Current Microbiology (2024) 81:314
https://doi.org/10.1007/s00284-024-03829-z

ORIGINAL PAPER

Propane Monooxygenases in Soil Associated Metagenomes Align Most Closely to those in the Genera *Kribbella, Amycolatopsis, Bradyrhizobium, Paraburkholderia* and *Burkholderia*

Alison M. Cupples[1]

Received: 24 June 2024 / Accepted: 2 August 2024

Abstract
Propanotrophs are a focus of interest because of their ability to degrade numerous environmental contaminants. To explore the phylogeny of microorganisms containing the propane monooxygenase gene cluster (*prmABCD*), NCBI bacterial genomes and publicly available soil associated metagenomes (from soils, rhizospheres, tree roots) were both examined. Nucleic acid sequences were collected only if all four subunits were located together, were of the expected length and were annotated as propane monooxygenase subunits. In the bacterial genomes, this resulted in data collection only from the phyla *Actinomycetota* and *Pseudomonadota.* For the soil associated metagenomes, reads from four studies were subject to quality control, assembly and annotation. Following this, the propane monooxygenase subunit nucleic acid sequences were collected and aligned to the collected bacterial sequences. In total, forty-two propane monooxygenase gene clusters were annotated from the soil associated metagenomes. The majority aligned closely to those from the *Actinomycetota,* followed by the *Alphaproteobacteria,* then the *Betaproteobacteria*. *Actinomycetota* aligning propane monooxygenase sequences were obtained from all four datasets and most closely aligned to the genera *Kribbella* and *Amycolatopsis*. *Alphaproteobacteria* aligning sequences largely originated from metagenomes associated with miscanthus and switchgrass rhizospheres and primarily aligned with the genera *Bradyrhizobium*, *Acidiphilium* and unclassified *Rhizobiales. Betaproteobacteria* aligning sequences were obtained from only the Red Oak root metagenomes and primarily aligned with the genera *Paraburkholderia, Burkholderia* and *Caballeronia*. Interestingly, sequences from the environmental metagenomes were not closely aligned to those from well-studied propanotrophs, such as *Mycobacterium* and *Rhodococcus*. Overall, the study highlights the previously unreported diversity of putative propanotrophs in environmental samples. The common occurrence of propane monooxygenase gene clusters has implications for their potential use for contaminant biodegradation.

Introduction

Propanotrophs grow on propane (C_3H_8) as a sole source of carbon and energy [1, 2] and are a focus of interest because of their ability to transform a range of environmental contaminants. For example, five *Mycobacterium* strains growing on propane degraded trichloroethene and *Mycobacterium vaccae* JOB5 degraded *cis*-dichloroethene, *trans*-dichloroethene, 1,1-dichloroethene and vinyl chloride [1]. Similarly, after growth on propane, several isolates (including *Mycobacterium vaccae* JOB5) degraded methyl tert-butyl ether (MTBE), ethyl tert-butyl ether and tert-amyl methyl ether [2]. Propanotrophs (*Rhodococcus* sp. RR1 and *Mycobacterium vaccae* JOB5) have also been associated with the biodegradation of n-nitrosodimethylamine (NDMA) [3]. Further, propane grown *Rhodococcus aetherivorans* TPA co-metabolized 1,1,2,2-tetrachloroethane [4], four propane oxidizing bacteria (*Rhodococus jostii* RHA1, *Mycobacterium vaccae* JOB5, *Rhodococcus rubber* ENV425, *Sphingopyxis* sp. AX-A) degraded 1,2,3-trichloropropane [5] and propane induced isolates (*Mycobacterium vaccae* JOB5 and *Rhodococcus jostii* RHA1) degraded 1,4-dioxane [6]. Propane amended mixed cultures have also been linked to the biodegradation of organic contaminants. Propane and inorganic nutrients stimulated the biodegradation of 1,2-dibromoethane in microcosms constructed with aquifer solids and groundwater [7] and a propanotrophic enrichment culture

✉ Alison M. Cupples
cupplesa@msu.edu

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

Published online: 20 August 2024

derived from the site materials also degraded 1,2-dichloroethane [7].

Propanotrophs have also facilitated the removal of contaminants in situ. At Vandenberg Space Force Base, propane biosparging and bioaugmentation (with *Rhodococcus ruber* ENV425) were successfully applied to promote in situ biodegradation of 1,4-dioxane [8, 9]. At another site (Lansing, Mi), propane biosparging resulted in approximately 70% to 99% reductions in 1,4-dioxane concentrations at key monitoring locations [10, 11]. The ability of propanotrophs to degrade 1,4-dioxane is particularly advantageous, because methanotrophs appear unable to degrade this chemical, both in pure cultures (*Methylosinus trichosporium OB3b, Methylomonas methanica strain 68–1, Methylocella palustris strain K*) and in methane-oxidizing mixed cultures enriched from aquifers [12]. Another key advantage to propanotrophs is their ability to degrade contaminants to low levels. Propane biosparging was effective in reducing groundwater NDMA to ng/L concentrations [13] and propane amendments resulted in 1,4-dioxane removal to below the detection limit (0.06 µg/L) in a pilot-scale water reuse system [14].

Propane monooxygenases have been found within the group 5 (propane-2-monooxygenase) [15] and group 6 (propane-1-monooxygenase) [16–18] soluble di-iron monooxygenases [19]. The enzyme is encoded by four subunits: propane monooxygenase large subunit (*prmA*), propane monooxygenase reductase subunit (*prmB*), propane monooxygenase small subunit (*prmC*) and propane monooxygenase coupling protein (*prmD*) [20]. Many studies have provided direct evidence for the role of propane monooxygenase in contaminant biodegradation. For example, an inducible propane monooxygenase was responsible for NDMA biodegradation in *Rhodococcus* sp. strain RHA1 [20]. Propane monooxygenase from *Mycobacterium vaccae* JOB5 was associated with the biodegradation of MTBE [21] and was also linked to the biodegradation of 1,4-dioxane in *Mycobacterium dioxanotrophicus* PH-06 [8, 16]. Propane monooxygenase sequences similar to *Rhodococcus* sp. RR1 *prmA* were associated with 1,4-dioxane biodegradation in mixed microbial communities [22].

Here, the overall objective was to determine the diversity of microorganisms (bacterial genomes and in environmental metagenomes) containing all four propane monooxygenase subunits. To date, the majority of research has focussed on propane monooxygenases from the *Actinomycetota,* primarily the genera *Rhodococcus, Mycobacterium* and *Gordonia*. The approach adopted was two-fold, first propane monooxygenase subunit nucleic sequences were collected from bacterial genomes available from NCBI. Then, sequencing reads were downloaded from four publicly available whole genome sequencing (WGS) projects involving soils, rhizospheres and tree roots. This involved downloading Sequence Read Archives (SRA) files from NCBI, quality control, assembly and annotation. The sequences obtained from the WGS sequencing projects were then aligned to those from the NCBI bacterial genomes. The data generated represent a wealth of information on the propane monooxygenase gene cluster in bacterial genomes and their common occurrence in environmental samples. The analysis indicates these genes are present in both the *Alphaproteobacteria* and the *Betaproteobacteria* and in many previously unreported genera in the *Actinomycetota.* The results are particularly relevant due to the potential of propane monooxygenases to degrade environmental contaminants.

Methods

Propane Monooxygenase Subunit Sequences from NCBI Bacterial Genomes

Nucleic acid sequences were collected from NCBI (https://www.ncbi.nlm.nih.gov/) using the search words "propane monooxygenase". Subunit sequences were only collected if 1) all four subunits were close together, 2) all subunits were within a range of expected sequence length and 3) all were annotated as propane monooxygenase subunits. The majority of sequences (except for those found in 4 *Mycobacterium*, 1 *Mycolicibacterium* and 2 *Rhodococcus* strains) occurred in the following order: propane monooxygenase large subunit (*prmA*), propane monooxygenase reductase subunit (*prmB*), propane monooxygenase small subunit (*prmC*), propane monooxygenase coupling protein (*prmD*). The range of typical subunits lengths were *prmA* 1650–1670 bp, *prmB* 1030–1100 bp, *prmC* 1085–1200 bp and *prmD* 350–380 bp. The subunit order in the *Mycobacterium*, *Mycolicibacterium* and *Rhodococcus* strains was *prmA, prmC, prmD, prmB.* Subunit sequences from the bacterial genomes were downloaded from NCBI and stored as text files. Sequences from NCBI from uncultured microorganisms were not included in the analysis. Additional propane monooxygenase sequences were collected from the literature (or the corresponding authors), e.g. those from *Mycobacterium* sp. ENV421 and *Rhodococcus* sp. ENV425 [23–25], *Mycobacterium dioxanotrophicus* strain PH-06 [16, 26] and *Mycobacterium vaccae* JOB5 [21].

Whole Genome Sequencing (WGS) Datasets

Four soil associated WGS datasets were selected (from Red Oak roots, agricultural soil, rhizosphere soil) for analysis primarily based on the larger sizes of the sequencing files (> 30 Gbases). Preliminary data analysis indicated smaller

sequencing files often did not result in the assembly of the four subunits onto the same contig. Three datasets were from studies conducted within Michigan and the fourth was from data collected in the United Kingdom. Select metadata (e.g. Run, BioProject, BioSample numbers, bases, bytes, instrument) for the four WGS datasets are summarized in the supplementary section (Supplementary Tables 1–4). Additional metadata is available at https://www.ncbi.nlm.nih.gov/. Brief details on the WGS datasets are provided below.

The forest soil WGS dataset was derived from a study that examined interactions between inorganic nitrogen availability, soil organic matter and fungal community composition [27]. The sequencing files were from DNA extracted from root tips in soils from Manistee National Forest (northern Lower Michigan). The plots sampled were in even-aged (~ 100 year-old) second-growth forests on uniformly sandy soils (~ 85% sand) [27–29]. Details on soil sampling, DNA extraction and WGS have previously been reported [27]. Only a subset of the files (the largest files) were analysed in the current study and all were associated with Red Oak root tips.

The three agricultural WGS datasets included two from Michigan and one from the United Kingdom. One Michigan dataset was from a study that investigated the impact of nitrogen fertilizer rates on switchgrass soil microbial communities [30]. The switchgrass nitrogen rate experiment is part of the Great Lakes Bioenergy Research Center (GLBRC), located at the W. K. Kellogg Biological Station (KBS) study site (Michigan State University site for field experimental research). The study applied three nitrogen fertilizer rates (0, 56, 196 kg N/ha) annually in switchgrass grown for bioenergy production [30]. The predominant soil series at KBS is Kalamazoo loam (Fine-Loamy, Mixed, Semiactive, Mesic Typic Hapludalfs) [30]. The current study examined WGS files from the unfertilized sample (0N) and excess fertilized sample (196N), each with four replicates, at plant peak productivity stage. Again, details on soil sampling, DNA extraction and sequencing have previously been reported [30]. The other Michigan dataset also originated from research conducted at the GLBRC at KBS [31, 32]. In this case, the project examined soil rhizosphere microbial communities associated with three biofuel crops (switchgrass, corn and miscanthus) [31, 32]. The authors sampled rhizosphere of each crop at seven plot areas. The researchers considered soil closely attached to roots (< 1 mm) to be rhizosphere soil. Additional details on sample collection, DNA extraction and sequencing have previously been provided [31, 32]. The third agricultural soil WGS dataset originated from a soil microbial community study involving three long term treatments at the Rothamsted Highfield Ley-Arable field experiment in the United Kingdom [33]. The authors reported the soil is a silty clay loam (25% clay, 62% silt, 13% sand) (Chromic Luvisol). DNA was extracted from plots that had been managed consistently as bare fallow, arable (continuous winter wheat, at the time of sampling) or mixed grass swards [33]. Additional details on the soil plots, DNA extraction and sequencing have also been previously reported [33].

WGS Data Analysis

The WGS sequencing files (SRA files) from the four studies were downloaded from NCBI using the SRA-Toolkit (Version 3.0.3) [34] to a directory in the High Performance Computing Cluster (HPCC) at MSU. The SRA files were then uploaded to the United States Department of Energy Systems Biology Knowledgebase (KBase) [35]. The uploaded files were subject to quality control and filtering using FastQC (Version 0.12.1) [36] and Trimmomatic (Version 0.36) (sliding window size equal to 4 and sliding window minimum quality equal to 15) [37]. The files were then assembled using Megahit (Version 1.2.9) [38] with a minimum contig length of 2000 bp. The Megahit metagenomes were annotated with Prokka (Version 1.14.16) [39]. When present, the propane monooxygenase subunit sequences were collected and stored in text files. Similar to the bacterial genome data collection, sequences were only collected if all four propane monooxygenase subunits were present on the same contig, all four subunits were the correct size and were correctly annotated. In many cases, the large subunit (*prmA*) was smaller than expected and in those cases, sequences were not collected.

As an additional check on the WGS analysis, SRA files from the Red Oak root tips dataset were also analysed on the Galaxy web platform [40, 41]. For this, within the platform, the SRA files were downloaded from NCBI, processed with Trimmomatic [37], assembled with Megahit [38] (minimum contig length of 2000 bp) and annotated with Prokka [39].

Phylogenetic Trees and Gene Arrow Plots

Phylogenetic trees using nucleic acid sequences of the propane monooxygenase alpha unit (*prmA*) were generated using MEGA 11 [42]. In MEGA 11, this first involved aligning the sequences using ClustalW with the default parameters [43]. The best substitution model was determined to be General Time Reversible for every alignment. The trees were then created using the Maximum Likelihood method with 500 bootstrap replications.

Each tree legend has additional information on the tree construction. Following each nucleic acid alignment, the amino acid sequence alignment was checked for the presence of both di-iron centres (DE*RH). Five final *prmA* trees were generated. The first tree involved representative *prmA* sequences from the major bacterial genera (from

NCBI) found to contain the four propane monooxygenase subunits. Another three trees involved an alignment of the bacterial NCBI collected *prmA* sequences to those from the WGS datasets. For this, the first step was to divide WGS *prmA* sequences by phylogeny (based on BLAST searches for all four subunits). Then, the WGS *prmA* sequences were aligned with all collected NCBI *prmA* sequences for that bacterial group (either *Actinomycetota, Alphaproteobacteria* or *Betaproteobacteria*). Following this, the closest aligned bacterial NCBI *prmA* sequences were selected for the generation of the final three trees (divided by bacterial group). A final tree was generated to compare *Betaproteobacteria prmA* sequences generated by the two analysis methods (KBase and Galaxy). The orientation and size of all propane monooxygenase subunits were both illustrated in gene arrow plots generated with R (Version 4.2.1) [44] in RStudio (Version 2022.12.0) [45] and the R packages ggplot2 (Verion 3.4.4) [46], readxl (Version 1.4.2) [47] and gggenes (Version 0.5.1.) [48]. Gene arrow plots were generated for all propane monooxygenase sequences collected from NCBI as well as those collected for the WGS datasets. Gene sequences for all phylogenetic tree alignments are shown at the end of the Supplementary Section.

Results

Propane Monooxygenase Subunits in Bacterial Genomes

The NCBI search using "propane monooxygenase" resulted in 553 matches (under the category "nucleotides") with the classification Bacteria. From this, 269 were primarily from uncultured bacteria (the others only contained the word propane in the title of the submission) and were not investigated further. The remaining 284 were associated with *Actinomycetota* (182), *Alphaproteobacteria* (63), *Betaproteobacteria* (19), *Gammaproteobacteria* (13), *Firmicutes* (5) and CFB group bacteria (2). Each search result was individually examined for the presence of the four propane monooxygenase subunits. All selections from the *Gammaproteobacteria, Firmicutes* and CFB group did not contain any strain with all four subunits annotated as propane monooxygenase subunits and were not examined further. A summary of the search results is provided for the remaining three classifications (*Actinomycetota, Alphaproteobacteria* and *Betaproteobacteria*) containing genera with all four subunits (Fig. 1). The classifications

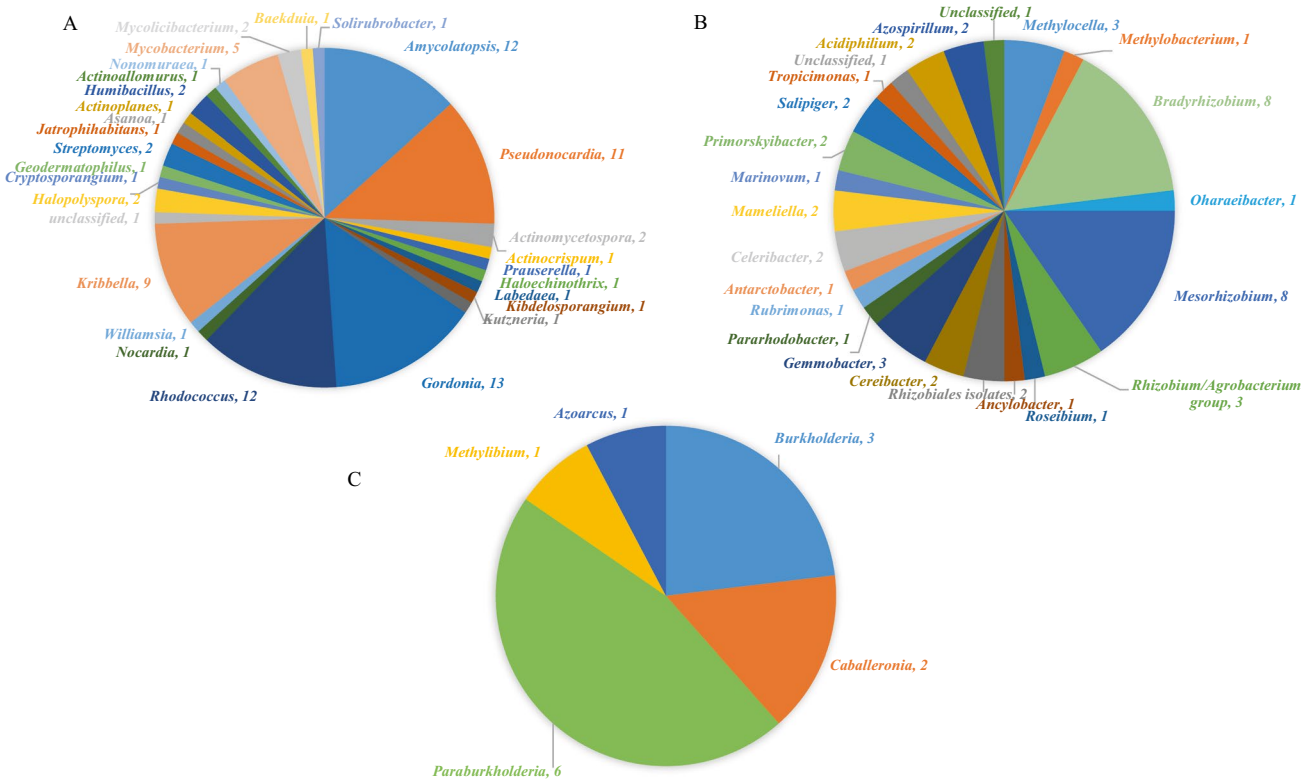

Fig. 1 The number of species or strains in each genus containing the four subunits of the propane monooxygenase operon in the *Actinomycetota* (**a**), *Alphaproteobacteria* (**b**) and *Betaproteobacteria* (**c**)

of each strain or species are also provided, for the *Actinomycetota* (Table 1) and the *Pseudomonadota* (Table 2).

The phylum *Actinomycetota* contained the largest number of species/strains with all four subunits (89 in total) (Fig. 1a), followed by the class *Alphaproteobacteria* (52) (Fig. 1b), then the class *Betaproteobacteria* (13) (Fig. 1c). In the *Actinomycetota,* the genera with the largest number of species/strains containing the four subunits included *Amycolatopsis* (12 strains, in the order *Pseudonocardiales*), *Pseudonocardia* (11, *Pseudonocardiales*), *Gordonia* (13, *Mycobacteriales*), *Mycobacterium* (5, *Mycobacteriales*), *Rhodococcus* (12, *Mycobacteriales*) and *Kribbella* (9, *Propionibacteriales*). All other genera contained only 1 or 2 strains/species and classified within numerous other orders and one other class (*Thermoleophilia*).

The class *Alphaproteobacteria* contained 52 species or strains with all four subunits (Fig. 1b). All genera were from three orders (*Hyphomicrobiales*, *Rhodobacterales*, *Rhodospirillales*) and multiple families (Table 2). The majority of genera only contained 1 or 2 stains/species with all four subunits. Five genera contained more than 2 species/strains with all four subunits, including *Bradyrhizobium* (8 species/strains), *Mesorhizobium* (8), *Gemmobacter* (3), *Methylocella* (3) and *Rhizobium* (3).

The class *Betaproteobacteria* only contained 13 species or strains with all four subunits (Fig. 1c). The majority classified within the order *Burkholderiales* and 1 classified within the order *Rhodocyclales* (Table 2). The genera with the largest number of species/strains within this class were *Paraburkholderia* (6), *Burkholderia* (3) and *Caballeronia* (2) (all within the family *Burkholderiaceae*).

Table 1 Phylogenetic classification and number of microorganisms containing the propane monooxygenase operon in the classes *Actinomycetes* and *Thermoleophilia*

Phylum	Class	Order	Family	Genus	Number of genomes with *prmABCD*
Actinomycetota	***Actinomycetes***	***Pseudonocardiales***	***Pseudonocardiaceae***	***Amycolatopsis***	***12***
	Actinomycetes	***Pseudonocardiales***	***Pseudonocardiaceae***	***Pseudonocardia***	***11***
	Actinomycetes	*Pseudonocardiales*	*Pseudonocardiaceae*	*Actinomycetospora*	*2*
	Actinomycetes	*Pseudonocardiales*	*Pseudonocardiaceae*	*Actinocrispum*	*1*
	Actinomycetes	*Pseudonocardiales*	*Pseudonocardiaceae*	*Prauserella*	*1*
	Actinomycetes	*Pseudonocardiales*	*Pseudonocardiaceae*	*Haloechinothrix*	*1*
	Actinomycetes	*Pseudonocardiales*	*Pseudonocardiaceae*	*Labedaea*	*1*
	Actinomycetes	*Pseudonocardiales*	*Pseudonocardiaceae*	*Kibdelosporangium*	*1*
	Actinomycetes	*Pseudonocardiales*	*Pseudonocardiaceae*	*Kutzneria*	*1*
	Actinomycetes	***Mycobacteriales***	***Gordoniaceae***	***Gordonia***	***13***
	Actinomycetes	***Mycobacteriales***	***Mycobacteriaceae***	***Mycobacterium***	***5***
	Actinomycetes	*Mycobacteriales*	*Mycobacteriaceae*	*Mycolicibacterium*	*2*
	Actinomycetes	***Mycobacteriales***	***Nocardiaceae***	***Rhodococcus***	***12***
	Actinomycetes	*Mycobacteriales*	*Nocardiaceae*	*Nocardia*	*1*
	Actinomycetes	*Mycobacteriales*	*Nocardiaceae*	*Williamsia*	*1*
	Actinomycetes	***Propionibacteriales***	***Kribbellaceae***	***Kribbella***	***9***
	Actinomycetes	*Propionibacteriales*	*unclassified*	*unclassified*	*1*
	Actinomycetes	*Actinomycetes incerta sedis*	*Actinomycetes incerta sedis*	*Halopolyspora*	*2*
	Actinomycetes	*Cryptosporangiales*	*Cryptosporangiaceae*	*Cryptosporangium*	*1*
	Actinomycetes	*Geodermatophilales*	*Geodermatophilaceae*	*Geodermatophilus*	*1*
	Actinomycetes	*Kitasatosporales*	*Streptomycetaceae*	*Streptomyces*	*2*
	Actinomycetes	*Jatrophihabitantales*	*Jatrophihabitantaceae*	*Jatrophihabitans*	*1*
	Actinomycetes	*Micromonosporales*	*Micromonosporaceae*	*Asanoa*	*1*
	Actinomycetes	*Micromonosporales*	*Micromonosporaceae*	*Actinoplanes*	*1*
	Actinomycetes	*Micrococcales*	*Intrasporangiaceae*	*Humibacillus*	*2*
	Actinomycetes	*Streptosporangiales*	*Thermomonosporaceae*	*Actinoallomurus*	*1*
	Actinomycetes	*Streptosporangiales*	*Streptosporangiaceae*	*Nonomuraea*	*1*
	Thermoleophilia	*Solirubrobacterales*	*Baekduiaceae*	*Baekduia*	*1*
	Thermoleophilia	*Solirubrobacterales*	*Solirubrobacteraceae*	*Solirubrobacter*	*1*

The genera in bold simply illustrate the largest numbers

Table 2 Phylogenetic classification and number of microorganisms containing the propane monooxygenase operon in the classes *Alphaproteobacteria* and *Betaproteobacteria*

Phylum	Class	Order	Family	Genus	Number of genomes with *prmABCD*
Pseudomonadota	*Alphaproteobacteria*	*Hyphomicrobiales*	*Beijerinckiaceae*	*Methylocella*	*3*
	Alphaproteobacteria	*Hyphomicrobiales*	*Methylobacteriaceae*	*Methylobacterium*	*1*
	Alphaproteobacteria	***Hyphomicrobiales***	***Nitrobacteraceae***	***Bradyrhizobium***	***8***
	Alphaproteobacteria	*Hyphomicrobiales*	*Pleomorphomonadaceae*	*Oharaeibacter*	*1*
	Alphaproteobacteria	***Hyphomicrobiales***	***Phyllobacteriaceae***	***Mesorhizobium***	***8***
	Alphaproteobacteria	*Hyphomicrobiales*	*Rhizobiaceae*	*Rhizobium/ Agrobacterium group*	*3*
	Alphaproteobacteria	*Hyphomicrobiales*	*Stappiaceae*	*Roseibium*	*1*
	Alphaproteobacteria	*Hyphomicrobiales*	*Xanthobacteraceae*	*Ancylobacter*	*1*
	Alphaproteobacteria	*Hyphomicrobiales*	*Unclassified Hyphomicrobiales*	*Rhizobiales isolates*	*2*
	Alphaproteobacteria	*Rhodobacterales*	*Paracoccaceae*	*Cereibacter*	*2*
	Alphaproteobacteria	*Rhodobacterales*	*Paracoccaceae*	*Gemmobacter*	*3*
	Alphaproteobacteria	*Rhodobacterales*	*Paracoccaceae*	*Pararhodobacter*	*1*
	Alphaproteobacteria	*Rhodobacterales*	*Paracoccaceae*	*Rubrimonas*	*1*
	Alphaproteobacteria	*Rhodobacterales*	*Roseobacteraceae*	*Antarctobacter*	*1*
	Alphaproteobacteria	*Rhodobacterales*	*Roseobacteraceae*	*Celeribacter*	*2*
	Alphaproteobacteria	*Rhodobacterales*	*Roseobacteraceae*	*Mameliella*	*2*
	Alphaproteobacteria	*Rhodobacterales*	*Roseobacteraceae*	*Marinovum*	*1*
	Alphaproteobacteria	*Rhodobacterales*	*Roseobacteraceae*	*Primorskyibacter*	*2*
	Alphaproteobacteria	*Rhodobacterales*	*Roseobacteraceae*	*Salipiger*	*2*
	Alphaproteobacteria	*Rhodobacterales*	*Roseobacteraceae*	*Tropicimonas*	*1*
	Alphaproteobacteria	*Rhodobacterales*	*Roseobacteraceae*	*Unclassified*	*1*
	Alphaproteobacteria	*Rhodospirillales*	*Acetobacteraceae*	*Acidiphilium*	*2*
	Alphaproteobacteria	*Rhodospirillales*	*Azospirillaceae*	*Azospirillum*	*2*
	Alphaproteobacteria	*Rhodospirillales*	*Unclassified*	*Unclassified*	*1*
	Betaproteobacteria	***Burkholderiales***	***Burkholderiaceae***	***Burkholderia***	***3***
	Betaproteobacteria	*Burkholderiales*	*Burkholderiaceae*	*Caballeronia*	*2*
	Betaproteobacteria	***Burkholderiales***	***Burkholderiaceae***	***Paraburkholderia***	***6***
	Betaproteobacteria	*Burkholderiales*	*Sphaerotilaceae*	*Methylibium*	*1*
	Betaproteobacteria	*Rhodocyclales*	*Zoogloeaceae*	*Azoarcus*	*1*

The genera in bold simply illustrate the largest numbers

The orientation and length of all collected subunits from the bacterial genomes are summarized in gene arrow plots (Supplementary Figs. 1–3). The plots for the class *Actinomycetota* are divided into four sections: *Pseudonocardiales* (Supplementary Fig. 1a), *Mycobacteriales* (Supplementary Fig. 1b), other orders (Supplementary Fig. 1c) and *Mycobacterium* and *Mycolicibacterium* (Supplementary Fig. 1d). The latter plot was separated from the rest, due to the order of the subunits (as discussed above, *prmACDB*). Interestingly, four *Mycobacterium* and one *Mycolicibacterium* illustrate the order *prmACDB* and the rest (two *Mycobacterium* and one *Mycolicibacterium*) follow the order *prmABCD*. Only two other operons illustrates the order *prmACDB*: one of the two *Rhodococcus* ENV425 operons (NODE_125_length_41067_cov_140.733337) and *Rhodococcus ruber* (Supplementary Fig. 1b). The gene arrow plots for the class *Alphaproteobacteria* are divided into two sections *Hyphomicrobiales* (Supplementary Fig. 2a) and *Rhodobacterales* or *Rhodospirillales* (Supplementary Fig. 2b). The gene arrow plots for the bacterial genomes classifying with the *Betaproteobacteria* are also shown (Supplementary Fig. 3). All of the *Alphaproteobacteria* and *Betaproteobacteria* propane monooxygenase subunits follow the expected order (*prmABCD*).

A phylogenetic tree was constructed with three propane monooxygenase alpha unit (*prmA*) sequences from the genera with larger numbers of strains/species with the propane monooxygenase operon (Supplementary

Fig. 4). Not surprisingly, sequences from the two classes (*Alphaproteobacteria* and *Betaproteobacteria*) group separately from the phylum *Actinomycetota*. There is also a clear separation between both *Pseudomonadota* classes. In general, sequences from the same genus group together. The *Mycobacterium prmA* sequences form a separate branch, appearing distinct from all others.

Di-*Iron* Centre Amino Acid Sequences

The amino acid sequences of both di-iron centres were determined from the *prmA* nucleic acid sequences for both the bacterial genomes and the soil metagenomes (Table 3). All of the *Actinomycetota prmA* sequences from the soil metagenomes and the majority of the *Actinomycetota* bacterial genome *prmA* sequences illustrated the following amino acid sequences: DE V RH, DE S RH. However, for seven *Actinomycetota* bacterial genome *prmA* sequences, a different amino acid was present in the second di-iron centre (DE A RH). These were the same seven microorganisms illustrating the subunit order *prmACBD*, as discussed above. Differences were also noted for the *Alphaproteobacteria* and *Betaproteobacteria*, involving a change in an amino acid in the first di-iron centre (from DE F RH to DE L RH) in a small number of cases.

Propane Monooxygenase Subunits from the Metagenomes

Many propane monooxygenase gene clusters were obtained from each WGS dataset. The orientation and length of each has been summarized in gene arrow plots (Fig. 2). In many cases, all subunits were present, but one or more subunit was not the correct length, therefore these sequences were not collected. As stated above, each set of sequences were subject to a BLAST search to determine which bacterial class illustrated the closest alignment. The majority classified within the *Actinomycetota,* followed by the *Alphaproteobacteria,* then the *Betaproteobacteria*. Phylogenetic trees of *prmA* from the metagenomes with the most closely aligned bacterial *prmA* sequences for each group were then created (Figs. 3–5). Sequences aligning to those from *Actinomycetota* were obtained from all four WGS datasets (Fig. 3), suggesting their common occurrence in these environmental samples. The majority of metagenome *prmA* sequences aligned most closely to the genera *Kribbella* and *Amycolatopsis.* One soil metagenome sequence (from UK soil study) aligned most closely to *prmA* sequences from *Pseudonocardia*.

Interestingly, the majority of *Alphaproteobacteria* aligning *prmA* sequences originated from metagenomes associated with the miscanthus and switchgrass treatments (Fig. 4). There were no *prmA* sequences aligning to the *Alphaproteobacteria* from the UK soils study, there was only one from the corn treatment and only one from the forest soil study. The metagenome *prmA* sequences most closely aligned to those from the genera *Bradyrhizobium*, *Acidiphilium* as well as unclassified *Rhizobiales.* The *Betaproteobacteria* aligning *prmA* sequences were all associated with the forest soil metagenomes (Fig. 5). The majority of these sequences aligned with the genera *Paraburkholderia, Burkholderia* and *Caballeronia* (all in the family *Burkholderiaceae*). Two sequences aligned most closely with the genera *Azoarcus* and *Methylibium*. A summary of the subunit order and length of the bacterial *prmA* sequences most closely aligning with the metagenome *prmA* sequences is shown (Supplementary Fig. 5).

As an additional check on the sequencing analysis approach, the SRA files from the forest study were also analysed with the Galaxy web platform [40, 41]. Overall, both analyses procedures produced similar results, with two exceptions. The Galaxy platform did

Table 3 Amino acid composition of both di-iron centers (DE*RH) in the bacterial genomes and soil metagenomes

	First di-iron center	Second di-iron center
Actinobacteria		
All soil metagenomes	DE **V** RH	DE **S** RH
Rhodococcus ENV425, NODE_125_length_41067_cov_140.733337 *Rhodococcus ruber* *Mycobacterium chubuense* strain NBB4 *Mycobacterium dioxanotrophicus* strain PH-06 *Mycobacterium* sp. ENV421 NODE_68_length_8680_cov_111.355 *Mycobacterium* sp. TY-6 *Mycolicibacterium vanbaalenii* strain JOB5	DE **V** RH	DE **A** RH
All remaining *Actinobacteria* bacteria genomes	DE **V** RH	DE **S** RH
Alphaproteobacteria		
All soil metagenomes	DE **F** RH	DE **S** RH
Majority (50/52) of *Alphaproteobacteria* genomes	DE **F** RH	DE **S** RH
Bradyrhizobium erythrophlei strain GAS138 *Bradyrhizobium erythrophlei* strain GAS242	DE **L** RH	DE **S** RH
Betaproteobacteria		
Majority (7/8) of soil metagenomes	DE **F** RH	DE **S** RH
Minority (1/8) of soil metagenomes	DE **L** RH	DE **S** RH
Majority of bacterial genomes*	DE **F** RH	DE **S** RH

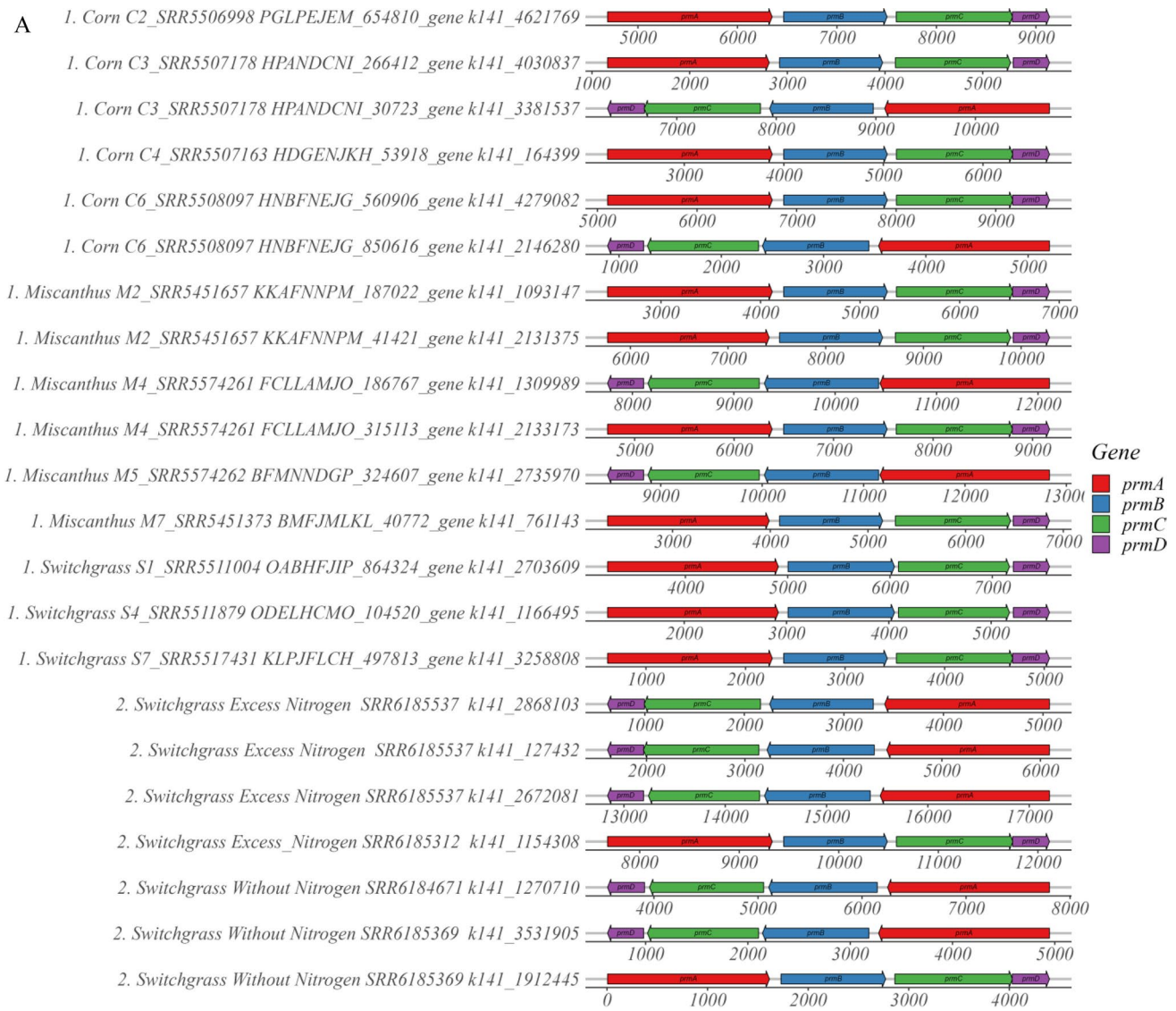

Fig. 2 Propane monooxygenase subunit length and orientation from the four metagenome datasets. Both KBS soil studies are shown together (**a**) and the UK soils and forest soils studies are shown together (**b**)

not retrieve the two gene clusters associated with the outliers in the *Betaproteobacteria prmA* tree (bottom of tree, Supplementary Fig. 6). Also, an extra propane monooxygenase gene cluster was detected by the Galaxy platform for one sample. Impressively, both platforms produced gene clusters with similar alignments, as illustrated for *Betaproteobacteria prmA* sequences (Supplementary Fig. 6). Sequences from both methods align beside each other on this tree, providing confidence on the reproducibility of both approaches.

The percent identities of each metagenome propane monooxygenase subunit to the closest bacterial genome subunit are summarized for those aligning with the *Actinomycetota* (Supplementary Table 5), *Alphaproteobacteria* (Supplementary Table 6) and the *Betaproteobacteria* (Supplementary Table 7). From this set of data, the highest average percent identities were from the *Betaproteobacteria* (91.7% for *prmA*, 88.7% for *prmB*, 88.9% for *prmC*, 91.6% for *prmD*), followed by the *Alphaproteobacteria* (88.7% for *prmA*, 80.8% for *prmB*, 84.0% for *prmC*, 85.6% for *prmD*), then the *Actinomycetota* (88.5% for *prmA*, 77.1% for *prmB*, 82.3% for *prmC*, 83.1% for *prmD*). Three gene clusters illustrated > 98% matches to strains of *Paraburkholderia aspalathi* for all subunits (Supplementary Table 7).

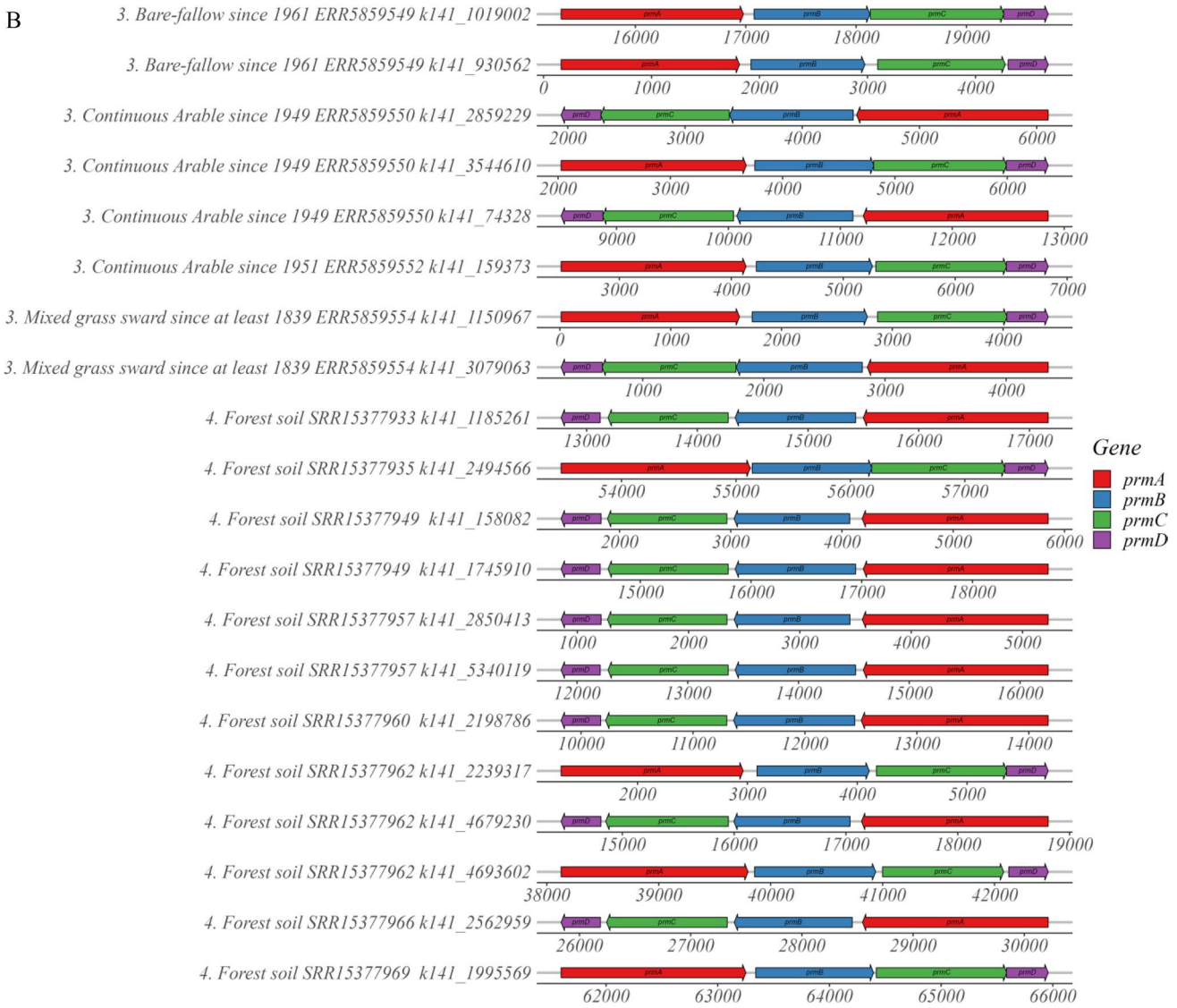

Fig. 2 (continued)

Discussion

The current study adopted a two-stage approach to investigate the phylogeny of microorganisms containing the four subunits of propane monooxygenase. First, nucleic acid sequences were collected from bacterial genomes, available through NCBI. Although much is known about several *Actinomycetota* genera (e.g. *Rhodococcus, Gordonia, Mycobacterium*) containing propane monooxygenase, less is known about other genera in this phylum containing all subunits. Further, limited information is available on microorganisms in other phyla containing these genes. The second stage involved data mining of WGS data for propane monooxygenase sequences from four previous studies. The data generated illustrated common trends in the phylogeny of the microorganisms containing these genes across samples. Significantly, the subunit sequences from environmental samples were not dominated by well-studied propanotrophs, such as *Rhodococcus, Gordonia* or *Mycobacterium*, but by other genera.

As stated above, in the *Actinomycetota*, the majority of metagenome *prmA* sequences aligned most closely to the genera *Kribbella* and *Amycolatopsis*. Both phylotypes were previously associated with 1,4-dioxane biodegradation in laboratory microcosms inoculated with agricultural soil [49]. In that work, multiple genera were statistically significantly enriched following 1,4-dioxane biodegradation compared to the live controls (no 1,4-dioxane), suggesting a growth benefit for 1,4-dioxane biodegradation. Notably, three of the four most enriched in that study were

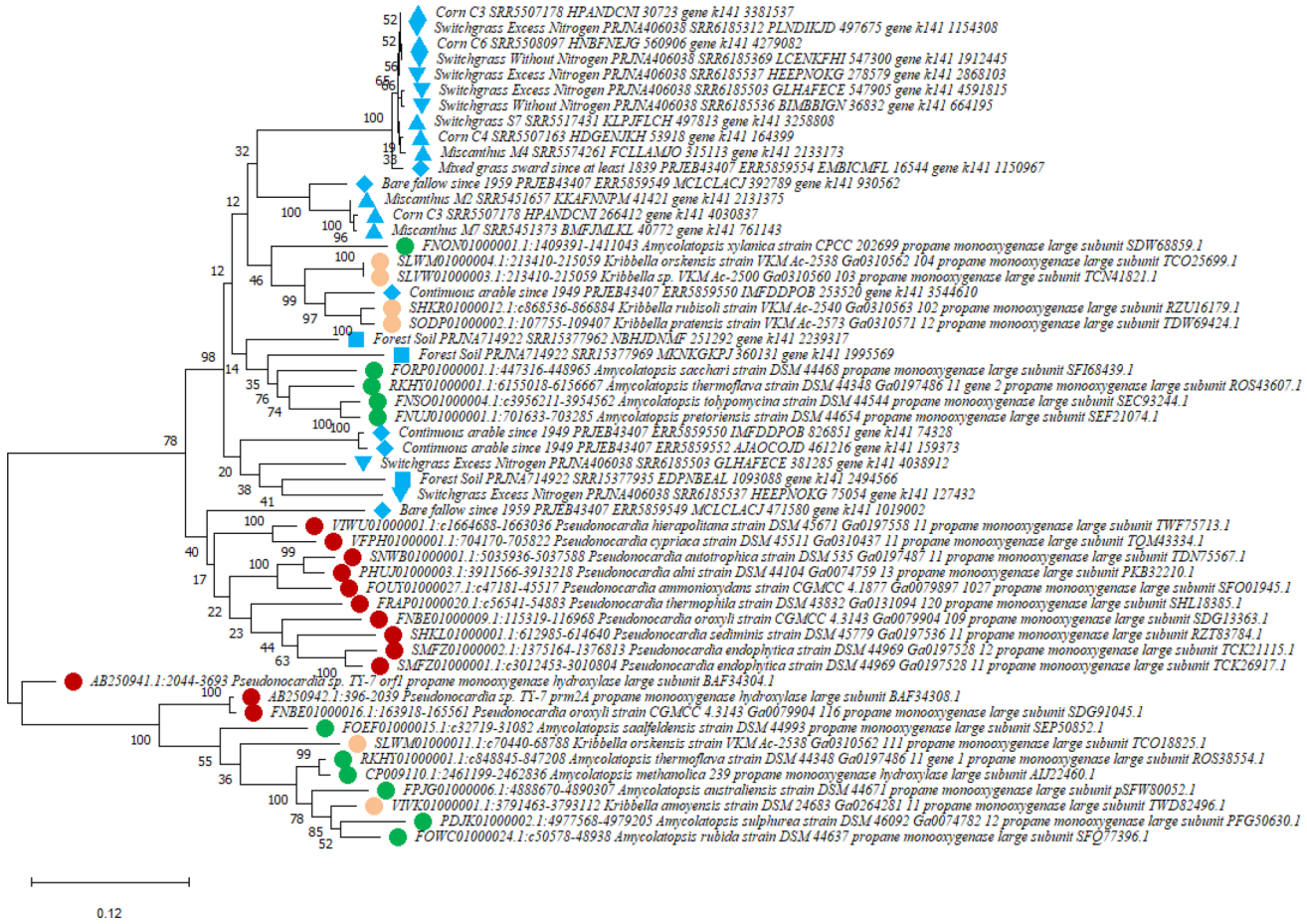

Fig. 3 Alignment of propane monooxygenase alpha unit (*prmA*) from forest soils (blue square), KBS biofuel soils (upward blue triangle), KBS switchgrass soils (downward blue triangle) and UK soils (diamond) with the closest sequences collected from NCBI classifying within the *Actinomycetota*. Only the closest matching *Actinomycetota* sequences are shown. The evolutionary history was inferred by using the Maximum Likelihood method and General Time Reversible model. The tree with the highest log likelihood (− 18,872.59) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.2959)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 54 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding. There were a total of 1676 positions in the final dataset. Evolutionary analyses were conducted in MEGA11

Mycobacterium, Kribbella and *Amycolatopsis*. Combined with the data generated in the current study, it is possible these microorganisms were degrading 1,4-dioxane via propane monooxygenases. Additional work is needed to confirm this hypothesis.

The genera with the most closely aligning *prmA* sequences to those from the metagenomes have all previously been associated the degradation of numerous organics. From the *Actinomycetota*, *Amycolatopsis* and *Kribbella* contained propane monooxygenases aligning to sequences from all four datasets, suggesting their importance in soil associated samples. Members of *Amycolatopsis* have been associated with the degradation of numerous organics, such as the bioplastic polylactic acid [50], the chlorinated organophosphate flame retardant, tris-(2-chloroisopropyl) phosphate [51], polycyclic aromatic hydrocarbons (PAHs) naphthalene and pyrene [52], pharmaceuticals naproxen and carbamazepine [53] and the novel herbicide ZJ0273 [54]. Limited information is available on the biodegradation characteristics of *Kribbella* strains, although a previous study linked this genus to the biodegradation of polychlorinated biphenyls [55] and this genus was implicated in 1,4-dioxane degradation [49] (as stated above).

Three of the four WGS datasets contained propane monooxygenase sequences aligning most closely with *Bradyrhizobium* (*Alphaproteobacteria*). *Bradyrhizobium*

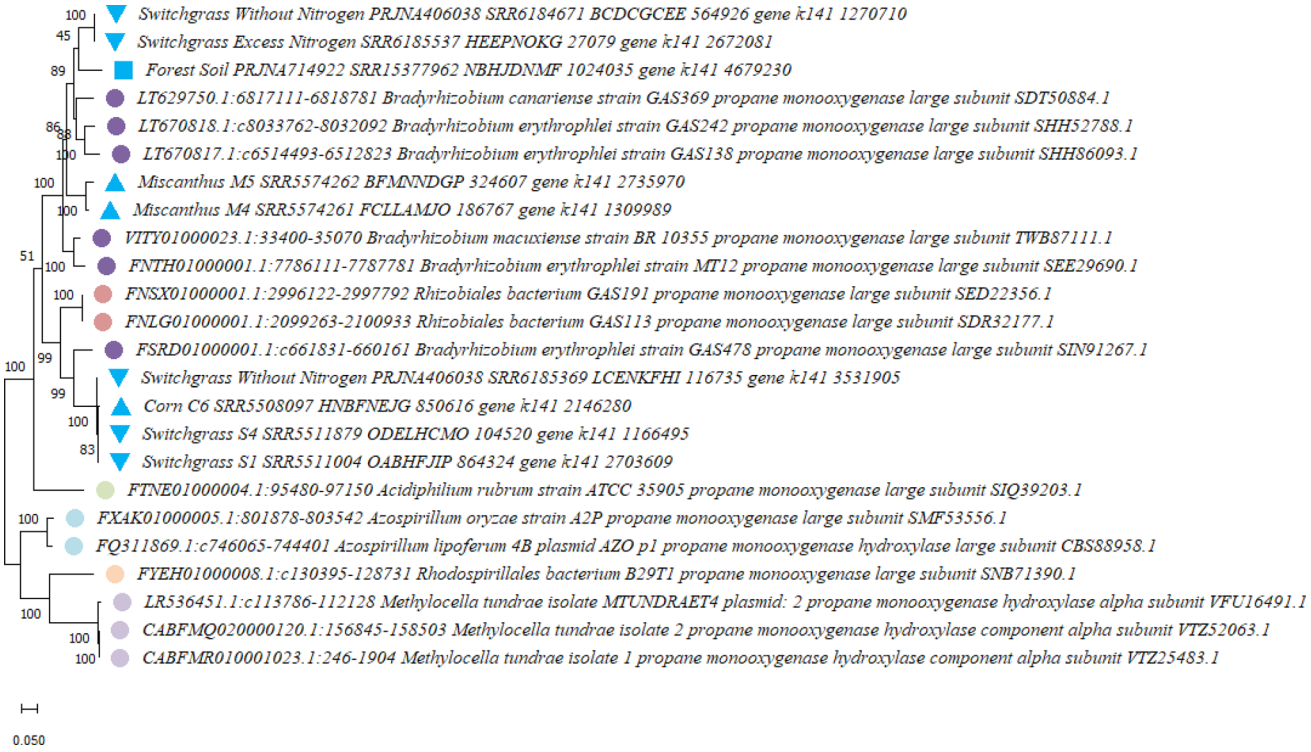

Fig. 4 Alignment of propane monooxygenase alpha unit (*prmA*) from forest soils (blue square), KBS biofuel soils (upward blue triangle), KBS switchgrass soils (downward blue triangle) with the closest sequences collected from NCBI classifying within the *Alphaproteobacteria*. Only the closest matching *Alphaproteobacteria* sequences are shown. The UK soil metagenomes did not contain similar sequences. The evolutionary history was inferred by using the Maximum Likelihood method and General Time Reversible model. The tree with the highest log likelihood (− 12,331.74) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.2097)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 39.87% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 24 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding. There were a total of 1671 positions in the final dataset. Evolutionary analyses were conducted in MEGA11

strains have also been linked to the biodegradation of numerous chemicals including the antibiotic ciprofloxacin [56], the herbicides sulcotrione, mesotrione [57], methoxychlor [58], atrazine [59], simazine [60] and other organics [61] such as 1,2-dibromoethane [62], phenol [63], anthracene [64] and tert-butyl alcohol [65]. From the *Betaproteobacteria*, *Paraburkholderia* and *Burkholderia* were primarily associated with the propane monooxygenase genes from the forest soil WGS dataset. Microorganisms within the *Paraburkholderia* are particularly interesting because of their ability to degrade aromatic chemicals and the ability of some to form root nodules that fix atmospheric nitrogen [66]. Members of this genus have been associated with biodegradation of benzene, toluene, xylene, ethylbenzene (BTEX), naphthalene [67], phenolic acids [66], 3-chlorobenzoate [68], the fungicide mandipropamid [69], PAHs such as phenanthrene and pyrene [70] and dibenzothiophene [71] and the pharmaceutical 17α-ethinylestradiol [72]. *Paraburkholderia xenovorans* LB400, isolated from a PCB-contaminated landfill [73], is a well-studied, effective polychlorinated biphenyl-degrader [74]. Members of the genus *Burkholderia* are often found in soil and water [75] and have also been linked to the biodegradation of a large number of organic chemicals [75], such as phenol [76], triclosan [77], the insecticides acephate [78] and chlorpyrifos [79], PAHs [80–83] and BTEX [84]. Additional work will be needed to determine if the degradative abilities described above are related to the propane monooxygenases.

The current study provides a wealth of information on the phylogeny of microorganisms containing all four propane monooxygenase subunits both from bacterial genomes and environmental metagenomes. The analysis produced several particularly interesting trends. First, the sequences annotating as propane monooxygenase from the environmental samples were not closely aligned

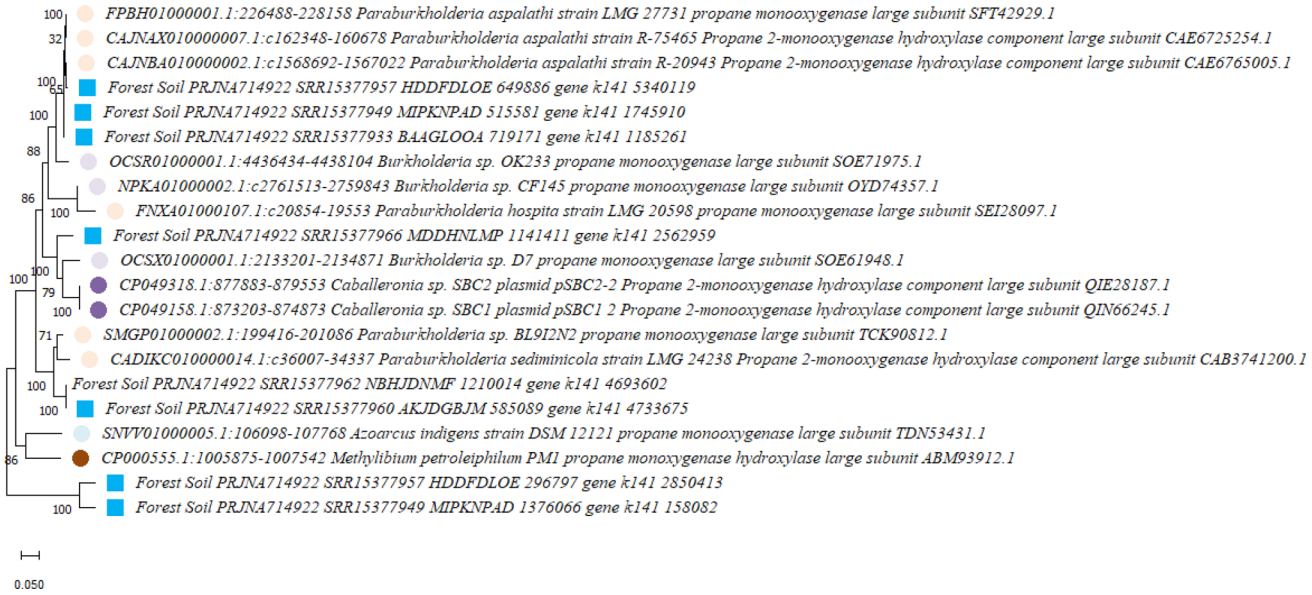

Fig. 5 Alignment of propane monooxygenase alpha unit (prmA) from forest soils (blue square) with the closest sequences collected from NCBI classifying within the *Betaproteobacteria*. The other metagenome studies did not contain similar sequences. The evolutionary history was inferred by using the Maximum Likelihood method and General Time Reversible model. The tree with the highest log likelihood (− 9579.52) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.6280)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 26.16% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 21 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding. There were a total of 1675 positions in the final dataset. Evolutionary analyses were conducted in MEGA11

to microorganisms typically associated with propane biodegradation (e.g. *Mycobacterium* and *Rhodococcus*). Laboratory studies will be required to determine if the propane monooxygenase associated genera from the environmental samples (*Amycolatopsis, Kribbella, Bradyrhizobium, Paraburkholderia* and *Burkholderia)* are indeed capable of propane metabolism. Another notable trend is the occurrence of *Betaproteobacteria* associated propane monooxygenase sequences in the forest soil roots (and not in the other WGS datasets). Root samples may therefore be useful as an inoculum to isolate novel *Betaproteobacteria* propanotrophs. Finally, the *Alphaproteobacteria* phylogenetic trees were dominated by sequences from the switchgrass and miscanthus soils and *Bradyrhizobium*, a genus containing many nitrogen-fixing microorganisms. This trend is consistent with the fact that switchgrass and miscanthus requires lower or no N-fertilizer inputs in comparison to conventional maize production [85]. Again, this may offer another opportunity to isolate novel propanotrophs from the *Alphaproteobacteria*. Overall, the current study highlights the previously untapped potential of diverse propanotrophs from natural samples.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00284-024-03829-z.

Author Contributions AC analysed the data and wrote the manuscript.

Funding This research was supported by a grant from NSF (Award Number 2129228).

Data Availability The sequencing data examined is available at NCBI (see supplementary section for more details).

Declarations

Conflict of interest The author has no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical Approval The study did not involve animal or human subjects.

References

1. Wackett LP, Brusseau GA, Householder SR, Hanson RS (1989) Survey of microbial oxygenases: trichloroethylene degradation by propane-oxidizing bacteria. Appl Environ Microbiol 55:2960–2964. https://doi.org/10.1128/aem.55.11.2960-2964.1989

2. Steffan RJ, McClay K, Vainberg S, Condee CW, Zhang D (1997) Biodegradation of the gasoline oxygenates methyl tert-butyl ether, ethyl tert-butyl ether, and tert-amyl methyl ether by propane-oxidizing bacteria. Appl Environ Microbiol 63:4216–4222. https://doi.org/10.1128/aem.63.11.4216-4222.1997
3. Sharp JO, Sales CM, Alvarez-Cohen L (2010) Functional characterization of propane-enhanced N-nitrosodimethylamine degradation by two actinomycetales. Biotechnol Bioeng 107:924–932. https://doi.org/10.1002/bit.22899
4. Cappelletti M, Pinelli D, Fedi S, Zannoni D, Frascari D (2018) Aerobic co-metabolism of 1,1,2,2-tetrachloroethane by *Rhodococcus aetherivorans* TPA grown on propane: kinetic study and bioreactor configuration analysis. J Chem Technol Biotechnol 93:155–165
5. Wang B, Chu KH (2017) Cometabolic biodegradation of 1,2,3-trichloropropane by propane-oxidizing bacteria. Chemosphere 168:1494–1497. https://doi.org/10.1016/j.chemosphere.2016.12.007
6. Hand S, Wang B, Chu KH (2015) Biodegradation of 1,4-dioxane: effects of enzyme inducers and trichloroethylene. Sci Total Environ 520:154–159. https://doi.org/10.1016/j.scitotenv.2015.03.031
7. Hatzinger PB, Streger SH, Begley JF (2015) Enhancing aerobic biodegradation of 1,2-dibromoethane in groundwater using ethane or propane and inorganic nutrients. J Contam Hydrol 172:61–70. https://doi.org/10.1016/j.jconhyd.2014.11.006
8. Lippincott D et al (2015) Bioaugmentation and propane biosparging for in situ biodegradation of 1,4-dioxane. Ground Water Monit Rem 35:81–92
9. Bell CH et al (2022) First full-scale *in situ* propane biosparging for co-metabolic bioremediation of 1,4-dioxane. Ground Water Monit Rem 42:54–66
10. Horst JF et al (2019) Bioremediation of 1,4- dioxane: Successful demonstration of in situ and ex situ approaches. Groundwater Monit Rem 39:15–24
11. Divine C et al (2024) Advances in remediation solutions: new developments and opportunities in 1,4- dioxane biological treatment. Groundwater Monit Rem. https://doi.org/10.1111/gwmr.12649
12. Hatzinger PB et al (2017) Potential for cometabolic biodegradation of 1,4-dioxane in aquifers with methane or ethane as primary substrates. Biodegradation 28:453–468
13. Hatzinger PB, Lippincott DR (2019) Field demonstration of N-Nitrosodimethylamine (NDMA) treatment in groundwater using propane biosparging. Water Res 164:114923. https://doi.org/10.1016/j.watres.2019.114923
14. Stohr H et al (2023) Cometabolic treatment of 1,4-dioxane in biologically active carbon filtration with tetrahydrofuran and propane at relevant concentrations for potable reuse. ACS ES&T Water 9:2948–2954
15. Kotani T, Yamamoto T, Yurimoto H, Sakai Y, Kato N (2003) Propane monooxygenase and NAD+-dependent secondary alcohol dehydrogenase in propane metabolism by *Gordonia* sp strain TY-5. J Bacteriol 185:7120–7128. https://doi.org/10.1128/JB.185.24.7120-7128.2003
16. Deng DY, Li F, Li MY (2018) A novel propane monooxygenase initiating degradation of 1,4-dioxane by *Mycobacterium dioxanotrophicus* PH-06. Environ Sci Technol Lett 5:86–91
17. Kotani T, Kawashima Y, Yurimoto H, Kato N, Sakai Y (2006) Gene structure and regulation of alkane monooxygenases in propane-utilizing *Mycobacterium* sp. TY-6 and *Pseudonocardia* sp. TY-7. J Biosci Bioeng 102:184–192. https://doi.org/10.1263/jbb.102.184
18. Masuda H, McClay K, Steffan RJ, Zylstra GJ (2012) Characterization of three propane-inducible oxygenases in *Mycobacterium* sp. strain ENV421. Lett Appl Microbiol 55:175–181. https://doi.org/10.1111/j.1472-765X.2012.03290.x
19. Yang SNN, Haritos V, Kertesz MA, Coleman NV (2024) A novel soluble di-iron monooxygenase from the soil bacterium *Solimonas soli*. Environ Microbiol 26:e16567. https://doi.org/10.1111/1462-2920.16567
20. Sharp JO et al (2007) An inducible propane monooxygenase is responsible for N-nitrosodimethylamine degradation by *Rhodococcus* sp strain RHA1. Appl Environ Microbiol 73:6930–6938
21. Chen Y, Ren H, Kong X, Wu H, Lu Z (2023) A multicomponent propane monooxygenase catalyzes the initial degradation of methyl tert-butyl ether in *Mycobacterium vaccae* JOB5. Appl Environ Microbiol 89:e0118723. https://doi.org/10.1128/aem.01187-23
22. Eshghdoostkhatami Z, Cupples AM (2024) Occurrence of *Rhodococcus* sp RR1 prmA and *Rhodococcus jostii* RHA1 *prmA* across microbial communities and their enumeration during 1,4-dioxane biodegradation. J Microbiol Methods 219:106908. https://doi.org/10.1016/j.mimet.2024.106908
23. Tupa PR, Masuda H (2018) Draft genome sequence of a propanotroph, *Rhodococcus* sp. Strain ENV425, capable of degrading Methyl tert-Butyl Ether and N-Nitrosodimethylamine. Genome Announc. https://doi.org/10.1128/genomeA.00051-18
24. Tupa PR, Masuda H (2018) Genomic analysis of propane metabolism in Methyl Tert-Butyl Ether-degrading *Mycobacterium* Sp. Strain ENV421. J Genomics 6:24–29. https://doi.org/10.7150/jgen.24929
25. Tupa PR, Masuda H (2018) Comparative proteomic analysis of propane metabolism in *Mycobacterium* sp. strain ENV421 and *Rhodococcus* sp. strain ENV425. J Mol Microbiol Biotechnol 28:107–115
26. He Y et al (2017) 1,4-Dioxane biodegradation by *Mycobacterium dioxanotrophicus* PH-06 is associated with a group-6 soluble di-iron monooxygenase. Environ Sci Technol Lett 4:494–499
27. Argiroff WA, Zak DR, Pellitier PT, Upchurch RA, Belke JP (2022) Decay by ectomycorrhizal fungi couples soil organic matter to nitrogen availability. Ecol Lett 25:391–404. https://doi.org/10.1111/ele.13923
28. Zak DR, Host GE, Pregitzer KS (1989) Regional variability in nitrogen mineralization, nitrification, and overstory biomass in northern Lower Michigan. Can J For Res 19:1521–1526
29. Zak DR, Pregitzer KS (1990) Spatial and temporal variability of nitrogen cycling in northern Lower Michigan. Forest Sci 36:367–338
30. Li B-B et al (2024) Long-term excess nitrogen fertilizer increases sensitivity of soil microbial community to seasonal change revealed by ecological network and metagenome analyses. Soil Biol Biochem 160:108349
31. Guo J, Cole JR, Zhang Q, Brown CT, Tiedje JM (2016) Microbial community analysis with ribosomal gene fragments from shotgun metagenomes. Appl Environ Microbiol 82:157–166. https://doi.org/10.1128/AEM.02772-15
32. Guo J (2016) Rhizosphere metagenomics of three biofuel crops *Michigan State University Thesis*, https://ezproxy.msu.edu/login?url=https://www.proquest.com/dissertations-theses/rhizosphere-metagenomics-three-biofuel-crops/docview/1849010897/se-1849010892
33. Neal AL, Hughes D, Clark IM, Jansson JK, Hirsch PR (2021) Microbiome aggregated traits and assembly are more sensitive to soil management than diversity. mSystems 6:e01056–e01020
34. Toolkit S. https://github.com/ncbi/sra-tools/wiki/01.-Downloading-SRA-Toolkit.
35. Arkin AP et al (2018) KBase: The United States department of energy systems biology knowledgebase. Nat Biotechnol 36:566–569. https://doi.org/10.1038/nbt.4163

36. Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
37. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
38. Li D, Liu CM, Luo R, Sadakane K, Lam TW (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033
39. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153
40. Community TG (2024) The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. Nucleic Acids Res. https://doi.org/10.1093/nar/gkae410
41. Galaxy C (2022) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. Nucleic Acids Res 50:W345–W351. https://doi.org/10.1093/nar/gkac247
42. Tamura K, Stecher G, Kumar S (2021) MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol 38:3022–3027
43. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680
44. R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria. , 2018).
45. RStudio: Integrated Development for R. RStudio, PBC (Boston, MA 2020).
46. Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. https://ggplot2.tidyverse.org.
47. Wickham H, Bryan J (2023) readxl: Read Excel Files_. R package version 1.4.2. <https://CRAN.R-project.org/package=readxl>
48. Wilkins D (2023) gggenes: Draw Gene Arrow Maps in 'ggplot2'_. R package version 0.5.1, <https://CRAN.R-project.org/package=gggenes>.
49. Ramalingam V, Cupples AM (2020) Enrichment of novel Actinomycetales and the detection of monooxygenases during aerobic 1,4-dioxane biodegradation with uncontaminated and contaminated inocula. Appl Microbiol Biotechnol 104:2255–2269. https://doi.org/10.1007/s00253-020-10376-7
50. Yasin NM, Pancho F, Yasin M, Van Impe JFM, Akkermans S (2024) Novel methods to monitor the biodegradation of polylactic acid (PLA) by *Amycolatopsis orientalis* and *Amycolatopsis thailandensis*. Front Bioeng Biotechnol 12:1355050. https://doi.org/10.3389/fbioe.2024.1355050
51. Feng M et al (2023) Efficient biodegradation of tris-(2-chloroisopropyl) phosphate by a novel strain *Amycolatopsis* sp. FT-1: Process optimization, mechanism studies and toxicity changes. J Hazard Mater 443:130149. https://doi.org/10.1016/j.jhazmat.2022.130149
52. Peralta H et al (2022) Determination of the metabolic pathways for degradation of naphthalene and pyrene in *Amycolatopsis* sp. Poz14. Comp Biochem Physiol C Toxicol Pharmacol 254:109268. https://doi.org/10.1016/j.cbpc.2022.109268
53. Alanis-Sanchez BM et al (2019) Utilization of naproxen by *Amycolatopsis* sp. Poz 14 and detection of the enzymes involved in the degradation metabolic pathway. World J Microbiol Biotechnol 35:186. https://doi.org/10.1007/s11274-019-2764-0
54. Cai Z et al (2012) Degradation of the novel herbicide ZJ0273 by *Amycolatopsis* sp. M3–1 isolated from soil. Appl Microbiol Biotechnol 96:1371–1379. https://doi.org/10.1007/s00253-011-3867-1
55. Leigh MB et al (2007) Biphenyl-utilizing bacteria and their functional genes in a pine root zone contaminated with polychlorinated biphenyls (PCBs). ISME J 1:134–148
56. Nguyen LN, Nghiem LD, Oh S (2018) Aerobic biotransformation of the antibiotic ciprofloxacin by *Bradyrhizobium* sp. isolated from activated sludge. Chemosphere 211:600–607. https://doi.org/10.1016/j.chemosphere.2018.08.004
57. Romdhane S et al (2016) Isolation and characterization of *Bradyrhizobium* sp. SR1 degrading two beta-triketone herbicides. Environ Sci Pollut Res Int 23:4138–4148. https://doi.org/10.1007/s11356-015-4544-1
58. Satsuma K, Masuda M, Sato K (2013) A role of *Bradyrhizobium elkanii* and closely related strains in the degradation of methoxychlor in soil and surface water environments. Biosci Biotechnol Biochem 77:2222–2227. https://doi.org/10.1271/bbb.130439
59. Zhang F et al (2023) Enhanced phytoremediation of atrazine-contaminated soil by vetiver (*Chrysopogon zizanioides* L.) and associated bacteria. Environ Sci Pollut Res Int 30:44415–44429. https://doi.org/10.1007/s11356-023-25395-w
60. Ozawa T, Yoshida R, Wakashiro Y, Hase H (2004) Improvement of simazine degradation by inoculation of corn and soybean plants with rhizobacteria. Soil Sci Plant Nutr 50:1295–1299
61. Vela S, Häggblom MM, Young LY (2002) Biodegradation of aromatic and aliphatic compounds by rhizobial species. Soil Sci 167:802–810
62. Vasileva E, Parvanova-Mancheva T, Beschkov V (2018) 1,2-Dibromoethane biodegradation capacity of *Bradyrhizobium japonicum* strain 273. J Int Sci Publ Ecol Safety 12:82–89
63. Vasileva E, Parvanova-Mancheva T, Beschkov V (2020) The *Bradyrhizobium japonicum* 273 strain' s ability to degrade phenol. Part 1 J Int Sci Publ Mater Methods Technol 14:165–176
64. Dunlevy SR, Singleton DR, Aitken MD (2013) Biostimulation reveals functional redundancy of anthracene-degrading bacteria in polycyclic aromatic hydrocarbon-contaminated soil. Environ Eng Sci 30:697–705. https://doi.org/10.1089/ees.2013.0067
65. Le Digabel Y, Demaneche S, Benoit Y, Fayolle-Guichard F, Vogel TM (2014) Ethyl tert-butyl ether (ETBE)-degrading microbial communities in enrichments from polluted environments. J Hazard Mater 279:502–510. https://doi.org/10.1016/j.jhazmat.2014.07.013
66. Wilhelm RC et al (2020) *Paraburkholderia madseniana* sp. nov., a phenolic acid-degrading bacterium isolated from acidic forest soil. Int J Syst Evol Microbiol 70:2137–2146. https://doi.org/10.1099/ijsem.0.004029
67. Lee Y, Jeon CO (2018) *Paraburkholderia aromaticivorans* sp. nov., an aromatic hydrocarbon-degrading bacterium, isolated from gasoline-contaminated soil. Int J Syst Evol Microbiol 68:1251–1257. https://doi.org/10.1099/ijsem.0.002661
68. Ara I et al (2023) Isolation and genomic analysis of 3-chlorobenzoate-degrading bacteria from soil. Microorganisms. https://doi.org/10.3390/microorganisms11071684
69. Han L et al (2022) Indigenous functional microbial degradation of the chiral fungicide mandipropamid in repeatedly treated soils: preferential changes in the R-enantiomer. J Hazard Mater 435:128961. https://doi.org/10.1016/j.jhazmat.2022.128961
70. Li J et al (2024) Responses of a polycyclic aromatic hydrocarbon-degrading bacterium, *Paraburkholderia fungorum* JT-M8, to Cd (II) under P-limited oligotrophic conditions. J Hazard Mater 465:133123. https://doi.org/10.1016/j.jhazmat.2023.133123
71. Ortega Ramírez CA, Ching T, Yoza B, Li QX (2022) Glycerol-assisted degradation of dibenzothiophene by *Paraburkholderia* sp. C3 is associated with polyhydroxyalkanoate granulation. Chemosphere 291:133054
72. Palma TL, Costa MC (2024) Biodegradation of 17α-ethinylestradiol by strains of *Aeromonas* genus isolated from acid mine drainage. Clean Technol 6:116–139

73. Bedard DL et al (1986) Rapid assay for screening and characterizing microorganisms for the ability to degrade polychlorinated biphenyls. Appl Environ Microbiol 51:761–768. https://doi.org/10.1128/aem.51.4.761-768.1986
74. Chain PS et al (2006) *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. Proc Natl Acad Sci U S A 103:15280–15287. https://doi.org/10.1073/pnas.0606924103
75. Morya R, Salvachua D, Thakur IS (2020) Burkholderia: an untapped but promising bacterial genus for the conversion of aromatic compounds. Trends Biotechnol 38:963–975. https://doi.org/10.1016/j.tibtech.2020.02.008
76. Chen J et al (2017) Characterization of *Burkholderia* sp. XTB-5 for phenol degradation and plant growth promotion and its application in bioremediation of contaminated soil. Land Degradation and Development 28:1091–1099
77. Qiu L et al (2023) Uncovering the metabolic pathway of novel *Burkholderia* sp. for efficient triclosan degradation and implication: Insight from exogenous bioaugmentation and toxicity pressure. Environ Pollut 334:122111. https://doi.org/10.1016/j.envpol.2023.122111
78. Wu X et al (2023) Rapid biodegradation of the organophosphorus insecticide acephate by a novel strain *Burkholderia* sp. A11 and its impact on the structure of the indigenous microbial community. J Agric Food Chem 71:5261–5274. https://doi.org/10.1021/acs.jafc.2c07861
79. Hossain MS et al (2015) Determination of selected pesticides in water samples adjacent to agricultural fields and removal of organophosphorus insecticide chlorpyrifos using soil bacterial isolates. Appl Water Sci 5:171–179
80. Zhao Z, Chen W, Cheng Y, Li J, Chen Z (2023) *Burkholderia cepacia* immobilized onto rGO as a biomaterial for the removal of naphthalene from wastewater. Environ Res 235:116663. https://doi.org/10.1016/j.envres.2023.116663
81. Cauduro GP et al (2021) New benzo(a)pyrene-degrading strains of the *Burkholderia cepacia* complex prospected from activated sludge in a petrochemical wastewater treatment plant. Environ Monit Assess 193:163. https://doi.org/10.1007/s10661-021-08952-z
82. Cauduro GP et al (2022) *Burkholderia vietnamiensis* G4 as a biological agent in bioremediation processes of polycyclic aromatic hydrocarbons in sludge farms. Environ Monit Assess 195:116. https://doi.org/10.1007/s10661-022-10733-1
83. Liu XX et al (2019) Biodegradation of phenanthrene and heavy metal removal by acid-tolerant *Burkholderia fungorum* FM-2. Front Microbiol 10:408. https://doi.org/10.3389/fmicb.2019.00408
84. Bacosa HP, Mabuhay-Omar JA, Balisco RAT, Omar DM Jr, Inoue C (2021) Biodegradation of binary mixtures of octane with benzene, toluene, ethylbenzene or xylene (BTEX): insights on the potential of *Burkholderia, Pseudomonas* and *Cupriavidus* isolates. World J Microbiol Biotechnol 37:122. https://doi.org/10.1007/s11274-021-03093-4
85. Mao Y, Yannarell AC, Davis SC, Mackie RI (2013) Impact of different bioenergy crops on N-cycling bacterial and archaeal communities in soil. Environ Microbiol 15:928–942. https://doi.org/10.1111/j.1462-2920.2012.02844.x