627.8, median = 14.8 (Tables S12 and S13)]. While viruses have been suggested to overcome large distances by traveling with rain or water sources (Rahlff et al. 2022), this process cannot explain the infiltration of pristine oil reservoirs like Filicudi. Nevertheless, the deep sea and the ocean, in general, harbor a global virome (Gregory et al. 2019, He et al. 2023) with a certain degree of exchange and connectivity through ocean currents (Jian et al. 2021), whose imprints also appear visible in the investigated hydrocarbon-associated sites and oil reservoirs.

The distribution of putative plasmids followed a substantially different pattern from those of potential viruses and *rpS3* gene sequences. Predicted plasmids from oil-associated sites were rarely shared with Tara Ocean data [21 from 386 (5.6%); depth of coverage between 2.6 and 210.6; median 12.4] compared to *rpS3* gene sequences (19.98%). By contrast, a larger subset was highly similar between globally distributed oil reservoirs and hydrocarbon-associated sites [26 from 386 (6.73%); mean depth of coverage between 1.9 and 2511.8; median 20.5 (Table S10), compared to host organisms (2.39%) (Fig. 5B)]. While host specificity and compatibility have previously been shown to selectively influence plasmid conjugation (Redondo-Salvo et al. 2020, Benz and Hall 2023), it is not the driving force for plasmid transfer in hydrocarbon-impacted sites, as shown by the deviating patterns of shared species (Fig. 3) and plasmids (Fig. 5B). This directly contradicts the usual dispersal pathway of plasmids with their host organisms, which should, in principle, result in a similar distribution pattern, while viruses may disperse independently. However, it should be noted that the Tara Ocean dataset only includes samples from the surface ocean, and a global dataset on deep ocean metagenomes is not available in the literature, which might be more representative for comparing genetic signatures from oil reservoirs with those from the ocean.

The metagenomic analysis of the pristine sample from the Filicudi reservoir revealed many bacteria and fungi involved in hydrocarbon degradation. Compared to other oil reservoirs or hydrocarbon-associated sites, we see three global patterns for prokaryotes, putative viruses, and putative plasmids. Shared prokaryotes and viral scaffolds with the Tara Ocean dataset and the fact that the volume of the entire ocean is estimated to be cir-

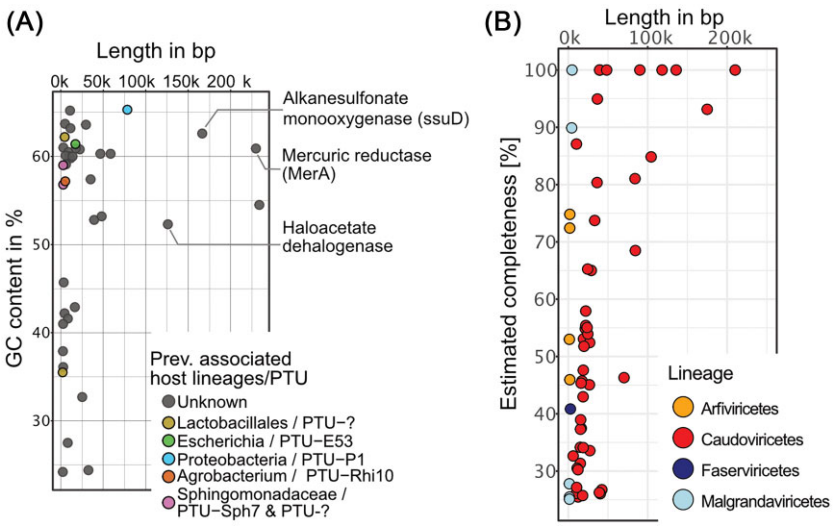

Figure 4. Plasmids and viruses in the pristine Filicudi reservoir. (A) Length in base pairs (bp) and GC content of predicted plasmids in the Filicudi oil reservoir across all three metagenomes. Colored circles represent plasmid taxonomic units (PTUs) and previously associated host lineages of said PTUs. Auxiliary metabolic genes related to important reservoir processes are labeled to the according putative plasmids. CheckV (Nayfach et al. 2020) completeness and length (bp) of putative viruses identified in the metagenomics dataset of the Filicudi oil reservoir (for the complete checkV output see Table S5). Lineages were predicted with geNomad (Camargo et al. 2023). Dot plots were created in R (R Core Team 2020).

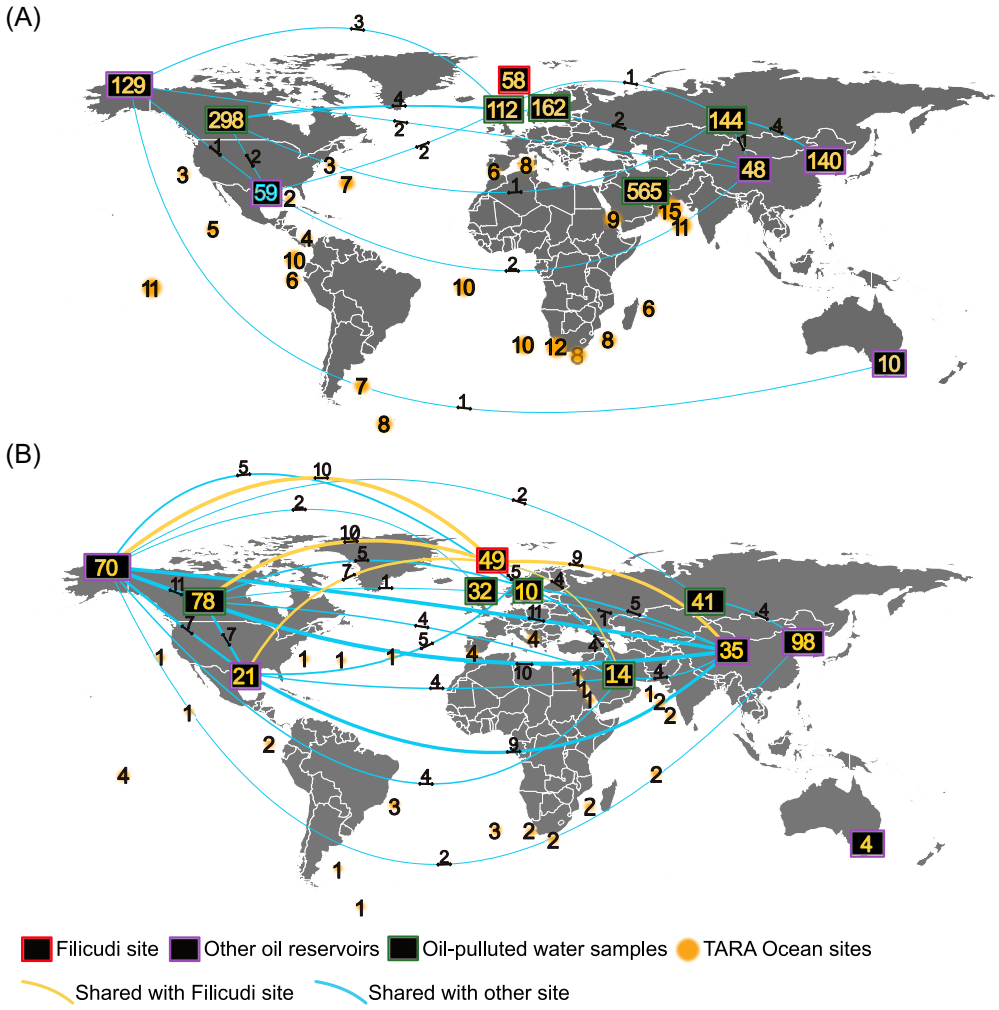

Figure 5. Global distribution of putative plasmids, and viruses from hydrocarbon-associated sites. Line thickness along with the number on top of the line indicate the number of shared viruses (A), and plasmids (B). The threshold for minimum mapping breadth (br) was 95%. The count of elements found in each location is represented by boxes. The size of the spheres represents the detection of elements from any oil-associated site in the TARA ocean database (indicated by the number within the spheres).

culated through the upper ocean crust within only 200 ka explain at least their global distribution between the studied sites (Johnson and Pruis 2003). However, the distribution pattern of plasmids (usually associated with hosts) remains enigmatic at this stage. Future research and many additional metagenomes of oil reservoirs are needed to fully grasp the picture of the global plasmidome in oil reservoirs and to confidently predict the dispersal routes of MGEs and prokaryotes.

Acknowledgment

The authors thank Ken Dreger for exemplary server administration.

Supplementary data

Supplementary data is available at *FEMSML Journal* online.

Conflict of interest: The authors declare no conflict of interest.

Funding

This project was funded by Aker BP (Norway) as part of the GeneOil Project. We appreciate support by the German Research Foundation (DFG) CRC 1439/1; project number 426547801. Tom L. Stach was supported by the German Academic Scholarship Foundation. We acknowledge support by the Open Access Publication Fund of the University of Duisburg-Essen

Data availability

All metagenome datasets and metagenome assembled genomes (MAGs) from the Filicudi oil reservoir are published under the BioProject number PRJNA1147926. The metagenomes have the individual BioSample accessions SAMN43174722, SAMN43174723, and SAMN43174724, respectively. The accession numbers of MAGs are listed in Tables S3 and S4.

References

Acosta-González A, Rosselló-Móra R, Marqués S. Diversity of benzylsuccinate synthase-like (*bssA*) genes in hydrocarbon-polluted marine sediments suggests substrate-dependent clustering. *Appl Environ Microbiol* 2013;**79**:3667–76. https://doi.org/10.1128/AEM.03934-12.

Alibrandi A, Di Primio R, Bartholomäus A *et al.* A modified isooctane-based DNA extraction method from crude oil. *mLife* 2023;**2**:328–38. https://doi.org/10.1002/mlf2.12081.

Alneberg J, Bjarnason BS, Bruijn Id *et al.* Binning metagenomic contigs by coverage and composition. *Nat Methods* 2014;**11**:1144–46. https://doi.org/10.1038/nmeth.3103.

Antipov D, Raiko M, Lapidus A *et al.* Metaviral SPAdes: assembly of viruses from metagenomic data. *Bioinformatics* 2020;**36**:4126–29. https://doi.org/10.1093/bioinformatics/btaa490.

Asemoloye MD, Tosi S, Daccò C *et al.* Hydrocarbon degradation and enzyme activities of *Aspergillus oryzae* and *Mucor irregularis* isolated from Nigerian crude oil-polluted sites. *Microorganisms* 2020;**8**:1912. https://doi.org/10.3390/microorganisms8121912.

Atakpa EO, Zhou H, Jiang L *et al.* Improved degradation of petroleum hydrocarbons by co-culture of fungi and biosurfactant-producing bacteria. *Chemosphere* 2022;**290**:133337. https://doi.org/10.1016/j.chemosphere.2021.133337.

Benedek T, Szentgyörgyi F, Gergócs V *et al.* Potential of *Variovorax paradoxus* isolate BFB1_13 for bioremediation of BTEX contaminated sites. *AMB Expr* 2021;**11**:126. https://doi.org/10.1186/s13568-021-01289-3.

Benz F, Hall AR. Host-specific plasmid evolution explains the variable spread of clinical antibiotic-resistance plasmids. *Proc Natl Acad Sci USA* 2023;**120**:e2212147120. https://doi.org/10.1073/pnas.2212147120.

Bian Z, Chen Y, Zhi Z *et al.* Comparison of microbial community structures between oil and water phases in a low-permeability reservoir after water flooding. *Energy Rep Vol* 2023;**9**:1054–61. https://doi.org/10.1016/j.egyr.2022.12.026.

Bornemann TLV, Esser SP, Stach TL. *et al.* uBin—a manual refining tool for metagenomic bins designed for educational purposes. *Environ Microbiol* 2023;**25**:1077–83. https://doi.org/10.1111/1462-2920.16351.

Brown CT, Hug LA, Thomas BC. *et al.* Unusual biology across a group comprising more than 15% of domain bacteria. *Nature* 2015;**523**:208–11.

Brůna T, Hoff KJ, Lomsadze A *et al.* BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics Bioinformatics* 2021;**3**:lqaa108. https://doi.org/10.1093/nargab/lqaa108.

Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015;**12**:59–60. https://doi.org/10.1038/nmeth.3176.

Bushnell B. BBMap: a fast, accurate, splice-aware aligner. In: *Proceedings of the 9th Annual Genomics of Energy & Environment Meeting*. Berkeley, CA: DOE Joint Genome Institute, 2014. https://sourceforge.net/projects/bbmap/ (08 July 2021, date last accessed)

Cai M, Nie Y, Chi C-Q *et al.* Crude oil as a microbial seed bank with unexpected functional potentials. *Sci Rep* 2015;**5**:16057. https://doi.org/10.1038/srep16057.

Camargo AP, Roux S, Schulz F *et al.* Identification of mobile genetic elements with geNomad. *Nat Biotechnol* 2023;**42**:1303–12. https://doi.org/10.1038/s41587-023-01953-y.

Chaumeil P-A, Mussig AJ, Hugenholtz P *et al.* GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 2022;**38**:5315–16. https://doi.org/10.1093/bioinformatics/btac672.

Chklovski A, Parks DH, Woodcroft BJ. *et al.* CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods* 2023;**20**:1203–12. https://doi.org/10.1038/s41592-023-01940-w.

Gao P, Tian H, Wang Y *et al.* Spatial isolation and environmental factors drive distinct bacterial and archaeal communities in different types of petroleum reservoirs in China. *Sci Rep* 2016;**6**:20174. https://doi.org/10.1038/srep20174.

Gregory AC, Zayed AA, Conceição-Neto N *et al.* Marine DNA viral macro- and microdiversity from pole to pole. *Cell* 2019;**177**:1109–23.e14.https://doi.org/10.1016/j.cell.2019.03.040.

Guo J, Bolduc B, Zayed AA. *et al.* VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* 2021;**9**:37. https://doi.org/10.1186/s40168-020-00990-y.

He T, Jin M, Cui P *et al.* Environmental viromes reveal the global distribution signatures of deep-sea DNA viruses. *J Adv Res* 2023;**57**:107–17. https://doi.org/10.1016/j.jare.2023.04.009.

Heider J, Szaleniec M, Sünwoldt K *et al.* Ethylbenzene dehydrogenase and related molybdenum enzymes involved in oxygen-independent alkyl chain hydroxylation. *J Mol Microbiol Biotechnol* 2016;**26**:45–62. https://doi.org/10.1159/000441357.

Hu P, Tom L, Singh A *et al.* Genome-resolved metagenomic analysis reveals roles for candidate phyla and other microbial community

members in biogeochemical transformations in oil reservoirs. *mBio* 2016;**7**:e01669–15. https://doi.org/10.1128/mBio.01669-15.

Hyatt D, Chen G-L, LoCascio PF *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf* 2010;**11**:119. https://doi.org/10.1186/1471-2105-11-119.

Jain C, Rodriguez-R LM, Phillippy AM. *et al.* High throughput ANI analysis of 90 K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;**9**:5114. https://doi.org/10.1038/s41467-018-07641-9.

James BT, Luczak BB, Girgis HZ. MeShClust: an intelligent tool for clustering DNA sequences. *Nucleic Acids Res* 2018;**46**:e83. https://doi.org/10.1093/nar/gky315.

Jian H, Yi Y, Wang J *et al.* Diversity and distribution of viruses inhabiting the deepest ocean on Earth. *ISME J* 2021;**15**:3094–110. https://doi.org/10.1038/s41396-021-00994-y.

Johnson HP, Pruis MJ. Fluxes of fluid and heat from the oceanic crustal reservoir. *Earth Planet Sci Lett* 2003;**216**:565–74. https://doi.org/10.1016/S0012-821X(03)00545-4.

Joshi NA, Fass J. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. GitHub, 2011.

Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–89. https://doi.org/10.1038/s41586-021-03819-2.

Kalyaanamoorthy S, Minh BQ, Wong TKF *et al.* ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;**14**:587–89. https://doi.org/10.1038/nmeth.4285.

Karthikeyan S, Rodriguez-R LM, Heritier-Robbins P *et al.* Genome repository of oiled systems (GROS): an interactive and searchable database that expands the catalogued diversity of crude oil-associated microbes. *Environ Microbiol* 2019;**22**:2094–2106. https://doi.org/10.1101/838573.

Khot V, Zorz J, Gittins DA. *et al.* CANT-HYD: a curated database of phylogeny-derived hidden Markov models for annotation of marker genes involved in hydrocarbon degradation. *Front Microbiol* 2022;**12**:764058. https://doi.org/10.3389/fmicb.2021.764058.

Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of virome function from genomic sequences. *Microbiome* 2019;**8**:90. https://doi.org/10.1101/855387.

Kim D, Gilchrist CLM, Chun J *et al.* UFCG: database of universal fungal core genes and pipeline for genome-wide phylogenetic analysis of fungi. *Nucleic Acids Res* 2023;**51**:D777–84. https://doi.org/10.1093/nar/gkac894.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;**9**:357–59. https://doi.org/10.1038/nmeth.1923.

Letunic I, Bork P. Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 2021;**49**:W293–96. https://doi.org/10.1093/nar/gkab301.

Li D, Greenfield P, Rosewarne CP. *et al.* Draft genome sequence of *Thermoanaerobacter* sp. strain A7A, reconstructed from a metagenome obtained from a high-temperature hydrocarbon reservoir in the Bass Strait, Australia. *Genome Announc* 2013;**1**. https://doi.org/10.1128/genomeA.00701-13.

Lin J, Hao B, Cao G *et al.* A study on the microbial community structure in oil reservoirs developed by water flooding. *J Pet Sci Eng* 2014;**122**:354–59. https://doi.org/10.1016/j.petrol.2014.07.030.

Liu H, Wu M, Gao H *et al.* Crude oil removal by *Meyerozyma* consortium and nitrogen supplement: hydrocarbon transformation, nitrogen fate, and enhancement mechanism. *J Environ Chem Eng* 2023;**11**:109034. https://doi.org/10.1016/j.jece.2022.109034.

Martin R, Dreßler H, Hattab G *et al.* MOSGA 2: comparative genomics and validation tools. *Comput Struct Biotechnol J* 2021;**19**:5504–9. https://doi.org/10.1016/j.csbj.2021.09.024.

Minh BQ, Schmidt HA, Chernomor O *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;**37**:1530–34. https://doi.org/10.1093/molbev/msaa015.

Moraru C, Varsani A, Kropinski AM. VIRIDIC—a novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. *Viruses* 2020;**12**:1268. https://doi.org/10.3390/v12111268.

Nayfach S, Pedro Camargo A, Eloe-Fadrosh E *et al.* CheckV: assessing the quality of metagenome-assembled viral genomes. *Nat Biotechnol* 2020;**39**:578–85. https://doi.org/10.1101/2020.05.06.081778.

Nesbø C, Swithers KS, Dahle H *et al.*, Evidence for extensive gene flow and *Thermotoga* subpopulations in subsurface and marine environments. *ISME J* 2015;**9**:1532–42. https://doi.org/10.1038/ismej.2014.238.

Nurk S, Meleshko D, Korobeynikov A *et al.* metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 2017;**27**:824–34. https://doi.org/10.1101/gr.213959.116.

Pannekens M, Kroll L, Müller H *et al.* Oil reservoirs, an exceptional habitat for microorganisms. *New Biotechnol* 2019;**49**:1–9. https://doi.org/10.1016/j.nbt.2018.11.006.

Park C, Shin B, Park W. Protective role of bacterial alkanesulfonate monooxygenase under oxidative stress. *Appl Environ Microbiol* 2020;**86**:e00692–20. https://doi.org/10.1128/AEM.00692-20.

Pellow D, Zorea A, Probst M *et al.* SCAPP: an algorithm for improved plasmid assembly in metagenomes. *Microbiome* 2021;**9**:144. https://doi.org/10.1186/s40168-021-01068-z.

Pesant S, Not F, Picheral M *et al.* Open science resources for the discovery and analysis of Tara Oceans data. *Sci Data* 2015;**2**:150023. https://doi.org/10.1038/sdata.2015.23.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;**26**:841–42. https://doi.org/10.1093/bioinformatics/btq033.

R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. 2020. https://www.R-project.org/ (22 May 2022, date last accessed)

Rahlff J, Esser SP, Plewka J *et al.* Heads in the clouds: marine viruses disperse bidirectionally along the natural water cycle. *Nat Commun* 2022;**14**:6354. https://doi.org/10.1101/2022.06.21.497027.

Redondo-Salvo S, Bartomeus-Peñalver R, Vielva L *et al.* COPLA, a taxonomic classifier of plasmids. *BMC Bioinf* 2021;**22**:390. https://doi.org/10.1186/s12859-021-04299-x.

Redondo-Salvo S, Fernández-López R, Ruiz R *et al.* Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. *Nat Commun* 2020;**11**:3602. https://doi.org/10.1038/s41467-020-17278-2.

Ren J, Song K, Deng C *et al.* Identifying viruses from metagenomic data using deep learning. *Quant Biol* 2020;**8**:64–77. https://doi.org/10.1007/s40484-019-0187-4.

Reyes-Avila J, Roldán-Carrillo T, Castorena-Cortés G. *et al.* Effect of sulphur species on the hydrocarbon biodegradation of oil sludge generated by a gas processing facility. *Int J Environ Sci Technol* 2013;**10**:551–58. https://doi.org/10.1007/s13762-013-0183-6.

Rodriguez-R LM, Gunturu S, Tiedje JM. *et al.* Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity. *mSystems* 2018;**3**:e00039–18. https://doi.org/10.1128/mSystems.00039-18.

Sanjuán R, Nebot MR, Chirico N *et al.* Viral mutation rates. *J Virol* 2010;**84**:9733–48. https://doi.org/10.1128/JVI.00694-10.

Seitz KW, Dombrowski N, Eme L *et al.* Asgard archaea capable of anaerobic hydrocarbon cycling. *Nat Commun* 2019;**10**:1822. https://doi.org/10.1038/s41467-019-09364-x.

Semenova EM, Ershov AP, Sh. Sokolova D. *et al.* Diversity and biotechnological potential of nitrate-reducing bacteria from heavy-oil reservoirs (Russia). *Microbiology* 2020;**89**:685–96. https://doi.org/10.1134/S0026261720060168.

Shaffer M, Borton MA, McGivern BB. *et al.* DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res* 2020;**48**:8883–900. https://doi.org/10.1101/2020.06.29.177501.

Sharon I, Kertesz M, Hug LA. *et al.* Accurate, multi-kb reads resolve complex populations and detect rare microorganisms. *Genome Res* 2015;**25**:534–43. https://doi.org/10.1101/gr.183012.114.

Sieber CMK, Probst AJ, Sharrar A *et al.* Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* 2018;**3**:836–43. https://doi.org/10.1038/s41564-018-0171-1.

Simão FA, Waterhouse RM, Ioannidis P *et al.* BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015;**31**:3210–12. https://doi.org/10.1093/bioinformatics/btv351.

Sobecky PA, Hazen TH. Horizontal gene transfer and mobile genetic elements in marine systems. In: Boekels Gogarten M, Peter Gogarten J, Olendzenski LC (eds), *Methods in Molecular Biology*. Vol. **532**. Totowa, NJ: Humana Press, 2009, 435–53. https://doi.org/10.1007/978-1-60327-853-9_25.

The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019;**47**:D506–15. https://doi.org/10.1093/nar/gky1049.

Van Kempen M, Kim SS, Tumescheit C *et al.* Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* 2023;**42**:243–6. https://doi.org/10.1038/s41587-023-01773-0.

Vigneron A, Alsop EB, Lomans BP *et al.* Succession in the petroleum reservoir microbiome through an oil field production lifecycle. *ISME J* 2017;**11**:2141–54. https://doi.org/10.1038/ismej.2017.78.

Wang X, Li X, Yu L *et al.* Characterizing the microbiome in petroleum reservoir flooded by different water sources. *Sci Total Environ* 2019;**653**:872–85. https://doi.org/10.1016/j.scitotenv.2018.10.410.

West PT, Probst AJ, Grigoriev IV. *et al.* Genome-reconstruction for eukaryotes from complex natural microbial communities. *Genome Res* 2018;**28**:569–80. https://doi.org/10.1101/gr.228429.117.

Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 2016;**32**:605–7. https://doi.org/10.1093/bioinformatics/btv638.

Xu X, Liu W, Tian S *et al.* Petroleum hydrocarbon-degrading bacteria for the remediation of oil pollution under aerobic conditions: a perspective analysis. *Front Microbiol* 2018;**9**:2885. https://doi.org/10.3389/fmicb.2018.02885.

Yang R, Zhang G, Li S *et al.* Degradation of crude oil by mixed cultures of bacteria isolated from the Qinghai–Tibet plateau and comparative analysis of metabolic mechanisms. *Environ Sci Pollut Res* 2019;**26**:1834–47. https://doi.org/10.1007/s11356-018-3718-z.

Youssef N, Elshahed MS, McInerney MJ. Chapter 6 microbial processes in oil fields. *Adv Appl Microbiol* 2009;**66**:141–251. https://doi.org/10.1016/S0065-2164(08)00806-X.

Zhou L, Li H, Zhang Y *et al.* Sphingomonas from petroleum-contaminated soils in Shenfu, China and their PAHs degradation abilities. *Braz J Microbiol* 2016;**47**:271–78. https://doi.org/10.1016/j.bjm.2016.01.001.

Received 21 January 2024; **revised** 23 October 2024; **accepted** 22 January 25