Predicting the Occurrence of Genes Relevant to Contaminant Biodegradation and Their Associated Phylotypes in Soil Microcosms

Alison M. Cupples* and Jean-Rene Thelusmond

Department of Civil and Environmental Engineering, Michigan State University, East Lansing, Michigan, USA

*Corresponding Author:

Alison M. Cupples

A135, 1449 Engineering Research Court, Michigan State University, East Lansing, MI 48824,

cupplesa@egr.msu.edu

Abstract

The overall aims were to 1) ascertain which genes encoding for monooxygenases (from methanotrophs, ammonia oxidizing bacteria and toluene/phenol oxidizers) and other key enzymes are present in soil microcosms and 2) determine which phylotypes are associated with those genes. The approach involved a predictive tool called PICRUSt2 and 16S rRNA gene amplicon datasets from two previous soil microcosm studies. The following targets from the KEGG database were examined: *pmo/amo, mmo, dmp/pox/tomA, tmo/tbu/tou, bssABC* (and downstream genes), *tod, xylM, xylA*, *gst, dhaA*, *catE*, *dbfA1, dbfA2* and phenol 2-monooxygenase.

A large number of phylotypes were associated with *pmo/amo*, while *mmo* was linked to only five. Several phylotypes were associated with both *pmo/amo* and *mmo*. The most dominant microorganism predicted for *mmoX* was *Mycobacterium* (also predicted for *pmo/amo*). A large number of phylotypes were associated with all six genes from the *dmp/pox/tomA* KEGG group. The taxonomic associations predicted for the *tmo/tbu/tou* KEGG group were more limited. In both datasets, *Geobacter* was a key phylotype for benzylsuccinate synthase. The dioxygenase-mediated toluene degradation pathway encoded by *todC1C2BA* was largely absent, as were the genes (*xylM, xylA*) encoding for xylene monooxygenase. All other genes investigated were predicted to be present and were associated with a number of microorganisms. Overall, the analysis predicted the genes encoding for sMMO (*mmo*), T3MO/T3MO/ToMO (*tmo/tbu/tou*) and benzylsuccinate synthase (*bssABC*) are present for a limited number of phylotypes compared to those encoding for pMMO/AMO (*pmo/amo*) and phenol monooxygenase/T2MO (*dmp/poxA/tomA*).

Keywords: PICRUSt2, monooxygenases, methanotrophs, toluene oxidation.

1. Introduction

Cometabolic oxidation involves the fortuitous oxidation of chemicals often by monooxygenases or dioxygenases (Horvath 1972; Wackett 1996). These enzymes are induced by growth substrates, such as methane, propane or toluene. During the oxidation process, oxygenases introduce one or two oxygen atoms, leading to metabolites which can then be further mineralized (Karigar and Rao 2011). Cometabolic oxidation can therefore be a detoxifying process for environmental contaminants such as trichloroethene (TCE) (Dolinova et al. 2017; Eguchi et al. 2001; Pfiffner et al. 1997; Semrini et al. 1991; Shao et al. 2019; Sutfin and Ramey 1997) or 1,4-dioxane (Mahendra et al. 2013). A number of oxygenases and microorganisms have been associated with this detoxification process. Knowledge on these microorganisms and their associated genes has the potential to enhance our understanding of the removal of the chlorinated solvents and other organic contaminants. Towards this goal, the current work applied a predictive tool called PICRUSt2 (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) (Douglas et al. 2020) to determine the occurrence of such genes (from the KEGG database, (Kanehisa 2002)) in soil communities. The following provides background on the biological markers examined.

The cometabolism of many environmental contaminants has frequently been associated with the methanotrophs. Methanotrophs use methane as a sole carbon and energy source, and are the only biological sink of methane. These organisms also play a critical role in the global carbon cycle because methane is an important greenhouse gas (IPCC 2007). Methanotrophs contain methane monooxygenase (MMO), which catalyzes the oxidation of methane to methanol. MMO exists in two forms, a soluble cytoplasmic form (sMMO) and a particulate membrane-associated form (pMMO) (Jiang et al. 2011). While pMMO has been found in all methanotrophs, except for the genera *Methylocella* (Theisen et al. 2005) and *Methyloferula* (Vorobev et al. 2011), sMMO is present in fewer strains (Murrell et al. 2000).

sMMO contains a hydroxylase with an $(\alpha\beta\gamma)_2$ structure, a regulatory protein and a reductase (Jiang et al. 2011). The genes encoding sMMO in *Methylococcus capsulatus* Bath (Csaki et al. 2003; Stainthorpe et al. 1990; Stainthorpe et al. 1989) and *Methylosinus trichosporium* OB3b (Cardy et al. 1991a; Cardy et al. 1991b) have been determined. They are found in a six-gene operon (*mmoXYBZDC*), which encodes the α, β, and γ subunits of the hydroxylase (*mmoXYZ*), the reductase (*mmoC*) and a regulatory or coupling protein (*mmoB*) (Jiang et al. 2011). MMOD is a regulatory element to repress expression of sMMO (Kim et al. 2019; Koo and Rosenzweig 2021; Merkx and Lippard 2002; Sazinsky et al. 2004; Semrau et al. 2013).

General Business

pMMO consists of three subunits encoded by the *pmoCAB* operon (Koo and Rosenzweig 2021). pMMO shares various similarities with ammonia monooxygenase (AMO), found in ammonia-oxidizing bacteria (AOB) (Holme et al. 1995). The enzymes typically contain three subunits (for pMMO: PmoA, PmoB, PmoC; for AMO: AmoA, AmoB, AmoC) (Wendeborn 2020). The substrate profile of pMMO is more limited compared to sMMO, including only methane, short linear hydrocarbons and TCE (Jiang et al. 2011). The substrate range for sMMO includes alkanes, alkenes, alicyclic hydrocarbons, halogenated aliphatics, monoaromatics, diaromatics and substituted methane derivatives (Jiang et al. 2011). For example, the sMMO in *Methylosinus trichosporium* OB3b was associated with the degradation of numerous chlorinated aliphatics as well as 1,4-dioxane (Mahendra and Alvarez-Cohen 2006; Oldenhuis et al. 1989).

AOB are capable of cometabolic TCE transformation via AMO (Alpaslan Kocamemi and Cecen 2007). Aerobic nitrification is an important nitrogen cycling process in many natural and engineered systems. The first step involves the oxidation of ammonia to hydroxylamine by the membrane-bound protein AMO (Hooper et al. 1997). The degradation of TCE and other halogenated hydrocarbons has been extensively studied in *Nitrosomonas europaea* (Arciero et al. 1989; Rasche et al. 1991; Vannelli et al. 1990). The KEGG database (Kanehisa 2002) places the genes for pMMO and AMO together (*pmo/amo*) and therefore the current analysis targeted both sets of genes simultaneously.

Other aerobic bacteria capable of cometabolic TCE oxidation include toluene oxidizers (Chang and Alvarez-Cohen 1995; Fries et al. 1997; Guo et al. 2001; Sun et al. 1997). For example, *Burkholderia vietnamiensis* G4 (formerly *Pseudomonas cepacia; Burkholderia cepacia*) oxidized TCE when induced on toluene or phenol (Nelson et al. 1987). TCE removal was associated with toluene 2-monooxygenase (T2MO) (Folsom et al. 1990; Nelson et al. 1986; Newman and Wackett 1997; Shields et al. 1995). T2MO is encoded by the operon *tomA012345* (Shields and Francesconi 1996) which is a three-component enzyme consisting of a hydroxylase (*tomA1A3A4*) a NADH oxidoreductase (*tomA5*), and a protein (*tomA2*) involved in electron transfer between the hydroxylase and reductase (Canada et al. 2002; Newman and Wackett 1995). In *B. vietnamiensis* G4, the toluene degradation genes are located on a large, self-transmissible plasmid (Parales et al. 2008; Shields and Francesconi 1996; Shields et al. 1995). T2MO also catalyzes the oxidation of *o*-cresol, dichloroethylenes, phenol, chloroform, 1,4-dioxane, aliphatic ethers, and diethyl sulfide (Hur et al. 1997; Mahendra and Alvarez-Cohen 2006; Parales et al. 2008; Shim and Wood 2000). The KEGG database (Kanehisa 2002) places the genes encoding for T2MO in a group with those encoding for phenol monooxygenases, which have been found in *Ralstonia eutropha* strain E2 (*poxABCDEF*) (Hino et al. 1998) and *Pseudomonas* sp. CF600 (*dmpKLMNOP*) (Nordlund et al. 1990).

General Business

As the current analysis utilized the KEGG database (Kanehisa 2002), it targeted all three sets of genes simultaneously.

TCE degradation by *Ralstonia (*formerly *Pseudomonas) pickettii* PKO1 has been attributed to toluene 3-monooxygenase (T3MO) (Leahy et al. 1996; Olsen et al. 1994), with sequence analysis of the T3MO-encoding region illustrating six structural genes, *tbuA1UBVA2C* (Byrne et al. 1995). Substrates for T3MO include toluene, benzene, ethylbenzene, *o*-xylene, *m*-xylene, and *p*-xylene, alkenes, TCE, 1,4-dioxane, nitrobenzene and N-nitrosodimethylamine (Haigler and Spain 1991; Leahy et al. 1996; Mahendra and Alvarez-Cohen 2006; McClay et al. 2000; Olsen et al. 1997; Parales et al. 2008; Sharp et al. 2005). A comparison of the deduced amino acid sequences revealed significant overall homology to peptides from the toluene 4-monooxygenase (T4MO) from *Pseudomonas mendocina* KR1 (Byrne et al. 1995). The genes encoding T4MO involve a cluster of five genes, *tmoABCDE* (Yen et al. 1991). T4MO has been associated with the oxidation of alkenes, N-nitrosodimethylamine, various polycyclic aromatic substrates, TCE, chloroform, 1,4-dioxane and substituted benzenes (Mahendra and Alvarez-Cohen 2006; McClay et al. 1996; McClay et al. 2000; Oppenheim et al. 2001; Parales et al. 2008; Pikus et al. 1997; Sharp et al. 2005; Winter et al. 1989). Recently, researchers provided evidence that a toluene monooxygenase in *Azoarcus* sp.DD4, encoded by the *tmoABCDEF* gene cluster, was the key enzyme for the cometabolism of dioxane, 1,1-DCE and cDCE (Deng et al. 2020; Li et al. 2021).

Another toluene degrading monooxygenase (toluene/o-xylene monooxygenase, ToMO) was identified in *P. stutzeri* OX1 and is similar to T4MO, although the enzyme is not regiospecific and has no clear preference for the position of oxidation on the toluene ring (Parales et al. 2008). ToMO contains a three-component hydroxylase (encoded by *touABE*) a NADH-ferredoxin oxidoreductase (*touF*), a mediating protein (*touD*), and a Rieske-type ferredoxin (from *touC*) (Bertoni et al. 1998; Ryoo et al. 2001). ToMO has a broad substrate range including *o*-xylene, *m*-xylene, *p*-xylene, toluene, benzene, ethylbenzene, styrene, and naphthalene (Bertoni et al. 1996), TCE, chloroform, and 1,1-dichloroethylene (Chauhan et al. 1998; Shim and Wood 2000) and tetrachloroethene (Ryoo et al. 2000). The gene order and the deduced amino acid sequences of ToMO are similar to those of T3MO and T4MO (Chauhan et al. 1998). The KEGG database (Kanehisa 2002) places the genes encoding for T3MO, T4MO and ToMO in one group, thus, the current analysis targeted all three sets of genes simultaneously.

Additional genes targeted in the current study include the dioxygenase-mediated toluene degradation pathway in *Pseudomonas putida F1*, encoded by *todC1C2BA* (Parales et al. 2008; Zylstra and Gibson 1989; Zylstra et al. 1988) as well as *todD* and *todE* (for the next steps in the toluene degradation pathway)

General Business

(Gibson et al. 1970; Klecka and Gibson 1981). The genes (*xylM, xplA*) encoding for another toluene degrading enzyme, xylene monooxygenase, a two-component enzyme consisting of XylM and XylA (Shaw and Harayama 1995; Suzuki et al. 1991) were also examined here. In addition to the genes discussed above, a number of other genes associated with the biodegradation of environmental contaminants were also investigated, including benzylsuccinate synthase, glutathione *S*-transferase, haloalkane dehalogenase, catechol 2,3-dioxygenase, dibenzofuran dioxygenase and phenol 2-monooxygenase.

The objective of this study was to apply a predictive tool, PICRUSt2 (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) (Douglas et al. 2020) to 1) ascertain which genes encoding for monooxygenases (from methanotrophs, ammonia oxidizing bacteria and toluene/phenol oxidizers) and other enzymes are present in soil communities and 2) determine the phylotypes associated with those genes. The analysis involved an examination of 16S rRNA amplicon gene data from two previous studies (Thelusmond et al. 2018; Thelusmond et al. 2016). An overall hypothesis was that phylotypes less typically associated with these functional genes would be identified. Also, we hypothesized that some genes would be associated with a larger number phylotypes compared to others. The data in the current study provides novel insights as to the potential occurrence of these functional genes and their associated phylotypes in soil microbial communities.

2. Methods

2.1. Soil Communities

High throughput sequencing datasets (16S rRNA gene amplicon, Illumina MiSeq) from two previous studies involving different agricultural soils (Thelusmond et al. 2018; Thelusmond et al. 2016) were examined in the current study using PICRUSt2 (Douglas et al. 2020). PICRUSt2 was developed recently to predict the functional potential of a bacterial community based on marker gene sequencing profiles. The two previous studies investigated the biodegradation of Carbamazepine (CBZ), Diclofenac and Triclocarban in soil microcosms. Details on sample collection, DNA extraction and high throughput sequencing were previously described (Thelusmond et al. 2018; Thelusmond et al. 2016). Briefly, in one study (herein called the CBZ Study), two agricultural soils (herein soils 1 and 2) were amended with different concentrations (0, 50 ng/g, 500 ng/g, 5000 ng/g) of CBZ. Soil microcosms were either kept aerobic or were purged with oxygen-free nitrogen and sealed with septa and crimps (herein called saturated). Soil characteristics in both studies were determined by A & L Great Lakes Laboratories, Inc. (Fort Wayne, IN). Soil 1 was a loamy sand at pH 7.6 and 1.5 % organic matter. Soil 2 was a sandy loam at pH 6.7 and 8% organic matter. DNA was extracted after 14 days of incubation with CBZ (Thelusmond

et al. 2016). In the other study (herein called the Multiple Chemicals Study) another set of agricultural soils were amended with CBZ, Diclofenac or Triclocarban and DNA was extracted following different incubation periods (called the samples). Additionally, soils were treated in the same manner except no chemicals were added (called the controls). The study included two sandy loam soils (herein soils A, B) and one loamy sand soil (herein soil C). Soil pH values were 6.9 for soils A and B and 6.6 for soil C. The soil organic matter values were 2.8%, 2.4% and 1.4% for soils A, B, and C, respectively (Thelusmond et al. 2018).

2.2. R Packages

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), 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. 2021), 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) 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.

2.3. Most Abundant Phylotypes

In the current work, the amplicon sequencing data in the fastq file format was re-analyzed with Mothur (version 1.44.2) (Schloss 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. Phyloseq was used to determine relative abundance values and ampvis2 was used to generate the heatmaps illustrating the most abundant phylotypes for each set of samples. For all figures, the R package patchwork combined plots, combined legends (when appropriate) and created letter annotations.

General Business

2.4. PICRUSt2 and R Analysis

PICRUSt2 (Douglas et al. 2020) was used to analyze Mothur generated files on the High Performance Computing Cluster (HPCC) at Michigan State University (MSU). 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 were examined for the presence of genes and phylotypes associated with each of the following genes from the KEGG database (Kanehisa 2002): *pmoA/amoA* (K10944), *pmoB/amoB* (K10945), *pmoC/amoC* (K10946), *mmoX* (K16157), *mmoY* (K16158), *mmoZ* (K16159), *mmoB* (K16160), *mmoC* (K16161), *mmoD* (K16162), *dmpK/poxA/tomA0* (K16249), *dmpL/poxB/tomA1* (K16243), *dmpM/poxC/tomA2* (K16244), *dmpN/poxD/tomA3* (K16242), *dmpO/poxE/tomA4* (K16245), *dmpP/poxF/tomA5* (K16246), *tmoA/tbuA1/touA* (K15760), *tmoB/tbuU/touB* (K15761), *tmoC/tbuB/touC* (K15762), *tmoD/tbuV/touD* K15763), *tmoE/tbuA2/touE* (K15764), *tmoF/tbuC/touF* (K15765), benzylsuccinate synthase (BSS) (*bssABC*) (K07540), *bbsA* (K07549), *bbsB* (K07550), *bbsC* (K07547), *bbsD* (K07548), *bbsE* (K07543), *bbsF* (K07544), *bbsG* (K07545), *bbsH* (K07546), *todC1* (K03268), *todC2* (K16268), *todB* (K18089), *todA* (K18090), *todD* (K16269), *todE* (K16270), *xylM* (K15757), *xylA* (K15758), *gst* (K00799), *dhaA* (K01563), *catE* (K07104), *dbfA1* (K14599), *dbfA2* (K14600) and phenol 2-monooxygenase (K03380).

The analysis was performed using R (version 4.0.4) (R Core Team 2018) with RStudio (version 1.1.456) (RStudio Team 2020) and a number of R packages. RStudio on the HPCC at MSU 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 1) a file containing gene numbers and descriptions and 2) a taxonomy file (from Mothur), using the R packages data.table, dplyr, tidyr, ggplot2 and patchwork. Chord diagrams to illustrate the relationships between phylotypes and genes were created with the chordDiagram function in the R package circlize. One or two genes (*pmoA/amoA, mmoX, dmpK/poxA/tomA0, bssABC, tmoA/tbuA1/touA, gst*, *dhaA*, *catE*, *dbfA1*, *dbfA2* and phenol 2-monooxygenase) were selected for the creation of bar charts illustrating the dominant phylotypes, faceted (in ggplot2) for different treatments and soils for each study.

3. Results

Heatmaps of the most abundant phylotypes were generated (Figure 1) to enable a comparison to those associated with the functional genes (as discussed below). For the CBZ Study, the most abundant included an unclassified phylotype within the family *Methylophilaceae*, *Subgroup 6* (*Acidobacteria*), an uncultured *Bacteroidetes*, and an uncultured *Gemmatimonadetes* (Figure 1A). The *Methylophilaceae*

phylotype appeared particularly impacted by CBZ. *Methylophilaceae* are methylotrophs capable of utilizing methanol or methylamine (but not methane) as a sole source of carbon and energy (Doronina et al. 2014). In the Multiple Chemicals Study, *Subgroup 6* (*Acidobacteria)* was dominant in soils A and B (but not C) (Figure 1B). Instead, *Sphingomonas* and an uncultured *Proteobacteria* were particularly dominant across all treatments in soil C. From these abundant microorganisms, only *Sphingomonas* was associated with the genes investigated (see below, dibenzofuran dioxygenase).

The *pmo/amo* KEGG group was predicted for a large num ber of microorganisms in both studies (Figure 2). Further, all three genes (*pmoA/amoA, pmoB/amoB, pmoA/amoC*) were detected for every phylotype. Focusing on *pmoA/amoA* alone, the most dominant phylotypes of the CBZ Study included unclassified *Beijerinckiaceae* (order *Rhizobiales*), unclassified *Methylomonaceae* (*Methylococcales*), *MND1* (*Betaproteobacteriales*), *Nitrosomonas (Betaproteobacteriales)* and *Nitrospira (Nitrospirales)* (Figure 3A). In soil 2, *Methylobacter (Methylococcales)* was also dominant and *MND1* was less important (Figure 3A). In the Multiple Chemicals Study, *Nitrosomonas* and *Nitrospira* were dominant in soils A and C, respectively, with others being present at lower levels (Figure 3B).

The genes encoding for sMMO were present in only four phylotypes in the CBZ Study (Figure 4A) and two phylotypes in the Multiple Chemicals Study (Figure 4B). Two phylotypes (unclassified *Gammaproteobacteria*, unclassified *Methylomonaceae*) were associated with all six subunits (in red, Figure 4A). For each study, two phylotypes were associated with five subunits (all except *mmoD*), *Mycobacterium* (both studies), unclassified *Corynebacteriales* (CBZ Study only) and unclassified *Actinobacteria* (Multiple Chemicals Study). However, as stated above, MMOD is not necessary for sMMO function. Focusing only on *mmoX, Mycobacterium* was the most dominant microorganism linked with this gene in both studies (Figure 5).

Considering both sets of genes encoding for MMO (sMMO and pMMO), for the CBZ Study, all four phylotypes (unclassified *Gammaproteobacteria,* unclassified *Methylomonaceae,* unclassified *Corynebacteriales* and *Mycobacterium*) detected for *mmo* were also detected for *pmo/amo.* For the Multiple Chemicals Study, both phylotypes (unclassified *Actinobacteria* and *Mycobacterium*) associated with *mmo* were also associated with *pmo/amo.*

A large number of microorganisms were associated with all six genes from the *dmp/pox/tomA* KEGG group in both datasets (Figure 6A and B). In the CBZ Study (Figure 6A), the majority classified within the families *Burkholderiaceae (Massilia, Burkholderiaceae_unclass, Polaromonas, Burkholderia.*

Caballeronia.Paraburkholderia, Hydrogenophaga, Cupriavidus), Rhodocyclaceae (unclassified Rhodocyclaceae, Dechloromonas, Ferribacterium, Uliginosibacterium) and *Nitrosomonadaceae (IS-44, mle1-7, unclassified Nitrosomonadaceae)*. A few phylotypes belonged to other families (in parenthesis): *A21b_ge (A21b), TRA3-20_ge (TRA3-20), SC-I-84_ge (SC-I-84), Pseudomonas (Pseudomonadaceae)* and *Acinetobacter (Moraxellaceae)*. Also, many phylotypes were unclassified *(*uncultured, *unclassified Betaproteobacteriales, unclassified Bacteria, unclassified Proteobacteria, unclassified Gammaproteobacteria)*. All of the above families, except the unclassified, *Pseudomonadaceae (Pseudomonadales) and Moraxellaceae (Pseudomonadales),* are within the order *Betaproteobacteriales*.

Many of the same phylotypes were predicted for all six genes from the *dmp/pox/tomA* KEGG group in the Multiple Chemicals Study (Figure 6B). The majority classified within the families *Burkholderiaceae (Burkholderia.Caballeronia.Paraburkholderia*, *unclassified Burkholderiaceae, Hydrogenophaga, Massilia, Cupriavidus), Rhodocyclaceae (unclassified Rhodocyclaceae, Dechloromonas, Thauera)* and *Nitrosomonadaceae (MND1, unclassified Nitrosomonadaceae, mle1-7)*. A few belonged to other families (in parenthesis): *A21b_ge (A21b), SC-I-84_ge (SC-I-84), TRA3-20_ge (TRA3-20), Pseudomonas (Pseudomonadaceae)*. Further, some were unclassified (*unclassified Betaproteobacteriales, uncultured, unclassified Gammaproteobacteria)*. Again, all of the above families, except the unclassified and *Pseudomonadaceae (Pseudomonadales)*, are within the order *Betaproteobacteriales.*

To determine the importance of each phylotype for each soil and set of conditions, a more detailed analysis was performed for the phylogenetic associations of *dmpK/poxA*/*tomA0* (Figure 7). In soil 1, under both aerobic and saturated conditions, *unclassified Betaproteobacteria*, *unclassified Burkholderiaceae, Hydrogenophaga, Massilia* and *Polaromonas* were particularly important (Figure 7A), whereas *unclassified Burkholderiaceae, unclassified Rhodocyclaceae* and *Massilia* were the dominant phylotypes for soil 2 (Figure 7A). For the Multiple Chemicals Study, *Burkholderia.Caballeronia.Paraburkholderia, TRA3-20_ge* and *Massilia* were important for soil A, *unclassified Betaproteobacteria* was dominant for soil B, whereas *TRA3-20_ge* and *uncultured* was important for soil C (Figure 7B). The trends were similar between the aerobic and saturated (Figure 7A) and between the samples and controls (Figure 7B).

The taxonomic associations for the *tmo/tbu/tou* KEGG group were more limited compared to the *dmp/pox/tomA* KEGG group. For the CBZ Study, although 39 phylotypes were predicted for one or more of the genes, only four (*Nevskia, Rhodococcus, unclassified Corynebacteriales* and *unclassified Nocardiaceae*) were associated with all six genes (Figure 8A). Three of these classify within

General Business

Actinobacteria (Corynebacteriales) and *Nevskia* classifies within the *Gammaproteobacteria (Salinisphaerales).* For the Multiple Chemicals Study, 32 phylotypes contained one or more of the genes and only three (*Rhodococcus, Labrys* and *Burkholderia.Caballeronia.Paraburkholderia*) were predicted to contain all six (Figure 8B). *Rhodococcus, Labrys* and *Burkholderia.Caballeronia.Paraburkholderia* classify within *Actinobacteria, Alphaproteobacteria and Gammaproteobacteria*, respectively.

A large number of phylotypes were associated with one or more of the following genes in both studies, *bssABC, bbsA, bbsB, bbsC, bbsD, bbsE, bbsF, bbsG* or *bbsH* (Figure 9). In the CBZ Study, only *Geobacter, unclassified Geobacteraceae* and *unclassified Desulfuromonadales* were associated with all and *unclassified Syntrophaceae, Syntrophus* and *Smithella* were associated with benzylsuccinate synthase (Figure 9A). In the Multiple Chemical Study, only *Geobacter* was associated with benzylsuccinate synthase or with all genes (Figure 9B).

Focusing on *tmoA* only, from the phylotypes associated with all six genes, *Rhodococcus* and *Nevskia* were dominant in the CBZ Study (Figure 10A). However, *Nevskia* was present at low levels in soil 2 under saturated conditions. Further, *unclassified Corynebacteriales* and *unclassified Nocardiaceae* were present only at low levels for the majority of the treatments. In the Multiple Chemicals Study, *Rhodococcus* was important only in soils A (control and samples) and B (samples only) and *Labrys* was important only in soil A (control and samples) (Figure 10B). *Burkholderia.Caballeronia.Paraburkholderia* was only important in soil C samples (Figure 10B). A similar analysis for benzylsuccinate synthase (BSS) demonstrated the importance of *Geobacter* in all datasets (Figure 10C and 10D). *Unclassified Syntrophaceae, Syntrophus* and *Smithella* were only strongly associated with benzylsuccinate synthase in soil 2 under aerobic conditions (Figure 10C).

The supplementary section illustrates the phylotypes associated with glutathione S-transferase (Figure S1), haloalkane dehalogenase (Figure S2), catechol 2,3-dioxygenase (Figure S3), dibenzofuran dioxygenase (Figure S4) and phenol 2-monooxygenase (Figure S5). Unclassified *Sphingomonadaceae* and unclassified *Myxococcales* were the dominant phylotypes associated with both glutathione S-transferase and haloalkane dehalogenase (Figures S1 and S2). *Microvirga,* unclassified *Xanthobacteraceae* and unclassified *Solirubrobacteraceae* were important for catechol 2,3-dioxygenase (Figure S3). Unclassified *Sphingomonadaceae* and *Sphingomonas* were key phylotypes for dibenzofuran dioxygenase (Figure S4). Finally, unclassified *Micrococcaceae*, unclassified *Microbacteriaceae,* unclassified *Xanthobacteraceae* and *Nitrobacter* were dominant phylotypes for phenol 2-monooxygenase (Figure S5).

General Business

The dioxygenase-mediated toluene degradation pathway in *Pseudomonas putida F1*, encoded by the *todC1C2BA* genes (Parales et al. 2008; Zylstra and Gibson 1989; Zylstra et al. 1988), was absent in all samples, except for a small number of predictions for *todB* which was associated with *Sphingobium* in the CBZ Study samples. The genes (*todD, todE*) that encode for the next steps in the toluene degradation pathway (Gibson et al. 1970; Klecka and Gibson 1981) were also absent in all samples (except for one prediction for *todE* associated with *Mycobacterium*). The genes (*xylM, xylA*) encoding for another toluene degrading enzyme, xylene monooxygenase, a two-component enzyme consisting of XylM and XylA (Shaw and Harayama 1995; Suzuki et al. 1991), were also absent in both datasets in the current study.

4. Discussion

The current analysis generated novel data for the genes and phylotypes potentially linked to contaminant biodegradation in the soil communities. As noted by the developers of PICRUSt2, there are two criticisms of this approach; the data are biased toward existing reference genomes and amplicon-based predictions cannot provide resolution to distinguish strain-specific functionality (Douglas et al. 2020). Nevertheless, the approach provides a platform for other studies to examine the functional abilities of microbial communities without the expense of shotgun sequencing. The results generated are valuable for hypotheses development towards future research.

Aerobic methanotrophs have been found within the *Gammaproteobacteria*, *Alphaproteobacteria* and *Verrucomicrobia* (Knief 2015; Koo and Rosenzweig 2021). Here, no phylotypes classifying with the *Verrucomicrobia* were detected for either sMMO or pMMO. In the current study, both sMMO and pMMO were present in all datasets. The genes encoding sMMO were associated with fewer phylotypes compared to pMMO. These results are consistent with the literature illustrating the presence of sMMO in fewer methanotrophs compared to pMMO (Murrell et al. 2000).

Mycobacterium was the dominant phylotype associated with the genes encoding for sMMO. This finding is supported by previous research that identified a sMMO-like enzyme in two *Mycobacterium* strains (NBB3 and NBB4) (Martin et al. 2014). Also, an NCBI BLAST search by the authors (1000 max. target sequences) with the MULTISPECIES: methane monooxygenase component A alpha chain (Sequence ID: WP_003609337.1) resulted in numerous alignments to sequences from the genus *Mycobacterium*. In the CBZ Study, unclassified *Methylomonaceae (Methylococcales, Gammaproteobacteria)* was also associated with the genes encoding for sMMO. Consistent with this, a DNA-based stable isotope probing study reported predominantly active methanotrophs belonged to *Methylomonaceae* (Kauppera et al.

2021). Other researchers found methane oxidizers within the order *Methylococcales* were composed of bacteria belonging to the family *Methylomonaceae* (Broman et al. 2020). Another report found *Methylomonaceae* had the highest average relative abundance of bacterial cDNA transcripts during drought in two restored fens (Unger et al. 2021). The other two phylotypes associated with *mmo* (unclassified *Corynebacteriales* and unclassified *Actinobacteria*) represent the order and phylum of the genus *Mycobacterium*. Other genera associated with sMMO, such as *Methylosinus* and *Methylococcus*, were not predicted to be important in the soils of the current study.

The *pmo/amo* KEGG group was associated with an unclassified member of the *Beijerinckiaceae,* a family known to contain methanotrophs (Knief 2015). In the CBZ Study, in soil 2, *Methylobacter (Methylococcales)* was a key phylotype for the *pmo/amo* KEGG group. Others reported that three methanotroph genomes from the genus *Methylobacter* represented the most abundant methanotrophs across the wetland (Smith et al. 2018). The authors concluded that *Methylobacter* may represent important mediators of methane fluxes in freshwater saturated sediments and soils worldwide (Smith et al. 2018). Interestingly, in the current work this phylotype was only predicted to be important for the *pmo/amo* KEGG group in one of five soils.

A notable finding was that there were several phylotypes associated with both *mmo* and *pmo/amo* datasets. For the CBZ Study, the four phylotypes (unclassified *Gammaproteobacteria,* unclassified *Methylomonaceae,* unclassified *Corynebacteriales* and *Mycobacterium*) associated with *mmo* were also detected for *pmo/amo.* For the Multiple Chemicals Study, the two phylotypes (unclassified *Actinobacteria* and *Mycobacterium*) associated with *mmo* were also associated with *pmo/amo.* These results again emphasize the potential importance of *Mycobacterium* and unclassified *Methylomonaceae* for methane oxidation and contaminant biodegradation.

Here, the ammonia oxidizing phylotypes associated with the *pmo/amo* KEGG group included *MND1, Nitrosomonas* and *Nitrosospira.* AOB are found within the *Betaproteobacteria* (genera *Nitrosomonas, Nitrosospira*) and *Gammaproteobacteria* (genus *Nitrosococcus*), with terrestrial AOB generally being restricted to the *Betaproteobacteria* (Norton 2011). *Nitrosomonas* and *Nitrosospira* were key phylotypes for *pmoA/amoA* in soil A and C, respectively in the Multiple Chemicals Study. *MND1* was abundant in soil 1 of the CBZ Study under aerobic conditions for *pmoA/amoA*. *MND1* belongs to the family *Nitrosomonadaceae* and it was previously reported that all cultivated representatives of the *Nitrosomonadaceae* are lithoautotrophic ammonia oxidizers (Prosser et al. 2014).

General Business

The current analysis predicted the importance of phylotypes primarily in the families *Burkholderiaceae, Rhodocyclaceae* and *Nitrosomonadaceae* for the six genes from the *dmp/pox/tomA* KEGG group. Consistent with these results, an NCBI BLAST performed on the complete sequence of *Burkholderia cepacia* G4 toluene ortho-monooxygenase operon (accession number AF349675.1) generated matches primarily within the family *Burkholderiaceae.* Further both, *Burkholderia vietnamiensis* G4 (containing *tomA012345*) (Parales et al. 2008; Shields and Francesconi 1996; Shields et al. 1995) and *Ralstonia eutropha* strain E2 (containing *poxABCDEF*) (Hino et al. 1998) classify within the *Burkholderiaceae.* The prediction of *Pseudomonas* for these genes in both studies is consistent with previous reports of *dmpKLMNOP in Pseudomonas* sp. CF600 (Nordlund et al. 1990). The current research introduces the possibility that phylotypes within *Burkholderiaceae, Pseudomonadaceae, Rhodocyclaceae* and *Nitrosomonadaceae* and a small number of other families may be associated with these genes in agricultural soils.

Here, *Rhodococcus, Nevskia* and *Labrys* were primarily associated with the six genes from the *tmo/tbu/tou* KEGG group. Consistent with these results, open reading frames in *Rhodococcus* sp. strain AD45 illustrated high sequence similarity with the *tmoABCDEF* gene cluster encoding the toluene 4-monooxygenase of *Pseudomonas mendocina* KR1 (van Hylckama Vlieg et al. 2000). Further, an NCBI BLAST search with TmoA from *Thaurea* sp. 17 (ENO79309) resulted in a 75% identity match to an aromatic/alkene/methane monooxygenase hydroxylase/oxygenase subunit alpha (WP_029920984.1) in *Nevskia soli*, as well as matches to several *Burkholderia* and *Paraburkholderia* sequences.

The BSS-based toluene pathway, first identified in the genera *Thauera* and *Azoarcus*, was previously reported to be a common mechanism for anaerobic toluene degradation by phylogenetically diverse organisms (Chakraborty and Coates 2004; Parales et al. 2008; Spormann and Widdel 2000; Widdel and Rabus 2001). The first step is the addition of toluene to the double bond of fumarate to form benzylsuccinate by benzylsuccinate synthase (BssABC) (Leuthner et al. 1998). The next steps are catalyzed by BbsEF, BbsG, BbsH, BbsCD and BbsAB to convert benzylsuccinate to benzoyl-CoA (Leuthner and Heider 2000). In the CBZ study, six phylotypes (*Geobacter, unclassified Geobacteraceae, unclassified Desulfuromonadales, unclassified Syntrophaceae, Syntrophus, Smithella*) were predicted to be associated with benzylsuccinate synthase. Consistent with these results, toluene degradation has been reported for *Geobacter toluenoxydans* (Kunapuli et al. 2010), *G. metallireducens* GS-15 and *G. grbiciae* TACP-2 (Coates et al. 2001; Lovley et al. 1993). Further, using time-resolved RNA stable isotope probing and RT-qPCR, researchers reported that organisms within the family *Syntrophaceae* appeared to play an important role in toluene metabolism (Fowler et al. 2014). The family *Syntrophaceae* contains

General Business

four genera *Syntrophus, Smithella, Desulfobacca,* and *Desulfomonile* (Kuever 2014). In the Multiple Chemicals Study, only *Geobacter* was associated with benzylsuccinate synthase.

5. Conclusions

PICRUSt2 was utilized to investigate the occurrence of genes associated with contaminant biodegradation from methanotrophs, ammonia oxidizing bacteria and toluene/phenol oxidizers. From these, genes encoding for sMMO (*mmo*), T34MO/T3MO/ToMO (*tmo/tbu/tou*) and benzylsuccinate synthase (*bssABC*) were detected for a limited number of phylotypes. In contrast, the genes encoding for pMMO/AMO (*pmo/amo*) and phenol monooxygenase/T2MO (*dmp/poxA/tomA*) were detected for a larger number of phylotypes, suggesting their occurrence may be more widespread in soil communities. Unclassified *Methylomonaceae* was linked to both *pmo/amo* as well as *mmo,* indicating this family may have a key role in methane oxidation in soil communities. *Mycobacterium* and *Geobacter* were particularly dominant for sMMO and benzylsuccinate synthase respectively, however, additional data are needed to confirm these findings. This work offers a platform to study the potential functional capabilities of microbial communities. Future research could focus on how these trends differ between environmental samples (such as contaminated soil, sediment or groundwater) or site conditions (such as pH, redox potential).

Acknowledgements

Funding: This work was supported by the USDA (Agriculture and Food Research Initiative, Number 2014-67019-24024) and NSF (Award Number 2129228).

Availability of data: Illumina sequencing datasets for the CBZ Study were deposited in the NCBI Sequence Read Archive under Bioproject: PRJNA311080 and Biosample: SAMN04461763 (Thelusmond et al. 2016). Illumina sequencing datasets for the Multiple Chemical Study were deposited in the NCBI Sequence Read Archive under BioProject: PRJNA429625 and BioSample: SAMN08348582 (Thelusmond et al. 2018).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Alpaslan Kocamemi B, Cecen F (2007) Kinetic analysis of the inhibitory effect of trichloroethylene (TCE) on nitrification in cometabolic degradation. Biodegrad 18(1):71-81 doi:10.1007/s10532-005-9037-3

Ammar R (2019) randomcoloR: Generate Attractive Random Colors. doi: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 doi: https://doi.org/10.1101/299537

Arciero D, Vannelli T, Logan M, Hooper AB (1989) Degradation of trichloroethylene by the ammonia-oxidizing bacterium *Nitrosomonas europaea*. Biochem Biophys Res Commun 159(2):640-3 doi:10.1016/0006-291x(89)90042-9

Attali D (2021) Colourpicker: A colour picker tool for shiny and for selecting colours in plots doi: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 doi:10.1093/sysbio/syy054

Bertoni G, Bolognese F, Galli E, Barbieri P (1996) Cloning of the genes for and characterization of the early stages of toluene and o-xylene catabolism in *Pseudomonas stutzeri* OX1. Appl Environ Micro 62(10):3704-11 doi:10.1128/aem.62.10.3704-3711.1996

Bertoni G, Martino M, Galli E, Barbieri P (1998) Analysis of the gene cluster encoding toluene/o-xylene monooxygenase from *Pseudomonas stutzeri* OX1. Appl Environ Micro 64(10):3626-32 doi:10.1128/AEM.64.10.3626-3632.1998

Broman E, Sun X, Stranne C, Salgado MG, Bonaglia S, Geibel M, Jakobsson M, Norkko A, Humborg C, Nascimento FJA (2020) Low abundance of methanotrophs in sediments of shallow Boreal coastal zones with high water methane concentrations. Front Microbiol 11:1536 doi:10.3389/fmicb.2020.01536

Byrne AM, Kukor JJ, Olsen RH (1995) Sequence analysis of the gene cluster encoding toluene-3-monooxygenase from *Pseudomonas pickettii* PKO 1. Gene 154:65-70

Canada KA, Iwashita S, Shim H, Wood TK (2002) Directed evolution of toluene ortho-monooxygenase for enhanced 1-naphthol synthesis and chlorinated ethene degradation. Journal of Bacteriology 184(2):344-9

Cardy DL, Laidler V, Salmond GP, Murrell JC (1991a) The methane monooxygenase gene cluster of *Methylosinus trichosporium*: cloning and sequencing of the mmoC gene. Arch Microbiol 156(6):477-83 doi:10.1007/BF00245395

Cardy DL, Laidler V, Salmond GP, Murrell JC (1991b) Molecular analysis of the methane monooxygenase (MMO) gene cluster of *Methylosinus trichosporium* OB3b. Mol Microbiol 5(2):335-42 doi:10.1111/j.1365-2958.1991.tb02114.x

Chakraborty R, Coates JD (2004) Anaerobic degradation of monoaromatic hydrocarbons. Appl Microbiol Biotechnol 64(4):437-46 doi:10.1007/s00253-003-1526-x

Chang HL, Alvarez-Cohen L (1995) Transformation capacities of chlorinated organics by mixed cultures enriched on methane, propane, toluene, or phenol. Biotechnol Bioeng 45(5):440-9 doi:10.1002/bit.260450509

Chauhan S, Barbieri P, Wood TK (1998) Oxidation of trichloroethylene, 1,1-dichloroethylene, and chloroform by toluene/o-xylene monooxygenase from *Pseudomonas stutzeri* OX1. Appl Environ Microbiol 64(8):3023-4 doi:10.1128/AEM.64.8.3023-3024.1998

Coates JD, Bhupathiraju VK, Achenbach LA, McLnerney MJ, Lovley DR (2001) *Geobacter hydrogenophilus, Geobacter chapellei* and *Geobacter grbiciae*, three new, strictly anaerobic,

558 dissimilatory Fe(III)-reducers. Int J Syst Evol Microbiol 51(Pt 2):581-588 doi:10.1099/00207713-
559 51-2-581
560 Csaki R, Bodrossy L, Klem J, Murrell JC, Kovacs KL (2003) Genes involved in the copper-dependent
561 regulation of soluble methane monooxygenase of *Methylococcus capsulatus* (Bath): cloning,
562 sequencing and mutational analysis. Micro 149(Pt 7):1785-1795 doi:10.1099/mic.0.26061-0
563 Czech L, Barbera P, Stamatakis A (2020) Genesis and Gappa: processing, analyzing and visualizing
564 phylogenetic (placement) data. Bioinformatics 36(10):3263-3265
565 doi:10.1093/bioinformatics/btaa070
566 Deng D, Pham DN, ,, Li F, Li M (2020) Discovery of an inducible toluene monooxygenase that co-oxidizes
567 1, 4-dioxane and 1, 1-dichloroethylene in propanotrophic *Azoarcus* sp. DD4. Appl Environ
568 Microbiol 86 e01163-20
569 Dolinova I, Strojsova M, Cernik M, Nemecek J, Machackova J, Sevcu A (2017) Microbial degradation of
570 chloroethenes: a review. Environ Sci Pollut Res Int 24(15):13262-13283 doi:10.1007/s11356-
571 017-8867-y
572 Doronina N, Kaparullina E, Trotsenko Y (2014) The Family *Methylophilaceae*. In: Rosenberg E, DeLong EF,
573 Lory S, Stackebrandt E, Thompson F (eds) The Prokaryotes: *Alphaproteobacteria* and
574 *Betaproteobacteria*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 869-880
575 Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, Huttenhower C, Langille MGI
576 (2020) PICRUSt2 for prediction of metagenome functions. Nat Biotechnol 38(6):685-688
577 doi:10.1038/s41587-020-0548-6
578 Dowle M, Srinivasan A (2021) data.table: Extension of `data.frame`. R package version 1.14.0.
579 doi:https://CRAN.R-project.org/package=data.table
580 Eguchi M, Kitagawa M, Suzuki Y, Nakamuara M, Kawai T, Okamura K, Sasaki S, Miyake Y (2001) A field
581 evaluation of in situ biodegradation of trichloroethylene through methane injection. Water
582 Res.: 2145-2152
583 Folsom BR, Chapman PJ, Pritchard PH (1990) Phenol and trichloroethylene degradation by *Pseudomonas*
584 *cepacia* G4: kinetics and interactions between substrates. Appl Environ Micro 56(5):1279-85
585 doi:10.1128/aem.56.5.1279-1285.1990
586 Fowler SJ, Gutierrez-Zamora ML, Manefield M, Gieg LM (2014) Identification of toluene degraders in a
587 methanogenic enrichment culture. FEMS Microbiol Ecol 89(3):625-36 doi:10.1111/1574-
588 6941.12364
589 Fries MR, Forney LJ, Tiedje JM (1997) Phenol-and toluene degrading microbial populations from an
590 aquifer in which successful trichloroethene cometabolism occurred. Appl Environ Microbiol
591 63:1523–1530
592 Gibson DT, Hensley M, Yoshioka H, Mabry TJ (1970) Formation of (+)-cis-2,3-dihydroxy-1-
593 methylcyclohexa-4,6-diene from toluene by Pseudomonas putida. Biochem 9:1626–1630
594 Gu Z, Gu L, Eils R, Schlesner M, Brors B (2014) circlize Implements and enhances circular visualization in
595 R. Bioinformatics 30(19):2811-2 doi:10.1093/bioinformatics/btu393
596 Guo GL, Tseng DH, Huang SL (2001) Co-metabolic degradation of trichloroethylene by *Pseudomonas*
597 *putida* in a fibrous bed bioreactor. Biotechnol Lett 23:1653–1657
598 Haigler BE, Spain JC (1991) Biotransformation of nitrobenzene by bacteria containing toluene
599 degradative pathways. Appl Environ Micro 57(11):3156-62 doi:10.1128/aem.57.11.3156-
600 3162.1991
601 Hino S, Watanabe K, Takahashi N (1998) Phenol hydroxylase cloned from *Ralstonia eutropha* strain E2
602 exhibits novel kinetic properties. Micro 144 (Pt 7):1765-1772 doi:10.1099/00221287-144-7-
603 1765

Holme AJ, Costello A, Lidstrom ME, Murrell JC (1995) Evidence that particulate methane monooxygenase and ammonia monooxygenase may be evolutionarily related FEMS Micro Lett 132:203-208
Hooper AB, Vannelli T, Bergmann DJ, Arciero DM (1997) Enzymology of the oxidation of ammonia to nitrite by bacteria. Ant Van Lee 71(1-2):59-67 doi:10.1023/a:1000133919203
Horvath RS (1972) Microbial co-metabolism and the degradation of organic compounds in nature. Bac Reviews 36(2):146
Hur H, Newman LM, Wackett LP, Sadowsky MJ (1997) Toluene 2-monooxygenase-dependent growth of *Burkholderia cepacia* G4/PR1 on diethyl ether. Appl Environ Micro 63(4):1606-9 doi:10.1128/aem.63.4.1606-1609.1997
IPCC (2007) Fourth Assessment Report: Climate Change PUblisher.
Jiang H, Chen Y, Murrell JC, Jiang P, Zhang C, Xing X-H, Smith TJ (2011) Methanotrophs: multifunctional Bacteria with promising applicationsin environmental bioengineering. In: Moo-Young M (ed) Comprehensive Biotechnology. vol Volume 6: Environmental Biotechnology and Safety. Elsevier B.V.
Kanehisa M (2002) The KEGG database. In Silico Simulation of Biological Processes 247:91-103
Karigar CS, Rao SS (2011) Role of microbial enzymes in the bioremediation of pollutants: a review. Enzy Res 2011
Kassambara A (2020) ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. doi:https://CRAN.R-project.org/package=ggpubr
Kassambara A (2021) rstatix: Pipe-Friendly Framework for Basic Statistical Tests R package version 0.7.0. doi:https://CRAN.R-project.org/package=rstatix
Kauppera T, Mendesb LW, Harniszc M, Kraused SMB, Horn MA, Adrian Ho A (2021) Recovery of Methanotrophic Activity Is Not Reflected in the Methane-Driven Interaction Network after Peat Mining. Appl Env Micro 87
Kim H, An S, Park YR, Jang H, Yoo H, Park SH, Lee SJ, Cho US (2019) MMOD-induced structural changes of hydroxylase in soluble methane monooxygenase. Sci Adv 5(10):eaax0059 doi:10.1126/sciadv.aax0059
Klecka GM, Gibson DT (1981) Inhibition of catechol 2,3-dioxygenase from *Pseudomonas putida* by 3-chlorocatechol. Appl Environ Microbiol 41:1159-1165
Knief C (2015) Diversity and habitat preferences of cultivated and uncultivated aerobic methanotrophic bacteria evaluated based on *pmoA* as molecular marker. Front Micro 6:1346
Koo CW, Rosenzweig AC (2021) Biochemistry of aerobic biological methane oxidation. Chem Soc Rev 50(5):3424-3436 doi:10.1039/d0cs01291b
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 Micro 79(17):5112-20 doi:10.1128/AEM.01043-13
Kuever J (2014). In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (eds) The Prokaryotes. pp 281-288
Kunapuli U, Jahn MK, Lueders T, Geyer R, Heipieper HJ, Meckenstock RU (2010) *Desulfitobacterium aromaticivorans* sp. nov. and *Geobacter toluenoxydans* sp. nov., iron-reducing bacteria capable of anaerobic degradation of monoaromatic hydrocarbons. Int J Syst Evol Microbiol 60(Pt 3):686-695 doi:10.1099/ijs.0.003525-0
Lahti L, Shetty S (2017) Tools for microbiome analysis in R. doi:URL: http://microbiome.github.io
Leahy JG, Byrne AM, Olsen RH (1996) Comparison of factors influencing trichloroethylene degradation by toluene-oxidizing bacteria. Appl Environ Micro 62(3):825-33 doi:10.1128/aem.62.3.825-833.1996

651 Leuthner B, Heider J (2000) Anaerobic toluene catabolism of *Thauera aromatica*: the bbs operon codes
652 for enzymes of beta oxidation of the intermediate benzylsuccinate. Journal of Bacteriology
653 182(2):272-7 doi:10.1128/JB.182.2.272-277.2000
654 Leuthner B, Leutwein C, Schulz H, Horth P, Haehnel W, Schiltz E, Schagger H, Heider J (1998) Biochemical
655 and genetic characterization of benzylsuccinate synthase from *Thauera aromatica*: a new glycyl
656 radical enzyme catalysing the first step in anaerobic toluene metabolism. Mol Microbiol
657 28(3):615-28 doi:10.1046/j.1365-2958.1998.00826.x
658 Li F, Deng D, Zeng L, Abrams S, Li M (2021) Sequential anaerobic and aerobic bioaugmentation for
659 commingled groundwater contamination of trichloroethene and 1,4-dioxane. Sci Total Environ
660 774:145118 doi:10.1016/j.scitotenv.2021.145118
661 Louca S, Doebeli M (2018) Efficient comparative phylogenetics on large trees. Bioinformatics 34(6):1053-
662 1055 doi:10.1093/bioinformatics/btx701
663 Lovley DR, Giovannoni SJ, White DC, Champine JE, Phillips EJ, Gorby YA, Goodwin S (1993) *Geobacter*
664 *metallireducens* gen. nov. sp. nov., a microorganism capable of coupling the complete oxidation
665 of organic compounds to the reduction of iron and other metals. Arch Microbiol 159(4):336-44
666 doi:10.1007/BF00290916
667 Mahendra S, Alvarez-Cohen L (2006) Kinetics of 1,4-dioxane biodegradation by monooxygenase-
668 expressing bacteria. Environmental Science & Technology 40(17):5435-42
669 doi:10.1021/es060714v
670 Mahendra S, Grostern A, Alvarez-Cohen L (2013) The impact of chlorinated solvent co-contaminants on
671 the biodegradation kinetics of 1, 4-dioxane. Chemosphere 91(1):88-92
672 Martin KE, Ozsvar J, Coleman NV (2014) SmoXYB1C1Z of *Mycobacterium* sp. strain NBB4: a soluble
673 methane monooxygenase (sMMO)-like enzyme, active on C2 to C4 alkanes and alkenes. Appl
674 Environ Micro 80(18):5801-6 doi:10.1128/AEM.01338-14
675 McClay K, Fox BG, Steffan RJ (1996) Chloroform mineralization by toluene-oxidizing bacteria. Appl
676 Environ Micro 62(8):2716-22 doi:10.1128/aem.62.8.2716-2722.1996
677 McClay K, Fox BG, Steffan RJ (2000) Toluene monooxygenase-catalyzed epoxidation of alkenes. Appl
678 Environ Micro 66(5):1877-82 doi:10.1128/AEM.66.5.1877-1882.2000
679 McMurdie PJ, Holmes S (2013) phyloseq: an R package for reproducible interactive analysis and graphics
680 of microbiome census data. Plos One 8(4):e61217 doi:10.1371/journal.pone.0061217
681 Merkx M, Lippard SJ (2002) Why OrfY? Characterization of MMOD, a long overlooked component of the
682 soluble methane monooxygenase from *Methylococcus capsulatus* (Bath). Journal of Biological
683 Chemistry 277(8):5858-65 doi:10.1074/jbc.M107712200
684 Murrell JC, McDonald IR, Gilbert B (2000) Regulation of expression of methane monooxygenases by
685 copper ions. Trends Microbiol 8(5):221-5 doi:10.1016/s0966-842x(00)01739-x
686 Nelson MJ, Montgomery SO, Mahaffey WR, Pritchard PH (1987) Biodegradation of trichloroethylene and
687 involvement of an aromatic biodegradative pathway. Appl Environ Micro 53(5):949-54
688 doi:10.1128/aem.53.5.949-954.1987
689 Nelson MJ, Montgomery SO, O'Neill E J, Pritchard PH (1986) Aerobic metabolism of trichloroethylene by
690 a bacterial isolate. Appl Environ Micro 52(2):383-4 doi:10.1128/aem.52.2.383-384.1986
691 Neuwirth E (2014) RColorBrewer: ColorBrewer Palettes. doi:https://CRAN.R-
692 project.org/package=RColorBrewer
693 Newman LM, Wackett LP (1995) Purification and characterization of toluene 2-monooxygenase from
694 *Burkholderia cepacia* G4. Biochemistry 34(43):14066-76 doi:10.1021/bi00043a012
695 Newman LM, Wackett LP (1997) Trichloroethylene oxidation by purified toluene 2-monooxygenase:
696 products, kinetics, and turnover-dependent inactivation. J Bacter 179:90-97

697 Nordlund I, Powlowski J, Shingler V (1990) Complete nucleotide sequence and polypeptide analysis of
698 multicomponent phenol hydroxylase from *Pseudomonas* sp. strain CF600. Journal of
699 Bacteriology 172(12):6826-33 doi:10.1128/jb.172.12.6826-6833.1990
700 Norton JM (2011) Diversity and environmental distribution of ammonia-oxidizing bacteria. In: Ward BB,
701 Arp DJ, Klotz MG (eds) Nitrification. Amer Soc Micro, pp 39-55
702 Oldenhuis R, Vink RL, Janssen DB, Witholt B (1989) Degradation of chlorinated aliphatic hydrocarbons by
703 *Methylosinus trichosporium* OB3b expressing soluble methane monooxygenase. Appl Environ
704 Micro 55(11):2819-26 doi:10.1128/aem.55.11.2819-2826.1989
705 Olsen RH, Kukor JJ, Byrne AM, Johnson GR (1997) Evidence for the evolution of a single component
706 phenol/cresol hydroxylase from a multicomponent toluene monooxygenas. J Ind Microbiol
707 Biotechnol 19:360-368
708 Olsen RH, Kukor JJ, Kaphammer B (1994) A novel toluene-3-monooxygenase pathway cloned from
709 *Pseudomonas pickettii* PKO1. Journal of Bacteriology 176(12):3749-56
710 doi:10.1128/jb.176.12.3749-3756.1994
711 Oppenheim SF, Studts JM, Fox BG, Dordick JS (2001) Aromatic hydroxylation catalyzed by toluene 4-
712 monooxygenase in organic solvent/aqueous buffer mixtures. Appl Biochem Biotechnol
713 90(3):187-97 doi:10.1385/abab:90:3:187
714 Parales RE, Parales JV, Pelletier DA, Ditty JL (2008) Diversity of microbial toluene degradation pathways.
715 64:1-73, 2 p following 264 doi:10.1016/S0065-2164(08)00401-2
716 Pedersen TL (2020) patchwork: The Composer of Plots. doi:https://CRAN.R-
717 project.org/package=patchwork
718 Pfiffner SM, Palumbo AV, Phelps TJ, Hazen TC (1997) Effects of nutrient dosing on subsurface
719 methanotrophic populations and trichloroethylene degradation. J Ind Microbiol Biotechnol 18(2-
720 3):204-12 doi:10.1038/sj.jim.2900350
721 Pikus JD, Studts JM, McClay K, Steffan RJ, Fox BG (1997) Changes in the regiospecificity of aromatic
722 hydroxylation produced by active site engineering in the diiron enzyme toluene 4-
723 monooxygenase. Biochemistry 36(31):9283-9 doi:10.1021/bi971049t
724 Prosser JI, Head IM, Stein LY (2014) The Family *Nitrosomonadaceae*. In: Rosenberg E, DeLong EF, Lory S,
725 Stackebrandt E, Thompson F (eds) The Prokaryotes *Alphaproteobacteria* and
726 *Betaproteobacteria*. Springer-Verlag, Berlin, pp 901–918
727 Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glockner FO (2007) SILVA: a comprehensive
728 online resource for quality checked and aligned ribosomal RNA sequence data compatible with
729 ARB. Nucl Acid Res 35(21):7188-96 doi:10.1093/nar/gkm864
730 R Core Team (2018) R: A language and environment for statistical computing. R Foundation for
731 Statistical Computing, Vienna, Austria.
732 Rasche ME, Hyman MR, Arp DJ (1991) Factors limiting aliphatic chlorocarbon degradation by
733 *Nitrosomonas europaea*: cometabolic inactivation of ammonia monooxygenase and substrate
734 specificity. Appl Environ Micro 57(10):2986-94 doi:10.1128/aem.57.10.2986-2994.1991
735 RStudio Team (2020) RStudio: Integrated Development for R. RStudio, PBC. Boston, MA
736 Ryoo D, Shim H, Arenghi FL, Barbieri P, Wood TK (2001) Tetrachloroethylene, trichloroethylene, and
737 chlorinated phenols induce toluene-o-xylene monooxygenase activity in *Pseudomonas stutzeri*
738 OX1. Appl Microbiol Biotechnol 56(3-4):545-9 doi:10.1007/s002530100675
739 Ryoo D, Shim H, Canada K, Barberi P, Wood TK (2000) Aerobic degradation of tetrachloroethylene by
740 toluene-o-monooxygenase of *Pseudomonas stutzeri* OX1. Nat Biotechnol 18:775–778.
741 Sazinsky MH, Merkx M, Cadieux E, Tang S, Lippard SJ (2004) Preparation and X-ray structures of metal-
742 free, dicobalt and dimanganese forms of soluble methane monooxygenase hydroxylase from
743 *Methylococcus capsulatus* (Bath). Biochemistry 43(51):16263-76 doi:10.1021/bi048140z
744 Schloss PD (2009) A high-throughput DNA sequence aligner for microbial ecology studies. Plos One 4(12)

745 Semrau JD, Jagadevan S, DiSpirito AA, Khalifa A, Scanlan J, Bergman BH, Freemeier BC, Baral BS, Bandow
746 NL, Vorobev A, Haft DH, Vuilleumier S, Murrell JC (2013) Methanobactin and MmoD work in
747 concert to act as the 'copper-switch' in methanotrophs. 15(11):3077-86 doi:10.1111/1462-
748 2920.12150
749 Semrini L, Hopkins GD, Roberts PV, Grbicgalic D, McCarty PL (1991) A field evaluation of in situ
750 biodegradation of chlorinated ethens 3. studies of competitive inhibition. Groundwater 29:239-
751 250
752 Shao Y, Hatzinger PB, Streger SH, Rezes RT, Chu KH (2019) Evaluation of methanotrophic bacterial
753 communities capable of biodegrading trichloroethene (TCE) in acidic aquifers. Biodegradation
754 30(2-3):173-190 doi:10.1007/s10532-019-09875-w
755 Sharp JO, Wood TK, Alvarez-Cohen L (2005) Aerobic biodegradation of N-nitrosodimethylamine (NDMA)
756 by axenic bacterial strains. Biotechnol Bioeng 89(5):608-18 doi:10.1002/bit.20405
757 Shaw JP, Harayama S (1995) Characterization in vitro of the hydroxylase component of xylene
758 monooxygenase, the first enzyme of the TOL-plasmid-encoded pathway for the mineralization
759 of toluene and xylenes. J Ferment Bioeng 79:195-199
760 Shields MS, Francesconi SC (1996) Microbial degradation of trichloroethylene, dichloroethylenes, and
761 aromatic pollutants.,
762 Shields MS, Reagin MJ, Gerger RR, Campbell R, Somerville C (1995) TOM, a new aromatic degradative
763 plasmid from *Burkholderia (Pseudomonas) cepacia* G4. Appl Environ Micro 61(4):1352-6
764 doi:10.1128/aem.61.4.1352-1356.1995
765 Shim H, Wood TK (2000) Aerobic degradation of mixtures of chlorinated aliphatics by cloned toluene-o-
766 xylene monooxygenase and toluene o-monooxygenase in resting cells. Biotechnol Bioeng
767 70(6):693-8 doi:10.1002/1097-0290(20001220)70:6<693::aid-bit12>3.0.co;2-w
768 Smith GJ, Angle JC, Solden LM, Borton MA, Morin TH, Daly RA, Johnston MD, Stefanik KC, Wolfe R, Gil B,
769 Wrighton KC (2018) Members of the genus *Methylobacter* are inferred to account for the
770 majority of aerobic methane oxidation in oxic soils from a freshwater wetland. mBio 9(6)
771 doi:10.1128/mBio.00815-18
772 Spormann AM, Widdel F (2000) Metabolism of alkylbenzenes, alkanes, and other hydrocarbons in
773 anaerobic bacteria. Biodegradation 11(2-3):85-105 doi:10.1023/a:1011122631799
774 Stainthorpe AC, Lees V, Salmond GP, Dalton H, Murrell JC (1990) The methane monooxygenase gene
775 cluster of *Methylococcus capsulatus* (Bath). Gene 91(1):27-34 doi:10.1016/0378-1119(90)90158-
776 n
777 Stainthorpe AC, Murrell JC, Salmond GP, Dalton H, Lees V (1989) Molecular analysis of methane
778 monooxygenase from *Methylococcus capsulatus* (Bath). Arch Microbiol 152(2):154-9
779 doi:10.1007/BF00456094
780 Sun AK, Hong J, Wood TK (1997) Trichloroethylene mineralization in a fixed-film bioreactor using a pure
781 culture expressing constitutively toluene ortho-monooxygenase. Biotechnol Bioeng 55: 674–685
782 Sutfin JA, Ramey D (1997) In situ biological treatment of TCE-impacted soil and groundwater:
783 Demonstration results. Env Progress 16:287-296
784 Suzuki M, Hayakawa T, Shaw JP, Rekik M, Harayama S (1991) Primary structure of xylene
785 monooxygenase: similarities to and differences from the alkane hydroxylation system. Journal of
786 Bacteriology 173(5):1690-5 doi:10.1128/jb.173.5.1690-1695.1991
787 Theisen AR, Ali MH, Radajewski S, Dumont MG, Dunfield PF, McDonald IR, Dedysh SN, Miguez CB,
788 Murrell JC (2005) Regulation of methane oxidation in the facultative methanotroph *Methylocella*
789 *silvestris* BL2. Mol Microbiol 58(3):682-92 doi:10.1111/j.1365-2958.2005.04861.x
790 Thelusmond JR, Kawka E, Strathmann TJ, Cupples AM (2018) Diclofenac, carbamazepine and triclocarban
791 biodegradation in agricultural soils and the microorganisms and metabolic pathways affected.
792 640-641:1393-1410 doi:10.1016/j.scitotenv.2018.05.403

Thelusmond JR, Strathmann TJ, Cupples AM (2016) The identification of carbamazepine biodegrading phylotypes and phylotypes sensitive to carbamazepine exposure in two soil microbial communities. 571:1241-52 doi:10.1016/j.scitotenv.2016.07.154

Unger V, Liebner S, Koebsch F, Yang S, Horn F, Sachs T, Kallmeyer J, Knorr K-H, Rehderf G, Gottschalk P, Jurasinski G (2021) Congruent changes in microbial community dynamics and ecosystem methane fluxes following natural drought in two restored fens. Soil Bio Biochem 160

van Hylckama Vlieg JE, Leemhuis H, Spelberg JH, Janssen DB (2000) Characterization of the gene cluster involved in isoprene metabolism in *Rhodococcus* sp. strain AD45. Journal of Bacteriology 182(7):1956-63 doi:10.1128/JB.182.7.1956-1963.2000

Vannelli T, Logan M, Arciero DM, Hooper AB (1990) Degradation of halogenated aliphatic compounds by the ammonia- oxidizing bacterium *Nitrosomonas europaea*. Appl Environ Micro 56(4):1169-71 doi:10.1128/aem.56.4.1169-1171.1990

Vorobev AV, Baani M, Doronina NV, Brady AL, Liesack W, Dunfield PF, Dedysh SN (2011) *Methyloferula stellata* gen. nov., sp. nov., an acidophilic, obligately methan otrophic bacterium that possesses only a soluble methane monooxygenase. Int J Syst Evol Microbiol 61(10):2456–2463

Wackett LP (1996) Co-metabolism: is the emperor wearing any clothes? Curr Opin Biotechnol 7(3):321-325

Wendeborn S (2020) The chemistry, biology, and modulation of ammonium nitrification in soil. Angew Chem Int Ed Engl 59(6):2182-2202 doi:10.1002/anie.201903014

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York

Wickham H (2021a) Forcats: tools for working with categorical variables (factors). doi:https://CRAN.R-project.org/package=forcats

Wickham H (2021b) tidyr: Tidy Messy Data. doi: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. 4(43), 1686, doi:https://doi.org/10.21105/joss.01686

Wickham H, Bryan J (2019) readxl: read excel files. R package version 1.3.1 ed. doi: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. doi:https://CRAN.R-project.org/package=dplyr

Widdel F, Rabus R (2001) Anaerobic biodegradation of saturated and aromatic hydrocarbons. Curr Opin Biotechnol 12(3):259-76 doi:10.1016/s0958-1669(00)00209-3

Winter RB, Yen K-M, Ensley BD (1989) Efficient degradation of trichloroethylene by a recombinant *Escherichia coli*. Bio/Technology 7:282–285

Ye Y, Doak TG (2009) A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. PLoS Comput Biol 5(8):e1000465 doi:10.1371/journal.pcbi.1000465

Yen KM, Karl MR, Blatt LM, Simon MJ, Winter RB, Fausset PR, Lu HS, Harcourt AA, Chen KK (1991) Cloning and characterization of a *Pseudomonas mendocina* KR1 gene cluster encoding toluene-4-monooxygenase. Journal of Bacteriology 173(17):5315-27 doi:10.1128/jb.173.17.5315-5327.1991

Zylstra GJ, Gibson DT (1989) Toluene degradation by *Pseudomonas putida* F1: Nucleotide sequence of the todC1C2BADE genes and their expression in *E. coli*. J Biol Chem 264:14940–14946

Zylstra GJ, McCombie WR, Gibson DT, Finette BA (1988) Toluene degradation by *Pseudomonas putida* F1: genetic organization of the tod operon. Appl Environ Microbiol 54(6):1498-503 doi:10.1128/aem.54.6.1498-1503.1988

840

841