Environmental Science
Processes & Impacts

Microbial Genetic Potential for Xenobiotic Metabolism Increases with Depth During Biofiltration

Journal:	*Environmental Science: Processes & Impacts*
Manuscript ID	EM-ART-06-2020-000254.R1
Article Type:	Paper

Environmental Significance Statement

Managed aquifer recharge has been used globally to augment water supplies. In addition to storage, the infiltration process improves water quality, which can be particularly important when impaired waters such as stormwater runoff or reclaimed water are used. Past research suggests that biotransformation rates of a wide variety of trace organic chemicals may be enhanced in deeper regions when contrasted with shallower saturated zones. Here, we conduct a targeted exploration of a subsurface microbiome that has been exposed to organic substrate limited conditions. Despite lower overall biomass, these zones contain genetic signatures consistent with enhanced trace organic metabolism. Hence, both infiltration depth and influent properties should be considered to optimize the biodegradation of trace organics during managed aquifer recharge.

Microbial Genetic Potential for Xenobiotic Metabolism Increases with Depth During Biofiltration

Dong Li[1], Jonathan O. Sharp[1,3], Jörg E. Drewes[1,2]*

[1]NSF Engineering Research Center *ReNUWIt*, Department of Civil and Environmental Engineering, Colorado School of Mines, Golden, CO, USA

[2]Chair of Urban Water Systems Engineering, Technische Universität München, Garching, Germany

[3]Hydrologic Science and Engineering Program, Colorado School of Mines

*Corresponding author:

JE Drewes, Chair of Urban Water Systems Engineering, Technische Universität München, Am Coulombwall 8, 85748 Garching, Germany

E-mail: jdrewes@tum.de;

Abstract

Water infiltration into the subsurface can result in pronounced biogeochemical depth gradients. In this study, we assess metabolic potential and properties of the subsurface microbiome during water infiltration by analyzing sediments from spatially-segmented columns. Past work in these laboratory constructs demonstrated that removal efficiencies of trace organic pollutants were enhanced by limited availability of biodegradable dissolved organic carbon (BDOC) associated with higher humic ratios and deeper sediment regions. Distinct differences were observed in the microbial community when contrasting shallow versus deeper profile sediments. Metagenomic analyses revealed that shallow sediments contained an enriched potential for bacterial growth and division processes. In contrast, deeper sediments harbored a significant increase in genes associated with the metabolism of secondary metabolites and the biotransformation of xenobiotic water pollutants. Metatranscripts further supported this trend, with increased potential for metabolic attributes associated with the biotransformation of xenobiotics and antibiotic resistance within deeper sediments. Furthermore, increasing ratios of humics in feed solutions correlated to enhanced expression of genes associated with xenobiotic biodegradation. These results provide genetic support for the interplay of dissolved organic carbon limitation and enhanced trace organic biotransformation by the subsurface microbiome.

Keyword: Managed aquifer recharge; Microbiology; Trace organics; Emerging contaminants; Biotransformation

Introduction

The removal of dissolved organic matter (DOM) by indigenous microorganisms in the subsurface has been utilized as a component of water treatment in managed natural systems and engineered analogs including riverbank filtration, soil-aquifer treatment, and aquifer recharge and recovery.[1] These passive systems, collectively named managed aquifer recharge, have advantages when contrasted with more conventional treatment and storage approaches with respect to energy demand, localized recycling, and limiting aquifer compaction in arid and semiarid regions. They have been used in Europe and North America for decades to augment local water supplies. Impaired waters such as stormwater runoff or reclaimed water are promising candidates for these natural water treatment systems.

DOM is typically degraded or assimilated by indigenous microorganisms present in the subsurface during water infiltration.[2] This assimilation largely occurs in the shallow sediments as typified by DOM concentrations that decline steeply with depth. However, refractory organic compounds including emerging trace organic chemicals such as pharmaceuticals, personal care products, household chemicals, endocrine disruptors, and pesticides are recalcitrant and transformed slowly in saturated infiltration zones with transport ranging from meters to kilometers from discharge localities.[3] The persistence of these trace organic chemicals has raised concern due to their potential impact on human health when groundwater is used as drinking water sources,[4-5] and mandates a better understanding of their fate and transport as the microbial processes involved in attenuation are not well understood.[6] Reliable

enhanced removal of these emerging trace organic chemicals during infiltration could lead to more effective utilization of subsurface systems for both treatment and storage in managed aquifer recharge systems.[7]

Microorganisms residing in saturated regions during managed aquifer recharge may be spatially dynamic because of dissolved organic carbon (DOC), nutrient and redox profiles. However, our understanding of microbial community characteristics and responses with depth during infiltration is limited. Because of field access limitations, many prior studies have focused on shallow sediments[8-10] with less understanding about microbial characteristics within deeper infiltration zones.[11-15] Compared to shallow sediments, there is a lower density of microorganisms residing in the deeper reaches of infiltration zones.[11] Although prior work has documented spatial shifts in microbial communities associated with organic loading and consumption,[13,16] details of the metabolic functionality of these microbial communities within these deeper regions is not well understood. Functional results in tracking contaminant attenuation suggest that biotransformation processes of trace organic chemicals are enhanced in these deeper regions when contrasted with adjacent shallower zones in managed aquifer systems.[17] To this end, we hypothesized that the subsurface microbiome in the deeper portions of an infiltration regime differs from those found in shallower sediments with genetic signatures that correlate to enhanced potential for trace organic metabolism.

To investigate the genetic potential for metabolic shifts during water infiltration, we utilized a series of laboratory sediment columns that simulated subsurface

infiltration gradients. Analogous laboratory experiments have been used to simulate water infiltration in sediments to better understand intertwined hydrological process, sorption, bulk organic and trace organic chemical attenuation mechanisms, as well as microbial community phylogenetic structure.[17-19] Feed water with various ratios of refractory to labile organics were introduced into of the different column systems over a 12-month study period. This resulted in depth gradients that in turn created distinct eutrophic and oligotrophic zones. Prior work established enhanced trace organic chemical degradation in the more oligotrophic zones within these columns.[17] To better understand how the subsurface microbiome correlated with these functional attributes, we undertook a multipronged approach that quantified changes in microbial biomass, phylogenetic composition, and metabolic potential by using metagenomic and metatranscriptomic analyses of depth resolved biofilms. These microbial queries were then coupled to past reports of trace organic chemical removal to provide tools and insights into the optimization of trace organic chemical degradation during managed aquifer recharge.

Materials and methods

Laboratory-Scale Sediment Columns

Four laboratory-scale sediment column systems were established using homogenized sediments (sieve fraction 0.2-2 mm) derived from a field riverbed site near Taif, Saudi Arabia.[17] Each of the four column systems was further comprised of four (30×4 cm

I.D.) glass chromatography columns (Kimble Chase Kontes) linked in series with Viton tubing (Masterflex) as shown in the experimental blueprint art (Table of contents entry). Hence, each column system contained an overall infiltration zone of 120 cm that could be sampled at 30 cm gradients. The sediments were characterized as sand (94.8%) with small fractions of gravel, silt and clay, a low organic carbon content (f_{oc}=0.10±0.01%), and a porosity of 0.32±0.03. Columns were operated under saturated up-flow conditions at ambient room temperature (around 20 ºC) with a hydraulic loading rate of 1.44 m d^{-1} and wrapped with aluminum foil to minimize light exposure. The hydraulic retention time was determined to be approximately 16 hours for each system (120 cm transect) using a conservative tracer (KBr). More details about the column design and operation can be found in Alidina *et al*.[17]

To evaluate the influence of organic matter composition on microorganisms, the column arrays were exposed to synthetic feed solutions with the same approximate concentration of biodegradable dissolved organic carbon (BDOC) but containing differing ratios of more recalcitrant humics versus more bioavailable substrates. Two Ismatec IPC8 8-channel peristaltic pumps (Ismatec, Switzerland) were utilized to continuously deliver the feed solution. The feed solution consisted of peptone, yeast extract and humic acid based on a modified OECD recipe.[17] Peptone (BD Difco) and yeast extract (BD Difco) representing readily biodegradable dissolved organic matter were mixed in a ratio of 2:1 (w/w), and then blended with humic acid sodium salt (Sigma-Aldrich) representing a more refractory carbon source. These were introduced at ratios of 100:0, 60:40, 40:60, and 0:100 (peptone and yeast extract versus humic

acid in w/w) to the four different column systems (C1-C4), respectively. A target of approximately 1.5 mg/L BDOC in the influent (determined by calculating the DOC consumed in the column) was iteratively maintained across the experimental column systems by introducing different quantities of absolute DOC. Other salt and trace metal components of the feed solution mimicking water quality in soil pore space have been described before.[17]

To monitor the functional attributes of these columns after establishment, 18 different trace organic chemicals (listed in Table 1) known to be normally present in surface or impaired waters were spiked into the feed solution at environmentally relevant concentrations (300-500 ng/L) and fed into column systems during the whole operating period. Analysis of trace organic chemical behavior across and within these columns and associated methods has been reported previously.[17] Adsorption of trace organic chemicals in these sediments was assessed and determined by a short-term biologically repressed control conducted in a separate sediment column (30 cm length). For this control, riverbed sediments were conditioned with the 40:60 influent followed by 2 mM sodium azide to inhibit respiration.[17] Water quality in column effluents approached stability within one month of operation as assessed by DOC and nitrate concentrations. Influent and effluent water quality parameters including DOC, nitrate, UV absorbance, pH, and others are listed in Table S1. Nitrate was measured periodically and did not change significantly during the study period.

Phylogenetic and Biomass Analyses

5-10 grams of column sediments were harvested from column interfacial depths of 1 cm, 30 cm, 60 cm, 90 cm and 120 cm using sterile sampling spoons (disinfected with 70% ethanol and evaporated before use). This sampling for microbial analysis was conducted after both ten and twelve months of operation. Triplicate DNA extractions were performed and pooled for each sediment sample using the PowerSoil kit (MO BIO laboratories, Carlsbad, CA). DNA samples from the same sampling site were further combined and quantified using Nanodrop spectrophotometer (Thermo Scientific, Wilmington, DE) for subsequent microbial phylogenetic, biomass, and metagenomic analyses. The details of microbial 16S rRNA sequencing for phylogenetic analysis and quantitative PCR (qPCR) for biomass measurement were the same as described previously.[13] Briefly, a portion of the 16S rRNA gene was amplified using the forward primer 515F (GTGYCAGCMGCCGCGGTAA) combined with the 454 Adaptor A, a CA linker sequence, and a unique 12-bp error-correcting Golay barcode, as well as the reverse primer 806R (GGACTACHVGGGTWTCTAAT) combined with the 454 Adaptor B and a 2-base linker sequence (TC). After amplification (denaturation at 94°C for 3 min, then amplification with 35 cycles of 94°C for 45 s, 50°C for 30 s, and 72°C for 90 s, and a final extension of 10 min at 72°C), PCR products were submitted to the Genomics Core Laboratory at King Abdullah University of Science and Technology for pyrosequencing on a 454 FLX Titanium genome sequencer (Roche). 16S rRNA sequencing data were processed using the Quantitative Insights Into Microbial Ecology (QIIME; version 1.5.0) pipeline with default settings.[20] A total of 71,107

sequences obtained were grouped into operational taxonomic units (OTUs) and further assigned with taxonomic names by using the Greengenes database. The OTU table created was rarefied so each sample contained 1270 reads. The weighted UniFrac distance matrix was used for Non-metric multidimensional scaling (NMDS) analysis with PRIMER 6.[21] The raw pyrosequencing data of 16S rRNA gene are deposited in the NCBI Sequence Read Archive with the accession number SRR1106900.

Bacterial biomass in sediments was estimated by qPCR of 16S rRNA gene using the forward primer 341F (CCTACGGGAGGCAGCAG) and the reverse primer 518R (ATTACCGCGGCTGCTGG) with iQ SYBR green Supermix (Bio-Rad) in a CFX96 Touch real-time PCR and associated software system (Bio-Rad). qPCR was performed in a 50-µl final reaction volume (approximately 20ng DNA per the vendor's protocol) and carried out as follows: 95°C for 3 min, followed by 40 cycles of 95°C for 15 s and 55°C for 30 s with a final melting curve period of 50 to 95°C with a heating increment rate of 0.5°C per 5 s and a continuous fluorescence measurement. Standard curves were generated by serial dilutions of the pCR2.1 plasmid vector (Invitrogen) containing cloned insert of bacterial 16S rRNA sequence as standard template DNA. All qPCR reactions were performed in triplicate. The range of efficiency was 95.9% - 99.8%, and R^2 of the standard curves was above 0.98.

Metagenomic and Metatranscriptomic Analyses

Sediment samples collected from 1 cm and 90 cm depths of all column systems

(C1-C4) were utilized for metagenomic and metatranscriptomic analyses. Additionally, higher depth resolution including 30 cm, 60 cm and 120 cm derived samples from column C3 were included. Metagenomic analysis was conducted with a total of 5-6 μg of DNA pooled from multiple sediment extractions for each sampling location. DNA was nebulized and tagged using the GS-FLX-Titanium Rapid Library MID Adapters Kit (454 Life Sciences, Branford, CT) before submitting for pyrosequencing on a 454 FLX Titanium sequencer (Roche). For metatranscriptomic analysis of sediment samples, 2.5 volumes of LifeGuard Soil Preservation Solution (MO BIO laboratories) were added to sediments immediately after harvesting from columns to stabilize microbial RNA. Sediment samples were stored at -20ºC for less than one week before triplicate RNA extractions were performed using the PowerSoil Total RNA Isolation Kit (MO BIO laboratories). Possible DNA residual was removed using the Turbo DNA-free kit (Ambion, Austin, TX) followed by treatment with mRNA-ONLY Prokaryotic mRNA Isolation Kit (Epicentre, Madison, WI). MICROBExpress and MICROBEnrich kits (Ambion) were then utilized to reduce the proportion of rRNA. mRNA was amplified using the MessageAmp II-Bacteria Kit (Ambion). Amplified RNA was converted to cDNA using the Universal RiboClone cDNA Synthesis System with random primers (Promega, Madison, WI) and purified using the QIAquick PCR purification kit (Qiagen). cDNA samples were subjected to gel electrophoresis and the fragments of 250–800bp was selected before submitting for pyrosequencing on a 454 FLX Titanium sequencer (Roche).

The metagenomic run yielded 1.11 million reads in total, with an average read

length of 431bp for 10 samples. The sample collected from 1 cm depth of column 2 was not included in the following analysis due to much lower number of sequences obtained compared to others. The metatranscriptomic run yielded 0.94 million reads in total, with an average read length of 406bp for 12 samples. The detailed sample list for metagenomic and metatranscriptomic analyses was included in Table S2. All metagenomic and metatranscriptomic data were trimmed and filtered using Trimmomatic according to default parameters.[22] Obtained sequences were phylogenetically assigned using the MetaPhlAn2.[23] Relative abundances of microbial genera were obtained for further statistical analyses. As the sequencing depth limited robust contig assembly and scaffolding, functional profiling of the metagenomic and metatranscriptomic reads was performed using the HUMAnN2 pipeline.[24] Functional gene family abundance was normalized to copies per million and further regrouped to enzyme commission (EC) values. Raw sequences were archived in the US Department of Energy Joint Genome Institute IMG/M system (http://img.jgi.doe.gov/cgi-bin/m/main.cgi) (ID: 46444-46472).

Statistical analyses

The relative abundance of EC genes in each library were transformed (Log(x+1)) for normalization, and then similarity matrices for all samples were calculated using Bray–Curtis distance. Non-metric multidimensional scaling (NMDS) and hierarchical clustering based on sample similarity matrices were performed using PRIMER 6. The SIMPROF test was conducted to validate the clustering of samples with 999

simulation permutations.

The identification of individual EC varying significantly between two clustering groups of samples was conducted using the R package ShotgunFunctionalizeR based on the Poisson model[25] and the algorithm LEfSe (the linear discriminant analysis effect size) using non-parametric factorial Kruskal-Wallis test followed by Linear Discriminant Analysis.[26] *p* values in the package ShotgunFunctionalizeR were corrected for multiple tests using the Benjamini–Hochberg correction factor, and default parameter settings of the algorithm LEfSe (alpha parameter of 0.05 for pairwise tests set and the threshold on the logarithmic LDA score as 2.0) were used. The heatmap was generated using the function Heatplot of the R package Made4.[27] Microbial taxonomic groups that are significantly associated to each group of samples were determined using the algorithm LEfSe.

The undirected, weighted microbial ECs expression networks were constructed by calculating the pairwise Pearson correlations between the relative abundance of ECs across all samples using the R package WGCNA.[28] Only the ECs detected in all metatranscriptomic libraries of this study and with average relative abundance above 0.1% were included in this analysis. ECs were organized into modules using the topological overlap measurement in a hierarchical cluster analysis. The data network was further exported to Cytoscape for fundamental statistics calculation.[29] The global descriptors of the modules including network density, network centralities, and clustering coefficient were calculated.

Additional statistical analyses such as Kendall's W Test and PERMANOVA

were performed using the SPSS package (version 16.0), PAST[30] or PRIMER 6. *p* values less than 0.05 were considered significant.

Results

Water and Biomass Profiles

Sediment columns received synthetic feed solutions with the same concentration of BDOC (1.5 mg/L) but with differing ratios (w/w) of refractory humic acid (0%, 40%, 60%, and 100%, respectively). The majority of this introduced BDOC (approximately 80+%) was assimilated within the first 30 cm of infiltration in all column systems, whereas more refractory carbon sources with higher SUVA values remained and were consumed slowly in the deeper infiltration zone as reported previously (Table S1).[17] The introduction of trace organic chemicals after 10 and 12 months of operation resulted in similar removal ratios at the 30 cm infiltration depth among the four systems (Kendall's W Test, p=0.779) despite differences in BDOC components (Table 1). Fractional removal of many, but not all, of the trace organic chemicals increased in association with increasing humic ratios (100% humic acid versus 0% humic acid additions; Kendall's W Test, p=0.023) at a transit depth of 120 cm.

The highest density of microorganisms was detected at the surficial 1 cm depth of sediment in all column systems with the 16S rRNA gene copy number ranging from 7.3E+08 to 2.5E+09 copies/g (Table S1), indicating higher growth in this region. Microbial density declined along the column profiles where the concentrations at the

remaining sampling zones from 30-120 cm were similar among different column systems at the same depth but contained less than 10% of the biomass recorded at the 1 cm depth. Only 3.4E+05 to 1.7E+07 copies/g of 16S rRNA gene was detected in effluent.

Phylogenetic Insights

Microbial community composition was grouped using the weighted UniFrac distance matrix for 16S rRNA sequencing results or Bray–Curtis distance matrix of microbial species through NMDS for metagenomic and metatranscriptomic data sets. Metagenomic reconstruction was consistent with microbial community structure as determined by 16S rRNA genes (Figure S1) (PERMANOVA, p=0.977). However, divergence was observed when contrasting phylogenetic compositions derived from metagenomic versus metatranscriptomic data sets (Figure 1a) (PERMANOVA, p=0.001). Shallow sediments (1 cm depth) were significantly different from samples collected in deeper sediments (30 cm to 120 cm) at both the genomic and transcriptomic levels (PERMANOVA, both p=0.001). In contrast, the introduction of different refractory substrate ratios did not have a significant impact on microbial community composition. Hence, for subsequent analysis, we established a binning approach that contrasted shallow (1cm) with deeper (30-120cm) sediment-associated communities.

Microbial genera that were significantly enriched in shallow sediment samples (1 cm depth) as compared to the deeper samples (30-120 cm depth) were identified using

the LEfSe algorithm for metagenomic sequencing data, with representative genera depicted in Figure 2a. Most of the genera enriched in shallow sediments belonged to the orders *Burkholderiales* and *Methylophilales* of the *Betaproteobacteria* superclass as well as the order *Rhodobacterales* within *Alphaproteobacteria*. In contrast, the genera significantly enriched in deeper sediments were mainly grouped within the *Bacteroidetes* and *Firmicutes* phyla (Figure 2a). Phylogenetic assignments from metatranscriptomic sequences, which can provide insights into active microorganisms, exhibited an ecological dominance of genera in shallow sediments that belonged to *Betaproteobacteria* and *Bacteroidetes*, such as *Chitinophaga*, *Spirosoma*, and *Marivirga*, while microbial genera within *Firmicutes* and *Gammaproteobacteria* were more significantly observed in deeper sediments, such as *Pseudomonas*, *Staphylococcus*, and *Pseudoflavonifractor* (Figure 2b).

Trends in Metabolic Potential and Properties

Microbial communities based on functional gene presence and expression profiles derived from metagenomic and metatranscriptomic analyses were also grouped through NMDS using Bray–Curtis similarity matrices (Figure 1b), and were analogous to those derived from microbial phylogenetic composition (Figure 1a). As shown in Figure 1b, microbial communities based on metagenomic and metatranscript profiles were distinctly different from each other (PERMANOVA, p=0.001). Of particular importance to our guiding research questions, gene presence and transcripts were significantly different when contrasting shallow (1 cm) versus deeper sediments

(30-120 cm) (PERMANOVA, p=0.012 and 0.01). These results indicate that infiltration depth, and by extension BDOC quantity, is a major driver of microbial metabolic properties. It also provides support for the spatial binning approach used in our analyses.

A contrast of the metabolic potential of the shallow versus deeper sediments revealed that 123 enzyme commissions (ECs) differed significantly within the metagenomic dataset using ShotgunFunctionalizeR. Of these, 86 ECs were significantly more abundant in shallow sediments (p<0.0001, Table S3). The algorithm LefSe identified 105 ECs that were best at discriminating microbial metabolic potential between shallow and deeper sediments, which have already been identified using ShotgunFunctionalizeR. Shallow sediments harbored a higher representation of ECs related to fundamental respiration and growth (Figure 3) including the metabolism of nitrogen, sulfur, oxidative phosphorylation, lipids, nucleotides, as well as synthesis of cofactors and vitamins. Exodeoxyribonuclease V (EC3.1.11.5) that participates in homologous recombination associated with cellular growth and repair was also overrepresented in the shallow sediments. In contrast, the potential for the metabolism of amino acids, secondary metabolites, and xenobiotics was enriched in deeper sediments. This included higher ratios of genes encoding for beta-lactamase (EC3.5.2.6) for penicillin and cephalosporin resistance. With respect to trace organic chemical degradation, a hydrogenase (EC1.12.99.6) important for nitrotoluene degradation and 4-hydroxybenzoyl-CoA reductase (EC1.3.7.9) that participates in benzoate degradation were also enriched in these deeper sediment

regions. Interestingly, acyl-homoserine-lactone acylase (EC3.5.1.97) known as quorum-quenching enzyme by hydrolyzing acyl-homoserine lactones also followed this trend.

More prominent differences were observed when contrasting metatranscripts between shallow and deeper sediments (Figure 4) where 550 ECs were identified as significantly different between shallow and deeper sediments. This was achieved using ShotgunFunctionalizeR (p<0.0001, Table S4) with further confirmation using LEfSe. Consistent with the metagenomic insights, the majority of these transcripts (456 of 550) were more abundant in shallow sediments, and metabolism categories were overrepresented in shallow sediments. Notably, this included the metabolism of carbohydrates such as pyruvate (represented by EC1.13.12.4 in Figure 4), nitrogen (EC1.7.2.6), sulfur (EC1.8.1.2), lipopolysaccharide (EC2.3.1.191), amino acids (EC6.3.5.4), lipids (EC2.7.7.41), nucleotides (EC2.7.7.8 and EC6.3.4.2), cofactors and vitamins (EC6.3.3.3), as well as terpenoids and polyketides (EC1.17.1.2). Some ECs responsible for microbial genome replication and repair such as DNA-(apurinic or apyrimidinic site) lyase (EC4.2.99.18) and crossover junction endodeoxyribonuclease (EC3.1.22.4) were also more abundant in shallow sediments, together with alanine-tRNA ligase (EC6.1.1.7) for translation, chemotaxisprotein-glutamate O-methyltransferase (EC2.1.1.80) for two-component system, and type I site-specific deoxyribonuclease (EC3.1.21.3).

In contrast, genes encoding enzymes associated with xenobiotic biodegradation and secondary metabolite metabolism were more prominent within the deeper

sediments (Figure 4). This included (S)-mandelate dehydrogenase (EC1.1.99.31) involved in the biodegradation of aminobenzoate, carbazole 1,9a-dioxygenase (EC1.14.12.22), which catalyzes the first step of the carbazole degradation pathway, as well as salicylate 1-monooxygenase (EC1.14.13.1) and protocatechuate 4,5-dioxygenase (EC1.13.11.8) vital for polycyclic aromatic hydrocarbon degradation. In accordance with the metagenomic results, penicillin and cephalosporin resistance (EC3.5.2.6), as well as streptomycin biosynthesis (EC1.1.1.133) were overrepresented in deeper sediments (Figure 4). Organomercurylyase (EC4.99.1.2) and mercury(II) reductase (EC1.16.1.1), which are crucial for bacterial organomercury resistance by converting highly toxic methyl mercury to ionic mercury and further reducing ionic mercury to elemental mercury, were also found in higher quantities in deeper sediments.

When comparing ratios of metabolic potential (genomic) versus expression (transcriptomic) within shallow sediments, only 28 ECs were identified to have significantly higher proportional representations on the transcriptomic level (p<0.0001, Table S5). In contrast, 106 ECs were significantly more abundant in deeper sediments when contrasted with their metagenomic presence (p<0.0001, Table S6). The majority of these were associated with the biodegradation and metabolism of diverse xenobiotics (Figure S2), including vanillate monooxygenase (EC1.14.13.82) which participates in the biodegradation of aminobenzoate, as well as 2-hydroxyhexa-2,4-dienoate hydratase (EC4.2.1.132) and 2-oxopent-4-enoate hydratase (EC4.2.1.80) involved in xylene degradation. Beta-lactamase (EC3.5.2.6),

organomercurylyase (EC4.99.1.2), and mercury(II) reductase (EC1.16.1.1) were also overexpressed in deeper sediments when contrasted with their relative presence in the metagenome (Figure S2).

While the overall initial BDOC concentration between the different columns was largely uniform, it should be noted that when comparing the expression of genes in deeper sediments among the four different column set-ups, ECs involved in xenobiotic biodegradation were generally more abundant in association with higher proportions of humics to labile carbon (Figure S2). Such ECs included salicylate 1-monooxygenase (EC1.14.13.1), 4-hydroxy 2-oxovalerate aldolase (EC4.1.3.39) involved in benzoate degradation, and carboxymethylenebutenolidase (EC3.1.1.45) for chlorocyclohexane and chlorobenzene degradation. Beta-lactamase (EC3.5.2.6) for penicillin and cephalosporin biosynthesis and resistance also increased with increasing humic ratios (Figure S2).

Network Analysis

To further interpret these trends, a network analysis of identified genes was achieved by clustering genes into modules reflecting functional associations in the resident microbial community. Five functional gene expression modules were identified, and fundamental statistics describing the network modules and the hubs have been listed in Table 2, with the full list of ECs in each module shown in Table S7. Module 4 was significantly associated with shallow sediments (p=0.007) and consisted of ECs involved in fundamental respiratory processes (Figure 4), including starch and sucrose

metabolism (EC3.2.1.1), alanine, aspartate and glutamate metabolism (EC6.3.5.4), lipopolysaccharide biosynthesis (EC2.4.99.12), glycerophospholipid metabolism (EC3.1.4.46), riboflavin metabolism (EC3.5.4.25), biotin metabolism (EC2.1.1.199), two-component system (EC2.1.1.80), and aminoacyl-tRNA biosynthesis (EC6.1.1.1) (Table S7). In contrast, ECs in module 1, which were significantly correlated with deeper sediments (p=0.05), participated in xenobiotic biodegradation, such as (S)-mandelate dehydrogenase (EC1.1.99.31) and vanillate monooxygenase (EC1.14.13.82) degrading aminobenzoate, as well as carboxymethylenebutenolidase (EC3.1.1.45), which participates in toluene degradation (Table S7). ECs responsible for nucleotide metabolism and genome replication and repair were prevalent in module 2. In module 3, ECs associated with lipid metabolisms were dominant. Module 5 was overrepresented by metabolisms of cofactors, vitamins and amino acids.

Discussion

Biofiltration processes play a vital role in the removal of dissolved organic matter in both natural and engineered systems. It has been estimated that global freshwater sediments receive organic carbon at 0.2 Pg annually.[31] Allochthonous organic matter from terrestrial sources, particularly from anthropogenic sources like municipal, industrial, and agricultural wastewater contributes significantly to the total amount of organic matter in freshwater.[32] Mineralization of these organic compounds by

indigenous terrestrial microorganisms is thus an important component of global carbon and nutrient cycles. The results of this study provide genetic support for microbial consumption of BDOC within the top layer of sediments, regardless of different ratios of refractory to more labile organic carbon in the feed solutions. Genomic and transcriptomic identification of higher densities of genes involved in replication were supported by higher quantities of overall biomass in shallow sediments. This laboratory result is consistent with field observations where the top layer (0-20 cm) of sediment contained approximately 65% of the total microbial biomass residing in a 2-meter vertical profile of sediments.[11]

In contrast, prior work demonstrated that comparably recalcitrant trace organic chemicals were preferentially removed in the deeper infiltration zones of these columns (Table 1 and results in Alidina *et al*.[17]). This indicates that the biodegradation capabilities of the resident microbial community are distinctly higher within deeper infiltration zones compared to shallow sediments. Importantly this increased activity exists despite a lower density of biomass. With respect to ecological differences, there was a clear contrast between shallow and deeper sediments. Microbial genera significantly enriched in shallow sediments harbored a diverse array of metabolic capabilities for the assimilation of labile organics such as methylotrophic genera (*Methylotenera*, *Methylovorus*, and *Methylobacillus*), *Alicycliphilus* sp. capable of utilizing cyclic hydrocarbons, *Comamonas* sp. capable of using diverse organic compounds, *Marivirga* sp. capable of using glycerol, glucose, galactose, and sucrose, and *Mucilaginibacter* sp. capable of utilizing polysaccharides (Figure 2a).[33] In

contrast, microbial genera overrepresented in deeper sediments were generally not linked with the assimilation of labile organics (Figure 2b).

In support of our guiding hypothesis, the comparatively oligotrophic conditions within the deeper percolation zone resulted in higher genetic signatures for xenobiotic biodegradation potential as well as secondary metabolite metabolism. While limitations in sequencing depth and a lack of metagenome assembled genomes (MAGs) limits our ability to link increases in candidate enzyme synthesis to particular microorganisms, it was sufficient to demonstrate differences in the presence and expression of a broader suite of putative enzymes that could in turn be correlated to increased trace organic biotransformation.

A myriad of approaches to enhance the biodegradation of recalcitrant organic chemicals in contaminated groundwater, such as adding carbon sources or nutrients, enhancing dissolved oxygen or other electron acceptor levels, and bioaugmenting clades of target organisms such as *Dehalococcoides* have been actively explored.[34-37] Our results build upon this broader theme by providing genetic support for how BDOC limitation can disproportionately enhance the biotransformation rates of trace organic chemicals. This indicates that modifications of organic availability in influent or at target depths can help to achieve this goal.

The results of this study also suggest that the deeper infiltration zone could be a reservoir of novel biodegrading microorganisms and genes. Previous studies support this idea, where the biodegradation capabilities in subsurface saturated soils have been observed for diverse chemical pollutants such as benzene, toluene, methyl tert-butyl

ether, polycyclic aromatic hydrocarbons, or petroleum.[38-41] Approaches such as ours that integrate high throughput genomics techniques and network analyses have been anticipated to bring further insights.[42]

Furthermore, biosynthesis and resistance of antibiotics were overrepresented in the deeper infiltration zones of this study, in accordance with previous reports that antibiotic resistance was prevalent in deep subsurface bacteria[43]. This could represent an interesting region for bioprospecting for novel antibiotics[44] and an unwanted source of antibiotic resistant bacteria if used for drinking water applications.[45-46] The quorum-quenching enzyme acyl-homoserine-lactone acylase was also enriched in the deeper infiltration zone. Quorum sensing can regulate community density and coordinate gene expression for biofilm formation, virulence, and antibiotic resistance[47]. Though not directly relevant to potable water reuse, organomercurylyase (MerB) and mercury(II) reductase (MerA)[48] were found to be overexpressed in deeper sediments of this study, suggesting that microorganisms within the deeper infiltration zone could harbor the capacity of cleaning up methylmercury contamination. More broadly, these results suggest that enhanced contaminant attenuation potential is not dependent upon the presence of target contaminants of concern but rather is selected by environmental factors such as organic carbon availability.

Conclusions

Biofiltration processes, utilized intentionally or not during managed aquifer recharge, play a vital role in the removal of DOM in both natural and engineered systems.

Concern about the persistence of trace organic chemicals in biofiltration systems and downstream drinking water mandates a better understanding of microbiological processes controlling the attenuation of water pollutants in these systems. Through the comprehensive analysis of a series of laboratory sediment columns that simulated infiltration processes using microbial phylogenic, metagenomic, and transcriptomic methods, our results indicate that spatial variations in the metabolic profiles of microbial communities selected during infiltration at depth are unique and characterized by having an increased genetic potential for pollutant biotransformation despite lower overall biomass. In doing so, this study provides genetic correlations to past findings of increased rates of trace organic degradation within these more oligotrophic saturated zones.[16-17] These insights into the metabolic characteristics can inform our understanding of fundamental microbial responses to biofiltration process as well as provide tools for the optimization of trace organic chemical biodegradation. Hence, we conclude that both infiltration depth and influent properties should be considered to sustain a microbial community more capable of the degradation of xenobiotic trace organics during managed aquifer recharge.

Supplementary Information

Supplementary figures (Figure S1-S2) and tables (Table S1-S7) are available in the online version of this article.

Acknowledgements

This material is based upon work supported by the US National Science Foundation under grants EEC-1028968 and CBET-1055396 and discretionary investigator funds at King Abdullah University of Science and Technology (KAUST). The authors thank Dr. Kristin Mikkelson for valuable insight and edits during manuscript preparation and Dr. Pascal Saikaly at KAUST for technical assistance in experimental design and interpretation.

References

1. G. Amy, J. E. Drewes, Soil aquifer treatment (SAT) as a natural and sustainable wastewater reclamation/reuse technology: fate of wastewater effluent organic matter (EfOM) and trace organic compounds, Environ. Monit. Assess. 2007, **129**, 19–26

2. A. Konopka, R. Turco, Biodegradation of organic compounds in vadose zone and aquifer sediments, Appl. Environ. Microbiol. 1991, **57**, 2260–2268

3. P. A. Holden, N. Fierer, Microbial processes in the vadose zone, Vadose Zone J. 2005, **4**, 1–21

4. C. Hoppe-Jones, G. Oldham, J. E. Drewes, Attenuation of total organic carbon and unregulated trace organic chemicals in U.S. riverbank filtration systems, Water Res. 2010, **44**, 4643-4659

5. R. P. Schwarzenbach, B. I. Escher, K. Fenner, T. B. Hofstetter, C. A. Johnson, U. von Gunten, B. Wehrli, The challenge of micropollutants in aquatic systems, Science 2006, **313**, 1072-1077

6. J. E. Drewes, S. Khan, Water reuse for drinking water augmentation, J. Edzwald, (ed.) Water Quality and Treatment, 2011, 6th Edition. 16.1-16.48. American Water Works Association. Denver, Colorado.

7. D. R. Lovley, Cleaning up with genomics: applying molecular biology to bioremediation, Nat. Rev. Microbiol. 2003, **1**, 35-44

8. S. Spring, R. Schulze, J. Overmann, K.-H. Schleifer, Identification and characterization of ecologically significant prokaryotes in the sediment of

freshwater lakes: molecular and cultivation studies, FEMS Microbiol. Rev. 2000, **24**, 573–590

9. N. Tšertova, A. Kisand, H. Tammert, V. Kisand, Low seasonal variability in community composition of sediment bacteria in large and shallow lake. Environ. Microbiol. Rep. 2011, **3**, 270–277
10. S. A. Wakelin, M. J. Colloff, R. S. Kookana, Effect of wastewater treatment plant effluent on microbial function and community structure in the sediment of a freshwater stream with variable seasonal flow, Appl. Environ. Microbiol. 2008, **74**, 2659–2668
11. E. Blume, M. Bischoff, J. M. Reichert, T. Moorman, A. Konopka, R. F. Turco, Surface and subsurface microbial biomass, community structure and metabolic activity as a function of soil depth and season, Appl. Soil. Ecol. 2002, **20**, 171–181
12. J. K. Fredrickson, T. R. Garland, R. J. Hicks, J. M. Thomas, S. W. Li, K. M. McFadden, Lithotrophic and heterotrophic bacteria in deep subsurface sediments and their relation to sediment properties, Geomicrobiol. J. 1989, **7**, 53-66
13. D. Li, J. O. Sharp, P. E. Saikaly, S. Ali, M. Alidina, M. S. Alarawi, S. Keller, C. Hoppe-Jones, J. E. Drewes, Dissolved organic carbon influences microbial community composition and diversity in managed aquifer recharge systems. Appl. Environ. Microbiol. 2012, **78**, 6819-6828
14. D. P. Martino, E. L. Grossman, G. A. Ulrich, K. C. Burger, J. L. Schlichenmeyer, J. M. Suflita, J. W. Ammerman, Microbial abundance and activity in a

low-conductivity aquifer system in east-central Texas, Microb. Ecol. 1998, **35**, 224-234

15. J. M. Yagi, E. F. Neuhauser, J. A. Ripp, D. M. Mauro, E. L. Madsen, Subsurface ecosystem resilience: long-term attenuation of subsurface contaminants supports a dynamic microbial community. ISME J. 2010, **4**, 131–143
16. D. Li, M. Alidina, J. E. Drewes, Role of primary substrate composition on microbial community structure and function and trace organic chemical attenuation in managed aquifer recharge systems, Appl. Microbiol. Biotechnol. 2014, **98**, 5747-5756
17. M. Alidina, D. Li, M. Ouf, J. E. Drewes, Role of primary substrate composition and concentration on attenuation of trace organic chemicals in managed aquifer recharge systems, J. Environ. Manage. 2014, **144**, 58-66
18. K. M. Onesios, E. J. Bouwer, Biological removal of pharmaceuticals and personal care products during laboratory soil aquifer treatment simulation with different primary substrate concentrations, Water Res. 2012, **46**, 2365-2375
19. J. Regnery, D. Li, J. Lee, K. M. Smits, J. O. Sharp, Hydrogeochemical and microbiological effects of simulated recharge and drying within a 2D meso-scale aquifer, Chemosphere 2020, **241**, 125116
20. J. G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A.

Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, R. Knight, QIIME allows analysis of high-throughput community sequencing data, Nat. Methods 2010, **7**, 335-336

21. K. R. Clarke, R. N. Gorley, PRIMER v6: User Manual/Tutorial. PRIMER-E, Plymouth. 2006

22. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 2014, **30**, 2114–2120

23. N. Segata, L. Waldron, A. Ballarini, V. Narasimhan, O. Jousson, C. Huttenhower, Metagenomic microbial community profiling using unique clade-specific marker genes, Nat. Methods 2012, **8**, 811–814

24. E. A. Franzosa, L. J. McIver, G. Rahnavard, L. R. Thompson, M. Schirmer, G. Weingart, L. K. Schwarzberg, R. Knight, J. G. Caporaso, N. Segata, C. Huttenhower, Species-level functional profiling of metagenomes and metatranscriptomes, Nat. Methods 2018, **15**, 962-968

25. E. Kristiansson, P. Hugenholtz, D. Dalevi, ShotgunFunctionalizeR: an R-package for functional comparison of metagenomes, Bioinformatics 2009, **25**, 2737–2738

26. N. Segata, J. Izard, L. Walron, D. Gevers, L. Miropolsky, W. Garrett, C. Huttenhower, Metagenomic biomarker discovery and explanation, Genome Biol. 2011, **12**, R60

27. A. C. Culhane, J. Thioulouse, G. Perriere, D. G. Higgins, MADE4: an R package for multivariate analysis of gene expression data, Bioinformatics 2005, **21**, 2789-2790

28. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinformatics 2008, **9**, 559

29. R. Saito, M. E. Smoot, K. Ono, J. Ruscheinski, P. L. Wang, S. Lotia, A. R. Pico, G. D. Bader, T. Ideker, A travel guide to Cytoscape plugins, Nat. Methods 2012, **9**, 1069-1076

30. Ø. Hammer, D. A. T. Harper, P. D. Ryan, Past: paleontological statistics software package for education and data analysis, Palaeont. Elec. 2001, **4**, e1–e9

31. J. J. Cole, Y. T. Prairie, N. F. Caraco, W. H. McDowell, L. J. Tranvik, R. G. Striegl, C. M. Duarte, P. Kortelainen, J. A. Downing, J. J. Middelburg, J. Melack, Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget, Ecosystems 2007, **10**, 171–184

32. J. T. Lennon, L. E. Pfaff, Source and supply of terrestrial organic matter affects aquatic microbial metabolism, Aquat. Microb. Ecol. 2005, **39**, 107–119

33. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, E. Stackebrandt, (Eds.) The Prokaryotes: A Handbook on the Biology of Bacteria. Springer 3rd ed. 2006

34. J. He, K. M. Ritalahti, K.-L. Yang, S. S. Koenigsberg, F. E. Löffler, Detoxification of vinyl chloride to ethene coupled to growth of an anaerobic bacterium, Nature 2003, **424**, 62-65

35. F. E. Löffler, E. A. Edwards, Harnessing microbial activities for environmental cleanup, Curr. Opin. Biotechnol. 2006, **17**, 274-284

36. J. Pandey, A. Chauhan, R. K. Jain, Integrative approaches for assessing the ecological sustainability of in situ bioremediation, FEMS Microbiol. Rev. 2009,

33, 324-375

37. K. Rossmassler, S. Kim, C. D. Broeckling, S. Galloway, J. Prenni, S. K. De Long, Impact of primary carbon sources on microbiome shaping and biotransformation of pharmaceuticals and personal care products, Biodegradation 2019, **30**, 127-145
38. I. M. Head, D. M. Jones, S. R. Larter, Biological activity in the deep subsurface and the origin of heavy oil, Nature 2003, **426**, 344-352
39. D. Moreels, L. Bastiaens, F. Ollevier, R. Merckx, L. Diels, D. Springael, Evaluation of the intrinsic methyl tert-butyl ether (MTBE) biodegradation potential of hydrocarbon contaminated subsurface soils in batch microcosm systems, FEMS Microbiol. Ecol. 2004, **49**, 121-128
40. E. J. O'Loughlin, G. K. Sims, S. J. Traina, Biodegradation of 2-methyl, 2-ethyl, and 2-hydroxypyridine by an *Arthrobacter* sp. isolated from subsurface sediment, Biodegradation 1999, **10**, 93-104
41. W. F. Röling, H. W. van Verseveld, Natural attenuation: what does the subsurface have in store? Biodegradation 2002, **13**, 53-64
42. R. Vilchez-Vargas, H. Junca, D. H. Pieper, Metabolic networks, microbial ecology and 'omics' technologies: towards understanding in situ biodegradation processes, Environ. Microbiol. 2010, **12**, 3089-3104
43. M. G. Brown, D. L. Balkwill, Antibiotic resistance in bacteria isolated from the deep terrestrial subsurface, Microb. Ecol. 2009, **57**, 484-493
44. C. Borsetto, E. M. H. Wellington, Bioprospecting Soil Metagenomes for Antibiotics. In: R. Paterson, N. Lima, (eds) Bioprospecting. Topics in

Biodiversity and Conservation, vol 16. Springer, Cham. 2017

45. I. Vaz-Moreira, O. C. Nunes, C. M. Manaia, Diversity and antibiotic resistance patterns of Sphingomonadaceae isolates from drinking water, Appl. Environ. Microbiol. 2011, **77**, 5697–5706

46. C. Xi, Y. Zhang, C. F. Marrs, W. Ye, C. Simon, B. Foxman, J. Nriagu, Prevalence of antibiotic resistance in drinking water treatment and distribution systems, Appl. Environ. Microbiol. 2009, **75**, 5714–5718

47. Z. Li, S. K. Nair, Quorum sensing: how bacteria can coordinate activity and synchronize their response to external signals? Protein Sci. 2012, **21**, 1403-1417

48. J. Lafrance-Vanasse, M. Lefebvre, P. Di Lello, J. Sygusch, J. G. Omichinski, Crystal structures of the organomercuriallyase MerB in its free and mercury-bound forms: insights into the mechanism of methylmercury degradation, J. Biol. Chem. 2009, **284**, 938-944

Table 1. Fractions[a] of trace organic pollutants remaining during transit through sediment column systems

Compound	Column 1 (100:0)[b]		Column 2 (60:40)		Column 3 (40:60)		Column 4 (0:100)	
	30 cm	Effluent	30 cm	Effluent	30 cm	Effluent	30 cm	Effluent
Acetaminophen	0.71	0.60	0.70	0.32	0.03	0.00	0.00	0.00
Atenolol	0.94	0.39	0.84	0.56	0.91	0.57	0.70	0.06
Atrazine	0.98	0.96	0.98	0.95	0.97	0.90	1.01	0.94
Caffeine	0.45	0.24	0.53	0.15	0.53	0.09	0.80	0.05
Carbamazepine	0.95	0.93	0.94	0.92	0.97	0.93	0.98	0.96
DEET	0.96	0.91	0.98	0.88	0.99	0.84	1.00	0.90
Dilantin	0.99	0.95	1.00	0.91	1.00	0.80	0.99	0.80
Fluoxetine	0.03	0.01	0.11	0.02	0.16	0.08	0.13	0.05
Primidone	0.98	1.06	1.07	1.01	1.07	1.16	1.06	1.10
Sulfamethoxazole	0.71	0.67	0.79	0.76	0.90	0.90	0.95	0.75
Trimethoprim	0.76	0.58	0.79	0.32	0.80	0.20	0.77	0.33
TCPP	1.10	1.48	1.17	1.29	1.07	1.26	1.13	1.64
TDCPP	1.31	2.81	1.61	2.98	1.24	2.21	1.16	2.08
BPA	1.09	1.10	0.99	0.76	1.06	0.83	1.01	0.57
Diclofenac	0.98	0.90	0.95	0.92	0.97	0.88	0.99	0.53
Gemfibrozil	0.99	0.88	0.92	0.79	0.93	0.59	0.91	0.28
Ibuprofen	0.79	0.78	0.76	0.37	0.70	0.30	0.07	0.00
Naproxen	0.91	0.87	0.91	0.86	0.96	0.82	0.96	0.46

[a] The fraction was calculated by the concentration in 30 cm or effluent divided by the initial spiking concentration. All the fractions were the average of two sampling results. Removal fractions of a subset of these chemicals have been previously reported in Alidina et al.[17] [b] Peptone/yeast extract was mixed with humic acid in ratios

of 100:0, 60:40, 40:60, and 0:100 (w/w) that were in turn introduced to the four different column systems (C1-C4), respectively.

Table 2. Fundamental statistics describing network modules of metatranscriptomic data[a]

Module	Clustering coefficient	Centralization	Density	Neighbors	Heterogeneity	Nodes	Hubs	Overall functions
1	0.734	0.297	0.330	11.9	0.347	37	4-hydroxyphenylpyruvate dioxygenase (EC1.13.11.27)	Xenobiotic biodegradation
2	0.830	0.285	0.696	16.0	0.343	24	Inositol-phosphate phosphatase (EC3.1.3.25)	Genome replication and repair
3	0.723	0.254	0.209	5.6	0.559	28	5'-nucleotidase (EC3.1.3.5)	Lipid metabolism
4	0.730	0.453	0.436	27.9	0.477	65	Dihydrolipoamide dehydrogenase (EC1.8.1.4)	Fundamental respiratory processes
5	0.813	0.243	0.684	14.4	0.380	22	Arginine decarboxylase (EC4.1.1.19)	Cofactor, vitamin and amino acid metabolisms

[a] Only the ECs detected in all metatranscriptomic libraries of this study and with average relative abundance above 0.1% were included in the analysis.

Figure 1. The NMDS clustering of microbial communities derived from metagenomic and metatranscriptomic sequences based on (a) phylogenetic species assignment and (b) functional genes and expression across sediment samples (binned into 1 cm and 30-120 cm samples). Sample names represent the columns 1-4 and sampling depths 1 cm, 30 cm, 60 cm, 90 cm, and 120 cm. For example, C1-90 denotes the sample collected from column 1 (100:0) at a depth of 90 cm.

Figure 2. The relative abundance of the 20 most abundant microbial genera derived from (a) metagenomic and (b) metatranscriptomic sequencing that are significantly associated with shallow (1cm, red) or deeper sediments (30-120 cm, blue).

Figure 3. Representative ECs from metagenomic analyses that display the largest differences in the functional gene profiles of the microbial community between shallow and deeper sediments of this study. The ECs were shown as the relative abundances of encoding genes (normalized by Z-score across all data sets) across columns and depths as clarified in figure 1.

Figure 4. Representative ECs from metatranscriptomic analyses that showed the largest differences in the functional gene expression profiles of microbial community between shallow and deeper sediments of this study. The ECs were shown as the relative abundances of encoding genes (normalized by Z-score across all data sets).

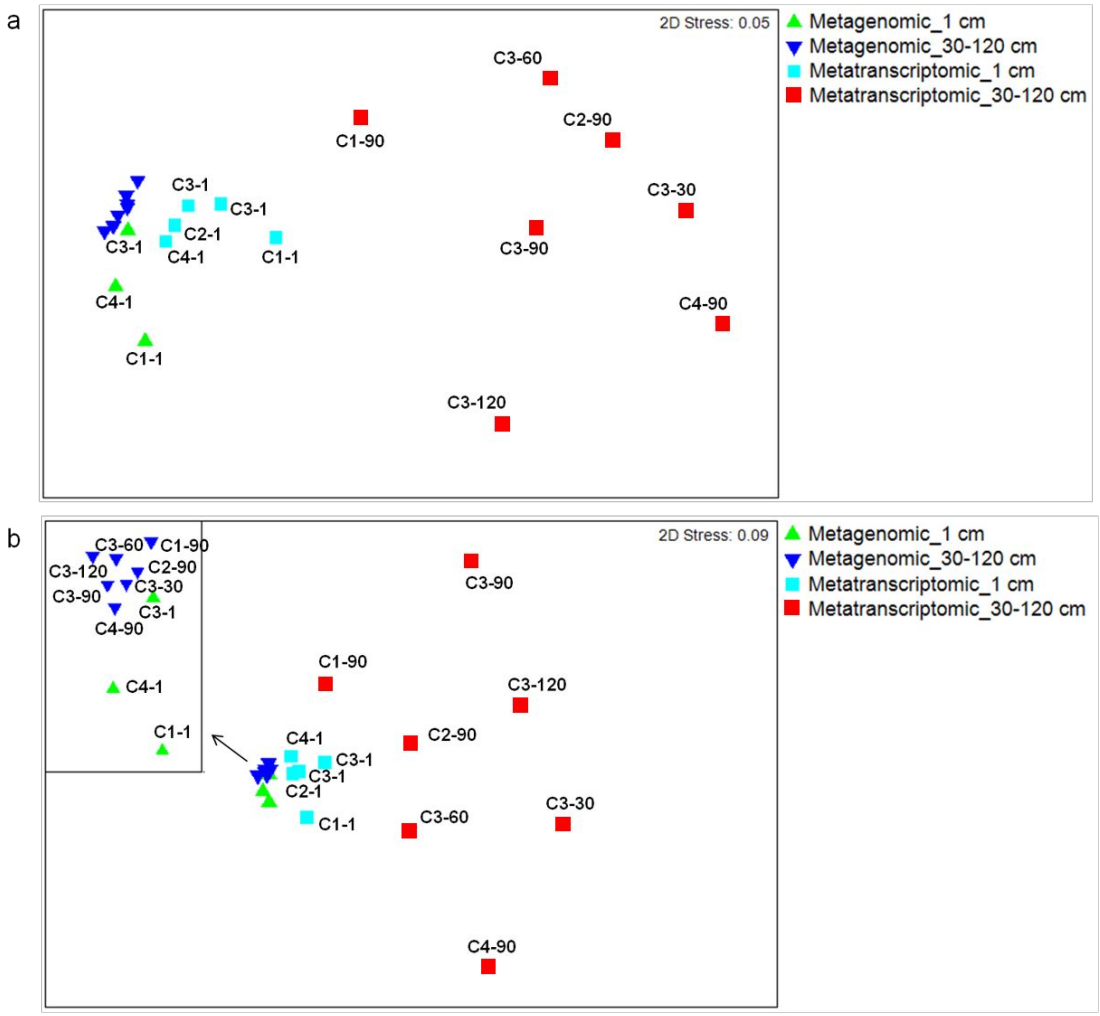

a
2D Stress: 0.05
Metagenomic_1 cm
Metagenomic_30-120 cm
Metatranscriptomic_1 cm
Metatranscriptomic_30-120 cm
C3-60
C1-90
C2-90
C3-1
C3-1
C3-30
C2-1
C3-1
C4-1
C1-1
C3-90
C4-1
C4-90
C1-1
C3-120
b
2D Stress: 0.09
Metagenomic_1 cm
Metagenomic_30-120 cm
Metatranscriptomic_1 cm
Metatranscriptomic_30-120 cm
C3-60
C1-90
C3-120
C2-90
C3-30
C3-90
C3-1
C4-90
C3-90
C1-90
C4-1
C3-120
C1-1
C4-1
C2-90
C3-1
C3-1
C2-1
C3-30
C1-1
C3-60
C4-90

Figure 1

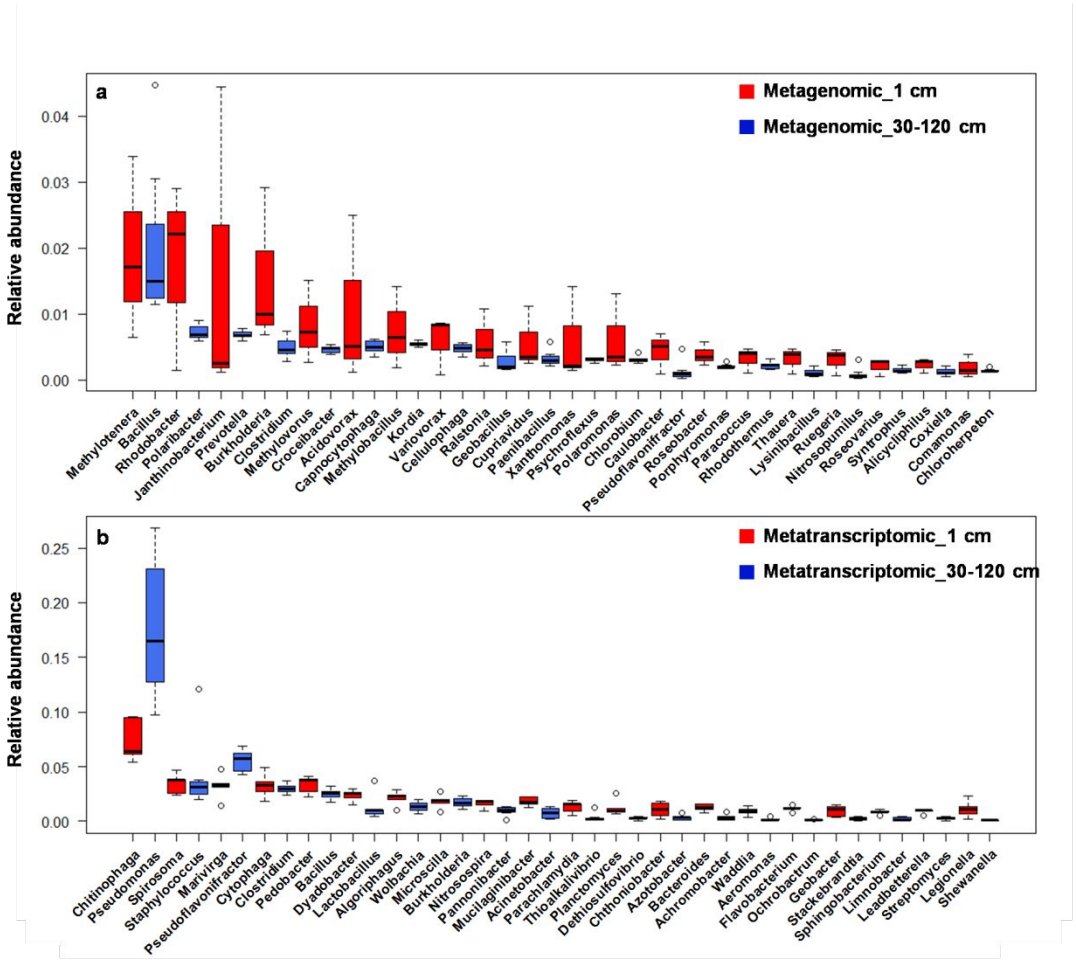

a
Metagenomic_1 cm
Metagenomic_30-120 cm
Relative abundance
0.04
0.03
0.02
0.01
0.00
Methylotenera
Bacillus
Rhodobacter
Polaribacter
Janthinobacterium
Prevotella
Burkholderia
Clostridium
Methylovorus
Crocebacter
Acidovorax
Caprocytophaga
Methylobacillus
Kordia
Variovorax
Cellulophaga
Ralstonia
Geobacillus
Cupriavidus
Paenibacillus
Xanthomonas
Psychroflexus
Polaromonas
Chlorobium
Caulobacter
Pseudoflavonifractor
Roseobacter
Porphyromonas
Paracoccus
Rhodothermus
Thauera
Lysinibacillus
Ruegeria
Nitrosopumilus
Roseovarius
Syntrophus
Alicycliphilus
Coxiella
Comamonas
Chlororherpeton
b
Metatranscriptomic_1 cm
Metatranscriptomic_30-120 cm
Relative abundance
0.25
0.20
0.15
0.10
0.05
0.00
Chitinophaga
Pseudomonas
Spirosoma
Staphylococcus
Marivirga
Pseudoflavonifractor
Cytophaga
Clostridium
Pedobacter
Bacillus
Dyadobacter
Lactobacillus
Algoriphagus
Wolbachia
Microscilla
Burkholderia
Nitrosospira
Pannonibacter
Mucilaginibacter
Acinetobacter
Parachlamydia
Thioalkalivibrio
Planctomyces
Dethiosulfovibrio
Chthoniobacter
Azotobacter
Bacteroides
Achromobacter
Waddlia
Aeromonas
Flavobacterium
Ochrobactrum
Geobacter
Stackebrandtia
Sphingobacterium
Limnobacter
Leadbetterella
Streptomyces
Legionella
Shewanella

Figure 2

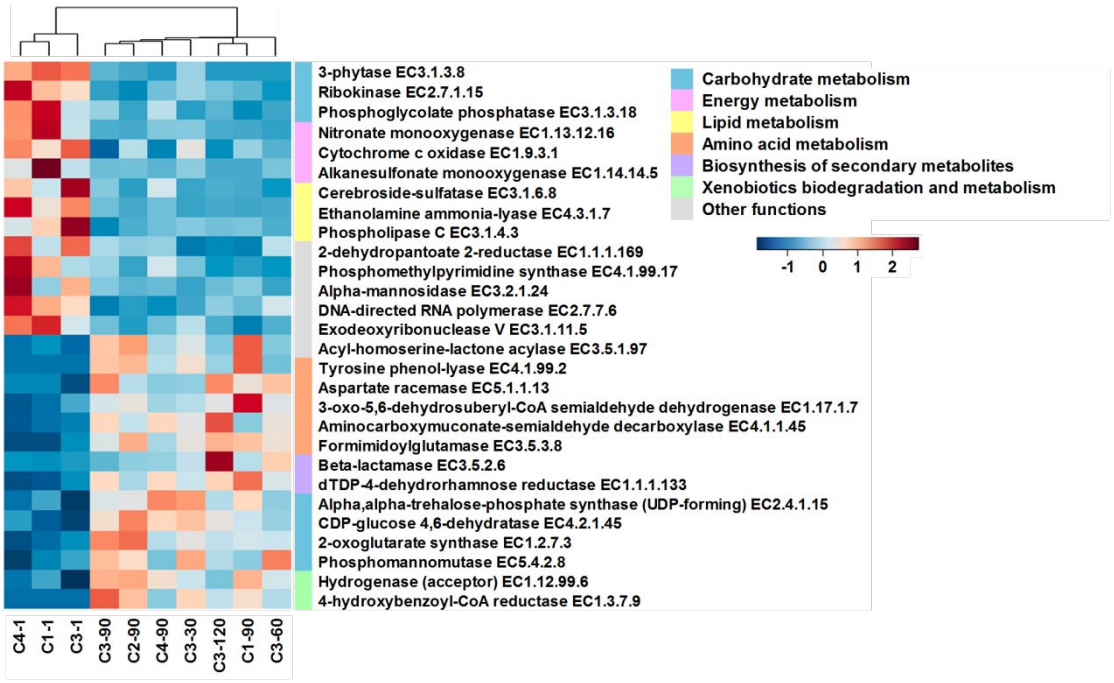

3-phytase EC3.1.3.8
Ribokinase EC2.7.1.15
Phosphoglycolate phosphatase EC3.1.3.18
Nitronate monooxygenase EC1.13.12.16
Cytochrome c oxidase EC1.9.3.1
Alkanesulfonate monooxygenase EC1.14.14.5
Cerebroside-sulfatase EC3.1.6.8
Ethanolamine ammonia-lyase EC4.3.1.7
Phospholipase C EC3.1.4.3
2-dehydropantoate 2-reductase EC1.1.1.169
Phosphomethylpyrimidine synthase EC4.1.99.17
Alpha-mannosidase EC3.2.1.24
DNA-directed RNA polymerase EC2.7.7.6
Exodeoxyribonuclease V EC3.1.11.5
Acyl-homoserine-lactone acylase EC3.5.1.97
Tyrosine phenol-lyase EC4.1.99.2
Aspartate racemase EC5.1.1.13
3-oxo-5,6-dehydrosuberyl-CoA semialdehyde dehydrogenase EC1.17.1.7
Aminocarboxymuconate-semialdehyde decarboxylase EC4.1.1.45
Formimidoylglutamase EC3.5.3.8
Beta-lactamase EC3.5.2.6
dTDP-4-dehydrorhamnose reductase EC1.1.1.133
Alpha,alpha-trehalose-phosphate synthase (UDP-forming) EC2.4.1.15
CDP-glucose 4,6-dehydratase EC4.2.1.45
2-oxoglutarate synthase EC1.2.7.3
Phosphomannomutase EC5.4.2.8
Hydrogenase (acceptor) EC1.12.99.6
4-hydroxybenzoyl-CoA reductase EC1.3.7.9
Carbohydrate metabolism
Energy metabolism
Lipid metabolism
Amino acid metabolism
Biosynthesis of secondary metabolites
Xenobiotics biodegradation and metabolism
Other functions
-1
0
1
2
C4-1
C1-1
C3-1
C3-90
C2-90
C4-90
C3-30
C3-120
C1-90
C3-60

Figure 3

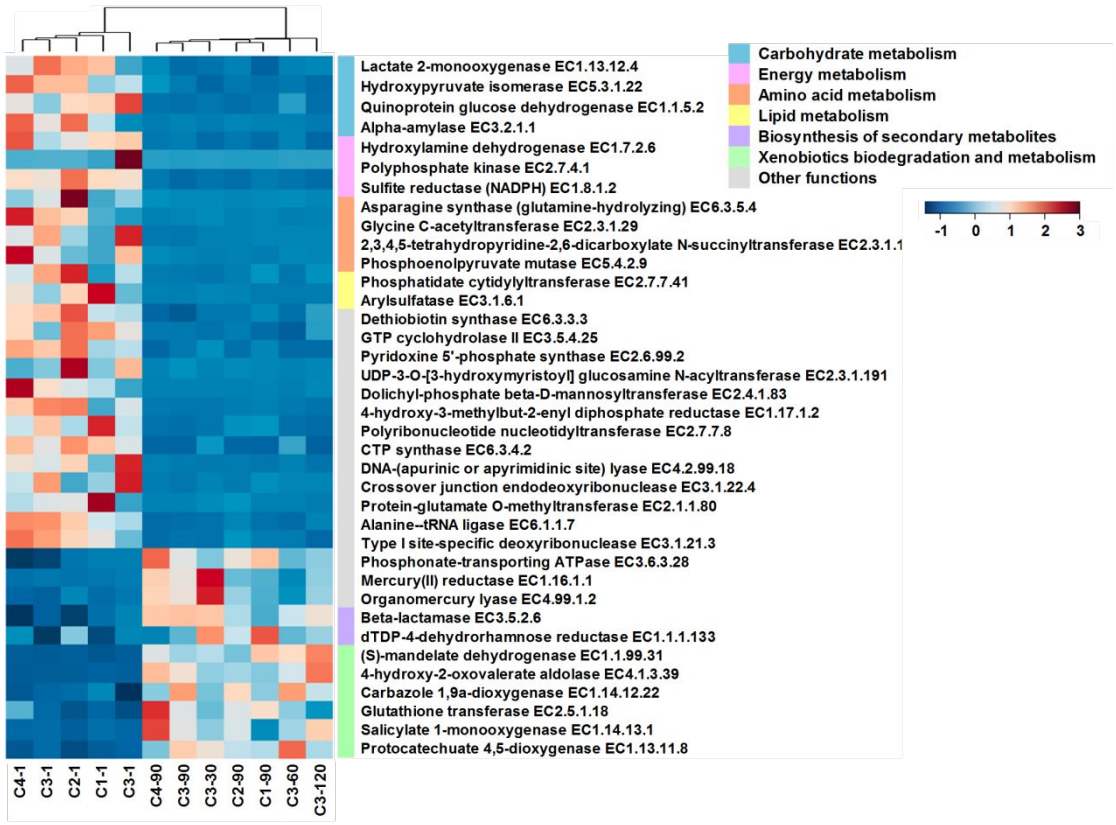
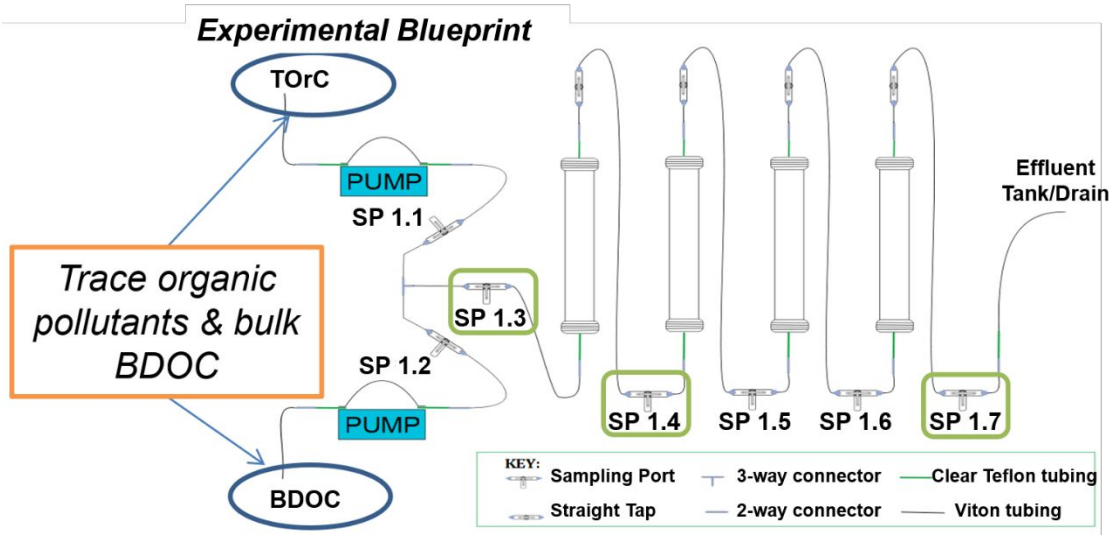

Lactate 2-monooxygenase EC1.13.12.4
Hydroxypyruvate isomerase EC5.3.1.22
Quinoprotein glucose dehydrogenase EC1.1.5.2
Alpha-amylase EC3.2.1.1
Hydroxylamine dehydrogenase EC1.7.2.6
Polyphosphate kinase EC2.7.4.1
Sulfite reductase (NADPH) EC1.8.1.2
Asparagine synthase (glutamine-hydrolyzing) EC6.3.5.4
Glycine C-acetyltransferase EC2.3.1.29
2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase EC2.3.1.1
Phosphoenolpyruvate mutase EC5.4.2.9
Phosphatidate cytidylyltransferase EC2.7.7.41
Arylsulfatase EC3.1.6.1
Dethiobiotin synthase EC6.3.3.3
GTP cyclohydrolase II EC3.5.4.25
Pyridoxine 5'-phosphate synthase EC2.6.99.2
UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase EC2.3.1.191
Dolichyl-phosphate beta-D-mannosyltransferase EC2.4.1.83
4-hydroxy-3-methylbut-2-enyl diphosphate reductase EC1.17.1.2
Polyribonucleotide nucleotidyltransferase EC2.7.7.8
CTP synthase EC6.3.4.2
DNA-(apurinic or apyrimidinic site) lyase EC4.2.99.18
Crossover junction endodeoxyribonuclease EC3.1.22.4
Protein-glutamate O-methyltransferase EC2.1.1.80
Alanine--tRNA ligase EC6.1.1.7
Type I site-specific deoxyribonuclease EC3.1.21.3
Phosphonate-transporting ATPase EC3.6.3.28
Mercury(II) reductase EC1.16.1.1
Organomercury lyase EC4.99.1.2
Beta-lactamase EC3.5.2.6
dTDP-4-dehydrorhamnose reductase EC1.1.1.133
(S)-mandelate dehydrogenase EC1.1.99.31
4-hydroxy-2-oxovalerate aldolase EC4.1.3.39
Carbazole 1,9a-dioxygenase EC1.14.12.22
Glutathione transferase EC2.5.1.18
Salicylate 1-monooxygenase EC1.14.13.1
Protocatechuate 4,5-dioxygenase EC1.13.11.8
Carbohydrate metabolism
Energy metabolism
Amino acid metabolism
Lipid metabolism
Biosynthesis of secondary metabolites
Xenobiotics biodegradation and metabolism
Other functions
-1 0 1 2 3
C4-1 C3-1 C2-1 C1-1 C3-1 C4-90 C3-90 C3-30 C2-90 C1-90 C3-60 C3-120

Figure 4



Microbial genetic potential for the biotransformation of xenobiotics and antibiotic resistance increases with depth during biofiltration