Water Research 221 (2022) 118767

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

A metagenomic portrait of the microbial community responsible for two decades of bioremediation of poly-contaminated groundwater

Ernestina Hauptfeld [a], Jordi Pelkmans [b], Terry T. Huisman [a], Armin Anocic [a], Basten L. Snoek [a], F.A. Bastiaan von Meijenfeldt [a], Jan Gerritse [c], Johan van Leeuwen [g], Gert Leurink [d], Arie van Lit [e], Ruud van Uffelen [e], Margot C. Koster [b], Bas E. Dutilh [a,f,*]

[a] Theoretical Biology and Bioinformatics, Science for Life, Utrecht University, the Netherlands
[b] Department of Molecular Microbiology, Science for Life, Utrecht University, the Netherlands
[c] Deltares, Utrecht, the Netherlands
[d] City of Utrecht, the Netherlands
[e] Royal Haskoning DHV, the Netherlands
[f] Institute of Biodiversity, Faculty of Biological Sciences, Cluster of Excellence Balance of the Microverse, Friedrich-Schiller-University Jena, Germany
[g] KWR, Groningenhaven 7, 3433 PE, Nieuwegein, The Netherlands

ARTICLE INFO

Keywords:
Biodegradation
Bioremediation
Anaerobic microbiology
Groundwater contamination
BTEX
PAH

ABSTRACT

Biodegradation of pollutants is a sustainable and cost-effective solution to groundwater pollution. Here, we investigate microbial populations involved in biodegradation of poly-contaminants in a pipeline for heavily contaminated groundwater. Groundwater moves from a polluted park to a treatment plant, where an aerated bioreactor effectively removes the contaminants. While the biomass does not settle in the reactor, sediment is collected afterwards and used to seed the new polluted groundwater via a backwash cycle. The pipeline has successfully operated since 1999, but the biological components in the reactor and the contaminated park groundwater have never been described. We sampled seven points along the pipeline, representing the entire remediation process, and characterized the changing microbial communities using genome-resolved metagenomic analysis. We assembled 297 medium- and high-quality metagenome-assembled genome sequences representing on average 46.3% of the total DNA per sample. We found that the communities cluster into two distinct groups, separating the anaerobic communities in the park groundwater from the aerobic communities inside the plant. In the park, the community is dominated by members of the genus *Sulfuricurvum*, while the plant is dominated by generalists from the order *Burkholderiales*. Known aromatic compound biodegradation pathways are four times more abundant in the plant-side communities compared to the park-side. Our findings provide a genome-resolved portrait of the microbial community in a highly effective groundwater treatment system that has treated groundwater with a complex contamination profile for two decades.

1. Introduction

Around the world, soil and water pollution with hydrocarbons threaten food and drinking water safety (Rodríguez-Eugenio et al., 2018). Anthropogenic activities, such as oil or gas factories and city dump sites release toxic chemicals like benzene, toluene, ethylbenzene, xylenes (BTEX), and polycyclic aromatic hydrocarbons (PAHs) into the environment (Rodríguez-Eugenio et al., 2018). BTEX and PAH pollution in soil and groundwater limits plant growth and is toxic to animals (Salanitro et al., 1997). For humans, BTEX and PAHs are carcinogens, and exposure to these compounds can lead to reproductive problems, respiratory diseases, and cardiovascular diseases (Bolden et al., 2015). In recent years, microbial biodegradation of pollutants has been established as a sustainable and cost-effective strategy to remediate hydrocarbon-polluted terrestrial and aquatic environments (Guerra et al., 2018). Many microorganisms are known to degrade hydrocarbons (Chakraborty and Coates, 2004; Gallego et al., 2001; Head et al., 2006; Izmalkova et al., 2013; Lee et al., 2019; Liu et al., 2017; Meckenstock et al., 2004; Tian et al., 2002; Wang et al., 2020), and pathways involved in aerobic as well as anaerobic degradation have been identified (Dhar

* Corresponding author
E-mail address: bedutilh@gmail.com (B.E. Dutilh).

https://doi.org/10.1016/j.watres.2022.118767
Received 15 March 2022; Received in revised form 18 May 2022; Accepted 15 June 2022
Available online 16 June 2022

et al., 2020; Ghosal et al., 2016). However, complex mixtures and high concentrations of hydrocarbons, combined with limitations in nutrients and electron acceptors inhibit microbial growth and thus their ability to effectively remove contaminants. This leads to persistence of BTEX and PAHs, e.g. in groundwater and soil (Johnsen et al., 2005).

At sites where the native microbial community cannot effectively metabolize the contaminants, human assistance may be required to remediate the contaminated environment. Several strategies for assisted hydrocarbon bioremediation have been successfully applied. Biostimulation with surfactants increases the bioavailability of petroleum hydrocarbons, because it makes them more soluble and therefore less likely to stick to soil particles (Calvo et al., 2009; Xu et al., 2018). Biostimulation can also be performed with nutrients that facilitate the growth of biodegrading microbes (Powell et al., 2006; Röling et al., 2002). Bioaugmentation helps remove contaminants by introducing micro-organisms into the environment that have known biodegradation capabilities (Vogel, 1996; Wu et al., 2016). In bioreactor-based remediation, contaminated material is moved into a reactor with favorable conditions for biodegradation (such as a specific temperature or pH, or with stable aerobic conditions) and the contaminants are removed *ex situ* (Safdari et al., 2018). Bioremediation techniques have been shown to effectively remove hydrocarbons from groundwater and soil: natural attenuation (that is, unassisted bioremediation), biostimulation, and bioaugmentation effectively degraded diesel-oil in different soils from Long Beach and Hong Kong in a microcosm experiment (Bento et al., 2005). In another microcosm experiment, soil from different regions in China all showed strong biodegradation abilities for hydrocarbons from crude oil (Liu et al., 2017). Bacteria in soil from seven different regions in the US were all able to degrade >80% of crude-oil contamination, with different microbes responding to the contamination depending on soil properties (Hamamura et al., 2006).

The microbial response to hydrocarbon pollution can differ depending on the environment: contamination with hydrocarbons has been associated with an increase of members belonging to *Proteobacteria* compared to uncontaminated samples after an oil spill in a lake in Italy (D'Ugo et al., 2021). Specifically, enrichments of *Alpha-* and *Betaproteobacteria* were observed, both of which contain known hydrocarbon-degrading bacteria (D'Ugo et al., 2021). Beneath a former petrochemical plant in China, the most abundant microorganisms belonged to the genera *Amycolatopsis (Actinobacteria), Rhodocyclaceae (Betaproteobacteria), Sulfurimonas*, and *Sulfuricurvum* (both *Epsilonproteobacteria,* Sun et al., 2017). In formation waters from a Canadian oil sands reservoir, a massive dominance of *Epsilonbacteria*, particularly belonging to the genus *Sulfuricurvum* was observed (Hubert et al., 2012).

Metagenomic profiling of the microbiome, both in contaminated native environments and in bioremediation-based wastewater treatment systems, provides the basic understanding of the microbial organisms and their functional potential. This knowledge could contribute to improved bioremediation efforts by enabling knowledge transfer between contaminated sites. In this study, we use genome-resolved metagenomics to investigate the microbiome of a heavily hydrocarbon-contaminated city park in Utrecht, the Netherlands, as well as the highly effective bioremediation-based wastewater treatment system that removes the contaminants from the park groundwater (see Box 1 for details about the site). The bioremediation system has been in operation since 1999, continuously pumping water out of the park, to protect surrounding drinking water reserves that serve several towns in the province. The water is transported via an underground pipeline to a wastewater treatment plant where most of the pollutants are removed by microorganisms in an Anti-Bulking Reactor (ABR, Leurink et al., 2008). Although the bioremediation capacity of the microorganisms in the ABR has been known for a long time, their identity as well as the details of their biodegradation potential are still unknown.

To identify the changes that occur within the microbial community during transport and aeration and zoom in on the microbial hydrocarbon degradation process, we determined the taxonomic and functional composition of the microbes through genome-resolved metagenomics. We present detailed analyses of the microbiome at different locations within the bioremediation system (Fig. 2A), including from the groundwater directly emerging from the park and the water in the ABR. Our results show that the microbiomes form distinct groups outside and within the treatment plant and that each group is dominated by few, highly abundant taxa. For the first time, we provide insight into the genomic potential of the microbial communities in the transport pipeline and the Anti-Bulking Reactor.

2. Materials & Methods

2.1. The Griftpark Groundwater Treatment Pipeline

The Griftpark Groundwater Treatment Pipeline (Fig. 2A) has enabled effective removal of pollutants from groundwater since 1999 (Fig. 2C). Pollutant and geochemical measurements were carried out by the Department for City Development, Environment and Sustainability in Utrecht, independent of water sampling for DNA extraction. The entire treatment process takes 26.5 hours. Groundwater from the contaminated site is gathered in a collection basin, which is infused with small amounts of oxygen (p2) and pumped over 3km (3.5 hours) to a treatment plant. The groundwater enters via the influent buffer (p4), where it is inoculated with bacteria from the sludge thickener (p7), which is located at the end of the treatment plant. Then, the water proceeds to the aeration tank (p5), where it spends 13 hours while it is oxygenated and most of the contaminants are removed. After passing into an overflow (p6), the water enters the settling tank, where it separates from the sludge. The sludge phase is pumped into the sludge thickener (p7), while the water phase undergoes treatment via sand filters (optional) and activated carbon (optional) before entering the sludge thickener as well. Finally, a portion of the sludge from the sludge thickener is pumped back into the influent buffer, while the remainder is sufficiently remediated to allow discharge into the communal wastewater influent.

2.2. DNA isolation, sequencing, and quality control

Seven different points (Fig. 2) along the pipeline were sampled 1-5 times for a total of 24 metagenomic samples. Sampling dates were between March 22nd and August 29th 2016. Bacteria were obtained from 700 to 10 000ml of water per sample (Supplementary Table S1) by centrifugation and filtration. DNA was isolated using the MoBio PowerWater DNA Isolation Kit (MoBio Laboratories, Carlsbad, USA) followed by a phenol:chloroform extraction step to remove polluting residues (Sambrook et al., 1989). To increase the DNA yield of the samples taken in March and April, an amplification step was added using the DNA Nanokit for 6 cycles. All samples were sequenced on an Illumina HiSeq platform with a read length of 150bp. One sample from the aeration tank (p5) was additionally sequenced on the Oxford Nanopore Technologies MinION platform to enable hybrid assembly. Reads were filtered using bbduk from the BBTools v36.64 toolbox, and general quality control of sequencing reads was performed with FastQC v.0.11.5 (Andrews, 2010), statistics are provided in Supplementary Table S1. The datasets have been deposited in the sequence read archive (SRA) under Bioproject PRJNA816463.

2.3. Metagenome assembly and binning

The metagenomic datasets were assembled using short/long read hybrid *de novo* sequence assembly (Miller et al., 2017) with metaSPAdes v3.11.1 (Nurk et al., 2017), except 2-100817, which was assembled using only short reads, due to technical limitations. Assembly quality was assessed using QUAST v4.6.3 (Gurevich et al., 2013). Reads from each Illumina-sequenced metagenomic sample were mapped back to the assembled contigs using bwa-mem v0.7.12-r1039 (Li, 2013). The

Figure 1. Aerial view of the Griftpark in 1967 (panel A) and 2018 (panel B). The white outline indicates the area that is isolated by the underground bentonite wall.

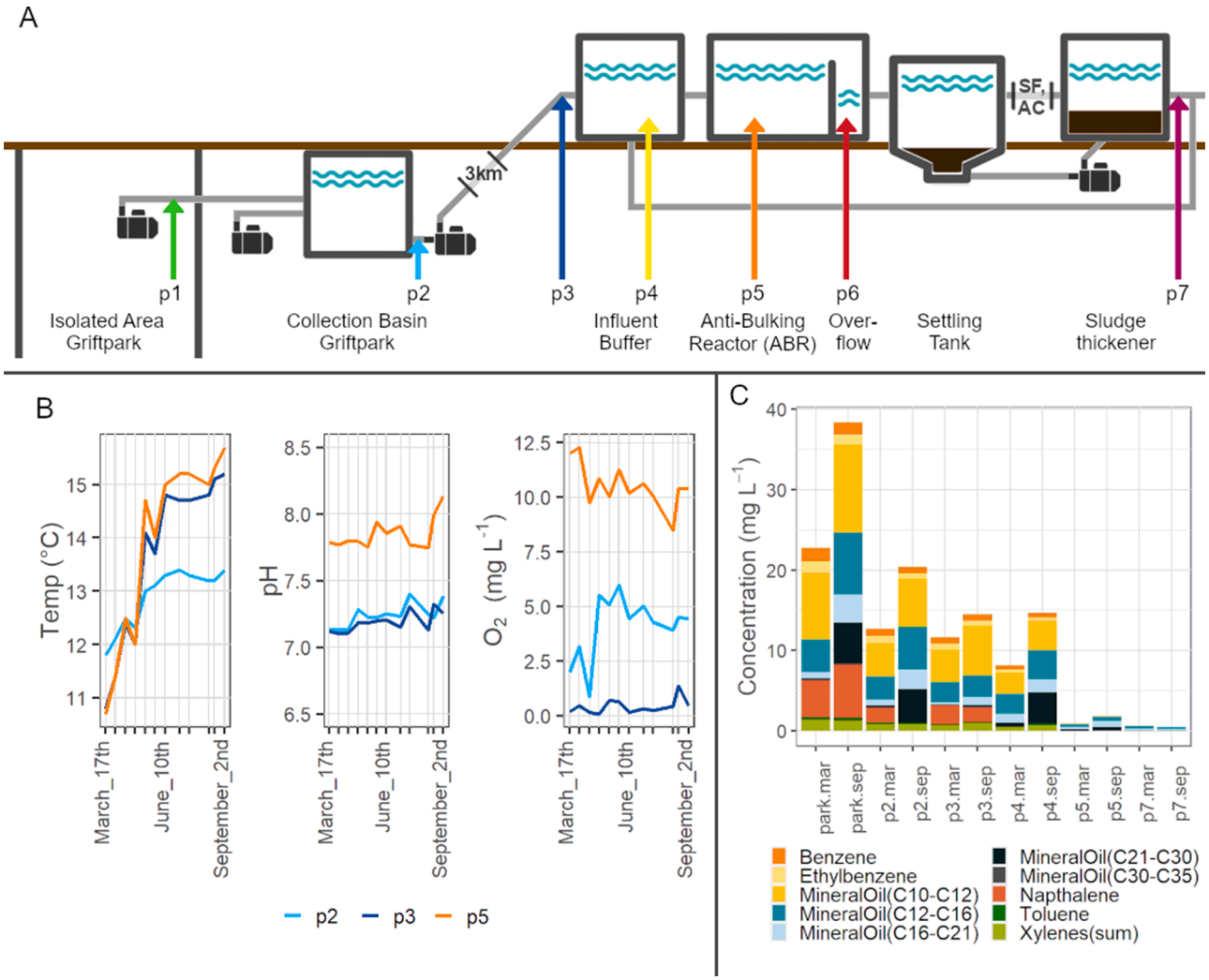

Figure 2. A. Schematic overview of the Griftpark Groundwater Treatment Pipeline. Seven sampled points are indicated with colored arrows. B. Temperature, pH, and oxygen content of the pipelines between March and September 2016 (data collected by municipality). C. Total pollutant concentration in the park groundwater and at sampling points p2, p3, p4, p5, and p7 on the 18th of March, and the 16th of September 2016, as provided by the municipality.

mapped reads made up 63-83.1% of reads per sample. Read alignments were sorted with samtools v1.7 (Li et al., 2009). The contigs were binned using metaBAT v2.12.1 (Lin and Liao, 2016) into 675 metagenome-assembled genomes (MAGs, see Supplementary Table S2) across all samples.

Completeness and contamination of MAGs was assessed with CheckM v.1.0.11 (Parks et al., 2015) and taxonomic classification performed with BAT v5.0.4 (Meijenfeldt et al., 2019). Phylogenetic analysis

Table 1
Maximum measured concentration of hydrocarbon compounds in the Griftpark soil. The table also lists the intervention value, i.e. the pollutant concentration at which intervention to remove the pollutant is required by Dutch law.

Compound	Concentration in mg kg^{-1} dry soil	Intervention value soil	Concentration in µg L^{-1} groundwater	Intervention value groundwater
PAHs (sum)	13 200	40	49 000	0.5 per PAH
BTEX (sum)	613	160.1	270 000	1 250
Benzene	280	1.1	100 000	30
Mineral Oil	9 900	5 000	220 000	600
Naphtalene	12 900	40	40 400	70
Phenols	300	14	14 400	1 500

of robust-quality MAGs (>50% complete, <10% contamination) was performed based on marker genes as identified by CheckM as follows. The MAGs of all samples were dereplicated using dRep (Olm et al., 2017). Closest known relatives of the dereplicated genomes were selected from the Genome Taxonomy Database r202 (GTDB, 34) by aligning the protein sequences of known marker gene to a set of GTDB proteins from representative genomes with diamond blastp v2.0.5 (Buchfink et al., 2015). For each robust-quality MAG, the top two reference genomes with the highest number of best hits over all the gene sets were selected as closest known relatives. The protein sequences were aligned with mafft v7.453 (Katoh and Standley, 2013) in e-ins-i mode. The resulted MSAs were trimmed using trimal v1.4rev15 (Capella-Gutiérrez et al., 2009) and alignment blocks concatenated. A phylogenetic tree was inferred using IQ-TREE v2.1.2 (Nguyen et al., 2015) and plotted with iTOL (Letunic and Bork, 2019). Terrabacteria was selected as outgroup for rooting the tree (Hug et al., 2016).

2.4. Microbial community profiling

All assembled contigs were classified using CAT v5.0.4 (Meijenfeldt et al., 2019) using the nr database version of 4th of March 2020 (Sayers et al., 2020). Relative abundance of taxa/MAGs was estimated per taxonomic rank by calculating the fraction of illumina reads that were mapped to contigs classified within each detected taxon/MAG. Shannon diversity was calculated using the diversity function of package *vegan* (Oksanen et al., 2019) in R v3.6.1 (R Core Team, n.d.). Multiple linear regression was performed using the lm function of the R standard library. Principle Component Analysis (PCA) was done with the R standard library and plotted with the packages *ggplot2* (Wickham, 2009) and *ggfortify* (Tang et al., 2016).

2.5. Analysis of anaerobic/aerobic marker genes

We analyzed the abundance of five different genes that are typical for aerobic/anaerobic metabolism: cytochrome-c-oxidase, ubiB, menaquinone biosynthesis decarboxylase, nrdD, and nrdG. We downloaded all homologues of the five aforementioned proteins from RefSeq, built a diamond database, and mapped the quality filtered reads using diamond blastx v2.0.5 (Buchfink et al., 2015). We then calculated the fraction of reads that mapped to each gene per sampling point.

2.6. Microbial co-response networks

Microbial co-response networks were constructed at genus rank by using the relative abundances of genera per sample. Genera with an abundance of >0.1% in at least two samples were selected (n=94). The Spearman rank correlation across samples was calculated using R (R Core Team, n.d.). From the correlation matrix pairs with a correlation >0.75, were selected for visualization in the co-response network. The network layout was made using the *igraph* R-package (Csardi and Nepusz, 2006) and visualized using *ggplot2* (Wickham, 2009). Correlations with environmental factors were calculated using the environmental measurements taken closest to the sampling point and date. Spearman correlation was calculated between the relative abundance of all selected genera and the environmental factors and visualized as node colors in the co-response networks.

2.7. Functional profiling

All proteins identified by Prodigal during CAT classification (see Microbial Community Profiling) were mapped to the UniRef50 database (Suzek et al., 2007) using DIAMOND v0.8.22 (Buchfink et al., 2015) where hits were defined as having >50% identity over >80% of the query. The hit with the highest bitscore was selected for queries with multiple hits. Metagenomic reads were mapped to a custom database containing these annotated proteins and the corresponding gene sequences with HUMAnN2 v0.11.2 (Franzosa et al., 2018). Pathways were annotated based on the MetaCyc pathway database (Caspi et al., 2019). Pathway abundance was normalized to pathway copies per million pathways (CoPM). All pathway abundances are provided in Supplementary Table S3.

The distribution of pathway reactions across MAGs was determined using the output of HUMAnN2. Contigs were translated and mapped against UniRef50 gene families by MetaPhlAn v2.1.0 (Truong et al., 2015). Based on the UniRef50 annotations of the contigs, reactions were annotated as being present or absent in each MAG (Supplementary Table S4).

3. Results and Discussion

We aimed to determine the identity and genomic potential of microorganisms from the contaminated groundwater beneath an urban recreational park and trace their abundance as the groundwater is pumped through a transport system and into an aerated bioreactor (Anti-Bulking Reactor, ABR) where the contaminants are removed. We

Box 1
The Griftpark in Utrecht, the Netherlands

The Griftpark (Fig. 1), today a popular urban recreational park, was a municipal gas plant that produced gas from coal between 1859 and 1960. In 1980, heavy soil and groundwater pollution with BTEX, PAHs, and crude oil was discovered at the plant's former dump site (see Table 1). To prevent lateral diffusion of the pollutants, an underground bentonite wall was built around the most heavily polluted area, 80cm thick and reaching on average 64m into the ground. Since 1999, groundwater is continuously pumped out of the park, keeping the local water level low and preventing dissolved contamination of the surrounding drinking water reserves. The water is transported to a local wastewater treatment plant (WTP) where the pollutants are biologically removed in an aerated Anti-bulking reactor (ABR, Fig. 2, Leurink et al., 2008). As the pipeline has never been augmented with external microorganisms, the bacteria responsible for biodegradation were named "Griftpark bacteria" and highlighted in the popular press over a decade ago ("Unieke bacterie peuzelt gif op in vervuild Griftpark," 2008). Still, their identity, as well as the details of their biodegradation potential have remained unknown.

performed a full metagenomic analysis charting the taxonomic and functional composition, reconstructed draft genome sequences of the most important players in the system and identified known biodegradation pathways. Over the course of six months, we sampled seven points of the groundwater treatment pipeline that represent the entire treatment process (Fig. 2A). Each point was sampled 1-5 times, adding up to a total of 24 samples. We extracted DNA from every sample and sequenced the metagenome, resulting in over 400 million reads (Supplementary Table S1). *De novo* assembly of the metagenomes yielded over 7.5 million contigs (Supplementary Table S1).

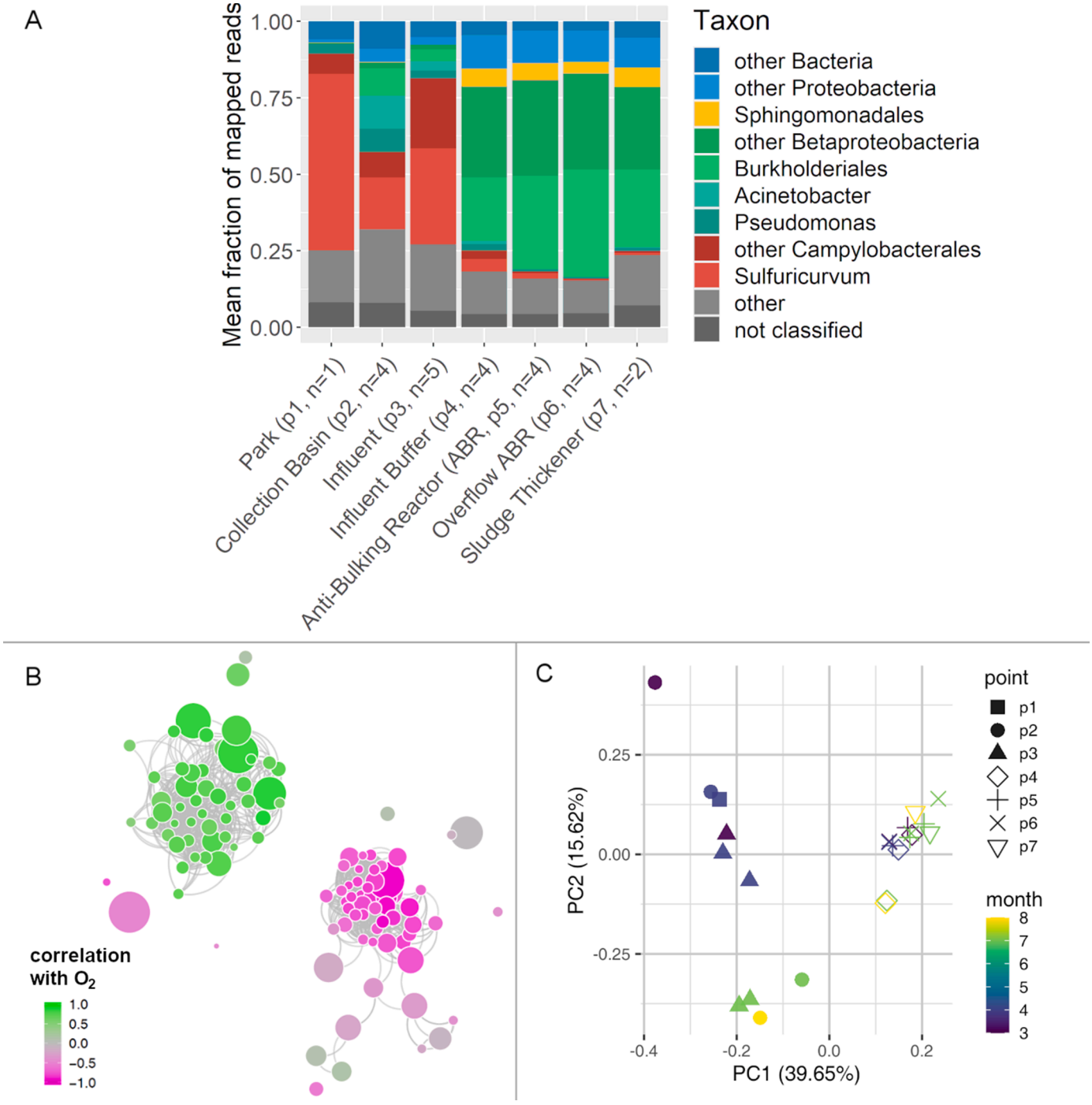

Figure 3. The microorganisms in the treatment pipeline cluster into two distinct groups. A. Relative abundance of the most abundant taxa in the treatment pipeline in mean fraction of mapped reads belonging to that taxon. B. Co-response network of the microbial communities. Nodes represent organisms with a relative abundance of at least 0.1% in a minimum of two samples (n=94). Edges connect organisms that have a Spearman correlation >=0.75. The nodes are colored according to the Spearman correlation of the taxa with oxygen concentration. The full network with organism annotation can be found in Supplementary Figure S3. Similar plots for other environmental variables including oxygen, pH, and pollutant concentrations are shown in Supplementary Figures S4-S18. C. Principal component analysis based on relative abundance of taxa in all 24 samples showing high similarity between plant-side communities (p4-p7) and structuring of the park-side communities by date along PC2.

3.1. The microbial community transforms as the water flows from the park into the WTP

To estimate the relative abundance of microbial taxa along the bioremediation pipeline, we determined the fraction of reads that mapped to contigs that were classified by CAT (Fig. 3A, Supplementary Table S5). In the Anti-Bulking Reactor (ABR), where most of the contamination is degraded (Fig. 2C), the microbial community was dominated by *Betaproteobacteria*, at the order level being represented by almost equal abundances of a diverse community of *Burkholderiales* (30.6% in the ABR) and unclassified *Betaproteobacteria* (31.2% in the ABR), most of which we also identified as belonging to the order *Burkholderiales* in our subsequent analysis (see below). The remaining ABR community consisted of members of *Proteobacteria* that could not be further classified (10.5%), *Sphingomonadales* (5.6%), and unclassified *Bacteria* (3.1%). *Burkholderiales* include strictly aerobic and facultative anaerobic chemoorganotrophs that are often capable of metabolizing aromatic compounds and are known as versatile degraders (Pérez-Pantoja et al., 2012). The most abundant *Burkholderiales* species in the ABR was *Macromonas bipunctata* (6.7%), a sulfur-oxidizing bacterium that so far has not been linked to biodegradation (Willems, 2014). *Sphingomonadales*, including the strictly aerobic genus *Novosphingobium,* have been shown to biodegrade PAHs (Baboshin et al., 2008; Colquhoun et al., 2012; Kertesz et al., 2019; Pinyakong et al., 2003).

The microbial community was markedly different in groundwater extracted from the park. Here, the community was dominated by *Epsilonproteobacteria* (Fig. 3A, 68.2% of mapped reads in p1), a class that has been shown to thrive in a Canadian oil sands reservoir (Hubert et al., 2012). The dominant genus in the park was *Sulfuricurvum* (57.5% in p1), that may occupy a similar niche as the cultured species, *Sulfuricurvum kujiense* (Kodama and Watanabe, 2004). *S. kujiense* is a facultatively anaerobic chemolithoautotrophic sulfur-oxidizing bacterium that was originally isolated from oil-contaminated groundwater from a crude-oil storage cavity in Japan. Other abundant microorganisms in the park belong to *Campylobacterales* (5.9%), unclassified *Bacteria* (4.5%, likely multiple bacteria), and *Pseudomonas* (3.2%), a versatile genus found in many environments (Gomila et al., 2015; Spiers et al., 2000) that has been associated to hydrocarbon-degradation including diesel-contaminated soils (Gallego et al., 2001; Izmalkova et al., 2013; Tian et al., 2002). While changes in the microbial community are expected along the treatment process (Tiwari et al., 2021), it is striking that only 0.38% of the community in the park belonged to *Betaproteobacteria*, which are massively enriched in the ABR. On average, 16.2% of reads from the ABR mapped to the assembly of the park microbiome, showing that the biodegradation community in the WTP dominates the bacteria derived from the park. We tested the abundance of a set of marker genes for aerobic metabolism: i) cytochrome-c-oxidase, which catalyzes the transport of electrons to molecular oxygen (Ludwig, 1987), mixed metabolism ii) ubiB, which produces ubiquinone, which is present in all bacteria, but the relative abundance of which should be lower in anaerobic environments compared to aerobic environments, as anaerobic gram-negative bacteria produce menaquinone as well (Collins and Jones, 1981), and anaerobic metabolism iii) menaquinone biosynthesis decarboxylase, which produces a quinone present in anaerobic gram-negative bacteria, and gram-positive bacteria (Collins and Jones, 1981) iv) nrdG and v) nrdD which catalyze ribonucleotide reduction in anaerobic growth (Garriga et al., 1996). The fraction of reads mapping to cytochrome-c-oxidase increased more than threefold from the park to the aeration tank (from $2.0*10^{-4}$ to $6.4*10^{-4}$), indicating the abundance of aerobic bacteria in the aeration tank (Supplementary Figure S1). NrdD decreased tenfold between the park and the aeration tank ($1.5*10^{-4}$ to $1.3*10^{-5}$), indicating an aerobic microbial community, while nrdD remained stagnant. Menaquinone biosynthesis decarboxylase and ubiB showed mirroring trends, with ubiB decreasing between the park and the influent (p3), and then increasing in the WTP. This also indicates an aerobic microbial community in the WTP.

To further investigate the taxonomic classification and interrelationships of the reconstructed MAGs, especially those that only had a classification at high taxonomic rank, we generated a phylogenetic tree based on 60 universal marker genes (Supplementary Figure S2). We dereplicated the 675 assembled MAGs, resulting in a non-redundant set of 57 high-quality MAGs. We chose the closest relatives of our MAGs from the Genome Taxonomy Database to provide context (Parks et al., 2018). Twelve of the dereplicated MAGs belong to the order *Burkholderiales*. Five MAGs were classified as *Betaproteobacteria* by BAT, but did not receive a robust classification at a lower taxonomic rank. In the phylogenetic tree, the same five genomes were placed within *Burkholderiales* (Supplementary Figure S2). *Burkholderiales* and *Sphingomonadales* genomes were generally more abundant in the WTP compared to the park, while *Sulfuricurvum* and *Pseudomonas* were more abundant in or close to the park. As the metagenomic datasets are compositional in nature we could not assess whether the bacteria from the park decreased in absolute abundance in the WTP (Gloor et al., 2017), i.e. which of them were able to grow under aeration.

3.2. Backwash cycle drives shift in microbial communities from park-side to the WTP

To explore the effect of oxygen on the microbial community, we built a microbial co-response network to reveal significant correlations in the abundance of different microbial taxa, and between the taxa and chemical parameters (Fig. 3B, Supplementary Figure S3). The network showed two distinct clusters of bacteria. The first cluster included *Alphaproteobacteria, Betaproteobacteria, Flavobacteria, Gammaproteobacteria (Xanthomonadales)*, and *Oligoflexia* that correlate positively with oxygen and were abundant in the WTP (WTP cluster, Fig. 3B). The second cluster included *Actinobacteria, Deltaproteobacteria, Epsilonproteobacteria, Erysipelotrichia*, and *Gammaproteobacteria (Pseudomonadales)* that correlated inversely with oxygen and were abundant in or close to the park (park-side cluster). The separation of the clusters was further supported by principal component analysis (Fig. 3C), where principal component PC1 correlated with the sampling point in the pipeline. Small amounts of oxygen were pumped into the collection basin at p2 by the municipality. This basin is never emptied completely and could therefore contain an adapted microbial community, explaining the increased abundance of *Pseudomonas* and *Acinetobacter*. In the WTP, the added oxygen in the ABR facilitated the change from an anaerobic to an aerobic microbial community.

Sampling points p3 and p4 are located within 20m of each other, yet p4 samples fell among the WTP communities in the PCA (Fig. 4C), showing that inoculation of sludge from the sludge thickener (p7) into the influent buffer (p4) via the backwash cycle (see Fig. 2A) is a major driver of the shift in microbial composition. To investigate the relative contribution of activated sludge and groundwater derived from the park to the p4 microbial community, we used multiple linear regression of the 98 taxa that contributed more than 0.1% of the community to estimate the relative abundance of taxa in p4 from their abundances in p3 and p7. Based on this analysis, we estimate that p7 contributes nine times as many microbes to the p4 community as p3 (89.9% and 10.1%, respectively, adjusted R^2 0.965), consistent with a much higher bacterial density in sludge (p7) than in the influent water (p3).

3.3. Diversity decreases as the microbes enter the treatment plant

The diversity of the microbial communities in the park-side is higher than in the WTP (mean species level Shannon diversity 3.04±0.41 and 2.7±0.45, respectively, see Supplementary Table S6). This higher diversity might explain (i) the shorter contigs in the park relative to the WTP (mean contig length 447nt and 615nt, Supplementary Table S1) although the number of sequencing reads was similar, (ii) the lower mean depth of contigs (3.24 and 3.65-fold coverage after read mapping),

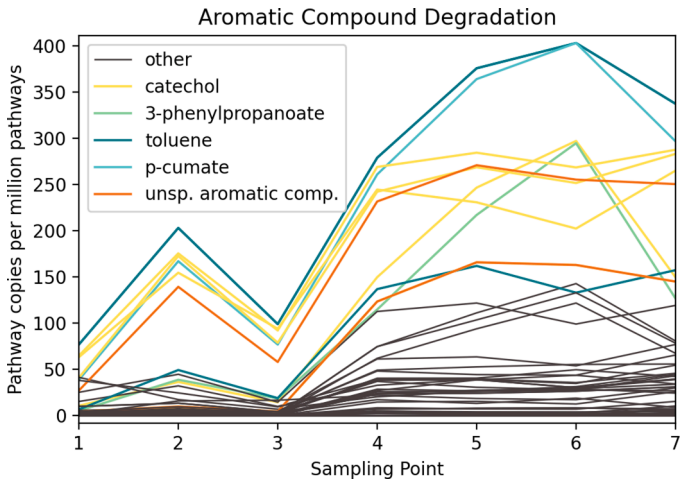

Figure 4. Abundance of MetaCyc Aromatic Compound Degradation pathways across the treatment pipeline. Each line represents the relative abundance of one pathway in copies per million pathways. HUMAnN2 assigns the relative abundance of a pathway as the relative abundance of the second least abundant gene in that pathway (i.e. one gene from the pathway can be missing or undetected).

(iii) the fact that a higher fraction of reads did not map to contigs (27.8 ±8.8% and 13.7±1.3%).

3.4. Known aerobic aromatic compound degradation pathways are more abundant in the treatment plant

To identify the shifts in functional potential of the microbial community, we mapped metagenomic reads to MetaCyc pathways. We observed a marked difference between the park- and WTP-side of the pipeline in the number of reads per sample that mapped to the proteins identified in that assembly (22.46-55.6% and 49.44-61.22% of reads, respectively). While many of the proteins encoded by the microorganisms in the treatment plant had homologs with well described functions, the organisms in the park microbiome remain under-studied (Meijenfeldt et al., 2019), resulting in a lower annotation rate.

To analyze the shifts in metabolic potential along the pipeline, we searched for known pathways using HUMAnN2 (Caspi et al., 2019). The direction and magnitude of the change in relative pathway abundance varied throughout the treatment pipeline and among different pathways (Fig. 4). For example, the sampling points at the beginning and the end of the 3km transport pipe (p2 and p3), the influent buffer (p4), and the sludge thickener (p7) showed the largest changes in pathway abundance relative to the previous points. Here, we address Aromatic Compound Degradation (ACD) pathways, information about all other pathways involved in e.g. Nucleotide/Amino Acid Metabolism and Energy/Metabolism can be found in Supplementary Table S3.

The relative abundance of biodegradation pathway genes, particularly all subcategories of ACD increased dramatically in the treatment plant (Fig. 4). As most of the well-described ACD pathways are aerobic, this increase in relative abundance is consistent with the oxygen availability in the ABR and the backwash cycle from p7 to p4 (Fig. 2) (Caspi et al., 2019; Weelink et al., 2010).

The most abundant MetaCyc ACD pathways identified are related to the degradation of toluene, catechol, p-cumate, 3-phenylpropanoate, and unspecified aromatic compounds (Supplementary Table S3). Catechol is a common intermediate metabolite in aromatic compound degradation (Surendra et al., 2017), which could imply that aromatic compounds from the park are degraded via catechol in the ABR. Pathways for degradation of toluene, catechol, p-cumate and unspecified aromatic compounds appear to increase at p2, then decrease from p2 to p3, before abruptly increasing at p4 (Fig. 4, Supplementary Table S3).

The peak in p2 could have two explanations. First, the water in the collection basin next to the park is a mixture of water from all around the park and not just from sampling point p1. Aromatic compound degradation pathways could be more abundant in other areas of the park, which would lead to a higher abundance of ACD pathways in the collection basin. Second, small amounts of oxygen are pumped into the collection basin, which could lead to an increase in the abundance of aerobic aromatic compound degradation pathways. Three and a half hours later, before the water enters the influent buffer in the treatment plant at p3, the oxygen is all but depleted (Fig. 2B), which is mirrored in the decreased relative abundance of degradation pathways at p3 (Fig. 5). Based on the decrease of ACD pathways in sampling point p3, we suggest that oxygen availability rather than mixture of communities drove the increase of ACD pathways in p2.

3.5. ACD pathways prevalent in key taxa of the plant side

Next, we investigated whether the MAGs that were enriched in the WTP preferentially contained ACD pathways. Of the 57 high-quality, non-redundant MAGs, 34 contained complete ACD pathways, 19 of which were enriched in the WTP compared to the park, while the remaining 15 MAGs where more abundant in the park. 14 of the 34 MAGs belonged to the class *Betaproteobacteria* (eleven *Burkholderiales*, three *Rhodocyclales*). Other MAGs that contained complete ACD pathways belonged to the orders *Campylobacterales* (4 *Sulfuricurvum*, 3 *Arcobacter*), *Pseudomonadales* (3 *Pseudomonas*, 2 *Acinetobacter*), *Desulfuromonadales* (1 *Geobacter*), and *Sphingomonadales* (1 unknown genus). MAGs that contain ACD pathways make up 24.2% of the community in the park, due to some *Sulfuricurvum* genomes containing a syringate degradation pathway. In the ABR, MAGs that contain ACD pathways add up to 42% of the community. These results suggest that the taxa enriched in the WTP (*Burkholderiales, Rhodocyclales, Sphingomonadales*) are responsible for the rapid degradation of the pollutants in the ABR. The only MAGs that were more abundant in the park-side and contained more than one ACD pathway belonged to *Pseudomonadales. Pseudomonadales* are most abundant in the collection basin, aerobic biodegradation is possible. The most ubiquitous ACD pathway among the park-enriched MAGs was the syringate degradation pathway. Syringate is an intermediate metabolite in the degradation of lignin, which is part of the woody plant cell wall and the most abundant aromatic compound in nature (Janusz et al., 2017). While lignin itself is not a toxic compound, it bears structural resemblance to some PAHs (Dhar et al., 2020), so these proteins annotated as belonging to the syringate degradation pathway may play a role in biodegradation. In *Sulfuricurvum*, which can thrive in hydrocarbon-polluted soil without breaking down the pollutants (Kodama and Watanabe, 2004), no known ACD pathways other than syringate degradation were found.

Fewer ACD pathways were found in the park-side bacteria compared to the bacteria in the WTP. However, the absence of known ACD pathways does not rule out biodegradation in the park groundwater. Among the many genes of unknown function that we recovered (92.3% in the park versus 86.3% in the WTP), undescribed anaerobic mechanisms for biodegradation of BTEX and PAH may exist. Currently, 179 out of the 194 ACD pathways in the MetaCyc database depend on oxygen, whereas the conditions in the Griftpark groundwater are mainly anaerobic. Anaerobic biodegradation pathways have been found for BTEX compounds (Weelink et al., 2010), but we found no known complete anaerobic biodegradation pathways in any of the MAGs in this study.

4. Conclusions and outlook

In this study, we investigated the microorganisms responsible for the biodegradation of aromatic compounds in the Griftpark groundwater treatment pipeline. The pipeline microbiome consists of two distinct groups of microorganisms, shifting from mainly anaerobic organisms in the park groundwater to a highly specialized aerobic community in the

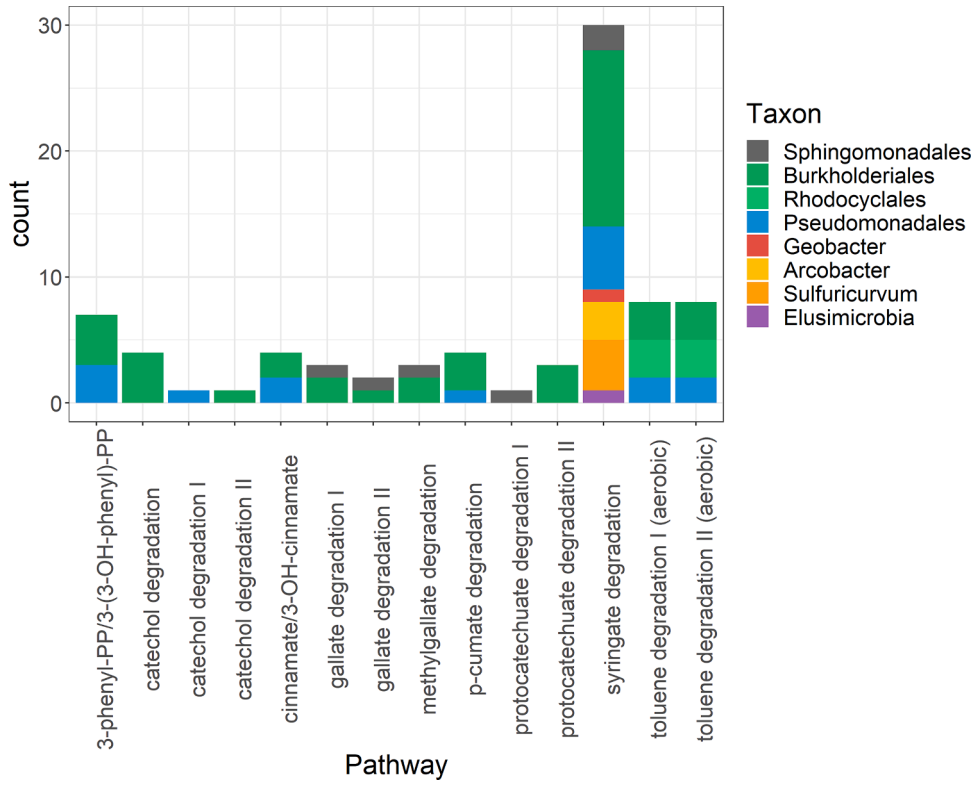

Figure 5. Number of complete MetaCyc pathways of the category Aromatic Compound Degradation found in the dereplicated MAGs. Note that some pathways overlap (e.g. catechol degradation pathways). If a microorganism contains the overlapping genes and the non-overlapping genes for two pathways, both pathways will be counted as present.

WTP. The microbial communities are dominated by a few highly abundant bacterial taxa, both in the park (*Sulfuricurvum*) and in the WTP (*Burkholderiales*). *Burkholderiales*, which account for only 0.3% of the community in the park, make up half of the community in the ABR and contain many known aromatic hydrocarbon degradation pathways. The highly specialized microbial communities in the ABR are enriched and maintained by a backwash cycle that inoculates the influent of the WTP with bacteria from the previous treatment cycle.

Our results provide insight into the biological processes involved in a highly effective groundwater bioremediation system from before treatment to the effluent of the pipeline. We show that *Burkholderiales*, already well-known degraders of various pollutants, are likely the driving force behind the degradation in the ABR as well. As the conditions in the ABR are aerobic, possible ACD pathways are much better described than for the anaerobic conditions in the park. Whether biodegradation of the pollutants is also happening in the anaerobic groundwater in the park is still an open question and will require further research. Due to the thorough sampling of the treatment process from beginning to end, our results improve the understanding of how the microbial community changes at each step in bioremediation. This additional insight will aid in the interpretation of biological processes active in other bioremediation pipelines as well as the design of new bioremediation pipelines.

Declaration of Competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Ernestina Hauptfeld: no competing interests.

Jordi Pelkmans: no competing interests.

Terry T. Huisman: no competing interests.

Armin Anocic: no competing interests.

Basten Snoek: no competing interests.

FA Bastiaan von Meijenfeldt: no competing interests.

Jan Gerritse reports financial support was provided by Deltares. Jan Gerritse reports a relationship with Deltares that includes: employment.

Johan van Leeuwen reports financial support was provided by Deltares. Johan van Leeuwen reports a relationship with Deltares that includes: former employment.

Gert Leurink reports financial support was provided by Municipality of Utrecht. Gert Leurink reports a relationship with Municipality of Utrecht that includes: employment.

Arie van Lit reports financial support was provided by Royal HaskoningDHV. Arie van Lit reports a relationship with Royal HaskoningDHV that includes: former employment.

Ruud van Uffelen reports financial support was provided by Royal HaskoningDHV. Ruud van Uffelen reports a relationship with Royal HaskoningDHV that includes: former employment.

Margot C. Koster: no competing interests.

Bas E. Dutilh: no competing interests.

Acknowledgements

This research was supported by the Netherlands Ministry for Infrastructure and Environment. EH and BED were supported by the European Research Council (ERC) Consolidator grant [grant number 865694, DiversiPHI]; Research was supported by the Enabling Technologies Hotels programme of ZonMw [grant number 40-43500-98-239] and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2051 – Project-ID 390713860. JG and JvL were supported by the municipality of Utrecht, The Netherlands [grant number 4281188/170821/1001-gl].

Supplementary materials

Supplementary material associated with this article can be found, in

the online version, at doi:10.1016/j.watres.2022.118767.

References

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.

Baboshin, M., Akimov, V., Baskunov, B., Born, T.L., Khan, S.U., Golovleva, L., 2008. Conversion of polycyclic aromatic hydrocarbons by Sphingomonas sp. VKM B-2434. Biodegradation 19, 567–576. https://doi.org/10.1007/s10532-007-9162-2.

Bento, F.M., Camargo, F.A.O., Okeke, B.C., Frankenberger, W.T., 2005. Comparative bioremediation of soils contaminated with diesel oil by natural attenuation, biostimulation and bioaugmentation. Bioresour. Technol. 96, 1049–1055. https://doi.org/10.1016/j.biortech.2004.09.008.

Bolden, A.L., Kwiatkowski, C.F., Colborn, T., 2015. New look at BTEX: Are ambient levels a problem. Environ. Sci. Technol. 49, 5261–5276. https://doi.org/10.1021/es505316f.

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

Calvo, C., Manzanera, M., Silva-Castro, G.A., Uad, I., González-López, J., 2009. Application of bioemulsifiers in soil oil bioremediation processes. Future prospects. Sci. Total Environ. 407, 3634–3640. https://doi.org/10.1016/j.scitotenv.2008.07.008.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Caspi, R., Billington, R., Keseler, I.M., Kothari, A., Krummenacker, M., Midford, P.E., Ong, W.K., Paley, S., Subhraveti, P., Karp, P.D., 2019. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. Nucleic Acids Res 48, D445–D453. https://doi.org/10.1093/nar/gkz862.

Chakraborty, R., Coates, J.D., 2004. Anaerobic degradation of monoaromatic hydrocarbons. Appl. Microbiol. Biotechnol. 64, 437–446. https://doi.org/10.1007/s00253-003-1526-x.

Collins, M.D., Jones, D., 1981. Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implications. Microbiol. Rev. 45, 316–354. https://doi.org/10.1128/mmbr.45.2.316-354.1981.

Colquhoun, D.R., Hartmann, E.M., Halden, R.U., 2012. Proteomic profiling of the dioxin-degrading bacterium Sphingomonas wittichii RW1. J. Biomed. Biotechnol. 2012 408690. https://doi.org/10.1155/2012/408690.

Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research. InterJournal Complex Sy 1695.

D'Ugo, E., Bruno, M., Mukherjee, A., Chattopadhyay, D., Giuseppetti, R., De Pace, R., Magurano, F., 2021. Characterization of microbial response to petroleum hydrocarbon contamination in a lacustrine ecosystem. Environ. Sci. Pollut. Res. Int. 28, 26187–26196. https://doi.org/10.1007/s11356-021-13885-8.

Dhar, K., Subashchandrabose, S.R., Venkateswarlu, K., Krishnan, K., Megharaj, M., 2020. Anaerobic Microbial Degradation of Polycyclic Aromatic Hydrocarbons: A Comprehensive Review, in: de Voogt, P. (Ed.), Reviews of Environmental Contamination and Toxicology Volume 251. Springer International Publishing, Cham, pp. 25–108. https://doi.org/10.1007/398_2019_29.

Franzosa, E.A., McIver, L.J., Rahnavard, G., Thompson, L.R., Schirmer, M., Weingart, G., Lipson, K.S., Knight, R., Caporaso, J.G., Segata, N., Huttenhower, C., 2018. Species-level functional profiling of metagenomes and metatranscriptomes. Nat. Methods 15, 962–968. https://doi.org/10.1038/s41592-018-0176-y.

Gallego, J.L.R., Loredo, J., Llamas, J.F., Vázquez, F., Sánchez, J., 2001. Bioremediation of diesel-contaminated soils: Evaluation of potential in situ techniques by study of bacterial degradation. Biodegradation 12, 325–335. https://doi.org/10.1023/A:1014397732435.

Garriga, X., Eliasson, R., Torrents, E., Jordan, A., Barbe, J., Gibert, I., Reichard, P., 1996. nrdD and nrdG Genes Are Essential for Strict Anaerobic Growth of Escherichia coli Escherichia coli contains the genetic information for three separate ribonucleotide reduc- tases . The nrdA and nrdB genes code for the enzyme that provides deoxyribonucleot. Biochem. Biophys. Res. Commun. 192, 189–192.

Ghosal, D., Ghosh, S., Dutta, T.K., Ahn, Y., 2016. Current State of Knowledge in Microbial Degradation of Polycyclic Aromatic Hydrocarbons (PAHs): A. Review. Front. Microbiol. 7, 1369. https://doi.org/10.3389/fmicb.2016.01369.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., Egozcue, J.J., 2017. Microbiome datasets are compositional: And this is not optional. Front. Microbiol. 8, 1–6. https://doi.org/10.3389/fmicb.2017.02224.

Gomila, M., Peña, A., Mulet, M., Lalucat, J., García-Valdés, E., 2015. Phylogenomics and systematics in Pseudomonas. Front. Microbiol.

Guerra, A.B., Oliveira, J.S., Silva-Portela, R.C.B., Araújo, W., Carlos, A.C., Vasconcelos, A.T.R., Freitas, A.T., Domingos, Y.S., de Farias, M.F., Fernandes, G.J.T., Agnez-Lima, L.F., 2018. Metagenome enrichment approach used for selection of oil-degrading bacteria consortia for drill cutting residue bioremediation. Environ. Pollut. 235, 869–880. https://doi.org/10.1016/j.envpol.2018.01.014.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

Hamamura, N., Olson, S.H., Ward, D.M., Inskeep, W.P., 2006. Microbial population dynamics associated with crude-oil biodegradation in diverse soils. Appl. Environ. Microbiol. 72, 6316–6324. https://doi.org/10.1128/AEM.01015-06.

Head, I.M., Jones, D.M., Röling, W.F.M., 2006. Marine microorganisms make a meal of oil. Nat. Rev. Microbiol. 4, 173–182. https://doi.org/10.1038/nrmicro1348.

Hubert, C.R.J., Oldenburg, T.B.P., Fustic, M., Gray, N.D., Larter, S.R., Penn, K., Rowan, A. K., Seshadri, R., Sherry, A., Swainsbury, R., Voordouw, G., Voordouw, J.K., Head, I. M., 2012. Massive dominance of Epsilonproteobacteria in formation waters from a Canadian oil sands reservoir containing severely biodegraded oil. Environ. Microbiol. 14, 387–404. https://doi.org/10.1111/j.1462-2920.2011.02521.x.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D.A., Finstad, K.M., Amundson, R., Thomas, B.C., Banfield, J.F., 2016. A new view of the tree of life. Nat. Microbiol. 1, 16048. https://doi.org/10.1038/nmicrobiol.2016.48.

Izmalkova, T.Y., Sazonova, O.I., Nagornih, M.O., Sokolov, S.L., Kosheleva, I.A., Boronin, A.M., 2013. The organization of naphthalene degradation genes in Pseudomonas putida strain AK5. Res. Microbiol. 164, 244–253. https://doi.org/10.1016/j.resmic.2012.12.007.

Janusz, G., Pawlik, A., Sulej, J., Swiderska-burek, U., Jarosz-wilkołazka, A., Paszczy, A., 2017. Lignin degradation : microorganisms, enzymes involved, genomes analysis and evolution 941–962. https://doi.org/10.1093/femsre/fux049.

Johnsen, A.R., Wick, L.Y., Harms, H., 2005. Principles of microbial PAH-degradation in soil. Environ. Pollut. 133, 71–84. https://doi.org/10.1016/j.envpol.2004.04.015.

Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.

Kertesz, M.A., Kawasaki, A., Stolz, A., 2019. Aerobic Hydrocarbon-Degrading Alphaproteobacteria : Sphingomonadales 105–124.

Kodama, Y., Watanabe, K., 2004. Sulfuricurvum kujiense gen. nov., sp. nov., a facultatively anaerobic, chemolithoautotrophic, sulfur-oxidizing bacterium isolated from an underground crude-oil storage cavity. Int. J. Syst. Evol. Microbiol. 54, 2297–2300. https://doi.org/10.1099/ijs.0.63243-0.

Lee, Yunho, Lee, Yunhee, Jeon, C.O., 2019. Biodegradation of naphthalene, BTEX, and aliphatic hydrocarbons by Paraburkholderia aromaticivorans BN5 isolated from petroleum-contaminated soil. Sci. Rep. 9, 860. https://doi.org/10.1038/s41598-018-36165-x.

Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47, W256–W259. https://doi.org/10.1093/nar/gkz239.

Leurink, G., Lit, A.van, Moolenaar, S., Manders, B., 2008. Anti Bulking Reactor reinigt grondwater biologisch tegen lage kosten. H 2 O 41 32.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Lin, H.-H., Liao, Y.-C., 2016. Accurate binning of metagenomic contigs via automated clustering sequences using information of genomic signatures and marker genes. Sci. Rep. 6 https://doi.org/10.1038/srep24175.

Liu, Q., Tang, J., Gao, K., Gurav, R., Giesy, J.P., 2017. Aerobic degradation of crude oil by microorganisms in soils from four geographic regions of China. Sci. Rep. 7, 14856. https://doi.org/10.1038/s41598-017-14032-5.

Ludwig, B., 1987. Cytochrome c oxidase in prokaryotes. FEMS Microbiol. Lett. 46, 41–56. https://doi.org/10.1016/0378-1097(87)90185-6.

Meckenstock, R.U., Safinowski, M., Griebler, C., 2004. Anaerobic degradation of polycyclic aromatic hydrocarbons. FEMS Microbiol. Ecol 49, 27–36. https://doi.org/10.1016/j.femsec.2004.02.019.

Meijenfeldt, F.A.B.Von, Arkhipova, K., Cambuy, D.D., Coutinho, F.H., Dutilh, B.E., 2019. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. Genome Biol 20, 217.

Miller, J.R., Zhou, P., Mudge, J., Gurtowski, J., Lee, H., Ramaraj, T., Walenz, B.P., Liu, J., Stupar, R.M., Denny, R., Song, L., Singh, N., Maron, L.G., McCouch, S.R., McCombie, W.R., Schatz, M.C., Tiffin, P., Young, N.D., Silverstein, K.A.T., 2017. Hybrid assembly with long and short reads improves discovery of gene family expansions. BMC Genomics 18, 1–12. https://doi.org/10.1186/s12864-017-3927-8.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300.

Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., 2017. MetaSPAdes: A new versatile metagenomic assembler. Genome Res 27. https://doi.org/10.1101/gr.213959.116.

Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2019. Community Ecology Package.

Olm, M.R., Brown, C.T., Brooks, B., Banfield, J.F., 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J 11, 2864–2868. https://doi.org/10.1038/ismej.2017.126.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., Hugenholtz, P., 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. 36, 996–1004. https://doi.org/10.1038/nbt.4229.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25, 1043–1055. https://doi.org/10.1101/gr.186072.114.

Pérez-Pantoja, D., Donoso, R., Agulló, L., Córdova, M., Seeger, M., Pieper, D.H., González, B., 2012. Genomic analysis of the potential for aromatic compounds biodegradation in Burkholderiales. Environ. Microbiol. 14, 1091–1117. https://doi.org/10.1111/j.1462-2920.2011.02613.x.

Pinyakong, O., Habe, H., Yoshida, T., Nojiri, H., Omori, T., 2003. Identification of three novel salicylate 1-hydroxylases involved in the phenanthrene degradation of

Sphingobium sp. strain P2. Biochem. Biophys. Res. Commun. 301, 350–357. https://doi.org/10.1016/s0006-291x(02)03036-x.
Powell, S.M., Ferguson, S.H., Snape, I., Siciliano, S.D., 2006. Fertilization stimulates anaerobic fuel degradation of antarctic soils by denitrifying microorganisms. Environ. Sci. Technol. 40, 2011–2017. https://doi.org/10.1021/es051818t.
R Core Team, 2022. R: A Language and Environment for Statistical Computing.
Rodríguez-Eugenio, N., McLaughlin, M., Pennock, D., 2018. Soil Pollution: A hidden reality. Rome.
Röling, W.F.M., Milner, M.G., Jones, D.M., Lee, K., Daniel, F., Swannell, R.J.P., Head, I. M., 2002. Robust hydrocarbon degradation and dynamics of bacterial communities during nutrient-enhanced oil spill bioremediation. Appl. Environ. Microbiol. 68, 5537–5548. https://doi.org/10.1128/aem.68.11.5537-5548.2002.
Safdari, M.-S., Kariminia, H.-R., Rahmati, M., Fazlollahi, F., Polasko, A., Mahendra, S., Wilding, W.V., Fletcher, T.H., 2018. Development of bioreactors for comparative study of natural attenuation, biostimulation, and bioaugmentation of petroleum-hydrocarbon contaminated soil. J. Hazard. Mater. 342, 270–278. https://doi.org/10.1016/j.jhazmat.2017.08.044.
Salanitro, J.P., Dorn, P.B., Huesemann, M.H., Moore, K.O., Rhodes, I.A., Rice Jackson, L. M., Vipond, T.E., Western, M.M., Wisniewski, H.L., 1997. Crude oil hydrocarbon bioremediation and soil ecotoxicity assessment. Environ. Sci. Technol. 31, 1769–1776. https://doi.org/10.1021/es960793i.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular cloning: a laboratory manual. Cold spring harbor laboratory press.
Sayers, E.W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K.D., Karsch-Mizrachi, I., 2020. GenBank. Nucleic Acids Res. 48, D84–D86. https://doi.org/10.1093/nar/gkz956.
Spiers, A.J., Buckling, A., Rainey, P.B., 2000. The causes of Pseudomonas diversity. Microbiology 146, 2345–2350. https://doi.org/10.1099/00221287-146-10-2345.
Sun, Y., Lu, S., Zhao, X., Ding, A., Wang, L., 2017. Long-Term Oil Pollution and In Situ Microbial Response of Groundwater in Northwest China. Arch. Environ. Contam. Toxicol. 72, 519–529. https://doi.org/10.1007/s00244-017-0405-x.
Surendra, V., Mahalingam, B.L., Velan, M., 2017. Degradation of monoaromatics by Bacillus pumilus MVSV 3 Sheeba.
Suzek, B.E., Huang, H., McGarvey, P., Mazumder, R., Wu, C.H., 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics 23, 1282–1288. https://doi.org/10.1093/bioinformatics/btm098.
Tang, Y., Horikoshi, M., Li, W., 2016. ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages. R J. 8.
Tian, L., Ma, P., Zhong, J.-J., 2002. Kinetics and key enzyme activities of phenanthrene degradation by Pseudomonas mendocina. Process Biochem 37, 1431–1437. https://doi.org/10.1016/S0032-9592(02)00032-8.
Tiwari, A., Hokajärvi, A.M., Domingo, J.S., Elk, M., Jayaprakash, B., Ryu, H., Siponen, S., Vepsäläinen, A., Kauppinen, A., Puurunen, O., Artimo, A., Perkola, N., Huttula, T., Miettinen, I.T., Pitkänen, T., 2021. Bacterial diversity and predicted enzymatic function in a multipurpose surface water system – from wastewater effluent discharges to drinking water production. Environ. Microbiomes 16, 1–17. https://doi.org/10.1186/s40793-021-00379-w.
Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., Segata, N., 2015. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat. Methods 12, 902–903. https://doi.org/10.1038/nmeth.3589.
Unieke bacterie peuzelt gif op in vervuild Griftpark, 2008. Trouw.
Vogel, T.M., 1996. Bioaugmentation as a soil bioremediation approach. Curr. Opin. Biotechnol. 7, 311–316. https://doi.org/10.1016/S0958-1669(96)80036-X.
Wang, W., Li, Z., Zeng, L., Dong, C., Shao, Z., 2020. The oxidation of hydrocarbons by diverse heterotrophic and mixotrophic bacteria that inhabit deep-sea hydrothermal ecosystems. ISME J. https://doi.org/10.1038/s41396-020-0662-y.
Weelink, S.A.B., van Eekert, M.H.A., Stams, A.J.M., 2010. Degradation of BTEX by anaerobic bacteria: physiology and application. Rev. Environ. Sci. Bio/Technology 9, 359–385. https://doi.org/10.1007/s11157-010-9219-2.
Wickham, H., 2009. Use R! ggplot2 - Elegant Graphics for Data Analysis, Media. https://doi.org/10.1007/978-0-387-78171-6.
Willems, A., 2014. The Family Comamonadaceae BT - The Prokaryotes: Alphaproteobacteria and Betaproteobacteria, in: Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F. (Eds.), Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 777–851. https://doi.org/10.1007/978-3-642-30197-1_238.
Wu, M., Dick, W.A., Li, W., Wang, X., Yang, Q., Wang, T., Xu, L., Zhang, M., Chen, L., 2016. Bioaugmentation and biostimulation of hydrocarbon degradation and the microbial community in a petroleum-contaminated soil. Int. Biodeterior. Biodegradation 107, 158–164. https://doi.org/10.1016/j.ibiod.2015.11.019.
Xu, R., Zhang, Z., Wang, L., Yin, N., Zhan, X., 2018. Surfactant-enhanced biodegradation of crude oil by mixed bacterial consortium in contaminated soil. Environ. Sci. Pollut. Res. 25, 14437–14446. https://doi.org/10.1007/s11356-018-1604-3.