

Diversity and metagenome analysis of a hydrocarbon-degrading bacterial consortium from asphalt lakes located in Wietze, Germany

Michael O. Eze^{1,2*}, Grant C. Hose³, Simon C. George² and Rolf Daniel¹

¹Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Georg-August University of Göttingen, 37077 Göttingen, Germany.

²Department of Earth and Environmental Sciences, Macquarie University, Sydney, NSW 2109, Australia.

³Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia.

*Correspondence: meze@gwdg.de

Abstract

The pollution of terrestrial and aquatic environments by petroleum contaminants, especially diesel fuel, is a persistent environmental threat requiring cost-effective and environmentally sensitive remediation approaches. Bioremediation is one such approach, but is dependent on the availability of microorganisms with the necessary metabolic abilities and environmental adaptability. The aim of this study was to examine the microbial community in a petroleum contaminated site, and isolate organisms potentially able to degrade hydrocarbons. Through successive enrichment of soil microorganisms from samples of an historic petroleum contaminated site in Wietze, Germany, we isolated a bacterial consortium using diesel fuel hydrocarbons as sole carbon and energy source. The 16S rRNA gene analysis revealed the dominance of *Alphaproteobacteria*. We further reconstructed a total of 18 genomes from both the original soil sample and the isolated consortium. The analysis of both the metagenome of the consortium and the reconstructed metagenome-assembled genomes show that the most abundant bacterial genus in the consortium, *Acidocella*, possess many of the genes required for the degradation of diesel fuel aromatic hydrocarbons, which are often the most toxic component. This can explain why this genus proliferated in all the enrichment cultures. Therefore, this study reveals that the microbial consortium isolated in this study and its dominant genus, *Acidocella*, could potentially serve as an effective inoculum for the bioremediation of sites polluted with diesel fuel or other organic contaminants.

Keywords: Bacterial diversity, diesel fuel, consortium, metagenome, bioremediation.

INTRODUCTION

Petroleum pollution is a recurring environmental threat resulting from oil and gas exploration, production, transport and storage (Eze and George 2020). Spills have occurred in terrestrial as well as aquatic environments, and they are often caused by human error, corrosion and equipment failure (Dalton and Jin 2010; Errington et al. 2018; Hassler 2016; Hong et al. 2010). This is a major threat to both the environment and human health, due to the phytotoxicity and carcinogenicity of petroleum hydrocarbons.

In view of the diversity of pollutants, a range of *ex situ* and *in situ* bioremediation techniques have been developed (Azubuike et al. 2016). *Ex situ* techniques involve the excavation and off-site treatment of contaminated soils or water, while *in situ* strategies involve on-site treatment of contaminants. As a result, *ex situ* techniques are often more expensive than *in situ* techniques owing to the additional costs associated with contaminant excavation and relocation (USEPA 2000). The United States Environmental Protection Agency indicated that implementing *in situ* degradation will result in cost savings of 50 to 80% over traditional methods such as excavation and landfill incineration (USEPA 2001). Moreover, *ex situ* methods are environmentally problematic as they alter the soil matrix and associated microbiomes.

The success of any oil spill remediation approach depends on environmental conditions such as temperature, pH and nutritional constraints in contaminated sites (Joner et al. 2002; Kleinstuber et al. 2006; Leahy and Colwell 1990; Rohrbacher and St-Arnaud 2016). The hydrophobic nature of petroleum hydrocarbons limits their availability to biodegradation. Hence, the presence of microorganisms with the metabolic capability to degrade petroleum and the ability to adapt to a range of environmental conditions is a crucial factor (Das and Chandran 2011). Organisms capable of degrading diesel fuel and other organic contaminants are diverse and present in many natural habitats, including extreme ones (Gemmell and Knowles 2000; Hara and Uchiyama 2013; Lohi et al. 2008; Nie et al. 2014; Stapleton et al. 1998). Microorganisms from polluted environments hold the key to unlocking most of the challenges associated with bioremediation (Eze et al. 2020; Liang et al. 2019; Liang et al. 2016). One such environment is the heavily polluted oil field in Wietze, Germany.

Wietze is an important historical site of crude-oil production. In Germany, pre-industrial oil production started in the 17th century, followed by industrial oil extraction beginning in 1859 (Craig et al. 2018). Between 1900 and 1920, Wietze was the most productive oil field in Germany, with almost 80% of German oil produced there. Oil production in Wietze was discontinued in 1963, but the former oil field continues to witness considerable amounts of oil seepage, with several heavily polluted sites, contaminated ponds, and organic debris from surrounding plants (Figure 1). Therefore, it is an ideal site for obtaining microorganisms with the potential for bioremediation of petroleum hydrocarbons. Samples investigated in this study were taken from three sites around a small asphalt pond (Figure 1).

Due to the so-called uncultivability of many environmental microorganisms (Steen et al. 2019), several studies have concentrated on remediation by indigenous microorganisms (Kumar and Gopal 2015; Sarkar et al. 2016). More recent studies have shown that the inoculation of carefully cultivated hydrocarbon-degrading bacterial consortia or isolates enhances the effectiveness of various remediation techniques (Atashgahi et al. 2018; Garrido-Sanz et al. 2019). Therefore, it is important to discover novel microbes capable of degrading petroleum hydrocarbons either as single isolates, consortia, or synergistically with plants. The aim of this study was to investigate the diversity and genomic potential of bacterial consortia derived from a hydrocarbon contaminated asphalt lake in Wietze, Germany. We also aimed to reconstruct metagenome-assembled genomes, and to examine the potential of the reconstructed genomes for bioremediation of diesel fuel contaminated sites.



Figure 1. Sampling site in Wietze, Germany (52°39'0''N, 09°50'0''E).

MATERIALS AND METHODS

Soil sampling

Topsoil samples (10 g each) and water samples (approximately 50 mL each) were taken in November 2019 from three heavily polluted sites located at the historical oil field in Wietze (52°39'0''N, 09°50'0''E), Germany. In addition, two reference samples were taken from nearby unpolluted soils. Samples were placed into 50 mL Eppendorf conical tubes. The samples were transported to the laboratory on ice.

Enrichment cultures and growth conditions

Approximately 1 g of each of the crude oil-polluted soil samples was added to Erlenmeyer flasks (300 mL) containing 100 mL of a liquid mineral medium (MM) composed of KH_2PO_4 (0.5 g/L), NaCl (0.5 g/L), and NH_4Cl (0.5 g/L). Sterile-filtered trace elements (1 mL/L) (Atlas 2010), vitamin solution (1 mL/L) (Atlas 2010) and $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ (5 mL/L of a 100 mg/mL solution) were added

to the MM, post MM-autoclaving. One mL of sterile-filtered diesel fuel (C₁₀-C₂₅) was added to each flask as the sole carbon and energy source. The cultures were grown at 30°C with shaking at 110 rpm (INFORS HT shaker, model CH-4103, Infors AG, Bottmingen, Switzerland) and subcultured every five days. After three successive subculturing steps, 30 mL aliquots (OD₆₀₀, 0.635) were centrifuged for 10 min at 4,000 × *g*.

DNA Extraction

Microbial cells from approximately 30 mL of the enrichment cultures and water samples were harvested by centrifugation at 4,000 × *g* for 10 min. The supernatant was subsequently discarded. DNA from the cell pellets and 100 mg of each of the original samples were extracted using the PowerSoil® DNA Extraction kit as recommended by the manufacturer (Qiagen, Hilden, Germany). DNA from one of the original soil samples and one of the three final enrichments (S3S and S3E3 respectively, Supplementary Figure S1) were used for metagenome studies.

Sequencing of bacterial 16S rRNA genes

Bacterial 16S rRNA genes were amplified using the forward primer S-D-Bact-0341-b-S-17 (5'-CCT ACG GGN GGC WGC AG-3') and the reverse primer S-D-Bact-0785-a-A-21 (5'-GAC TAC HVG GGT ATC TAA TCC-3') (Klindworth et al. 2013) containing adapters for Illumina MiSeq sequencing. The PCR reaction (25 µL final volume) contained 5 µL of five-fold Phusion HF buffer, 200 µM of each of the four deoxynucleoside triphosphates, 4 µM of each primer, 1 U of Phusion high fidelity DNA polymerase (Thermo Scientific, Waltham, MA, USA), and approximately 50 ng of the extracted DNA as the template. Negative controls were performed using the reaction mixture without a template. The following thermal cycling scheme was used: initial denaturation at 98 °C for 30 s, 30 cycles of denaturation at 98 °C for 15 s, annealing at 53 °C for 30 s, followed by extension at 72 °C for 30 s. The final extension was carried out at 72 °C for 2 min. The PCR products that were obtained were controlled for appropriate size, and then purified using the MagSi-NGS Plus kit according to the manufacturer's protocol (Steinbrenner Laborsysteme GmbH, Germany). Quantification of the PCR products was performed using the Quant-iT dsDNA HS assay kit and a Qubit fluorometer, as recommended by the manufacturer (Thermo Scientific). The DNA samples were barcoded using the Nextera XT-Index kit (Illumina, San Diego, USA) and the Kapa HIFI Hot Start polymerase (Kapa Biosystems, USA). Sequencing

was performed at the Göttingen Genomics Laboratory using an Illumina MiSeq Sequencing platform (paired-end 2×300 bp) and the MiSeq reagent kit v3, as recommended by the manufacturer (Illumina).

Processing of the 16S rRNA gene data

Trimmomatic version 0.39 (Bolger et al. 2014) was initially used to truncate low-quality reads if quality dropped below 12 in a sliding window of 4 bp. Datasets were subsequently processed with Usearch version 11.0.667 (Edgar 2010) as described in Wemheuer et al. (2020). In brief, paired-end reads were merged and quality-filtered. Filtering included the removal of low-quality reads and reads shorter than 200 bp. Processed sequences of all samples were joined, dereplicated and clustered in zero-radius operational taxonomic units (zOTUs) using the UNOISE algorithm implemented in Usearch. A *de novo* chimera removal was included in the clustering step. Afterwards, zOTU sequences were taxonomically classified using the SINTAX algorithm against the SILVA database (SILVA SSURef 138 NR99). All non-bacterial zOTUs were removed based on taxonomic classification. Subsequently, processed sequences were mapped on final zOTU sequences to calculate the distribution and abundance of each OTU in every sample.

Metagenome sequencing, assembly and analysis

Sequencing libraries were generated from environmental DNA. These were barcoded using the Nextera XT-Index kit (Illumina, San Diego, USA) and the Kapa HIFI Hot Start polymerase (Kapa Biosystems, Wilmington, USA). Sequencing was performed by employing an Illumina HiSeq 2500 system and the HiSeq Rapid SBS kit V2 (2x250 bp) as recommended by the manufacturer (Illumina). Metagenomic reads were further processed as described previously (Eze et al. 2020). In brief, reads were processed with the Trimmomatic tool version 0.39 (Bolger et al. 2014) and assembled using metaSPAdes version 3.13.2 (Bankevich et al. 2012). Coverage information for each scaffold was determined using Bowtie2 version 2.3.2 (Langmead and Salzberg 2012) and SAMtools version 1.7 (Li et al. 2009). Metagenome-assembled genomes (MAGs) were reconstructed with MetaBAT version 2.12.1 (Kang et al. 2015). MAG quality was determined using CheckM version 1.0.13 (Parks et al. 2015). Only MAGs with a completeness minus contamination of more than 50% and a contamination rate of less than 7% were considered for further analysis. MAGs were classified taxonomically using GTDB-Tk version 1.0.2 and the

Genome Taxonomy Database (release 86) (Chaumeil et al. 2019; Parks et al. 2019). Coding DNA sequences (CDSs) were identified with prodigal version 2.6.3 (Hyatt et al. 2010). Functional annotation was performed with diamond version v0.9.29 (Buchfink et al. 2015) and the KEGG database (October release 2018) (Kanehisa and Goto 2000).

Data analysis

Data analysis was performed in R (RCoreTeam 2018). Richness, diversity, evenness, and coverage based on the Chao1 richness estimator were estimated in R using the vegan package (RCoreTeam 2018). In addition, richness was estimated using the Michaelis-Menten equation in R with the drc package (RCoreTeam 2018). Prior to alpha diversity analysis, the zOTU table was rarefied to 12,924 per sample. Beta-diversity was calculated in R using the vegan package. Non-metric multidimensional scaling plots were generated based on Bray-Curtis dissimilarities. Dissimilarities were calculated based on the raw zOTU table.

RESULTS

Bacterial diversity of the sampling sites and the diesel-degrading cultures

The 16S rRNA gene amplicon sequencing resulted in 242,025 16S rRNA gene sequences across all samples (36,441–10,309 reads per sample, average 22,002 per sample). Clustering resulted in a total of 6,453 zOTUs (average: 587) ranging from 225 to 813 zOTUs per sample. The highest bacterial richness and diversity were observed in the reference samples, the lowest in the enrichment samples. Calculated coverage values indicate that the majority of the bacterial diversity (>80.9%, see Supplementary Table S1) was recovered by the surveying effort.

Non-metric multidimensional scaling revealed clear differences between the microbial community composition of the polluted soil and water samples, enrichment cultures, and reference unpolluted soil samples (Figure 2).

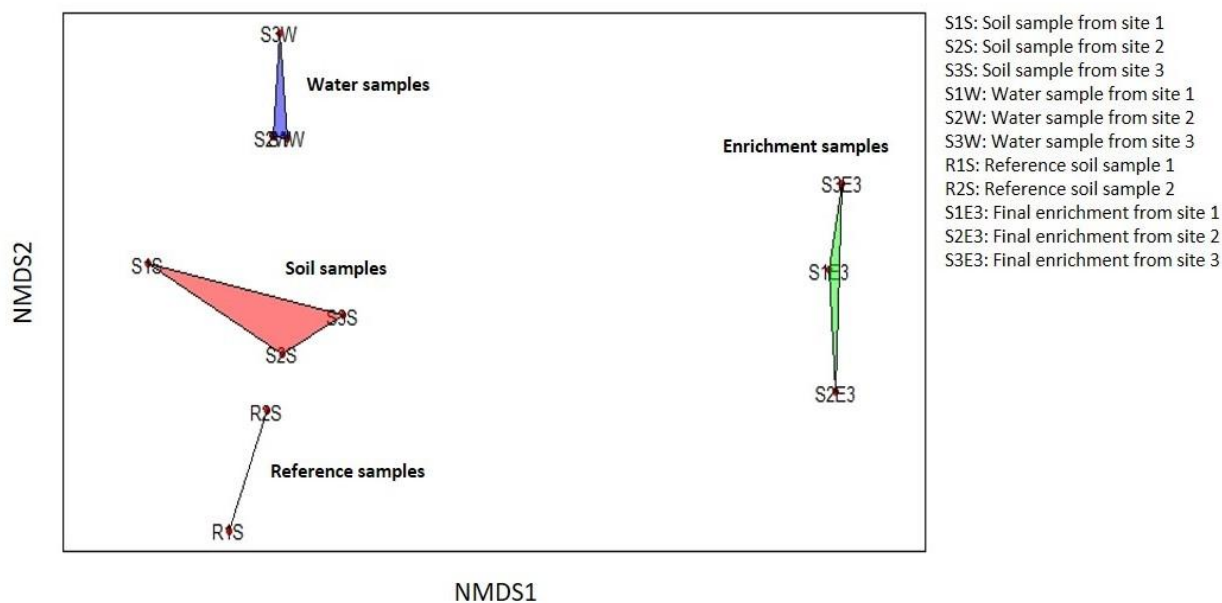


Figure 2. Non-metric multidimensional scaling (NMDS) ordination plot showing differences in microbial community compositions of the water, soil, enrichment, and reference unpolluted soil samples based on community composition at the genus level.

The relative abundances at the bacterial class level (Figure 3a) showed the dominance of *Gammaproteobacteria* in the polluted water sample (90.6%), followed by *Alphaproteobacteria* (3.2%). The polluted soil samples contain similar relative abundances for *Gammaproteobacteria*, *Alphaproteobacteria* and *Acidobacteriae* (26.4%, 21.4% and 19.1%, respectively). The enrichment cultures are dominated by members of the *Alphaproteobacteria*, with a relative abundance of 75.8%. Other bacterial classes present in the enrichment culture include *Gammaproteobacteria* and *Acidobacteriae* (15.4% and 8.6%, respectively). A higher diversity and richness (Supplementary Table S1) was recorded in the unpolluted reference sample in which *Actinobacteria* (17.0%), *Alphaproteobacteria* (14.6%), *Acidobacteriae* (13.5%), and *Bacteroidia* (10.1%) are dominant. Other less abundant classes include *Phycisphaerae* and *Verrucomicrobiae*. At genus level, *Acidocella* are dominant in all the enrichment cultures from the three sites (87.4% to 75.4%). Other genera present in the enrichment cultures include *Acidobacterium* and *Paraburkholderia* (Figure 3b).

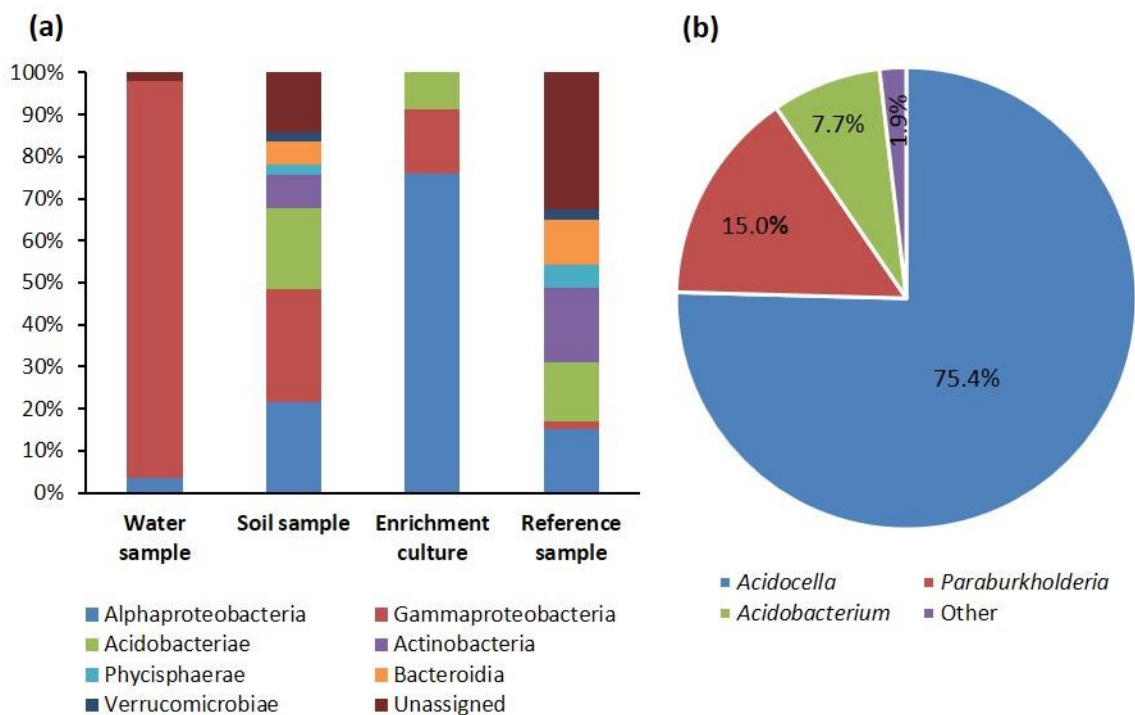


Figure 3. (a) Bacterial community composition in selected water, soil, enrichment, and unpolluted reference samples. (b) relative abundance of the enrichment culture at the genus level. Only taxa with a relative abundance of >1% across all samples are presented. For details on relative abundances and 16S rRNA gene amplicon data, see Supplementary Figure S1.

Identification of aliphatic and aromatic hydrocarbon-degrading coding DNA sequences

Functional analysis of the metagenome derived from the microbial diesel enrichment revealed the presence of 42 potential enzymatic classes represented by 186 coding DNA sequences (CDSs) involved in the degradation of aliphatic and aromatic hydrocarbons (Figure 4).

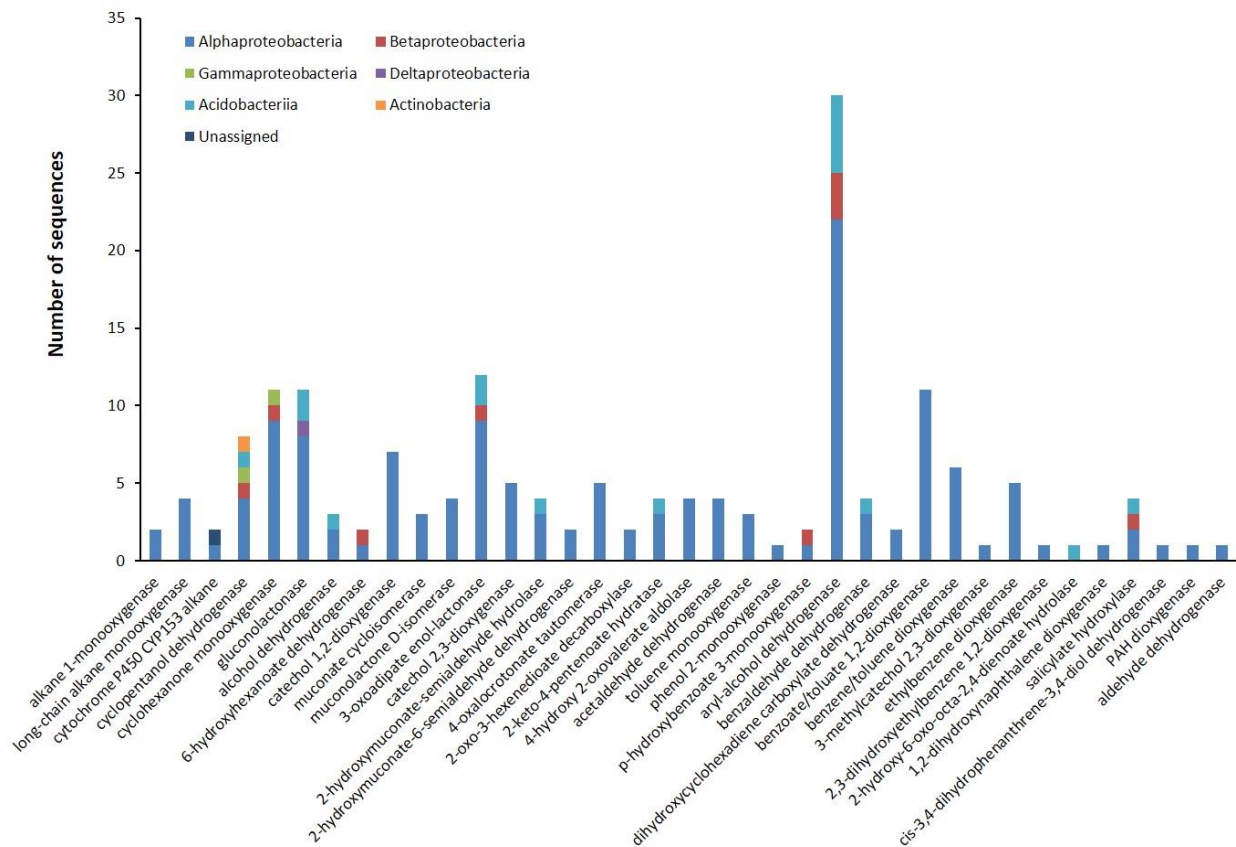


Figure 4. The number of sequences associated with specific hydrocarbon-degrading enzymes in each taxonomic group. The analysis was based on the metagenome of the S3E3 enrichment culture.

The enzymes considered as responsible for the degradation of aliphatic hydrocarbons included alkane 1-monoxygenase, long-chain alkane monoxygenase, cytochrome P450 CYP153 alkane hydroxylase, cyclopentanol dehydrogenase, cyclohexanone monoxygenase, gluconolactonase, alcohol dehydrogenase, and 6-hydroxyhexanoate dehydrogenase. Forty-three CDSs were detected that are considered to play a role in aliphatic hydrocarbon degradation. The majority of the genes that putatively code for aliphatic hydrocarbon degradation are involved in cycloalkane degradation. These include the *cpnA*, *chnB*, *gnl*, *adh* and *chnD* genes, which are involved in the Baeyer-Villiger oxidation reactions (Figure 5).

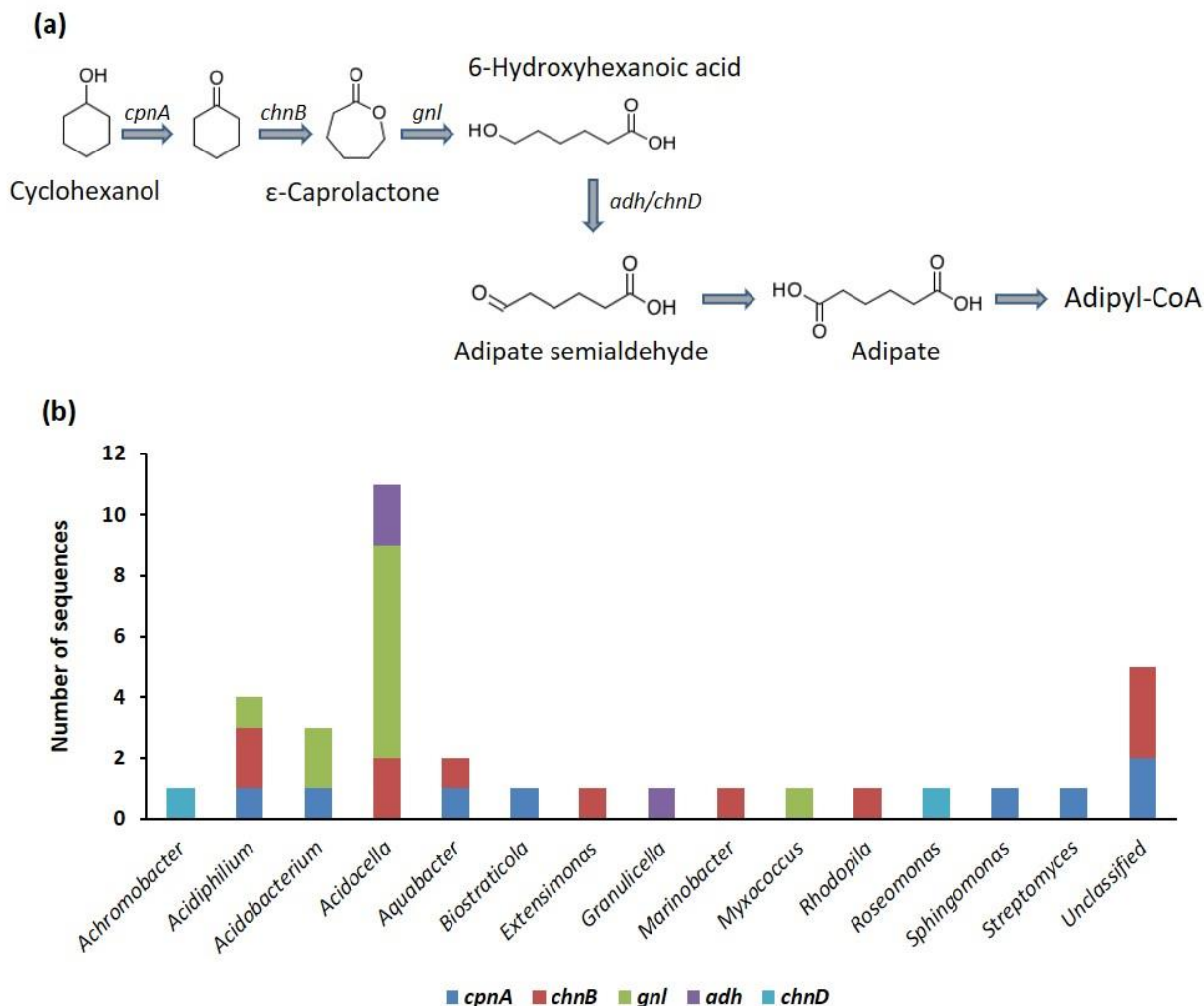


Figure 5. (a) Ring cleavage via the Baeyer-Villiger oxidation pathway for the degradation of cycloalkanes, and (b) genus assignment of the putative genes involved in the Baeyer-Villiger oxidation pathway identified in the diesel-degrading consortium.

The degradation of aromatic hydrocarbons occurs through a series of reactions involving oxidation, hydroxylation, dehydrogenation and ring cleavage. Out of the 186 CDSs putatively linked to diesel degradation, 143 CDSs are potentially involved in aromatic hydrocarbon degradation. Among the 48 CDSs belonging to the aromatic ring dioxygenases, eleven were annotated as benzoate/toluene 1,2-dioxygenase, six as biphenyl 2,3-dioxygenase, six as benzene/toluene/chlorobenzene dioxygenase, five as ethylbenzene dioxygenase, and three as naphthalene 1,2-dioxygenase (Table 1).

Table 1. Key monooxygenases and dioxygenases involved in the activation and ring cleavage of aromatic hydrocarbons in the diesel-degrading consortium.

| Genes | Enzyme | Function | No. of CDSs |
|------------------|---|-------------------|-------------|
| <i>tmoCF</i> | Toluene monooxygenase | Activation | 3 |
| <i>pobA</i> | p-Hydroxybenzoate 3-monooxygenase | Activation | 2 |
| <i>todABC1C2</i> | Benzene/toluene/chlorobenzene dioxygenase | Activation | 6 |
| <i>etbAaAbAc</i> | Ethylbenzene dioxygenase | Activation | 5 |
| <i>benABC</i> | Benzoate/toluate 1,2-dioxygenase | Activation | 11 |
| <i>bphA</i> | Biphenyl 2,3-dioxygenase | Activation | 6 |
| <i>nahAb</i> | Naphthalene 1,2-dioxygenase | Activation (PAHs) | 3 |
| <i>nahC</i> | 1,2-Dihydroxynaphthalene dioxygenase | Activation (PAHs) | 1 |
| <i>nidA</i> | PAH dioxygenase | Activation (PAHs) | 1 |
| <i>catA</i> | Catechol 1,2-dioxygenase | Ortho-cleavage | 7 |
| <i>dmpB</i> | Catechol 2,3-dioxygenase | Meta-cleavage | 5 |
| <i>todE</i> | 3-Methylcatechol 2,3-dioxygenase | Meta-cleavage | 1 |
| <i>etbC</i> | 2,3-Dihydroxyethylbenzene 1,2-dioxygenase | Meta-cleavage | 1 |
| <i>bphC</i> | Biphenyl-2,3-diol 1,2-dioxygenase | Meta-cleavage | 1 |

Reconstruction of metagenome-assembled genomes

We were able to reconstruct fifteen nearly complete genomes from the whole-metagenome sequence of the original soil samples, and three nearly complete genomes from the enrichment culture (Supplementary Table S2). Quality analysis of the MAGs showed that the average completeness and contamination level for the MAGs were 85% and 2% respectively (Supplementary Table S3). The majority of the metagenome-assembled genomes (MAGs) were classified as belonging to the *Gammaproteobacteria* (8 MAGs), followed by *Alphaproteobacteria* (4 MAGs), *Acidobacteriae* (3 MAGs), *Actinobacteria* (2 MAGs) and *Caldisericia* (1 MAG). The three metagenome-assembled genomes from the enrichment culture were classified as *Acidocella aminolytica*, *Acidobacterium capsulatum*, and *Acidocella* sp., with a completeness of 72.4%, 99.8% and 100%, respectively.

A comparison of the three nearly complete genomes reconstructed from the metagenome of the enrichment culture shows that the genes encoding enzymes involved in the activation and degradation of petroleum hydrocarbons are more abundant in *Acidocella* than in *Acidobacterium* (Supplementary Table S2). For example, while the two MAGs classified as *Acidocella* contain an average of 18 CDSs involved in aromatic ring activation, *Acidobacterium* had only 7 CDSs encoding for the activation of aromatic hydrocarbons. Key enzymes that are encoded by the reconstructed MAGs belonging to *Acidocella* but are missing in those belonging to *Acidobacterium* include long-chain alkane monooxygenase, cyclohexanone monooxygenase, ethylbenzene dioxygenase, and benzoate/toluene 1,2-dioxygenase.

Further comparisons performed between the MAGs assembled from the metagenome data of the enrichment culture and those obtained from a previous study of a crude oil bore hole (Eze et al. 2020) revealed that the *Acidocella* MAGs obtained from this study exhibit a higher abundance of genes that putatively encode the degradation of cycloalkanes. For example, in the 36 MAGs from Eze et al. (2020), genes that encode for cyclopentanol dehydrogenase (*cpnA*) and cyclohexanone monooxygenase (*chnB*) were present in 16 and 11 MAGs, respectively. In this study, MAGs reconstructed from both the enrichment culture and the original soil samples were rich in genes that encode these enzymes with more than 6 CDSs per gene in some MAGs. The reconstructed MAGs were also found to be richer in CDSs that encode for aromatic degradation than those in the previous study. For example, aryl alcohol dehydrogenase, an enzyme vital for the degradation of aromatic hydrocarbons was missing in all of the 36 assembled MAGs from the crude oil bore hole study (Eze et al. 2020). Potential genes encoding the enzyme were present in two of the three MAGs from the enrichment culture of this study.

DISCUSSION

The successive enrichment of the different experimental samples using diesel fuel resulted in the dominance of *Alphaproteobacteria*. The dominance of *Alphaproteobacteria* in the bacterial communities, especially *Acidocella* and *Paraburkholderia*, indicates the tolerance of these genera to high concentrations of petroleum hydrocarbons and their potential degradative capacity for organic contaminants. The taxa that are abundant in the polluted water and soil, and in the

enrichment cultures, were also associated with hydrocarbon pollution in other locations (Lee et al. 2019; Röling et al. 2006; Stapleton et al. 1998). The biodegradative ability of these taxa and their tolerance to heavy metals (Giovanella et al. 2020) indicate that they are potentially suitable for the remediation of multiple contaminants such as hydrocarbon-polluted acidic mine sites.

Diesel fuel contains aliphatic and aromatic hydrocarbons. The aliphatic hydrocarbon fraction is predominantly composed of normal-, iso- and cyclo-alkanes, while the aromatic hydrocarbon fraction is composed primarily of alkylbenzenes, naphthalene, alkylnaphthalenes, biphenyl and alkylbiphenyls (Woolfenden et al. 2011). The degradation of n-alkanes is primarily carried out by alkane 1-monooxygenase (*alkB*), cytochrome P450 CYP153 alkane hydroxylase (*CYP153*) and long-chain alkane monooxygenases (*ladA*) genes, and their roles in the degradation of n-alkanes and iso-alkanes have been extensively studied (Ji et al. 2013; Li et al. 2008; van Beilen et al. 2006). The degradation of n-alkanes and iso-alkanes by the consortium is indicated by the presence of potential *alkB*, *CYP153* and *ladA* genes. The low number of the corresponding gene sequences (eight) can be explained by the taxonomic composition of the consortium. Previous studies have shown that n-alkane degrading genes are often associated with *Betaproteobacteria* and *Gammaproteobacteria* especially the *Pseudomonas* genus (Garrido-Sanz et al. 2019; Liu et al. 2014; Shao and Wang 2013; van Beilen et al. 2001; van Beilen et al. 1994). In our study, the diesel-degrading consortium in the enrichment cultures was dominated by *Alphaproteobacteria* (Figures 2 and 3). Thus, the majority of CDSs in our metagenome consortium belong to the *Alphaproteobacteria*, especially the *Acidocella* genus and not to *Pseudomonas*.

Of the genes that putatively code for aliphatic hydrocarbon degradation, the majority are involved in cycloalkane degradation. These enzymes include cyclopentanol dehydrogenase (*cpnA*), cyclohexanone monooxygenase (*chnB*), gluconolactonase (*gnl*), alcohol dehydrogenase (*adh*), and 6-hydroxyhexanoate dehydrogenase (*chnD*) (Bohren et al. 1989; Iwaki et al. 1999; Iwaki et al. 2002; Kanagasundaram and Scopes 1992). This is interesting since cycloalkanes are moderately resistant to biodegradation (Connan 1984). The degradation of cycloalkanes involves ring cleavage via Baeyer-Villiger oxidation (Perkel et al. 2018; Sheng et al. 2001), which requires an initial oxidation of cyclohexane to cyclohexanol by cyclohexane monooxygenase, and then a dehydrogenation reaction to cyclohexanone. This step is followed by another monooxygenase attack to form epsilon-caprolactone, followed by ring cleavage that is carried out by

gluconolactonase (Figure 5a). All the genes involved in this degradation pathway are present in the metagenome of the enrichment culture, but a single taxon in the bacterial community that possess all the genes involved in this pathway was not detected (Figure 5b). This indicates a synergistic interaction of different bacterial genera in the degradation of recalcitrant hydrocarbons. The high number of *cpnA*, *chnB*, *gnl*, *adh* and *chnD* genes (35 CDSs) in the metagenome of the enrichment culture indicates the significant potential of the microbial community for the degradation of cycloalkanes present in diesel fuel.

The degradation of aromatic hydrocarbons requires initial activation by oxygenases resulting in the formation of oxygenated intermediates such as catechol (Atashgahi et al. 2018; Das and Chandran 2011; Peters et al. 2004). The bacterial consortium contains more genes that putatively encode dioxygenases than those that encode monooxygenases (Table 1). The genes that encode dioxygenases include the *todABC1C2*, *etbAaAbAc* and *benABCD* genes (Fong et al. 1996; Werlen et al. 1996; Zylstra and Gibson 1989). The higher abundance of genes encoding dioxygenases indicates that the activation of alkylbenzenes and phenolic compounds by the microbial consortium predominantly follows the dioxygenase pathway rather than the monooxygenase pathway.

The central metabolism of aromatic hydrocarbons that follows initial activation involves ortho- and meta-cleavage of catechol or methylcatechol (Benjamin et al. 1991; Ehrt et al. 1995; Hidalgo et al. 2020; Liang et al. 2019; Neidle et al. 1988; Peters et al. 2004; Rohrbacher and St-Arnaud 2016). Functional analysis reveals that genes encoding enzymes putatively involved in the central metabolism of aromatic hydrocarbons are present in the microbial community. The most abundant CDSs in our diesel-degrading community that are responsible for this reaction are catechol 1,2-dioxygenase and catechol 2,3-dioxygenase (7 and 5 CDSs, respectively) (Table 1). Other enzymes that are present include 3-oxoadipate enol-lactonase, muconolactone D-isomerase (a decarboxylating dehydrogenase), 4-oxalocrotonate tautomerase, and acetaldehyde dehydrogenase. Most of the corresponding genes are affiliated to *Alphaproteobacteria*.

Polycyclic aromatic hydrocarbons (PAHs) are more resistant to microbial attack than smaller aromatic hydrocarbons, and when biodegradation is possible, this often proceeds through oxidation and ring cleavage by dioxygenases (Sipilä et al. 2008). The metagenome contains genes

that encode enzymes putatively involved in the degradation of PAHs and other recalcitrant hydrocarbons, such as biphenyl and alkylbiphenyls. These enzymes include naphthalene 1,2-dioxygenase (*nahAb*) and 1,2-dihydroxynaphthalene dioxygenase (*nahC*) for naphthalene and alkyl-naphthalenes (Peng et al. 2008), biphenyl 2,3-dioxygenase (*bphA*) for biphenyl and alkylbiphenyls, and PAH dioxygenase (*nidA*) for phenanthrene, alkylphenanthrenes, and other high molecular weight PAHs (Iwasaki et al. 2006; Robrock et al. 2011) (Table 1). Since crude oil and oil spills often contain significant amount of polycyclic aromatic hydrocarbons such as naphthalene, alkyl-naphthalenes, phenanthrene and alkylphenanthrenes (Ahmed and George 2004; Eze and George 2020), the presence of putative genes encoding PAH dioxygenases in the metagenome of the consortium indicates the potential of the consortium for the remediation and reclamation of petroleum-contaminated soils.

Interestingly, the majority of previous studies on microbially-enhanced rhizoremediation of petroleum hydrocarbons have focused on *Pseudomonas* (de Lima-Morales et al. 2015; Di Martino et al. 2012), *Burkholderia* (Okoh et al. 2001), and *Paraburkholderia* (Dias et al. 2019; Lee and Jeon 2018), but these organisms often do not have the enzymes to run the complete metabolic pathways for the degradation of all hydrocarbons present in diesel fuel, especially the aromatic constituents. For example, in a study of rhizoremediation of diesel-contaminated soils, a scarcity of ring-hydroxylating and ring-cleavage dioxygenases among *Gammaproteobacteria* was reported by Garrido-Sanz et al. (2019). These researchers also noted that none of the *nahA* genes in the metagenome was affiliated to *Pseudomonas* or even to the *Gammaproteobacteria* class that dominated the PAH-degrading consortium. In contrast, the consortium reported here contains the CDSs required for the complete degradation of these aromatic components in diesel fuel.

The comparison made between the MAGs assembled from the metagenome data of the enrichment culture and those obtained from studies of similar sites (Eze et al. 2020) revealed the relatively higher abundance, in the consortium, of genes involved in hydrocarbon degradation. These include the *adhP* and *yiaY* genes encoding alcohol dehydrogenases (Drewke and Ciriacy 1988; Glasner et al. 1995; Williamson and Paquin 1987), and the *cpnA* and *chnB* involved in the degradation of cycloalkanes (Iwaki et al. 2002; Sheng et al. 2001). These genes are also involved in the degradation of other organic contaminants such as haloalkanes (Belkin 1992; Yokota et al. 1986). This difference in potential degradative capacity between the MAGs from the two studies

can be explained by the taxonomic differences between the MAGs obtained in both cases. In the study of a crude oil bore hole (Eze et al. 2020), majority of the reconstructed MAGs were affiliated to *Gammaproteobacteria*. In contrast, *Alphaproteobacteria*, especially *Acidocella* was the dominant genus in both the enrichment culture and the MAGs from the enrichment culture.

The potential of the enrichment culture to degrade recalcitrant hydrocarbons was also revealed by the presence of genes encoding enzymes involved in degradation of recalcitrant organic compounds. For example, one of the three MAGs from the enrichment culture contained genes encoding 2-halobenzoate 1,2-dioxygenase (*cbdA*), an enzyme that activates the oxidation of 2-chlorobenzoate to catechol. In contrast, none of the 36 MAGs from the previous study contains this gene. Since the enrichment culture is composed of predominantly *Acidocella* strains, the abundance of genes that putatively encode for the degradation of cycloalkanes and aromatic hydrocarbons in the MAGs classified as *Acidocella* is an indication for the potential of the consortium for petroleum hydrocarbon biodegradation.

Conclusions

The degradation of petroleum hydrocarbons requires several microorganisms with both the ability to withstand toxicity and to harbour the required metabolic pathways. Therefore, a foremost step in establishing a successful microbially-mediated bioremediation approach is the selective cultivation of a suitable microbial consortium with the required degradative capability for the target contaminants. Through successive enrichment using soil samples taken from an historical oil-contaminated site in Germany, we successfully generated a bacterial consortium capable of degrading diesel fuel. We further reconstructed a total of 18 genomes from both the original soil sample and the isolated consortium. The analysis of both the metagenome of the consortium and the reconstructed metagenome-assembled genomes shows that the most abundant bacterial genus in the consortium, *Acidocella*, possess many of the coding DNA sequences required for the degradation of diesel fuel aromatic hydrocarbons, which are often the most toxic components. This can explain why this genus proliferated in all the enrichment cultures. Therefore, this study revealed that the microbial consortium isolated in this study or its dominant genus, *Acidocella*, could potentially serve as an effective inoculum for biotechnological applications in the reclamation of soils contaminated with diesel fuel.

Conflict of Interest

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

Author Contributions

Conceptualization and design: MOE, GCH, SCG and RD. Planning and implementation: MOE and RD. Experiments and bioinformatics analyses: MOE. Writing – original draft: MOE. Writing – review and editing: GCH, SCG and RD. Supervision: GCH, SCG and RD. All authors interpreted the results, and agreed to the final version of the manuscript.

Acknowledgments

The authors would like to thank Macquarie University and the Commonwealth Government of Australia for supporting this research project by providing M.O.E. with an international Research Training Program (iRTP) scholarship, and the German Academic Exchange Service (DAAD) for providing M.O.E. with a DAAD scholarship (Allocation Numbers: 2017561 and 91731339, respectively). This publication was supported financially by the Open Access Publication Fund of the University of Göttingen. The funders had no role in study design, data collection, and interpretation, or the decision to submit the work for publication. We also thank Dr. Anja Poehlein and Melanie Heinemann for assistance during the sequencing.

Data Availability

Raw sequencing data has been deposited in the sequence read archive of the National Center for Biotechnology Information under BioProject number [PRJNA612814](#).

References

- Ahmed M, George SC (2004) Changes in the molecular composition of crude oils during their preparation for GC and GC–MS analyses. *Org. Geochem.* 35:137-155. doi:<https://doi.org/10.1016/j.orggeochem.2003.10.002>.
- Atashgahi S, Hornung B, van der Waals MJ, da Rocha UN, Hugenholtz F, Nijssse B, Molenaar D, van Spanning R, Stams AJM, Gerritse J, Smidt H (2018) A benzene-degrading nitrate-reducing microbial consortium displays aerobic and anaerobic benzene degradation pathways. *Sci. Rep.* 8:4490. doi:10.1038/s41598-018-22617-x.
- Atlas RM (2010) Handbook of microbiological media, Fourth edition edn. CRC Press, Boca Raton, Florida.
- Azubuikwe CC, Chikere CB, Okpokwasili GC (2016) Bioremediation techniques–classification based on site of application: principles, advantages, limitations and prospects. *World J. Microbiol. Biotechnol.* 32:180. doi:10.1007/s11274-016-2137-x.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19:455-477. doi:10.1089/cmb.2012.0021.
- Belkin S (1992) Biodegradation of haloalkanes. *Biodegradation* 3:299-313. doi:10.1007/BF00129090.
- Benjamin RC, Voss JA, Kunz DA (1991) Nucleotide sequence of *xyIE* from the TOL pDK1 plasmid and structural comparison with isofunctional catechol-2,3-dioxygenase genes from TOL, pWW0 and NAH7. *J. Bacteriol.* 173:2724. doi:10.1128/jb.173.8.2724-2728.1991.
- Bohren KM, Bullock B, Wermuth B, Gabbay KH (1989) The aldo-keto reductase superfamily. cDNAs and deduced amino acid sequences of human aldehyde and aldose reductases. *J. Biol. Chem.* 264:9547-51.
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120. doi:10.1093/bioinformatics/btu170.
- Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12:59-60. doi:10.1038/nmeth.3176.
- Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH (2019) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925-1927. doi:10.1093/bioinformatics/btz848.
- Connan J (1984) Biodegradation of crude oils in reservoirs. In: Advances in Petroleum Geochemistry, volume 1. Academic Press, London.
- Craig J, Gerali F, MacAulay F, Sorkhabi R (2018) The history of the European oil and gas industry (1600s–2000s). *Geological Society, London, Special Publications* 465. doi:10.1144/SP465.23.
- Dalton T, Jin D (2010) Extent and frequency of vessel oil spills in US marine protected areas. *Mar. Pollut. Bull.* 60:1939-1945. doi:<https://doi.org/10.1016/j.marpolbul.2010.07.036>.
- Das N, Chandran P (2011) Microbial degradation of petroleum hydrocarbon contaminants: An overview. *Biotechnol. Res. Int.* 2011. doi:10.4061/2011/941810.

- de Lima-Morales D, Chaves-Moreno D, Wos-Oxley ML, Jáuregui R, Vilchez-Vargas R, Pieper DH (2015) Degradation of benzene by *Pseudomonas veronii* 1YdBTEX2 and 1YB2 is catalyzed by enzymes encoded in distinct catabolism gene clusters. *Appl. Environ. Microbiol.* 82:167-173. doi:10.1128/AEM.03026-15.
- Di Martino C, López NI, Raiger Iustman LJ (2012) Isolation and characterization of benzene, toluene and xylene degrading *Pseudomonas* sp. selected as candidates for bioremediation. *Int. Biodeterior. Biodegradation* 67:15-20 doi:<https://doi.org/10.1016/j.ibiod.2011.11.004>
- Dias GM, de Sousa Pires A, Grilo VS, Castro MR, de Figueiredo Vilela L, Neves BC (2019) Comparative genomics of *Paraburkholderia kururiensis* and its potential in bioremediation, biofertilization, and biocontrol of plant pathogens. *MicrobiologyOpen* 8:e00801. doi:10.1002/mbo3.801.
- Drewke C, Ciriacy M (1988) Overexpression, purification and properties of alcohol dehydrogenase IV from *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* 950:54-60. doi:[https://doi.org/10.1016/0167-4781\(88\)90072-3](https://doi.org/10.1016/0167-4781(88)90072-3).
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460-2461. doi:10.1093/bioinformatics/btq461.
- Ehrt S, Schirmer F, Hillen W (1995) Genetic organization, nucleotide sequence and regulation of expression of genes encoding phenol hydroxylase and catechol 1,2-dioxygenase in *Acinetobacter calcoaceticus* NCIB8250. *Mol. Microbiol.* 18:13-20. doi:10.1111/j.1365-2958.1995.mmi_18010013.x.
- Errington I, King CK, Wilkins D, Spedding T, Hose GC (2018) Ecosystem effects and the management of petroleum-contaminated soils on subantarctic islands. *Chemosphere* 194:200-210. doi:<https://doi.org/10.1016/j.chemosphere.2017.11.157>.
- Eze MO, George SC (2020) Ethanol-blended petroleum fuels: implications of co-solvency for phytotechnologies. *RSC Adv.* 10:6473-6481. doi:10.1039/C9RA10919F.
- Eze MO, Lütgert SA, Neubauer H, Balouri A, Kraft AA, Sieven A, Daniel R, Wemheuer B (2020) Metagenome assembly and metagenome-assembled genome sequences from a historical oil field located in Wietze, Germany. *Microbiol. Resour. Announc.* 9:e00333-20. doi:10.1128/MRA.00333-20.
- Fong KP, Goh CB, Tan HM (1996) Characterization and expression of the plasmid-borne *bedD* gene from *Pseudomonas putida* ML2, which codes for a NAD⁺-dependent cis-benzene dihydrodiol dehydrogenase. *J. Bacteriol.* 178:5592. doi:10.1128/jb.178.19.5592-5601.1996.
- Garrido-Sanz D, Redondo-Nieto M, Guirado M, Pindado Jiménez O, Millán R, Martín M, Rivilla R (2019) Metagenomic insights into the bacterial functions of a diesel-degrading consortium for the rhizoremediation of diesel-polluted soil. *Genes* 10. doi:10.3390/genes10060456.
- Gemmell RT, Knowles CJ (2000) Utilisation of aliphatic compounds by acidophilic heterotrophic bacteria. The potential for bioremediation of acidic wastewaters contaminated with toxic organic compounds and heavy metals. *FEMS Microbiol. Lett.* 192:185-190. doi:10.1111/j.1574-6968.2000.tb09380.x.
- Giovanella P, Vieira GAL, Ramos Otero IV, Pais Pellizzer E, de Jesus Fontes B, Sette LD (2020) Metal and organic pollutants bioremediation by extremophile microorganisms. *J. Hazard. Mater.* 382:121024. doi:<https://doi.org/10.1016/j.jhazmat.2019.121024>

- Glasner JD, Kocher TD, Collins JJ (1995) *Caenorhabditis elegans* contains genes encoding two new members of the Zn-containing alcohol dehydrogenase family. *J. Mol. Evol.* 41:46-53 doi:10.1007/BF00174040.
- Hara E, Uchiyama H (2013) Degradation of petroleum pollutant materials by fungi. In: Goltapeh EM, Danesh YR, Varma A (eds) *Fungi as Bioremediators*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 117-133.
- Hassler B (2016) Oil Spills from Shipping: A Case Study of the Governance of Accidental Hazards and Intentional Pollution in the Baltic Sea. In: Gilek M, Karlsson M, Linke S, Smolarz K (eds) *Environmental Governance of the Baltic Sea*. Springer International Publishing, Cham, pp 125-146.
- Hidalgo KJ, Sierra-Garcia IN, Dellagnezze BM, de Oliveira VM (2020) Metagenomic insights into the mechanisms for biodegradation of polycyclic aromatic hydrocarbons in the oil supply chain. *Front. Microbiol.* 11. doi:10.3389/fmicb.2020.561506.
- Hong X, Chen W, Zhang L (2010) A probabilistic risk forecast of accidental oil spills from vessels in Luoyuan Bay, Fujian Province, PRC. *Procedia Environ. Sci.* 2:49-56. doi:<https://doi.org/10.1016/j.proenv.2010.10.008>.
- Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* 11:119-119. doi:10.1186/1471-2105-11-119.
- Iwaki H, Hasegawa Y, Teraoka M, Tokuyama T, Bergeron H, Lau PCK (1999) Identification of a transcriptional activator (ChnR) and a 6-Oxohexanoate dehydrogenase (ChnE) in the cyclohexanol catabolic pathway in *Acinetobacter* sp. strain NCIMB 9871 and localization of the genes that encode them. *Appl. Environ. Microbiol.* 65:5158. doi:10.1128/AEM.65.11.5158-5162.1999.
- Iwaki H, Hasegawa Y, Wang S, Kayser MM, Lau PCK (2002) Cloning and characterization of a gene cluster involved in cyclopentanol metabolism in *Comamonas* sp. strain NCIMB 9872 and biotransformations effected by *Escherichia coli*-expressed cyclopentanone 1,2-monooxygenase. *Appl. Environ. Microbiol.* 68:5671. doi:10.1128/AEM.68.11.5671-5684.2002.
- Iwasaki T, Miyauchi K, Masai E, Fukuda M (2006) Multiple-subunit genes of the aromatic-ring-hydroxylating dioxygenase play an active role in biphenyl and polychlorinated biphenyl degradation in *Rhodococcus* sp. strain RHA1. *Appl. Environ. Microbiol.* 72:5396. doi:10.1128/AEM.00298-06.
- Ji Y, Mao G, Wang Y, Bartlam M (2013) Structural insights into diversity and n-alkane biodegradation mechanisms of alkane hydroxylases. *Front. Microbiol.* 4:58
- Joner EJ, Corgié SC, Amellal N, Leyval C (2002) Nutritional constraints to degradation of polycyclic aromatic hydrocarbons in a simulated rhizosphere. *Soil Biol. Biochem.* 34:859-864. doi:[https://doi.org/10.1016/S0038-0717\(02\)00018-4](https://doi.org/10.1016/S0038-0717(02)00018-4).
- Kanagasundaram V, Scopes R (1992) Isolation and characterization of the gene encoding gluconolactonase from *Zymomonas mobilis*. *Biochim. Biophys. Acta* 1171:198-200. doi:[https://doi.org/10.1016/0167-4781\(92\)90120-O](https://doi.org/10.1016/0167-4781(92)90120-O).
- Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28:27-30. doi:10.1093/nar/28.1.27.
- Kang DD, Froula J, Egan R, Wang Z (2015) MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3:e1165-e1165. doi:10.7717/peerj.1165.

- Kleinstaub S, Riis V, Fetzer I, Harms H, Müller S (2006) Population dynamics within a microbial consortium during growth on diesel fuel in saline environments. *Appl. Environ. Microbiol.* 72:3531. doi:10.1128/AEM.72.5.3531-3542.2006.
- Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41:e1-e1. doi:10.1093/nar/gks808.
- Kumar BL, Gopal DVRS (2015) Effective role of indigenous microorganisms for sustainable environment. *3 Biotech* 5:867-876. doi:10.1007/s13205-015-0293-6.
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357-359. doi:10.1038/nmeth.1923.
- Leahy JG, Colwell RR (1990) Microbial degradation of hydrocarbons in the environment. *Microbiol. Rev.* 54:305-315.
- Lee Y, Jeon CO (2018) *Paraburkholderia aromaticivorans* sp. nov., an aromatic hydrocarbon-degrading bacterium, isolated from gasoline-contaminated soil. *Int. J. Syst. Evol. Microbiol.* 68:1251-1257. doi:<https://doi.org/10.1099/ijsem.0.002661>.
- Lee Y, Lee Y, Jeon CO (2019) Biodegradation of naphthalene, BTEX, and aliphatic hydrocarbons by *Paraburkholderia aromaticivorans* BN5 isolated from petroleum-contaminated soil. *Sci. Rep.* 9:860. doi:10.1038/s41598-018-36165-x.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078-2079. doi:10.1093/bioinformatics/btp352.
- Li L, Liu X, Yang W, Xu F, Wang W, Feng L, Bartlam M, Wang L, Rao Z (2008) Crystal structure of long-chain alkane monooxygenase (LadA) in complex with coenzyme FMN: unveiling the long-chain alkane hydroxylase. *J. Mol. Biol.* 376:453-465. doi:<https://doi.org/10.1016/j.jmb.2007.11.069>.
- Liang C, Huang Y, Wang H (2019) *pahE*, a functional marker gene for polycyclic aromatic hydrocarbon-degrading bacteria. *Appl. Environ. Microbiol.* 85:e02399-18. doi:10.1128/AEM.02399-18.
- Liang J-L, Jiang Yang J-H, Nie Y, Wu X-L (2016) Regulation of the alkane hydroxylase CYP153 Gene in a gram-positive alkane-degrading bacterium, *Dietzia* sp. strain DQ12-45-1b. *Appl. Environ. Microbiol.* 82:608. doi:10.1128/AEM.02811-15.
- Liu H, Xu J, Liang R, Liu J (2014) Characterization of the medium- and long-chain n-alkanes degrading *Pseudomonas aeruginosa* strain SJTD-1 and its alkane hydroxylase genes. *PLOS ONE* 9:e105506. doi:10.1371/journal.pone.0105506.
- Lohi A, Alvarez Cuenca M, Anania G, Upreti SR, Wan L (2008) Biodegradation of diesel fuel-contaminated wastewater using a three-phase fluidized bed reactor. *J. Hazard. Mater.* 154:105-111. doi:<https://doi.org/10.1016/j.jhazmat.2007.10.001>.
- Neidle EL, Hartnett C, Bonitz S, Ornston LN (1988) DNA sequence of the *Acinetobacter calcoaceticus* catechol 1,2-dioxygenase I structural gene *catA*: evidence for evolutionary divergence of intradiol dioxygenases by acquisition of DNA sequence repetitions. *J. Bacteriol.* 170:4874. doi:10.1128/jb.170.10.4874-4880.1988.
- Nie Y, Chi C-Q, Fang H, Liang J-L, Lu S-L, Lai G-L, Tang Y-Q, Wu X-L (2014) Diverse alkane hydroxylase genes in microorganisms and environments. *Sci. Rep.* 4:4968. doi:10.1038/srep04968.

- Okoh A, Ajisebutu S, Babalola G, Trejo-Hernandez M (2001) Potential of *Burkholderia cepacia* RQ1 in the biodegradation of heavy crude oil. *Int. Microbiol.* 4:83-87. doi:10.1007/s101230100018.
- Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P (2019) Selection of representative genomes for 24,706 bacterial and archaeal species clusters provide a complete genome-based taxonomy. *bioRxiv*:771964. doi:10.1101/771964.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25:1043-1055. doi:10.1101/gr.186072.114.
- Peng R-H, Xiong A-S, Xue Y, Fu X-Y, Gao F, Zhao W, Tian Y-S, Yao Q-H (2008) Microbial biodegradation of polyaromatic hydrocarbons. *FEMS Microbiol. Rev.* 32:927-955. doi:10.1111/j.1574-6976.2008.00127.x.
- Perkel AL, Voronina SG, Borkina GG (2018) The role of the Baeyer–Villiger reaction in the liquid-phase oxidation of organic compounds. *Russ. Chem. Bull.* 67:779-786. doi:10.1007/s11172-018-2137-0.
- Peters KE, Walters CC, Moldowan JM (2004) *The Biomarker Guide: Volume 2: Biomarkers and Isotopes in Petroleum Systems and Earth History.* Cambridge University Press.
- R Core Team (2018) *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.
- Robrock KR, Mohn WW, Eltis LD, Alvarez-Cohen L (2011) Biphenyl and ethylbenzene dioxygenases of *Rhodococcus jostii* RHA1 transform PBDEs. *Biotechnol. Bioeng.* 108:313-321. doi:10.1002/bit.22952.
- Rohrbacher F, St-Arnaud M (2016) Root exudation: the ecological driver of hydrocarbon rhizoremediation. *Agronomy* 6. doi:10.3390/agronomy6010019.
- Röling WFM, Ortega-Lucach S, Larter SR, Head IM (2006) Acidophilic microbial communities associated with a natural, biodegraded hydrocarbon seepage. *J. Appl. Microbiol.* 101:290-299. doi:10.1111/j.1365-2672.2006.02926.x.
- Sarkar J, Kazy SK, Gupta A, Dutta A, Mohapatra B, Roy A, Bera P, Mitra A, Sar P (2016) Biostimulation of indigenous microbial community for bioremediation of petroleum refinery sludge. *Front. Microbiol.* 7:1407. doi:10.3389/fmicb.2016.01407.
- Shao Z, Wang W (2013) Enzymes and genes involved in aerobic alkane degradation. *Front. Microbiol.* 4:116. doi:10.3389/fmicb.2013.00116.
- Sheng D, Ballou DP, Massey V (2001) Mechanistic studies of cyclohexanone monooxygenase: chemical properties of intermediates involved in catalysis. *Biochemistry* 40:11156-11167. doi:10.1021/bi011153h.
- Sipilä TP, Keskinen A-K, Åkerman M-L, Fortelius C, Haahtela K, Yrjälä K (2008) High aromatic ring-cleavage diversity in birch rhizosphere: PAH treatment-specific changes of I.E.3 group extradiol dioxygenases and 16S rRNA bacterial communities in soil. *ISME J.* 2:968-981. doi:10.1038/ismej.2008.50.
- Stapleton RD, Savage DC, Sayler GS, Stacey G (1998) Biodegradation of aromatic hydrocarbons in an extremely acidic environment. *Appl. Environ. Microbiol.* 64:4180-4184.
- Steen AD, Crits-Christoph A, Carini P, DeAngelis KM, Fierer N, Lloyd KG, Cameron Thrash J (2019) High proportions of bacteria and archaea across most biomes remain uncultured. *ISME J.* 13:3126-3130. doi:10.1038/s41396-019-0484-y.

- USEPA (2000) EPA/600/R-99/107: Introduction to Phytoremediation. United States Environmental Protection Agency.
- USEPA (2001) EPA 542-R-01-006: Brownfields Technology Primer: Selecting and Using Phytoremediation for Site Cleanup. United States Environmental Protection Agency.
- van Beilen JB, Funhoff EG, van Loon A, Just A, Kaysser L, Bouza M, Holtackers R, Röthlisberger M, Li Z, Witholt B (2006) Cytochrome P450 Alkane Hydroxylases of the CYP153 Family Are Common in Alkane-Degrading Eubacteria Lacking Integral Membrane Alkane Hydroxylases. *Appl. Environ. Microbiol.* 72:59. doi:10.1128/AEM.72.1.59-65.2006.
- van Beilen JB, Panke S, Lucchini S, Franchini AG, Röthlisberger M, Witholt B (2001) Analysis of *Pseudomonas putida* alkane-degradation gene clusters and flanking insertion sequences: evolution and regulation of the *alk* genes. *Microbiology* 147:1621-1630. doi:10.1099/00221287-147-6-1621.
- van Beilen JB, Wubbolts MG, Witholt B (1994) Genetics of alkane oxidation by *Pseudomonas oleovorans*. *Biodegradation* 5:161-174. doi:10.1007/BF00696457.
- Wemheuer F, Berkelmann D, Wemheuer B, Daniel R, Vidal S, Bisseleua Daghela HB (2020) Agroforestry management systems drive the composition, diversity, and function of fungal and bacterial endophyte communities in *Theobroma cacao* leaves. *Microorganisms* 8:405. doi:10.3390/microorganisms8030405.
- Werlen C, Kohler HP, van der Meer JR (1996) The broad substrate chlorobenzene dioxygenase and cis-chlorobenzene dihydrodiol dehydrogenase of *Pseudomonas* sp. strain P51 are linked evolutionarily to the enzymes for benzene and toluene degradation. *J. Biol. Chem.* 271:4009-16. doi:10.1074/jbc.271.8.4009.
- Williamson VM, Paquin CE (1987) Homology of *Saccharomyces cerevisiae* ADH4 to an iron-activated alcohol dehydrogenase from *Zymomonas mobilis*. *Mol. Gen. Genet.* 209:374-381. doi:10.1007/BF00329668.
- Woolfenden ENM, Hince G, Powell SM, Stark SC, Snape I, Stark JS, George SC (2011) The rate of removal and the compositional changes of diesel in Antarctic marine sediment. *Sci. Total Environ.* 410-411:205-216. doi:<https://doi.org/10.1016/j.scitotenv.2011.09.013>.
- Yokota T, Fuse H, Omori T, Minoda Y (1986) Microbial dehalogenation of haloalkanes mediated by oxygenase or halohydrolyase. *Agric. Biol. Chem.* 50:453-460. doi:10.1080/00021369.1986.10867402.
- Zylstra GJ, Gibson DT (1989) Toluene degradation by *Pseudomonas putida* F1. Nucleotide sequence of the *todC1C2BADE* genes and their expression in *Escherichia coli*. *J. Biol. Chem.* 264:14940-6.