Applied Microbiology and Biotechnology
https://doi.org/10.1007/s00253-021-11756-3

ENVIRONMENTAL BIOTECHNOLOGY

Predicted functional genes for the biodegradation of xenobiotics in groundwater and sediment at two contaminated naval sites

Andrea Vera[1] · **Fernanda Paes Wilson**[1] · **Alison M. Cupples**[1]

Received: 6 August 2021 / Revised: 23 November 2021 / Accepted: 27 December 2021

Abstract

The goals of this study were to predict the genes associated with the biodegradation of organic contaminants and to examine microbial community structure in samples from two contaminated sites. The approach involved a predictive bioinformatics tool (PICRUSt2) targeting genes from twelve KEGG xenobiotic biodegradation pathways (benzoate, chloroalkane and chloroalkene, chlorocyclohexane and chlorobenzene, toluene, xylene, nitrotoluene, ethylbenzene, styrene, dioxin, naphthalene, polycyclic aromatic hydrocarbons, and metabolism of xenobiotics by cytochrome P450). Further, the predicted phylotypes associated with functional genes early in each pathway were determined. Phylogenetic analysis indicated a greater diversity in the sediment compared to the groundwater samples. The most abundant genera for sediments/microcosms included *Pseudomonas*, *Methylotenera*, *Rhodococcus*, *Stenotrophomonas*, and *Brevundimonas*, and the most abundant for the groundwater/microcosms included *Pseudomonas*, *Cupriavidus*, *Azospira*, *Rhodococcus*, and unclassified *Burkholderiaceae*. Genes from all twelve of the KEGG pathways were predicted to occur. Seven pathways contained less than twenty-five genes. The predicted genes were lowest for xenobiotics metabolism by cytochrome P450 and ethylbenzene biodegradation and highest for benzoate biodegradation. Notable trends include the occurrence of the first genes for trinitrotoluene and 2,4-dinitrotoluene degradation. Also, the complete path from toluene to benzoyl-CoA was predicted. Twenty-two of the dioxin pathway genes were predicted, including genes within the first steps. The following phylotypes were associated with the greatest number of pathways: unclassified *Burkholderiaceae*, *Burkholderia-Caballeronia-Paraburkholderia*, *Pseudomonas*, *Rhodococcus*, unclassified *Betaproteobacteria*, and *Polaromonas.* This work illustrates the value of PICRUSt2 for predicting biodegradation potential and suggests that a subset of phylotypes could be important for the breakdown of organic contaminants or their metabolites.

Key points

- *The approach is a low-cost alternative to shotgun sequencing.*
- *The genes and phylotypes encoding for xenobiotic degradation were predicted.*
- *A subset of phylotypes were associated with many pathways.*

Keywords Xenobiotic · PICRUSt2 · Nitrotoluene · Toluene · RDX

Introduction

Bioremediation is a widely used remediation approach for contaminated sites that greatly relies on microbial degradation. To determine the feasibility of this method, groundwater or sediment samples can be subject to quantitative PCR (qPCR) targeting the microorganisms and functional genes of interest (Fuller et al. 2010; Richards et al. 2019; Wilson and Cupples 2016). However, unless high-throughput qPCR is utilized, only a limited number of genes can be investigated. Another approach commonly adopted is 16S rRNA gene amplicon sequencing, providing information on which microorganisms are present (Dang and Cupples 2021; Ramalingam and Cupples 2020a; Wilson et al. 2016), with no information on which functional genes are present. Alternatively, shotgun sequencing provides both phylogenetic and functional information for site samples (Dang et al. 2018; Ramalingam and Cupples 2020b). Yet, shotgun sequencing

✉ Alison M. Cupples
cupplesa@egr.msu.edu

[1] Department of Civil and Environmental Engineering, Michigan State University, A135, 1449 Engineering Research Court, East Lansing, MI 48824, USA

Published online: 11 January 2022

is expensive and requires more advanced bioinformatics skills for data analysis.

In 2013, researchers developed the bioinformatics software package PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) to expand the investigative abilities of 16S rRNA gene amplicon sequencing (Langille et al. 2013). Specifically, the tool predicts functional profiles of microbial communities based on marker gene (16S rRNA gene) data. The tool was recently (2020) updated (PICRUSt2) and, according to the authors, contains an updated and larger database, provides interoperability with any operational taxonomic unit-picking or denoising algorithm, and supports phenotype predictions (Douglas et al. 2020). The authors validated their predictions using seven datasets generated from 16S rRNA gene and shotgun metagenomic sequencing (Douglas et al. 2020). One option for PICRUSt2 predictions includes the Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologs (KO) (Kanehisa and Goto 2000), with the total number of KOs in PICRUSt2 being 10,543 (Douglas et al. 2020). For those interested in the biodegradation of environmental contaminants, the KEGG pathways on xenobiotic biodegradation are particularly useful. For example, researchers have used PICRUSt or PICRUSt2 to make predictions on the xenobiotic degrading pathways from alkaline hot springs in India (Choure et al. 2021), sediment from contaminated and pristine locations from the Yucatan Peninsula (Navarrete-Euan et al. 2021), heavy-metal-contaminated soil from Korea (Hur and Park 2019), agricultural soils from Michigan exposed to pharmaceuticals and personal care products (Thelusmond et al. 2018), as well as soil and sediment samples from rivers and a site contaminated by hospital waste in Northeast India (DeMandal et al. 2019).

The current work expands on the use of PICRUSt2, adopting this method to examine groundwater and sediment from two contaminated sites. The work targets a subset of the KEGG xenobiotic degradation pathways using 16S rRNA gene sequencing data from a previous study that involved sequencing DNA from sediments, groundwater, and microcosms from two naval sites (Wilson and Cupples 2016). The previous study primarily focused on qPCR for the functional genes involved in RDX biodegradation with only a brief analysis of 16S rRNA gene data (phylotypes that increased in abundance in microcosms amended with RDX). Here, the overall composition of the microbial communities is examined. More importantly, PICRUSt2 (Douglas et al. 2020) was used to predict the abundance of genes in a subset of the KEGG xenobiotic biodegradation pathways (benzoate, chloroalkane and chloroalkene, chlorocyclohexane and chlorobenzene, toluene, xylene, nitrotoluene, ethylbenzene, styrene, dioxin, naphthalene, polycyclic aromatic hydrocarbons, and metabolism of xenobiotics by cytochrome P450) (Kanehisa and Goto 2000). Additionally, also using PICRUSt2, the current study predicted the phylotypes associated with the functional genes early in each pathway, as these are typically critical genes for the initiation of contaminant removal.

The current work determined the occurrence of these genes in sediment at different depths, in groundwater (upstream and downstream of a biobarrier), and in microcosms incubated over time (with or without RDX). Although the xenobiotic degrading pathways have previously been examined for various environmental samples, this work is novel because it is the first to provide a detailed analysis of the phylotypes linked to key pathway genes. Further, to our knowledge, this is the first attempt to use PICRUSt2 to examine these pathways in sediment over a range of depths at a contaminated site or in contaminated site groundwater. We hypothesize that certain phylotypes will be associated with multiple genes and will likely be good biomarkers for contaminant biodegradation potential. Additionally, it is likely that some pathways (e.g., benzoate) will have numerous genes detected and others will have less (e.g., dioxin), and this will vary over depth. An overall goal is to provide data to improve our understanding of biodegradation potential in sediment and groundwater at contaminated sites. As many contaminated sites contain a large number of co-contaminants, our aim was to understand the diversity of microorganisms and functional genes in situ regarding a wide number of contaminants.

Methods

Sites and samples

Previously, DNA was extracted and sequenced from sediment, groundwater, and laboratory microcosms (Wilson and Cupples 2016). For that study, sediment samples were collected from Naval Base Kitsap (RDX-contaminated site, Washington). The samples originated from two wells located on the site, hereafter called wells 58 and 61. Samples at five different depths (5, 10, 20, 25, and 30 ft deep) were collected from well 58, while samples at three different depths (5, 10, and 20 feet deep) were collected from well 61. The groundwater samples were collected from another RDX-contaminated naval site, the US Department of Defense explosives testing range in Virginia. Samples originated from a location 2.5 ft downstream, well 1 (hereafter called downstream well), and 10 ft upstream, well 10 (hereafter called upstream well), of a 100-ft installed buffered emulsified biobarrier. Microcosms were inoculated with groundwater and sediments, with or without the addition of RDX, and DNA was extracted at different times. Details on both sites, the microcosm studies, DNA extraction, and 16S rRNA gene amplicon sequencing were previously provided (Wilson and

Cupples 2016). The decision to use these samples was based on the availability of both groundwater and sediment from two contaminated sites.

R packages

In the current study, data analyses and the generation of all figures were achieved using the following R packages in R (version 4.0.4) (R Core Team 2018) within RStudio (version 1.1.456) (RStudio Team 2020): microbiome (version 1.10.0) (Lahti and Shetty 2017), phyloseq (version 1.32.0) (McMurdie and Holmes 2013), ampvis2 (version 2.6.5) (Andersen et al. 2018), ggplot2 (version 3.3.2) (Wickham 2016), ggpubr (version 0.4.0) (Kassambara 2020), vegan (version 2.5–7) (Oksanen et al. 2020), colourpicker (version 1.1.0.9000) (Attali 2021), readxl (version 1.3.1) (Wickham and Bryan 2019), rstatix (version 0.7.0) (Kassambara 2021), forcats (version 0.5.1) (Wickham 2021a), data.table (version 1.14.0) (Dowle and Srinivasan 2021), dplyr (version 1.0.6) (Wickham et al. 2021a, b), patchwork (version 1.1.1) (Pedersen 2020), tidyr (version 1.1.3) (Wickham 2021b), randomcoloR (version 1.1.0.1) (Ammar 2019), RColorBrewer (version 1.1–2) (Neuwirth 2014), circlize (version 0.4.13) (Gu et al. 2014), ComplexHeatmap (version 2.8.0) (Gu et al. 2016), and tidyverse (version 1.3.1) (Wickham et al. 2019). The R package versions and citations are not shown in the following text to improve clarity.

Microbial community analysis

In the current work, the amplicon sequencing data in the fastq file format was re-analyzed with Mothur (version 1.44.2) (Schloss 2009; Schloss et al. 2009) using the MiSeq Standard Operating Procedure (Kozich et al. 2013). The procedure included trimming the raw sequences and quality control. The database used for alignment was SILVA bacteria database (Release 138) for the V4 region (Pruesse et al. 2007). Chimeras, mitochondrial, and chloroplast lineage sequences were removed. Two Mothur generated files (shared file and taxonomy file) were combined with a metadata file using the package microbiome. The packages phyloseq and ggplot2 were used to rarefy the data and create the principle component analysis plots. Vegan was used to test for differences between microbial communities (between wells, treatments, time, depth) for the sediment and groundwater samples with Permutational Multivariate Analysis of Variance (PERMANOVA). Phyloseq was used to transform the rarified data to relative abundance values for plotting the bar charts by class. Bar charts were created separately for each well for both the sediment samples and the groundwater samples. The package ampvis2 was used to generate the heatmaps illustrating the most abundant phylotypes for each well. In each case, the package patchwork combined plots created a common legend and letter annotations.

PICRUSt2 analysis

PICRUSt2 (Douglas et al. 2020) was used to analyze Mothur generated files on the High Performance Computing Cluster (HPCC) at the Michigan State University. PICRUSt2 was applied with EPA-NG (Barbera et al. 2019) and gappa (Czech et al. 2020) for phylogenetic placement of reads, castor (Louca and Doebeli 2018) for hidden state prediction, and MinPath (Ye and Doak 2009) for pathway inference. The PICRUSt2 generated files for the KEGG genes (pred_metagenome_unstrat.tsv and pred_metagenome_contrib.tsv within the folder KO_metagenome_out) were examined for the presence of genes and phylotypes associated with the biodegradation of a group of xenobiotics. For this, data was first collected on a subset of KEGG xenobiotic degradation pathways (Kanehisa and Goto 2000) from the KEGG website (https://www.genome.jp/kegg/pathway.html). Specifically, gene numbers and descriptions were obtained from the following twelve pathways: benzoate, chloroalkane and chloroalkene, chlorocyclohexane and chlorobenzene, toluene, xylene, nitrotoluene, ethylbenzene, styrene, dioxin, naphthalene, polycyclic aromatic hydrocarbons, and metabolism of xenobiotics by cytochrome P450. Additionally, genes associated with methane and propane metabolism were selectively obtained from the KEGG website. The PICRUSt2 files were examined for the genes and associated phylotypes using R (version 4.0.4) (R Core Team 2018) with R studio (version 1.1.456) (RStudio Team 2020) and a number of R packages, as described below.

Heatmap and KEGG maps

The genes present in the sediment, groundwater, and microcosm samples were determined for each KEGG pathway. For this, the PICRUSt2 output file pred_metagenome_unstrat.tsv was unzipped and then combined with the appropriate KEGG file using the function inner_join (from dplyr). The R package ComplexHeatmap was then used to generate heatmaps of the 25 most abundant genes for each pathway for both the sediment and groundwater samples. For the cases when less than 25 genes were present for a pathway, all genes were included in the heatmaps. The package readxl was used to import the appropriate metafiles and dplyr (slice_max) was utilized to determine which genes had the highest mean value of the samples examined. Heatmap colors and break values were selected using the R package circlize. In addition, various annotations were added to each heatmap using the function HeatmapAnnotation from ComplexHeatmap. Following copyright permission, three pathways (nitrotoluene, toluene, dioxin) were selected for the creation of KEGG

map figures, with each displaying all genes present as well as the genes selected for phylotype examination (see below). During the generation of the heatmaps, a list of all the genes present beyond the 25 most abundant within each sample was compiled for each pathway.

Phylotypes associated with xenobiotic degrading genes

RStudio on the HPCC was used to generate a file that contained which genes and phylotypes were present using the PICRUSt2 output file pred_metagenome_contrib.tsv (unzipped). The approach involved combining this file with the KEGG pathway files and a taxonomy file (from Mothur), using the R packages data.table, dplyr, tidyr, ggplot2, and patchwork. The analysis involved only the sediment samples and sediment-inoculated microcosms. Only genes at the start of each pathway were selected for this analysis, as these are often the most important genes for contaminant biodegradation. For each pathway, four genes were selected for the creation of bar charts, faceted (in ggplot2) to illustrate the data over two wells, at different DNA extraction times and at different depths.

Genes enriched in RDX-amended microcosms

The final analysis involved determining if any genes were statistically significantly enriched in the RDX-amended microcosms compared to the no RDX-amended controls. For each pathway, the relative abundance of each gene was determined for each sample, and the data were combined with metadata (e.g., well number, depth, DNA extraction time, replicate). The Wilcoxon rank test was used to determine genes statistically more significant (higher relative abundance) in the RDX-amended microcosms compared to the microcosms not amended with RDX. Only the eight most abundant are shown for each pathway (if eight were statistically significantly different). If less than eight were significant, then all were included in the box and whisker plots. The plots were faceted (in ggplot2) to illustrate the data from different wells at different depths. The following packages were utilized for data manipulation and the creation of the plots: tidyverse, ggpubr, rstatix, forcats, readxl, forcats, colourpicker, and patchwork.

Results

The dominant microbial class in the majority of the sediment and groundwater microbial communities was *Gammaproteobacteria* (Figs. 1 and 2). This trend was particularly apparent in the groundwater samples, with eight of the ten communities being dominated by this class (Fig. 2). In the sediment samples, other important classes included *Actinobacteria*, *Bacteroidia*, *Alphaproteobacteria*, *Bacilli*, and *Clostridia* (Fig. 1). For well 58, no clear trends were observable by incubation time, treatment, or depth (Fig. 1A). Overall, the communities from well 61 microcosms illustrated less *Gammaproteobacteria* (Fig. 1B), for example, four of the five communities from 10 ft deep were dominated by *Bacteroidia*, *Bacilli*, or *Actinobacteria.* For well 61, there was also a trend of a larger percentage of *Bacilli* in the RDX amended samples from the Time 1 communities (but not Time 2). The upstream well-groundwater communities before incubation (Time 0) exhibited a number of dominant classes (*Actinobacteria*, *Alphaproteobacteria*, *Bacilli*, *Bacteroidia*, *Deltaproteobacteria*, and *Gammaproteobacteria*) (Fig. 2A). In contrast, the downstream well community before incubation (Time 0) was dominated only by *Gammaproteobacteria* (Fig. 2B). For both the upstream and downstream wells, when glucose was added, the classes observed were similar. When RDX was added with no glucose to the upstream well, the community became dominated by *Bacteroidia* and *Deltaproteobacteria* (Fig. 2A). However, this trend did not occur for the downstream well. Comparing all of the sediment to all of the groundwater microbial communities visually, one clear trend is a greater diversity in dominant classes in the sediment samples.

Principle component analysis (PCA) of the microbial communities indicated a separation of communities between the sediments and sediment-inoculated microcosms between the two wells (Fig. 3A), between different time points for the two wells (Fig. 3B), and between treatments (Fig. 3A and B). Separations were also noted individually for each well over depth (Fig. 3C and D), over time (Fig. 3E and F), and over treatment (Fig. 3C, D, E, and F). The differences between communities were confirmed with PERMANOVA (Table S1). Similarly, differences were observed via PCA and confirmed with PERMANOVA for the groundwater wells by well, treatment, and time (Fig. 4 and Table S2). The most abundant phylotypes across the samples and microcosms for sediments/microcosms for both wells 58 and 61 included *Pseudomonas*, *Methylotenera*, *Rhodococcus*, *Stenotrophomonas*, and *Brevundimonas* (Fig. 5). The most abundant for the groundwater/microcosms for both wells included *Pseudomonas*, *Cupriavidus*, *Azospira*, *Rhodococcus*, and unclassified *Burkholderiaceae* (Fig. 6).

All predicted genes (Tables S3 and S4), along with the twenty-five most abundant predicted genes, were determined for each pathway (Tables S3 and S4, Figs. 7 and 8). The heatmaps all illustrate a distinction between the most abundant compared to the rest. For example, the most abundant predicted genes from the benzoate biodegradation pathway included those encoding for acetyl-CoA C-acetyltransferase, enoyl-CoA hydratase, 3-hydroxybutyryl-CoA dehydrogenase, 4-oxalocrotonate tautomerase, glutaryl-CoA

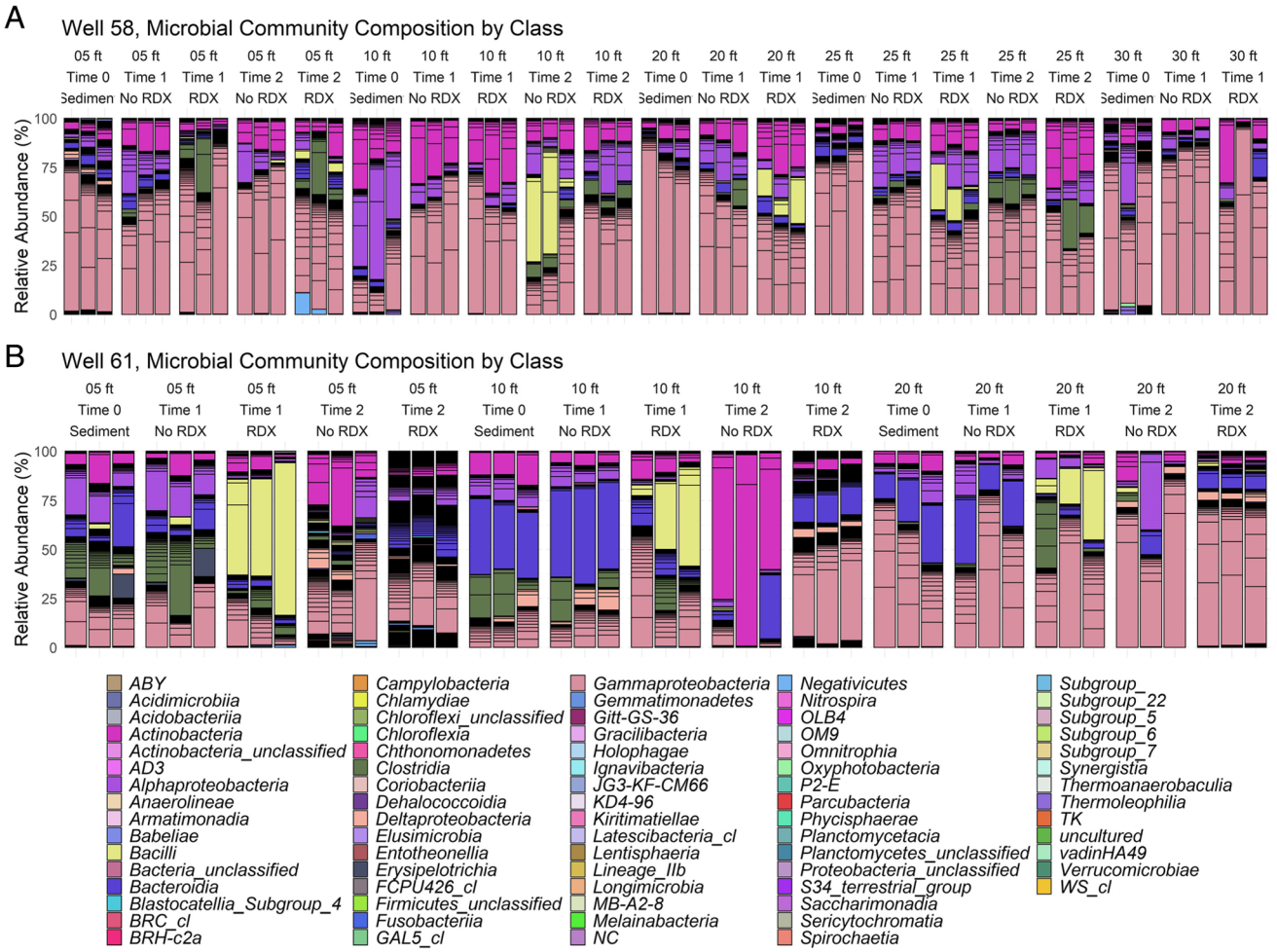


Fig. 1 Bar charts illustrating the microbial community composition (by class) over different sampling times and treatments for the sediment samples and microcosms from well 58 (**A**) and well 61 (**B**). For well 58, the results are from sediment at day 0 and from microcosms at days 90 and 130. For well 61, the results are from sediment at day 0 and from microcosms at days 45 and 90

dehydrogenase, and 4-carboxymuconolactone decarboxylase (Fig. 7A). Many of these predicted genes are part of the lower degradation pathway for benzoyl-CoA. The enzymes associated with the gene numbers shown in the heatmaps have been summarized (Tables S3 and S4). Seven pathways contained less than 25 predicted genes (Fig. 7B, D, E, and F and Fig. 8A, C, and E). The number of predicted genes was lowest for xenobiotics degradation by cytochrome P450 (Fig. 8E) and ethylbenzene degradation (Fig. 7D) and highest for degradation of benzoate (Tables S3 and S4). Overall, the trends between both wells and at different depths were similar. The heatmaps for the groundwater samples and microcosms are also shown (Figures S1 and S2).

Permission was granted to include three KEGG pathways (nitrotoluene, toluene, and dioxin degradation) in this publication (KEGG Copyright Permission 210,692) (Fig. 9). The nitrotoluene pathway was selected because of similarities in structure to RDX (a contaminant at both sites). Eighteen genes from the nitrotoluene pathway were predicted to be present (shown in pink and red; some boxes represent more than one gene, Fig. 9A). Notably, genes for the entire degradation pathway were not predicted to be present, indicating complete degradation of nitrotoluene may not occur. However, the first genes in the pathways of both trinitrotoluene and 2,4-dinitrotoluene were present, suggesting partial degradation of the parent contaminants is potentially feasible. More than thirty of the toluene degradation pathway genes were predicted to be present (Tables S3 and S4), including many of the genes in the first steps of the pathway (Fig. 9B). The complete path from toluene to benzoyl-CoA was predicted to occur, suggesting this contaminant would be susceptible to biodegradation. Twenty-two of the genes within the dioxin degradation pathway were predicted in the samples analyzed (Tables S3 and S4), including many of the genes encoding for the enzymes within the first steps (Fig. 9C).

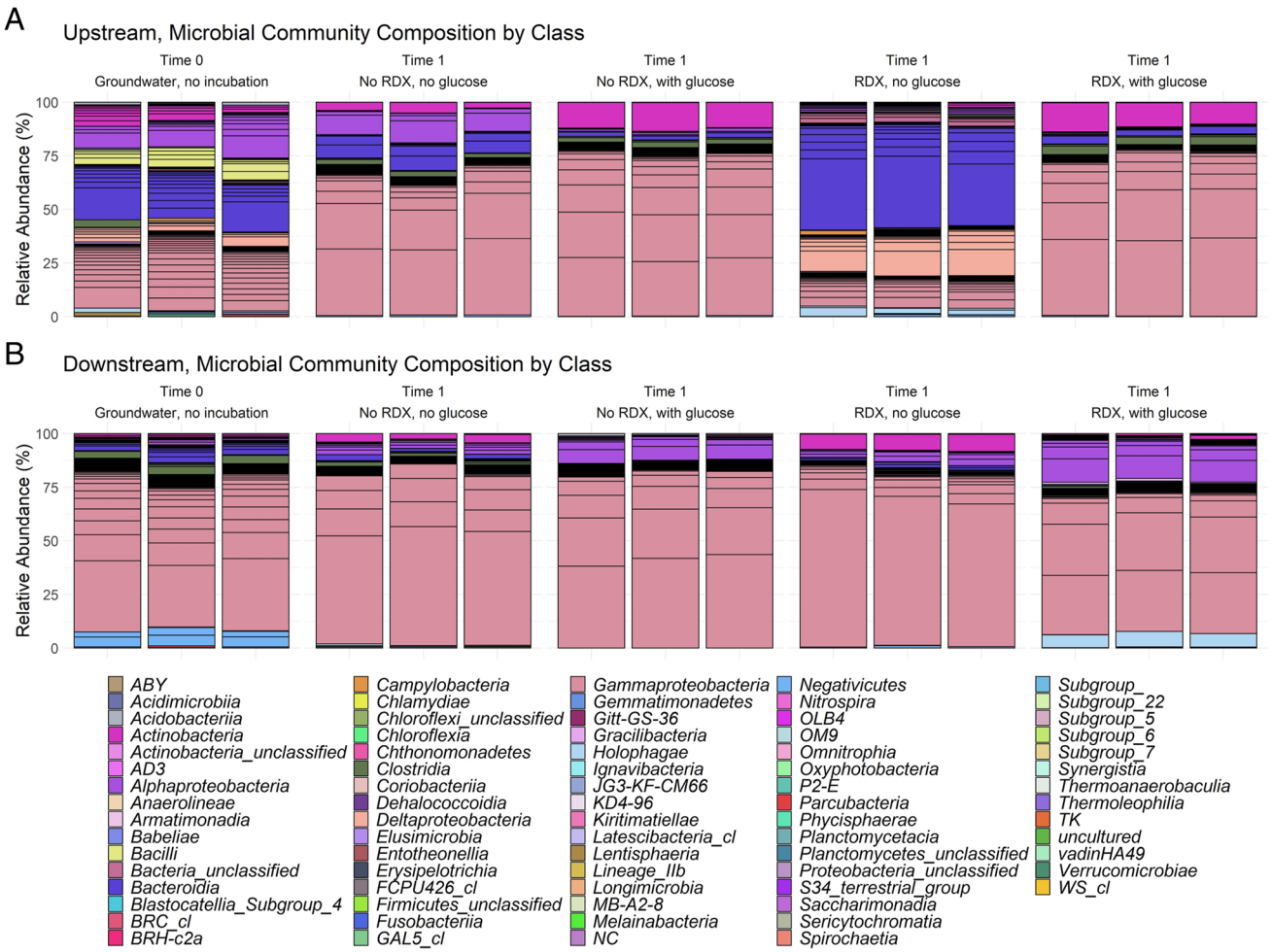


Fig. 2 Bar charts illustrating the microbial community composition (by class) over different sampling times and treatments for the groundwater samples and microcosms from the upstream well (**A**) and the downstream well (**B**). For the upstream well, the results are from sediment at day 0 and from microcosms at day 67. For the downstream well, the results are from sediment at day 0 and from microcosms at day 100

The PICRUSt2 generated files also enabled the identification of the phylotypes associated with the predicted genes. This analysis also involved the genes associated with methane and propane biodegradation (Figures S3 and S4), as these genes have been linked to the degradation of other contaminants (such as trichloroethene and 1,4-dioxane). Unclassified *Betaproteobacteria*, unclassified *Burkholderiaceae*, unclassified *Galliellales*, and *Rhodococcus* were primarily associated with propane 2-monooxygenase large subunit (Figures S3A, S4A). *Mycobacterium* was responsible for the detection of *mmoX* methane monooxygenase component A alpha chain (Figures S4B and S5B). Methane/ammonia monooxygenase subunit A was associated with *Roseiarcus* and unclassified *Betaproteobacteria* (Figures S3C, S4C).

The phylotype analyses for the other degradation pathways only focused on the early genes, as these are critical for initiating the removal of the parent contaminant. While a number of microorganisms were associated with these genes in each case, the following discussion highlights only those that were dominant. For each pathway, the trends of these phylotypes for each gene varied over time, depth, and between wells (Figs. 10, S5–S13). For nitrotoluene, unclassified *Enterobacteriaceae*, unclassified *Rhodocyclaceae*, unclassified *Burkholderiaceae*, *Burkholderia-Caballeronia-Paraburkholderia* (hereafter called *Burk-Caball-Paraburk*), *Polaromonas*, *Pseudomonas*, and *Achromobacter* were important (Fig. 10A). For toluene, *Geobacter*, unclassified *Burkholderiaceae*, unclassified *Rhodocyclaceae*, *Cupriavidus*, *Nevskia*, unclassified *Cellulomonadaceae*, and *Pseudomonas* were dominant (Fig. 10B). For the dioxin early genes, unclassified *Betaproteobacteria*, unclassified *Gammaproteobacteria*, *Rhodococcus*, unclassified *Bacteria*, *Burk-Caball-Paraburk*, and unclassified *Burkholderiaceae* were important (Fig. 10C).

For ethylbenzene biodegradation, the phylotypes JG30-KF-CM66 (from *Chloroflexi*), *Sphingomonas*,

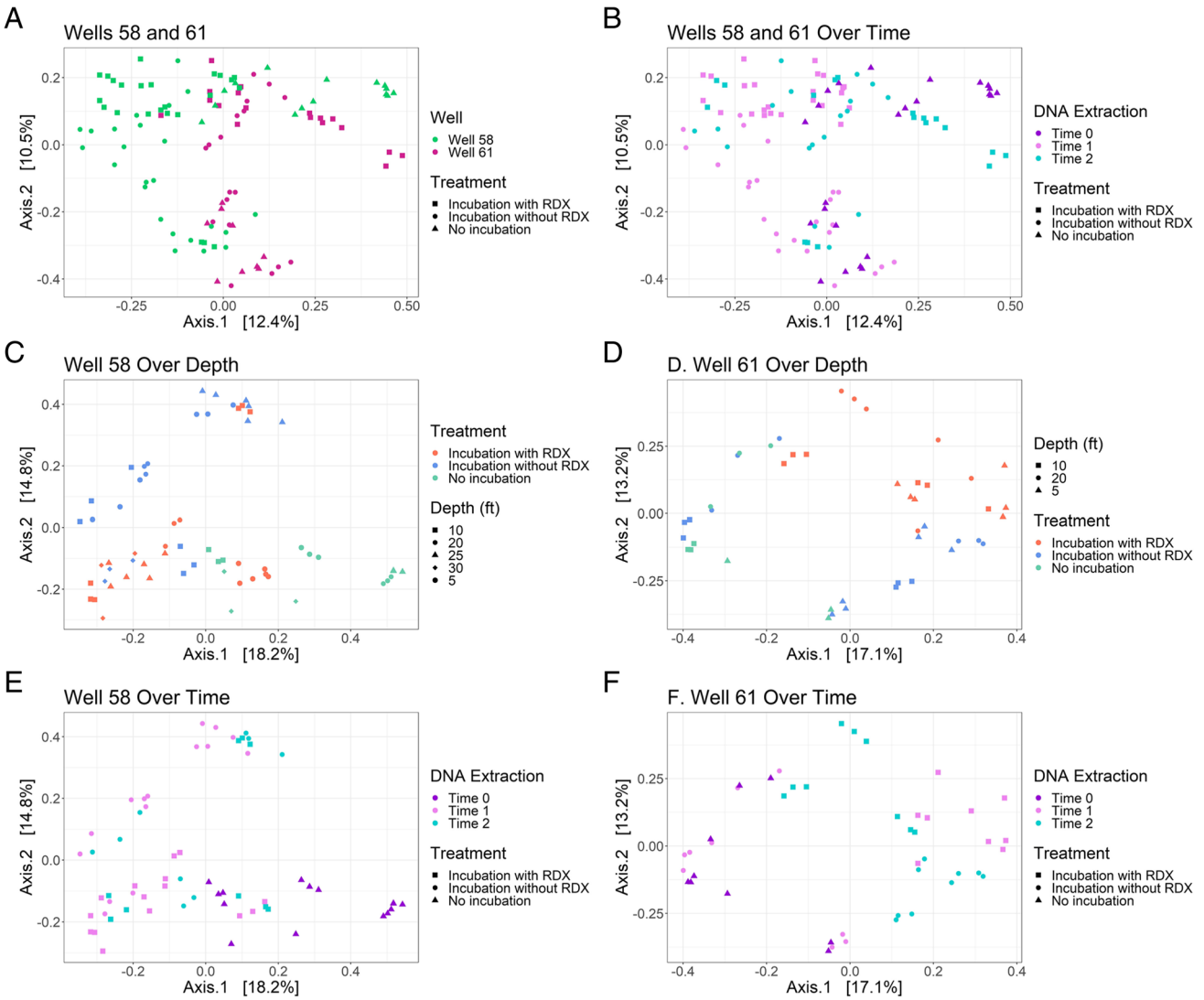


Fig. 3 Principle component analysis of the microbial communities from the sediments and sediment-inoculated microcosms over different sampling times and treatments for wells 58 and 61. For well 58, the results are from sediment at day 0 and from microcosms at days 90 and 130. For well 61, the results are from sediment at day 0 and from microcosms at days 45 and 90

unclassified *Sphingomonadaceae*, unclassified *Nocardiaceae*, unclassified *Corynebacteriales*, and *Rhodococcus* were important (Figure S5). For xylene, *Pseudomonas*, unclassified *Gammaproteobacteria*, unclassified *Betaproteobacteria*, unclassified *Burkholderiaceae*, unclassified *Nocardiaceae*, *Rhodococcus*, *Stenotrophomonas*, and *Burk-Caball-Paraburk* were associated with the first four genes (Figure S6). For styrene, unclassified *Betaproteobacteria*, unclassified *Gammaproteobacteria*, unclassified *Burkholderiaceae*, unclassified *Sphingomonadaceae*, *Sphingobium*, *Nocardioides*, unclassified *Microbacteriaceae*, and *Pseudomonas* were the key phylotypes (Figure S7). For the chloroalkane and chloroalkene pathway, unclassified *Galliellales*, *Rhodococcus*, *Xanthomonas*, unclassified *Xanthomonadaceae*, unclassified *Xanthobacteraceae*, unclassified *Burkholderiaceae*, *Pseudomonas*, unclassified *Rhodocyclaceae*, *Subgroup_6*, unclassified *Betaproteobacteria*, *Methylotenera*, and *Methylophilus* were associated with the early pathway genes (Figures S8). For benzoate, *Acinetobacter*, *Burk-Caball-Paraburk*, unclassified *Betaproteobacteria*, *Pseudomonas*, *Ralstonia*, unclassified *Burkholderiaceae*, and unclassified *Xanthobacteraceae* were important for the first genes (Figure S9). Phylotypes linked to the early PAH genes included *Polaromonas*, unclassified *Microbacteriaceae*, *Achromobacter*, and unclassified *Burkholderiaceae* (Figure S10). For the chlorocyclohexane and chlorobenzene early genes, unclassified *Galliellales*, *Rhodococcus*, *Xanthomonas*, unclassified *Xanthomonadaceae*, unclassified

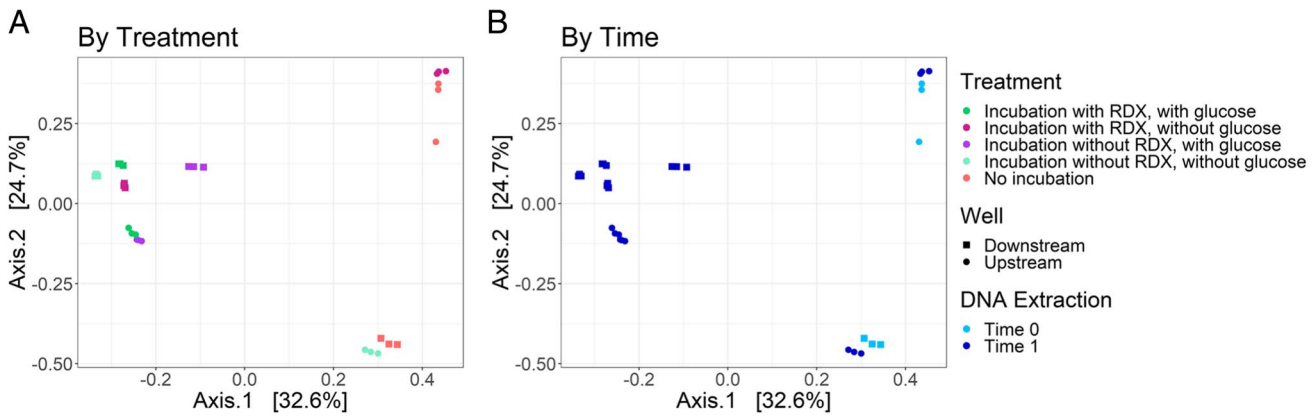


Fig. 4 Principle component analysis of the microbial communities from the groundwater and groundwater-inoculated microcosms over different sampling times and treatments for the upstream and downstream wells. For the upstream well, the results are from sediment at day 0 and from microcosms at day 67. For the downstream well, the results are from sediment at day 0 and from microcosms at day 100

Burkholderiaceae, *Stenotrophomonas*, unclassified *Rhodocyclaceae*, *Burk-Caball-Paraburk*, and unclassified *Betaproteobacteria* were dominant (Fig. 11S). The pathway involving the degradation of xenobiotics by cytochrome P450 primarily involved unclassified *Burkholderiaceae*, *Burk-Caball-Paraburk*, unclassified *Bacteria*, unclassified *Rhizobiaceae*, and *Rhodococcus* (Figure S12). Finally, for naphthalene, *Achromobacter*, unclassified *Burkholderiaceae*, *Polaromonas*, unclassified *Betaproteobacteria*, *Duganella*, and *Novosphingobium* were important (Figure S13). Considering all of the twelve KEGG pathways examined here together, the following phylotypes were associated with the greatest number of pathways (the number of pathways are in parenthesis), unclassified *Burkholderiaceae* (eleven), *Burkholderia-Caballeronia-Paraburkholderia* (six), *Pseudomonas* (six), *Rhodococcus* (six), unclassified *Betaproteobacteria* (five), and *Polaromonas* (three).

RDX positively impacted the relative abundance of some genes in each pathway (Figures S14–S17). For the nitrotoluene pathway, *nemA*, N-ethylmaleimide reductase (K10680) exhibited the highest relative abundance of the genes predicted to be statistically significantly positively impacted (Figure S17A). For toluene, the predicted abundance of carboxymethylenebutenolidase (K01061) was particularly higher in RDX-amended samples compared to those not amended with RDX (Figure S17B). For the dioxin pathway, two genes were predicted to be markedly more abundant (acetaldehyde dehydrogenase [K04073] and 2-oxo-3-hexenedioate decarboxylase [K01617]) in the RDX-amended samples compared to the no RDX microcosms (Figure S17C). The statistically significantly impacted genes in the other nine pathways are shown in the supplementary section (Figures S14–S16).

Discussion

The current work utilized a newly updated bioinformatics tool (PICRUSt2) (Douglas et al. 2020) to examine the functional genes involved in organic contaminant biodegradation and their associated phylotypes in contaminated site samples and microcosms. Genes from each of the twelve KEGG xenobiotic degradation pathways investigated were predicted to be present, although not all of the genes for each pathway were predicted. The approach offers a low-cost alternative to shotgun sequencing for examining the potential for the biodegradation of organic contaminants in environmental samples. The data generated are novel as this is the first report predicting genes from these pathways over a range of sediment depths at a contaminated site. Further, the research provides an in-depth analysis of the phylotypes associated with the first genes in each of the twelve pathways for each depth and well investigated.

It is important to note that others have also examined the KEGG xenobiotic degradation pathways using other approaches. For example, a metatranscriptome shotgun sequencing study of wheat rhizosphere reported transcripts of sixteen different enzymes from six xenobiotic degradation pathways (chlorocyclohexane and chlorobenzene, benzoate, aminobenzoate, nitrotoluene, styrene, and caprolactam) (Singh et al. 2018). Another metatranscriptome project reported a larger number (twenty-one) of the KEGG xenobiotic biodegradation pathways in soil samples from a wastewater-contaminated mangrove, with the most abundant being benzoate degradation, chloroalkane and chloroalkene degradation, drug metabolism—other enzymes, and naphthalene degradation (Isaza et al. 2021). The main genera associated with the most abundant genes were *Exiguobacterium*, *Bacillus*, *Rhodopseudomonas*, *Amphibacillus*, and *Nocardioides*.

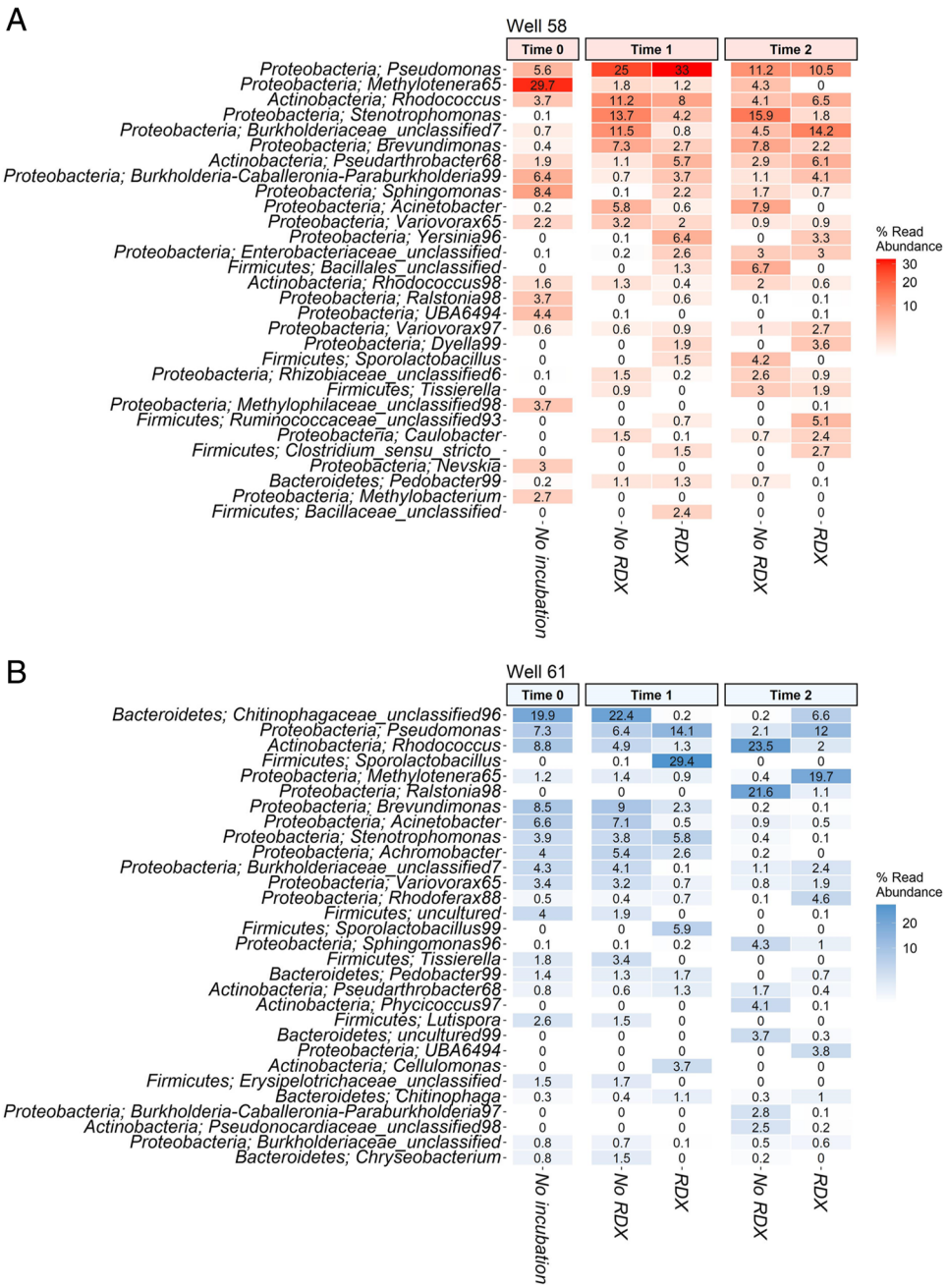


Fig. 5 Heatmaps of the most abundant genera across samples from well 58 (**A**) and well 61 (**B**) over different sampling times and treatments. The phylum for each genus is also shown. For well 58, the results are from sediment at day 0 and from microcosms at days 90 and 130. For well 61, the results are from sediment at day 0 and from microcosms at days 45 and 90

Others have adopted metagenomics approaches to examine KEGG xenobiotic degradation pathways. For example, twenty-one pathways were detected in a semi-arid mangrove in Colombia (Munoz-Garcia et al. 2019). The researchers reported that sixteen of these pathways were influenced by salinity. Comparing the abundance of each pathway, the degradation of benzoate showed the highest abundance, followed by the drug metabolism and the degradation of chloroalkanes and chloroalkenes. Another shotgun metagenomic sequencing study (from cave soil samples) detected genes from twenty-six pathways classifying as xenobiotics biodegradation and metabolism, among them, benzoate degradation via CoA ligation being the most abundant (Wiseschart et al. 2019). Further, seventeen KEGG xenobiotic degrading pathways were identified during a metagenomic analysis of a commercial scale biofilter (Li et al. 2019). Another paper reported eighteen KEGG xenobiotic degradation pathways in bacterial communities in a soda lake in India before and after monsoon using shotgun sequencing and reported nitrotoluene and benzoate degradation pathway as being predominant (Chakraborty et al. 2021).

Previously, our group used shotgun sequencing to examine the most abundant phylotypes associated with a subset of xenobiotic degrading KEGG pathways in four agricultural

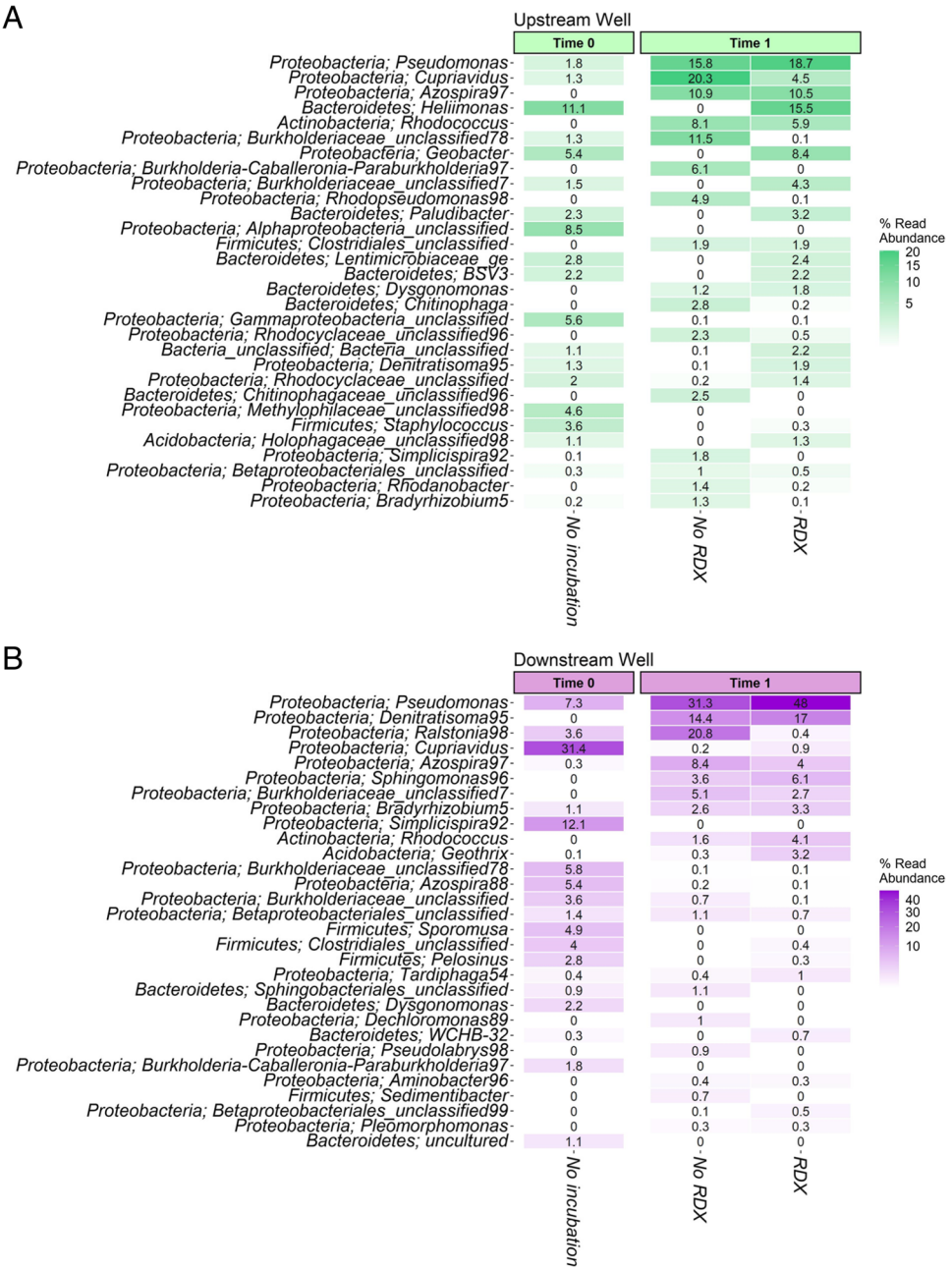


Fig. 6 Heatmaps of the most abundant genera across samples from the groundwater upstream well (**A**) and downstream well (**B**) over different sampling times and treatments. The phylum for each genus is also shown. For the upstream well, the results are from sediment at day 0 and from microcosms at day 67. For the downstream well, the results are from sediment at day 0 and from microcosms at day 100

soils (Thelusmond et al. 2019) using the first version of PICRUSt (Langille et al. 2013). In that work, the most abundant phylotypes encoding for the genes involved in benzoate degradation were *Rhodopseudomonas*, *Polaromonas*, *Cupriavidus*, *Bradyrhizobium*, *Burkholderia*, *Pseudomonas*, and *Ralstonia*. These results are partially consistent with the current study, which documented the importance of *Burkholderia*, *Pseudomonas*, *Ralstonia*, and *Cupriavidus* for the genes early in the benzoate degradation pathway. Previously, for the chlorocyclohexane and chlorobenzene biodegradation pathway genes, three phylotypes (*Bradyrhizobium*, *Candidatus Solibacter*, and *Burkholderia*) dominated (Thelusmond et al. 2019). Overlapping phylotypes for this pathway between the previous and current study include *Burkholderia* and *Rhodococcus*. In the previous study, the phylotypes associated with dioxin biodegradation primarily involved *Bradyrhizobium*, *Polaromonas*, *Aromatoleum*, and *Rhizobium* (Thelusmond et al. 2019). In the current study, for this pathway, unclassified *Betaproteobacteria*, unclassified *Gammaproteobacteria*, *Rhodococcus*, unclassified *Bacteria*, *Burk-Caball-Paraburk*, and unclassified *Burkholderiaceae* were important. Notably, *Rhodococcus* and *Burkholderia* were also associated with the dioxin pathway genes in the previous study (to a lesser extent that the four

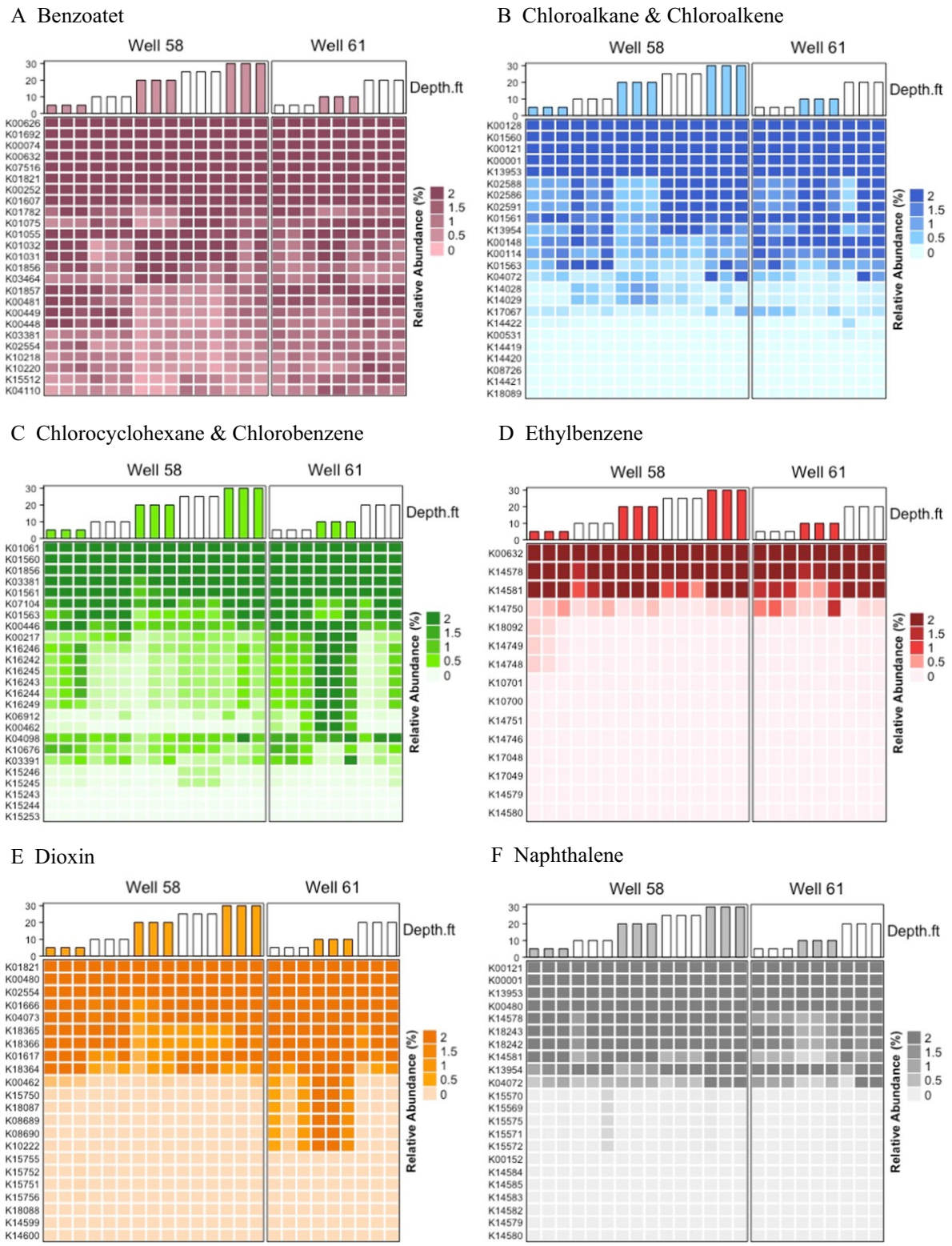


Fig. 7 Heatmaps of the most abundant genes present in each pathway for the microcosms inoculated with sediment for well 58 or well 61 incubated with (solid bar) and without RDX (white bar). Data are shown for triplicate samples

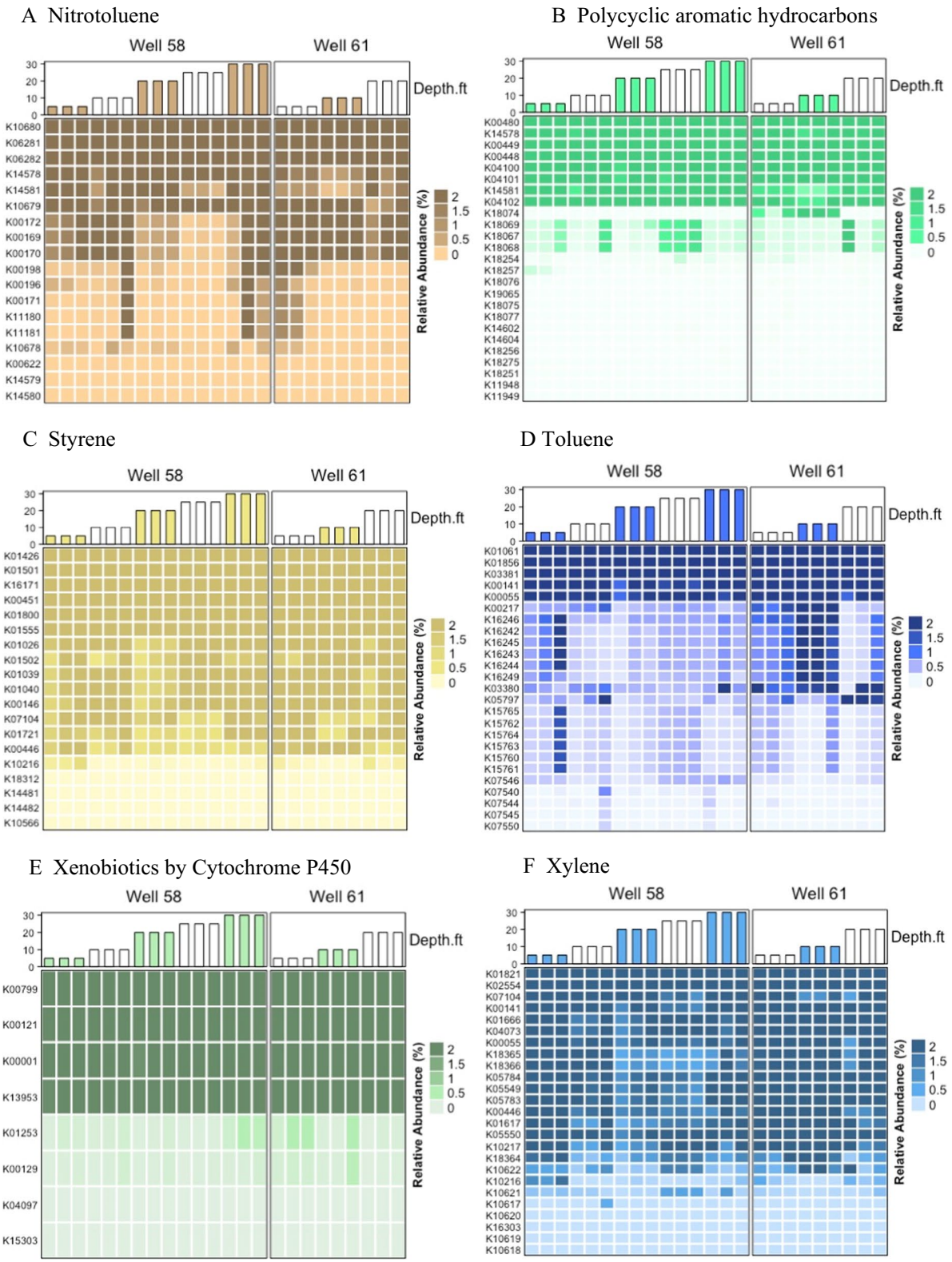


Fig. 8 Heatmaps of the most abundant genes present in each pathway for the microcosms inoculated with sediment for well 58 or well 61 incubated with (solid bar) and without RDX (white bar). Data are shown for triplicate samples

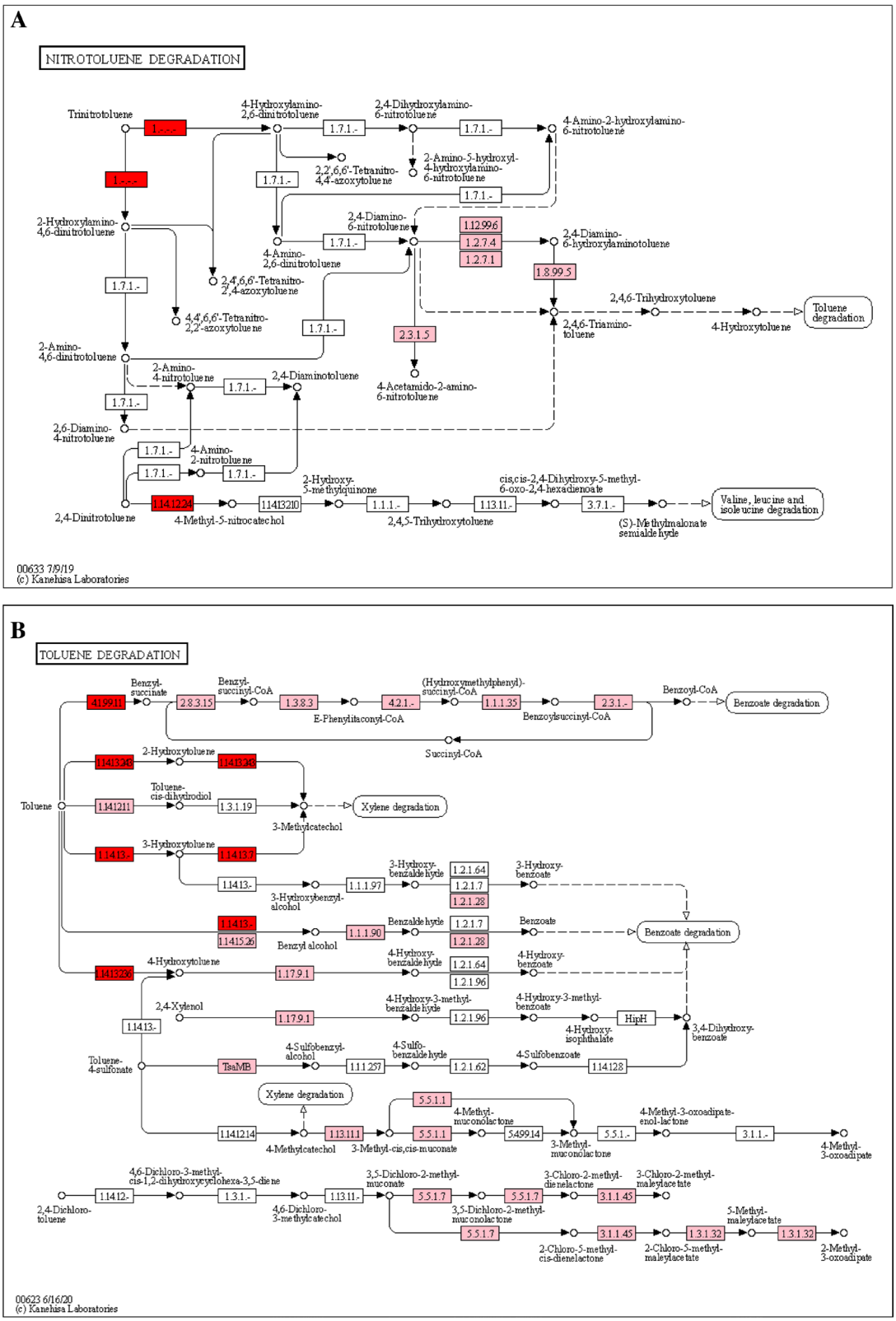


Fig. 9 **A** Genes present (pink) in all samples and those analyzed for phylotypes (red) for the KEGG nitrotoluene degradation pathway (used with permission [KEGG Copyright Permission] 210,692). **B** Genes present (pink) in all samples and those analyzed for phylotypes (red) for the KEGG toluene degradation pathway (used with permission [KEGG Copyright Permission] 210,692). **C** Genes present (pink) in all samples and those analyzed for phylotypes (red) for the KEGG dioxin degradation pathway (used with permission [KEGG Copyright Permission] 210,692)

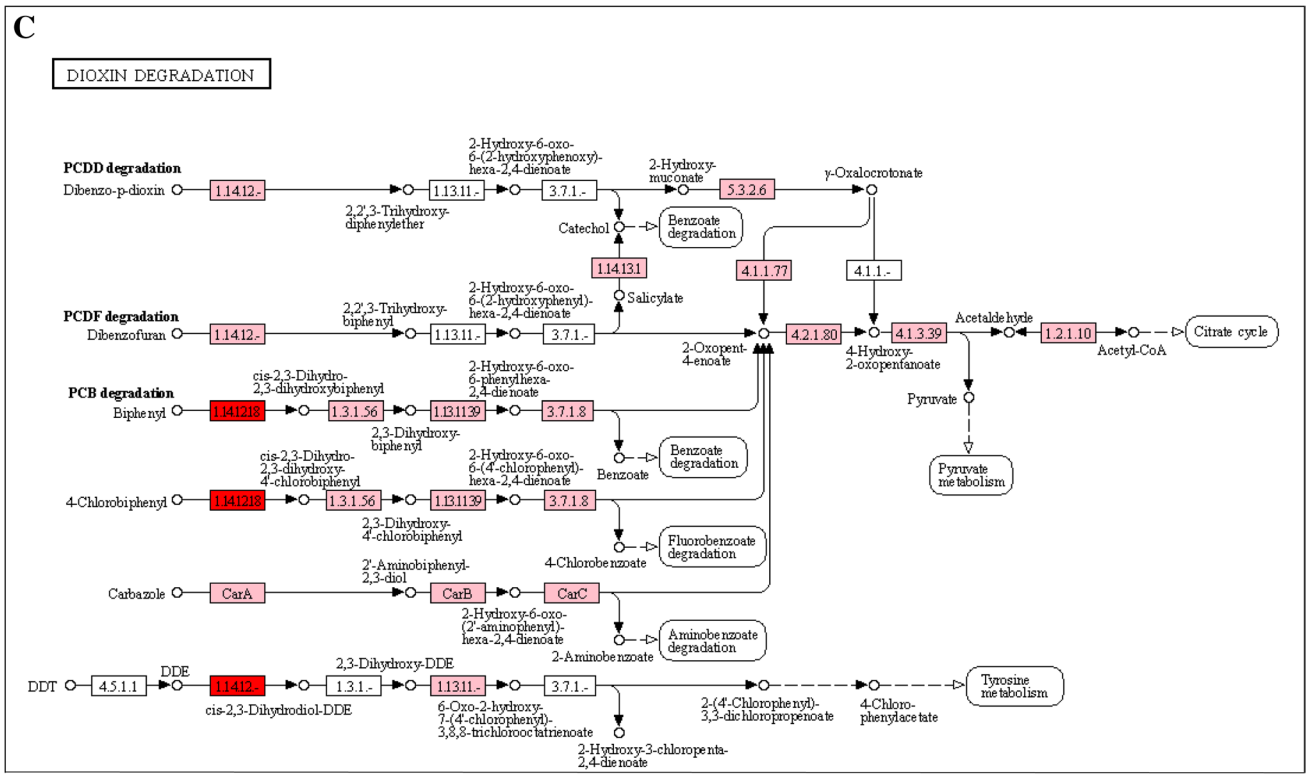


Fig. 9 (continued)

noted above). The fourth common pathway between the previous and current study was for nitrotoluene. In our previous work, *Cupriavidus*, *Candidatus Solibacter*, *Mycobacterium*, *Streptomyces*, *Rhodococcus*, *Methylobacillus*, and *Frankia* were the dominant phylotypes for that pathway (Thelusmond et al. 2019). While in the current study, for nitrotoluene, unclassified *Enterobacteriaceae*, *Rhodocyclaceae*, *Burkholderiaceae*, *Burk-Caball-Paraburk*, *Polaromonas*, *Pseudomonas*, and *Achromobacter* were important. Differences between the results of the two studies for these four pathways could be attributed to the data analysis approach (considering all genes, instead of only four), the sequencing/analysis approach (shotgun vs. 16S rRNA gene/PICRUSt2), the version of PICRUSt, or the samples themselves (agricultural soil vs. contaminated site sediments).

The current study detected more genes compared to the previous shotgun sequencing study (Thelusmond et al. 2019) for each of the four common KEGG pathways examined. Previously, forty-two genes associated with the benzoate degradation pathway were detected (Thelusmond et al. 2019), compared to eighty-eight in the current work. Previously, only three genes (biphenyl 2,3-dioxygenase subunit alpha, cis-2,3-dihydrobiphenyl-2,3-diol dehydrogenase, salicylate hydroxylase) associated with dioxin biodegradation were detected (Thelusmond et al. 2019), and in current study, twenty-two were detected (including the three just listed). Similarly, fifteen and twenty-four genes were associated with chlorocyclohexane and chlorobenzene metabolism, in previous (Thelusmond et al. 2019) and current work, respectively. Both projects detected the gene encoding carboxymethylenebutenolidase as being the most abundant. Only six genes were previously detected as part of the nitrotoluene degradation pathway (Thelusmond et al. 2019) compared to eighteen in the current study. The top three genes detected from both projects were the same (*hyaB*, *hybC*; hydrogenase large subunit, *hyaA*, *hybO*; hydrogenase small subunit, *nemA*; N-ethylmaleimide reductase). The differences in genes detected are likely a result of the different sequencing and analysis techniques (shotgun vs. 16S rRNA gene/PICRUSt2).

As discussed above, others have used PICRUSt or PICRUSt2 to investigate the importance of xenobiotic degrading pathways (Choure et al. 2021; Hur and Park 2019; Navarrete-Euan et al. 2021; Thelusmond et al. 2018). To our knowledge, only limited efforts have been made to identify the phylotypes associated with the specific pathway genes. One such example examined soil and sediment samples from three rivers and one dumping site contaminated by hospital domestic waste in Northeast India using 16S rRNA gene amplicon sequencing (DeMandal et al. 2019) and the first PICRUSt release tool (Langille et al. 2013). They found a smaller number of xenobiotic degradation genes for each

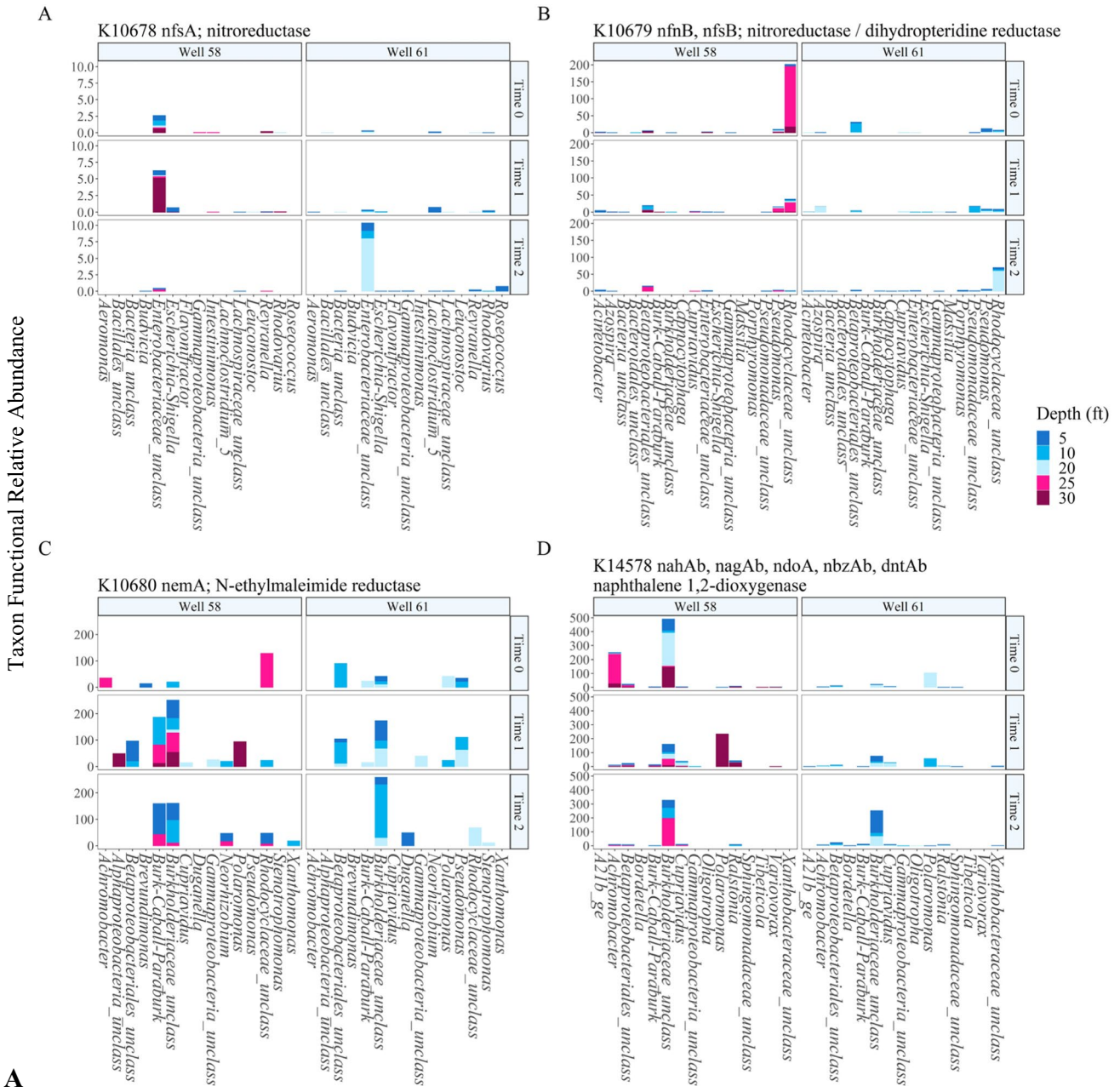


Fig. 10 **A** Predicted taxon functional relative abundance for four early genes in the KEGG nitrotoluene degradation pathway for sediment and microcosms inoculated with sediment from two wells at different depths. For well 58, the results are from sediment at day 0 and from microcosms at days 90 and 130. For well 61, the results are from sediment at day 0 and from microcosms at days 45 and 90. For well 61, sediment was only collected and examined from 5, 10, and 20 ft. **B** Predicted taxon functional relative abundance for four early genes in the KEGG toluene degradation pathway for sediment and microcosms inoculated with sediment from two wells at different depths. For well 58, the results are from sediment at day 0 and from microcosms at days 90 and 130. For well 61, the results are from sediment at day 0 and from microcosms at days 45 and 90. For well 61, sediment was only collected and examined from 5, 10, and 20 ft. **C** Predicted taxon functional relative abundance for four early genes in the KEGG dioxin degradation pathway for sediment and microcosms inoculated with sediment from two wells at different depths. For well 58, the results are from sediment at day 0 and from microcosms at days 90 and 130. For well 61, the results are from sediment at day 0 and from microcosms at days 45 and 90. For well 61, sediment was only collected and examined from 5, 10, and 20 ft

pathway, as follows: benzoate (eight), naphthalene (twelve), chloroalkane and chloroalkene (twelve), chlorocyclohexane and chlorobenzene (eight), and nitrotoluene (twelve), and toluene (eight), in the soil and sediment samples (DeMandal et al. 2019). The lower number of genes may be related to the version of PICRUSt used (version 1, compared to

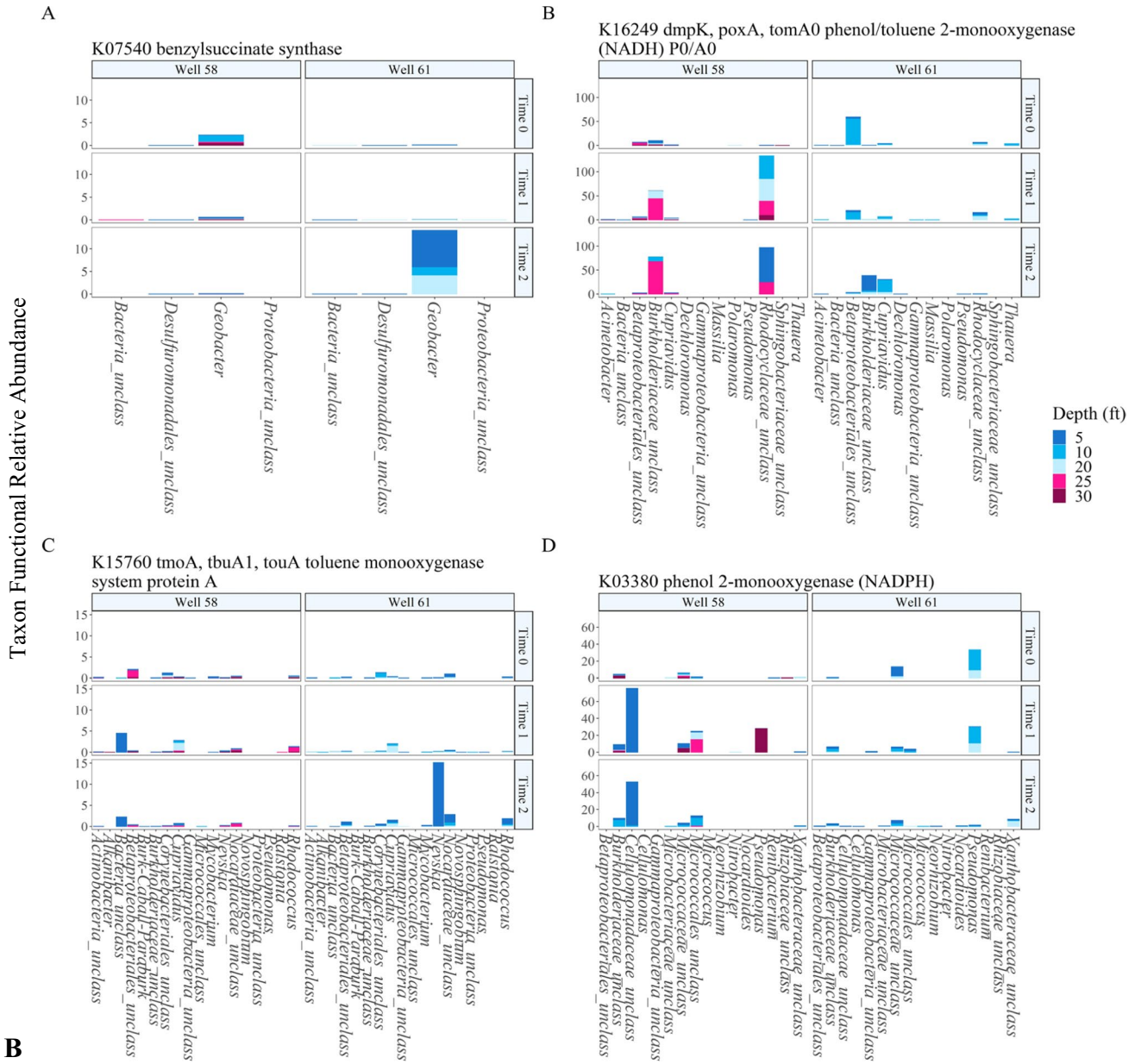
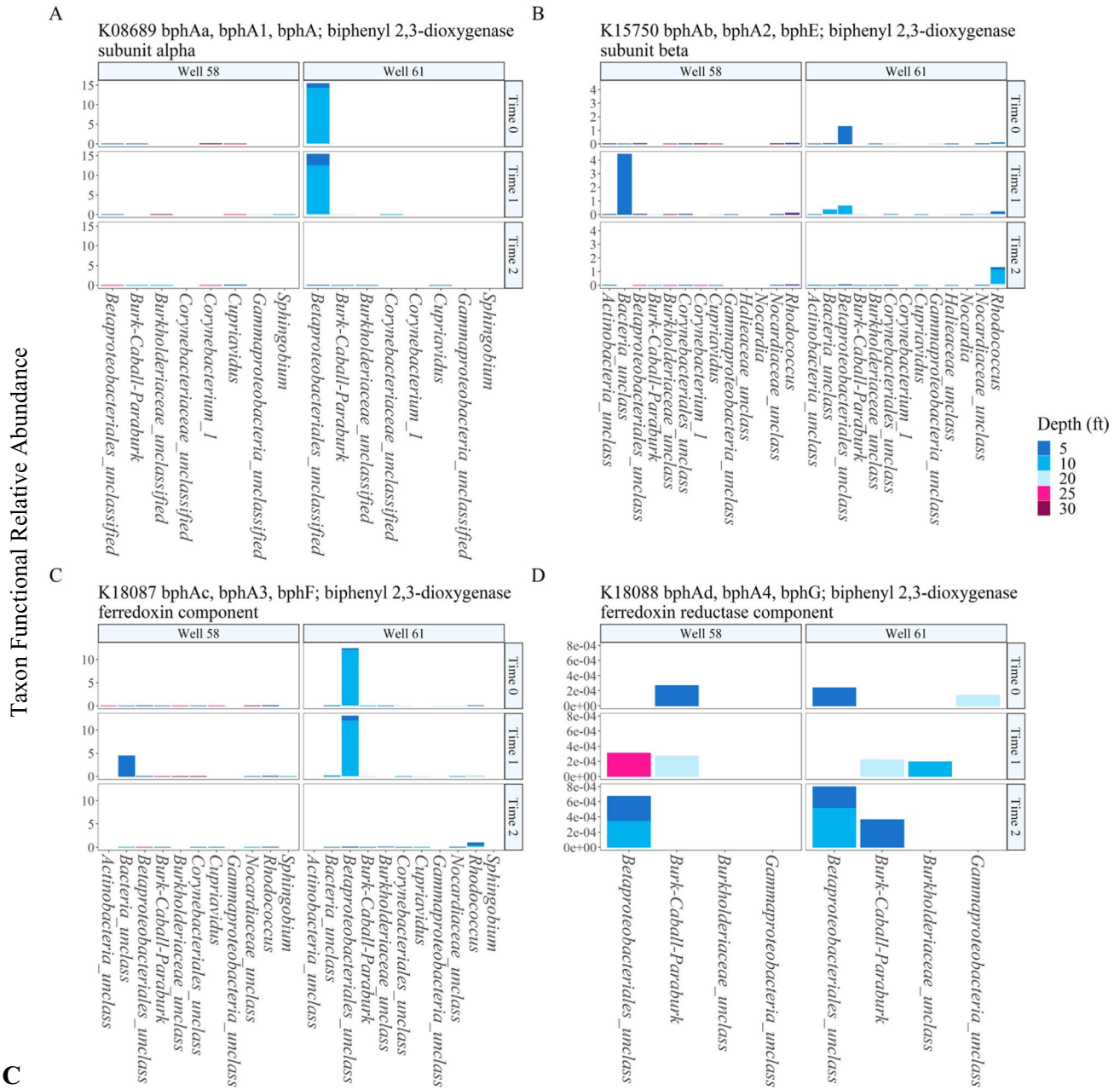


Fig. 10 (continued)

PICRUSt2 in the current study), the sequencing analysis approach (QIIME (Caporaso et al. 2010) compared to Mothur (Schloss 2009) or the sequencing depth (averages of ~ 280,000 compared to ~ 500,000 reads). The researchers listed the phylotypes associated with the genes involved in the following KEGG xenobiotic degradation pathways: benzoate, aminobenzoate, naphthalene, and fluorobenzoate degradation. We can only compare the results from the current study for benzoate and naphthalene biodegradation pathways, as these are the only two common pathways. In their study, the following phylotypes were associated with benzoate degradation: *Acinetobacter*, *Arthrobacter*, *Comamonas*, *Diaphorobacter*, *Geobacter*, *Novispirillum*, *Phenylobacterium*, *Pseudoxanthomonas*, and *Rhodoplanes*. In the current study, for the first four genes examined for benzoate degradation, from these, only *Acinetobacter* and *Geobacter* were detected. They also associated the following with naphthalene degradation: *Arthrobacter*, *Comamonas*, *Diaphorobacter*, *Geobacter*, *Klebsiella*, *Leptothrix*, and *Novosphingobium.* In the current study, for the first four genes examined for naphthalene degradation, from these, only *Novosphingobium* was detected.

In summary, this work predicted the occurrence of genes involved in the biodegradation of a range of organic



C

Fig. 10 (continued)

contaminants. The approach is a low-cost alternative for investigating genes associated with contaminant biodegradation and could be adopted for any site investigation. However, it is important to note that the results of this approach are only predictions and will therefore not be as robust as the results from a metatranscriptomics study. Nevertheless, the research provides key insights as to which microorganisms may be the most important for in situ bioremediation of this set of contaminants and/or their degradation products. Given the lack of previous research linking specific genes to specific phylotypes at contaminated sites for these pathways, the work fills a key knowledge gap in the current literature.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00253-021-11756-3.

Acknowledgements Thanks to Mark E. Fuller and Paul Hatzinger (CB&I Federal Services) for providing groundwater samples. Thanks to David Liu (Naval Facilities Engineering Command Northwest) and Mandy Michalsen (USACE, Seattle District) for providing the sediment samples.

Author contribution AV: Writing, editing, methodology, programming, formal analysis, and investigation. FW: Methodology, investigation, and editing. AC: Conceptualization, resources, data curation, writing (review and editing), visualization, supervision, project administration, and funding acquisition.

Funding Strategic Environmental Research and Development Program (SERDP Project ER1606), Michigan State University College of Engineering ENSURE program and NSF (grant number 1902250).

Availability of data Illumina sequencing data were deposited in the NCBI Sequence Read Archive under Bioproject Number PRJ302752.

Declarations

Ethical statement This work was supported by the College of Engineering ENSURE program, SERDP, and NSF.

Conflict of interest The authors declare no competing interests.

Human and animal rights and informed consent This article does not contain any studies with human participants or animals performed by any of the authors.

References

Ammar R (2019) randomcoloR: generate attractive random colors. https://CRAN.R-project.org/package=randomcoloR

Andersen KS, Kirkegaard RH, Karst SM, Albertsen M (2018) ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. bioRxiv. https://doi.org/10.1101/299537

Attali D (2021) Colourpicker: a colour picker tool for shiny and for selecting colours in plots . https://github.com/daattali/colourpicker

Barbera P, Kozlov AM, Czech L, Morel B, Darriba D, Flouri T, Stamatakis A (2019) EPA-ng: massively parallel evolutionary placement of genetic sequences. Syst Biol 68(2):365–369. https://doi.org/10.1093/sysbio/syy054

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7(5):335–336. https://doi.org/10.1038/nmeth.f.303

Chakraborty J, Rajput V, Sapkale V, Kamble S, Dharne M (2021) Spatio-temporal resolution of taxonomic and functional microbiome of Lonar soda lake of India reveals metabolic potential for bioremediation. Chemosphere 264:128574. https://doi.org/10.1016/j.chemosphere.2020.128574

Choure K, Parsai S, Kotoky R, Srivastava A, Tilwari A, Rai PK, Sharma A, Pandey P (2021) Comparative metagenomic analysis of two alkaline hot springs of Madhya Pradesh, India and deciphering the extremophiles for industrial enzymes. Front Genet 12:643423. https://doi.org/10.3389/fgene.2021.643423

Czech L, Barbera P, Stamatakis A (2020) Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. Bioinformatics 36(10):3263–3265. https://doi.org/10.1093/bioinformatics/btaa070

Dang H, Cupples AM (2021) Identification of the phylotypes involved in cis-dichloroethene and 1,4-dioxane biodegradation in soil microcosms. Sci Total Environ 794:148690. https://doi.org/10.1016/j.scitotenv.2021.148690

Dang H, Kanitkar YH, Stedtfeld RD, Hatzinger PB, Hashsham SA, Cupples AM (2018) Abundance of chlorinated solvent and 1,4-dioxane degrading microorganisms at five chlorinated solvent contaminated sites determined via shotgun sequencing. Environ Sci Technol 52:13914–13924. https://doi.org/10.1021/acs.est.8b04895

DeMandal S, Mathipi V, Muthukumaran RB, Gurusubramanian G, Lalnunmawii E, Kumar NS (2019) Amplicon sequencing and imputed metagenomic analysis of waste soil and sediment microbiome reveals unique bacterial communities and their functional attributes. Environ Monit Assess 191:778

Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, Huttenhower C, Langille MGI (2020) PICRUSt2 for prediction of metagenome functions. Nat Biotechnol 38(6):685–688. https://doi.org/10.1038/s41587-020-0548-6

Dowle M, Srinivasan A (2021) data.table: extension of 'data.frame'. R package version 1.14.0. https://CRAN.R-project.org/package=data.table

Fuller ME, McClay K, Higham M, Hatzinger PB, Steffan RJ (2010) Hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) bioremediation in groundwater: are known RDX-degrading bacteria the dominant players? Bioremed J 14(3):121–134. https://doi.org/10.1080/10889868.2010.495367

Gu Z, Eils R, Schlesner M (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32(18):2847–2849. https://doi.org/10.1093/bioinformatics/btw313

Gu Z, Gu L, Eils R, Schlesner M, Brors B (2014) circlize Implements and enhances circular visualization in R. Bioinformatics 30(19):2811–2812. https://doi.org/10.1093/bioinformatics/btu393

Hur M, Park SJ (2019) Identification of microbial profiles in heavy-metal-contaminated soil from full-length 16S rRNA reads sequenced by a PacBio system. Microorganisms 7(9). https://doi.org/10.3390/microorganisms7090357

Isaza JP, Sandoval-Figueredo V, Rodelo MC, Muñoz-García A, Figueroa-Galvis I, Vanegas J (2021) Metatranscriptomic characterization of the bacterial community of a contaminated mangrove from the Caribbean. Reg Sudies Mar Sci 44(101724)

Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28(1):27–30. https://doi.org/10.1093/nar/28.1.27

Kassambara A (2020) ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. https://CRAN.R-project.org/package=ggpubr

Kassambara A (2021) rstatix: pipe-friendly framework for basic statistical tests R package version 0.7.0. https://CRAN.R-project.org/package=rstatix

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol 79(17):5112–5120. https://doi.org/10.1128/AEM.01043-13

Lahti L, Shetty S (2017) Tools for microbiome analysis in R. URL: http://microbiome.github.io

Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Thurber RLV, Knight R, Beiko RG, Huttenhower C (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol 31(9):814–821. https://doi.org/10.1038/nbt.2676

Li W, Ni J, Cai S, Liu Y, Shen C, Yang H, Chen Y, Tao J, Yu Y, Liu Q (2019) Variations in microbial community structure and functional gene expression in bio-treatment processes with

odorous pollutants. Sci Rep 9(1):17870. https://doi.org/10.1038/s41598-019-54281-0
Louca S, Doebeli M (2018) Efficient comparative phylogenetics on large trees. Bioinformatics 34(6):1053–1055. https://doi.org/10.1093/bioinformatics/btx701
McMurdie PJ, Holmes S (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 8(4):e61217. https://doi.org/10.1371/journal.pone.0061217
Munoz-Garcia A, Mestanza O, Isaza JP, Figueroa-Galvis I, Vanegas J (2019) Influence of salinity on the degradation of xenobiotic compounds in rhizospheric mangrove soil. Environ Pollut 249:750–757
Navarrete-Euan H, Rodriguez-Escamilla Z, Perez-Rueda E, Escalante-Herrera K, Martinez-Nunez MA (2021) Comparing sediment microbiomes in contaminated and pristine wetlands along the Coast of Yucatan. Microorganisms 9(4). https://doi.org/10.3390/microorganisms9040877
Neuwirth E (2014) RColorBrewer: ColorBrewer Palettes. https://CRAN.R-project.org/package=RColorBrewer
Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2020) vegan: Community Ecology Package. R package version 2.5–7 edn
Pedersen TL (2020) patchwork: The Composer of Plots. https://CRAN.R-project.org/package=patchwork
Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glockner FO (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucl Acid Res 35(21):7188–7196. https://doi.org/10.1093/nar/gkm864
R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
Ramalingam V, Cupples AM (2020a) Anaerobic 1,4-dioxane biodegradation and microbial community analysis in microcosms inoculated with soils or sediments and different electron acceptors. Appl Microbiol Biotechnol 104(9):4155–4170. https://doi.org/10.1007/s00253-020-10512-3
Ramalingam V, Cupples AM (2020b) Enrichment of novel *Actinomycetales* and the detection of monooxygenases during aerobic 1,4-dioxane biodegradation with uncontaminated and contaminated inocula. Appl Microbiol Biotechnol 104(5):2255–2269. https://doi.org/10.1007/s00253-020-10376-7
Richards PM, Liang Y, Johnson RL, Mattes TE (2019) Cryogenic soil coring reveals coexistence of aerobic and anaerobic vinyl chloride degrading bacteria in a chlorinated ethene contaminated aquifer. Water Res 157:281–291. https://doi.org/10.1016/j.watres.2019.03.059
RStudio Team (2020) RStudio: integrated development for R. RStudio, PBC. Boston, MA
Schloss PD (2009) A high-throughput DNA sequence aligner for microbial ecology studies. Plos One 4(12)
Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75(23):7537–7541
Singh DP, Prabha R, Gupta VK, Verma MK (2018) Metatranscriptome analysis deciphers multifunctional genes and enzymes linked with the degradation of aromatic compounds and pesticides in the wheat rhizosphere. Front Microbiol 9:1331. https://doi.org/10.3389/fmicb.2018.01331
Thelusmond JR, Kawka E, Strathmann TJ, Cupples AM (2018) Diclofenac, carbamazepine and triclocarban biodegradation in agricultural soils and the microorganisms and metabolic pathways affected. Sci Total Environ 640–641:1393–1410. https://doi.org/10.1016/j.scitotenv.2018.05.403
Thelusmond JR, Strathmann TJ, Cupples AM (2019) Carbamazepine, triclocarban and triclosan biodegradation and the phylotypes and functional genes associated with xenobiotic degradation in four agricultural soils. Sci Total Environ 657:1138–1149. https://doi.org/10.1016/j.scitotenv.2018.12.145
Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer-Verlag, New York
Wickham H (2021a) Forcats: tools for working with categorical variables (factors). https://CRAN.R-project.org/package=forcats
Wickham H (2021b) tidyr: Tidy Messy Data. https://CRAN.R-project.org/package=tidyr
Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019) Welcome to the Tidyverse. J Open Source Softw 4(43):1686. https://doi.org/10.21105/joss.01686
Wickham H, Bryan J (2019) readxl: read excel files. R package version 1.3.1 ed. https://CRAN.R-project.org/package=readxl
Wickham H, François R, Henry L, Müller K (2021) dplyr: a grammar of data manipulation. R package version 1.0.6. https://CRAN.R-project.org/package=dplyr
Wilson FP, Cupples AM (2016) Microbial community characterization and functional gene quantification in RDX-degrading microcosms derived from sediment and groundwater at two naval sites. Appl Microbiol Biotechnol 100(16):7297–7309. https://doi.org/10.1007/s00253-016-7559-8
Wilson FP, Liu X, Mattes TE, Cupples AM (2016) *Nocardioides, Sediminibacterium, Aquabacterium, Variovorax,* and *Pseudomonas* linked to carbon uptake during aerobic vinyl chloride biodegradation. Environ Sci Pollut Res Int 23(19):19062–19070. https://doi.org/10.1007/s11356-016-7099-x
Wiseschart A, Mhuantong W, Tangphatsornruang S, Chantasingh D, Pootanakit K (2019) Shotgun metagenomic sequencing from Manao-Pee cave, Thailand, reveals insight into the microbial community structure and its metabolic potential. BMC Microbiology 19(144)
Ye Y, Doak TG (2009) A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. PLoS Comput Biol 5(8):e1000465. https://doi.org/10.1371/journal.pcbi.1000465