

BIODEGRADATION AND MOLECULAR ANALYSIS OF 1,4-DIOXANE AND OTHER  
ORGANIC CONTAMINANTS IN SOILS USING METAGENOMIC TOOLS

By

Vidhya Ramalingam

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Environmental Engineering – Doctor of Philosophy

2021

## **ABSTRACT**

### **BIODEGRADATION AND MOLECULAR ANALYSIS OF 1,4-DIOXANE AND OTHER ORGANIC CONTAMINANTS IN SOILS USING METAGENOMIC TOOLS**

By

Vidhya Ramalingam

Historically, 1,4-dioxane, a potential human carcinogen, was used as a stabilizer in 1,1,1-trichloroethane (1,1,1-TCA) formulations and is now frequently detected at chlorinated solvent contaminated sites. Bioremediation has emerged as an effective strategy to treat 1,4-dioxane. However, the distribution of 1,4-dioxane degrading species across various environmental samples is generally unknown. Additionally, 1,4-dioxane contamination typically occurs in groundwater under highly reducing conditions. There is a significant knowledge gap and a lack of information on the susceptibility of 1,4-dioxane to biodegradation under such reducing conditions. The success of organic contaminant bioremediation is often linked to the abundance of functional genes present in the soil that are associated with the degradation process. Although some information exists on the presence of these genes in contaminated soils, there is limited knowledge on the presence and diversity of these genes in uncontaminated soils. To address all these knowledge gaps, a series of studies were conducted.

The first study aims at identifying which 1,4-dioxane degrading functional genes are present in soil communities and which genera may be using 1,4-dioxane and/or metabolites to support growth across different microbial communities. For this, laboratory sample microcosms and abiotic control microcosms (containing media) were inoculated with four uncontaminated soils and sediments from two contaminated sites. The sample microcosms were amended with 1,4-dioxane thrice and live control microcosms were treated in the same manner, except 1,4-

dioxane was not added. Biodegradation was observed and whole genome shotgun sequencing was carried out. Although some degraders previously linked to 1,4-dioxane degradation were detected, *Nocardioides*, *Gordonia* and *Kribbella* were found to be potentially novel degraders. The functional genes associated with 1,4-dioxane demonstrated three genes were present at higher relative abundance values, including *Rhodococcus sp. RR1 prmA*, *Rhodococcus jostii RHA1 prmA* and *Burkholderia cepacia G4 tomA3*.

The second study is focused on anaerobic biodegradation of 1,4-dioxane. The potential for 1,4-dioxane biodegradation was examined using multiple inocula and electron acceptor amendments. Compound specific isotope analysis (CSIA) was used to further investigate biodegradation in a subset of the microcosms. DNA was extracted from microcosms exhibiting 1,4-dioxane biodegradation for microbial community analysis using 16S rRNA gene amplicon high throughput sequencing. 1,4-dioxane biodegradation was most commonly observed in the nitrate amended and no electron acceptor treatments. However, it is important to note that the degradation was slow (approximately one year).

The third study examines a set of genes associated with organic contaminant degradation in four uncontaminated (agricultural) soils. The abundance and diversity of *benA*, *bph*, *dbfA*, *dxnA*, *etnC*, *etnE*, *ppaH*, *npaH*, *vcrA*, *xenA*, *xenB* and *xplA* were investigated using protein sequences from the Functional Gene Pipeline and Repository (FunGene). The phylogenetic trees created indicate many genera may potentially be associated with each gene including *Pseudomonas*, *Rhodococcus*, *Mycobacterium* and *Nocardioides*. From these, some strains are well studied and are known to be involved in the biodegradation of organic contaminants and others are potentially new genera that may be associated with the biodegradation of the targeted group of contaminants.

To  
Deepak and Advay  
Thank you for Everything

## ACKNOWLEDGEMENTS

I would like to thank Dr. Alison Cupples, for guiding and supporting me in my research work. Her patience and trust in me enabled me to complete this work. I would also like to thank members of my committee, Dr. Syed Hashsham, Dr. Irene Xagorarakis and Dr. Steve Safferman for their support and valuable inputs. A special thanks to Lori Lerner, for her prompt help with placing research supply orders. Also, thanks to the other support and administrative staff in the Department of Civil and Environmental Engineering for their support.

A special thanks to my previous supervisors Dr. Janahvive Morgan, Mr. Jason Smith, Mr. Anthony Ingle, and Dr. Tom Voice for providing me the opportunities to teach challenging courses. I am indebted to the various funding sources that enabled this work. I would like to thank the funding agency SERDP for supporting a significant portion of this research. A special thanks to KBS LTER for the summer fellowship and the College of Engineering along with the Department of Civil and Environmental Engineering (CEE) for multiple summer fellowships. Thanks to the Department of CEE for the Jerry McCowan endowed fellowship and the College of Engineering for the Dissertation Completion Fellowship. A special thanks to Dr. Katy Luchini Colbry for tuition scholarships provided through EGR 993.

This work would not have been possible without the support of the Research Technology and Support Facility (RTSF) at Michigan State University. I would like to thank Dr. Dan Jones and his team for providing analytical support in the measurement of chemical compounds. I extend my thanks to the genomics facility for their support in DNA sequencing. Also, a special thanks to Dr. Anthony Danko from NAVFAC for providing sediment samples collected from

various contaminated sites. I am grateful to the members of Dr. Hashsham's and Dr. Xagorarakis's laboratory for sharing instruments.

Finally, I would like to thank my family for standing by me on this journey. First, my husband Dr. Deepak Gunasekaran, his love and support single-handedly pushed me to the finish line. Thank you for showing me the strength that I never knew I possessed. I dedicate this dissertation to him and our son, Advay. I want to thank my parents for their unconditional love and support. They supported me through the toughest of times and I will forever be grateful to them. I would like to thank my brother for his constant support and his ability to make me smile even when I felt challenged. I would also like to thank my parents-in-law and sister-in-law for their constant encouragement and support in all my endeavors. They have supported me through challenging times and I am truly blessed to have them all as my family. A special thanks to my dear friend, Dr. Huiyun Wu for supporting me in more ways than one during my Ph.D. journey. I am thankful for the day I picked my office space next to you. Finally, I am thankful to Volt, my dog, for keeping me active.

## TABLE OF CONTENTS

<b>LIST OF TABLES</b> .....	ix
<b>LIST OF FIGURES</b> .....	xi
<b>KEY TO ABBREVIATIONS</b> .....	xvii
<b>Chapter 1 Introduction</b> .....	1
1.1 Biodegradation of 1,4-Dioxane .....	2
1.2 Other Organic Contaminants and Functional Genes .....	3
1.3 Next Generation Sequencing Techniques .....	4
1.4 Dissertation Outline and Objectives .....	5
<b>REFERENCES</b> .....	8
<b>Chapter 2 Enrichment of Novel Actinomycetales and the Detection of Monooxygenases during Aerobic 1,4-Dioxane Biodegradation with Uncontaminated and Contaminated Inocula</b> .....	12
2.1 Abstract .....	12
2.2 Introduction .....	13
2.3 Methods .....	17
2.3.1 Chemicals and Inocula .....	17
2.3.2 Experimental Setup, DNA Extraction, 1,4-Dioxane Analysis .....	17
2.3.3 Library Preparation, Sequencing, MG-RAST and DIAMOND analysis .....	19
2.4 Results .....	23
2.4.1 Biodegradation of 1,4-Dioxane .....	23
2.4.2 Taxonomic Analysis of Metagenomes .....	25
2.4.3 Genera Associated with 1,4-Dioxane Degradation .....	26
2.4.4 Relative Abundance of Genera Associated with 1,4-dioxane Biodegradation .....	33
2.4.5 Genes Associated with 1,4-Dioxane Degradation .....	35
2.5 Discussion .....	40
<b>APPENDIX</b> .....	45
<b>REFERENCES</b> .....	49
<b>Chapter 3 Anaerobic 1,4-Dioxane Biodegradation and Microbial Community Analysis in Microcosms Inoculated with Soils or Sediments and Different Electron Acceptors</b> .....	56
3.1 Abstract .....	56
3.2 Introduction .....	57
3.3 Methods .....	61
3.3.1 Chemicals and Inocula .....	61
3.3.2 Experimental Setup .....	62
3.3.3 1,4-Dioxane Analysis .....	64
3.3.4 Compound Specific Isotope Analysis (CSIA) .....	65

3.3.5 DNA Extraction and High Throughput Amplicon Sequencing.....	66
3.3.6 Analysis of Sequencing Data.....	67
3.3.7 Statistical Analysis .....	67
3.4. Results .....	68
3.4.1 1,4-Dioxane Removal Trends.....	68
3.4.2 Compound Specific Isotope Analysis (CSIA).....	79
3.4.3 Putative 1,4-Dioxane Degraders.....	81
3.5 Discussion .....	87
REFERENCES.....	93
Chapter 4 Analysis of Functional Genes Associated with Organic Contaminant Biodegradation in Agricultural Soils using Shotgun Sequencing Data.....	100
4.1 Abstract .....	100
4.2 Introduction .....	101
4.3 Methods.....	106
4.3.1 Experimental Setup and DNA Extraction .....	106
4.3.2 Shotgun Sequencing and Trimmomatic .....	106
4.3.3 FunGene Pipeline and DIAMOND .....	107
4.3.4 Relative Abundance of Biodegradation Genes.....	107
4.3.5 Statistical Analysis: Linear Regression and EstimateS Diversity Indices.....	108
4.3.6 Phylogenetic Trees Based on Contigs .....	109
4.4 Results .....	110
4.4.1 Relative abundance of Biodegradation Genes in Soils.....	110
4.4.2 Statistical Tests - ANOVA and Kruskal Wallis .....	112
4.4.3 Linear Regression of Data .....	114
4.4.4 Diversity Indices determined by EstimateS.....	118
4.4.5 Phylogenetic Trees .....	121
4.5 Discussion .....	131
APPENDIX.....	135
REFERENCES.....	146
Chapter 5 Conclusions .....	156
5.1. Future Work.....	157

## LIST OF TABLES

Table 2.1. Summary of sequencing information processed by MG-RAST .....	21
Table 2.2. Classification of genera statistically significantly enriched ( $p < 0.05$ ) in the samples compared to the controls (no 1,4-dioxane) following the degradation of 1,4-dioxane in all soils collectively and when the soils were analyzed individually. The last column also illustrates the difference in means between the controls and the samples for each genera. Genera in bold were identified in the BLASTP search as containing genes similar to <i>Rhodococcus jostii</i> RHA1 prmA and <i>Rhodococcus</i> sp. prmA (as discussed in the results section for the functional gene analysis). .....	28
Table 3.1. Experimental design (inocula and amendments) of the five microcosm studies.....	62
Table 3.2. Percent removal in samples compared to the controls at the last time point (when the difference was statistically significant, determined with student's t-test for independent variables with equal variance, $p < 0.05$ ).....	72
Table 4.1. Summary of data using Shapiro test to check normality of each gene. The values in bold ( $p$ -values $> 0.05$ ) indicating normal data .....	136
Table 4.2. Summary of linear regression results for normal data with equal variance.....	136
Table 4.3. Summary of the $p$ values from Spearman's rank correlation tests with gene relative abundance data. Values in bold indicate a significant difference ( $p \leq 0.05$ ). .....	137
Table 4.4. Summary of Spearman's correlation coefficient ( $\rho$ ) for Spearman's rank correlation test with gene relative abundance data. $\rho$ values in bold indicate a statistically significant correlation ( $p \leq 0.05$ ), as shown above. ....	137
Table 4.5. P-values for statistical tests with the richness index chao 1 of gene copies associated with xenobiotic degradation. "N/A" indicates the test was not appropriate and $p$ -values in bold indicate a significant difference ( $p \leq 0.05$ ). .....	138
Table 4.6. P-values for Tukey's HSD test with the richness index chao 1 of counts of xenobiotic genes. "N/A" indicates the test was not appropriate and $p$ -values in bold indicate a significant difference ( $p \leq 0.05$ ). .....	139
Table 4.7. P-values for Dunn's test with the richness index chao 1 of counts of genes associated with xenobiotic degradation. "N/A" indicates the test was not appropriate and $p$ -values in bold indicate a significant difference ( $p \leq 0.05$ ). .....	139

Table 4.8. P-values for statistical tests with the richness index chao 2 of gene copies associated with xenobiotic degradation. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ). ..... 140

Table 4.9 P-values for Tukey’s HSD test with the richness index chao 2 of counts of xenobiotic genes. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ). ..... 141

Table 4.10 P-values for Dunn’s test with the richness index chao 2 of counts of genes associated with xenobiotic degradation. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ). ..... 141

Table 4.11. P-values for statistical tests with the Shannon diversity of genes associated with xenobiotic degradation. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ). ..... 142

Table 4.12. P-values for Tukey’s HSD test with the Shannon diversity of genes associated with xenobiotic degradation. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ). ..... 143

Table 4.13. P-values for Dunn’s test with the Shannon diversity of genes associated with xenobiotic degradation. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ). ..... 143

Table 4.14. P-values for statistical tests with the richness index chao 2 of gene copies associated with xenobiotic degradation. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ). ..... 144

Table 4.15. P-values for Tukey’s HSD test with the richness index chao 2 of counts of xenobiotic genes. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ). ..... 144

Table 4.16. P-values for Dunn’s test with the richness index chao 2 of counts of genes associated with xenobiotic degradation. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ). ..... 145

## LIST OF FIGURES

Figure 2.1. Average 1,4-dioxane concentrations (mg/L) in triplicate samples and abiotic controls with different inocula, including four agricultural soils and sediments from two contaminated sites (bars represent standard deviations). 1,4-dioxane was reamended to the samples microcosms twice (arrows). ..... 24

Figure 2.2. Extended error bar plot illustrating genera statistically significantly different in relative abundance (Welch's two sided t-test,  $p < 0.05$ ) between the samples ( $n=9$ ) and the live controls (no 1, 4-dioxane,  $n=8$ ) following 1,4-dioxane degradation (A). The symbols to the left of the dashed line (yellow) indicate a higher relative abundance in the samples compared to the controls and the symbols to the right (blue) indicate the reverse. A comparison of the relative abundance values (%) for the genera enriched in the samples is also shown in a box plot format (B). The insert illustrates the relative abundance of these enriched genera in the contaminated site sample (C7A) with a different y-axis scale. .... 27

Figure 2.3. Extended error bar plots illustrating genera statistically significantly different in relative abundance (Welch's two sided t-test,  $p < 0.05$ ) between the samples and the live controls following 1,4-dioxane degradation in soil 1 (A) and 2 (B). The symbols to the left of the dashed line (in yellow) indicate a higher relative abundance in the samples compared to the controls and the symbols to the right (in blue) indicate the reverse. .... 30

Figure 2.4. Extended error bar plots illustrating genera statistically significantly different in relative abundance (Welch's two sided t-test,  $p < 0.05$ ) between the samples and the live controls following 1,4-dioxane degradation in soil F (A) and soil G. The symbols to the left of the dashed line (in yellow) indicate a higher relative abundance in the samples compared to the controls and the symbols to the right (in blue) indicate the reverse. .... 31

Figure 2.5. Summary of the relative abundance of statistically significantly enriched genera in the samples compared to the controls (no 1,4-dioxane) for soils 1, 2, F and G. The insert illustrates the relative abundance of these genera in the contaminated site sample (C7A) with a different scale on the y-axis. .... 32

Figure 2.6. Relative abundance (%) of genera associated with metabolic and co-metabolic degradation of 1,4-dioxane in live controls ( $n=8$ ) and samples ( $n=9$ ) in four soils and one contaminated site sample (C7A). The value "a" indicates a significant difference ( $p < 0.05$ ) in a two tailed student's t-test between the samples and controls. The insert illustrates the same data with a different y-axis. .... 34

Figure 2.7. Relative abundance (%) of reads aligning ( $\geq 60\%$  identity for  $\geq 49$  amino acids) to genes previously associated with the metabolic and co-metabolic degradation of 1,4- dioxane in Soil F and C7A (A), Soil G (B), Soil 1 (C) and Soil 2 (D). .... 36

Figure 2.8. Phylogenetic tree of *Rhodococcus jostii* RHA1 *prmA* and *Rhodococcus* sp. *prmA* and BLASTP results (>94.8% similar to the two query sequences). Only genera that were enriched following 1,4-dioxane degradation (compared to the controls) are shown (Table 2). The evolutionary history was inferred by using the Maximum Likelihood method based on the Jones-Taylor-Thornton (JTT) matrix-based model. The tree with the highest log likelihood (-2731.06) is shown. 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 a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 48 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 439 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 ..... 39

Figure 2.9. Phylogram (created with MEGAN6, version 6.11.7) illustrating the relative abundance and classification (Class Level) of all bacteria across all metagenomes (samples and controls). ..... 46

Figure 2.10. Phylogram (created with MEGAN6, version 6.11.7) illustrating the most abundant genera (ranked by average relative abundance, then selected if average relative abundance >0.5%) across all metagenomes (samples and controls). ..... 47

Figure 2.11. The twenty-five most common genera (by relative abundance, %), ranked by the averages of the samples and controls, in soil 1 (A), soil 2 (B), soil F (C), soil G (D) and the contaminated site sediment 7A (E). ..... 48

Figure 3.1. 1,4-dioxane concentrations in sample and abiotic control microcosms of three soils amended with EDTA/iron and humic acid. The values and bars represent averages and standard deviations from triplicates. The circles represent a significant difference (two-tailed t-test,  $p < 0.05$ ) between the samples and abiotic controls (when abiotic controls > samples). ..... 69

Figure 3.2. 1,4-dioxane concentrations in sample and abiotic control microcosms of three soils with different electron acceptor amendments (A-D). The values and bars represent averages and standard deviations from triplicates (except soil G,  $n = 2$  for three cases). The circles represent a significant different (two-tailed t-test,  $p < 0.05$ ) between the samples and abiotic controls (when controls > samples). The arrows indicate the samples subject to DNA extraction. CSIA was performed on the samples and abiotic controls from the soil F, nitrate amended treatments. .... 71

Figure 3.3. 1,4-dioxane concentrations in Sediment J and H sample and abioitic control microcosms over time with different amendments. The points illustrate average values from triplicates (except nitrate amended controls, sediment H,  $n=1$ ) and the bars illustrate standard deviations. The circles represent a significant different (two-tailed t-test,  $p < 0.05$ ) between the samples and abiotic controls. The arrow indicates the samples subject to DNA extraction. .... 73

Figure 3.4. 1,4-dioxane concentrations in Maine and California sample and abiotic control microcosms over time with different amendments. The points illustrate average values from triplicates (except sulfate amended Maine sample, n=2 and sulfate amended CA control, n=2) and the bars illustrate standard deviations. The circles represent a significant different (two-tailed t-test, p<0.05) between the samples and abiotic controls (controls > samples). The arrows indicate the samples subject to DNA extraction. CSIA was performed on the samples and abiotic controls from both of the no electron acceptor treatment (D). ..... 75

Figure 3.5. 1,4-dioxane concentrations in sample and abiotic control microcosms over time with different inocula. The points illustrate average values from triplicates (except Soil E and KBS Soil 2 samples, which are replicates) and the bars illustrate standard deviations. The circles represent a significant different (two-tailed t-test, p<0.05) between the samples and abiotic controls. The arrows indicate the samples subject to DNA extraction. .... 78

Figure 3.6. Comparison of  $\delta^{13}C$  (A) and  $\delta^2H$  (H) values in samples and abiotic controls of a sub-set of the experimental microcosms. The error bars represent standard deviations from triplicate samples and control microcosms and the p values were derived from the student's t-test (two tailed, independent variables, equal variance)..... 80

Figure 3.7. Phylotypes with a statistically significant difference (Welch's t-test, two sided, p<0.05) between the 1,4-dioxane degrading samples (n=3) compared to the live controls (n=3). The comparisons are shown for the Contaminated Site (CA) microcosms (A), the Sediment H microcosms (B) and KBS Soil 3 microcosms (C). No differences were noted for Soil F. The data points to the left of the dashed line indicate phylotypes more abundant in the 1,4-dioxane degrading samples compared to the controls and those to the right indicate the reverse. For each, the x-axis have different scales. .... 82

Figure 3.8. Relative abundance (%) of genera previously associated with 1,4-dioxane biodegradation in replicates of 1,4-dioxane degrading microcosms or live controls (C, no 1,4-dioxane) inoculated with different soils. The inserts have a different axis to better illustrate the relative abundance of Rhodanobacter. .... 85

Figure 3.9. Relative abundance (%) of genera previously associated with 1,4-dioxane biodegradation in replicates of 1,4-dioxane degrading microcosms or live controls (C, no 1,4-dioxane) inoculated with different sediments (A, B, C). The inserts have a different axis to better illustrate the relative abundance of Pseudomonas. .... 86

Figure 4.1. Box and whisker plot of relative abundance of genes (A) and Principle Component Analysis of the genes across the four agricultural soils (with or without added 1,4-dioxane) (B). ..... 111

Figure 4.2. Average relative abundance values (% , as determined by DIAMOND) for each soil (n=5 for soil F, n=4 for soil G and soil 1 and n=3 for soil 2) with standard deviations illustrated with the error bars. Values that are statistically significantly different (ANOVA or Kruskal-

Wallis test,  $p < 0.05$ ) are shown with different letters. Letters are missing for *xenA* because the statistical assumptions were not met for either test (unequal variance). Note, all y-axis have different scales. .... 113

Figure 4.3. Scatterplots comparing relative abundance values of 8 genes (non-normal data) across all samples. Correlations that were statistically significant (Spearman’s rank test,  $p < 0.05$ ) are boxed in red..... 115

Figure 4.4. Regression line scatter plots with 95% CI for genes with normal relative abundance data and equal variance. .... 117

Figure 4.5. Average index diversity values for each soil as determined by EstimateS ( $n=4$ ) with standard deviations illustrated with the bars. Values that are statistically significantly different (ANOVA or Kruskal-Wallis test,  $p < 0.05$ ) are shown with different letters. Note, the scale on the y-axis differs between graphs. .... 119

Figure 4.6. Average index diversity values for each soil as determined by EstimateS ( $n=4$ ) with standard deviations illustrated with the bars. Values that are statistically significantly different (ANOVA or Kruskal-Wallis test,  $p < 0.05$ ) are shown with different letters. Note, the scale on the y-axis differs between graphs. .... 120

Figure 4.7. Phylogenetic tree for gene *benA* created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The tree with the highest log likelihood (-9563.06) is shown. Initial tree(s) for the heuristic Phylogenetic tree for gene *benA* created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The tree with the highest log likelihood (-9563.06) is shown. 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 a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 41 amino acid sequences. There were a total of 492 positions in the final dataset. Evolutionary analyses were conducted in MEGA X. .... 123

Figure 4.8. Phylogenetic tree for genes (A) *bph* and (B) *dbfA* created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The trees with the highest log likelihood (A) (-15779.72) and (B) (-12459.37) are shown. 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 a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 37 (A) and 27 (B) amino acid sequences. There were a total of (A) 550 and (B) 586 positions in the final dataset. Evolutionary analyses were conducted in MEGA X..... 124

Figure 4.9. Phylogenetic tree for gene *dxnA* created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The tree with the highest log likelihood (-16429.21) is shown. 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 a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 27 amino acid sequences. There were a total of 595 positions in the final dataset. Evolutionary analyses were conducted in MEGA X. .... 125

Figure 4.10. Phylogenetic tree for gene *etnC* created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The tree with the highest log likelihood (-7690.39) is shown. 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 a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 46 amino acid sequences. There was a total of 599 positions in the final dataset. Evolutionary analyses were conducted in MEGA X. .... 126

Figure 4.11. Phylogenetic tree for genes (A) *etnE* and (B) *npaH* created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The trees with the highest log likelihood (A) (-8311.56) and (B) (-9047.49) are shown. 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 a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved (A) 33 and (B) 23 amino acid sequences. There were a total of (A) 726 and (B) 570 positions in the final dataset. Evolutionary analyses were conducted in MEGA X. .... 127

Figure 4.12. Phylogenetic tree for genes (A) *ppaH* and (B) *xenA* created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The trees with the highest log likelihood (A) (-10521.2) and (B) (-10642.31) are shown. 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 a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved (A) 28 and (B) 30 amino acid sequences. There were a total of (A) 561 and (B) 632 positions in the final dataset. Evolutionary analyses were conducted in MEGA X. .... 128

Figure 4.13. Phylogenetic tree for gene *xenB* created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The tree with the highest log likelihood (-11304.53) is shown. 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 a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of

substitutions per site. This analysis involved 29 amino acid sequences. There was a total of 458 positions in the final dataset. Evolutionary analyses were conducted in MEGA X. .... 129

Figure 4.14. Phylogenetic tree for gene xplA created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The tree with the highest log likelihood (-10292.94) is shown. 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 a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 30 amino acid sequences. There was a total of 809 positions in the final dataset. Evolutionary analyses were conducted in MEGA X. .... 130

## KEY TO ABBREVIATIONS

CSIA – Compound Specific Isotope Analysis

EDTA – Ethylenediamine tetraacetic Acid

ETBE – Ethyl Tertiary Butyl Ether

IARC – International Agency for Research on Cancer

KBS – Kellogg Biological Station

MG-RAST – Metagenomics Rapid Annotation using Subsystem Technology

MTBE – Methyl Tertiary Butyl Ether

PAH – Polycyclic Aromatic Hydrocarbon

PCB – Polychlorinated Biphenyls

RTSF – Research Technology and Support Facility

RDX – 1,3,5-Trinitro-1,3,5-triazinane

SDIMO – Soluble Di-Iron Monooxygenase

STAMP – Statistical Analyses of Metagenomic Profiles

TCA – 1,1,1-Trichloroethane

TCE – Trichloroethylene

THF - Tetrahydrofuran

TNT – 2,4,6-Trinitrotoluene

USEPA – United States Environmental Protection Agency

VC – Vinyl Chloride

## Chapter 1 Introduction

Chlorinated solvents have been used in industrial applications for many decades. Often, solvent stabilizers are used to prevent the degradation or breakdown of these chlorinated solvents (Mohr et al. 2010a). The cyclic ether 1,4-dioxane (1,4-diethylene dioxide) is a solvent stabilizer and is also formed as an undesired by-product in some industrial processes (Adams et al. 1994; Mahendra and Alvarez-Cohen 2005; Parales et al. 1994; Son et al. 2009). It is commonly detected at sites contaminated with chlorinated compounds owing to its wide usage as a solvent stabilizer for 1,1,1-trichloroethane (TCA) (Adamson et al. 2015). The degreasing formulations of TCA and TCA-based solvents may contain up to 8% 1,4-dioxane (Grimmett and Munch 2009). 1,4-Dioxane is also present in some household and industrial products including paints, varnishes, resins, lacquers, oils, dyes, waxes, insecticides, aircraft deicing fluids, antifreeze and fumigants (Anderson et al. 2012; Mohr et al. 2010a; USEPA 2017).

1,4-Dioxane has been classified as a possible human carcinogen (B2) by the International Agency for Research on Cancer and a probable human carcinogen (Group 2B) by the USEPA based on inadequate evidence of human studies but sufficient experimental evidence on animal studies (IARC 1999; USEPA 2017; Zenker et al. 2003). The classification was due to increased incidences of gall bladder carcinomas in guinea pigs and nasal carcinomas in rats (ATSDR 2012). The first research study on hepatocarcinogenicity of 1,4-dioxane in rats was published in 1935 (Argus et al. 1965). Thus, there is enough evidence in the existing literature pointing to a critical need for remediation of 1,4-dioxane.

## 1.1 Biodegradation of 1,4-Dioxane

A major challenge in treating 1,4-dioxane is its chemical characteristics. It is highly miscible in water and tends to contaminate water supplies. It has low vapor pressure, low octanol-water partition coefficient and high solubility. Hence, it is challenging to treat using carbon adsorption or air stripping (USEPA 2017). Treatments using hydrogen peroxide, ozone, UV light, sonication are possible (Adams et al. 1994; Coleman et al. 2007; Son et al. 2009; Stefan and Bolton 1998). However, these ex-situ treatment processes are expensive.

Early reports in 1970-1980s indicated 1,4-dioxane as a recalcitrant compound showing high resistance to biodegradation (Alexander 1973; Klečka and Gonsior 1986). This compound was tested for biodegradation in sewage cultures for up to a year. The experiment was designed with 1,4-dioxane concentrations varying from 100 to 900 mg/L. There was no removal of 1,4-dioxane in the samples (Klečka and Gonsior 1986). This led to conclusions of 1,4-dioxane being non-biodegradable.

A research study conducted in Germany in the late 1980s focused on the enrichment and isolation of bacterial strains that could grow on aliphatic cyclic ethers including 1,4-dioxane, tetrahydrofuran (THF) and 1,3-dioxalane. A *Rhodococcus* strain (strain 219) was identified as being capable of degrading 1,4-dioxane (Bernhardt and Diekmann 1991). This strain was selected for enrichment because of its fast growth in minimal, complex media in addition to its ability to degrade 1,4-dioxane. The THF and 1,4-dioxane degrading ability was not impacted during the growth on complex media. At concentrations higher than 10 mM THF, the lag phase

prolonged and growth rate reduced. However, as the growth phase ended, there was no THF or 1,4-dioxane detected (Bernhardt and Diekmann 1991).

The first pure culture capable of 1,4-dioxane degradation was isolated in 1994 when Parales et al. identified a strain that could degrade 1,4-dioxane aerobically by using it as a sole source of carbon and energy (Parales et al. 1994). In that study, a pure culture of a nocardioform *Actinomyces* capable of sustained growth on 1,4-dioxane was developed. This strain showed the ability to mineralize 1,4-dioxane. The organism designated as CB1190 could degrade 1,4-dioxane to CO<sub>2</sub> with no organic products other than biomass. This was later identified as *Pseudonocardia dioxanivorans* CB1190 (Mahendra and Alvarez-Cohen 2005).

There exists a significant knowledge gap in anaerobic biodegradation of 1,4-dioxane. In fact, to date, there is only one published study documenting anaerobic biodegradation of 1,4-dioxane (Shen et al. 2008). The authors reported biodegradation when Fe(III)-EDTA was introduced as electron acceptor. However, the microbes/microbial community responsible for biodegradation were never identified. This represents a major knowledge gap because 1,4-dioxane is commonly found at anaerobic chlorinated solvent contaminated sites.

## 1.2 Other Organic Contaminants and Functional Genes

A multitude of organic contaminants are observed at contaminated sites apart from 1,4-dioxane discussed above. These include Polycyclic Aromatic Hydrocarbons (PAHs), dioxins, vinyl chloride, chlorinated compounds, munition related contaminants including RDX and TNT and often a mixture of one or more of these (Giusti 2009; Haritash and Kaushik 2009; Hawari et al.

2000; Hunkeler et al. 2005; Soclo et al. 2000; Tuppurainen et al. 1998; Weber et al. 2008). One method of studying the organic contaminant degradation potential in native soils would be through the analysis of functional genes present in the soils. Often, the functional gene abundance correlates with the degradation potential (Li et al. 2014; Mattes 2018). An effective way to analyze biodegradation and genes associated with biodegradation is through next generation sequencing.

### 1.3 Next Generation Sequencing Techniques

The traditional culture techniques can only reveal a small proportion of the microbial community in an environmental sample since about 99% of prokaryotes in the environment cannot be grown in growth media (Fang et al. 2013; Schloss and Handelsman 2005). DNA sequencing overcomes this limitation and enables the study of microbial population without culturing requirements. The DNA sequencing approach increased in popularity after the introduction of Sanger sequencing (Sanger et al. 1977). In the following years, faster sequencing techniques were developed, including the two next generation sequencing methods that were used in this dissertation work : 1) 16S ribosomal RNA amplicon sequencing (chapter 3) and 2) whole genome shotgun sequencing (chapter 2 and 4).

The 16S rRNA amplicon sequencing is based on the phylogenetically informative gene 16S, which is ubiquitous in bacterial populations. In this method, the 16S rRNA region is amplified using primers that can recognize the highly conserved region. Typically, the analysis can yield data at phyla and genera level but holds less precision at the species level (Ranjan et al. 2016).

Whole genome shotgun sequencing, unlike 16S rRNA amplicon sequencing, provides a more global overview of the microbial community (Chen and Pachter 2005; Venter et al. 2004). In this method, random strands of DNA sequences are re-assembled with regions of overlap. This permits a more thorough phylogenetic analysis along with the ability to study the functional genes. It also has the potential to uncover new genes that may have diverged too far from the known genes to be amplified by PCR (Chen and Pachter 2005).

#### 1.4 Dissertation Outline and Objectives

The overall objective of this dissertation is to examine aerobic and anaerobic 1,4-dioxane biodegradation in soils along with functional genes involved in biodegradation of organic contaminants using metagenomic analysis. The dissertation work is described in the chapters below:

Chapter 2: This chapter (published : Ramalingam V, Cupples A (2020) Enrichment of novel Actinomycetales and the detection of monooxygenases during aerobic 1,4-dioxane biodegradation with uncontaminated and contaminated inocula. Applied Microbiology and Biotechnology) details the biodegradation experiments conducted with 1,4-dioxane using soils and sediments. The experiment was conducted with inocula collected from agricultural soils, river sediments and contaminated sediments. The microcosms were spiked with 1,4-dioxane and aerobic biodegradation was observed in microcosms spiked with 1,4-dioxane and the nucleic acid was extracted. The metagenomes were analyzed using whole genome shotgun sequencing (Illumina HiSeq 4000), MG-RAST, STAMP, Trimmomatic, DIAMOND, and MEGA7.

The key objectives of this chapter were:

1. To identify which 1,4-dioxane degrading functional genes are present across different microbial communities.
2. To determine which genera may be using 1,4-dioxane and/or metabolites to support growth. The research focused on laboratory microcosms inoculated with four uncontaminated soils and sediment samples from two 1,4-dioxane contaminated sites.

Chapter 3: This work (published : Ramalingam V, Cupples A (2020). Anaerobic 1,4-Dioxane Biodegradation and Microbial Community Analysis in Microcosms Inoculated with Soils or Sediments and Different Electron Acceptors. Applied Microbiology and Biotechnology) examines the anaerobic biodegradation of 1,4-dioxane using different electron acceptors. For this study, microcosms were established with different inocula (soils or sediments) and various electron acceptor amendments (nitrate, iron-EDTA/humic acid, sulfate and no amendment). Additionally, compound specific isotope analysis (CSIA) was used to determine changes in  $^{13}\text{C}/^{12}\text{C}$  ratios in a subset of the samples. The work is novel as it is the first to document the frequency of 1,4-dioxane biodegradation over a range of redox conditions and inocula types and provides critical evidence for the feasibility of anaerobic 1,4-dioxane bioremediation. A 16S rRNA amplicon sequencing approach was utilized along with CSIA, Mothur and STAMP software.

The key objectives of this chapter were:

1. Determine the susceptibility of 1,4-dioxane to biodegradation under anaerobic conditions using inocula from contaminated and uncontaminated sites.
2. Use Compound Specific Isotope Analysis (CSIA) to determine changes in  $^{13}\text{C}/^{12}\text{C}$  ratios in some samples.

3. Compare the anaerobic biodegradation potential of 1,4-dioxane with different electron acceptors.

Chapter 4: This chapter (Ramalingam V, Cupples A Analysis of Biodegradation Genes with Shotgun Sequencing in Uncontaminated Agricultural Soils), explores a set of biodegradation genes using protein sequences downloaded from the Functional Gene Pipeline and Repository (FunGene). Although previous studies have examined these biodegradation genes in contaminated environmental samples, limited information is available on the presence and abundance of these genes in uncontaminated / agricultural soils. Agricultural soils are known to possess a diverse microbial community both phylogenetically and functionally. The genes were analyzed using whole genome shotgun sequencing data, Trimmomatic, DIAMOND, RStudio, EstimateS and MEGA X.

The key objectives of this chapter were:

1. Examine the presence of a set of biodegradation genes associated with organic contaminant degradation in four uncontaminated (agricultural) soils.
2. Compare the relative abundance and diversity of *benA*, *bph*, *dbfA*, *dxnA*, *etnC*, *etnE*, *ppaH*, *npaH*, *vcrA*, *xenA*, *xenB* and *xplA* in soils.

Chapter 5: The conclusions of this dissertation are outlined in this chapter along with the direction for future work.

## REFERENCES

## REFERENCES

- Adams CD, Scanlan PA, Secrist ND (1994) Oxidation and biodegradability enhancement of 1,4-dioxane using hydrogen peroxide and ozone. *Environ Sci Technol* 28(11):1812-1818
- Adamson DT, Anderson RH, Mahendra S, Newell CJ (2015) Evidence of 1,4-dioxane attenuation at groundwater sites contaminated with chlorinated solvents and 1,4-dioxane. *Environ Sci Technol* 49(11):6510-6518 doi:10.1021/acs.est.5b00964
- Alexander M (1973) Nonbiodegradable and other recalcitrant molecules. *Biotechnology and Bioengineering* 15(4):611-647 doi:10.1002/bit.260150402
- Anderson RH, Anderson JK, Bower PA (2012) Co-occurrence of 1,4-dioxane with trichloroethylene in chlorinated solvent groundwater plumes at US Air Force installations: Fact or fiction. *Integrated Environmental Assessment and Management* 8(4):731-737 doi:10.1002/ieam.1306
- Argus MF, Arcos JC, Hochligeti C (1965) Studies on the carcinogenic activity of protein-denaturing agents: hepatocarcinogenicity of dioxane. *J Natl Cancer Inst* 35(6):949-58
- ATSDR AFTSaDR (2012) Toxicological profile for 1,4-dioxane. In: <https://www.atsdr.cdc.gov/toxprofiles/TP.asp?id=955&tid=199> (ed).
- Bernhardt D, Diekmann H (1991) Degradation of dioxane, tetrahydrofuran and other cyclic ethers by an environmental *Rhodococcus* strain. *Appl Microbiol Biot* 36(1):120-123
- Chen K, Pachter L (2005) Bioinformatics for Whole-Genome Shotgun Sequencing of Microbial Communities. *PLOS Computational Biology* 1(2):e24 doi:10.1371/journal.pcbi.0010024
- Coleman HM, Vimonses V, Leslie G, Amal R (2007) Degradation of 1,4-dioxane in water using TiO<sub>2</sub> based photocatalytic and H<sub>2</sub>O<sub>2</sub>/UV processes. *J Hazard Mater* 146(3):496-501 doi:<https://doi.org/10.1016/j.jhazmat.2007.04.049>
- Fang H, Cai L, Yu Y, Zhang T (2013) Metagenomic analysis reveals the prevalence of biodegradation genes for organic pollutants in activated sludge. *Bioresour Technol* 129:209-218 doi:<https://doi.org/10.1016/j.biortech.2012.11.054>
- Giusti L (2009) A review of waste management practices and their impact on human health. *Waste Management* 29(8):2227-2239 doi:10.1016/j.wasman.2009.03.028
- Grimmett PE, Munch JW (2009) Method development for the analysis of 1,4-dioxane in drinking water using solid-phase extraction and gas chromatography-mass spectrometry. *J Chromatogr Sci* 47(1):31-9 doi:10.1093/chromsci/47.1.31

Haritash AK, Kaushik CP (2009) Biodegradation aspects of Polycyclic Aromatic Hydrocarbons (PAHs): A review. *J Hazard Mater* 169(1-3):1-15 doi:10.1016/j.jhazmat.2009.03.137

Hawari J, Beaudet S, Halasz A, Thiboutot S, Ampleman G (2000) Microbial degradation of explosives: biotransformation versus mineralization. *Appl Microbiol Biot* 54(5):605-618 doi:10.1007/s002530000445

Hunkeler D, Aravena R, Berry-Spark K, Cox E (2005) Assessment of degradation pathways in an aquifer with mixed chlorinated hydrocarbon contamination using stable isotope analysis. *Environ Sci Technol* 39(16):5975-5981 doi:10.1021/es048464a

IARC (1999) Re-evaluation of some organic chemicals, hydrazine and hydrogen peroxide. Proceedings of the IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, France, 17-24 February 1998. IARC monographs on the evaluation of carcinogenic risks to humans 71 Pt 1:1-315

Klečka GM, Gonsior SJ (1986) Removal of 1,4-dioxane from wastewater. *J Hazard Mater* 13(2):161-168 doi:https://doi.org/10.1016/0304-3894(86)80016-4

Li M, Mathieu J, Liu Y, Van Orden ET, Yang Y, Fiorenza S, Alvarez PJJ (2014) The Abundance of Tetrahydrofuran/Dioxane Monooxygenase Genes (*thmA/dxmA*) and 1,4-Dioxane Degradation Activity Are Significantly Correlated at Various Impacted Aquifers. *Environ Sci Tech Let* 1(1):122-127 doi:10.1021/ez400176h

Mahendra S, Alvarez-Cohen L (2005) *Pseudonocardia dioxanivorans* sp nov., a novel actinomycete that grows on 1,4-dioxane. *Int J Syst Evol Micr* 55:593-598 doi:DOI 10.1099/ijs.0.63085-0

Mattes T (2018) Validation of Biotechnology for Quantifying the Abundance and Activity of Vinyl-chloride Oxidizers in Contaminated Groundwater ESTCP Project ER-201425. The University of Iowa

Mohr T, Stickney J, Diguseppi B (2010) Environmental investigation and remediation : 1,4-dioxane and other solvent stabilizers. CRC Press/Taylor & Francis, Boca Raton, FL

Parales RE, Adamus JE, White N, May HD (1994) Degradation of 1,4-dioxane by an *Actinomycete* in pure culture. *Appl Environ Microb* 60(12):4527-4530

Ranjan R, Rani A, Metwally A, McGee HS, Perkins DL (2016) Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochemical and Biophysical Research Communications* 469(4):967-977 doi:https://doi.org/10.1016/j.bbrc.2015.12.083

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74(12):5463-7 doi:10.1073/pnas.74.12.5463

Schloss PD, Handelsman J (2005) Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. *Genome Biology* 6(8):229 doi:10.1186/gb-2005-6-8-229

Shen W, Chen H, Pan S (2008) Anaerobic biodegradation of 1,4-dioxane by sludge enriched with iron-reducing microorganisms. *Bioresource Technol* 99(7):2483-2487  
doi:10.1016/j.biortech.2007.04.054

Soclo HH, Garrigues P, Ewald M (2000) Origin of polycyclic aromatic hydrocarbons (PAHs) in coastal marine sediments: Case studies in Cotonou (Benin) and Aquitaine (France) areas. *Marine Pollution Bulletin* 40(5):387-396 doi:10.1016/s0025-326x(99)00200-3

Son H-S, Im J-K, Zoh K-D (2009) A Fenton-like degradation mechanism for 1,4-dioxane using zero-valent iron (Fe<sup>0</sup>) and UV light. *Water Res* 43(5):1457-1463  
doi:https://doi.org/10.1016/j.watres.2008.12.029

Stefan MI, Bolton JR (1998) Mechanism of the degradation of 1,4-dioxane in dilute aqueous solution using the UV hydrogen peroxide process. *Environ Sci Technol* 32(11):1588-1595

Tuppurainen K, Halonen I, Ruokojarvi P, Tarhanen J, Ruuskanen J (1998) Formation of PCDDs and PCDFs in municipal waste incineration and its inhibition mechanisms: A review. *Chemosphere* 36(7):1493-1511 doi:10.1016/s0045-6535(97)10048-0

USEPA (2017) Technical Fact Sheet 1,4-Dioxane. In: EPA US (ed).

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH, Smith HO (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304(5667):66-74  
doi:10.1126/science.1093857

Weber R, Gaus C, Tysklind M, Johnston P, Forter M, Hollert H, Heinisch E, Holoubek I, Lloyd-Smith M, Masunaga S, Moccarelli P, Santillo D, Seike N, Symons R, Torres JPM, Verta M, Varbelow G, Vijgen J, Watson A, Costner P, Woelz J, Wycisk P, Zennegg M (2008) Dioxin- and POP-contaminated sites-contemporary and future relevance and challenges. *Environmental Science and Pollution Research* 15(5):363-393 doi:10.1007/s11356-008-0024-1

Zenker MJ, Borden RC, Barlaz MA (2003) Occurrence and treatment of 1,4-dioxane in aqueous environments. *Env Eng Sci* 20(5):423-432

Zhou J, He Z, Yang Y, Deng Y, Tringe SG, Alvarez-Cohen L (2015) High-Throughput Metagenomic Technologies for Complex Microbial Community Analysis: Open and Closed Formats. *mBio* 6(1):e02288-14 doi:10.1128/mBio.02288-14

## Chapter 2 Enrichment of Novel Actinomycetales and the Detection of Monooxygenases during Aerobic 1,4-Dioxane Biodegradation with Uncontaminated and Contaminated Inocula

### 2.1 Abstract

1,4-dioxane, a co-contaminant at many chlorinated solvent sites, is a problematic groundwater pollutant because of risks to human health and characteristics which make remediation challenging. *In situ* 1,4-dioxane bioremediation has recently been shown to be an effective remediation strategy. However, the presence/abundance of 1,4-dioxane degrading species across different environmental samples is generally unknown. Here, the objectives were to identify which 1,4-dioxane degrading functional genes are present and which genera may be using 1,4-dioxane and/or metabolites to support growth across different microbial communities. For, this laboratory sample microcosms and abiotic control microcosms (containing media) were inoculated with four uncontaminated soils and sediments from two contaminated sites. The sample microcosms were amended with 1,4-dioxane three times. Live control microcosms were treated in the same manner, except 1,4-dioxane was not added. 1,4-dioxane decreased in live microcosms with all six inocula, but not in the abiotic controls, suggesting biodegradation occurred. A comparison of live sample microcosms and live controls (no 1,4-dioxane) indicated nineteen genera were enriched following exposure to 1,4-dioxane, suggesting a growth benefit for 1,4-dioxane biodegradation. The three most enriched were *Mycobacterium*, *Nocardioides*, *Kribbella* (classifying as *Actinomycetales*). There was also a higher level of enrichment for *Arthrobacter*, *Nocardia* and *Gordonia* (all three classifying as *Actinomycetales*) in one soil, *Hyphomicrobium* (*Rhizobiales*) in another soil, *Clavibacter* (*Actinomycetales*) and *Bartonella* (*Rhizobiales*) in another soil and *Chelativorans* (*Rhizobiales*) in another soil. Although

*Arthrobacter*, *Mycobacterium* and *Nocardia* have previously been linked to 1,4-dioxane degradation, *Nocardioides*, *Gordonia* and *Kribbella* are potentially novel degraders. The analysis of the functional genes associated with 1,4-dioxane demonstrated three genes were present at higher relative abundance values, including *Rhodococcus sp. RRI prmA*, *Rhodococcus jostii RHA1 prmA* and *Burkholderia cepacia G4 tomA3*. Overall, this study provides novel insights into the identity of the multiple genera and functional genes associated with the aerobic degradation of 1,4-dioxane in mixed communities.

## 2.2 Introduction

1,4-dioxane, a probable human carcinogen (DeRosa et al. 1996), was commonly used as a stabilizer in 1,1,1-trichloroethane formulations and is now frequently detected at sites where the chlorinated solvents are present (Adamson et al. 2015; Adamson et al. 2014b; ATSDR 2012; Mohr et al. 2010a). For example, 1,4-dioxane was found at 195 sites in California with 95% containing one or more of the chlorinated solvents (Adamson et al. 2014b). 1,4-dioxane has been classified as a probable carcinogen (Group 2B) by the U.S. EPA and a possible human carcinogen (B2) by the International Agency for Research on Cancer based on animal studies (IARC 1999; USEPA 2017). No federal maximum contaminant level for 1,4-dioxane in drinking water has been established (EPA 2017), however, several states have set low advisory action levels (e.g. California, Florida, Michigan and North Carolina have levels <5 ppb). A major challenge to 1,4-dioxane remediation concerns chemical characteristics that result in migration and persistence (Adamson et al. 2015; Mohr et al. 2010a). A low organic carbon partition coefficient ( $\log K_{oc} = 1.23$ ) and Henry's Law Constant ( $5 \times 10^{-6} \text{ atm. m}^3\text{mol}^{-1}$ ), make traditional remediation methods such as air stripping or activated carbon largely ineffective (Mahendra and

Alvarez-Cohen 2006; Steffan et al. 2007; Zenker et al. 2003). *Ex-situ* oxidation methods including ozone and hydrogen peroxide (Adams et al. 1994) or hydrogen peroxide and ultraviolet light (Stefan and Bolton 1998) have been commercially applied, however these can be expensive at high concentrations (Steffan et al. 2007).

Many bacteria have been linked to the aerobic metabolic and co-metabolic degradation of 1,4-dioxane. Currently, *Pseudonocardia dioxanivorans* CB1190 (Parales et al. 1994), *Rhodococcus ruber* 219 (Bock et al. 1996), *Pseudonocardia benzennivorans* B5 (Kämpfer and Kroppenstedt 2004), *Mycobacterium* sp. PH-06 (Kim et al. 2008), *Afipia* sp. D1, *Mycobacterium* sp. D6, *Mycobacterium* sp. D11, *Pseudonocardia* sp. D17 (Sei et al. 2013a), *Acinetobacter baumannii* DD1 (Huang et al. 2014a), *Rhodanbacter* AYS5 (Pugazhendi et al. 2015b), *Xanthobacter flavus* DT8 (Chen et al. 2016a) and *Rhodococcus aetherivorans* JCM 14343 (Inoue et al. 2016a) are known to degrade 1,4-dioxane metabolically. A large number of microorganisms are known to co-metabolically degrade this contaminant. For example, *Pseudonocardia tetrahydrofuranoxydans* sp. K1 (Kohlweyer et al. 2000), *Pseudonocardia* sp. ENV478 (Vainberg et al. 2006), *Rhodococcus ruber* T1, *Rhodococcus ruber* T5 (Sei et al. 2013a), *Rhodococcus ruber* ENV 425 (Steffan et al. 1997a), *Rhodococcus* RR1 (Stringfellow and Alvarez-Cohen 1999), *Flavobacterium* sp. (Sun et al. 2011a), *Mycobacterium vaccae* (Burback and Perry 1993), *Mycobacterium* sp. ENV 421 (Masuda et al. 2012b), *Pseudomonas mendocina* KR1 (Whited and Gibson 1991), *Ralstonia pickettii* PKO1 (Kukor and Olsen 1990), *Burkholderia cepacia* G4 (Nelson et al. 1986), *Methylosinus trichosporium* OB3b (Whittenbury et al. 1970b), *Pseudonocardia acacia* JCM (Inoue et al. 2016a) and *Pseudonocardia asaccharolytica* JCM (Inoue et al. 2016a) are among those linked to co-metabolic degradation of 1,4-dioxane. Co-

metabolic 1,4-dioxane degradation has previously been reported with growth supporting substrates such as tetrahydrofuran, methane, propane, toluene, ethanol, sucrose, lactate, yeast extract and 2-propanol (Burbach and Perry 1993; Hand et al. 2015; Kohlweyer et al. 2000; Mahendra and Alvarez-Cohen 2006; Vainberg et al. 2006).

The initiation of 1,4-dioxane biodegradation has been associated with various groups of soluble di-iron monooxygenases (SDIMOs) (He et al. 2017). Monooxygenases are enzymes that facilitate bacterial oxidation through the introduction of oxygen. SDIMOs have been classified into 6 groups based on their preferred substrate and sequence similarity (Coleman et al. 2006). SDIMOs associated with metabolic and co-metabolic 1,4-dioxane degradation include [as summarized in (He et al. 2017)] *Burkholderia cepacia* G4 *tomA3* (Group 1) (Mahendra and Alvarez-Cohen 2006; Newman and Wackett 1995), *Pseudomonas pickettii* PKO1 *tbuA1* (Group 2) (Fishman et al. 2004; Mahendra and Alvarez-Cohen 2006), *Pseudomonas mendocina* KR1 *tmoA* (Group 2) (Mahendra and Alvarez-Cohen 2006; Yen et al. 1991), *Methylosinus trichosporium* OB3b *mmoX* (Group 3) (Mahendra and Alvarez-Cohen 2006; Oldenhuis et al. 1989), *Pseudonocardia dioxanivorans* CB1190 *prmA* (Group 5) (Parales et al. 1994; Sales et al. 2013; Sales et al. 2011), *Pseudonocardia tetrahydrofuranoxydans* K1 *thmA* (Group 5) (Kampfer et al. 2006; Thiemer et al. 2003), *Pseudonocardia* sp. strain ENV478 *thmA* (Group 5) (Masuda et al. 2012a), *Rhodococcus* sp. strain YYL *thmA* (Group 5) (Yao et al. 2009), *Rhodococcus jostii* RHA1 *prmA* (Group 5) (Hand et al. 2015; Sharp et al. 2007), *Rhodococcus* sp. RR1 *prmA* (Group 5) (Sharp et al. 2007), *Mycobacterium* sp. ENV421 *prmA* (Group 6) (Masuda 2009) and *Mycobacterium dioxanotrophicus* PH-06 *prmA* (Group 6) (He et al. 2017).

As the success of natural attenuation or biostimulation often depends on the population of native degraders present at the contaminated site, several studies have developed methods targeting these functional genes (Gedalanga et al. 2014; He et al. 2018; Li et al. 2013a; Li et al. 2013b). For example, methods have been developed for the functional genes associated with *Pseudonocardia* and *Mycobacterium* (Deng et al. 2018a; Gedalanga et al. 2014; He et al. 2017). Another study focused specifically on detecting functional genes of four 1,4-dioxane degraders (*Pseudonocardia dioxanivorans* CB1190, *Pseudonocardia* sp. strain ENV478, *Pseudonocardia tetrahydrofuranoxydans* K1, *Rhodococcus* sp. strain YYL). A larger number of functional genes were investigated with microarray-based technology (GeoChip 4.0) and denaturing gradient gel electrophoresis (Li et al. 2013b). More recently, high throughput shotgun sequencing was used to evaluate the presence of the functional genes listed above in impacted and non-impacted groundwater (Dang et al. 2018). This approach has the added advantage of enabling taxonomic as well as functional analysis of microbial communities. The current study adopted a similar approach to examine the microbial communities involved in 1,4-dioxane degradation in contaminated and uncontaminated sediment and soil inoculated microcosms.

In the current study, the objectives were 1) to identify which 1,4-dioxane degrading functional genes are present across different microbial communities and 2) to determine which genera may be using 1,4-dioxane and/or metabolites to support growth. The research focused on laboratory microcosms inoculated with four uncontaminated soils and sediment samples from two 1,4-dioxane contaminated sites. The media selected for the experiments followed the approach used to enrich *Pseudonocardia dioxanivorans* CB1190 from industrial sludge (Parales et al. 1994). The work is novel as it combines taxonomic and functional data to generate a more complete

picture of the multiple microorganisms and genes linked to 1,4-dioxane degradation in mixed communities.

## 2.3 Methods

### 2.3.1 Chemicals and Inocula

1,4-dioxane was purchased from Sigma-Aldrich (MO, USA) with 99.8% purity. All stock solutions and dilutions were prepared using DI water. The agricultural samples were collected from two locations on the campus of Michigan State University, East Lansing, Michigan (herein called soils F and G) and two locations at the Kellogg Biological Station, Hickory Corners, Michigan (soils 1 and 2). The contaminated site samples were obtained from California (contaminated with trichloroethene, 1,1-dichloroethene and 1,4-dioxane, herein called C7A) and Maine (contaminated with traces of 1,4-dioxane, herein called M10A). All samples were stored in the dark at 6 °C until use.

### 2.3.2 Experimental Setup, DNA Extraction, 1,4-Dioxane Analysis

Laboratory microcosms were established with soil or sediment (5g wet weight) and 25 mL of media in 30 mL serum bottles. For each of the six inocula (four uncontaminated soils or two contaminated sediments), the experiment design included triplicate sample microcosms, triplicate live control microcosms and triplicate abiotic control microcosms (autoclaved daily for three consecutive days). The triplicate live control microcosms were treated in the same manner as the sample microcosms except no 1,4-dioxane was added. This treatment was included to enable comparisons to the microbial communities exposed to 1,4-dioxane. Following the approach used to enrich *Pseudonocardia dioxanivorans* CB1190 from industrial sludge, each

liter of the final media contained 100 mL of a buffer stock [ $\text{K}_2\text{HPO}_4$  (32.4 g/L),  $\text{KH}_2\text{PO}_4$  (10 g/L),  $\text{NH}_4\text{Cl}$  (20 g/L)] and 100 mL of a trace metal stock [nitrilotriacetic acid (disodium salt) (1.23 g/L),  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  (2 g/L),  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$  (0.12 g/L),  $\text{MnSO}_4 \cdot \text{H}_2\text{O}$  (0.03 g/L),  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$  (0.03 g/L) and  $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$  (0.01 g/L)] (Parales et al., 1994). The nitrilotriacetic acid (within the trace metal stock solution) represents an additional carbon source. The live sample microcosms were re-spiked with 1,4-dioxane two additional times.

A GC/MS with Agilent 5975 GC/single quadrupole MS (Agilent Technologies, CA, USA) equipped with a CTC Combi Pal autosampler was used to determine 1,4-dioxane concentrations. Sterile syringes (1 mL) and needles (22 Ga 1.5 in.) were used to collect samples (0.7 mL) from each microcosm. The samples were filtered (0.22  $\mu\text{m}$  nylon filter) before being injected into an amber glass vial (40 mL) for GC/MS analysis. A method was developed to analyze 1,4-dioxane using solid phase micro extraction (SPME). The SPME fiber was inserted in the headspace of the vial and exposed to the analyte for 1 minute before being injected into the GC for thermal desorption. The fiber coating can adsorb the analytes in the vapor phase. Splitless injection was executed and the vials were maintained at 40 °C. The SPME fiber assembly involved 50/30 $\mu\text{m}$  Divinylbenzene/ Carboxen/ Polydimethylsiloxane (DVB/CAR/PDMS) and 24 Ga needle for injection. The initial oven temperature was 35 °C and was programmed to increase at a rate of 20 °C/min to 120 °C. Once it reached 120 °C, it increased at a rate of 40 °C/min to 250 °C, which was maintained for 3 min. A VF5MS column was used with helium as the carrier gas in constant flow mode at a flow rate of 1 ml/min. The conditioning of the SPME fiber was at 270 °C for 60 min at the beginning of each sequence.

DNA was extracted from the soil inoculated sample microcosms and live control microcosms (1.2 mL and 0.4 g soil) using QIAGEN DNeasy PowerSoil kit as per the manufacturer's instructions. The QIAGEN DNeasy Powermax Soil kit was used to extract DNA from the microcosms inoculated with the two contaminated sediments. For this, the entire content of each microcosm was sacrificed for DNA extraction. The DNA concentrations were determined using QUBIT dsDNA HS kit. The DNA extracts with the highest DNA yields were selected for shotgun sequencing.

### 2.3.3 Library Preparation, Sequencing, MG-RAST and DIAMOND analysis

Twenty-six samples were submitted for library generation and shotgun sequencing to the Research Technology Support Facility Genomics Core at Michigan State University. Libraries were prepared using the Takara SMARTer ThruPLEX DNA Library Preparation Kit following manufacturer's recommendations. Completed libraries were QC'd and quantified using a combination of Qubit dsDNA HS and Agilent 4200 TapeStation HS DNA1000 assays. Eight samples did not generate libraries of sufficient concentration for sequencing and were removed from further analysis. The remaining eighteen libraries were pooled in equimolar amounts for multiplexed sequencing. The pool was quantified using the Kapa Biosystems Illumina Library Quantification qPCR kit and loaded onto one lane of an Illumina HiSeq 4000 flow cell. Sequencing was performed in a 2x150 bp paired end format using HiSeq 4000 SBS reagents. Base calling was done by Illumina Real Time Analysis (RTA) v2.7.7 and output of RTA was demultiplexed and converted to FastQ format with Illumina Bcl2fastq v2.19.1.

The Meta Genome Rapid Annotation using Subsystem Technology (MG-RAST) (Meyer et al. 2008) version 4.0.2. was used for the taxonomic analysis of the metagenomes. The processing pipeline involved merging paired end reads, SolexaQA (Cox et al. 2010) to trim low-quality regions and dereplication to remove the artificial duplicate reads. Gene calling was performed using FragGeneScan (Rho et al. 2010). For the taxonomic profiles, the best hit classification at a maximum e-value of  $1e^{-5}$ , a minimum identity of 60% and a minimum alignment length of 15 against the ReqSeq database (Pruitt et al. 2005) were used. MG-RAST ID numbers and sequencing data have been summarized (Table 2.1) and the datasets are publicly available on MG-RAST. The number of sequences generated post quality control per sample was  $4.7 \pm 2.0$  million (ranging from ~1.2 to ~11 million) and the average length was  $237.3 \pm 2.9$  bp (averages ranging from 233 to 243 bp).

Table 2.1. Summary of sequencing information processed by MG-RAST

ID	Name	Upload: bp Count	Upload: Sequences Count	Artificial Duplicate Reads: Sequence Count	Post QC: bp Count	Post QC: Sequences Count	Post QC: Mean Sequence Length bp
mgm4846244.3	C7A_6_S10_L001_R	928,461,523	3,828,893	730,886	745,308,161	3,067,261	243 ± 33
mgm4842040.3	SF_2_S3_L001_R	1,249,466,781	5,278,200	741,761	1,057,029,663	4,469,282	237 ± 36
mgm4846245.3	SF_3_S6_L001_R	1,497,947,613	6,267,333	1,045,636	1,227,596,223	5,142,267	239 ± 35
mgm4846246.3	SF_4_S8_L001_R	698,047,178	2,986,100	484,606	572,423,699	2,454,542	233 ± 36
mgm4846247.3	SF_5_S11_L001_R	1,455,181,423	6,244,496	875,408	1,226,035,914	5,268,011	233 ± 37
mgm4846248.3	SF_6_S13_L001_R	1,276,229,786	5,446,704	791,730	1,069,530,685	4,570,187	234 ± 36
mgm4846291.3	SG_1_S15_L001_R	362,019,482	1,517,474	232,203	300,578,338	1,261,641	238 ± 35
mgm4841972.3	SG_3_S17_L001_R	1,237,320,023	5,151,224	739,217	1,042,837,451	4,346,084	240 ± 34
mgm4841973.3	SG_4_S1_L001_R	1,183,825,731	4,906,054	779,630	975,862,599	4,048,734	241 ± 34
mgm4842102.3	SG_6_S4_L001_R	1,097,936,207	4,598,102	612,987	933,115,160	3,911,767	239 ± 35
mgm4841974.3	ST1_2_S7_L001_R	1,722,580,822	7,248,808	1,205,707	1,409,132,207	5,932,625	238 ± 35
mgm4846290.3	ST1_3_S9_L001_R	687,193,817	2,855,664	394,713	580,636,546	2,415,127	240 ± 35
mgm4842023.3	ST1_4_S12_L001_R	1,720,695,830	7,269,781	1,124,424	1,433,388,134	6,053,609	237 ± 35
mgm4842024.3	ST1_6_S14_L001_R	1,165,586,999	4,879,142	683,336	983,135,037	4,119,069	239 ± 35
mgm4842104.3	ST2_1_S16_L001_R	3,144,582,833	13,429,024	2,172,387	2,579,894,248	11,026,268	234 ± 36
mgm4842103.3	ST2_3_S18_L001_R	1,274,658,549	5,369,262	727,123	1,080,887,053	4,557,307	237 ± 35
mgm4842107.3	ST2_5_S2_L001_R	1,622,482,462	6,854,685	982,861	1,362,403,381	5,762,840	236 ± 35
mgm4842106.3	ST2_6_S5_L001_R	1,463,433,476	6,249,417	846,173	1,252,082,629	5,349,932	234 ± 36

QC – Quality Control

The MG-RAST data files were downloaded and analyzed in Microsoft Excel 2016, STAMP (Statistical Analyses of Metagenomic Profiles, software version 2.1.3.) (Parks et al. 2014) and MEGAN6 (version 6.11.7) (Huson et al. 2016). STAMP was used to detect differences in the relative proportions of the taxonomic profiles between the live controls (no 1,4-dioxane) and the samples for each soil. This analysis included Welch's two sided t-test for two groups (samples and live controls) ( $p < 0.05$ ) to generate extended error bar figures for each soil. The same Welch's test was performed to compare the profiles of all samples ( $n = 9$ ) to all live controls ( $n = 8$ ). MEGAN6 was used to generate two phylograms. One phylogram illustrates the eighteen metagenomes classified to the Class Level. The other phylogram represents the most common genera (ranked by average relative abundance, then selected if average values  $> 0.5\%$ ) across all metagenomes.

The relative abundance of 1,4-dioxane degrading functional genes was determined using the alignment tool DIAMOND (double index alignment of next-generation sequencing data) (Buchfink et al. 2015). Specifically, reads aligning to the twelve genes previously associated with aerobic degradation (metabolic and co-metabolic) of 1,4-dioxane, as summarized previously (He et al. 2017), were determined. First, low quality sequences and Illumina adapters sequences were removed using Trimmomatic in the paired end mode (Bolger et al. 2014). The two paired output files were used for gene alignments in DIAMOND. Following alignment, the DIAMOND files were analyzed within Excel, which included combining the data from the two paired files and deleting duplicated data. The sort function was used to select reads that exhibited an identity of  $\geq 60\%$  and an alignment length  $\geq 49$  amino acids. For each gene, the relative abundance

values were calculated using the number of aligned reads divided by the total number of sequences for each sample (determined by Trimmomatic).

The above analysis indicated two functional genes (*Rhodococcus jostii* RHA1 *prmA* and *Rhodococcus* sp. *prmA*) were dominant in the soil and sediment metagenomes. Therefore, BLASTP 2.9.0+ (Altschul et al. 1997) (protein-protein BLAST) from the NCBI website was used to search for similar protein sequences to these two genes. The sequences obtained (>94.8% similar to the two query sequences) were used to create a phylogenetic tree in MEGAN7 using the Maximum Likelihood method (Jones et al. 1992; Kumar et al. 2016).

## 2.4 Results

### 2.4.1 Biodegradation of 1,4-Dioxane

The concentration of 1,4-dioxane declined in all of the live microcosms (inoculated with all four uncontaminated soils and with two contaminated site sediments), but not in the abiotic controls, indicating biological removal (Figure 2.1). All of the live microcosms, except M10A (contaminated site soil), demonstrated >50% removal in 1,4-dioxane in approximately 40 days. Following the initial biodegradation of the chemical, the microcosms were reamended with 1,4-dioxane twice. A steady depletion of 1,4-dioxane occurred in all four agricultural soils and the two contaminated site soil samples after each reamendment, while no removal was noted in the corresponding abiotic controls. However, limited biodegradation (only one replicate decreased) was observed for soil F after the last amendment. Overall, between approximately 220 and 245 days was required to reduce the majority of the amended 1,4-dioxane.

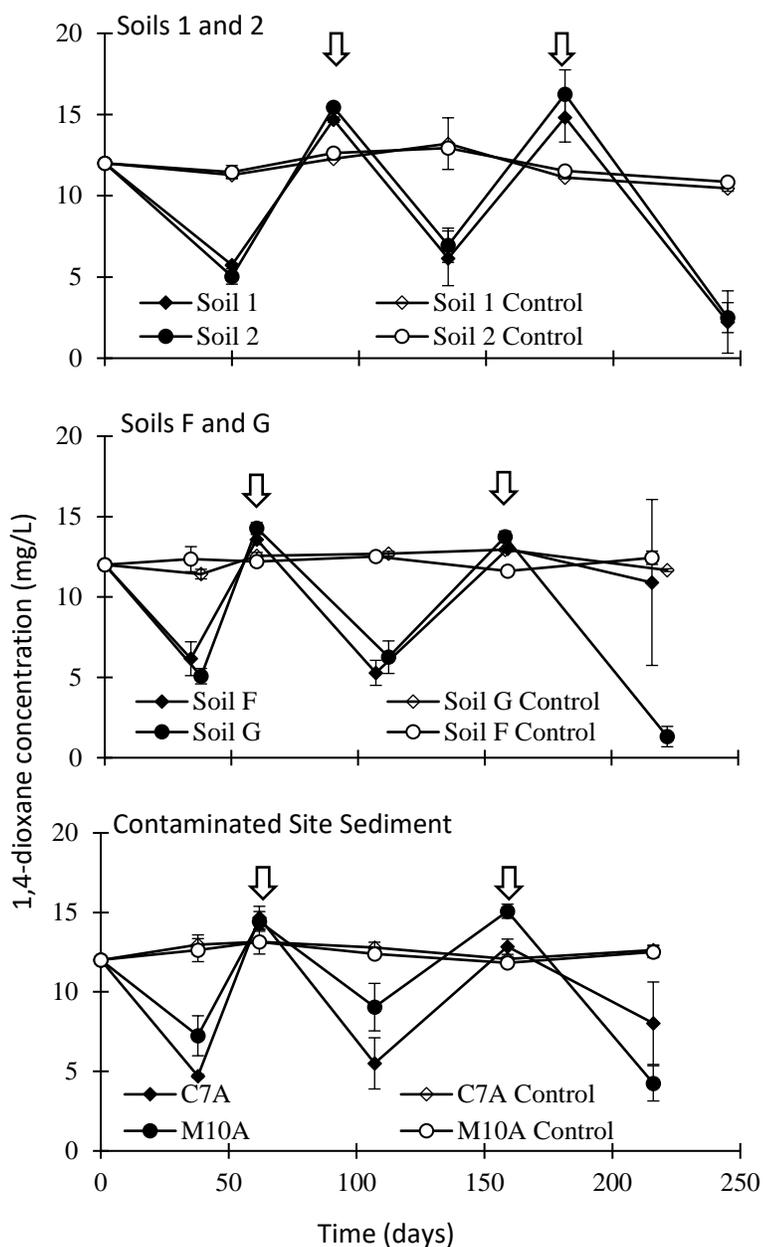


Figure 2.1. Average 1,4-dioxane concentrations (mg/L) in triplicate samples and abiotic controls with different inocula, including four agricultural soils and sediments from two contaminated sites (bars represent standard deviations). 1,4-dioxane was reamended to the samples microcosms twice (arrows).

#### 2.4.2 Taxonomic Analysis of Metagenomes

Although DNA was extracted from all microcosms and live controls (no 1,4-dioxane), in some cases insufficient DNA was extracted (and was not submitted for library generation) or did not generate libraries of sufficient concentration for sequencing. Unfortunately, this included all DNA extracts for the microcosms inoculated with sediment from one of the two 1,4-dioxane contaminated sites (M10A three samples and three live controls). Also, only one sample from the other contaminated site (C7A) generated enough DNA for sequencing. Overall, eighteen libraries were sequenced, which included two samples and two live controls for each soil (except soil F which included three samples) and one sample from one contaminated site (C7A).

The phylogenetic analysis of the eighteen soil and sediment metagenomes indicated the majority of the microorganisms classified within the classes *Acidobacteria*, *Alpha-*, *Beta-*, *Gamma-*, *Deltaproteobacteria*, *Actinobacteria*, *Bacilli* and *Clostridia* (Appendix Figure 2.9). The most abundant genera, averaged across all metagenomes, included *Candidatus Solibacter*, *Bradyrhizobium*, *Mesorhizobium*, *Burkholderia*, *Pseudomonas*, *Stenotrophomonas*, *Xanthomonas*, *Mycobacterium* and *Streptomyces* (Appendix Figure 2.10). The relative abundance (%) of the most abundant genera (25 most abundant) for each soil analyzed separately is also shown (Appendix Figure 2.11). The most abundant genera in all four soils were similar and included *Xanthomonas*, *Streptomyces*, *Mesorhizobium*, *Bradyrhizobium* and *Burkholderia*. In contrast, *Pseudomonas*, *Rhodococcus*, *Arthrobacter*, *Mycobacterium* and *Corynebacterium* were the most abundant genera in the contaminated site microcosms.

### 2.4.3 Genera Associated with 1,4-Dioxane Degradation

The metagenomes of the samples were compared to the live controls (no 1,4-dioxane) to determine which genera were positively influenced by 1,4-dioxane degradation. First, all of the samples ( $n=9$ ) were compared to all of the live controls ( $n=8$ ) (Figure 2.2). Overall, fifteen genera were statistically significantly enriched in the live samples compared to the controls. The greatest differences between the means were noted for *Mycobacterium* (0.304%,  $p=0.0029$ ), followed by *Nocardioides* (0.127%,  $p=0.023$ ), and *Kribbella* (0.079%,  $p=0.017$ ). The trends suggest these genera are obtaining a growth benefit from the presence of 1,4-dioxane. The relative abundance of these genera in the contaminated site microcosm is also shown (Figure 2.2B, insert). Except for *Ureaplasma*, the enriched genera all classify within the order *Actinomycetales* (Table 2.2).

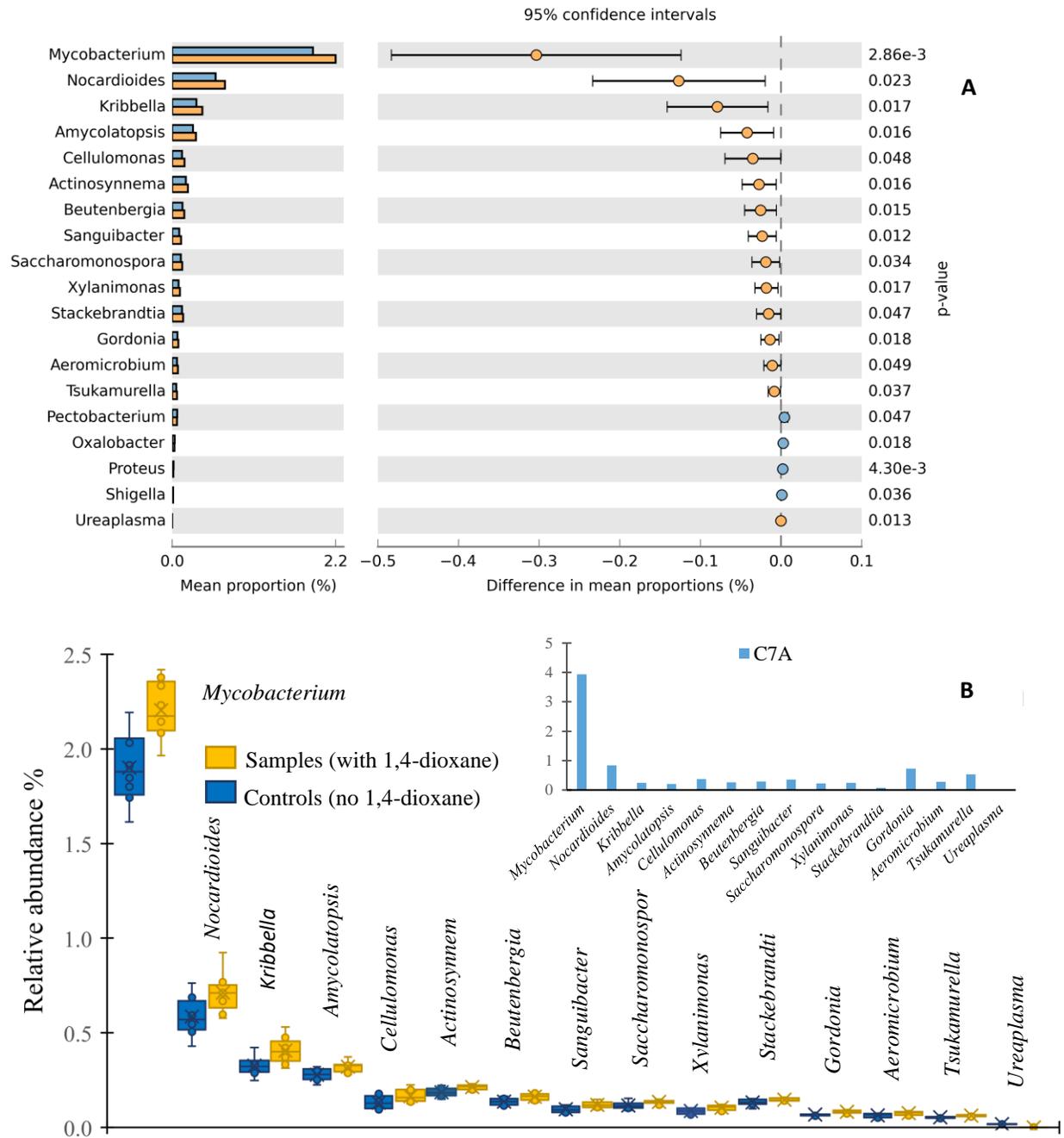


Figure 2.2. Extended error bar plot illustrating genera statistically significantly different in relative abundance (Welch's two sided t-test,  $p < 0.05$ ) between the samples ( $n=9$ ) and the live controls (no 1, 4-dioxane,  $n=8$ ) following 1,4-dioxane degradation (A). The symbols to the left of the dashed line (yellow) indicate a higher relative abundance in the samples compared to the controls and the symbols to the right (blue) indicate the reverse. A comparison of the relative abundance values (%) for the genera enriched in the samples is also shown in a box plot format (B). The insert illustrates the relative abundance of these enriched genera in the contaminated site sample (C7A) with a different y-axis scale.

Table 2.2. Classification of genera statistically significantly enriched ( $p < 0.05$ ) in the samples compared to the controls (no 1,4-dioxane) following the degradation of 1,4-dioxane in all soils collectively and when the soils were analyzed individually. The last column also illustrates the difference in means between the controls and the samples for each genera. Genera in bold were identified in the BLASTP search as containing genes similar to *Rhodococcus jostii* RHA1 prmA and *Rhodococcus* sp. prmA (as discussed in the results section for the functional gene analysis).

Phylum	Class	Order	Family	Genus	Difference in Means (%)
<b>All Soils: All samples (n=9) compared to all controls (n=8)</b>					
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Mycobacteriaceae</i>	<b><i>Mycobacterium</i></b>	0.304
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Nocardioideaceae</i>	<b><i>Nocardioides</i></b>	0.127
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Nocardioideaceae</i>	<b><i>Kribbella</i></b>	0.079
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Pseudonocardiaceae</i>	<i>Amycolatopsis</i>	0.042
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Cellulomonadaceae</i>	<i>Cellulomonas</i>	0.035
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Actinosynnemataceae</i>	<i>Actinosynnema</i>	0.027
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Beutenbergiaceae</i>	<i>Beutenbergia</i>	0.025
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Sanguibacteraceae</i>	<i>Sanguibacter</i>	0.023
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Pseudonocardiaceae</i>	<i>Saccharomonospora</i>	0.019
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Promicromonosporaceae</i>	<i>Xylanimonas</i>	0.018
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Glycomycetaceae</i>	<i>Stackebrandtia</i>	0.015
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Gordoniaceae</i>	<b><i>Gordonia</i></b>	0.014
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Nocardioideaceae</i>	<i>Aeromicrobium</i>	0.011
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Tsukamurellaceae</i>	<i>Tsukamurella</i>	0.008
<i>Tenericutes</i>	<i>Mollicutes</i>	<i>Mycoplasmatales</i>	<i>Mycoplasmataceae</i>	<i>Ureaplasma</i>	0.0002
<b>Soil 1: Samples (n=3) compared to controls (n=2)</b>					
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Microbacteriaceae</i>	<i>Clavibacter</i>	0.017
<i>Proteobacteria</i>	<i>Alphaproteobacteria</i>	<i>Rhizobiales</i>	<i>Bartonellaceae</i>	<i>Bartonella</i>	0.010
<b>Soil 2: Samples (n=2) compared to controls (n=2)</b>					
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Micrococcaceae</i>	<i>Arthrobacter</i>	0.276
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Nocardiaceae</i>	<b><i>Nocardia</i></b>	0.049
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Gordoniaceae</i>	<b><i>Gordonia</i></b>	0.019
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Micrococcaceae</i>	<i>Kocuria</i>	0.017
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Brevibacteriaceae</i>	<i>Brevibacterium</i>	0.015
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Micrococcaceae</i>	<i>Rothia</i>	0.005
<i>Firmicutes</i>	<i>Erysipelotrichi</i>	<i>Erysipelotrichales</i>	<i>Erysipelotrichaceae</i>	<i>Erysipelothrix</i>	0.0005
<b>Soil F: Samples (n=2) compared to controls (n=2)</b>					
<i>Proteobacteria</i>	<i>Alphaproteobacteria</i>	<i>Rhizobiales</i>	<i>Hyphomicrobiaceae</i>	<i>Hyphomicrobium</i>	0.033
<i>Proteobacteria</i>	<i>Alphaproteobacteria</i>	<i>Rhodospirillales</i>	<i>Acetobacteraceae</i>	<i>Acetobacter</i>	0.003
<i>Firmicutes</i>	<i>Negativicutes</i>	<i>Selenomonadales</i>	<i>Veillonellaceae</i>	<i>Veillonella</i>	0.002
<b>Soil G: Samples (n=2) compared to controls (n=2)</b>					
<i>Proteobacteria</i>	<i>Alphaproteobacteria</i>	<i>Rhizobiales</i>	<i>Phyllobacteriaceae</i>	<i>Chelativorans</i>	0.055

The metagenomes of the samples and controls were also compared for each soil individually. Two (*Clavibacter*, *Bartonella*) and seven genera (*Arthrobacter*, *Nocardia*, *Gordonia*, *Kocuria*, *Brevibacterium*, *Rothia*, *Erysipelothrix*) were statistically significantly enriched in the samples compared to the controls in soils 1 and 2, respectively (Figure 2.3). Three genera (*Hyphomicrobium*, *Acetobacter*, *Veillonella*) and one genus (*Chelativorans*) were statistically significantly enriched in the samples compared to the controls in soils F and G, respectively (Figure 2.4). Seven of the thirteen listed above classify within the *Actinomycetales* (Table 2.2). The differences between the means (for the individual soil analysis) were the highest ( $\geq 0.033\%$ ) for *Arthrobacter*, *Nocardia*, *Hyphomicrobium* and *Chelativorans* (Table 2.2). The relative abundance of the thirteen enriched genera in the samples compared to the controls (for the individual soil analysis) and for the contaminated site sample is shown (Figure 2.5). The contaminated site sample is shown with a different scale as it involved higher relative abundance values compared to the other microcosms (Figure 2.5 insert). Both *Arthrobacter* and *Nocardia* have relative abundance values of  $>1\%$  in the contaminated site sample.

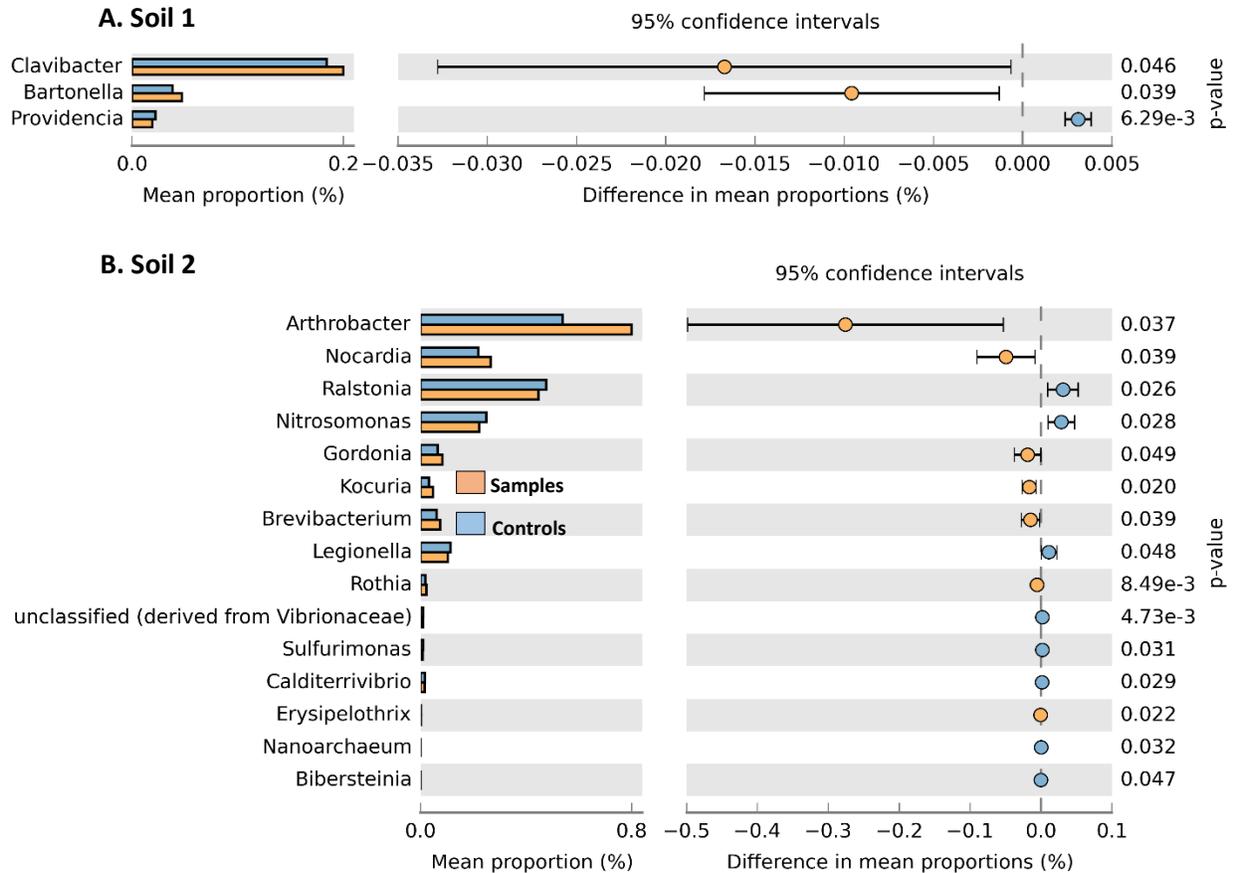


Figure 2.3. Extended error bar plots illustrating genera statistically significantly different in relative abundance (Welch's two sided t-test,  $p < 0.05$ ) between the samples and the live controls following 1,4-dioxane degradation in soil 1 (A) and 2 (B). The symbols to the left of the dashed line (in yellow) indicate a higher relative abundance in the samples compared to the controls and the symbols to the right (in blue) indicate the reverse.

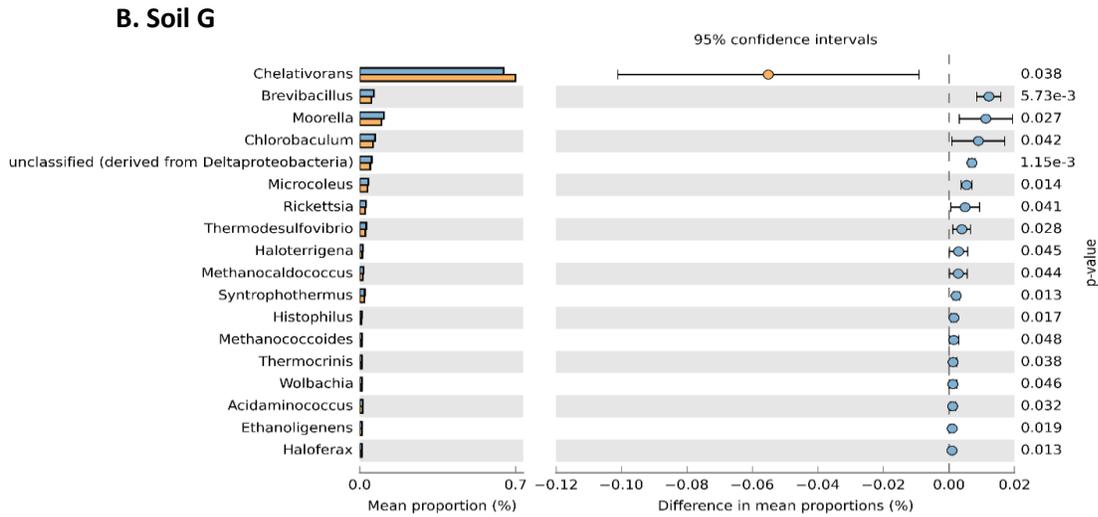
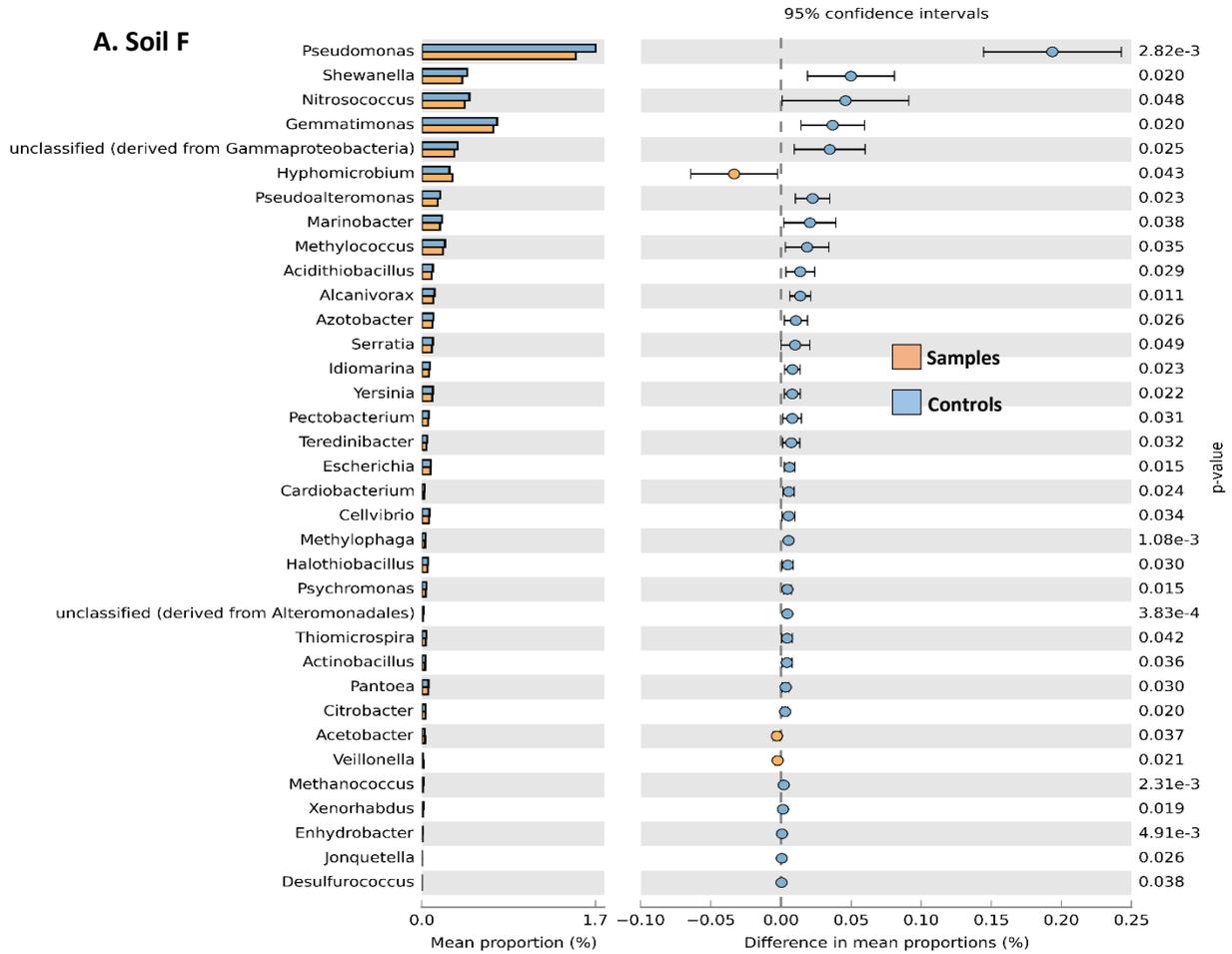


Figure 2.4. Extended error bar plots illustrating genera statistically significantly different in relative abundance (Welch's two sided t-test,  $p < 0.05$ ) between the samples and the live controls following 1,4-dioxane degradation in soil F (A) and soil G. The symbols to the left of the dashed line (in yellow) indicate a higher relative abundance in the samples compared to the controls and the symbols to the right (in blue) indicate the reverse.

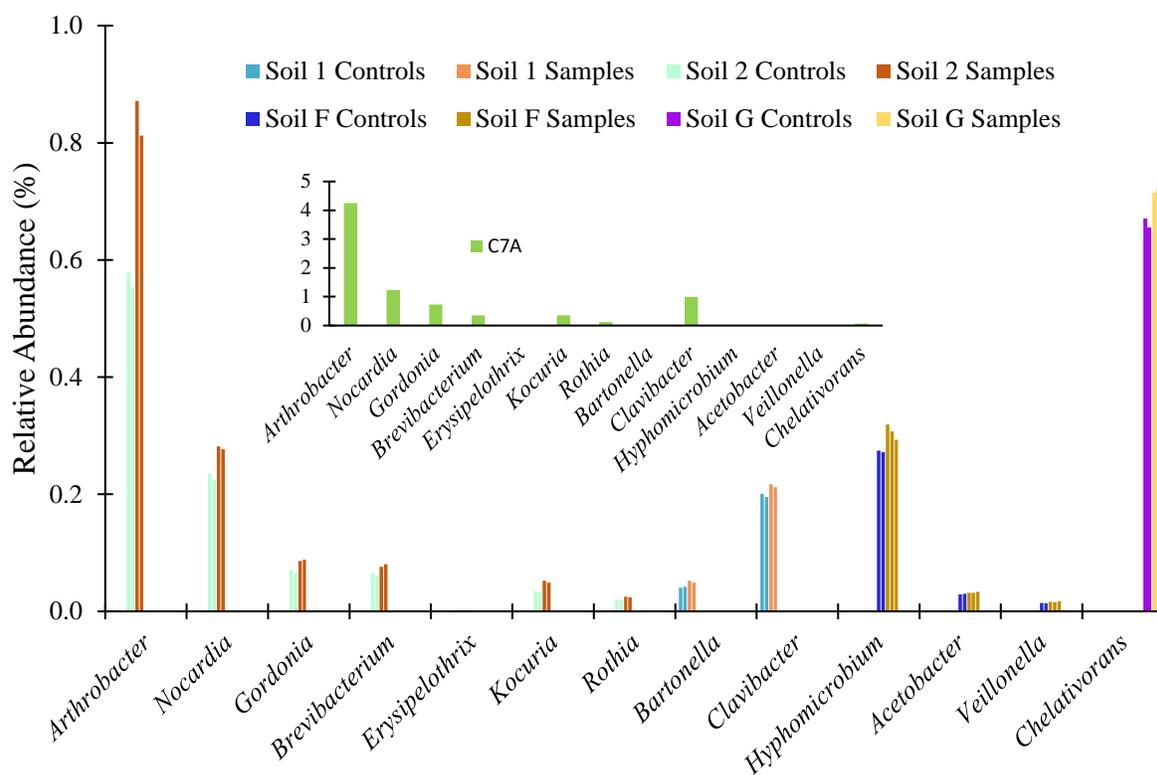


Figure 2.5. Summary of the relative abundance of statistically significantly enriched genera in the samples compared to the controls (no 1,4-dioxane) for soils 1, 2, F and G. The insert illustrates the relative abundance of these genera in the contaminated site sample (C7A) with a different scale on the y-axis.

#### 2.4.4 Relative Abundance of Genera Associated with 1,4-dioxane Biodegradation

The metagenomes were also investigated to determine the relative abundance (%) of fifteen genera previously associated with metabolic or co-metabolic 1,4-dioxane degradation (Figure 2.6). All except *Pseudonocardia* and *Rhodanbacter* were present in the samples and controls. *Burkholderia*, *Mycobacterium*, *Pseudomonas* and *Rhodococcus* were present at the highest relative abundance levels (0.84-2.45%). Only *Mycobacterium* was statistically significantly ( $p < 0.05$ ) enriched in the samples compare to the live controls. *Pseudonocardia* and *Rhodanbacter* were also absent in the contaminated site sample (Figure 2.6, insert). In the contaminated site metagenome, the four most abundant genera were *Pseudomonas* (49.0%), *Rhodococcus* (5.9%), *Mycobacterium* (3.9%) and *Nocardia* (1.2%). The contaminated site sample indicates a 25-fold higher relative abundance in *Pseudomonas* and almost a 6-fold higher relative abundance in *Rhodococcus* compared to the samples from agricultural sites.

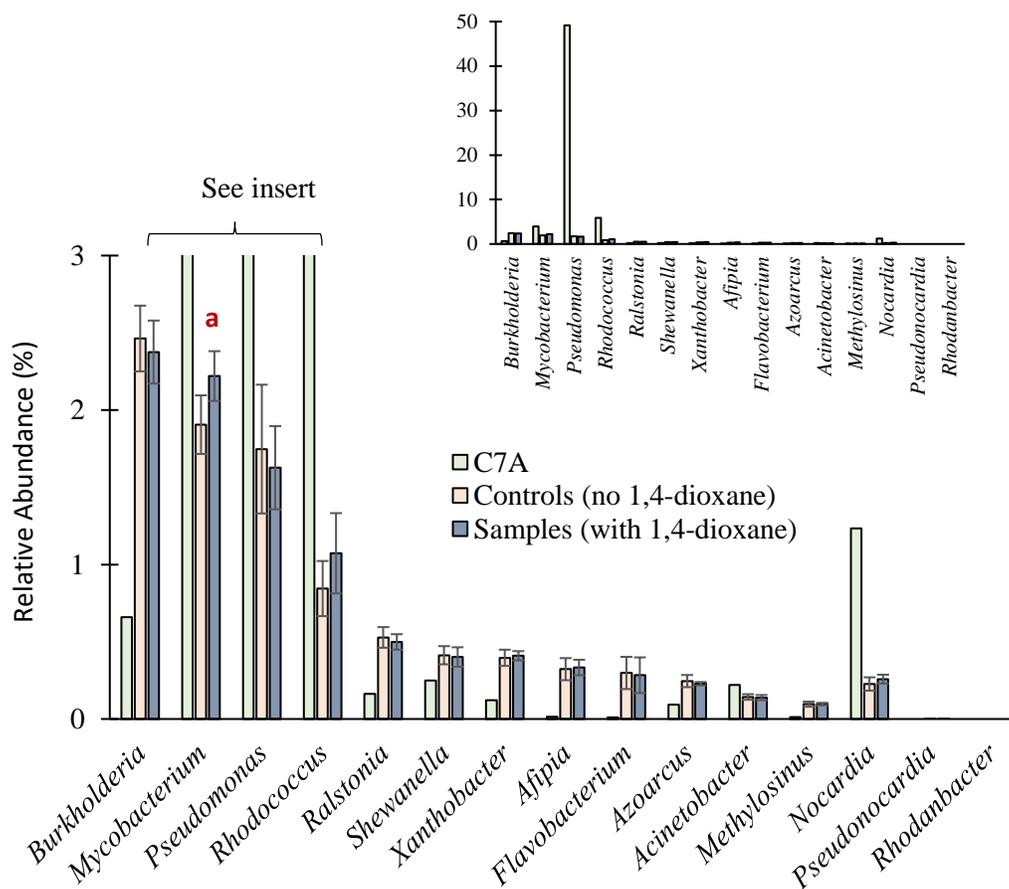


Figure 2.6. Relative abundance (%) of genera associated with metabolic and co-metabolic degradation of 1,4-dioxane in live controls (n=8) and samples (n=9) in four soils and one contaminated site sample (C7A). The value "a" indicates a significant difference (p<0.05) in a two tailed student's t-test between the samples and controls. The insert illustrates the same data with a different y-axis.

#### 2.4.5 Genes Associated with 1,4-Dioxane Degradation

The reads aligning to the genes previously associated with 1,4-dioxane degradation were determined using DIAMOND and the data were analyzed using Excel. Only the reads with  $\geq 60\%$  identity for  $\geq 49$  amino acids were included in the analysis. Among the twelve genes previously associated with 1,4-dioxane degradation, the majority were present in all the samples including the contaminated site sample (Figure 2.7).

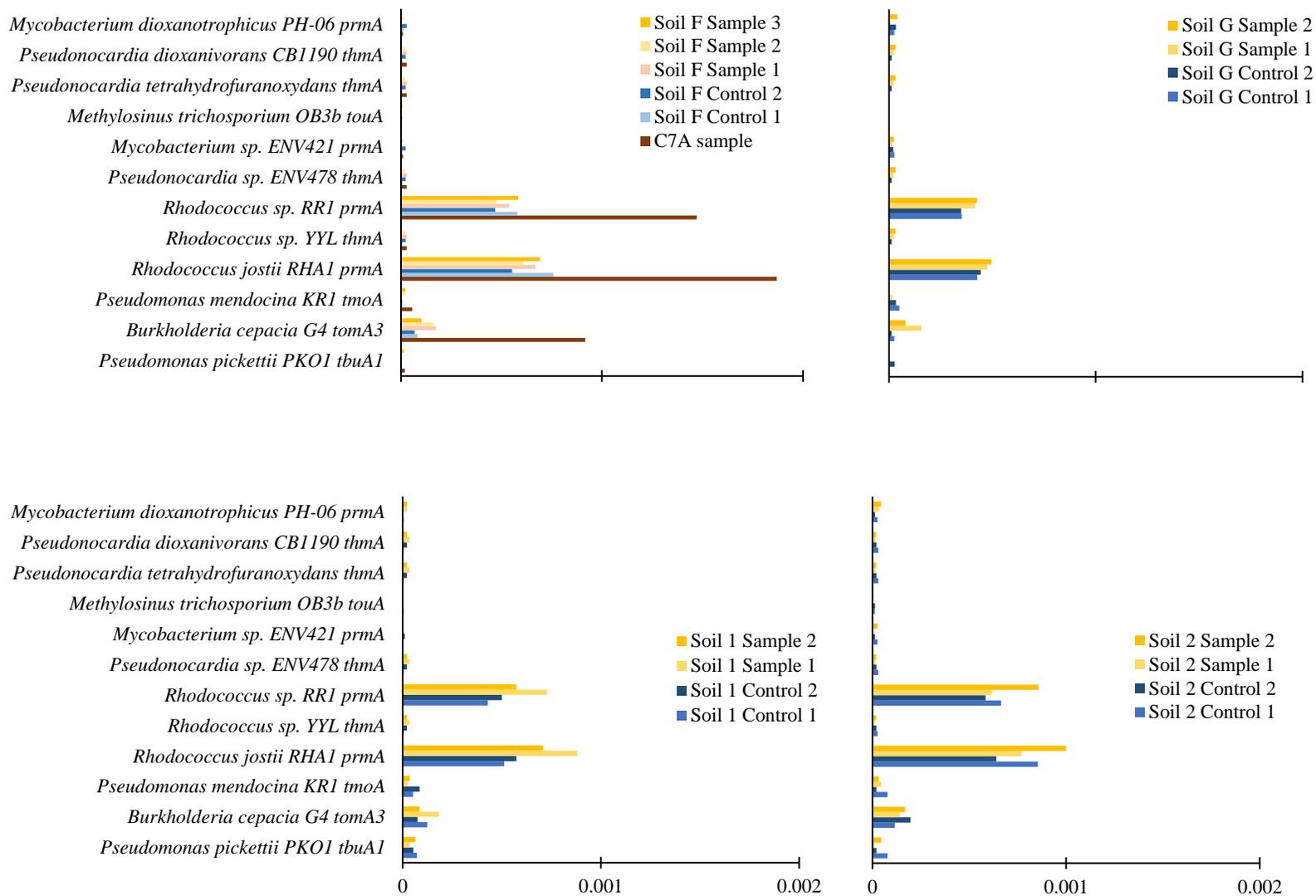


Figure 2.7. Relative abundance (%) of reads aligning ( $\geq 60\%$  identity for  $\geq 49$  amino acids) to genes previously associated with the metabolic and co-metabolic degradation of 1,4- dioxane in Soil F and C7A (A), Soil G (B), Soil 1 (C) and Soil 2 (D).

The analysis demonstrates a uniform trend of higher relative abundance values for *Rhodococcus* sp. RR1 *prmA* and *Rhodococcus jostii* RHA1 *prmA* in all four agricultural samples and the contaminated site sample compared to the other genes. The contaminated site sample demonstrates more than twice the relative abundance of these two genes compared to the samples from the agricultural sites. A high relative abundance of *Burkholderia cepacia* G4 *tomA3* was also noted in all of the samples, with higher abundance (10-fold increase) in the contaminated site sample compared to other agricultural samples (Figure 2.7). Only six and eleven metagenomes contained reads aligning with *Methylosinus trichosporium* OB3b *touA* and *Pseudomonas pickettii* PKO1 *tbuA1*, respectively. Seven functional genes (*Pseudomonas mendocina* KR1 *tmoA*, *Rhodococcus* sp. YYL *thmA*, *Pseudonocardia* sp. ENV478 *thmA*, *Mycobacterium* sp. ENV421 *prmA*, *Pseudonocardia tetrahydrofuranoxydans* *thmA*, *Pseudonocardia dioxanivorans* CB1190 *thmA*, *Mycobacterium dioxanotrophicus* PH-06 *prmA*) were present in between fourteen and eighteen metagenomes. All four soils generated similar trends for the functional genes and no statistically significant differences were noted between the live controls and samples. The contaminated site sample generated the same trend for the three most abundant genes.

Following the discovery of the dominance of *Rhodococcus jostii* RHA1 *prmA* and *Rhodococcus* sp. RR1 *prmA* in the soil metagenomes, a BLASTP search was performed to find the closest matching sequences in the NCBI database. The matching protein sequences, with number of microorganisms shown in parenthesis, belonged to the genera *Rhodococcus* (60), *Kribbella* (16), *Gordonia* (10), *Mycolicibacterium* (10), *Mycobacterium* (8), *Nocardia* (7), *Nocardioides* (6), *Hoyosella* (3), *Intrasporangium* (2), *Millisia* (1), *Cryptosporangium* (1) and *Acidobacteria* (1).

Interestingly, five of these genera (*Mycobacterium*, *Nocardioides*, *Kribbella*, *Gordonia* and *Nocardia*) were enriched in the samples compared to the live controls (as discussed above, Table 2.2). A phylogenetic tree was generated to illustrate the evolutionary relationships between the two query sequences and the enriched genera (Figure 2.8). *Rhodococcus jostii* RHA1 *prmA* clustered closest to *Nocardia* sequences and *Rhodococcus* sp. RR1 *prmA* clustered closest to *Kribbella* sequences.

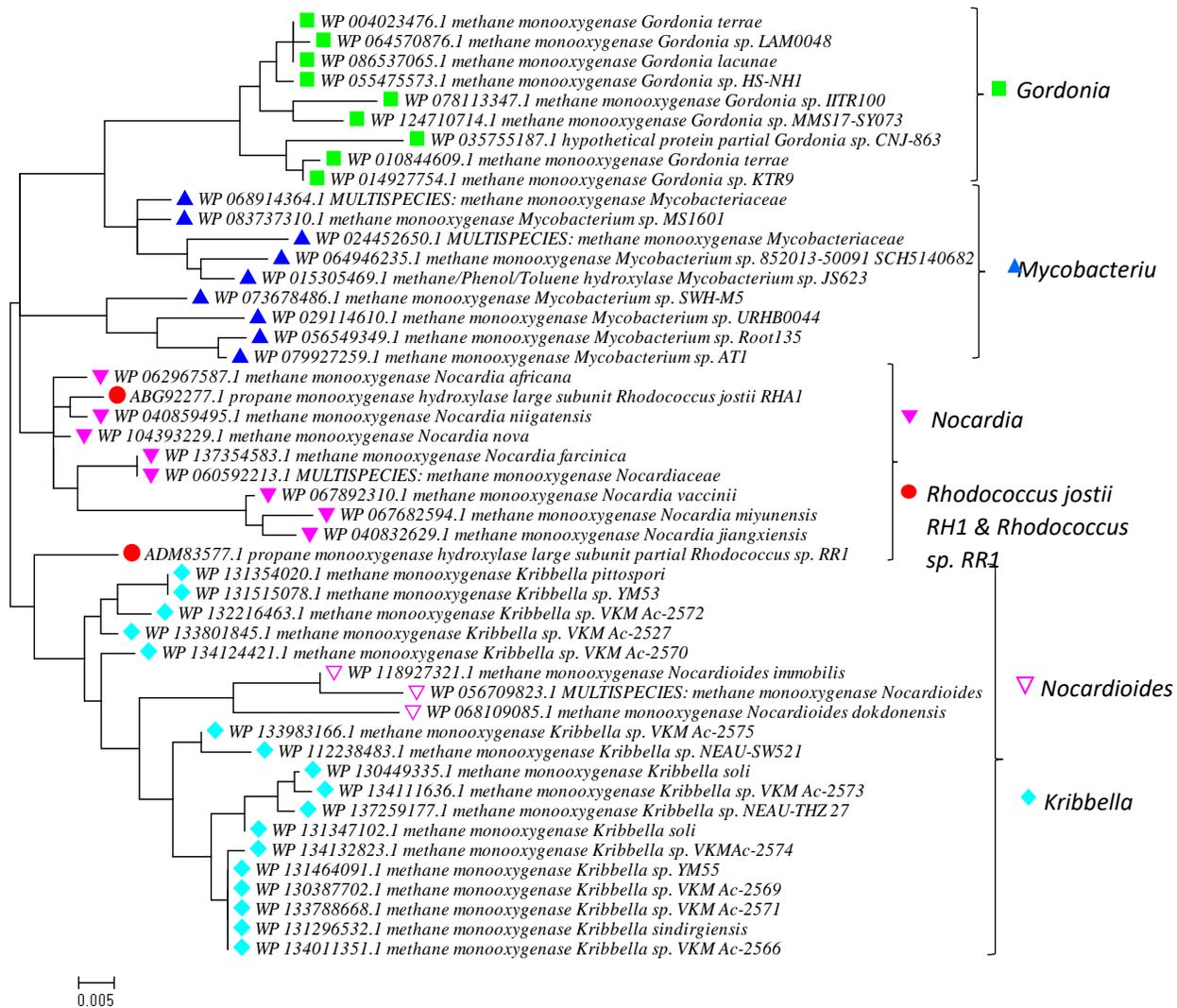


Figure 2.8. Phylogenetic tree of *Rhodococcus jostii* RHA1 *prmA* and *Rhodococcus* sp. *prmA* and BLASTP results (>94.8% similar to the two query sequences). Only genera that were enriched following 1,4-dioxane degradation (compared to the controls) are shown (Table 2). The evolutionary history was inferred by using the Maximum Likelihood method based on the Jones-Taylor-Thornton (JTT) matrix-based model. The tree with the highest log likelihood (-2731.06) is shown. 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 a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 48 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 439 positions in the final dataset. Evolutionary analyses were conducted in MEGA7

## 2.5 Discussion

The remediation of sites contaminated with 1,4-dioxane is challenging because of the physical and chemical properties of this chemical (Mohr et al. 2010a). Although bioremediation can be a viable option at some sites, it is still unclear which microorganisms and functional genes are linked to 1,4-dioxane degradation in mixed communities.

To address this knowledge gap, the current study identified which genera could obtain a growth benefit from 1,4-dioxane biodegradation. For this, the sample microcosms were supplied with media and 1,4-dioxane and the live control microcosms were supplied with the same media, but no 1,4-dioxane. Consequently, an increase in the relative abundance of any microorganism between the samples and live controls could be attributed to the presence of 1,4-dioxane. From this, a reasonable hypothesis would be that the enriched microorganisms are being exposed to growth supporting substrates from 1,4-dioxane degradation. Here, nineteen genera increased in abundance following 1,4-dioxane degradation compared to the live controls (no 1,4-dioxane). The three most enriched across all four soils were *Mycobacterium*, *Nocardioides*, *Kribbella* (all classifying as *Actinomycetales*). There was also a higher level of enrichment for *Arthrobacter*, *Nocardia* and *Gordonia* (*Actinomycetales*), *Hyphomicrobium* (*Rhizobiales*), *Clavibacter* (*Actinomycetales*) and *Bartonella* (*Rhizobiales*) and *Chelativorans* (*Rhizobiales*) in individual soils.

There are at least two hypotheses on why these genera increased in abundance in 1,4-dioxane amended samples compared to the live controls. One hypothesis being that these microorganisms are obtaining a growth benefit from consuming 1,4-dioxane biodegradation products. Several

studies have examined 1,4-dioxane biodegradation pathways (Grostern et al. 2012; Huang et al. 2014a; Kim et al. 2009; Mahendra et al. 2007; Sales et al. 2013; Vainberg et al. 2006). A study with *Pseudonocardia dioxanivorans* CB1190 provided evidence that carbon from 1,4-dioxane enters central metabolism via glyoxlate (Grostern et al. 2012). In contrast, *Pseudonocardia* sp. strain ENV478 produces 2-hydroxyethoxyacetic acid (HEAA) as a terminal product of 1,4-dioxane biodegradation (Vainberg et al. 2006). Conversely, 1,4-dioxane biodegradation by *Pseudonocardia dioxanivorans* CB1190 (metabolic 1,4-dioxane degrader), *Mycobacterium vaccae* JOB5, *Pseudomonas mendocina* KR1, *Pseudonocardia tetrahydrofuranoxydans* K1 (co-metabolic 1,4-dioxane degraders) produced HEAA transiently, but the chemical did not accumulate. They identified ethylene glycol, glycolic acid, glyoxylic acid and oxalic acid as 1,4-dioxane biodegradation intermediates by these isolates (Mahendra et al. 2007). Others have also identified ethylene glycol (Huang et al. 2014a; Kim et al. 2009), oxalic acid (Huang et al. 2014a) and ethane-1,2-diol (Kim et al. 2009) during 1,4-dioxane degradation. The enriched genera may have benefited from funneling these degradation intermediates into central metabolism.

A second hypothesis being that the enriched genera are responsible for both the initial attack on 1,4-dioxane and for the consumption of degradation products. Evidence for this concerns the similarity of genes belonging to the enriched genera (*Mycobacterium*, *Nocardioides*, *Kribbella*, *Nocardia* and *Gordonia*) to *Rhodococcus jostii* RHA1 *prmA* and *Rhodococcus* sp. RR1 *prmA* (as shown in the phylogenetic tree). Although *Rhodococcus jostii* RHA1 and *Rhodococcus* sp. RR1 co-metabolically degrade 1,4-dioxane, the enriched genera may also contain genes downstream in the pathway enabling growth on 1,4-dioxane. *Arthrobacter* did not contain genes similar to the *Rhodococcus* strains, although others have reported that *Arthrobacter* (ATCC 27779) can co-metabolically degrade 1,4-dioxane (Chu et al. 2009). *Arthrobacter*, *Mycobacterium* and

*Nocardia* have previously been linked to 1,4-dioxane degradation (Chu et al. 2009; Lan et al. 2013; Masuda 2009), whereas *Nocardioides*, *Gordonia* and *Kribbella* are potentially novel degraders. Certain species of *Gordonia* such as *G. terrae* are known to aid in degrading certain chemicals, including ethyl tertiary butyl ether (ETBE) metabolically, methyl tertiary butyl ether (MTBE) co-metabolically (Hernandez-Perez et al. 2001) as well as long chain hydrocarbons (Kubota et al. 2008). Overall, both hypotheses in this work suggests many genera (almost all classifying with the *Actinomycetales*) are likely involved in the degradation of 1,4-dioxane and/or 1,4-dioxane metabolites in the soil microcosms studied.

In the current study, reads from all of the 1,4-dioxane degrading function genes were observed in soil metagenomes. Consistent with the current study, others have detected SDIMOs from the majority (five from six groups) of SDIMO groups (Li et al. 2013b). In that research, the authors examined Arctic groundwater impacted by 1,4-dioxane using high-throughput microarrays and denaturing gradient gel electrophoresis and found an enrichment of *thmA*-like genes near the source zone (Li et al. 2013b). Also similar to the current work, a 1,4-dioxane degrading consortia contained a high percentage of group five SDIMOs (*Rhodococcus jostii* RHA1 *prmA* and *Rhodococcus* sp. RR1 *prmA* are group five SDIMOs), although the specific genes were not determined (He et al. 2018). Another study noted a correlation between *dxmA/thmA* (designed based on *Rhodococcus* sp. YYL *thmA*, *Pseudonocardia* sp. ENV478 *thmA*, *Pseudonocardia tetrahydrofuranoxydans* K1 *thmA* and *Pseudonocardia dioxanivorans* CB1190 *thmA*) and the amount of 1,4-dioxane degraded in groundwater inoculated microcosms (Li et al. 2013a). These genes were also present in the soil metagenomes (between fourteen and sixteen) of the current study.

Recently, shotgun sequencing was used to examine 1,4-dioxane degrading genes in groundwater from multiple chlorinated solvent sites (previously bioaugmented with SDC-9) (Dang et al. 2018). From the twelve genes examined, only six were found in the groundwater metagenomes. The six included the three most abundant genes in the current study; *Rhodococcus sp.* RR1 *prmA*, *Rhodococcus jostii* RHA1 *prmA* and *Burkholderia cepacia* G4 *tomA3*. From these, the *Rhodococcus* genes were both found in a only small number of metagenomes (~18%) and *B. cepacia* G4 *tomA3* was found in the majority (~68%). The occurrence of the three genes in both studies could suggest their importance across different environments (soil vs. groundwater, aerobic vs. oxygen depleted). Unlike the current study, the groundwater metagenomes contained high relative abundance values for *Methylosinus trichosporium* OB3b *touA* (up to 0.0031%) followed by *Pseudomonas mendocina* KR1 *tmoA* (up to 0.00022%) and *Pseudomonas pickettii* PKO1 *tbuA1* (up to 0.0013%). The different results between the two studies are likely due to variations in the conditions (redox potential, carbon availability, nutrient availability, soil vs. groundwater) from which the samples were obtained.

In summary, several key findings highly relevant for 1,4-dioxane bioremediation were generated here. Shotgun sequencing enabled both taxonomic and functional analyses to be performed on multiple mixed microbial communities. Multiple genera classifying (including novel and previously identified degraders) within the *Actinomycetales* were enriched during 1,4-dioxane degradation and may be associated with growth linked 1,4-dioxane degradation.

The three most enriched were *Mycobacterium*, *Nocardioides*, *Kribbella* (classifying as *Actinomycetales*). There was also a higher level of enrichment of other genera in individual soils. The current research found that both previously reported genera as well as novel genera (e.g.

*Nocardioides*, *Gordonia* and *Kribbella*) were linked to 1,4-dioxane degradation. However, it is unknown if these microorganisms are benefiting from the complete degradation of the chemical or from the consumption of 1,4-dioxane degradation products, such as HEAA, ethylene glycol, glycolic acid, glyoxylic acid or oxalic acid. Finally, all of the functional genes associated with 1,4-dioxane were found in the soil and sediment metagenomes. Reads aligning to *Rhodococcus jostii* RHA1 *prmA* and *Rhodococcus* sp. RR1 *prmA* illustrated the highest relative abundance values and were present in all eighteen metagenomes. Future research should be directed towards similar molecular analyses of groundwater and sediment samples from 1,4-dioxane contaminated sites as well as comparisons to 1,4-dioxane removal rates for propane amended samples.

## APPENDIX

# APPENDIX

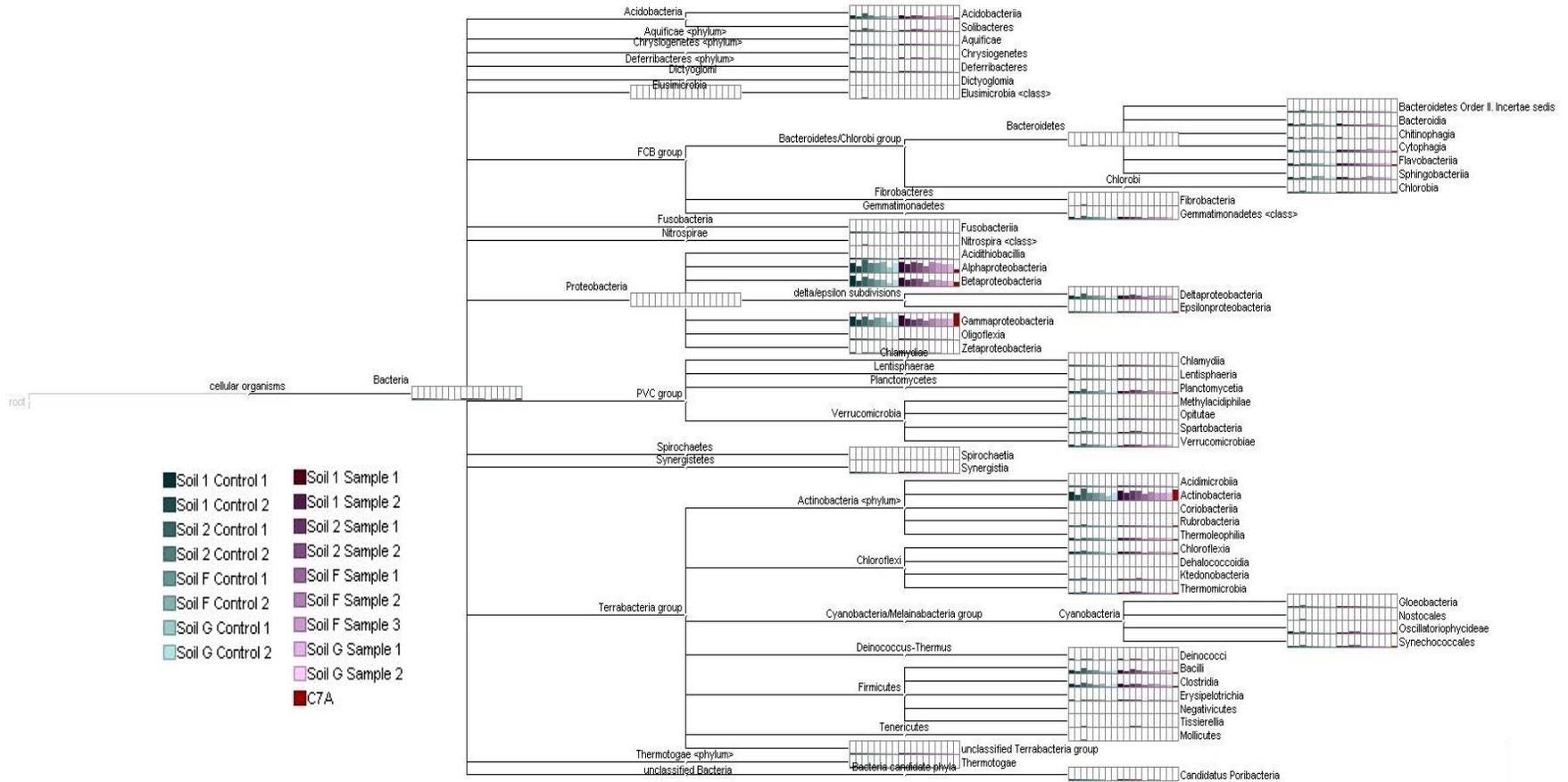


Figure 2.9. Phylogram (created with MEGAN6, version 6.11.7) illustrating the relative abundance and classification (Class Level) of all bacteria across all metagenomes (samples and controls).

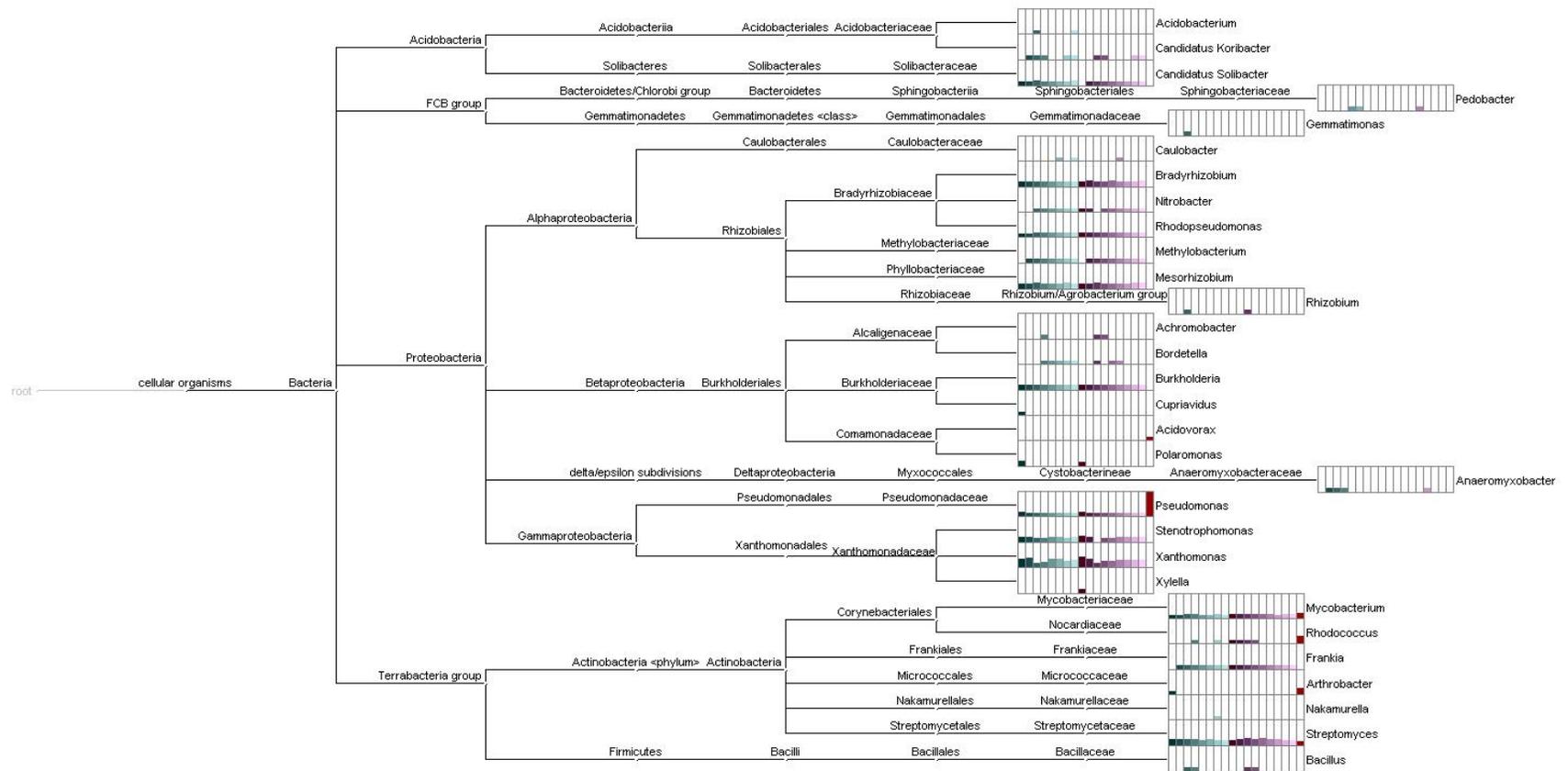


Figure 2.10. Phylogram (created with MEGAN6, version 6.11.7) illustrating the most abundant genera (ranked by average relative abundance, then selected if average relative abundance >0.5%) across all metagenomes (samples and controls).

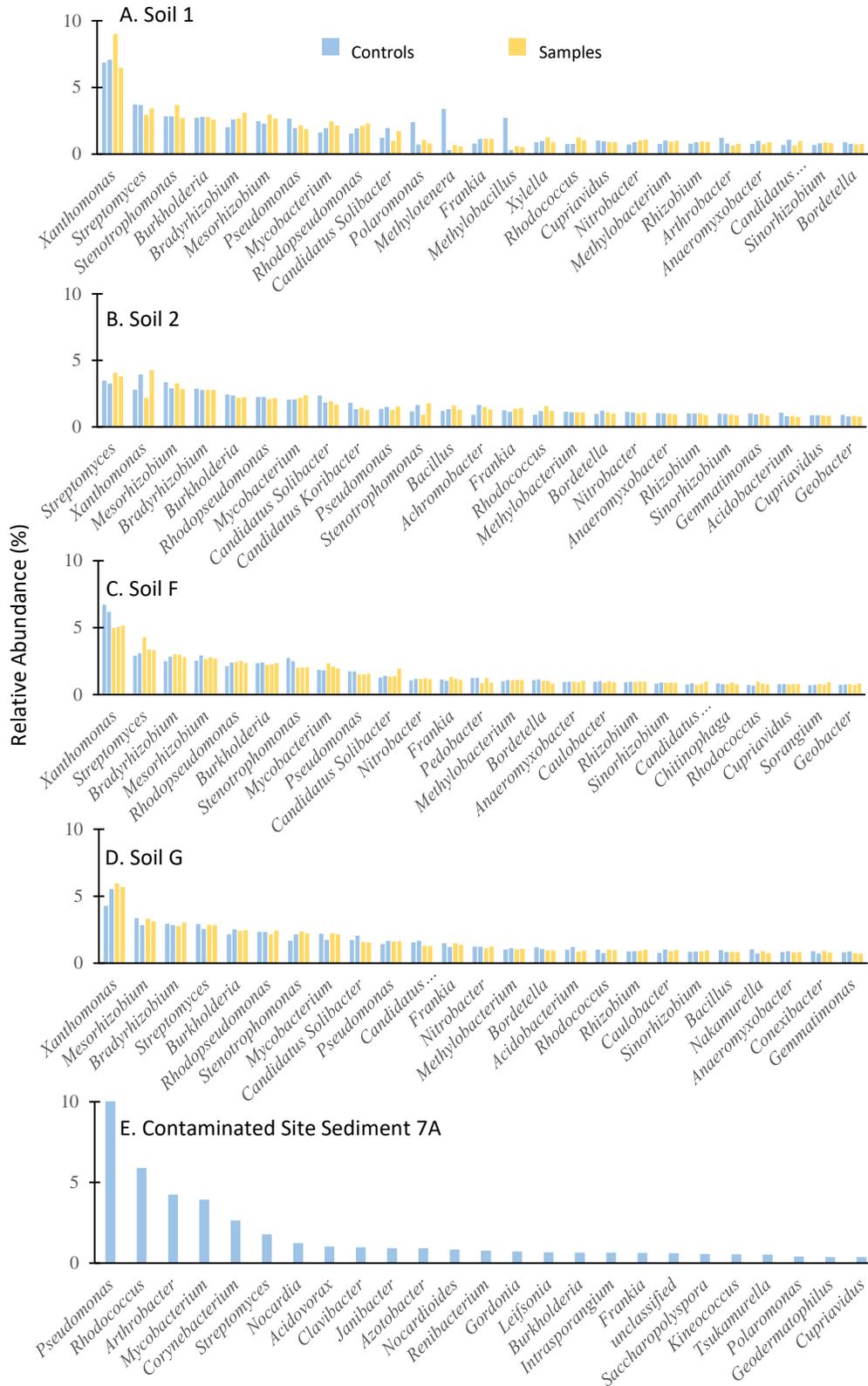


Figure 2.11. The twenty-five most common genera (by relative abundance, %), ranked by the averages of the samples and controls, in soil 1 (A), soil 2 (B), soil F (C), soil G (D) and the contaminated site sediment 7A (E).

## REFERENCES

## REFERENCES

- Adams CD, Scanlan PA, Secrist ND (1994) Oxidation and biodegradability enhancement of 1,4-dioxane using hydrogen peroxide and ozone. *Environ Sci Technol* 28(11):1812-1818
- Adamson DT, Anderson RH, Mahendra S, Newell CJ (2015) Evidence of 1,4-dioxane attenuation at groundwater sites contaminated with chlorinated solvents and 1,4-dioxane. *Environ Sci Technol* 49(11):6510-6518
- Adamson DT, Mahendra S, Walker KL, Rauch SR, Sengupta S, Newell CJ (2014) A multisite survey to identify the scale of the 1,4-dioxane problem at contaminated groundwater sites. *Environ Sci Tech Let* 1(5):254-258
- Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389-3402 doi:DOI 10.1093/nar/25.17.3389
- ATSDR AFTSaDR (2012) Toxicological profile for 1,4-dioxane. In: Registry AFTSaD (ed). Agency for Toxic Substances and Disease Registry. <https://www.atsdr.cdc.gov/toxprofiles/TP.asp?id=955&tid=199>, Atlanta, Georgia
- Bock C, Kroppenstedt RM, Diekmann H (1996) Degradation and bioconversion of aliphatic and aromatic hydrocarbons by *Rhodococcus ruber* 219. *Appl Microbiol Biotechnol* 45(3):408-410 doi:10.1007/s002530050704
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114-2120 doi:10.1093/bioinformatics/btu170
- Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12(1):59-60
- Burback BL, Perry JJ (1993) Biodegradation and biotransformation of groundwater pollutant mixtures by *Mycobacterium vaccae*. *Appl Environ Microb* 59(4):1025-1029
- Chen D-Z, Jin X-J, Chen J, Ye J-X, Jiang N-X, Chen J-M (2016) Intermediates and substrate interaction of 1,4-dioxane degradation by the effective metabolizer *Xanthobacter flavus* DT8. *Inter Biodeter & Biodeg* 106:133-140 doi:10.1016/j.ibiod.2015.09.018
- Chu MYJ, Bennett P, Dolan M, Hyman M, Anderson R, Bodour A, Peacock A (2009) Using aerobic cometabolic biodegradation and groundwater recirculation to treat 1,4-dioxane and co-contaminants in a dilute plume. Paper presented at the Tenth International Conference on the Remediation of Chlorinated and Recalcitrant Compounds, Baltimore, Maryland,
- Coleman NV, Bui NB, Holmes AJ (2006) Soluble di-iron monooxygenase gene diversity in soils, sediments and ethene enrichments. *8(7):1228-1239* doi:10.1111/j.1462-2920.2006.01015.x

- Cox MP, Peterson DA, Biggs PJ (2010) SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinform* 11:485:1-6
- Dang HY, Kanitkar YH, Stedtfeld RD, Hatzinger PB, Hashsham SA, Cupples AM (2018) Abundance of chlorinated solvent and 1,4-dioxane degrading microorganisms at five chlorinated solvent contaminated sites determined via shotgun sequencing. *Environ Sci Technol* 52(23):13914-13924 doi:10.1021/acs.est.8b04895
- Deng DY, Li F, Li MY (2018) A novel propane monooxygenase initiating degradation of 1,4-dioxane by *Mycobacterium dioxanotrophicus* PH-06. *Environ Sci Tech Lett* 5(2):86-91 doi:10.1021/acs.estlett.7b00504
- DeRosa CT, Wilbur S, Holler J, Richter P, Stevens YW (1996) Health evaluation of 1,4-dioxane. *Toxicol & Indust Health* 12(1):1-43
- EPA (2017) Technical fact sheet - 1,4-dioxane. vol EPA 505-F-17-011. EPA Office of Land and Emergency Management
- Fishman A, Tao Y, Wood TK (2004) Toluene 3-monooxygenase of *Ralstonia pickettii* PKO1 is a para-hydroxylating enzyme. *J Bacteriol* 186(10):3117-3123 doi:10.1128/Jb.186.12.3117-3123.2004
- Gedalanga PB, Pornwongthong P, Mora R, Chiang SYD, Baldwin B, Ogles D, Mahendra S (2014) Identification of biomarker genes To predict biodegradation of 1,4-dioxane. *Appl Environ Microb* 80(10):3209-3218 doi:10.1128/Abm.04162-1
- Grosterm A, Sales CM, Zhuang WQ, Erbilgin O, Alvarez-Cohen L (2012) Glyoxylate metabolism is a key feature of the metabolic degradation of 1,4-dioxane by *Pseudonocardia dioxanivorans* strain CB1190. *Appl Environ Microb* 78(9):3298-3308
- Hand S, Wang BX, Chu KH (2015) Biodegradation of 1,4-dioxane: Effects of enzyme inducers and trichloroethylene. *Sci Total Environ* 520:154-159
- He Y, Mathieu J, da Silva MLB, Li MY, Alvarez PJJ (2018) 1,4-Dioxane-degrading consortia can be enriched from uncontaminated soils: prevalence of *Mycobacterium* and soluble di-iron monooxygenase genes. *Microb Biotechnol* 11(1):189-198 doi:10.1111/1751-7915.12850
- He Y, Mathieu J, Yang Y, Yu P, da Silva MLB, Alvarez PJJ (2017) 1,4-Dioxane biodegradation by *Mycobacterium dioxanotrophicus* PH-06 is associated with a Group-6 soluble di-Iron monooxygenase. *Environ Sci & Tec Lett* 4(11):494-499 doi:10.1021/acs.estlett.7b00456
- Hernandez-Perez G, Fayolle F, Vandecasteele J-P (2001) Biodegradation of ethyl t-butyl ether (ETBE), methyl t-butyl ether (MTBE) and t-amyl methyl ether (TAME) by *Gordonia terrae*. *Appl Micro & Biotec* 55(1):117-121 doi:10.1007/s002530000482
- Huang H, Shen D, Li N, Shan D, Shentu J, Zhou Y (2014) Biodegradation of 1,4-dioxane by a novel strain and its biodegradation pathway. *Water Air Soil Poll* 225(9) doi:10.1007/s11270-014-2135-2

- Huson DH, Beier S, Flade I, Gorska A, El-Hadidi M, Mitra S, Ruscheweyh HJ, Tappu R (2016) MEGAN Community Edition - Interactive exploration and analysis of large-scale microbiome sequencing data. *Plos Comput Biol* 12(6) doi:ARTN e1004957 10.1371/journal.pcbi.1004957
- IARC (1999) Re-evaluation of some organic chemicals, hydrazine and hydrogen peroxide. Proceedings of the IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, France, 17-24 February 1998. 71 Pt 1:1-315
- Inoue D, Tsunoda T, Sawada K, Yamamoto N, Saito Y (2016) 1,4-Dioxane degradation potential of members of the genera *Pseudonocardia* and *Rhodococcus*. *Biodegrad* 27(4-6):277-286 doi:10.1007/s10532-016-9772-7
- Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8(3):275-282
- Kampfer P, Kohlweyer U, Thiemer B, Andreesen JR (2006) *Pseudonocardia tetrahydrofuranoxydans* sp nov. *Int J Syst Evol Micr* 56:1535-1538 doi:10.1099/ijs.0.64199-0
- Kämpfer P, Kroppenstedt RM (2004) *Pseudonocardia benzenivorans* sp. nov. *Int J Sys Evol Micro* 54(3):749-751 doi:10.1099/ijs.0.02825-0
- Kim Y-M, Jeon J-R, Murugesan K, Kim E-J, Chang Y-S (2008) Biodegradation of 1,4-dioxane and transformation of related cyclic compounds by a newly isolated *Mycobacterium* sp. PH-06. *Biodegrad* 20(4):511 doi:10.1007/s10532-008-9240-0
- Kim YM, Jeon JR, Murugesan K, Kim EJ, Chang YS (2009) Biodegradation of 1,4-dioxane and transformation of related cyclic compounds by a newly isolated *Mycobacterium* sp PH-06. *Biodegrad* 20(4):511-519
- Kohlweyer U, Thiemer B, Schrader T, Andreesen JR (2000) Tetrahydrofuran degradation by a newly isolated culture of *Pseudonocardia* sp strain K1. *Fems Microbiol Lett* 186(2):301-306 doi:DOI 10.1111/j.1574-6968.2000.tb09121.x
- Kubota K, Koma D, Matsumiya Y, Chung SY, Kubo M (2008) Phylogenetic analysis of long-chain hydrocarbon-degrading bacteria and evaluation of their hydrocarbon-degradation by the 2,6-DCPIP assay. *Biodegrad* 19(5):749-57 doi:10.1007/s10532-008-9179-1
- Kukor JJ, Olsen RH (1990) Molecular cloning, characterization, and regulation of a *Pseudomonas pickettii* PKO1 gene encoding phenol hydroxylase and expression of the gene in *Pseudomonas aeruginosa* PAO1c. *J Bac* doi:10.1128/jb.172.8.4624-4630.1990
- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33(7):1870-1874 doi:10.1093/molbev/msw054
- Lan RS, Smith CA, Hyman MR (2013) Oxidation of cyclic ethers by alkane-grown *Mycobacterium vaccae* JOB5. *Remediation* 23(4):23-42 doi:10.1002/rem.21364

Li M, Mathieu J, Liu Y, Van Orden E.T., Yang Y, Fiorenza S, Alvarez PJ (2013a) The abundance of tetrahydrofuran/dioxane monooxygenase genes (*thmA/dxmA*) and 1,4-dioxane degradation activity are significantly correlated at various impacted aquifers. *Env Sci Tech Lett* 1:122-127

Li MY, Mathieu J, Yang Y, Fiorenza S, Deng Y, He ZL, Zhou JZ, Alvarez PJJ (2013b) Widespread distribution of soluble di-iron monooxygenase (SDIMO) genes in Arctic groundwater impacted by 1,4-dioxane. *Environ Sci Technol* 47(17):9950-9958 doi:10.1021/es402228x

Mahendra S, Alvarez-Cohen L (2006) Kinetics of 1,4-dioxane biodegradation by monooxygenase-expressing bacteria. *Environ Sci Technol* 40(17):5435-5442 doi:10.1021/es060714v

Mahendra S, Petzold CJ, Baidoo EE, Keasling JD, Alvarez-Cohen L (2007) Identification of the intermediates of in vivo oxidation of 1,4-dioxane by monooxygenase-containing bacteria. *Environ Sci Technol* 41(21):7330-7336 doi:10.1021/es0705745

Masuda H (2009) Identification and characterization of monooxygenase enzymes involved in 1,4-dioxane degradation in *Pseudonocardia* sp. strain ENV478, *Mycobacterium* sp. strain ENV421, and *Nocardia* sp. strain ENV425. Rutgers The State University of New Jersey and University of Medicine and Dentistry of New Jersey

Masuda H, McClay K, Steffan RJ, Zylstra GJ (2012a) Biodegradation of tetrahydrofuran and 1,4-dioxane by soluble diiron monooxygenase in *Pseudonocardia* sp. strain ENV478. *J Mol Microb Biotech* 22(5):312-316

Masuda H, McClay K, Steffan RJ, Zylstra GJ (2012b) Characterization of three propane-inducible oxygenases in *Mycobacterium* sp. strain ENV421. *Lett in App Micro* 55(3):175-181 doi:10.1111/j.1472-765x.2012.03290.x

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9 doi:Artn 386 10.1186/1471-2105-9-386

Mohr T, Stickney J, Diguseppi B (2010) Environmental investigation and remediation : 1,4-dioxane and other solvent stabilizers. CRC Press/Taylor & Francis, Boca Raton, FL

Nelson MJ, Montgomery SO, O'Neill EJ, Pritchard PH (1986) Aerobic metabolism of trichloroethylene by a bacterial isolate. *Appl Environ Microbiol* 52(2):383-384

Newman LM, Wackett LP (1995) Purification and characterization of toluene 2-monooxygenase from *Burkholderia cepacia* G4. *Biochemistry-Us* 34(43):14066-14076 doi:DOI 10.1021/bi00043a012

- Oldenhuis R, Vink RLJM, Janssen DB, Witholt B (1989) Degradation of chlorinated aliphatic hydrocarbons by *Methylosinus trichosporium* OB3b expressing soluble methane monooxygenase. *Appl Environ Microb* 55(11):2819-2826
- Parales RE, Adamus JE, White N, May HD (1994) Degradation of 1,4-dioxane by an actinomycete in pure culture. *Appl Environ Microb* 60(12):4527-4530
- Parks DH, Tyson GW, Hugenholtz P, Beiko RG (2014) STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30(21):3123-3124
- Pruitt KD, Tatusova T, Maglott DR (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 33:D501-D504
- Pugazhendi A, Rajesh Banu J, Dhavamani J, Yeom IT (2015) Biodegradation of 1,4-dioxane by *Rhodanobacter* AYS5 and the role of additional substrates. *Annal of Micro* 65(4):2201-2208 doi:10.1007/s13213-015-1060-y
- Rho MN, Tang HX, Ye YZ (2010) FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res* 38(20)
- Sales CM, Grostern A, Parales JV, Parales RE, Alvarez-Cohen L (2013) Oxidation of the cyclic ethers 1,4-dioxane and tetrahydrofuran by a monooxygenase in two *Pseudonocardia* species. *Appl Environ Microb* 79(24):7702-7708 doi:10.1128/Aem.02418-13
- Sales CM, Mahendra S, Grostern A, Parales RE, Goodwin LA, Woyke T, Nolan M, Lapidus A, Chertkov O, Ovchinnikova G, Sczyrba A, Alvarez-Cohen L (2011) Genome sequence of the 1,4-dioxane-degrading *Pseudonocardia dioxanivorans* Strain CB1190. *J Bacteriol* 193(17):4549-4550 doi:10.1128/Jb.00415-11
- Sei K, Miyagaki K, Kakinoki T, Fukugasako K, Inoue D, Ike M (2013) Isolation and characterization of bacterial strains that have high ability to degrade 1,4-dioxane as a sole carbon and energy source. *Biodegrad* 24(5):665-674
- Sharp JO, Sales CM, LeBlanc JC, Liu J, Wood TK, Eltis LD, Mohn WW, Alvarez-Cohen L (2007) An inducible propane monooxygenase is responsible for N-nitrosodimethylamine degradation by *Rhodococcus* sp strain RHA1. *Appl Environ Microb* 73(21):6930-6938 doi:10.1128/Aem.01697-07
- Stefan MI, Bolton JR (1998) Mechanism of the degradation of 1,4-dioxane in dilute aqueous solution using the UV hydrogen peroxide process. *Env Sci Tech* 32(11):1588-1595
- 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(11):4216-4222
- Steffan RJ, McClay KR, Masuda H, Zylstra GJ (2007) Biodegradation of 1,4-dioxane. Strategic Environmental Research and Development Program: ER-1422 Final Report

Stringfellow WT, Alvarez-Cohen L (1999) Evaluating the relationship between the sorption of PAHs to bacterial biomass and biodegradation. *Wat Res* 33(11):2535-2544 doi:10.1016/s0043-1354(98)00497-7

Sun B, Ko K, Ramsay JA (2011) Biodegradation of 1,4-dioxane by a *Flavobacterium*. *Biodegrad* 22(3):651-659 doi:10.1007/s10532-010-9438-9

Thiemer B, Andreesen JR, Schrader T (2003) Cloning and characterization of a gene cluster involved in tetrahydrofuran degradation in *Pseudonocardia* sp strain K1. *Arch Microbiol* 179(4):266-277 doi:10.1007/s00203-003-0526-7

USEPA (2017) Technical Fact Sheet 1,4-Dioxane. In: Management USEOoLaE (ed).

Vainberg S, McClay K, Masuda H, Root D, Condee C, Zylstra GJ, Steffan RJ (2006) Biodegradation of ether pollutants by *Pseudonocardia* sp strain ENV478. *Appl Environ Microb* 72(8):5218-5224

Whited GM, Gibson DT (1991) Separation and partial characterization of the enzymes of the toluene-4-monooxygenase catabolic pathway in *Pseudomonas mendocina* KR1. *J Bacteriol* 173(9):3017-3020 doi:10.1128/jb.173.9.3017-3020.1991

Whittenbury R, Phillips KC, Wilkinson JF (1970) Enrichment, isolation and some properties of methane-utilizing bacteria. *J of Gen Micro* 61(2):205-218 doi:10.1099/00221287-61-2-205

Yao YL, Lv ZM, Min H, Lv ZH, Jiao HP (2009) Isolation, identification and characterization of a novel *Rhodococcus* sp strain in biodegradation of tetrahydrofuran and its medium optimization using sequential statistics-based experimental designs. *Bioresource Technol* 100(11):2762-2769 doi:10.1016/j.biortech.2009.01.006

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. *J Bacteriol* 173(17):5315-5327 doi:DOI 10.1128/jb.173.17.5315-5327.1991

Zenker MJ, Borden RC, Barlaz MA (2003) Occurrence and treatment of 1,4-dioxane in aqueous environments. *Env Eng Sci* 20(5):423-432

## Chapter 3 Anaerobic 1,4-Dioxane Biodegradation and Microbial Community Analysis in Microcosms Inoculated with Soils or Sediments and Different Electron Acceptors

### 3.1 Abstract

1,4-dioxane, a probable human carcinogen, is a co-contaminant at many chlorinated solvent contaminated sites. Although numerous 1,4-dioxane degrading aerobic bacteria have been isolated, almost no information exists on the microorganisms able to degrade this chemical under anaerobic conditions. Here, the potential for 1,4-dioxane biodegradation was examined using multiple inocula and electron acceptor amendments. The inocula included uncontaminated agricultural soils and river sediments as well as sediments from two 1,4-dioxane contaminated sites. Five separate experiments involved the examination of triplicate live microcosms and abiotic controls for approximately one year. Compound specific isotope analysis (CSIA) was used to further investigate biodegradation in a subset of the microcosms. Also, DNA was extracted from microcosms exhibiting 1,4-dioxane biodegradation for microbial community analysis using 16S rRNA gene amplicon high throughput sequencing.

Given the long incubation periods, it is likely that electron acceptor depletion occurred and methanogenic conditions eventually dominated. The iron/EDTA/humic acid or sulfate amendments did not result in 1,4-dioxane biodegradation in the majority of cases. 1,4-dioxane biodegradation was most commonly observed in the nitrate amended and no electron acceptor treatments. Notably, both contaminated site sediments illustrated removal in the samples compared to the abiotic controls in the no electron acceptor treatment. However, it is important to note that the degradation was slow (with concentration reductions occurring over approximately one year). In two of the three cases examined, CSIA provided additional evidence

for 1,4-dioxane biodegradation. In one case, the reduction in 1,4-dioxane in the samples compared the controls was likely too low for the method to detect a significant  $^{13}\text{C}/^{12}\text{C}$  enrichment. Further research is required to determine the value of measuring  $^2\text{H}/^1\text{H}$  for generating evidence for the biodegradation of this chemical.

The microbial community analysis indicated the phylotypes unclassified *Comamonadaceae* and 3 *genus incertae sedis* were more abundant in 1,4-dioxane degrading microcosms compared to the live controls (no 1,4-dioxane) in microcosms inoculated with contaminated and uncontaminated sediment, respectively. The relative abundance of known 1,4-dioxane degraders was also investigated at the genus level. The soil microcosms were dominated primarily by *Rhodanobacter* with lower relative abundance values for *Pseudomonas*, *Mycobacterium* and *Acinetobacter*. The sediment communities were dominated by *Pseudomonas* and *Rhodanobacter*. Overall, the current study indicates 1,4-dioxane biodegradation under anaerobic, and likely methanogenic conditions, is feasible. Therefore, natural attenuation may be an appropriate clean-up technology at sites where time is not a limitation.

### 3.2 Introduction

There is a critical need to develop remediation strategies for the contaminant 1,4-dioxane due to its widespread occurrence (Adamson et al. 2015) and its classification as a probable human carcinogen (DeRosa et al. 1996). Historically, 1,4-dioxane was used as a stabilizer in 1,1,1-trichloroethane (1,1,1-TCA) formulations and is now frequently detected at chlorinated solvent contaminated sites (Adamson et al. 2015; Adamson et al. 2014a; Mohr et al. 2010b). As 1,4-dioxane was typically not on the EPA's target compound lists, it is likely that closed sites will

require re-opening to address contamination. In fact, a multisite survey aimed at examining the extent of the 1,4-dioxane problem, indicated a primary risk is the large number of sites where this chemical is likely to be present but has yet to be identified (Adamson et al. 2014a).

The remediation of 1,4-dioxane contaminated sites is problematic because of chemical characteristics that result in movement and persistence (Adamson et al. 2015; Mohr et al. 2010b). Traditional methods such as air stripping or activated carbon can be ineffective because of the chemical's low organic carbon partition coefficient ( $\log K_{OC} = 1.23$ ) and low Henry's Law Constant ( $5 \times 10^{-6} \text{ atm. m}^3\text{mol}^{-1}$ ) (Mahendra and Alvarez-Cohen 2006; Steffan et al. 2007; Zenker et al. 2003). *Ex situ* oxidation methods such as ozone and hydrogen peroxide (Adams et al. 1994) or hydrogen peroxide and ultraviolet light (Stefan and Bolton 1998) have been commercially applied, although at high concentrations these methods can be cost-prohibitive (Steffan et al. 2007). Given the limitations associated with traditional remediation methods, interest has turned to the use of microorganisms to biodegrade this problematic contaminant (Chu et al. 2009; Li et al. 2010; Lippincott et al. 2015).

Numerous bacteria have been associated with the biodegradation (co-metabolic or metabolic) of 1,4-dioxane. Those linked to aerobic metabolic biodegradation classify within the genera *Pseudonocardia* (Kampfer and Kroppenstedt 2004; Mahendra and Alvarez-Cohen 2005; Mahendra and Alvarez-Cohen 2006; Matsui et al. 2016; Parales et al. 1994; Sei et al. 2013a; Yamamoto et al. 2018), *Mycobacterium* (Kim et al. 2009; Sei et al. 2013a), *Afipia* (Isaka et al. 2016; Sei et al. 2013a), *Xanthobacter* (Chen et al. 2016b), *Acinetobacter* (Huang et al. 2014b),

*Rhodococcus* (Bernhardt and Diekmann 1991; Inoue et al. 2016b; Inoue et al. 2018) and *Rhodanobacter* (Pugazhendi et al. 2015a).

Genera associated with aerobic co-metabolic 1,4-dioxane degradation include *Pseudonocardia* (Kohlweyer et al. 2000; Mahendra and Alvarez-Cohen 2006; Vainberg et al. 2006)], *Mycobacterium* (Lan et al. 2013; Masuda 2009) and *Rhodococcus* (Hand et al. 2015; Lippincott et al. 2015; Mahendra and Alvarez-Cohen 2006; Sei et al. 2013b; Steffan et al. 1997b; Stringfellow and Alvarez-Cohen 1999), *Flavobacterium* (Sun et al. 2011b), *Burkholderia* (Mahendra and Alvarez-Cohen 2006), *Nocardia* (Masuda 2009), *Ralstonia* (Mahendra and Alvarez-Cohen 2006), *Pseudomonas* (Mahendra and Alvarez-Cohen 2006), *Methylosinus* (Mahendra and Alvarez-Cohen 2006; Whittenbury et al. 1970a) and *Azoarcus* (Deng et al. 2018b). However, it is unlikely that these aerobic 1,4-dioxane degraders will be effective at chlorinated solvent sites, as these sites are typically highly reducing. Under such conditions, tetrachloroethene and trichloroethene undergo sequential reductive dechlorination to *cis*-1,2-dichloroethene and vinyl chloride, finally forming the non-toxic end product, ethene (Freedman and Gossett 1989). Reduction is commonly associated with microbial taxa such as *Dehalococcoides mccartyi* and *Dehalobacter* (Cupples 2008; Cupples et al. 2003; He et al. 2003; Holliger et al. 1998; Löffler et al. 2013; Maymó-Gatell et al. 1997; Sung et al. 2006). Commercially available reductive dechlorinating mixed cultures containing such strains are frequently used for bioaugmenting contaminated groundwater aquifers (Major et al. 2002; Steffan and Vainberg 2013; Vainberg et al. 2009). The lack of information on the susceptibility of the chlorinated solvent co-contaminant 1,4-dioxane to biodegradation under such highly reducing conditions is a significant knowledge gap.

No anaerobic 1,4-dioxane degrading isolates have been identified and, limited research has addressed 1,4-dioxane biodegradation under anaerobic conditions. One project investigated 1,4-dioxane degradation over a range of redox conditions (aerobic, nitrate reducing, iron reducing, sulfate reducing and methanogenic) (Steffan 2007). The work involved microcosm experiments with soil and groundwater from a site heavily contaminated with the chlorinated solvents and 1,4-dioxane. In these tests, nitrate, nitrite, sulfate and ferric iron were added as electron acceptors. In another set of experiments, samples from across a vegetable oil biobarrier were investigated, without the addition of electron acceptors, as it was expected that the biobarrier had resulted in a range of redox conditions. Notably, 1,4-dioxane was not degraded in any of the anaerobic microcosms during >400 days.

Another study produced more promising results, documenting 1,4-dioxane biodegradation under iron reducing conditions using an enrichment originating from wastewater treatment plant sludge (Shen et al. 2008). The researchers found that when Fe(III) was supplied as Fe(III)-EDTA, biodegradation was stimulated. Also, they reported that humic acids stimulated the activity of the Fe(III) reducing bacteria and 1,4-dioxane biodegradation. The authors hypothesized that humic acids promoted electron shuttling to Fe(III) in a catalytic manner. Others have also reported stimulation of bacterial growth and biodegradation (cyclic nitroamines RDX and HMX) in the presence of humic acids and Fe(III) (Bhushan et al. 2006).

The overall objective of the current study was to determine the susceptibility of 1,4-dioxane to biodegradation under anaerobic conditions using inocula from contaminated and uncontaminated sites. For this, microcosms were established with different inocula (soils or sediments) and

various electron acceptor amendments (nitrate, iron-EDTA/humic acid, sulfate and no amendment). Additionally, compound specific isotope analysis (CSIA) was used to determine changes in  $^{13}\text{C}/^{12}\text{C}$  ratios in a subset of the samples. The work is novel as it is the first to document the frequency of 1,4-dioxane biodegradation over a range redox conditions and inocula types and provides critical evidence for the feasibility of anaerobic 1,4-dioxane bioremediation.

### 3.3 Methods

#### 3.3.1 Chemicals and Inocula

1,4-dioxane was purchased from Alfa Aesar (MA, USA, 99+ % purity) and Sigma-Aldrich (MO, USA, 99.8% purity). Ethylenediaminetetraacetic acid (EDTA), iron (III) sodium salt, sodium sulfate, sodium nitrate and humic acid were purchased from Sigma-Aldrich (MO, USA). All stock solutions and dilutions were prepared using DI water. The uncontaminated agricultural soil samples were collected from two locations on the campus of Michigan State University (MSU), East Lansing, MI (herein called Soils E, F and G) and two locations at the Kellogg Biological Station (MSU), Hickory Corners, MI (KBS Soils 1, 2 and 3). The uncontaminated river sediment samples (sediments H and J) were collected from Red Cedar River, Okemos, MI and from a river leading to Lake Lansing in Haslett, MI. The contaminated site samples were sent to MSU from California (contaminated with trichloroethene, 1,1-dichloroethene and 1,4-dioxane) and Maine (contaminated with traces of 1,4-dioxane). All of the samples were stored in the dark at 6 °C until use.

### 3.3.2 Experimental Setup

Five sets of experiments were performed to examine the susceptibility of 1,4-dioxane to biodegradation under different potential electron acceptors (Table 3.1). All five experiments contained triplicates of live samples and abiotic controls (autoclaved daily for three consecutive days). The final set of experiments (experiment 5) also included live control microcosms (no 1,4-dioxane) to enable comparisons to the 1,4-dioxane degrading microbial communities.

Table 3.1. Experimental design (inocula and amendments) of the five microcosm studies.

<b>Experiment</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Soils E, F, G	X	X			X
KBS Soils 1, 2, 3					X
River Sediments H, J			X		H only
Contaminated Sites				CA, MN	CA only
Iron/EDTA/humic acid	X	X	X	X	
Nitrate		X	X	X	
Sulfate		X	X	X	
No electron acceptor		X	X	X	X
Lactate	X			X	X
Basal salts media				X	X
Samples for microbial community analysis		17	3	11	24

The first set of set of experiments was designed based on the positive results generated previously for 1,4-dioxane under iron reducing conditions (Shen et al. 2008). For this, 18 microcosms (70 mL serum bottles) were each inoculated with one of three agricultural soils (Soils E, F and G, 20 g wet weight), and 55 mL of a solution consisting of EDTA iron (III) sodium salt (10 mM), sodium lactate (5 mM) and humic acid (0.5 g/L). The triplicate

microcosms and abiotic controls were amended with approximately 5 mg/L of 1,4-dioxane (Alfa Aesar).

The second set of microcosms contained the same three agricultural soils (Soils E, F and G) under four redox conditions (nitrate, iron, sulfate amended and no amendment). Here, 72 microcosms were established (3 soils, 4 treatments) containing soil (5 g wet weight), 10 mL of a solution with either  $\text{NaNO}_3$  (10mM),  $\text{NaSO}_4$  (10mM), EDTA iron(III) sodium salt (10mM) or DI water (no amendment) in 30 mL glass serum bottles. The microcosms and abiotic controls were amended with approximately 5 mg/L of 1,4-dioxane (Alfa Aesar). The third set of microcosms contained each of the uncontaminated river sediment samples (Sediments H and J) using the same experimental setup (four redox conditions). Sodium lactate was not added for both sets of experiments to encourage the metabolic degradation of 1,4-dioxane.

The fourth set of microcosms (48 bottles including 2 sediments, 4 treatments) contained each of the contaminated sediment samples (from CA and MN) also under four redox conditions (nitrate, iron, sulfate amended and no amendment). In this case, a media solution was added containing  $\text{NH}_4\text{Cl}$  (1.5 g/L),  $\text{NaH}_2\text{PO}_4$  (0.6 g/L),  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$  (0.1 g/L),  $\text{KCl}$  (0.1 g/L),  $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$  (0.002 g/L) and sodium lactate (5 mM). The electron acceptor and 1,4-dioxane amendments were as described above and 1,4-dioxane from Sigma Aldrich was used.

The final set of microcosms contained no electron acceptor amendment with the aim of creating methanogenic conditions (based on positive results from some of the microcosms described above). The 72 microcosms included 8 different inocula (Soils E, F, G, KBS Soils 1, 2 and 3,

Sediment H and the CA Contaminated Sediment) with triplicates of sample microcosms and abiotic controls. Additionally, triplicate live control microcosms were included and these were treated in the same manner as the sample microcosms except no 1,4-dioxane was added. As stated above, this treatment was included to enable comparisons to the microbial communities exposed to 1,4-dioxane. The microcosms contained 5 g (wet weight) of soil or sediment, 25 mL of a solution containing sodium lactate (5mM),  $\text{NH}_4\text{Cl}$  (1.5 g/L),  $\text{NaH}_2\text{PO}_4$  (0.6 g/L),  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$  (0.1 g/L),  $\text{KCl}$  (0.1 g/L) and  $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$  (0.002 g/L) in 30 mL amber glass serum bottles. 1,4-dioxane (Sigma Aldrich, between 5-14 mg/L) was added the sample microcosms and abiotic controls.

All solutions were purged under a stream of nitrogen before being introduced into the anaerobic chamber. The solutions and soils/sediments were added to each serum bottle in the anaerobic chamber and, after approximately 3 days, 1,4-dioxane was added. The microcosms, closed with septa, were incubated in the anaerobic chamber at 20 °C. The anaerobic chamber was maintained with gaseous mix of approximately 5%  $\text{H}_2$ , 90%  $\text{N}_2$  and 5%  $\text{CO}_2$ . The vials were sealed using BiMetal vial crimp with PTFE/silicone septas to maintain the cultures at anaerobic conditions. All microcosms were transferred on a shaker at 200 rpm and maintained at 20 °C. The nitrate amended microcosms were tested for methane after 200 days of incubation using a GC (Hewlett Packard 5890).

### 3.3.3 1,4-Dioxane Analysis

A GC/MS with Agilent 5975 GC/single quadrupole MS (Agilent Technologies, CA, USA) equipped with a CTC Combi Pal autosampler was used to determine 1,4-dioxane concentrations.

Sterile syringes (1 mL) and needles (22 Ga 1.5 in.) were used to collect samples (0.7 mL) from each microcosm. The samples were filtered (0.22  $\mu\text{m}$  nylon filter) before being injected into an amber glass vial (40 mL) for GC/MS analysis. A method was developed to analyze 1,4-dioxane using solid phase micro extraction (SPME). The SPME fiber was inserted in the headspace of the vial and exposed to the analyte for 1 minute before being injected into the GC for thermal desorption. The fiber coating can adsorb the analytes in the vapor phase. Splitless injection was executed and the vials were maintained at 40 °C. The SPME fiber assembly involved 50/30 $\mu\text{m}$  Divinylbenzene/ Carboxen/ Polydimethylsiloxane (DVB/CAR/PDMS) and 24 Ga needle for injection. The initial oven temperature was 35 °C and was programmed to increase at a rate of 20 °C/min to 120 °C. Once it reached 120 °C, it increased at a rate of 40 °C/min to 250 °C, which was maintained for 3 min. A VF5MS column was used with helium as the carrier gas in constant flow mode at a flow rate of 1 ml/min. The conditioning of the SPME fiber was at 270 °C for 60 min at the beginning of each sequence.

#### 3.3.4 Compound Specific Isotope Analysis (CSIA)

At the last sampling point, three sets of samples and abiotic controls were submitted for CSIA analysis, including Soil F (nitrate amended, experiment two), and the two contaminated site samples with no electron acceptor amendment (experiment four). CSIA was performed at the University of Waterloo Environmental Isotope Laboratory (UWEIL), Ontario, Canada. The ratios of  $^{13}\text{C}/^{12}\text{C}$  and  $^2\text{H}/^1\text{H}$  were measured using a recently developed method. For this, the dilute 1,4-dioxane samples were concentrated on to a sorbent and subjected to thermal desorption in a GC coupled with an isotope ratio mass spectrometer (Bennett et al., 2018). All three sets of samples were submit to  $^{13}\text{C}/^{12}\text{C}$  analysis, however only Soil F was subject to  $^2\text{H}/^1\text{H}$ .

### 3.3.5 DNA Extraction and High Throughput Amplicon Sequencing

DNA was extracted (2 mL liquid and 0.6 mg soil/sediment, QIAGEN DNeasy PowerSoil kit) from the sample microcosms and corresponding live control microcosms (no electron acceptor amendment, experiment five) that illustrated 1,4-dioxane degradation (Soil F, CA Contaminated Site, Sediment H and KBS Soil 3, 24 extracts). Additionally, DNA was extracted from 31 other microcosms that illustrated 1,4-dioxane biodegradation (from the other four experiments). The concentration of DNA was determined using QUBIT dsDNA HS kit. The DNA extracts were submitted for high throughput 16S rRNA gene amplicon sequencing following previously described protocols (Caporaso et al. 2012; Caporaso et al. 2011) at the Research Technology Support Facility at Michigan State University. Specifically, 55 samples of microbial metagenomic DNA were submitted for amplicon library preparation and sequencing. The V4 hypervariable region of the 16S rRNA gene was amplified using the Illumina compatible, dual indexed primers 515f/806r. The primers and the library construction protocol developed in Patrick Schloss' lab were previously described (Kozich et al. 2013). PCR products were batch normalized using Invitrogen SequelPrep DNA Normalization plates and product recovered from the plate pooled. This pool was cleaned up and concentrated using AmpureXP magnetic beads. The pool was QC'd and quantified using a combination of Qubit dsDNA HS, Agilent 4200 TapeStation HS DNA1000 and Kapa Illumina Library Quantification qPCR assays. The pool was then loaded onto an Illumina MiSeq v2 Standard flow cell and sequencing was performed in a 2x250bp paired end format using a MiSeq v2 500 cycle reagent kit. Custom sequencing and index primers complementary to the 515f/806r oligos were added to appropriate wells of the reagent cartridge as previously described (Kozich et al. 2013). Base calling was performed by

Illumina Real Time Analysis (RTA) v1.18.54 and the output of RTA was demultiplexed and converted to FastQ format with Illumina Bcl2fastq v2.19.1.

### 3.3.6 Analysis of Sequencing Data

The amplicon sequencing data in the fastq file format was analyzed with Mothur (version 1.41.3) from Patrick D. Schloss Laboratory (Schloss 2009) using the MiSeq standard operating procedure (Schloss 2013). The steps involved the removal of barcode information and contiguous sequences were created using the forward and reverse reads. These were analyzed for errors and then classified. This involved checking the samples for the proper read length (<275 bp), ambiguous bases and homopolymer length greater than 8 to eliminate such sequences. The sequences were then aligned with the SILVA bacteria database (Pruesse et al. 2007) for the V4 region. Chimeras, mitochondrial and chloroplast lineage sequences were removed and then the sequences were classified into OTU's. The summary file created by mothur was analyzed with Microsoft Excel 2016 and STAMP (Statistical Analyses of Metagenomic Profiles, software version 2.1.3.) (Parks et al. 2014). STAMP was used to detect differences in the relative proportions of the taxonomic profiles between the live controls (no 1,4-dioxane) and the samples for each of the four inocula in the fifth set of experiments (Soil F, Contaminated Site 10A, Sediment H and KBS Soil 3). This analysis included Welch's two-sided t-test for two groups (samples and live controls) ( $p < 0.05$ ) to generate extended error bar figures for each soil.

### 3.3.7 Statistical Analysis

1,4-dioxane concentrations in the samples and abiotic controls were compared using the student's t-test for independent variables with equal variance. Normality of the 1,4-dioxane

concentration data was confirmed using the Shapiro-Wilk test ( $p>0.05$ ) and the assumption of equal variance was confirmed using Levene's tests ( $p>0.05$ ). The same tests were used to confirm equal variance and normality for the CSIA carbon data. The CSIA hydrogen data were not normal, therefore the Mann Whitney was used. These statistical tests were performed in XLSTAT software for statistical analysis in Excel (2019, <http://www.xlstat.com>).

### 3.4. Results

#### 3.4.1 1,4-Dioxane Removal Trends

1,4-dioxane concentrations in samples and abiotic controls were monitored for approximately one year in each of the five sets of experiments. Given the long incubation period, it is likely that the electron acceptors (nitrate, iron, sulfate) were depleted for each treatment. This was confirmed in one case, by the production of methane in the nitrate amended, agricultural soil inoculated microcosms. The first set of microcosms (inoculated with three agricultural soils and an EDTA iron/humic acid amendment) illustrated no significant difference ( $p<0.05$ ) between the samples and abiotic controls for the majority of sampling times, suggesting no biological 1,4-dioxane degradation occurred with this treatment for these soils (Figure 3.1).

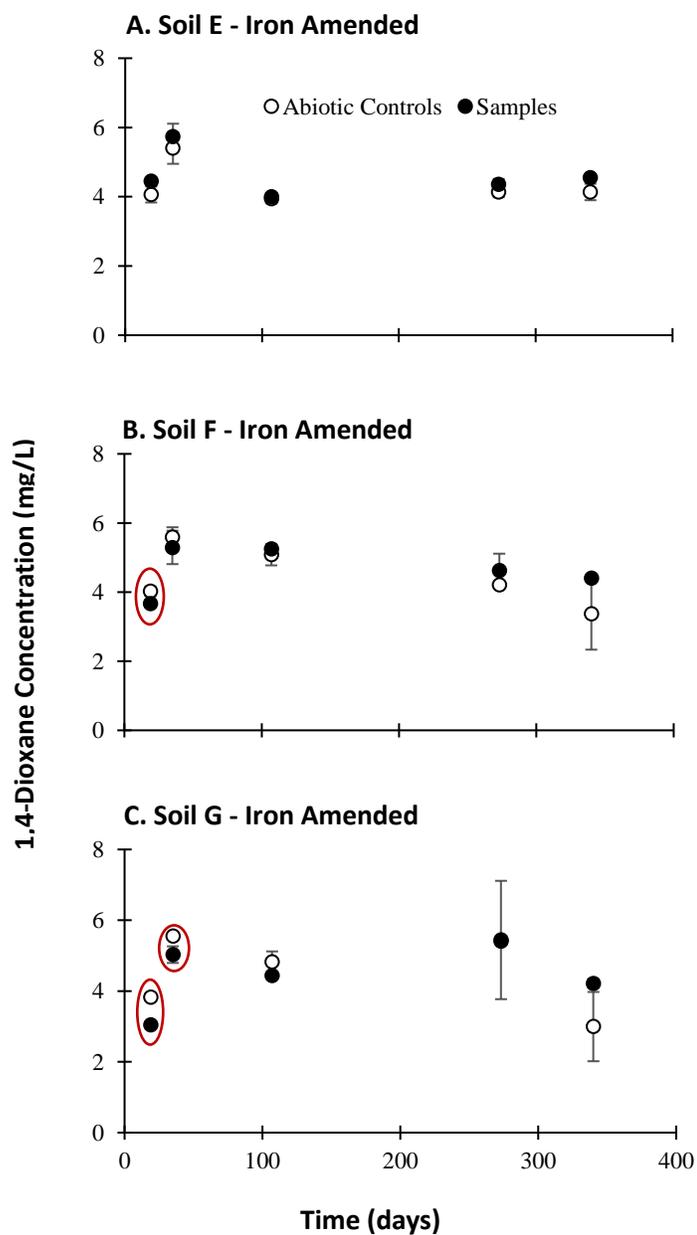


Figure 3.1. 1,4-dioxane concentrations in sample and abiotic control microcosms of three soils amended with EDTA/iron and humic acid. The values and bars represent averages and standard deviations from triplicates. The circles represent a significant difference (two-tailed t-test,  $p < 0.05$ ) between the samples and abiotic controls (when abiotic controls > samples).

The second set of experiments (agricultural soils and four electron acceptor treatments) produced a variety of trends between the soils and treatments (Figure 3.2). Comparing the treatments, the most consistent removal was in the nitrate amended microcosms (Figure 3.2A). For all three soils, there was a statistically significant difference ( $p < 0.05$ ) between the samples and abiotic controls at the last sampling point (day 295 to 316). The Soil F nitrate amended sample microcosms and abiotic controls were submitted for CSIA (as discussed later). Similar to the first set of microcosms, no sustained significant degradation was noted in the samples compared to the abiotic controls in the iron amended microcosms (Figure 3.2B).

In the sulfate amended microcosms, only Soil E resulted in a significant difference between the samples and abiotic controls over several time points (Figure 3.2C). When no electron acceptor was added, only Soils E and F illustrated consistent reduced concentrations in the samples compared to the abiotic controls (Figure 3.2D). Overall, the three soils illustrated a range of 1,4-dioxane biodegradation abilities, depending on the electron acceptor amendment. Specifically, Soil E most commonly produced significant differences between the samples and abiotic controls, followed by Soil F, then Soil G (Table 3.2).

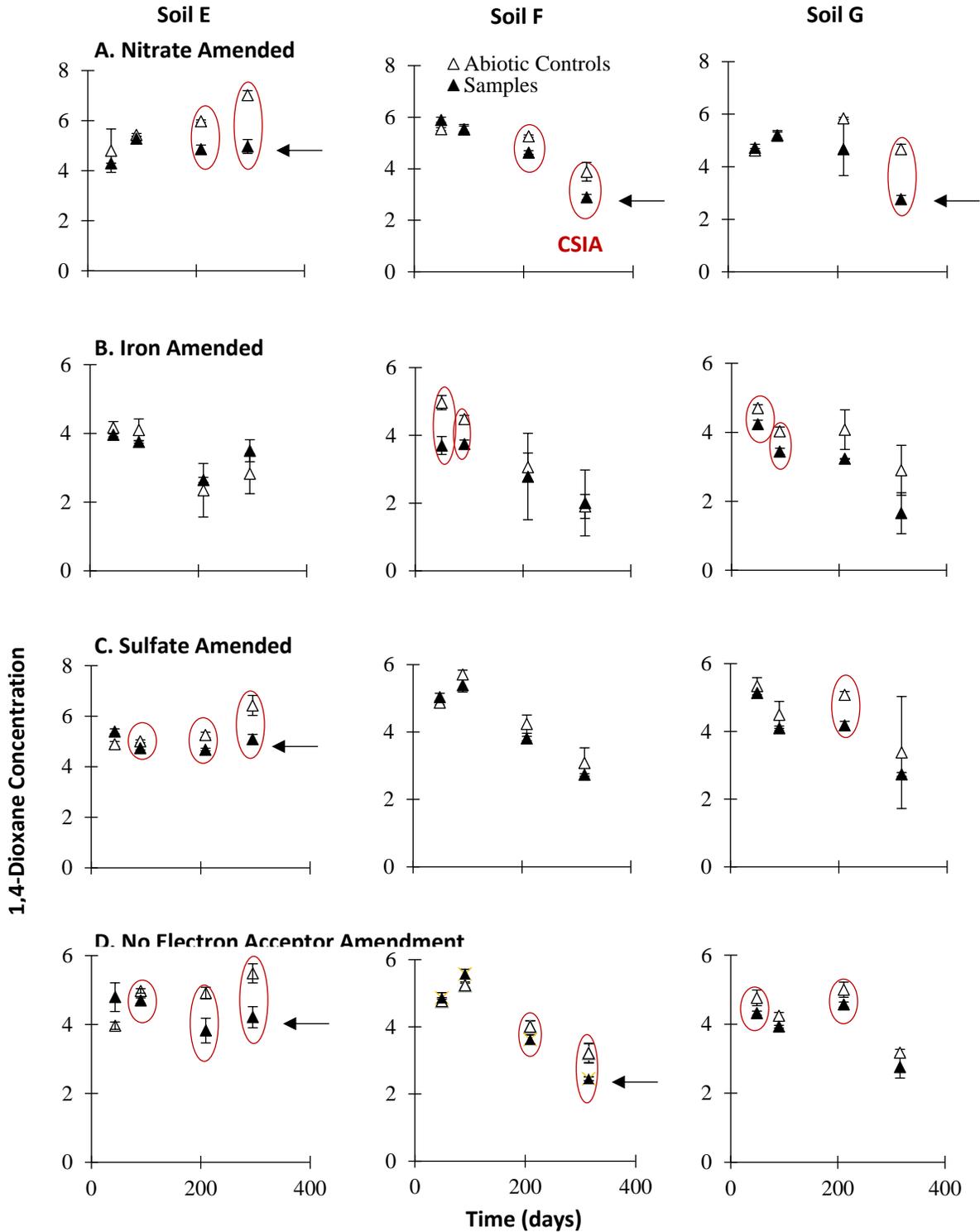


Figure 3.2. 1,4-dioxane concentrations in sample and abiotic control microcosms of three soils with different electron acceptor amendments (A-D). The values and bars represent averages and standard deviations from triplicates (except soil G, n = 2 for three cases). The circles represent a significant different (two-tailed t-test,  $p < 0.05$ ) between the samples and abiotic controls (when controls > samples). The arrows indicate the samples subject to DNA extraction. CSIA was performed on the samples and abiotic controls from the soil F, nitrate amended treatments.

Table 3.2. Percent removal in samples compared to the controls at the last time point (when the difference was statistically significant, determined with student's t-test for independent variables with equal variance,  $p < 0.05$ ).

	<b>Nitrate</b>	<b>Iron</b>	<b>Sulfate</b>	<b>No Amendment</b>
Soil E	29	0	21	23
Soil F	25	0	0	24
Soil G	41	0	0	0
Sediment H	0	0	0	43
Sediment J	0	0	0	0
Maine	0	0	29	50
California	16	0	0	16
% of sample types with decreases in samples compared to the controls	57	0	29	71
% average removal w/ std deviation	$16 \pm 16$	0	$7 \pm 12$	$22 \pm 19$

The third set of experiments involved Sediment H and Sediment J (both collected from uncontaminated rivers) as inocula. No significant decreases in 1,4-dioxane concentrations were seen in nitrate amended, sulfate amended or iron amended samples compared to the abiotic controls for both sediments (Figure 3.3A, 3.3B and 3.3C). Among the samples with no amendments, only Sediment H indicated a significant decrease in 1,4-dioxane concentrations compared to the abiotic controls (Figure 3.3D). This was one of the most dramatic reductions in 1,4-dioxane in the live sample microcosms compared to the abiotic controls over all treatments and inocula types.

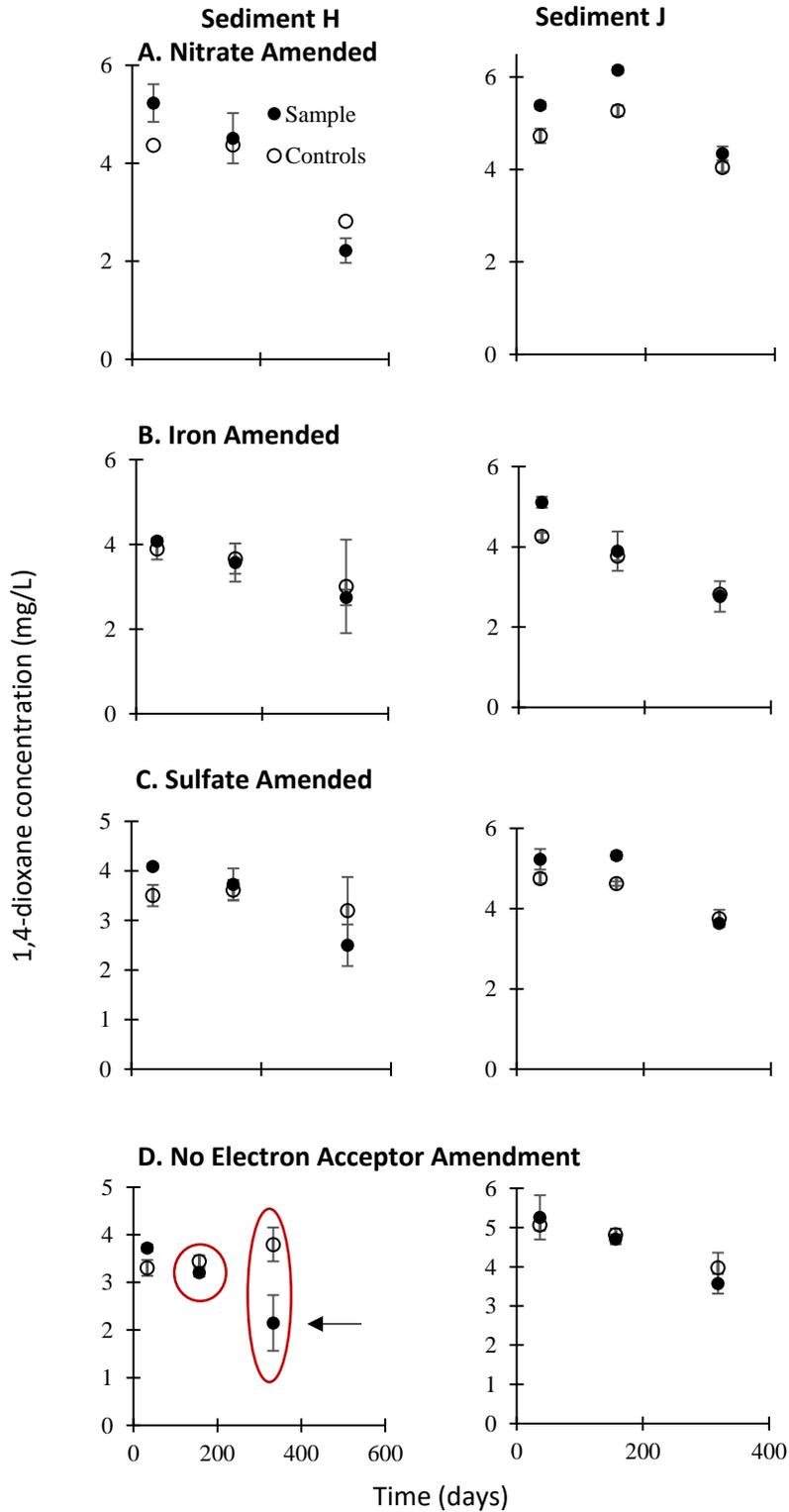


Figure 3.3. 1,4-dioxane concentrations in Sediment J and H sample and abiotic control microcosms over time with different amendments. The points illustrate average values from triplicates (except nitrate amended controls, sediment H, n=1) and the bars illustrate standard deviations. The circles represent a significant different (two-tailed t-test,  $p < 0.05$ ) between the samples and abiotic controls. The arrow indicates the samples subject to DNA extraction.

The fourth set of experiments (inoculated with sediments from two contaminated sites over four treatments) illustrated similar trends as those from the agricultural soils (Figure 3.4). Specifically, in one case for each of the nitrate amended and sulfate amended microcosms, there was a significant difference between the samples and abiotic controls (CA and MN, respectively) (Figure 3.4A and C). Further, no differences were observed between the samples and abiotic controls under the iron amended treatment (Figure 3.4B). Finally, in both cases, a substantial decrease (~17% and 50% for CA and MN) was observed in the samples compared to the abiotic controls when no electron acceptor was added (Figure 3.4D). The sample microcosms and abiotic control microcosms from the no amendment treatment with both sediments were submitted for CSIA analysis.

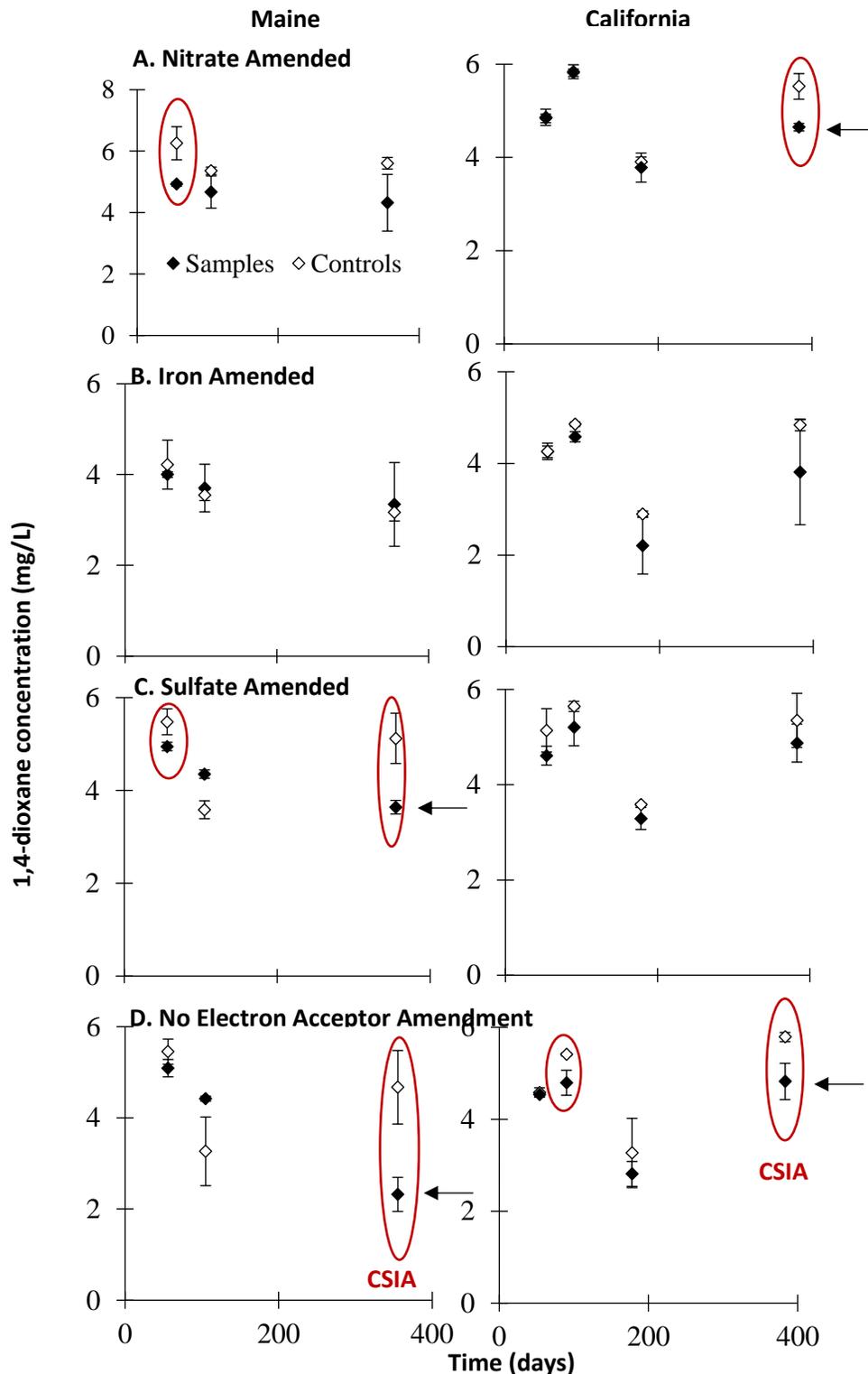


Figure 3.4. 1,4-dioxane concentrations in Maine and California sample and abiotic control microcosms over time with different amendments. The points illustrate average values from triplicates (except sulfate amended Maine sample,  $n=2$  and sulfate amended CA control,  $n=2$ ) and the bars illustrate standard deviations. The circles represent a significant different (two-tailed t-test,  $p < 0.05$ ) between the samples and abiotic controls (controls  $>$  samples). The arrows indicate the samples subject to DNA extraction. CSIA was performed on the samples and abiotic controls from both of the no electron acceptor treatment (D).

A summary of the results described in the four sets of experiments is provided, focusing only on the last sampling point for each (Table 3.2). In the nitrate amended treatments, four of the six inocula types (or 57%) produced a significant reduction in 1,4-dioxane in the samples compared to the abiotic controls. In contrast, no differences were noted between the samples and abiotic controls under any of the iron amendment treatments. Only two (Soil E and the MN contaminated site sediments) of the seven cases (or 29%) illustrated biodegradation under the sulfate amendment treatment. Finally, the most frequent (five of the seven inocula types or 71%) reduction of 1,4-dioxane was noted under the no amendment treatment. The average removal (again, only at the last time point) was greatest under the no amendment treatment ( $22 \pm 19 \%$ ), followed by the nitrate amended samples ( $16 \pm 16 \%$ ) and the sulfate amended samples ( $7 \pm 12 \%$ ). Overall, it appears the likelihood of 1,4-dioxane biodegradation is the greatest with no amendments, when the microbial populations are likely under methanogenic redox conditions.

The fifth set of experiments were established to compare the microbial communities between the live microcosms and the live control microcosms (no 1,4-dioxane amendment) and focused only on the no amendment treatment. For this, six agricultural soils, one contaminated (CA) and one uncontaminated sediment (Sediment H) sample were used as inocula (Figure 3.5). 1,4-dioxane concentrations were monitored in the samples and abiotic controls for more than 400 days. A significant reduction in 1,4-dioxane in the samples compared to the abiotic controls was only noted for two of six agricultural soils (Figure 3.5B, H). Consistent with the results described above (third and fourth set of experiments), a significant reduction in 1,4-dioxane was noted in the samples compared to the abiotic controls in Sediment H (Figure 3.5E) and the contaminated

samples from CA (Figure 3.5D). In this set of experiments, only 50% of the inocula types illustrated detectable levels of 1,4-dioxane biodegradation.

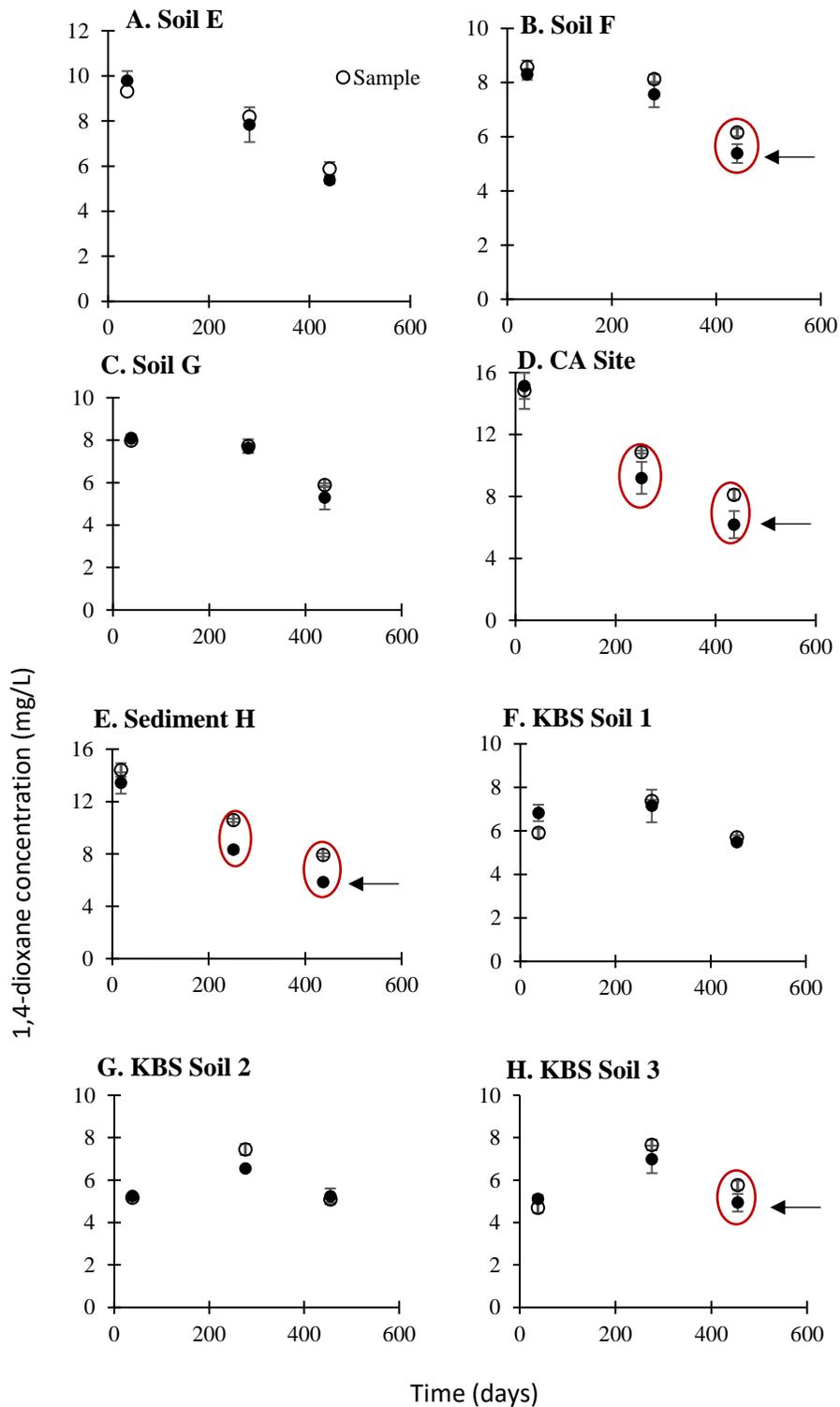


Figure 3.5. 1,4-dioxane concentrations in sample and abiotic control microcosms over time with different inocula. The points illustrate average values from triplicates (except Soil E and KBS Soil 2 samples, which are replicates) and the bars illustrate standard deviations. The circles represent a significant different (two-tailed t-test,  $p < 0.05$ ) between the samples and abiotic controls. The arrows indicate the samples subject to DNA extraction.

### 3.4.2 Compound Specific Isotope Analysis (CSIA)

CSIA was used to further investigate biodegradation in three sets of the anaerobic 1,4-dioxane degrading microcosms. For this, subsamples were collected from sample microcosms and abiotic controls from the Soil F/nitrate treatment (second set of experiments) and from the CA and MN contaminated site sediments/no electron acceptor treatment (fourth set of experiments) and these were sent to UWEIL for CSIA. This laboratory has already developed the methodology to measure  $\delta^{13}\text{C}$  and  $\delta^2\text{H}$  values for 1,4-dioxane (Bennett et al., 2018). All three sets of samples (six for each treatment, eighteen in total) were subject to  $^{13}\text{C}/^{12}\text{C}$  analysis, but only the six Soil F/nitrate microcosms were subject to  $^2\text{H}/^1\text{H}$  analysis. 1,4-dioxane degradation should result in more positive  $^{13}\text{C}/^{12}\text{C}$  and  $^2\text{H}/^1\text{H}$  ratios (or more positive  $\delta^{13}\text{C}$  and  $\delta^2\text{H}$  values) because bonds involving heavier isotopes are more difficult to break, and so bonds consisting of lighter isotopes are preferentially degraded, causing the residual, non-degraded contaminant to be heavy isotope enriched. In all cases, more positive  $\delta^{13}\text{C}$  and  $\delta^2\text{H}$  values were found in the live samples compared to the abiotic controls (Figure 3.6). However, the differences in  $\delta^{13}\text{C}$  values were only statistically significantly different between the samples and abiotic controls for Soil F/nitrate and contaminated site MN/no electron acceptor. It is likely that the decrease in 1,4-dioxane concentrations were too low for this method to detect a difference in  $\delta^{13}\text{C}$  values for the contaminated site CA/no electron acceptor. Specifically, for this treatment the average reduction in 1,4-dioxane concentration in the samples compared to the controls was 17% compared to 25% and 50% reductions in the other two treatments. The  $\delta^2\text{H}$  values were not significantly different between the samples and controls for Soil F (Figure 3.6B).

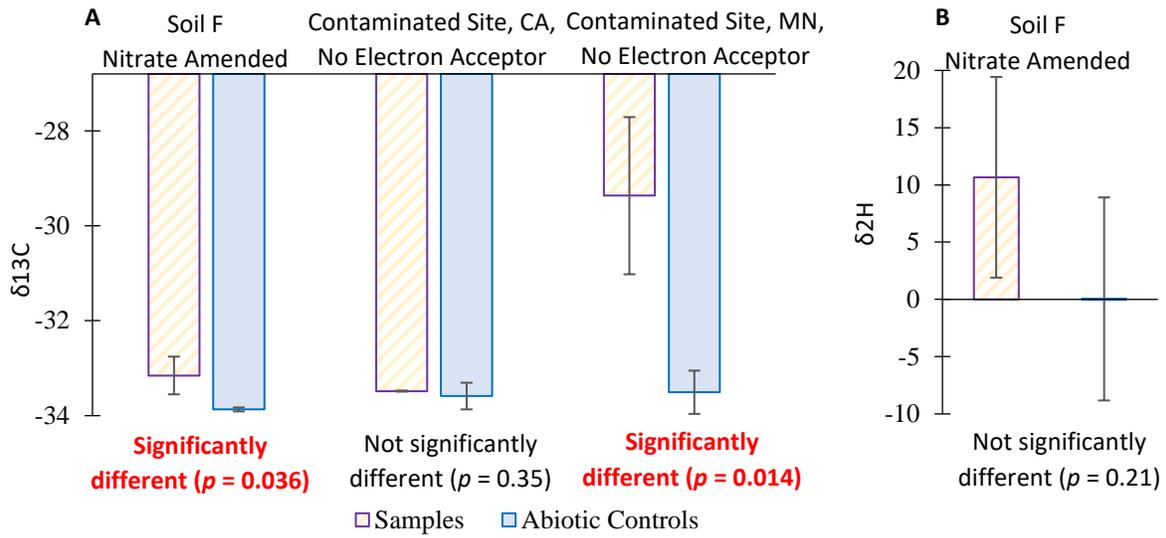


Figure 3.6. Comparison of  $\delta^{13}\text{C}$  (A) and  $\delta^2\text{H}$  (H) values in samples and abiotic controls of a sub-set of the experimental microcosms. The error bars represent standard deviations from triplicate samples and control microcosms and the p values were derived from the student's t-test (two tailed, independent

### 3.4.3 Putative 1,4-Dioxane Degraders

To determine if any microorganism could obtain a growth benefit from 1,4-dioxane degradation, sample microcosms were supplied with media and 1,4-dioxane and the live control microcosms were supplied with the same media, but no 1,4-dioxane (fifth set of microcosms). Therefore, an increase in the relative abundance of any microorganism between the samples and live controls could be attributed to the presence of 1,4-dioxane. From this, a reasonable hypothesis would be that the enriched microorganisms are being exposed to growth supporting substrates from 1,4-dioxane degradation. This comparison focused on the four inocula that illustrated 1,4-dioxane degradation (Figure 3.5B, D, E, H). Significant phylotypes enrichments ( $p < 0.05$ ) were noted for the microcosms inoculated with CA contaminated site sediments, Sediment F and KBS Soil 3 (Figure 3.7). Only one phylotype, unclassified *Comamonadaceae*, (phylum *Proteobacteria*, class *Betaproteobacteria*, order *Burkholderiales*) was enriched in the CA contaminated site microcosms (Figure 3.7A). Nine phylotypes were enriched in the uncontaminated site (Sediment H) microcosms, with 3 *genus incertae sedis* (phylum *Verrucomicrobia*, class *Subdivision3*) being dominant (Figure 3.7B). There was a small (yet significant) enrichment of one phylotype, *Pseudoxanthomonas*, (phylum *Proteobacteria*, class *Gammaproteobacteria*, order *Xanthomonadales*, family *Xanthomonadaceae*) in the KBS Soil 3 microcosms (Figure 3.7C) and no enrichment of any phylotypes in the Soil F microcosms.

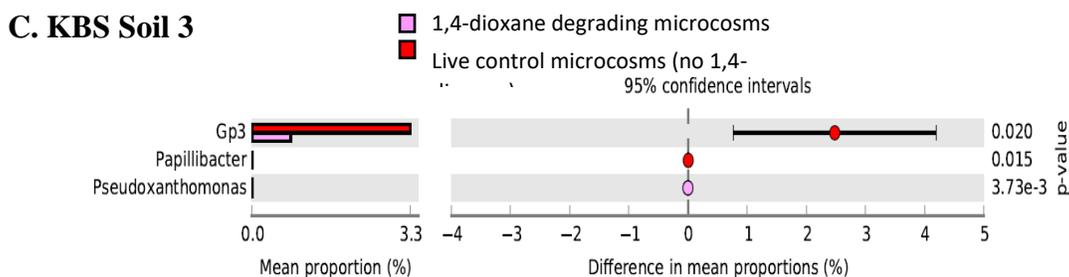
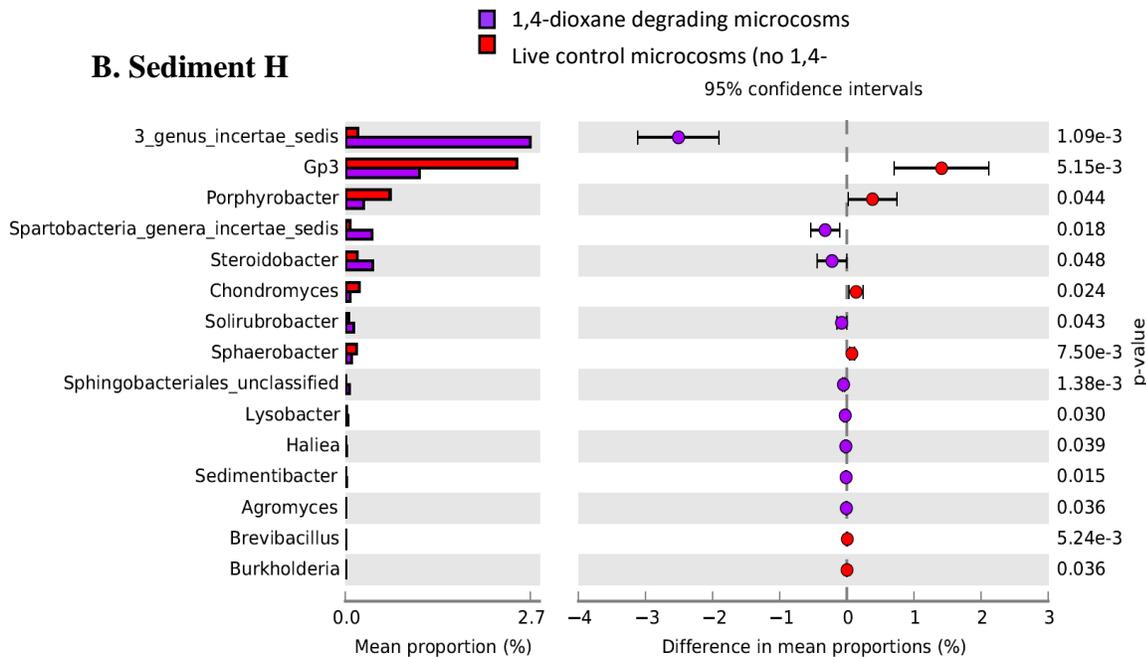
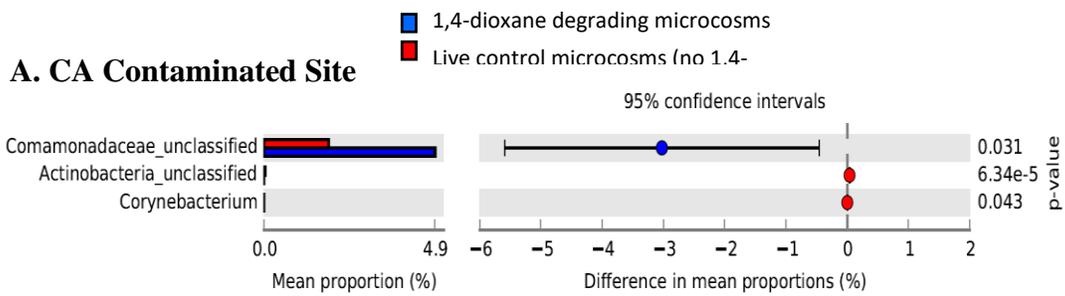


Figure 3.7. Phylotypes with a statistically significant difference (Welch's t-test, two sided,  $p < 0.05$ ) between the 1,4-dioxane degrading samples ( $n=3$ ) compared to the live controls ( $n=3$ ). The comparisons are shown for the Contaminated Site (CA) microcosms (A), the Sediment H microcosms (B) and KBS Soil 3 microcosms (C). No differences were noted for Soil F. The data points to the left of the dashed line indicate phylotypes more abundant in the 1,4-dioxane degrading samples compared to the controls and those to the right indicate the reverse. For each, the x-axis have different scales.

The sequencing data sets were also analyzed to determine the relative abundance of previously identified 1,4-dioxane degraders. It is important to note that the analysis could only be performed at the genus level, therefore it is not possible to determine if particular species or strains are present. The 55 DNA extracts examined were from the fifth set of microcosms that illustrated 1,4-dioxane degradation (Soil F, Contaminated Site 10A, Sediment H and KBS Soil 3) and from 31 other microcosms that illustrated 1,4-dioxane biodegradation (from the other four experiments) (Table 3.1). Interestingly, all four soil inoculated microcosms contained a high relative abundance (~5-45%) of the genus *Rhodanobacter* (Figure 3.8, inserts). All four soils contained the genera *Pseudomonas*, *Mycobacterium*, *Acinetobacter*, although the relative abundance values were much lower (< 0.15%) than *Rhodanobacter* (Figure 3.8). The microcosms inoculated with KBS Soil 3 illustrated the largest number of previously identified 1,4-dioxane degraders (11 from 15 genera) (Figure 3.8A). The Soils E, F and G inoculated microcosms primarily contained 4, 8 and 5 of these genera. The well-studied *Pseudonocardia* and *Rhodococcus* genera were only common in KBS Soil 3 microcosms (Figure 3.8A). The microcosms inoculated with contaminated site (CA) and uncontaminated site (Sediment H) sediments contained lower relative abundance values (<0.04%) for *Rhodanobacter* (Figure 3.9A and B). The CA contaminated site microcosms illustrated high levels of *Pseudomonas* (~20-70%) and lower levels of *Rhodanobacter* and *Acinetobacter* (Figure 3.9A). *Flavobacterium* and *Rhodococcus* were only present in two from the twelve CA microcosms examined. The Sediment H microcosms contained lower levels (<0.011%) of *Pseudomonas*, along with *Rhodanobacter* and four other genera (Figure 3.9B). The KBS Soil 3 microcosms illustrated higher values for *Rhodanobacter* (in one case, *Pseudomonas*) and lower levels of seven other genera (Figure 3.9C). *Rhodococcus* was present at low levels with all three inocula, but only in a small number

of microcosms. *Pseudonocardia* was absent from the CA microcosms, present at low levels in both a subset of the Sediment H microcosms and in all of the KBS Soil 3 microcosms. The average relative abundance values of the two enriched phylotypes discussed above (unclassified *Comamonadaceae* at 4.9% and 3 *genus incertae sedis* at 2.7%) were markedly higher than all of the previously identified dioxane degraders, except for *Rhodanobacter* and *Pseudomonas*.

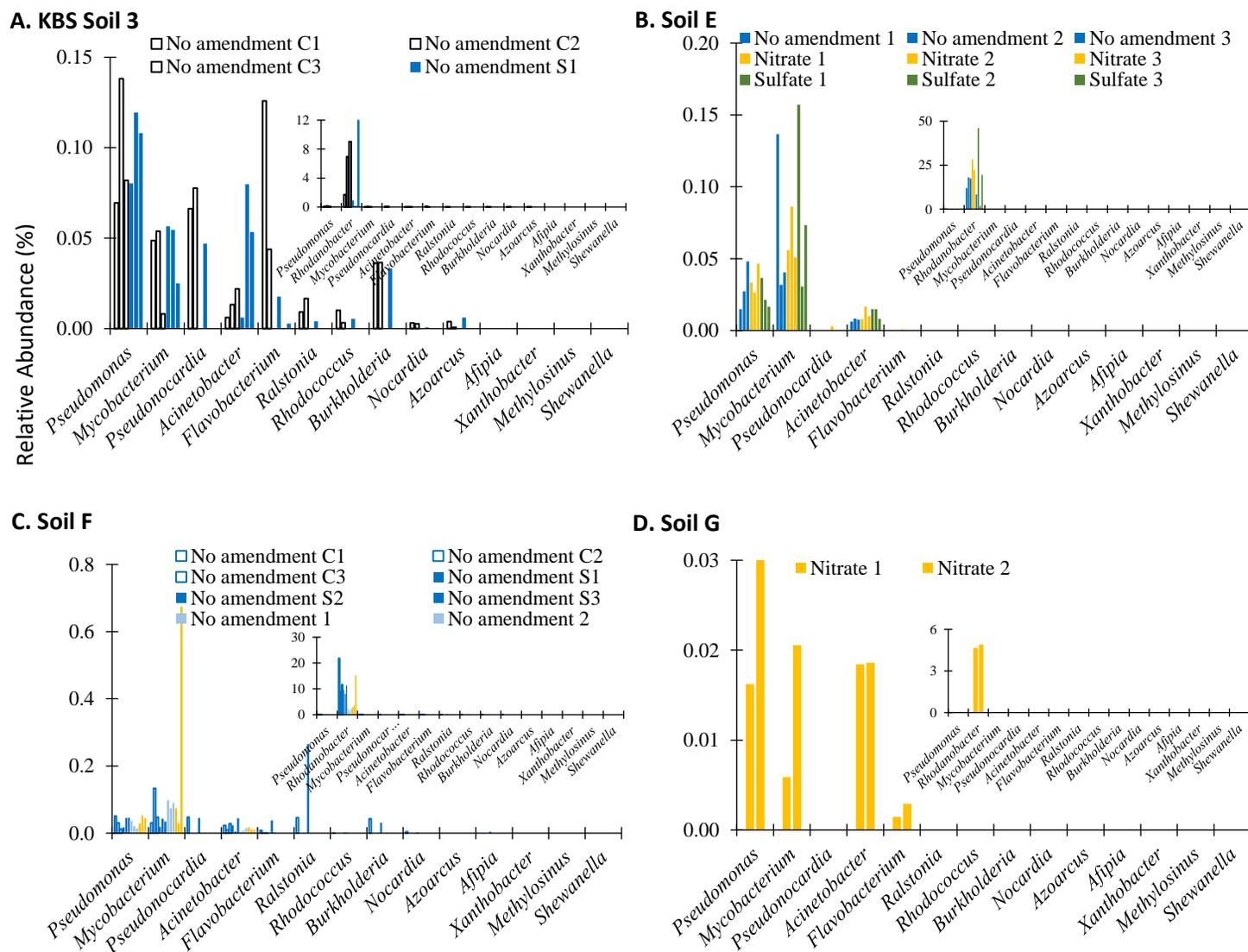
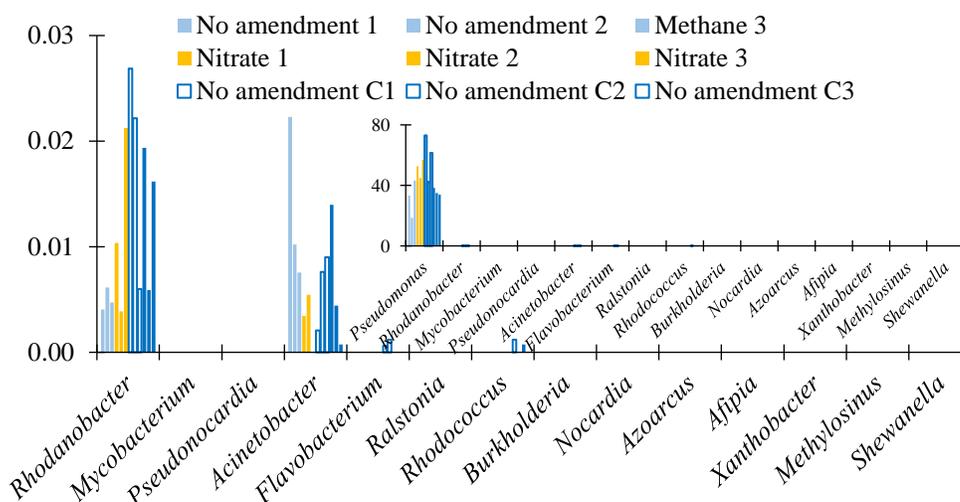
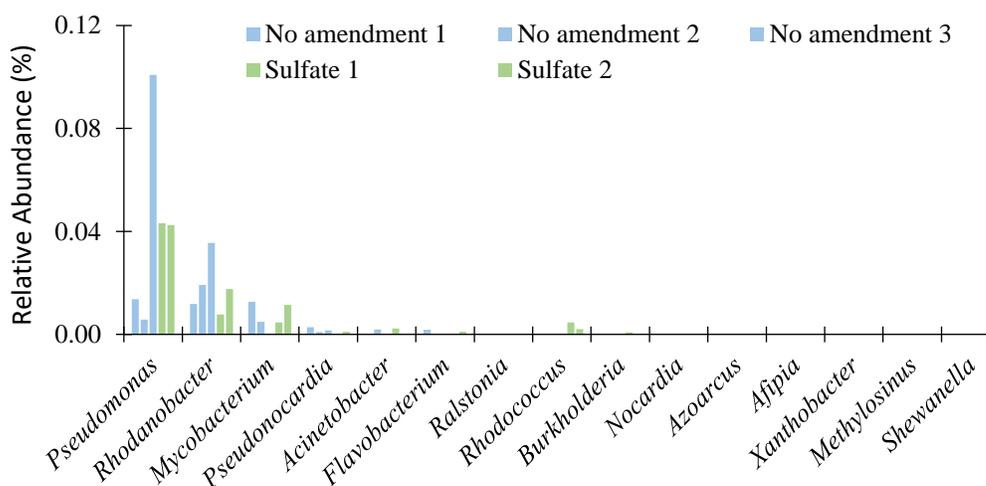


Figure 3.8. Relative abundance (%) of genera previously associated with 1,4-dioxane biodegradation in replicates of 1,4-dioxane degrading microcosms or live controls (C, no 1,4-dioxane) inoculated with different soils. The inserts have a different axis to better illustrate the relative abundance of *Rhodanobacter*.

### A. CA Contaminated Site



### B. MN Contaminated Site



### C. Sediment H

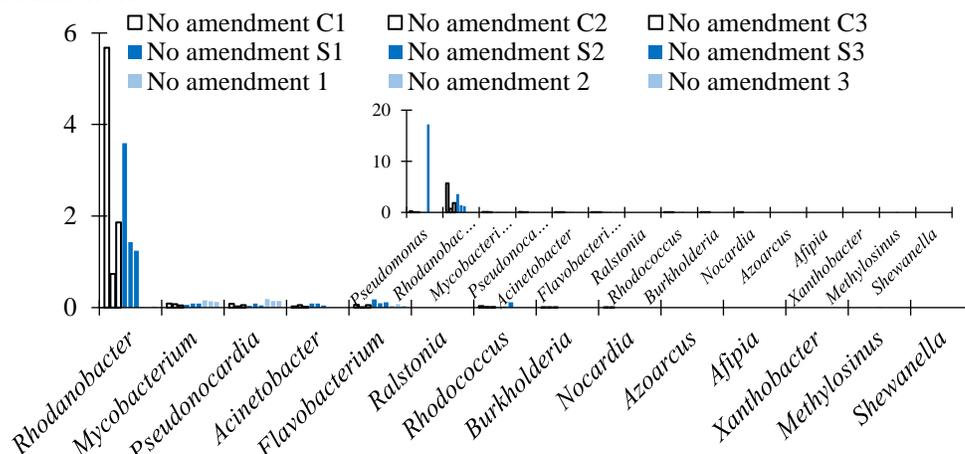


Figure 3.9. Relative abundance (%) of genera previously associated with 1,4-dioxane biodegradation in replicates of 1,4-dioxane degrading microcosms or live controls (C, no 1,4-dioxane) inoculated with different sediments (A, B, C). The inserts have a different axis to better illustrate the relative abundance of *Pseudomonas*.

### 3.5 Discussion

Although much is known about the microorganisms and functional genes associated with 1,4-dioxane degradation under aerobic conditions, limited research has addressed the susceptibility of 1,4-dioxane to biodegradation under anaerobic conditions (Rodriguez 2016; Shen et al. 2008; Steffan 2007). This research area is particularly important because 1,4-dioxane is a frequent co-contaminant at chlorinated solvent sites, which are commonly subject to remediation under highly reducing conditions (i.e. sulfate reducing or methanogenic). The current study examined the occurrence of 1,4-dioxane biodegradation using numerous inocula (soils/sediments from contaminated and uncontaminated sites) and a range of electron acceptor treatments (nitrate, iron/humic acid, sulfate and no amendment). Additionally, the phylotypes enriched during 1,4-dioxane biodegradation were investigated.

As discussed above, previous research reported no 1,4-dioxane biodegradation over a range of redox conditions (Steffan 2007) with biodegradation occurring under iron reducing conditions (Rodriguez 2016; Shen et al. 2008). Shen, Chen and Pan (2008) observed anaerobic 1,4-dioxane biodegradation using sludge from a wastewater treatment plant. Three cases were highlighted using samples which were amended with a) Fe(III) oxide b) Fe(III)-EDTA and c) Fe(III)-EDTA and humic acid. The reductions in 1,4-dioxane concentrations increased from case (a) through (c) within the same 40 day period (Shen et al. 2008). However, the microorganisms responsible for 1,4-dioxane removal were not identified. Rodriguez (2016) reported that 1,4-dioxane appeared to degrade under iron reducing conditions in the first 40 days. However, the concentration decreases could not be attributed to anaerobic biodegradation due to oxygen intrusion leading to oxidation of reduced iron (Rodriguez 2016). In the current work, the iron/EDTA/humic acid

amended microcosms did not result in any significant removal in the live samples compared to the abiotic controls. One hypothesis for the conflicting results (compared to Shen, Chan and Pan 2008) could be the different sources of inocula (anaerobic wastewater sludge compared to soils and sediments in the current study) for the two studies, resulting in microbial communities with varying degradative abilities. It is important to note that in the current study, the contaminated site microcosms did not illustrate any degradation under these conditions (iron/EDTA/humic acid), therefore it is unlikely that *in situ* iron amendments would be a successful 1,4-dioxane remediation approach.

An important limitation to the current study concerns the lack of data on electron acceptor concentrations. Given the long incubation periods, it is likely that electron acceptor depletion occurred and methanogenic conditions eventually dominated (which was confirmed in one case). Therefore, caution should be taken interpreting the impact of the addition of the various electron acceptors. However, it is reasonable to conclude that iron/EDTA/humic acid or sulfate amendments did not facilitate 1,4-dioxane biodegradation in the majority of cases and therefore these treatments would be ineffective *in situ* remediation techniques.

1,4-dioxane biodegradation most commonly occurred in the nitrate amended treatments and no electron acceptor treatments. Four and five (from seven) inocula sources examined in the first four sets of experiments produced significant reductions in 1,4-dioxane concentrations in the samples compared to the abiotic controls in the nitrate and no electron acceptor treatments, respectively. Further, both contaminated site sediments illustrated significant removal in the samples compared to the controls in the no electron acceptor treatment. However, the

biodegradation rates were slow, with concentration reductions occurring over approximately one year. Taken together, the data indicate *in situ* 1,4-dioxane bioremediation under methanogenic conditions may be a feasible remediation approach if sufficient time is permissible for site cleanup.

CSIA has been previously applied to examine aerobic 1,4-dioxane biodegradation (Bennett et al. 2018; Gedalanga et al. 2016; Pornwongthong 2014). Recently, CSIA enabled the discrimination between the activities of the bacteria *Rhodococcus rhodochrous* ATCC 21198 and *Pseudonocardia tetrahydrofuranoxidans* K1 during 1,4-dioxane biodegradation (Bennett et al. 2018). Another study examined 1,4-dioxane biodegradation using CSIA in both pure and mixed cultures and found more positive  $\delta^{13}\text{C}$  and the  $\delta^2\text{H}$  values after biodegradation (Pornwongthong 2014). Also, CSIA provided additional evidence of 1,4-dioxane biodegradation in a groundwater plume that contained both metabolic and co-metabolic 1,4-dioxane degrading bacteria (Gedalanga et al. 2016). Similar to the current work, CSIA has been adopted to characterize biodegradation of many organic compounds including chlorinated solvents, aromatic petroleum hydrocarbons and fuel oxygenates (Hunkeler et al. 2008; Pooley et al. 2009; Qian et al. 2019; Rosell et al. 2007; Wang et al. 2004). The applicability of CSIA has also been explored in in-situ biodegradation of organic contaminants such as hexachlorocyclohexane in contaminated soils (Qian et al. 2019).

To our knowledge, this is the first application of CSIA to investigate anaerobic 1,4-dioxane biodegradation. In two of the three cases examined, CSIA provided additional evidence for 1,4-dioxane biodegradation. In one case, the reduction in 1,4-dioxane in the samples compared the

controls was likely too low for the method to detect a significant  $^{13}\text{C}/^{12}\text{C}$  enrichment. The changes in  $^{13}\text{C}/^{12}\text{C}$  ratios in the current study were less than those reported previously for aerobic 1,4-dioxane biodegradation (Bennett et al. 2018), however, the changes in 1,4-dioxane concentrations were also less in the current study. The preliminary data set generated here indicate the value and limitations for  $^{13}\text{C}/^{12}\text{C}$  analysis for documenting anaerobic 1,4-dioxane biodegradation. Further research is required to determine the value of measuring  $^2\text{H}/^1\text{H}$  for generating evidence for the biodegradation of this chemical.

The greatest 1,4-dioxane decreases were observed in the microcosms inoculated with river sediment (Sediment H) and the contaminated sediment from Maine. Therefore, future research should examine the capabilities of other sediments to biodegrade this contaminant. The current research included experiments both with and without the addition of sodium lactate as an electron donor. Unfortunately, no clear trend was apparent as to the value of adding this substrate. Many chlorinated solvent sites are amended with an electron donor (e.g. vegetable soil, hydrogen release compound) so at least this particular substrate did not have any negative impact on 1,4-dioxane biodegradation. Additional research will be needed before any conclusions can be reached on the value of adding lactate to enhance biodegradation.

The analysis of the microbial communities from 1,4-dioxane degrading microcosms indicated unclassified *Comamonadaceae* were obtaining a growth benefit in microcosms inoculated with sediment from the CA contaminated site. *Comamonadaceae* is a large and diverse family (over 100 species in 29 genera) that illustrate an impressive level of phenotypic diversity, including aerobic organotrophs, anaerobic denitrifiers, iron reducing bacteria, hydrogen oxidizers,

photoautotrophic and photoheterotrophic bacteria, and fermentative bacteria (Willems 2014).

Additional research will be needed to determine which particular species within this family were enriched during 1,4-dioxane degradation.

For the uncontaminated sediment (Sediment H), nine phlotypes were enriched in the 1,4-dioxane degrading microcosms compared to the controls with 3 *genus incertae sedis* (phylum *Verrucomicrobia*, class *Subdivision3*) being more highly enriched compared to the others (average 2.7% relative abundance). This phlotype also exhibited a high relative abundance (average 2.9%) in the MN contaminated sediment microcosms that illustrated 1,4-dioxane degradation. *Verrucomicrobia* is a rarely cultivated phylum that is generally poorly studied (Naumoff and Dedysh 2018; Tanaka et al. 2017). Again, additional research is needed to confirm the role of these phlotypes in 1,4-dioxane biodegradation.

One genus, *Pseudoxanthomonas*, (phylum *Proteobacteria*, class *Gammaproteobacteria*, order *Xanthomonadales*, family *Xanthomonadaceae*) was enriched following 1,4-dioxane biodegradation in KBS Soil 3. This genus was recently linked to the biodegradation of anti-inflammatory drugs in the environment (Lu et al. 2019). The 1,4-dioxane degrading communities also contained sequences that classified in the genera of previously reported 1,4-dioxane degraders. The soil microcosms were dominated primarily by *Rhodanobacter* with lower abundance values for *Pseudomonas*, *Mycobacterium* and *Acinetobacter*. The sediment communities were dominated by *Pseudomonas* and *Rhodanobacter*. However, as stated above, it is important to note the classification is only to the genus level, therefore it is unknown if

specific strains or species are present. Again, additional work is needed to determine the importance of these genera in 1,4-dioxane degradation under these experimental conditions.

Overall, the current study indicates 1,4-dioxane biodegradation under anaerobic, and likely highly reducing conditions, is feasible. Therefore, natural attenuation may be an appropriate clean-up technology at sites where time is not a constraint. This work also provides novel insights on microbial community obtaining a growth benefit from 1,4-dioxane under anaerobic conditions. Future research should target sediments from other contaminated sites along with additional CSIA measurements to confirm biodegradation.

## REFERENCES

## REFERENCES

- Adams CD, Scanlan PA, Secrist ND (1994) Oxidation and biodegradability enhancement of 1,4-dioxane using hydrogen peroxide and ozone. *Environ Sci Technol* 28(11):1812-1818
- Adamson DT, Anderson RH, Mahendra S, Newell CJ (2015) Evidence of 1,4-dioxane attenuation at groundwater sites contaminated with chlorinated solvents and 1,4-dioxane. *Environ Sci Technol* 49(11):6510-6518 doi:10.1021/acs.est.5b00964
- Adamson DT, Mahendra S, Walker KL, Rauch SR, Sengupta S, Newell CJ (2014) A multisite survey to identify the scale of the 1,4-dioxane problem at contaminated groundwater sites. *Environ Sci Tech Let* 1(5):254-258
- Bennett P, Hyman M, Smith C, El Mugammar H, Chu MY, Nickelsen M, Aravena R (2018) Enrichment with carbon-13 and deuterium during monooxygenase-mediated biodegradation of 1,4-dioxane. *Environ Sci Tech Let* 5(3):148-153 doi:10.1021/acs.estlett.7b00565
- Bernhardt D, Diekmann H (1991) Degradation of dioxane, tetrahydrofuran and other cyclic ethers by an environmental *Rhodococcus* strain. *Appl Microbiol Biot* 36(1):120-123
- Bhushan B, Halasz A, Hawari J (2006) Effect of iron(III), humic acids and anthraquinone-2,6-disulfonate on biodegradation of cyclic nitramines by *Clostridium* sp. EDB2. *J Appl Microbiol* 100(3):555-563 doi:10.1111/j.1365-2672.2005.02819.x
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 6(8):1621-1624
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *P Natl Acad Sci USA* 108:4516-4522
- Chen DZ, Jin XJ, Chen J, Ye JX, Jiang NX, Chen JM (2016) Intermediates and substrate interaction of 1,4-dioxane degradation by the effective metabolizer *Xanthobacter flavus* DT8. *Int Biodeter Biodegr* 106:133-140 doi:10.1016/j.ibiod.2015.09.018
- Chu MYJ, Bennett P, Dolan M, Hyman M, Anderson R, Bodour A, Peacock A (2009) Using aerobic cometabolic biodegradation and groundwater recirculation to treat 1,4-dioxane and co-contaminants in a dilute plume. Paper presented at the Tenth International Conference on the Remediation of Chlorinated and Recalcitrant Compounds, Baltimore, Maryland,
- Cupples AM (2008) Real-time PCR quantification of *Dehalococcoides* populations: Methods and applications. *J Microbiol Methods* 72(1):1-11 doi:<http://dx.doi.org/10.1016/j.mimet.2007.11.005>

- Cupples AM, Spormann AM, McCarty PL (2003) Growth of a *Dehalococcoides*-like microorganism on vinyl chloride and *cis*-dichloroethene as electron acceptors as determined by competitive PCR. *Appl Environ Microbiol* 69(2):953-959
- Deng DY, Li F, Wu C, Li MY (2018) Synchronic biotransformation of 1,4-dioxane and 1,1-dichloroethylene by a gram-negative propanotroph *Azoarcus* sp DD4. *Environ Sci Tech Lett* 5(8):526 doi:10.1021/acs.estlett.8b00312
- DeRosa CT, Wilbur S, Holler J, Richter P, Stevens YW (1996) Health evaluation of 1,4-dioxane. *Toxicol & Indust Health* 12(1):1-43
- Freedman DL, Gossett JM (1989) Biological reductive dechlorination of tetrachloroethylene and trichloroethylene to ethylene under methanogenic conditions. *Appl Environ Microbiol* 55(9):2144-2151
- Gedalanga P, Madison A, Miao Y, Richards T, Hatton J, DiGuseppi WH, Wilson J, Mahendra S (2016) A multiple lines of evidence framework to evaluate intrinsic biodegradation of 1,4-dioxane. *Remediation* 27(1):93-114 doi:10.1002/rem.21499
- Hand S, Wang BX, Chu KH (2015) Biodegradation of 1,4-dioxane: Effects of enzyme inducers and trichloroethylene. *Sci Total Environ* 520:154-159
- He J, Ritalahti KM, Yang K-L, Koeningsberg SS, Löffler FE (2003) Detoxification of vinyl chloride to ethene coupled to growth of an anaerobic bacterium. *Nature* 424:62-65
- Holliger C, Wohlfarth G, Diekert G (1998) Reductive dechlorination in the energy metabolism of anaerobic bacteria. *Fems Microbiol Rev* 22(5):383-398
- Huang HL, Shen DS, Li N, Shan D, Shentu JL, Zhou YY (2014) Biodegradation of 1,4-dioxane by a novel strain and its biodegradation pathway. *Water Air Soil Poll* 225(9)
- Hunkeler D, Meckenstock RU, Lollar BS, Schmidt TC, Wilson JT (2008) A guide for assessing biodegradation and source identification of organic ground water contaminants using compound specific isotope analysis (CSIA). USEPA, Oklahoma
- Inoue D, Tsunoda T, Sawada K, Yamamoto N, Saito Y, Sei K, Ike M (2016) 1,4-Dioxane degradation potential of members of the genera *Pseudonocardia* and *Rhodococcus*. *Biodegrad* 27(4-6):277-286 doi:10.1007/s10532-016-9772-7
- Inoue D, Tsunoda T, Yamamoto N, Ike M, Sei K (2018) 1,4-Dioxane degradation characteristics of *Rhodococcus aetherivorans* JCM 14343. *Biodegrad* 29(3):301-310 doi:10.1007/s10532-018-9832-2
- Isaka K, Udagawa M, Sei K, Ike M (2016) Pilot test of biological removal of 1,4-dioxane from a chemical factory wastewater by gel carrier entrapping *Afipia* sp strain D1. *J Hazard Mater* 304:251-258 doi:10.1016/j.jhazmat.2015.10.066

- Kampfer P, Kroppenstedt RM (2004) *Pseudonocardia benzenivorans* sp nov. Int J Syst Evol Micr 54:749-751 doi:10.1099/ijs.0.02825-0
- Kim YM, Jeon JR, Murugesan K, Kim EJ, Chang YS (2009) Biodegradation of 1,4-dioxane and transformation of related cyclic compounds by a newly isolated *Mycobacterium* sp PH-06. Biodegrad 20(4):511-519
- Kohlweyer U, Thiemer B, Schrader T, Andreesen JR (2000) Tetrahydrofuran degradation by a newly isolated culture of *Pseudonocardia* sp strain K1. Fems Microbiol Lett 186(2):301-306 doi:DOI 10.1111/j.1574-6968.2000.tb09121.x
- 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 Microb 79(17):5112-5120 doi:10.1128/Aem.01043-13
- Lan RS, Smith CA, Hyman MR (2013) Oxidation of cyclic ethers by alkane-grown *Mycobacterium vaccae* JOB5. Remediation 23(4):23-42 doi:10.1002/rem.21364
- Li MY, Fiorenza S, Chatham JR, Mahendra S, Alvarez PJJ (2010) 1,4-Dioxane biodegradation at low temperatures in Arctic groundwater samples. Water Res 44(9):2894-2900 doi:10.1016/j.watres.2010.02.007
- Lippincott D, Streger SH, Schaefer CE, Hinkle J, Stormo J, Steffan RJ (2015) Bioaugmentation and propane biosparging for in situ biodegradation of 1,4-dioxane. Ground Water Monit R 35(2):81-92
- Löffler FE, Yan J, Ritalahti KM, Adrian L, Edwards EA, Konstantinidis KT, Müller JA, Fullerton H, Zinder SH, Spormann AM (2013) *Dehalococcoides mccartyi* gen. nov., sp. nov., obligately organohalide-respiring anaerobic bacteria relevant to halogen cycling and bioremediation, belong to a novel bacterial class, *Dehalococcoidia* classis nov., order *Dehalococcoidales* ord. nov. and family *Dehalococcoidaceae* fam. nov., within the phylum *Chloroflexi*. Int J Syst Evol Microbiol 63(Pt 2):625-635
- Lu ZD, Sun WJ, Li C, Ao XW, Yang C, Li SM (2019) Bioremoval of non-steroidal anti-inflammatory drugs by *Pseudoxanthomonas* sp. DIN-3 isolated from biological activated carbon process. Water Res 161:459-472 doi:10.1016/j.watres.2019.05.065
- Mahendra S, Alvarez-Cohen L (2005) *Pseudonocardia dioxanivorans* sp nov., a novel actinomycete that grows on 1,4-dioxane. Int J Syst Evol Micr 55:593-598 doi:DOI 10.1099/ijs.0.63085-0
- Mahendra S, Alvarez-Cohen L (2006) Kinetics of 1,4-dioxane biodegradation by monooxygenase-expressing bacteria. Environ Sci Technol 40(17):5435-5442 doi:10.1021/es060714v

- Major DW, McMaster ML, Cox EE, Edwards EA, Dworatzek SM, Hendrickson ER, Starr MG, Payne JA, Buonamici LW (2002) Field demonstration of successful bioaugmentation to achieve dechlorination of tetrachloroethene to ethene. *Environ Sci Technol* 36(23):5106-5116
- Masuda H (2009) Identification and characterization of monooxygenase enzymes involved in 1,4-dioxane degradation in *Pseudonocardia* sp. strain ENV478, *Mycobacterium* sp. strain ENV421, and *Nocardia* sp. strain ENV425. Rutgers The State University of New Jersey and University of Medicine and Dentistry of New Jersey
- Matsui R, Takagi K, Sakakibara F, Abe T, Shiiba K (2016) Identification and characterization of 1,4-dioxane-degrading microbe separated from surface seawater by the seawater-charcoal perfusion apparatus. *Biodegrad* 27(2-3):155-163 doi:10.1007/s10532-016-9763-8
- Maymó-Gatell X, Chien Y-T, Gossett JM, Zinder SH (1997) Isolation of a bacterium that reductively dechlorinates tetrachloroethene to ethene. *Science* 276(6):1568-1571
- Mohr T, Stickney J, DiGiuseppi W (2010) *Environmental Investigation and Remediation: 1,4-Dioxane and Other Solvent Stabilizers*. CRC Press, Boca Raton, ML
- Naumoff DG, Dedysh SN (2018) Bacteria from poorly studied phyla as a potential source of new enzymes: beta-galactosidases from *Planctomycetes* and *Verrucomicrobia*. *Microbiology* 87(6):796-805 doi:10.1134/S0026261718060127
- Parales RE, Adamus JE, White N, May HD (1994) Degradation of 1,4-dioxane by an *Actinomycete* in pure culture. *Appl Environ Microb* 60(12):4527-4530
- Parks DH, Tyson GW, Hugenholtz P, Beiko RG (2014) STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30(21):3123-3124
- Pooley KE, Blessing M, Schmidt TC, Haderlein SB, MacQuarrie KTB, Prommer H (2009) Aerobic biodegradation of chlorinated ethenes in a fractured bedrock aquifer: Quantitative assessment by compound-specific isotope analysis (CSIA) and reactive transport modeling. *Environ Sci Technol* 43(19):7458-7464 doi:10.1021/es900658n
- Pornwongthong P (2014) Stable isotopic and molecular biological tools to validate biodegradation of 1,4-dioxane. University of California, Los Angeles
- Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Gloeckner FO (2007) SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nuc Acids Res* 35(21):7188-7196 doi:10.1093/nar/gkm864
- Pugazhendi A, Banu JR, Dhavamani J, Yeom IT (2015) Biodegradation of 1,4-dioxane by *Rhodanobacter* AYS5 and the role of additional substrates. *Ann Microbiol* 65(4):2201-2208 doi:10.1007/s13213-015-1060-y
- Qian Y, Chen K, Liu Y, Li J (2019) Assessment of hexachlorocyclohexane biodegradation in contaminated soil by compound-specific stable isotope analysis. *Environ Pollut* 254:113008

Rodriguez FJB (2016) Evaluation of 1,4-dioxane biodegradation under aerobic and anaerobic conditions. Clemson University

Rosell M, Häggblom MM, Richnow H-H (2007) Compound-specific isotope analysis (CSIA) to characterise degradation pathways and to quantify in-situ degradation of fuel oxygenates and other fuel-derived contaminants. In: Barceló D (ed) Fuel Oxygenates. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 99-119

Schloss PD (2009) A high-throughput DNA sequence aligner for microbial ecology studies. PLOS One 4(12) doi:10.1371/journal.pone.0008230

Schloss PD (2013) MiSeq SOP. PUblisher. [http://www.mothur.org/wiki/MiSeq\\_SOP](http://www.mothur.org/wiki/MiSeq_SOP) Accessed Sept 15th 2013 2013

Sei K, Miyagaki K, Kakinoki T, Fukugasako K, Inoue D, Ike M (2013a) Isolation and characterization of bacterial strains that have high ability to degrade 1,4-dioxane as a sole carbon and energy source. Biodegrad 24(5):665-674

Sei K, Oyama M, Kakinoki T, Inoue D, Ike M (2013b) Isolation and characterization of tetrahydrofuran- degrading bacteria for 1,4-dioxane-containing wastewater treatment by co-metabolic degradation. J Water Environ Technol 11:11-19 doi:10.2965/jwet.2013.11

Shen W, Chen H, Pan S (2008) Anaerobic biodegradation of 1,4-dioxane by sludge enriched with iron-reducing microorganisms. Bioresource Technol 99(7):2483-2487 doi:10.1016/j.biortech.2007.04.054

Stefan MI, Bolton JR (1998) Mechanism of the degradation of 1,4-dioxane in dilute aqueous solution using the UV hydrogen peroxide process. Environ Sci Technol 32(11):1588-1595

Steffan RJ (2007) ER-1422: Biodegradation of 1,4-dioxane. vol Final Report. SERDP and ESTCP

Steffan RJ, McClay K, Vainberg S, Condee CW, Zhang DL (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 Microb 63(11):4216-4222

Steffan RJ, McClay KR, Masuda H, Zylstra GJ (2007) Biodegradatdion of 1,4-dioxane. Strategic Environmental Research and Development Program: ER-1422 Final Report

Steffan RJ, Vainberg S (2013) Production and handling of *Dehalococcoides* bioaugmentation cultures. In: Stroo HF, Leeson A, Ward CW (eds) Bioaugmentation for groundwater remediation. SERDP and ESTCP Remediation Technology. Springer, New York

Stringfellow WT, Alvarez-Cohen L (1999) Evaluating the relationship between the sorption of PAHs to bacterial biomass and biodegradation. Water Res 33(11):2535-2544 doi:Doi 10.1016/S0043-1354(98)00497-7

- Sun BZ, Ko K, Ramsay JA (2011) Biodegradation of 1,4-dioxane by a *Flavobacterium*. *Biodegrad* 22(3):651-659 doi:10.1007/s10532-010-9438-9
- Sung Y, Ritalahti KM, Apkarian RP, Löffler FE (2006) Quantitative PCR confirms purity of strain GT, a novel trichloroethene-to-ethene-respiring *Dehalococcoides* isolate. *Appl Environ Microbiol* 72(3):1980-1987
- Tanaka Y, Matsuzawa H, Tamaki H, Tagawa M, Toyama T, Kamagata Y, Mori K (2017) Isolation of novel bacteria including rarely cultivated phyla, *Acidobacteria* and *Verrucomicrobia*, from the roots of emergent plants by simple culturing method. *Microbes Environ* 32(3):288-292 doi:10.1264/jsme2.ME17027
- Vainberg S, Condee CW, Steffan RJ (2009) Large-scale production of bacterial consortia for remediation of chlorinated solvent-contaminated groundwater. *J Ind Microbiol Biot* 36(9):1189-1197 doi:10.1007/S10295-009-0600-5
- Vainberg S, McClay K, Masuda H, Root D, Condee C, Zylstra GJ, Steffan RJ (2006) Biodegradation of ether pollutants by *Pseudonocardia* sp strain ENV478. *Appl Environ Microb* 72(8):5218-5224
- Wang Y, Huang Y, Huckins JN, Petty JD (2004) Compound-specific carbon and hydrogen isotope analysis of sub-parts per billion level waterborne petroleum hydrocarbons. *Environ Sci Technol* 38(13):3689-3697 doi:10.1021/es035470i
- Whittenbury R, Phillips KC, Wilkinso.Jf (1970) Enrichment, isolation and some properties of methane-utilizing bacteria. *J Gen Microbiol* 61:205-& doi:Doi 10.1099/00221287-61-2-205
- Willems A (2014) The Family *Comamonadaceae*. In: Eugene Rosenberg (Editor-in-Chief) EFD, Stephen Lory, Erko Stackebrandt and Fabiano Thompson (Eds.) (ed) *The prokaryotes: Alphaproteobacteria and betaproteobacteria*. 4th edn. Springer-Verlag Berlin Heidelberg
- Yamamoto N, Saito Y, Inoue D, Sei K, Ike M (2018) Characterization of newly isolated *Pseudonocardia* sp N23 with high 1,4-dioxane-degrading ability. *J Biosci Bioeng* 125(5):552-558 doi:10.1016/j.jbiosc.2017.12.005
- Zenker MJ, Borden RC, Barlaz MA (2003) Occurrence and treatment of 1,4-dioxane in aqueous environments. *Env Eng Sci* 20(5):423-432

## Chapter 4 Analysis of Functional Genes Associated with Organic Contaminant Biodegradation in Agricultural Soils using Shotgun Sequencing Data

### 4.1 Abstract

Organic contaminants in soil and groundwater represent a significant problem as many are potentially hazardous to human health. Bioremediation can be an effective strategy to remediate these chemicals in both soil and groundwater. The success of bioremediation is often linked to the abundance of functional genes present in the soil that are associated with the degradation process. Although previous studies have explored these genes in contaminated environmental samples, limited information is available on the presence and abundance of these genes in uncontaminated / agricultural soils. Further, agricultural soils are known to contain diverse microbial communities both phylogenetically and functionally. In this study, the objective was to examine a set of genes associated with organic contaminant degradation in four uncontaminated (agricultural) soils. The abundance and diversity of *benA*, *bph*, *dbfA*, *dxnA*, *etnC*, *etnE*, *ppaH*, *npaH*, *vcrA*, *xenA*, *xenB* and *xplA* were investigated using protein sequences from the Functional Gene Pipeline and Repository (FunGene). A comparison of relative abundance of genes between four different soils indicated *dbfA*, *etnE*, *npaH*, *xenA* and *xplA* were more abundant in all of the soil metagenomes. Also, *vcrA* illustrated the the lowest relative abundance level. The phylogenetic trees created indicate many genera may potentially be associated with each gene including *Pseudomonas*, *Rhodococcus*, *Mycobacterium* and *Nocardioides*. From these, some strains are well studied and are known to be involved in the biodegradation of organic contaminants and others are potentially new genera that may be associated with the biodegradation of the targeted group of contaminants.

## 4.2 Introduction

The contamination of soil and groundwater with organic contaminants including hazardous hydrocarbons has resulted in public health issues. Many of these organic contaminants are known or potential carcinogens or pose toxic risks to humans including polychlorinated biphenyls (PCBs), polychlorinated dioxins and polychlorinated dibenzofurans (Brzuzy and Hites 1996; Iyer et al. 2016; Mayes et al. 1998; Van den Berg et al. 2006). This, in turn, demonstrates a need for better remediation strategies. Aerobic bacteria growing on organic contaminants have remained a topic of interest due to their role in bioremediation (Pieper 2005). The catabolic genes responsible for the production of degradative enzymes have been studied to investigate the biodegradation potential of such bacteria (Nyssönen et al. 2008; Widada et al. 2002). The native populations at contaminated sites could dictate the extent of bioremediation potential at the site (Widada et al. 2002) and this has led to studies focusing on functional genes associated with degradation.

The current study focuses primarily on the presence and abundance of genes in soils associated with organic contaminant biodegradation. To this aim, the abundance and diversity of the following 12 genes were investigated: *benA*, *bph*, *dbfA*, *dxnA*, *etnC*, *etnE*, *ppaH*, *npaH*, *vcrA*, *xenA*, *xenB* and *xplA*. For this, protein sequences were gathered from the Functional Gene Pipeline and Repository (FunGene). FunGene includes databases of multiple functional genes and proteins along with tools for further analysis with the ability to download aligned sequences (Fish et al. 2013).

The benzoate dioxygenase complex encoded by the catabolic genes *benABCDK* has been previously studied and is known to be associated with benzoate and biphenyl catabolism (Kitagawa et al. 2001; Patrauchan et al. 2005). Benzoate dioxygenase catalyzes the conversion of benzoate to catechol (Reiner 1971). The gene *benA* is associated with the biodegradation of benzoate under aerobic conditions to its non-aromatic form, cis-diol (Neidle et al. 1987). This compound is an intermediate in anaerobic metabolism of aromatic compounds but more commonly, benzoates and chlorobenzoates are intermediate metabolites of biphenyl and PCB degradation (Hernandez et al. 1991; Sondossi et al. 1992).

Biphenyl dioxygenases, encoded by *bph*, are known to be involved in (PCB)/biphenyl degradation (Erickson and Mondello 1992; Kumamaru et al. 1998; Ohtsubo et al. 2001; Seeger et al. 1995). The gene cluster (*bph* genes) was initially cloned from *Pseudomonas pseudoalcaligenes* KF707 (Furukawa and Miyazaki 1986). Many bacterial strains belonging to *Proteobacteria* and *Actinobacteria* can utilize biphenyl as a sole carbon source and co-metabolically utilize PCBs (Bedard et al. 1986; Deneff et al. 2004). PCBs have been widely used compounds in industrial applications and are known for their toxicity, mutagenicity, health risks and potential bioaccumulation (ATSDR 2000; Faroon and Ruiz 2016). Concerns about their persistence has led to many research studies on their biodegradation potential (Abraham et al. 2002; Kumamaru et al. 1998).

In general, there are three dioxygenases, encoded by *dxnA*, *dbfA* and *carA*, involved in catalysis of the initial reaction of dioxin catabolic pathway all of which attack at the angular position (Field and Sierra-Alvarez 2008). The presence and diversity of both *dbfA* and *dxnA* are included

in current work and are associated with angular dioxygenation of dibenzofuran and dioxin (Kasuga et al. 2013; Penton et al. 2013). The *dbfA* gene cluster encoding angular dioxygenases were amplified and sequenced previously (Peng et al. 2013). Although no extensive study has been conducted on dibenzofuran toxicity to humans, there is a low risk based on animal studies (EPA 2007). Chlorinated dioxins are known to consist of multiple congeners that are extremely toxic and sometimes the non-toxic lower chlorinated dioxins are biotransformation products of the more toxic versions (Field and Sierra-Alvarez 2008). This has led to increased interest in their biodegradation potential.

The enzymes alkene monooxygenases (encoded by *etnC*) and epoxyalkane: coenzyme M transferase (encoded by *etnE*) are involved in aerobic degradation of contaminants including vinyl chloride (VC), ethene and propene (Allen et al. 1999; Coleman and Spain 2003; Ensign 2001; Smith et al. 1999). Both ethene-assimilating bacteria and VC-assimilating bacteria utilize these enzymes in the initial process of aerobic biodegradation (Coleman and Spain 2003). VC, a reduction product of the groundwater contaminants tetrachloroethene (PCE) and trichloroethene (TCE), is a known human carcinogen and has been well-studied (ATSDR 2006; Wagoner 1983). This contaminant is highly prevalent at chlorinated solvent contaminated sites (Coleman et al. 2002a).

The degradation of polycyclic aromatic hydrocarbons (PAHs) depends on multiple genes and gene clusters grouped together into homologies including *pah*, *nah*, *ndo* and *dox* (Lu et al. 2021). Naphthalene dioxygenase and phthalate dioxygenase (*npaH* and *ppaH*) are associated with the degradation of contaminants including PAHs (Fang et al. 2013; Peng et al. 2008; Peng et al.

2010; Resnick et al. 1996; Selifonov et al. 1996). The PAHs are widespread in the environment and are widely regarded as hazardous due to their ability to induce mutagenicity and carcinogenicity (Dipple 1985; Talaska et al. 1996). Thus, the use of microbial action for the remediation of these contaminants paves a better way to dispose of them compared to disposal at waste sites. The ring hydroxylating oxygenases play a crucial role by catalyzing the first step of the degradation process (Peng et al. 2010). One example would be naphthalene dioxygenase, which is helpful in dehydroxylating low molecular weight PAHs (Peng et al. 2008). Phthalates are used commonly in the production of plastics, textiles, pesticides, munitions, cosmetics and insect repellents (Peakall 1975). Due to widespread usage, there is a growing concern about their release into the environment and the effects on human health (Kinzell et al. 1979).

Reductive dehalogenases are responsible for organohalide respiration and are associated with reductive dechlorination of chlorinated solvents, such as PCE and TCE (Löffler et al. 2013). The gene encoding for VC reductive dehalogenase, *vcrA*, is well-studied and plays an important role in VC remediation (Müller et al. 2004; van der Zaan et al. 2010). Also, *vcrA* is known to be up-regulated when exposed to TCE, cis-dichloroethylene and VC (Lee et al. 2006).

The enzymes encoded by *xenA*, *xenB* and *xplA* are associated with hexa-hydro-1,3,5-trinitro-1,3,5-triazine (RDX) biodegradation and play a crucial role in the biotransformation of RDX (Fuller et al. 2009; Indest et al. 2007; Rylott et al. 2006; Seth-Smith et al. 2002). Both *xenA* and *xenB* have been sequenced and characterized (Blehert et al. 1999). The gene *xplA* is the most well-studied among the three and has been identified in multiple genera (phylum *Actinobacteria*) including *Rhodococcus*, *Gordonia*, *Williamsia* (all under suborder *Corynebacterineae*) and

*Microbacterium* (under suborder *Micrococccineae*) (Andeer et al. 2009; Bernstein et al. 2011; Coleman et al. 1998; Indest et al. 2007; Nejidat et al. 2008; Rylott et al. 2006; Rylott et al. 2011; Seth-Smith et al. 2008; Seth-Smith et al. 2002; Thompson et al. 2005). Munition manufacturing sites have resulted in the release of munition-related chemicals including RDX in soil and groundwater. The toxicity of this compound is well established in humans and other mammals (Roberts and Hartley 1992).

The current study exploits the vast information available from high throughput shotgun sequencing to examine the occurrence and diversity of this subset of contaminant biodegradation genes in soil derived samples. Soil is well known as an excellent media to explore microbial diversity. The work takes advantage of a previous project (by the same authors) that produced the shotgun sequencing data (Ramalingam and Cupples 2020). Although previous studies by others have examined the occurrence of these genes in environmental samples, knowledge gaps exist concerning their diversity and abundance across different soils. Further, to date, shotgun sequencing has yet to be applied as a tool to investigate these genes between multiple samples. The overall objective of this work was to determine the relative abundance and diversity of twelve genes previously associated with contaminant biodegradation in multiple soil metagenomes. We hypothesized that 1) a large number of genera previously associated with these genes would be identified, 2) some genes would be more abundant and diverse than others, and 3) in some cases, positive correlations would exist between gene abundance values. The aim of the research is to improve our understanding of the biodegradation potential of organic contaminants across soils.

## 4.3 Methods

### 4.3.1 Experimental Setup and DNA Extraction

The microbial communities examined here originated from a previous study on the susceptibility of 1,4-dioxane to biodegradation under aerobic conditions (Ramalingam and Cupples 2020). Briefly, agricultural samples were collected from two locations on the campus of Michigan State University (MSU), East Lansing, Michigan (soils F and G) and two locations at the Kellogg Biological Station, Hickory Corners, Michigan (Treatments 1 and 2, called Soil T1 and T2). Each of the four soil metagenomes included replicated controls (no 1,4-dioxane) and samples (with 1,4-dioxane). Soil F included 2 controls and 3 samples and the other three agricultural soils (G, 1 and 2) included 2 samples and 2 controls. One sample sequencing file (within the set of 4) for soil T2 was deemed unusable since it appeared as an outlier and therefore not used in this study.

### 4.3.2 Shotgun Sequencing and Trimmomatic

Libraries were generated for shotgun sequencing at the Research Technology Support Facility Genomics Core at MSU. Libraries were prepared using the Takara SMARTer ThruPLEX DNA Library Preparation Kit following manufacturer's recommendations. Completed libraries were QC'd and quantified using a combination of Qubit dsDNA HS and Agilent 4200 TapeStation HS DNA1000 assays. The libraries were pooled in equimolar amounts for multiplexed sequencing. The pool was quantified using the Kapa Biosystems Illumina Library Quantification qPCR kit and loaded onto one lane of an Illumina HiSeq 4000 flow cell. Sequencing was performed using HiSeq 4000 SBS reagents in a 2x150 bp paired-end format. Base calling was done by Illumina Real Time Analysis (RTA) v2.7.7 and output of RTA was demultiplexed and converted to FastQ

format with Illumina Bcl2fastq v2.19.1. Trimmomatic was used to remove low quality sequences and Illumina adapters in paired-end mode (Version 0.39) (Bolger et al. 2014).

#### 4.3.3 FunGene Pipeline and DIAMOND

The MSU RDP Functional Gene Pipeline and Repository (FunGene) was used to select a group of genes associated with organic contaminant biodegradation for this study (Fish et al. 2013). For each gene, the sequences were downloaded with a minimum HMM coverage of 70% (95% was used for *xenA* and *xenB* due to the large number of sequences). The protein downloads were dereplicated using the FunGene pipeline. The dereplicated files were used with DIAMOND (double index alignment of next-generation sequencing data) (Buchfink et al. 2015) for aligning two paired output files from Trimmomatic to the twelve protein databases selected from the FunGene database. For each gene, a DIAMOND file was created to which the Trimmomatic files could be aligned.

#### 4.3.4 Relative Abundance of Biodegradation Genes

After the alignment was complete, the DIAMOND files were analyzed using R (Team 2018)(Version 3.5.1) in RStudio (Team 2020)(Version 0.9.24), which included combining the data from the two paired files and deleting duplicated data. RStudio was used to sort and select reads that exhibited an identity of  $\geq 60\%$  and an alignment length  $\geq 49$  amino acids. For each gene, the relative abundance values were calculated using the number of aligned reads divided by the total number of sequences for each sample (determined by Trimmomatic) and the number of sequences in the dereplicated file (from FunGene). Once the relative abundance was

computed, the resulting data was used to generate box and whisker plots along with Principal Component Analysis using Microsoft Excel (XLSTAT).

#### 4.3.5 Statistical Analysis: Linear Regression and EstimateS Diversity Indices

A correlation analysis between the different genes was performed using RStudio. One-way Analysis of Variance (ANOVA) was used as a parametric method for comparing the independent populations from the four soils. First, the data was checked for normality using the Shapiro Wilk test ( $p > 0.05$  shows normal distribution of data) with the R package rstatix (Kassambara 2020). The data was also checked for equal variance using Barlett's test ( $p > 0.05$  for equal variance) with R package stats (Team 2018). Based on the ANOVA tests, if the samples exhibited significant differences between the means ( $p < 0.05$ ), the results were further analyzed with Tukey's 'Honest Significant Difference' (HSD) method. For the samples where the data did not satisfy the normal distribution required for ANOVA, a non-parametric approach was used. For non-normal data, Kruskal Wallis One Way Analysis of Variance was used as an equivalent for the ANOVA test. The R package stats was used to carry out the test and the data was checked for homogeneity of variances using Levene's test with the R package car (Fox and Weisberg 2019). Any significant differences ( $p < 0.05$ ) were further tested using Dunn's test for multiple comparisons using the R package FSA (Ogle 2020).

Linear regression models were used to obtain information for the regression equations and p-values. The Shapiro Wilk test was used again to test for normality by checking the entire relative abundance dataset followed by testing for equal variance using F test (null hypothesis is that the ratio of variances of the populations is equal). For genes that satisfied both the normality and

equal variance, plots were made. For those showing a significant relationship ( $p < 0.05$ ), regression line plots were created using R package ggplot2 (Wickham 2016) with 95% CI and Pearson's coefficient being calculated. Non-parametric Spearman's Rank test was used for data sets that did not demonstrate normality. Spearman's correlation was performed using cor.test function with the R stats package.

EstimateS (Statistical Estimation of Species Richness and Shared Species from Samples) enabled the calculation of diversity indices (Version 9.1.0) (Colwell and Elsensohn 2014). This software computes biodiversity statistics, estimators, and indices based on sampling data. All computations were performed with sample-based incidence or abundance data settings. Fisher's alpha, Shannon and Inverse Simpson indices were calculated along with Chao1 and Chao2 bias corrections. The outputs were exported and analyzed in Microsoft Excel.

#### 4.3.6 Phylogenetic Trees Based on Contigs

Shotgun sequences processed by Trimmomatic were assembled into larger contiguous segments (contigs) using Megahit (Li et al. 2015; Li et al. 2016) (Version 1.2.4) with the pair end plus single end option (minimum and maximum kmer size were 27 and 127 with a kmer size step of 10). The final contigs output files were used further for creating phylogenetic trees. For this, DIAMOND files were created using 95% HMM from the FunGene database for each of the 12 genes (70% HMM for *xplA* since there were no hits at 95%). A higher percentage HMM was selected here (compared to the analysis of the reads) to produce alignments that were more specific to each target gene. The top 100 accession numbers for each gene (72 for *npaH* and none for *vcrA*) were taken from the files generated using Rstudio (with identity of  $\geq 60\%$  and an

alignment length  $\geq 49$  amino acids) and used in NCBI Constraint-based Multiple Alignment Tool (COBALT) (Papadopoulos and Agarwala 2007). After alignment, only the unique sequences were chosen and re-aligned. The files were downloaded and used to create the phylogenetic trees in MEGA X (Kumar et al. 2018) (Version 10.0.5). The trees were based on the most abundant sequences across all samples for all genes. The trees were constructed using Maximum Likelihood method based on the Jones-Taylor-Thornton (JTT) model (Jones et al. 1992). A Nearest-Neighbor-Interchange (NNI) ML Heuristic Method was used within MEGA X (Kumar et al. 2018).

## 4.4 Results

### 4.4.1 Relative abundance of Biodegradation Genes in Soils

Reads aligned to all 12 genes in the four agricultural soils, as shown in the box and whisker plot (Figure 4.1a). Genes including *dbfA*, *etnE*, *npaH*, *xenA* and *xplA* were more abundant than the rest in all four soils (Figure 4.1a). The lowest abundance levels were observed for *vcrA* and *xenB*. A PCA plot, created using soil characteristics and gene relative abundance values as active variables and soil samples as active observations, illustrates sample clusters suggesting similarities between various samples/genes (Figure 4.1b). For example, both soil F and soil G formed separate clusters (bottom and top left of plot), with soil G positively correlating with three soil variables (sand, potassium and phosphorus) and soil F illustrating no strong correlations. The majority of the soil T1 values clustered together, whereas soil T2 values did not form a strong cluster. The majority of the genes clustered together, except for *etnC* and *vcrA*. Further, the remaining soil variables positively correlated with values from soil T1.

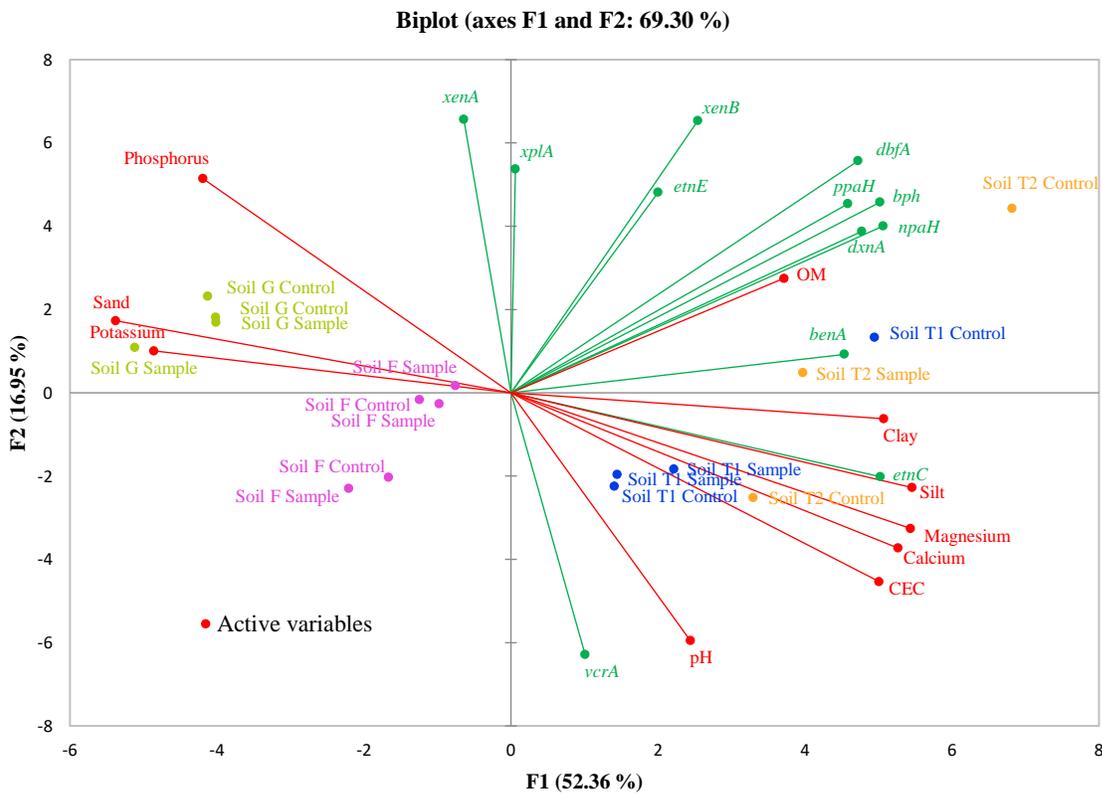
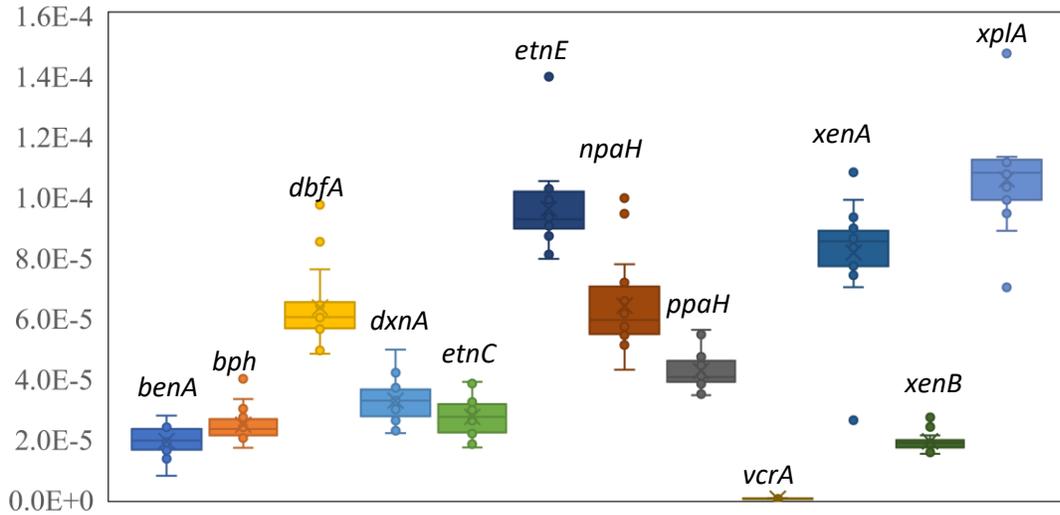


Figure 4.1. Box and whisker plot of relative abundance of genes (A) and Principle Component Analysis of the genes across the four agricultural soils (with or without added 1,4-dioxane) (B).

#### 4.4.2 Statistical Tests - ANOVA and Kruskal Wallis

The ANOVA or Kruskal-Wallis test was used to determine any significant differences in gene abundance between soils. Barlett's test indicated 8 genes exhibited equal variance (all except *etnE*, *vcrA*, *xenA* and *xplA* have  $p > 0.05$ ). Normality tests for these 8 genes indicated only *dxnA* and *xenB* had normal data for all 4 soils. The other genes had at least 1 soil with non-normal data. The Tukey HSD test indicated significant differences between the soils and are represented by different letter notations on the Figure 4.2. For the non-normal data, Kruskal Wallis test was carried out. The variances were checked using Levene's test and all but *xenA* indicated equal variance. The Kruskal Wallis test indicated significant differences for *etnE* and *xplA* and this was further investigated using Dunn's test for multiple comparisons. However, no significant differences between the soils were observed for *etnE* and *xplA*. The values for *xenA* did not meet the statistical assumptions made for ANOVA or Kruskal-Wallis test since the data showed unequal variance (thus, no letter notation was added) (Figure 4.2). The highest abundance values were noted for *dbfA*, *etnE*, *npaH*, *xenA* and *xplA* and the least abundant values were for *vcrA* in all four soils. For the relative abundance calculations, 3 samples of soil T2 (n=3) were included along with 4 samples for soil T1 (n=4), 4 samples for soil G (n=4) and 5 samples of soil F (n=5). The genes *benA*, *bph*, *dbfA*, *dxnA* and *npaH* indicated significant differences between soil G and soil T2. The gene *etnC* indicated significant differences between all the soil pairs except between two soil pairs (a) soil F and soil T1 and (b) soil T1 and soil T2. All the other genes showed no significant differences between the soils (Figure 4.2).

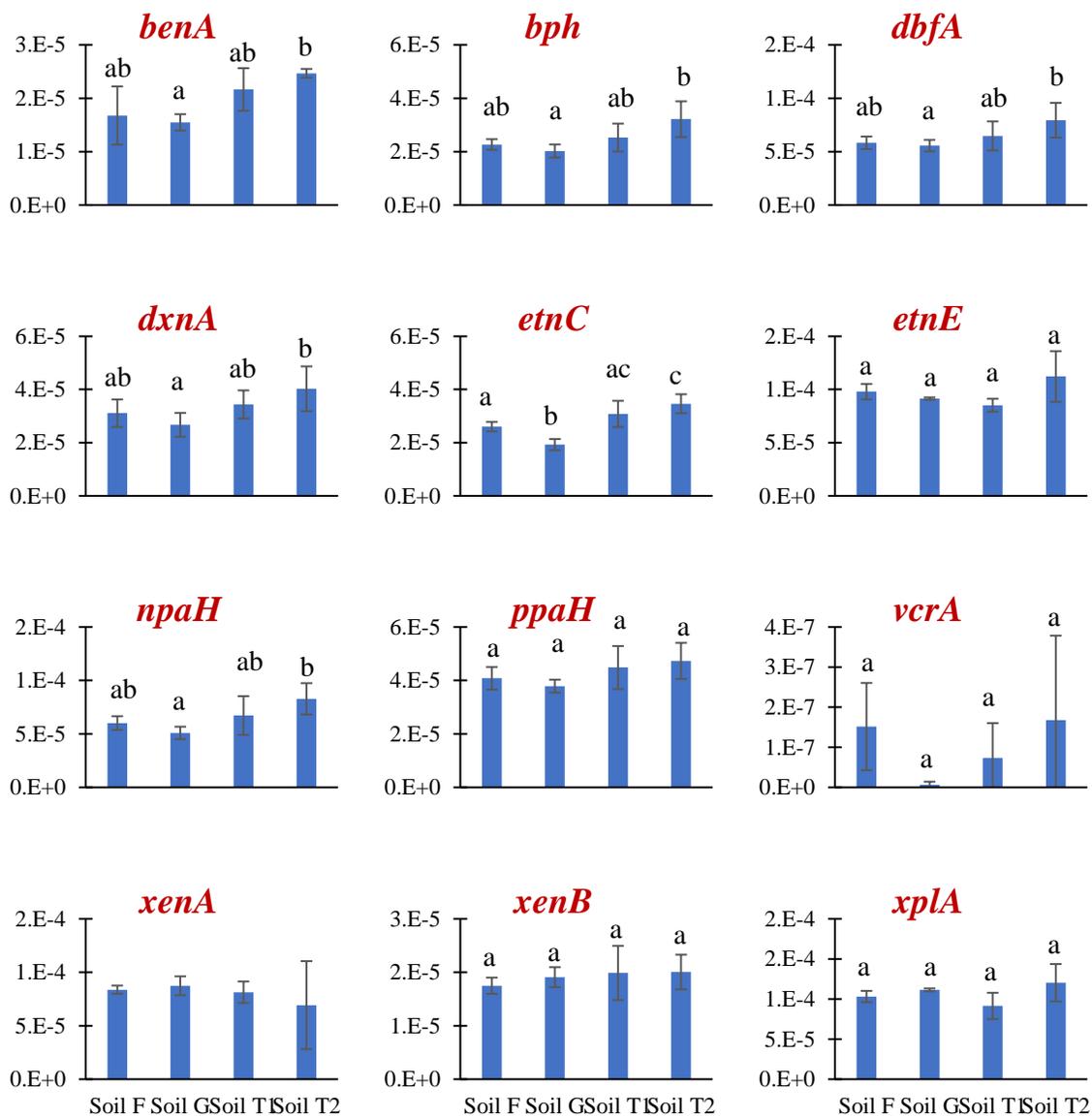


Figure 4.2. Average relative abundance values (% as determined by DIAMOND) for each soil (n=5 for soil F, n=4 for soil G and soil 1 and n=3 for soil 2) with standard deviations illustrated with the error bars. Values that are statistically significantly different (ANOVA or Kruskal-Wallis test,  $p < 0.05$ ) are shown with different letters. Letters are missing for *xenA* because the statistical assumptions were not met for either test (unequal variance). Note, all y-axis have different scales.

#### 4.4.3 Linear Regression of Data

A linear regression of the relative abundance of genes was performed using RStudio. Shapiro's test showed the 4 genes *benA*, *dxnA*, *etnC* and *ppaH* to have normal data, when all soil datasets were combined ( $p > 0.05$ ) (Appendix Table 4.1) and they were further tested by comparing variances using F test (Appendix Table 4.2). All of them showed equal variance and a linear model was plotted for each with the regression equation, p-values and F-test ratio of variances listed in the Appendix Table 4.2. All of the p-values for the linear model equations were below 0.05. Pearson's Correlation coefficient obtained for each regression varied from 0.55 to 0.88 (Appendix Table 4.2).

Eight genes including *bph*, *dbfA*, *etnE*, *npaH*, *vcrA*, *xenA*, *xenB* and *xplA* showed non-normal data and Spearman's rank test ( $p < 0.05$ ) was used to determine correlations that were statistically significant. The Spearman's rho values calculated based on comparisons between genes, show high correlation for *bph* vs *dbfA*, *bph* vs *npaH*, *dbfA* vs *npaH* and *xenA* vs *xenB*. A scatterplot comparing the relative abundance of the 8 genes across all the samples are shown (Figure 4.3). The resulting p-values and the Spearman's rho values are listed in Appendix Table 4.3 and 4.4 respectively.

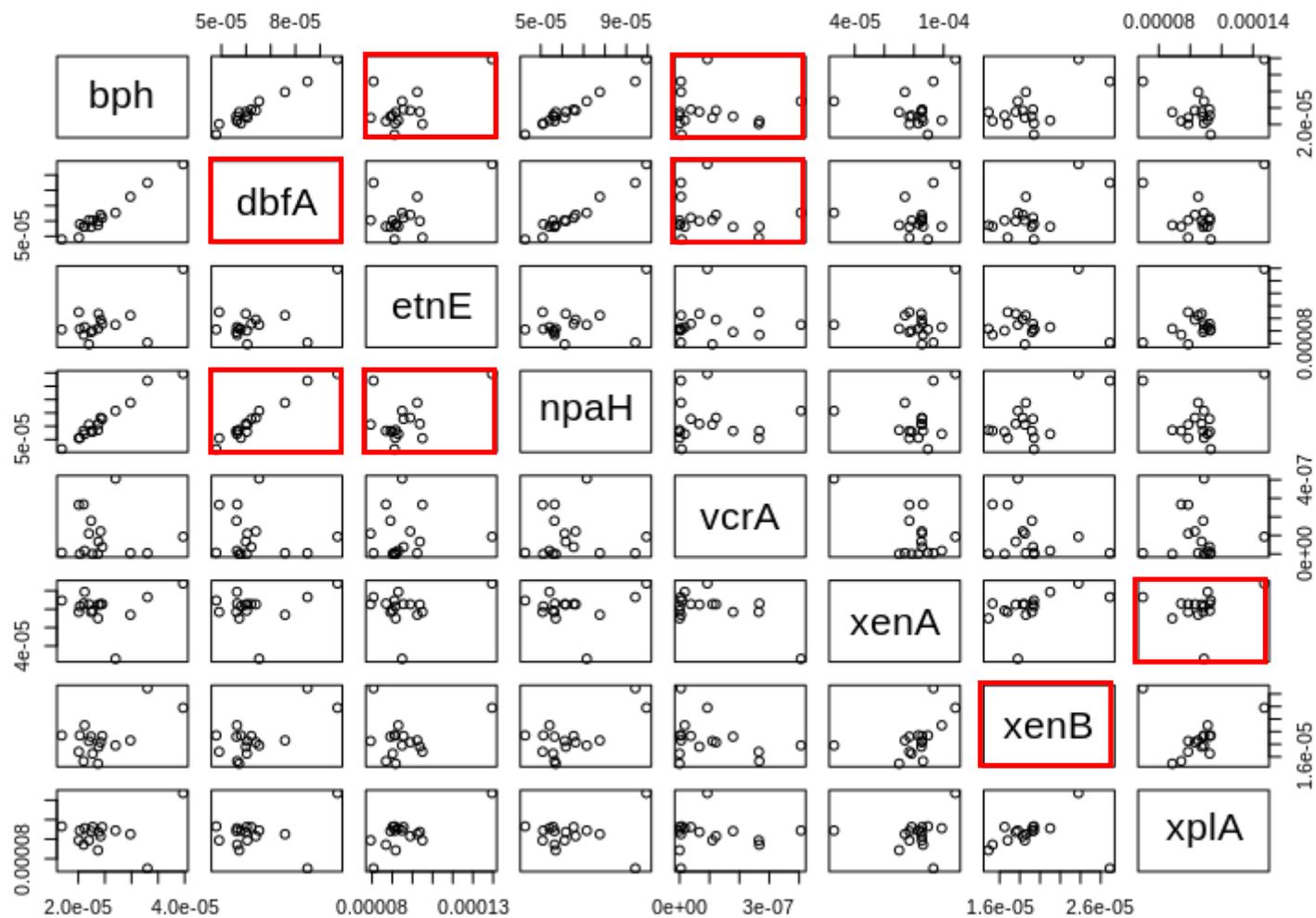


Figure 4.3. Scatterplots comparing relative abundance values of 8 genes (non-normal data) across all samples. Correlations that were statistically significant (Spearman's rank test,  $p < 0.05$ ) are boxed in red.

The regression line plots created for the genes with normal relative abundance data and equal variance indicate a positive relationship between the % relative abundance of each gene plotted against that of the other genes (Figure 4.4). The relative abundance % of *benA* vs *dxnA* indicates a significant relationship with p value of 0.02 (since p value < 0.05). Similarly, the relationships between the others including *benA* vs *etnC*, *benA* vs *ppaH*, *dxnA* vs *etnC*, *dxnA* vs *ppaH* and *etnC* vs *ppaH* also demonstrate significant relationships with p values < 0.05. Also, regression line plots with 95% CI indicate a higher slope and higher Pearson's correlation coefficient for *dxnA* vs *ppaH* plot compared to other plots (Figure 4.4). The lowest slope was noted for *etnC* vs *ppaH* linear regression plot.

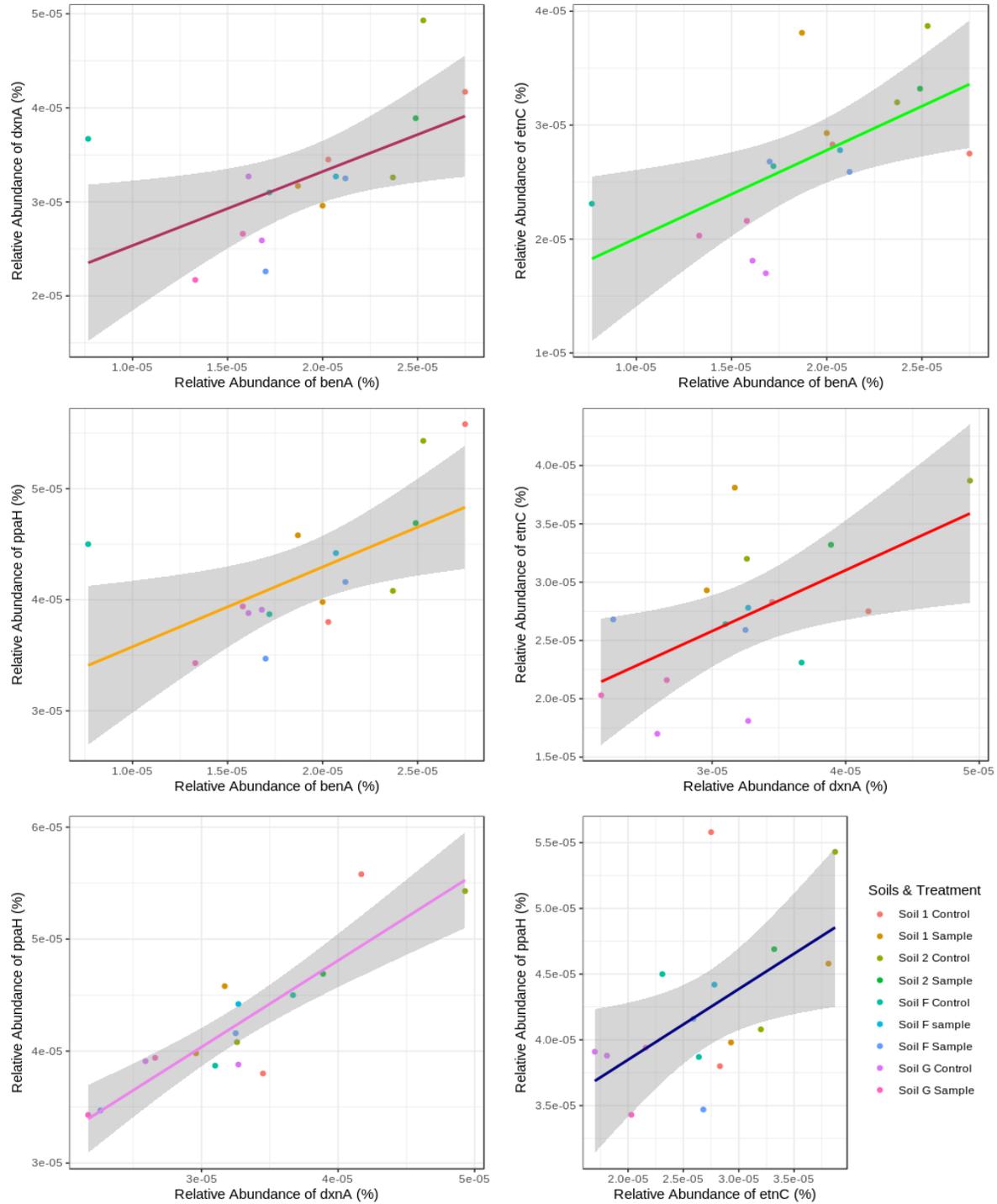


Figure 4.4. Regression line scatter plots with 95% CI for genes with normal relative abundance data and equal variance.

#### 4.4.4 Diversity Indices determined by EstimateS

The results from the Chao1, Chao2, Shannon and Simpson Inverse diversity indices were analyzed in excel (Figures 4.5 and 4.6). The Chao1 index diversity showed no significant differences between the 4 soils for each of the 12 genes except *dbfA* and *dxnA* (Appendix Table 4.5). However, differences between the genes were noted for each soil. The gene *xenB* indicated the highest diversity values among the 12 genes with values over 8000. The lowest diversity index values were observed for *vcrA* gene with values below 500. The genes *benA*, *bph*, *etnE* and *xenA* also indicated higher diversity values over 2000 (Figures 4.5 and 4.6). The Tukey HSD test results indicated a significant difference between soil T1 and soil G for both *dbfA* and *dxnA* (Appendix Table 4.6). The Chao2 diversity did not show any significant differences between the four soils for any of the genes. However, the Chao2 diversity index values followed a similar trend as Chao1 index diversity with differences between the genes for each soil (Figures 4.5 and 4.6). Neither the Tukey HSD nor the Dunn's test could be used appropriately in this case since none of the data satisfied the normal sample distribution criteria (Appendix Tables 4.8 – 4.10). The Shannon index indicated significant differences between all 4 soils in many of the genes including *bph*, *dbfA*, *etnC*, *etnE*, *npaH*, *ppaH*, *vcrA*, *xenB*, *xenA* and *xplA* (Appendix Tables 4.11 - 4.13). The Shannon diversity index indicated some variation between the genes for each soil with values around 5 to 7.5. The gene *xenB* indicated the highest diversity index values. The Simpson Inverse diversity index similarly showed a significant difference between the agricultural soils in all genes except *benA* (Appendix Tables 4.14 – 4.16). Also, differences were observed between the genes for each of the agricultural soils. The gene *benA* indicated the lowest diversity index values and the Tukey HSD and the Dunn's test were not appropriate for *benA* data for Simpson Inverse index (Figure 4.5).

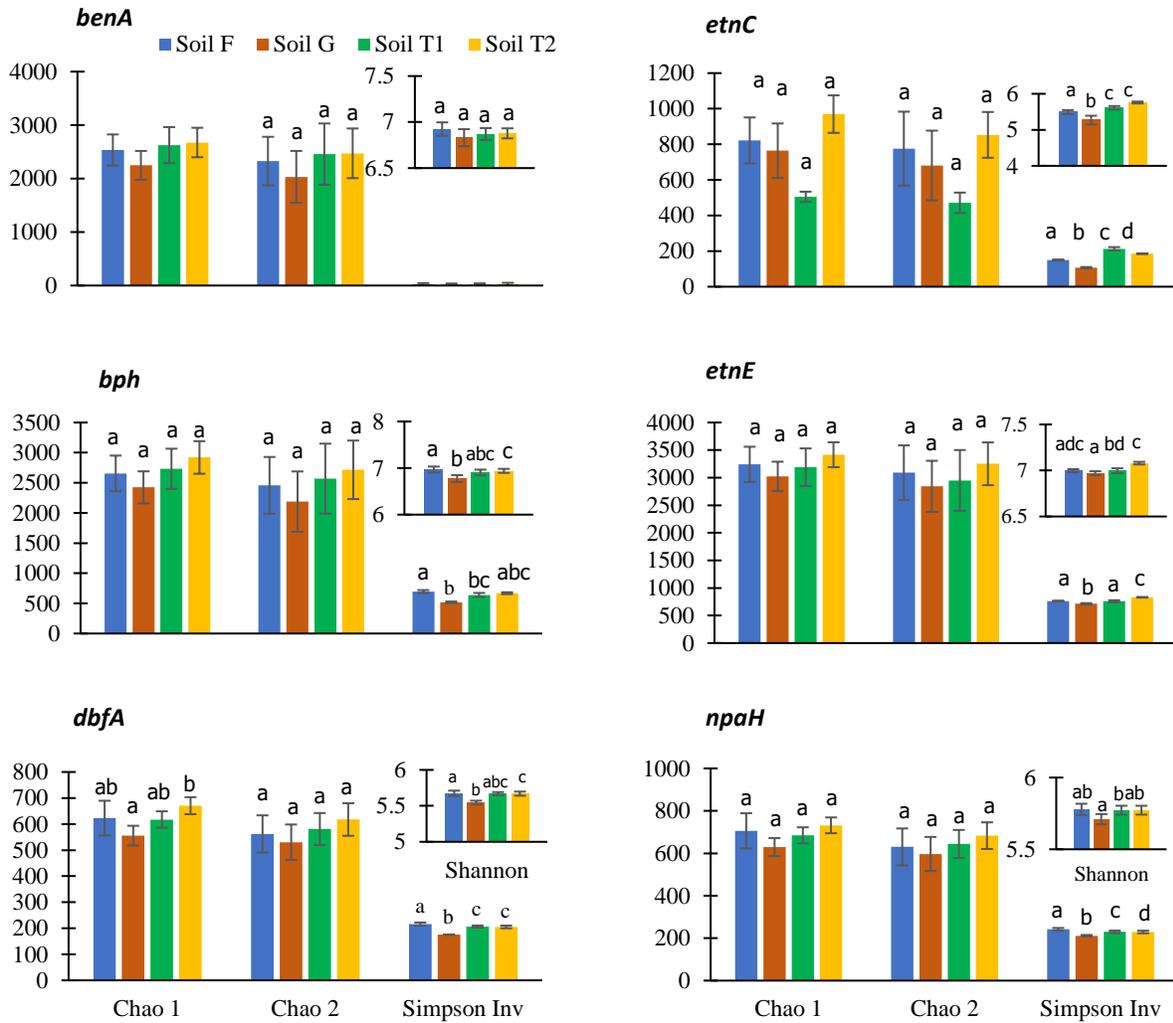


Figure 4.5. Average index diversity values for each soil as determined by EstimateS (n=4) with standard deviations illustrated with the bars. Values that are statistically significantly different (ANOVA or Kruskal-Wallis test,  $p < 0.05$ ) are shown with different letters. Note, the scale on the y-axis differs between graphs.

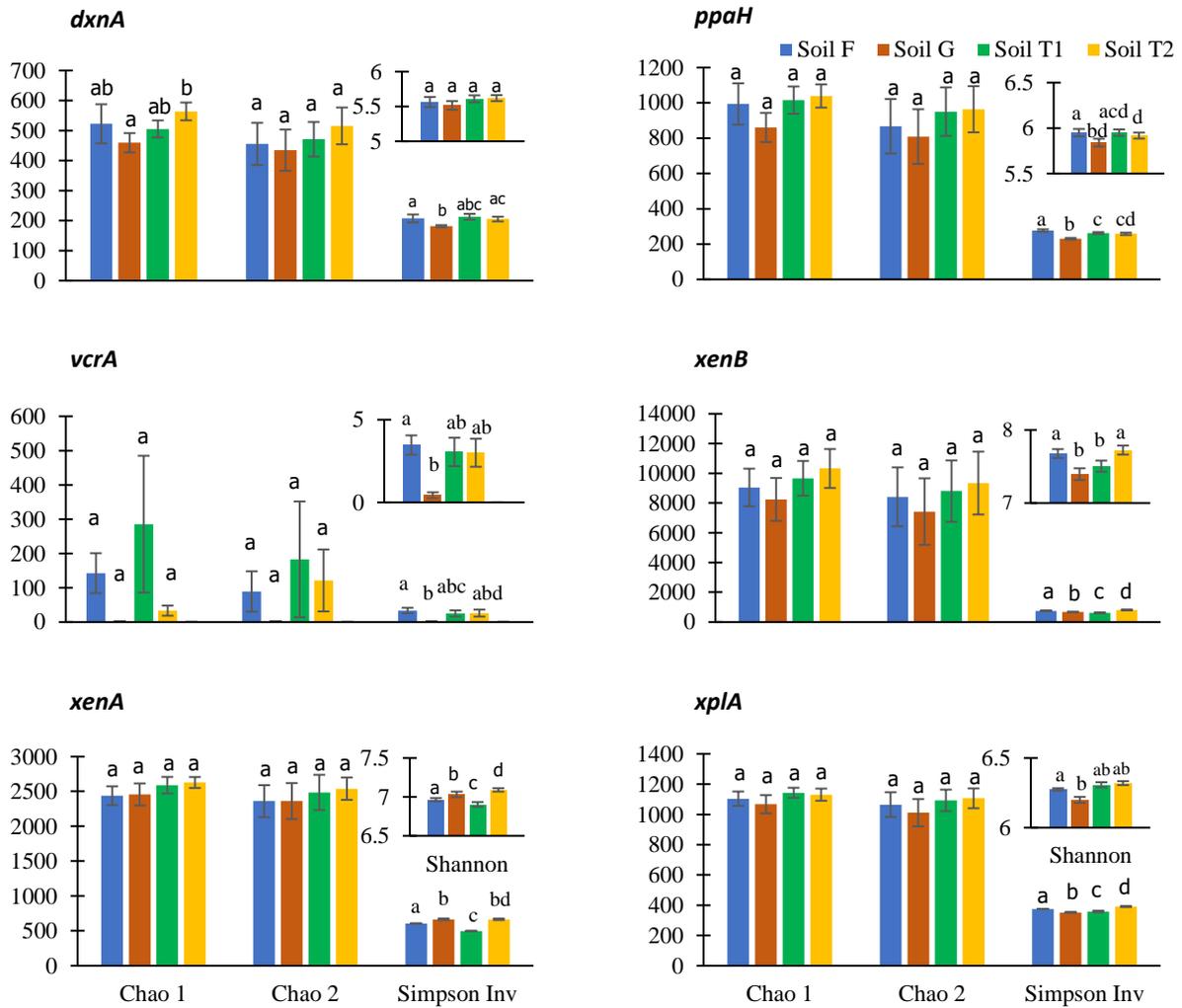


Figure 4.6. Average index diversity values for each soil as determined by EstimateS (n=4) with standard deviations illustrated with the bars. Values that are statistically significantly different (ANOVA or Kruskal-Wallis test,  $p < 0.05$ ) are shown with different letters. Note, the scale on the y-axis differs between graphs.

#### 4.4.5 Phylogenetic Trees

Phylogenetic trees were created using the top 100 accession numbers in NCBI COBALT for each gene (except *vcrA* and *npaH*). For *npaH*, only 72 accession numbers represented sequences with identity  $\geq 60\%$  and an alignment length  $\geq 49$  amino acids. For *vcrA*, no sequence exhibited an amino acid length  $\geq 49$  (and hence no tree was generated). The gene *benA* indicated an abundance of sequences classifying within the genera *Gordonia*, *Pseudomonas* and *Rhodococcus* with few classifying within the genera *Mycobacterium* and *Mycolibacterium* (Figure 4.7). The gene *bph* indicated an abundance of sequences classifying within the genera *Achromobacter*, *Bordetella* and *Mycolibacterium* with some classifying within the genera *Streptomyces* and *Herbaspirillum* (Figure 4.8a). Phylogenetic tree created for the gene *dxnA* indicated sequences classifying within the genera *Escherichia*, *Rhodococcus*, *Pseudomonas* and *Mycolibacterium* (Figure 4.9). The gene *etnC* indicated sequences classifying within the genera *Rhodococcus* and *Mesorhizobium* along with sequences classifying within the genera *Pseudomonas*, *Nocardioides*, *Mycobacterium* and *Pseudonocardia* (Figure 4.10). A number of sequences classifying within the genera *Achromobacter* and *Bradyrhizobium* were present in the phylogenetic tree created for the gene *etnE* along with some sequences classifying within the genera *Xanthomonas*, *Agrobacterium* and *Arthrobacter* (Figure 4.11a). The genes *npaH* and *ppaH* indicated several sequences classifying within the genera *Rhodococcus* and *Mycolibacterium* (Figure 4.11b and 4.12a). The gene *xenA* included sequences classifying within the genera *Pseudomonas*, *Rhodococcus* and *Bradyrhizobium* (Figure 4.12b). The gene *xenB* indicated an abundance of sequences classifying within the genera *Pseudomonas*, *Agrobacterium*, *Achromobacter*, *Fusarium*, *Streptomyces* and *Bradyrhizobium* (Figure 4.13). Similarly, the gene *xplA* indicated several sequences classifying within the genera *Bradyrhizobium* along with few sequences

classifying within the genera *Afipia* and *Mycobacterium* (Figure 4.14). Since only representative sequences were chosen for each gene, not all the strains of each species are listed in the trees created using MEGA X.

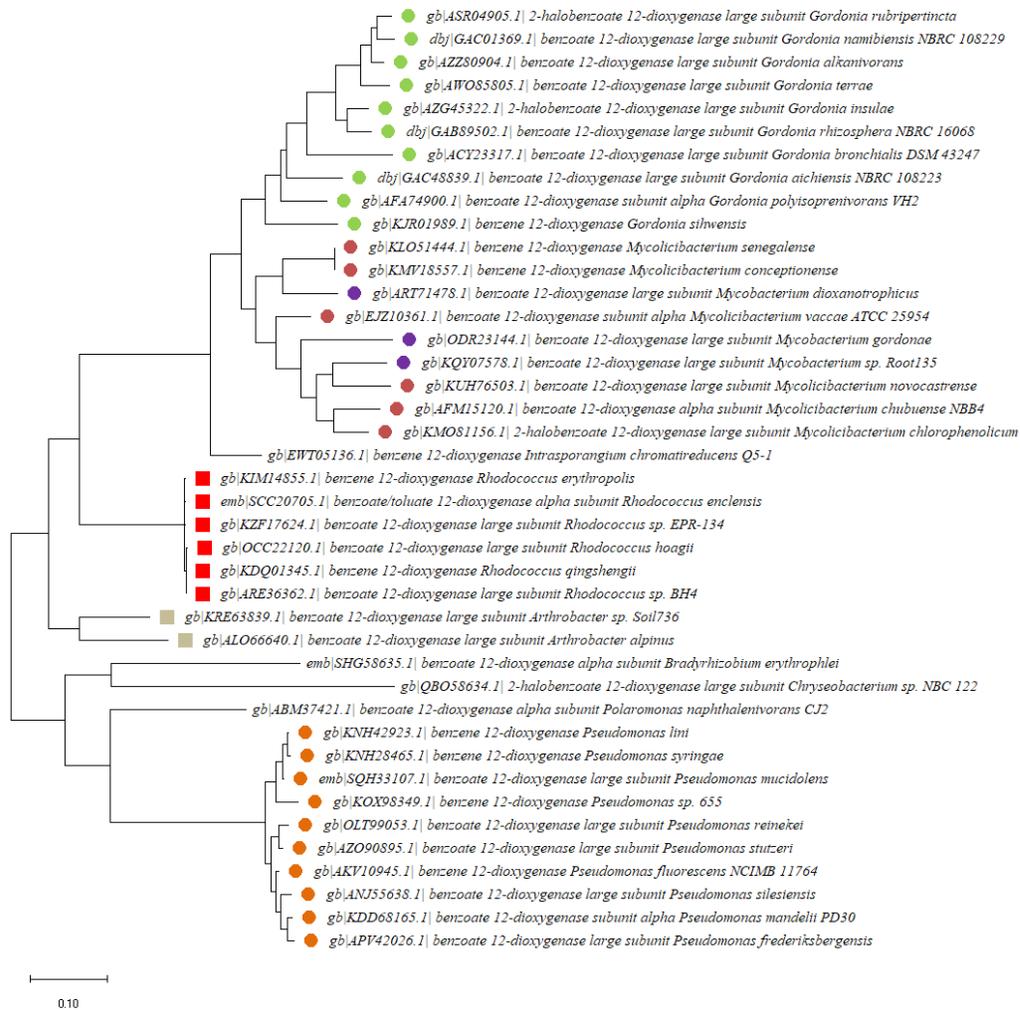


Figure 4.7. Phylogenetic tree for gene *benA* created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The tree with the highest log likelihood (-9563.06) is shown. Initial tree(s) for the heuristic Phylogenetic tree for gene *benA* created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The tree with the highest log likelihood (-9563.06) is shown. 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 a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 41 amino acid sequences. There were a total of 492 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.

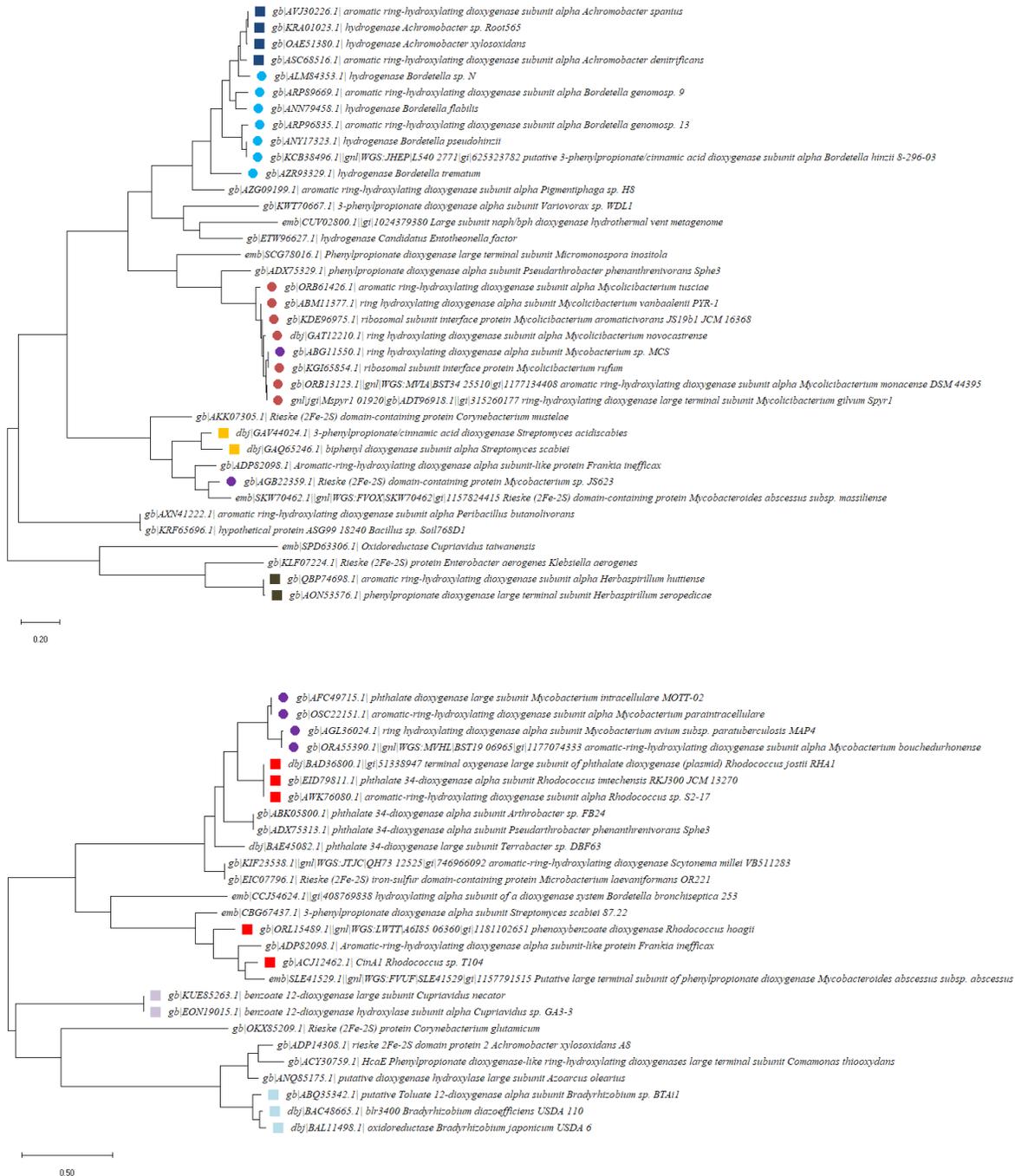


Figure 4.8. Phylogenetic tree for genes (A) bph and (B) dbfA created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The trees with the highest log likelihood (A) (-15779.72) and (B) (-12459.37) are shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 397 (A) and 27 (B) amino acid sequences. There were a total of (A) 550 and (B) 586 positions in the final dataset. Evolutionary analyses were conducted in MEGA X



Figure 4.9. Phylogenetic tree for gene *dxnA* created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The tree with the highest log likelihood (-16429.21) is shown. 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 a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 27 amino acid sequences. There were a total of 595 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.



Figure 4.10. Phylogenetic tree for gene *etnC* created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The tree with the highest log likelihood (-7690.39) is shown. 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 a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 46 amino acid sequences. There was a total of 599 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.

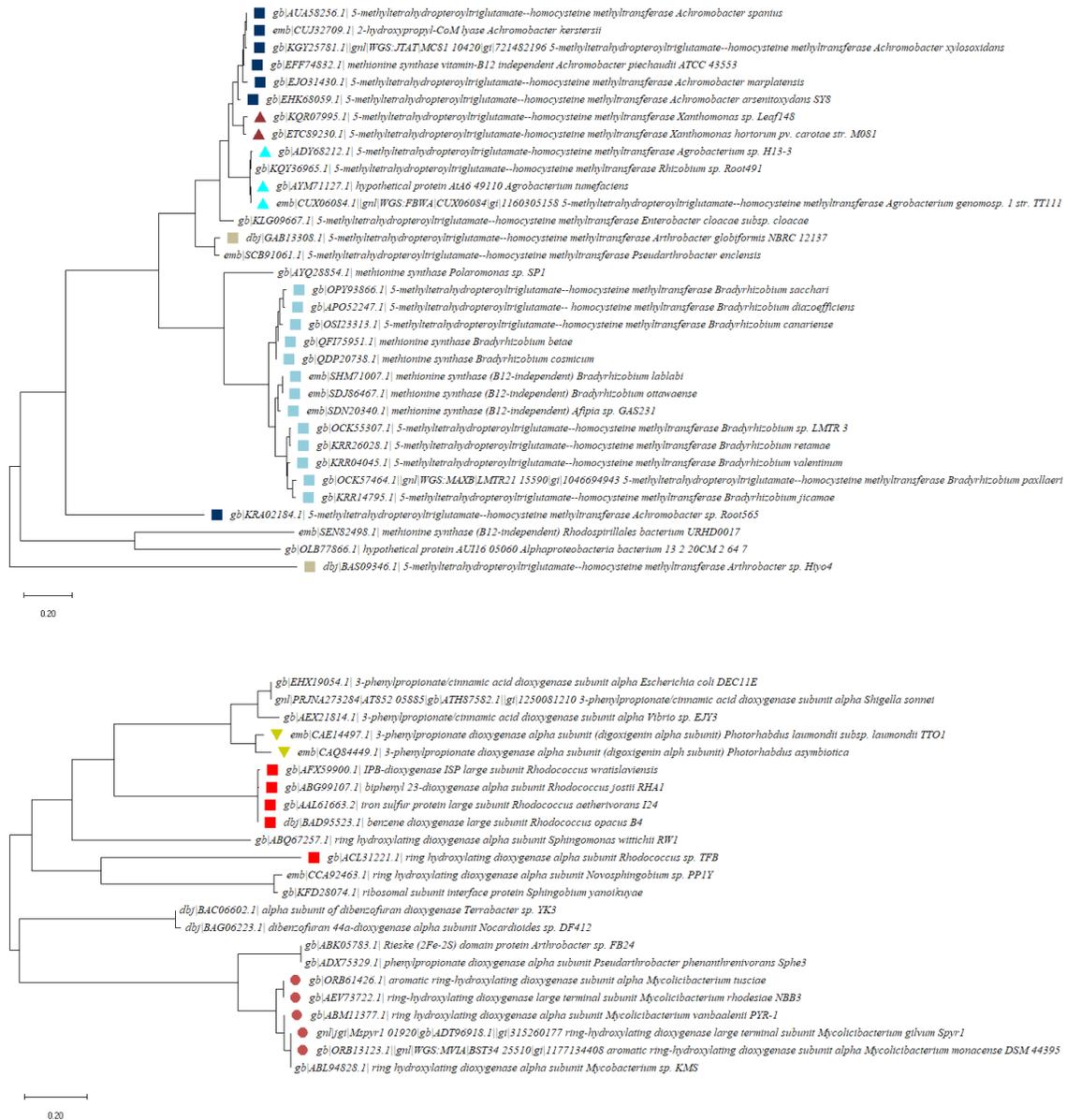


Figure 4.11. Phylogenetic tree for genes (A) *etnE* and (B) *npaH* created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The trees with the highest log likelihood (A) (-8311.56) and (B) (-9047.49) are shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved (A) 33 and (B) 23 amino acid sequences. There were a total of (A) 726 and (B) 570 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.

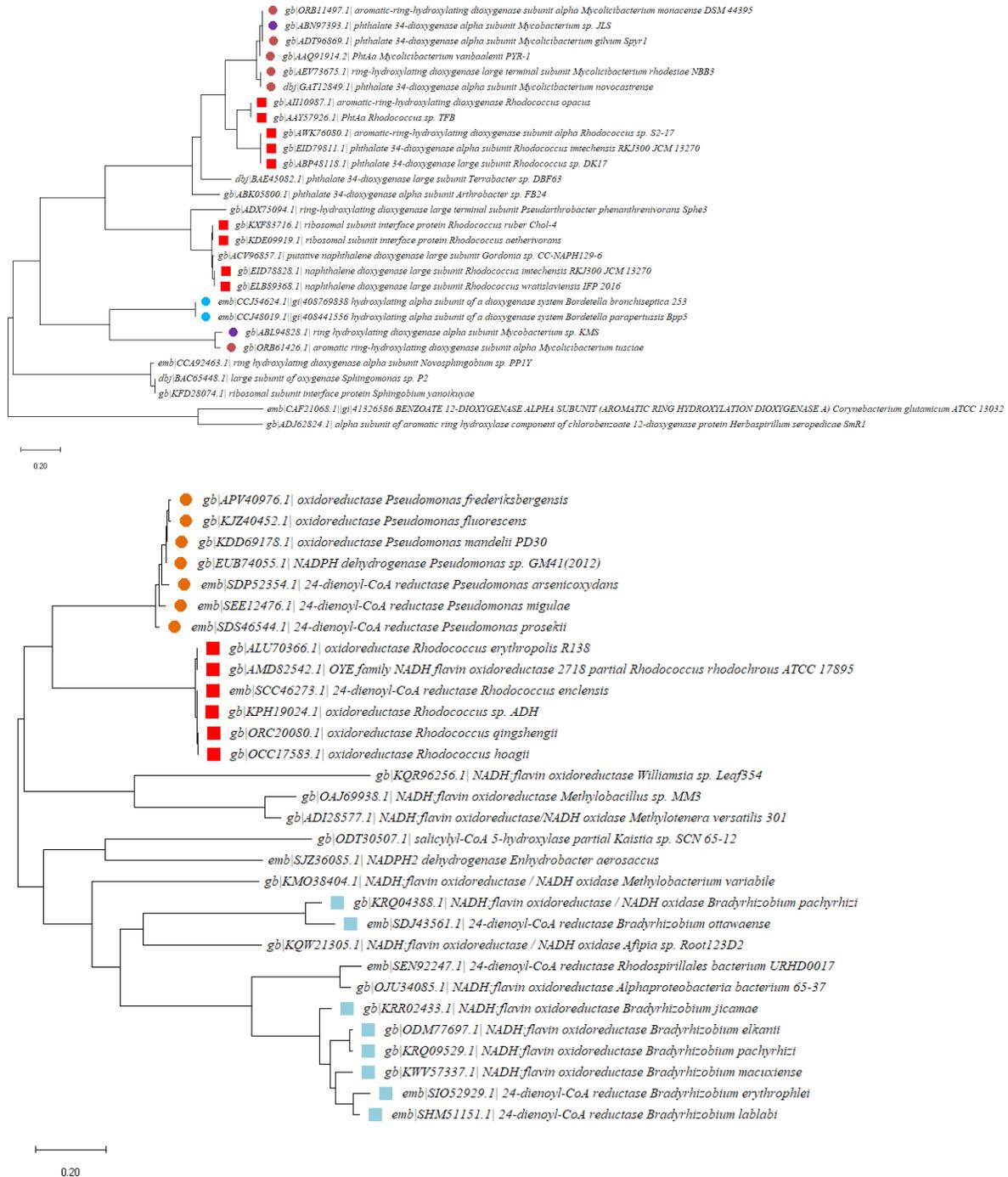


Figure 4.12. Phylogenetic tree for genes (A) ppaH and (B) xenA created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The trees with the highest log likelihood (A) (-10521.2) and (B) (-10642.31) are shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved (A) 28 and (B) 30 amino acid sequences. There were a total of (A) 561 and (B) 632 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.

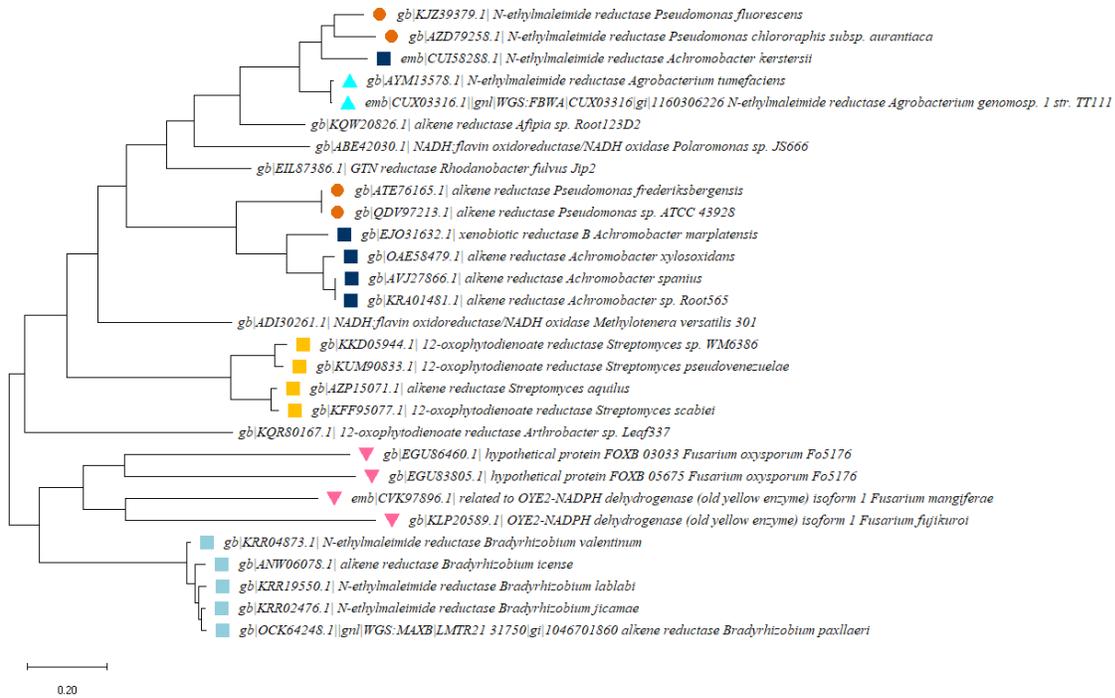


Figure 4.13. Phylogenetic tree for gene *xenB* created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The tree with the highest log likelihood (-11304.53) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 29 amino acid sequences. There was a total of 458 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.

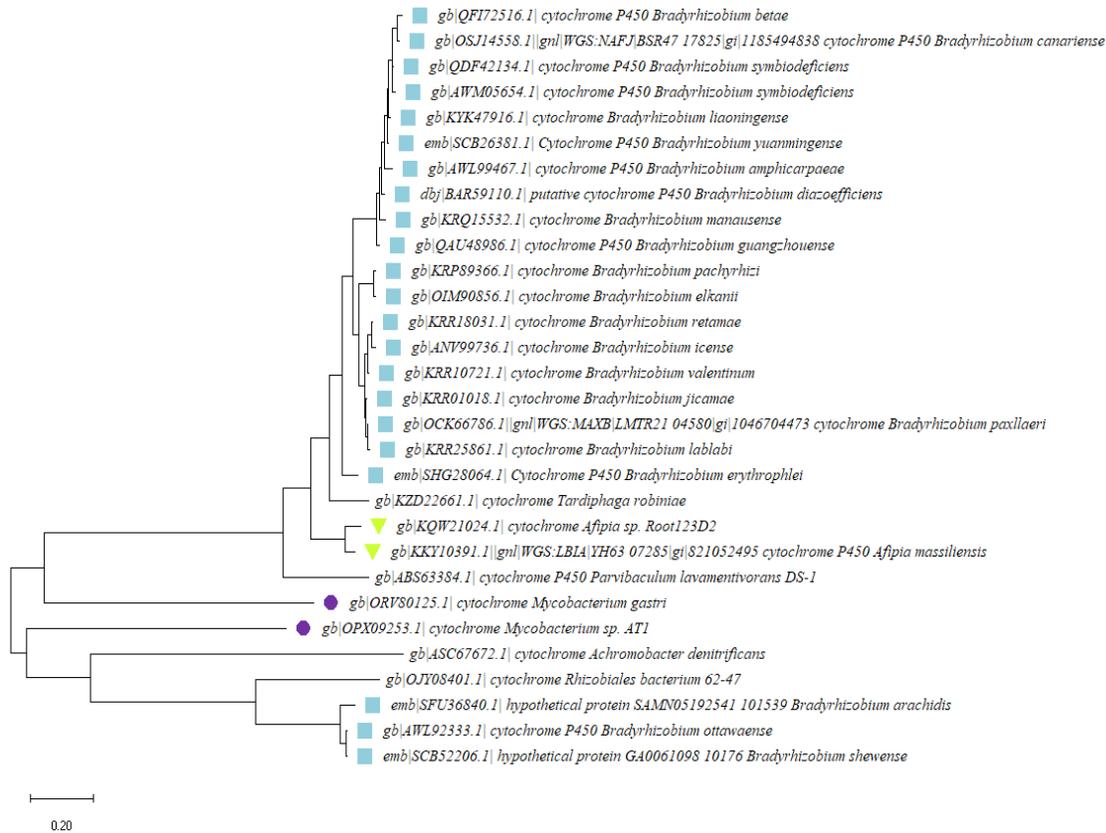


Figure 4.14. Phylogenetic tree for gene xplA created using MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based mode. The tree with the highest log likelihood (-10292.94) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 30 amino acid sequences. There was a total of 809 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.

## 4.5 Discussion

There are many challenges in remediating sites with multiple contaminants and bioremediation is often sought as a treatment option at these sites. Among the 12 genes investigated here, *dbfA*, *etnE*, *npaH*, *xenA* and *xplA* were the most abundant in all of the soil metagenomes. This trend could indicate a potential for the biodegradation of the associated contaminants as often the abundance of functional genes is positively correlated with the attenuation rates at contaminated sites (Mattes 2018). The gene *vcrA* is associated with anaerobic VC dechlorination and had the lowest presence in the soil cultures. Since the microcosms were maintained under aerobic conditions, this trend was not surprising.

Previous studies have explored the presence of these genes in soils and sediments, for example: *benA* from 3-chlorobenzoate dosed soil (Morimoto and Fujii 2009); *bph* from bacterial strains in biphenyl-contaminated soils (Hirose et al. 2019); *dbfA* in dibenzofuran degrading bacterial strains from insecticide/pesticide contaminated soils (Hassan et al. 2008); *etnC*, *etnE* and *vcrA* in samples collected from chlorinated ethene contaminated aquifer sediments (Patterson et al. 2013; Richards et al. 2019); *xenA*, *xenB* and *xplA* in uncontaminated soils and RDX contaminated aquifer sediments (Collier et al. 2019; Seth-Smith et al. 2002). However, in general, limited information is available on the presence or diversity of the targeted genes in uncontaminated soils.

In the current study, sequences associated with *xenA* classified with multiple genera including *Rhodococcus*, *Williamsia*, *Pseudomonas* and *Bradyrhizobium*. From these, both *Rhodococcus* and *Pseudomonas* have previously been linked to *xenA* (Coleman et al. 2002b; Fuller et al. 2009;

Li et al. 2014). The *Rhodococcus rhodochrous* strain identified here (as shown in the phylogenetic tree) is known to utilize RDX as a nitrogen source (Seth-Smith et al. 2002). The *xenB* tree is dominated by multiple genera including *Pseudomonas*, *Streptomyces*, *Achromobacter*, *Bradyrhizobium* and *Agrobacterium*. Among these, *Pseudomonas fluorescens* (as shown in the phylogenetic tree) was linked to *xenB* in a previous study (Blehert et al. 1999). Although the others have not previously been linked directly to RDX biodegradation or *xenB*, *Achromobacter* and *Streptomyces* have been associated with degradation of other nitro compounds such as 2,4,6-trinitrotoluene (TNT) (Pasti-Grigsby et al. 1996; Sheu et al. 2016). The phylogenetic tree developed for *xplA* in this study did not indicate the presence of RDX degrading *Gordonia* (Sabir et al. 2017) or *Rhodococcus* (Andeer et al. 2009) (both commonly reported RDX degraders) but predominantly displays *Bradyrhizobium*, *Afipia* and *Mycobacterium*. Among these, sequences related to *Afipia* have been previously recovered in a RDX degradation study (Fuller and Steffan 2008). *Mycobacterium* has also been linked to TNT degradation (Bernstein and Ronen 2012).

The gene *bph* was previously linked to biphenyl degradation with genera such as *Burkholderia*, *Achromobacter*, *Pseudomonas* and *Rhodococcus* (Pieper 2005). Although *Achromobacter* was associated with *bph* sequences in this study, the other commonly associated genera were not identified. *Pseudomonas* has frequently been associated with naphthalene degradation with extensive studies based on *Pseudomonas putida* as a model organism (Rhee et al. 2004; Simon et al. 1993). However, the phylogenetic tree developed here did not indicate *Pseudomonas* was associated with *npaH* in the soils studied.

Sequences classified with the genus *Sphingomonas* were observed in this study as shown in the *dxnA*, *npaH* and *ppaH* phylogenetic trees. *Sphingomonas* was previously associated with *dxnA* (Penton et al. 2013; Pinyakong et al. 2003). The most dominant genera in the *etnC* tree include: *Rhodococcus*, *Mycobacterium*, *Nocardioides*, *Amycolatopsis*, *Pseudonocardia* and *Mesorhizobium*. A subset of these microorganisms have been associated with *etnC* in previous studies including *Mycobacterium* (Coleman et al. 2011; McCarl et al. 2018), *Nocardioides* (Wilson et al. 2016) and *Rhodococcus* (Coleman and Spain 2003). Although, the *etnE* was linked with *Mycobacterium* (Coleman and Spain 2003), the phylogenetic tree in this study predominantly consists of *Bradyrhizobium*, *Xanthomonas*, *Achromobacter* and *Arthrobacter*.

Similar to the current work, previous studies on the distribution of biodegradation genes associated with organic pollutants, specifically in wastewater and activated sludge, have indicated a predominance of *Mycobacterium* (Fang et al. 2013). In that study, 87 bacterial genera capable of degrading organic pollutants were identified with the vast majority affiliating with *Proteobacteria*, *Bacteroidetes* and *Actinobacteria*.

The current study has several limitations that are important to state. A key concern is the extraction of DNA from enrichment cultures as this may have resulted in the selection of organisms that grow fastest versus those that occur naturally in the environment. Also, the percentage identity threshold during the analysis could impact the number of sequences similar to the target genes. Further, the composition of the reference database influences the diversity results with some uncertainty over the ideal filtering cut-off values since it varies between the functional genes. Another pattern of interest is the high positive correlation between some genes

including *dxnA* vs. *ppaH* (correlation coefficient > 0.8). These patterns could indicate homologs or similarity between the amino acid sequences. Further, the presence or abundance of certain genes may not translate to gene expression. For example, a previous study indicated the concentration of *merA* gene involved in Hg(II) reduction was not always proportional to the rate of Hg(II) reduction (Lovley 2003). Also, *ben* genes need benzoate as an inducer for gene expression (Moreno and Rojo 2008) and similarly the *bph* genes are typically induced when grown in the presence of biphenyl (Ohtsubo et al. 2001). Since the soils in the current study had no known previous exposure to these contaminants, the gene abundances may not indicate gene expression and enzyme activity.

The phylogenetic trees created in the current study suggest many genera are potentially associated with each gene. From these, some strains are well studied and are known to be involved in biodegradation of organic contaminants. However, from the current work there may be new genera associated with biodegradation of contaminants. Future research could focus on these genera and/or strains that could be involved in the biodegradation of contaminants. This study leads to a better understanding of biodegradation genes in soils and could potentially aid in manipulating contaminated sites to optimize contaminant removal.

### **Acknowledgements**

Support for this research was also provided by the NSF Long-term Ecological Research Program (DEB 1832042) at the Kellogg Biological Station and by Michigan State University

AgBioResearch

## APPENDIX

## APPENDIX

Table 4.1. Summary of data using Shapiro test to check normality of each gene. The values in bold (p-values > 0.05) indicating normal data

	<b>p-value</b>
<i>benA</i>	<b>0.8359</b>
<i>bph</i>	0.01994
<i>dbfA</i>	0.007037
<i>dxnA</i>	<b>0.5036</b>
<i>etnC</i>	<b>0.7091</b>
<i>etnE</i>	0.001807
<i>npaH</i>	0.03535
<i>ppaH</i>	<b>0.07593</b>
<i>vcrA</i>	0.003655
<i>xenA</i>	0.002165
<i>xenB</i>	0.04897
<i>xplA</i>	0.02699

Table 4.2. Summary of linear regression results for normal data with equal variance

	<i>Linear model equations</i>			<b>Pearson's correlation</b>	<b>F test</b>	
	<b>m slope</b>	<b>intercept</b>	<b>p value</b>		<b>p-value</b>	<b>ratio of variance</b>
<i>benA vs dxnA</i>	7.87E-01	1.75E-05	0.0243	0.559138	0.1979	0.5054
<i>benA vs etnC</i>	7.72E-01	1.23E-05	0.0128	0.606398	0.3598	0.6168
<i>benA vs ppaH</i>	7.17E-01	2.86E-05	0.018	0.581935	0.427	0.6579
<i>dxnA vs etnC</i>	6.38E-01	1.52E-05	0.0192	0.577182	0.7045	1.2205
<i>dxnA vs ppaH</i>	1.01E+00	-1.01E-05	6.24E-06	8.82E-01	6.16E-01	1.30E+00
<i>etnC vs ppaH</i>	5.74E-01	2.84E-06	0.0254	0.55568	0.9022	1.066

Table 4.3. Summary of the p values from Spearman's rank correlation tests with gene relative abundance data. Values in bold indicate a significant difference ( $p \leq 0.05$ ).

Genes	<i>bph</i>	<i>dbfA</i>	<i>etnE</i>	<i>npaH</i>	<i>vcrA</i>	<i>xenA</i>	<i>xenB</i>	<i>xplA</i>
<i>bph</i>		<b>1.68E-06</b>	2.40E-01	<b>1.22E-08</b>	8.79E-01	9.96E-01	4.18E-01	9.66E-01
<i>dbfA</i>	<b>1.68E-06</b>		4.44E-01	<b>4.32E-07</b>	7.94E-01	7.57E-01	3.21E-01	1.00E+00
<i>etnE</i>	2.40E-01	4.44E-01		3.62E-01	5.71E-01	6.44E-01	9.44E-01	3.16E-01
<i>npaH</i>	<b>1.22E-08</b>	<b>4.32E-07</b>	3.62E-01		6.63E-01	7.95E-01	4.71E-01	5.72E-01
<i>vcrA</i>	8.79E-01	7.94E-01	5.71E-01	6.63E-01		8.58E-01	4.16E-01	6.01E-01
<i>xenA</i>	9.96E-01	7.57E-01	6.44E-01	7.95E-01	8.58E-01		<b>9.72E-03</b>	3.19E-01
<i>xenB</i>	4.18E-01	3.21E-01	9.44E-01	4.71E-01	4.16E-01	<b>9.72E-03</b>		1.21E-01
<i>xplA</i>	9.66E-01	1.00E+00	3.16E-01	5.72E-01	6.01E-01	3.19E-01	1.21E-01	

Table 4.4. Summary of Spearman's correlation coefficient (rho) for Spearman's rank correlation test with gene relative abundance data. Rho values in bold indicate a statistically significant correlation ( $p \leq 0.05$ ), as shown above.

Genes	<i>bph</i>	<i>dbfA</i>	<i>etnE</i>	<i>npaH</i>	<i>vcrA</i>	<i>xenA</i>	<i>xenB</i>	<i>xplA</i>
<i>bph</i>		<b>9.03E-01</b>	3.12E-01	<b>9.53E-01</b>	-4.13E-02	-1.47E-03	2.18E-01	-1.18E-02
<i>dbfA</i>	<b>9.03E-01</b>		2.06E-01	<b>9.21E-01</b>	-7.08E-02	8.39E-02	2.65E-01	0.00E+00
<i>etnE</i>	3.12E-01	2.06E-01		2.44E-01	1.53E-01	-1.25E-01	-1.91E-02	2.68E-01
<i>npaH</i>	<b>9.53E-01</b>	<b>9.21E-01</b>	2.44E-01		1.18E-01	7.06E-02	1.94E-01	-1.53E-01
<i>vcrA</i>	-4.13E-02	-7.08E-02	1.53E-01	1.18E-01		-4.87E-02	-2.18E-01	-1.42E-01
<i>xenA</i>	-1.47E-03	8.39E-02	-1.25E-01	7.06E-02	-4.87E-02		<b>6.24E-01</b>	2.66E-01
<i>xenB</i>	2.18E-01	2.65E-01	-1.91E-02	1.94E-01	-2.18E-01	<b>6.24E-01</b>		4.03E-01
<i>xplA</i>	-1.18E-02	0.00E+00	2.68E-01	-1.53E-01	-1.42E-01	2.66E-01	4.03E-01	

Table 4.5. P-values for statistical tests with the richness index chao 1 of gene copies associated with xenobiotic degradation. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ).

Test	Shapiro-Wilk test				Levene's test	One-way ANOVA	Kruskal-Wallis test
Null Hypothesis	The sample distribution is normal				$\sigma_1 = \sigma_2$	$\mu_1 = \mu_2$	Median <sub>1</sub> = Median <sub>2</sub>
Soil groups	Soil F	Soil G	Soil T1	Soil T2			
benA	6.3E-2	5.0E-1	8.2E-1	8.6E-1	9.4E-1	2.2E-1	N/A
bph	1.9E-1	5.2E-1	9.3E-1	7.2E-1	9.6E-1	1.7E-1	N/A
dbfA1	8.1E-1	<b>2.3E-2</b>	6.1E-1	1.4E-1	4.6E-1	N/A	<b>3.9E-2</b>
dxnA	4.3E-1	7.6E-2	5.1E-1	<b>9.9E-3</b>	3.9E-1	N/A	<b>5.0E-2</b>
etnC	2.9E-1	3.4E-1	9.8E-1	5.1E-1	7.5E-1	1.4E-1	N/A
etnE	6.5E-1	5.9E-1	8.4E-1	2.8E-1	8.7E-1	3.5E-1	N/A
npaH	7.1E-1	9.7E-2	5.7E-1	1.3E-1	3.7E-1	1.1E-1	N/A
ppaH	1.3E-1	4.2E-1	3.9E-1	<b>8.3E-3</b>	8.5E-1	6.2E-2	N/A
vcrA	N/A	N/A	N/A	N/A	<b>6.6E-4</b>	N/A	<b>3.2E-2</b>
xenA	4.1E-1	2.8E-1	6.4E-1	3.7E-1	8.8E-1	1.1E-1	N/A
xenB	6.2E-1	6.6E-1	5.6E-1	6.5E-1	9.9E-1	1.8E-1	N/A
xplA	<b>3.5E-2</b>	5.2E-1	1.2E-1	2.6E-1	8.2E-1	1.5E-1	1.0E-1

Table 4.6. P-values for Tukey's HSD test with the richness index chao 1 of counts of xenobiotic genes. "N/A" indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ).

Test	Tukey's HSD test											
Null Hypothesis	$\mu_1 = \mu_2$											
Genes	benA	bph	dbfA1	dxnA	etnC	etnE	npaH	ppaH	vcrA	xenA	xenB	xplA
Soil G - Soil F	N/A	N/A	1.9E-1	1.9E-1	N/A							
Soil T1 - Soil F	N/A	N/A	1.0E+0	9.4E-1	N/A							
Soil T2 - Soil F	N/A	N/A	4.5E-1	5.2E-1	N/A							
Soil T1 - Soil G	N/A	N/A	2.9E-1	4.8E-1	N/A							
Soil T2 - Soil G	N/A	N/A	<b>1.8E-2</b>	<b>2.3E-2</b>	N/A							
Soil T2 - Soil T1	N/A	N/A	4.0E-1	2.8E-1	N/A							

Table 4.7. P-values for Dunn's test with the richness index chao 1 of counts of genes associated with xenobiotic degradation. "N/A" indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ).

Test	Dunn's test											
Null Hypothesis	$\mu_1 = \mu_2$											
Genes	benA	bph	dbfA1	dxnA	etnC	etnE	npaH	ppaH	vcrA	xenA	xenB	xplA
Soil G - Soil F	N/A	N/A	4.3E-1	4.6E-1	N/A	N/A	N/A	N/A	5.4E-2	N/A	N/A	N/A
Soil T1 - Soil F	N/A	N/A	1.0E+0	1.0E+0	N/A	N/A	N/A	N/A	1.0E+0	N/A	N/A	N/A
Soil T2 - Soil F	N/A	N/A	7.4E-1	1.0E+0	N/A	N/A	N/A	N/A	1.9E-1	N/A	N/A	N/A
Soil T1 - Soil G	N/A	N/A	1.0E+0	1.0E+0	N/A	N/A	N/A	N/A	4.0E-1	N/A	N/A	N/A
Soil T2 - Soil G	N/A	N/A	<b>2.5E-2</b>	<b>3.8E-2</b>	N/A	N/A	N/A	N/A	1.0E+0	N/A	N/A	N/A
Soil T2 - Soil T1	N/A	N/A	1.0E+0	6.4E-1	N/A	N/A	N/A	N/A	9.5E-1	N/A	N/A	N/A

Table 4.8. P-values for statistical tests with the richness index chao 2 of gene copies associated with xenobiotic degradation. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ).

Test	Shapiro-Wilk test				Levene's test	One-way ANOVA	Kruskal-Wallis test
Null Hypothesis	The sample distribution is normal				$\sigma_1 = \sigma_2$	$\mu_1 = \mu_2$	Median <sub>1</sub> = Median <sub>2</sub>
Soil groups	Soil F	Soil G	Soil T1	Soil T2			
benA	1.3E-1	4.8E-1	7.4E-1	5.2E-1	9.2E-1	5.8E-1	N/A
bph	2.0E-1	5.3E-1	8.1E-1	5.2E-1	9.3E-1	5.3E-1	N/A
dbfA1	7.2E-1	1.7E-1	7.9E-1	5.2E-1	1.0E+0	3.5E-1	N/A
dxnA	6.6E-1	1.2E-1	5.9E-1	6.8E-1	1.0E+0	3.8E-1	N/A
etnC	2.3E-1	1.3E-1	9.5E-1	6.4E-1	9.4E-1	5.9E-1	N/A
etnE	5.1E-1	6.6E-1	8.2E-1	5.5E-1	8.8E-1	6.6E-1	N/A
npaH	8.9E-1	3.1E-1	5.9E-1	5.2E-1	9.4E-1	4.7E-1	N/A
ppaH	6.3E-1	5.9E-1	3.3E-1	4.6E-1	9.9E-1	4.2E-1	N/A
vcrA	N/A	N/A	N/A	N/A	<b>1.1E-2</b>	N/A	N/A
xenA	3.9E-1	4.2E-1	5.6E-1	4.5E-1	9.2E-1	5.9E-1	N/A
xenB	3.5E-1	5.1E-1	5.1E-1	5.1E-1	1.0E+0	6.3E-1	N/A
xplA	1.4E-1	4.5E-1	3.7E-1	3.8E-1	9.7E-1	3.6E-1	N/A

Table 4.9 P-values for Tukey’s HSD test with the richness index chao 2 of counts of xenobiotic genes. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ).

Test	Tukey's HSD test											
Null Hypothesis	$\mu_1 = \mu_2$											
Genes	benA	bph	dbfA1	dxnA	etnC	etnE	npaH	ppaH	vcrA	xenA	xenB	xplA
Soil G - Soil F	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Soil T1 - Soil F	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Soil T2 - Soil F	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Soil T1 - Soil G	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Soil T2 - Soil G	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Soil T2 - Soil T1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 4.10 P-values for Dunn’s test with the richness index chao 2 of counts of genes associated with xenobiotic degradation. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ).

Test	Dunn's test											
Null Hypothesis	$\mu_1 = \mu_2$											
Genes	benA	bph	dbfA1	dxnA	etnC	etnE	npaH	ppaH	vcrA	xenA	xenB	xplA
Soil G - Soil F	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Soil T1 - Soil F	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Soil T2 - Soil F	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Soil T1 - Soil G	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Soil T2 - Soil G	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Soil T2 - Soil T1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 4.11. P-values for statistical tests with the Shannon diversity of genes associated with xenobiotic degradation. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ).

Test	Shapiro-Wilk test				Levene's test	One-way ANOVA	Kruskal-Wallis test
Null Hypothesis	The sample distribution is normal				$\sigma_1 = \sigma_2$	$\mu_1 = \mu_2$	Median <sub>1</sub> = Median <sub>2</sub>
Soil groups	Soil F	Soil G	Soil T1	Soil T2			
benA	1.0E-1	4.5E-1	5.8E-1	2.7E-1	8.9E-1	3.3E-1	N/A
bph	1.7E-1	2.1E-1	5.4E-1	5.8E-1	9.3E-1	<b>2.8E-3</b>	N/A
dbfA1	6.2E-2	6.9E-1	5.8E-1	1.6E-1	9.5E-1	<b>6.0E-5</b>	N/A
dxnA	1.0E-1	5.4E-1	3.3E-1	5.8E-1	9.4E-1	9.3E-2	N/A
etnC	3.5E-1	3.7E-1	7.1E-1	3.7E-1	2.5E-1	<b>2.8E-7</b>	N/A
etnE	5.4E-2	5.8E-1	3.7E-1	1.6E-1	8.3E-1	<b>1.8E-5</b>	N/A
npaH	9.6E-2	4.1E-1	5.4E-1	5.4E-1	1.0E+0	<b>4.6E-2</b>	N/A
ppaH	<b>4.5E-2</b>	5.8E-1	4.1E-1	1.2E-1	9.9E-1	N/A	<b>1.8E-2</b>
vcrA	N/A	N/A	N/A	N/A	1.2E-1	N/A	<b>4.4E-2</b>
xenA	1.7E-1	1.6E-1	5.4E-1	5.8E-1	8.7E-1	<b>1.6E-6</b>	N/A
xenB	2.9E-1	3.0E-1	4.6E-1	5.4E-1	9.7E-1	<b>3.8E-5</b>	N/A
xplA	<b>4.6E-2</b>	5.8E-1	8.7E-2	1.6E-1	7.2E-1	N/A	<b>3.3E-3</b>

Table 4.12. P-values for Tukey’s HSD test with the Shannon diversity of genes associated with xenobiotic degradation. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ).

Test	Tukey's HSD test											
Null Hypothesis	$\mu_1 = \mu_2$											
Genes	benA	bph	dbfA1	dxnA	etnC	etnE	npaH	ppaH	vcrA	xenA	xenB	xplA
Soil G - Soil F	N/A	<b>2.3E-3</b>	<b>1.2E-4</b>	N/A	<b>2.1E-3</b>	2.1E-1	5.3E-2	<b>4.5E-3</b>	N/A	<b>1.2E-2</b>	<b>2.1E-4</b>	<b>6.3E-5</b>
Soil T1 - Soil F	N/A	4.6E-1	1.0E+0	N/A	<b>2.2E-4</b>	1.0E+0	1.0E+0	1.0E+0	N/A	<b>2.5E-2</b>	<b>1.2E-2</b>	<b>3.8E-2</b>
Soil T1 - Soil G	N/A	9.1E-1	1.0E+0	N/A	<b>3.6E-4</b>	<b>1.8E-4</b>	<b>5.2E-3</b>	6.0E-1	N/A	<b>6.4E-5</b>	7.5E-1	<b>4.6E-3</b>
Soil T2 - Soil F	N/A	<b>4.1E-2</b>	<b>2.6E-4</b>	N/A	<b>8.0E-7</b>	2.5E-1	1.0E-1	<b>6.7E-3</b>	N/A	<b>8.0E-5</b>	1.7E-1	<b>1.9E-6</b>
Soil T2 - Soil G	N/A	<b>1.1E-2</b>	<b>2.6E-4</b>	N/A	<b>1.1E-6</b>	<b>1.5E-5</b>	1.0E-1	5.9E-2	N/A	<b>4.9E-2</b>	<b>7.7E-5</b>	<b>5.0E-7</b>
Soil T2 - Soil T1	N/A	8.5E-1	1.0E+0	N/A	9.9E-1	<b>2.9E-4</b>	1.0E+0	6.4E-1	N/A	<b>1.5E-6</b>	<b>3.1E-3</b>	7.0E-1

Table 4.13. P-values for Dunn’s test with the Shannon diversity of genes associated with xenobiotic degradation. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ).

Test	Dunn's test											
Null Hypothesis	$\mu_1 = \mu_2$											
Genes	benA	bph	dbfA1	dxnA	etnC	etnE	npaH	ppaH	vcrA	xenA	xenB	xplA
Soil G - Soil F	N/A	N/A	N/A	N/A	N/A	N/A	N/A	<b>3.1E-2</b>	<b>2.8E-2</b>	N/A	N/A	9.2E-1
Soil T1 - Soil F	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1.0E+0	1.0E+0	N/A	N/A	9.0E-1
Soil T2 - Soil F	N/A	N/A	N/A	N/A	N/A	N/A	N/A	<b>5.0E-2</b>	8.9E-1	N/A	N/A	<b>3.9E-2</b>
Soil T1 - Soil G	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1.0E+0	7.1E-1	N/A	N/A	9.2E-1
Soil T2 - Soil G	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1.0E+0	1.0E+0	N/A	N/A	<b>4.2E-3</b>
Soil T2 - Soil T1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1.0E+0	1.0E+0	N/A	N/A	1.0E+0

Table 4.14. P-values for statistical tests with the richness index chao 2 of gene copies associated with xenobiotic degradation. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ).

Test	Shapiro-Wilk test				Levene's test	One-way ANOVA	Kruskal-Wallis test
Null Hypothesis	The sample distribution is normal				$\sigma_1 = \sigma_2$	$\mu_1 = \mu_2$	Median <sub>1</sub> = Median <sub>2</sub>
Soil groups	Soil F	Soil G	Soil T1	Soil T2			
benA	9.4E-1	7.3E-1	8.6E-1	8.0E-1	9.7E-1	9.7E-1	N/A
bph	9.4E-2	2.2E-1	5.0E-1	2.6E-1	6.3E-1	<b>2.8E-7</b>	N/A
dbfA1	6.4E-2	6.8E-1	1.6E-1	1.3E-1	6.6E-1	<b>5.6E-8</b>	N/A
dxnA	7.9E-2	9.4E-1	5.4E-1	4.3E-1	7.3E-1	<b>2.2E-3</b>	N/A
etnC	5.1E-1	7.1E-1	2.7E-1	1.3E-1	3.9E-1	<b>3.3E-15</b>	N/A
etnE	1.6E-1	1.3E-1	1.7E-1	2.9E-1	5.9E-1	<b>3.8E-8</b>	N/A
npaH	5.1E-2	6.9E-1	1.9E-1	1.9E-1	9.6E-1	<b>2.4E-5</b>	N/A
ppaH	4.2E-2	4.0E-1	3.8E-1	1.3E-1	1.0E+0	N/A	<b>4.1E-3</b>
vcrA	2.0E-1	7.8E-2	8.0E-1	9.2E-1	7.5E-2	<b>6.5E-3</b>	N/A
xenA	7.1E-1	3.0E-1	3.9E-1	3.5E-1	5.6E-1	<b>1.7E-12</b>	N/A
xenB	8.2E-1	8.9E-1	6.5E-1	6.8E-1	5.2E-2	<b>5.1E-10</b>	N/A
xplA	8.5E-1	8.9E-1	8.9E-1	8.5E-1	2.2E-1	<b>2.8E-9</b>	N/A

Table 4.15. P-values for Tukey’s HSD test with the richness index chao 2 of counts of xenobiotic genes. “N/A” indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ).

Test	Tukey's HSD test											
Null Hypothesis	$\mu_1 = \mu_2$											
Genes	benA	bph	dbfA1	dxnA	etnC	etnE	npaH	ppaH	vcrA	xenA	xenB	xplA
Soil G - Soil F	N/A	<b>2.0E-7</b>	<b>0.0E+0</b>	<b>6.9E-3</b>	<b>0.0E+0</b>	<b>1.5E-4</b>	<b>1.1E-5</b>	<b>3.0E-7</b>	<b>3.5E-3</b>	<b>7.0E-7</b>	<b>1.3E-5</b>	<b>1.2E-6</b>
Soil T1 - Soil F	N/A	<b>1.6E-2</b>	<b>4.2E-2</b>	8.2E-1	<b>0.0E+0</b>	9.7E-1	<b>3.4E-2</b>	<b>2.0E-2</b>	2.4E-1	<b>0.0E+0</b>	<b>0.0E+0</b>	<b>4.0E-5</b>
Soil T2 - Soil F	N/A	3.4E-1	<b>1.7E-2</b>	9.9E-1	<b>0.0E+0</b>	<b>6.1E-6</b>	<b>2.3E-2</b>	<b>3.1E-3</b>	2.8E-1	<b>7.0E-7</b>	<b>3.8E-4</b>	<b>6.6E-5</b>
Soil T1 - Soil G	N/A	2.6E-5	<b>1.4E-6</b>	<b>2.3E-3</b>	<b>0.0E+0</b>	<b>4.8E-4</b>	<b>2.1E-3</b>	<b>3.0E-5</b>	1.5E-1	<b>0.0E+0</b>	<b>5.6E-4</b>	<b>9.9E-2</b>
Soil T2 - Soil G	N/A	2.7E-6	<b>2.5E-6</b>	<b>1.8E-2</b>	<b>0.0E+0</b>	<b>0.0E+0</b>	<b>3.1E-3</b>	<b>1.3E-4</b>	1.3E-1	1.0E+0	<b>0.0E+0</b>	<b>0.0E+0</b>
Soil T2 - Soil T1	N/A	3.5E-1	9.6E-1	6.8E-1	<b>3.1E-5</b>	<b>6.3E-6</b>	<b>1.0E+0</b>	7.6E-1	<b>1.0E+0</b>	<b>0.0E+0</b>	<b>0.0E+0</b>	<b>0.0E+0</b>

Table 4.16. P-values for Dunn's test with the richness index chao 2 of counts of genes associated with xenobiotic degradation. "N/A" indicates the test was not appropriate and p-values in bold indicate a significant difference ( $p \leq 0.05$ ).

Test	Dunn's test											
Null Hypothesis	$\mu_1 = \mu_2$											
Genes	benA	bph	dbfA1	dxnA	etnC	etnE	npaH	ppaH	vcrA	xenA	xenB	xplA
Soil G - Soil F	N/A	<b>4.1E-3</b>	N/A	N/A	N/A	N/A	N/A	<b>2.1E-3</b>	N/A	N/A	N/A	N/A
Soil T1 - Soil F	N/A	3.3E-1	N/A	N/A	N/A	N/A	N/A	1.0E+0	N/A	N/A	N/A	N/A
Soil T1 - Soil G	N/A	9.7E-1	N/A	N/A	N/A	N/A	N/A	2.1E-1	N/A	N/A	N/A	N/A
Soil T2 - Soil F	N/A	1.0E+0	N/A	N/A	N/A	N/A	N/A	2.2E-1	N/A	N/A	N/A	N/A
Soil T2 - Soil G	N/A	1.3E-1	N/A	N/A	N/A	N/A	N/A	9.7E-1	N/A	N/A	N/A	N/A
Soil T2 - Soil T1	N/A	1.0E+0	N/A	N/A	N/A	N/A	N/A	1.0E+0	N/A	N/A	N/A	N/A

## REFERENCES

## REFERENCES

- Abraham WR, Nogales B, Golyshin PN, Pieper DH, Timmis KN (2002) Polychlorinated biphenyl-degrading microbial communities in soils and sediments. *Current Opinion in Microbiology* 5(3):246-253 doi:10.1016/s1369-5274(02)00323-5
- Allen JR, Clark DD, Krum JG, Ensign SA (1999) A role for coenzyme M (2-mercaptoethanesulfonic acid) in a bacterial pathway of aliphatic epoxide carboxylation. *Proceedings of the National Academy of Sciences* 96(15):8432-8437
- Andeer PF, Stahl DA, Bruce NC, Strand SE (2009) Lateral Transfer of Genes for Hexahydro-1,3,5-Trinitro-1,3,5-Triazine (RDX) Degradation. *Appl Environ Microb* 75(10):3258 doi:10.1128/AEM.02396-08
- ATSDR (2000) Toxicological Profile for Polychlorinated Biphenyls. In: Registry AfTSaD (ed). U.S. Department of Health and Human Services, Public Health Service., Atlanta, GA
- ATSDR (2006) Toxicological profile for Vinyl Chloride. In: U.S. Department of Health and Human Services PHS (ed). Atlanta GA
- Bedard DL, Unterman R, Bopp LH, Brennan MJ, Haberl ML, Johnson C (1986) Rapid assay for screening and characterizing microorganisms for the ability to degrade polychlorinated biphenyls. *Appl Environ Microb* 51(4):761-768 doi:10.1128/AEM.51.4.761-768.1986
- Bernstein A, Adar E, Nejidat A, Ronen Z (2011) Isolation and characterization of RDX-degrading *Rhodococcus* species from a contaminated aquifer. *Biodegrad* 22(5):997-1005 doi:10.1007/s10532-011-9458-0
- Bernstein A, Ronen Z (2012) Biodegradation of the Explosives TNT, RDX and HMX. Springer Berlin Heidelberg, pp 135-176
- Blehert DS, Fox BG, Chambliss GH (1999) Cloning and sequence analysis of two *Pseudomonas* flavoprotein xenobiotic reductases. *Journal of Bacteriology* 181(20):6254-6263 doi:10.1128/jb.181.20.6254-6263.1999
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114-2120 doi:10.1093/bioinformatics/btu170
- Brzuzny LP, Hites RA (1996) Global Mass Balance for Polychlorinated Dibenzo-p-dioxins and Dibenzofurans. *Environ Sci Technol* 30(6):1797-1804 doi:10.1021/es950714n
- Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12(1):59-60

- Coleman NV, Mattes TE, Gossett JM, Spain JC (2002a) Phylogenetic and kinetic diversity of aerobic vinyl chloride-assimilating bacteria from contaminated sites. *Appl Environ Microb* 68(12):6162-6171 doi:10.1128/aem.68.12.6162-6171.2002
- Coleman NV, Nelson DR, Duxbury T (1998) Aerobic biodegradation of hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) as a nitrogen source by a *Rhodococcus* sp., strain DN22. *Soil Biology & Biochemistry* 30(8-9):1159-1167 doi:10.1016/s0038-0717(97)00172-7
- Coleman NV, Spain JC (2003) Epoxyalkane: coenzyme M transferase in the ethene and vinyl chloride biodegradation pathways of *Mycobacterium* strain JS60. *Journal of bacteriology* 185(18):5536-5545
- Coleman NV, Spain JC, Duxbury T (2002b) Evidence that RDX biodegradation by *Rhodococcus* strain DN22 is plasmid-borne and involves a cytochrome p-450. *J Appl Microbiol* 93(3):463-72 doi:10.1046/j.1365-2672.2002.01713.x
- Coleman NV, Yau S, Wilson NL, Nolan LM, Migocki MD, Ly MA, Crossett B, Holmes AJ (2011) Untangling the multiple monooxygenases of *Mycobacterium chubuense* strain NBB4, a versatile hydrocarbon degrader. *Environmental Microbiology Reports* 3(3):297-307 doi:10.1111/j.1758-2229.2010.00225.x
- Collier JM, Chai B, Cole JR, Michalsen MM, Cupples AM (2019) High throughput quantification of the functional genes associated with RDX biodegradation using the SmartChip real-time PCR system. *Appl Microbiol Biot* 103(17):7161-7175 doi:10.1007/s00253-019-10022-x
- Colwell RK, Elsensohn JE (2014) EstimateS turns 20: statistical estimation of species richness and shared species from samples, with non-parametric extrapolation. *Ecography* 37(6):609-613
- Denef VJ, Park J, Tsoi TV, Rouillard JM, Zhang H, Wibbenmeyer JA, Verstraete W, Gulari E, Hashsham SA, Tiedje JM (2004) Biphenyl and Benzoate Metabolism in a Genomic Context: Outlining Genome-Wide Metabolic Networks in *Burkholderia xenovorans* LB400. *Appl Environ Microb* 70(8):4961 doi:10.1128/AEM.70.8.4961-4970.2004
- Dipple A (1985) Polycyclic Aromatic Hydrocarbon Carcinogenesis Polycyclic Hydrocarbons and Carcinogenesis. ACS Symposium Series, vol 283. American Chemical Society, pp 1-17
- Ensign SA (2001) Microbial metabolism of aliphatic alkenes. *Biochemistry* 40(20):5845-5853
- EPA US (2007) Provisional Peer Reviewed Toxicity Values for Dibenzofuran. In: Center SHRTS (ed). National Center for Environmental Assessment, Cincinnati, OH 45268
- Erickson BD, Mondello FJ (1992) Nucleotide sequencing and transcriptional mapping of the genes encoding biphenyl dioxygenase, a multicomponent polychlorinated-biphenyl-degrading enzyme in *Pseudomonas* strain LB400. *Journal of bacteriology* 174(9):2903-2912 doi:10.1128/jb.174.9.2903-2912.1992

- Fang H, Cai L, Yu Y, Zhang T (2013) Metagenomic analysis reveals the prevalence of biodegradation genes for organic pollutants in activated sludge. *Bioresource Technol* 129:209-218 doi:<https://doi.org/10.1016/j.biortech.2012.11.054>
- Faroon O, Ruiz P (2016) Polychlorinated biphenyls: New evidence from the last decade. *Toxicol & Indust Health* 32(11):1825-1847 doi:10.1177/0748233715587849
- Field JA, Sierra-Alvarez R (2008) Microbial degradation of chlorinated dioxins. *Chemosphere* 71(6):1005-1018
- Fish J, Chai B, Wang Q, Sun Y, Brown CT, Tiedje J, R.Cole's J (2013) FunGene: the Functional Gene Pipeline and Repository. *Frontiers in microbiology* 4:291 doi:10.3389/fmicb.2013.00291
- Fox J, Weisberg S (2019) *An R Companion to Applied Regression, Third Edition.*, Thousand Oaks CA: Sage
- Fuller ME, McClay K, Hawari J, Paquet L, Malone TE, Fox BG, Steffan RJ (2009) Transformation of RDX and other energetic compounds by xenobiotic reductases XenA and XenB. *Appl Microbiol Biot* 84(3):535-544 doi:10.1007/s00253-009-2024-6
- Fuller ME, Steffan RJ (2008) *Groundwater Chemistry and Microbial Ecology Effects on Explosives Biodegradation.* vol SERDP Project ER-1378. Strategic Environmental Research and Development Program (SERDP)
- Furukawa K, Miyazaki T (1986) Cloning of a gene cluster encoding biphenyl and chlorobiphenyl degradation in *Pseudomonas pseudoalcaligenes*. *Journal of Bacteriology* 166(2):392-398 doi:10.1128/jb.166.2.392-398.1986
- Hassan H, Huu N, Wray V, Junca H, Pieper D (2008) Two Angular Dioxygenases Contribute to the Metabolic Versatility of Dibenzofuran-Degrading *Rhodococcus* sp. Strain HA01. *Appl Environ Microb* 74:3812-22 doi:10.1128/AEM.00226-08
- Hernandez BS, Higson FK, Kondrat R, Focht DD (1991) Metabolism of and inhibition by chlorobenzoates in *Pseudomonas putida* P111. *Appl Environ Microb* 57(11):3361-3366 doi:10.1128/AEM.57.11.3361-3366.1991
- Hirose J, Fujihara H, Watanabe T, Kimura N, Suenaga H, Futagami T, Goto M, Suyama A, Furukawa K (2019) Biphenyl/PCB Degrading bph Genes of Ten Bacterial Strains Isolated from Biphenyl-Contaminated Soil in Kitakyushu, Japan: Comparative and Dynamic Features as Integrative Conjugative Elements (ICEs). *Genes (Basel)* 10(5):404 doi:10.3390/genes10050404
- Indest KJ, Crocker FH, Athow R (2007) A TaqMan polymerase chain reaction method for monitoring RDX-degrading bacteria based on the *xplA* functional gene. *Journal of Microbiological Methods* 68(2):267-274 doi:<https://doi.org/10.1016/j.mimet.2006.08.008>
- Iyer R, Aggarwal J, Iken B (2016) A review of the Texas, USA San Jacinto Superfund site and the deposition of polychlorinated dibenzo-p-dioxins and dibenzofurans in the San Jacinto River

and Houston Ship Channel. *Environ Sci Pollut Res Int* 23(23):23321-23338 doi:10.1007/s11356-016-7501-8

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8(3):275-82 doi:10.1093/bioinformatics/8.3.275

Kassambara A (2020) R rstatix: Pipe-Friendly Framework for Basic Statistical Tests. package version 0.6.0.

Kasuga K, Nitta A, Kobayashi M, Habe H, Nojiri H, Yamane H, Omori T, Kojima I (2013) Cloning of *dfdA* genes from *Terrabacter* sp strain DBF63 encoding dibenzofuran 4,4a-dioxygenase and heterologous expression in *Streptomyces lividans*. *Appl Microbiol Biot* 97(10):4485-4498 doi:10.1007/s00253-012-4565-3

Kinzell L, Mckenzie R, Olson B, Kirsch D, Shull L (1979) ES&T Special Report: Priority pollutants: I-a perspective view. *Environ Sci Technol* 13:416-423

Kitagawa W, Miyauchi K, Masai E, Fukuda M (2001) Cloning and characterization of benzoate catabolic genes in the gram-positive polychlorinated biphenyl degrader *Rhodococcus* sp. strain RHA1. *Journal of Bacteriology* 183(22):6598 doi:10.1128/JB.183.22.6598-6606.2001

Kumamaru T, Suenaga H, Mitsuoka M, Watanabe T, Furukawa K (1998) Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase. *Nature Biotechnology* 16(7):663-666 doi:10.1038/nbt0798-663

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* 35(6):1547-1549 doi:10.1093/molbev/msy096

Lee PKH, Johnson DR, Holmes VF, He J, Alvarez-Cohen L (2006) Reductive dehalogenase gene expression as a biomarker for physiological activity of *Dehalococcoides* spp. *Appl Environ Microb* 72(9):6161-6168 doi:10.1128/AEM.01070-06

Li D, Liu C-M, Luo R, Sadakane K, Lam T-W (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31(10):1674-1676 doi:10.1093/bioinformatics/btv033

Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, Yamashita H, Lam TW (2016) MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* 102:3-11 doi:10.1016/j.ymeth.2016.02.020

Li RW, Giarrizzo JG, Wu S, Li W, Durringer JM, Craig AM (2014) Metagenomic Insights into the RDX-Degrading Potential of the Ovine Rumen Microbiome. *PLOS One* 9(11):e110505 doi:10.1371/journal.pone.0110505

Löffler FE, Ritalahti KM, Zinder SH (2013) *Dehalococcoides* and reductive dechlorination of chlorinated solvents Bioaugmentation for groundwater remediation. Springer, pp 39-88

Lovley DR (2003) Cleaning up with genomics: applying molecular biology to bioremediation. *Nature Reviews Microbiology* 1(1):35-44 doi:10.1038/nrmicro731

Lu C, Hong Y, Odinga ES, Liu J, Tsang DCW, Gao Y (2021) Bacterial community and PAH-degrading genes in paddy soil and rice grain from PAH-contaminated area. *Applied Soil Ecology* 158:103789 doi:10.1016/j.apsoil.2020.103789

Mattes T (2018) Validation of Biotechnology for Quantifying the Abundance and Activity of Vinyl-chloride Oxidizers in Contaminated Groundwater ESTCP Project ER-201425. The University of Iowa

Mayes BA, McConnell EE, Neal BH, Brunner MJ, Hamilton SB, Sullivan TM, Peters AC, Ryan MJ, Toft JD, Singer AW, Brown JF, Menton RG, Moore JA (1998) Comparative Carcinogenicity in Sprague–Dawley Rats of the Polychlorinated Biphenyl Mixtures Aroclors 1016, 1242, 1254, and 1260. *Toxicological Sciences* 41(1):62-76 doi:https://doi.org/10.1006/toxs.1997.2397

McCarl V, Somerville MV, Ly M-A, Henry R, Liew EF, Wilson NL, Holmes AJ, Coleman NV (2018) Heterologous Expression of Mycobacterium Alkene Monooxygenases in Gram-Positive and Gram-Negative Bacterial Hosts. *Appl Environ Microb* 84(15):e00397-18 doi:10.1128/AEM.00397-18

Moreno R, Rojo F (2008) The target for the *Pseudomonas putida* Crc global regulator in the benzoate degradation pathway is the BenR transcriptional regulator. *Journal of Bacteriology* 190(5):1539 doi:10.1128/JB.01604-07

Morimoto S, Fujii T (2009) A new approach to retrieve full lengths of functional genes from soil by PCR-DGGE and metagenome walking. *Appl Microbiol Biotechnol* 83(2):389-96 doi:10.1007/s00253-009-1992-x

Müller JA, Rosner BM, von Abendroth G, Meshulam-Simon G, McCarty PL, Spormann AM (2004) Molecular identification of the catabolic vinyl chloride reductase from *Dehalococcoides* sp. strain VS and its environmental distribution. *Appl Environ Microb* 70(8):4880 doi:10.1128/AEM.70.8.4880-4888.2004

Neidle EL, Shapiro MK, Ornston LN (1987) Cloning and expression in *Escherichia coli* of *Acinetobacter calcoaceticus* genes for benzoate degradation. *Journal of Bacteriology* 169(12):5496 doi:10.1128/jb.169.12.5496-5503.1987

Nejidat A, Kafka L, Tekoah Y, Ronen Z (2008) Effect of organic and inorganic nitrogenous compounds on RDX degradation and cytochrome P-450 expression in *Rhodococcus* strain YH1. *Biodegrad* 19(3):313-320 doi:10.1007/s10532-007-9137-3

Nyssönen M, Piskonen R, Itävaara M (2008) Monitoring aromatic hydrocarbon biodegradation by functional marker genes. *Environ Pollut* 154(2):192-202 doi:https://doi.org/10.1016/j.envpol.2007.10.009

Ogle DH, P. Wheeler, and A. Dinno (2020) FSA: Fisheries Stock Analysis. R package version 0.8.31.

Ohtsubo Y, Delawary M, Takagi M, Ohta A, Kimbara K, Nagata Y (2001) BphS, a Key Transcriptional Regulator of bph Genes Involved in Polychlorinated Biphenyl/Biphenyl Degradation in *Pseudomonas* sp. KKS102. *Journal of Biological Chemistry* 276(39):36146-36154 doi:10.1074/jbc.m100302200

Papadopoulos JS, Agarwala R (2007) COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23(9):1073-9 doi:10.1093/bioinformatics/btm076

Pasti-Grigsby MB, Lewis TA, Crawford DL, Crawford RL (1996) Transformation of 2,4,6-trinitrotoluene (TNT) by actinomycetes isolated from TNT-contaminated and uncontaminated environments. *Appl Environ Microb* 62(3):1120

Patrauchan MA, Florizone C, Dosanjh M, Mohn WW, Davies J, Eltis LD (2005) Catabolism of benzoate and phthalate in *Rhodococcus* sp. strain RHA1: redundancies and convergence. *Journal of bacteriology* 187(12):4050-4063

Patterson BM, Aravena R, Davis GB, Furness AJ, Bastow TP, Bouchard D (2013) Multiple lines of evidence to demonstrate vinyl chloride aerobic biodegradation in the vadose zone, and factors controlling rates. *Journal of Contaminant Hydrology* 153:69-77 doi:https://doi.org/10.1016/j.jconhyd.2013.07.008

Peakall DB (1975) Phthalate esters: occurrence and biological effects *Residue reviews*. Springer, pp 1-41

Peng P, Yang H, Jia R, Li L (2013) Biodegradation of dioxin by a newly isolated *Rhodococcus* sp with the involvement of self-transmissible plasmids. *Appl Microbiol Biot* 97(12):5585-5595 doi:10.1007/s00253-012-4363-y

Peng R-h, Xiong A-S, Xue Y, Fu X-Y, Gao F, Zhao W, Tian Y-S, Yao Q-H (2008) Microbial biodegradation of polyaromatic hydrocarbons. *Fems Microbiol Rev* 32:927-55 doi:10.1111/j.1574-6976.2008.00127.x

Peng R-H, Xiong A-S, Xue Y, Fu X-Y, Gao F, Zhao W, Tian Y-S, Yao Q-H (2010) A Profile of Ring-hydroxylating Oxygenases that Degrade Aromatic Pollutants. In: Whitacre DM (ed) *Reviews of Environmental Contamination and Toxicology* Volume 206. Springer New York, New York, NY, pp 65-94

Penton C, Johnson T, Quensen J, Iwai S, Cole J, Tiedje J (2013) Functional genes to assess nitrogen cycling and aromatic hydrocarbon degradation: primers and processing matter. *Frontiers in Microbiology* 4(279) doi:10.3389/fmicb.2013.00279

Pieper DH (2005) Aerobic degradation of polychlorinated biphenyls. *Appl Microbiol Biot* 67(2):170-191 doi:10.1007/s00253-004-1810-4

- Pinyakong O, Habe H, Omori T (2003) The unique aromatic catabolic genes in sphingomonads degrading polycyclic aromatic hydrocarbons (PAHs). *J Gen Appl Microbiol* 49(1):1-19 doi:10.2323/jgam.49.1
- Ramalingam V, Cupples A (2020) Enrichment of novel Actinomycetales and the detection of monooxygenases during aerobic 1,4-dioxane biodegradation with uncontaminated and contaminated inocula. *Appl Microbiol Biot* 104(5):2255-2269 doi:10.1007/s00253-020-10376-7
- Reiner AM (1971) Metabolism of benzoic acid by bacteria: 3,5-cyclohexadiene-1,2-diol-1-carboxylic acid is an intermediate in the formation of catechol. *Journal of bacteriology* 108(1):89-94 doi:10.1128/JB.108.1.89-94.1971
- Resnick S, Lee K, Gibson D (1996) Diverse reactions catalyzed by naphthalene dioxygenase from *Pseudomonas sp* strain NCIB 9816. *Journal of industrial microbiology* 17(5-6):438-457
- Rhee S-K, Liu X, Wu L, Chong SC, Wan X, Zhou J (2004) Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl Environ Microb* 70(7):4303-4317 doi:10.1128/AEM.70.7.4303-4317.2004
- Richards PM, Liang Y, Johnson RL, Mattes TE (2019) Cryogenic soil coring reveals coexistence of aerobic and anaerobic vinyl chloride degrading bacteria in a chlorinated ethene contaminated aquifer. *Water Res* 157:281-291 doi:https://doi.org/10.1016/j.watres.2019.03.059
- Roberts WC, Hartley WR (1992) *Drinking water health advisory: Munitions*. CRC Press
- Rylott EL, Jackson RG, Edwards J, Womack GL, Seth-Smith HM, Rathbone DA, Strand SE, Bruce NC (2006) An explosive-degrading cytochrome P450 activity and its targeted application for the phytoremediation of RDX. *Nature biotechnology* 24(2):216-219
- Rylott EL, Jackson RG, Sabbadin F, Seth-Smith HMB, Edwards J, Chong CS, Strand SE, Grogan G, Bruce NC (2011) The explosive-degrading cytochrome P450 XplA: Biochemistry, structural features and prospects for bioremediation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1814(1):230-236 doi:https://doi.org/10.1016/j.bbapap.2010.07.004
- Sabir DK, Grosjean N, Rylott EL, Bruce NC (2017) Investigating differences in the ability of XplA/B-containing bacteria to degrade the explosive hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX). *Fems Microbiol Lett* 364(14) doi:10.1093/femsle/fnx144
- Seeger M, Timmis KN, Hofer B (1995) Degradation of chlorobiphenyls catalyzed by the bph-encoded biphenyl-2,3-dioxygenase and biphenyl-2,3-dihydrodiol-2,3-dehydrogenase of *Pseudomonas sp.* LB400. *Fems Microbiol Lett* 133(3):259-264 doi:10.1111/j.1574-6968.1995.tb07894.x
- Selifonov S, Grifoll M, Eaton R, Chapman P (1996) Oxidation of Naphthenoaromatic and Methyl-Substituted Aromatic Compounds by Naphthalene 1,2-Dioxygenase. *Appl Environ Microb* 62:507-14 doi:10.1128/AEM.62.2.507-514.1996

Seth-Smith HMB, Edwards J, Rosser SJ, Rathbone DA, Bruce NC (2008) The Explosive-Degrading *Cytochrome P450* System Is Highly Conserved among Strains of *Rhodococcus spp.* Appl Environ Microbiol 74(14):4550 doi:10.1128/AEM.00391-08

Seth-Smith HMB, Rosser SJ, Basran A, Travis ER, Dabbs ER, Nicklin S, Bruce NC (2002) Cloning, Sequencing, and Characterization of the Hexahydro-1,3,5-Trinitro-1,3,5-Triazine Degradation Gene Cluster from *Rhodococcus rhodochrous*. Appl Environ Microbiol 68(10):4764 doi:10.1128/AEM.68.10.4764-4771.2002

Sheu YT, Lien PJ, Chen CC, Chang YM, Kao CM (2016) Bioremediation of 2,4,6-trinitrotoluene-contaminated groundwater using unique bacterial strains: microcosm and mechanism studies. International Journal of Environmental Science and Technology 13(5):1357-1366 doi:10.1007/s13762-016-0976-5

Simon MJ, Osslund TD, Saunders R, Ensley BD, Suggs S, Harcourt A, Wen-chen S, Cruder DL, Gibson DT, Zylstra GJ (1993) Sequences of genes encoding naphthalene dioxygenase in *Pseudomonas putida* strains G7 and NCIB 9816-4. Gene 127(1):31-37 doi:https://doi.org/10.1016/0378-1119(93)90613-8

Smith TJ, Lloyd JS, Gallagher SC, Fosdike WLJ, Murrell JC, Dalton H (1999) Heterologous expression of alkene monooxygenase from *Rhodococcus rhodochrous* B-276. European Journal of Biochemistry 260(2):446-452 doi:10.1046/j.1432-1327.1999.00179.x

Sondossi M, Sylvestre M, Ahmad D (1992) Effects of chlorobenzoate transformation on the *Pseudomonas testosteroni* biphenyl and chlorobiphenyl degradation pathway. Appl Environ Microb 58(2):485-495 doi:10.1128/AEM.58.2.485-495.1992

Talaska G, Underwood P, Maier A, Lewtas J, Rothman N, Jaeger M (1996) Polycyclic aromatic hydrocarbons (PAHs), nitro-PAHs and related environmental compounds: biological markers of exposure and effects. Environmental health perspectives 104 Suppl 5(Suppl 5):901-906 doi:10.1289/ehp.96104s5901

Team R (2020) RStudio: Integrated Development for R. RStudio, PBC, Boston, MA

Team RC (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Thompson KT, Crocker FH, Fredrickson HL (2005) Mineralization of the Cyclic Nitramine Explosive Hexahydro-1,3,5-Trinitro-1,3,5-Triazine by *Gordonia* and *Williamsia spp.* Appl Environ Microbiol 71(12):8265 doi:10.1128/AEM.71.12.8265-8272.2005

Van den Berg M, Birnbaum LS, Denison M, De Vito M, Farland W, Feeley M, Fiedler H, Hakansson H, Hanberg A, Haws L, Rose M, Safe S, Schrenk D, Tohyama C, Tritscher A, Tuomisto J, Tysklind M, Walker N, Peterson RE (2006) The 2005 World Health Organization Reevaluation of Human and Mammalian Toxic Equivalency Factors for Dioxins and Dioxin-Like Compounds. Toxicological Sciences 93(2):223-241 doi:10.1093/toxsci/kfl055

van der Zaan B, Hannes F, Hoekstra N, Rijnaarts H, de Vos WM, Smidt H, Gerritse J (2010) Correlation of *Dehalococcoides* 16S rRNA and chloroethene-reductive dehalogenase genes with geochemical conditions in chloroethene-contaminated groundwater. *Appl Environ Microbiol* 76(3):843-850 doi:10.1128/aem.01482-09

Wagoner JK (1983) Toxicity of vinyl chloride and poly(vinyl chloride): a critical review. *Environmental health perspectives* 52:61-66 doi:10.1289/ehp.835261

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

Widada J, Nojiri H, Omori T (2002) Recent developments in molecular techniques for identification and monitoring of xenobiotic-degrading bacteria and their catabolic genes in bioremediation. *Appl Microbiol Biotechnol* 60(1-2):45-59 doi:10.1007/s00253-002-1072-y

Wilson FP, Liu X, Mattes TE, Cupples AM (2016) *Nocardioideae*, *Sediminibacterium*, *Aquabacterium*, *Variovorax*, and *Pseudomonas* linked to carbon uptake during aerobic vinyl chloride biodegradation. *Environ Sci Pollut Res Int* 23(19):19062-70 doi:10.1007/s11356-016-7099-x

## Chapter 5 Conclusions

In both the aerobic and anaerobic studies, the biodegradation of 1,4-dioxane was observed.

However, the degradation rates were significantly slower under reducing conditions.

Aerobically, the degradation of 1,4-dioxane was observed within 30 days after each re-spike event. Anaerobically, in many cases, the timeline stretched to more than one year to observe any significant differences in concentrations between the samples and abiotic controls. The CSIA was found to be a useful tool to provide another line of evidence for anaerobic biodegradation of 1,4-dioxane.

All of the functional genes associated with biodegradation of 1,4-dioxane in agricultural soils were detected under aerobic conditions during 1,4-dioxane biodegradation (Chapter 2). The shotgun sequencing conducted as part of the work, has the added advantage of both taxonomic and functional analysis of microbial communities. This successfully links both taxonomic and functional data to create a better idea of the genes and microbes associated with 1,4-dioxane biodegradation.

Chapter 3 demonstrated 1,4-dioxane degradation under highly reducing conditions is feasible.

This, in turn, could indicate the potential of natural attenuation as a clean-up approach at anoxic contaminated sites. This is specifically useful at sites where time may not be a constraint. The microbial community benefitting from 1,4-dioxane under anaerobic conditions are also indicated.

Chapter 4 examines the presence, abundance, and diversity of biodegradation genes. The work also sheds light on the possibility of many new genera being potentially associated with the biodegradation of organic contaminants. As evidenced by this study, the presence of biodegradation genes linked to organic contamination are distributed in nature. However, this is not an indication of gene expression or enzyme activity. The work was conducted with uncontaminated agricultural soils and all of the genes examined were present, except for *vcrA*.

### 5.1. Future Work

- Shotgun sequencing of DNA can provide a general overview of genes and genetic diversity of the microbial cultures. However, it cannot differentiate between expressed or non-expressed genes in a given sample (Zhou et al. 2015). A combined omics approach can be useful as metatranscriptomics which involves expressed microbial community RNA.
- The microbial community identified as benefitting from 1,4-dioxane anaerobically could be examined further e.g. shotgun sequencing, stable isotope probing, isolations.
- A valuable addition to the existing work would be to investigate contaminated sediments with CSIA to confirm anaerobic biodegradation.
- Since 1,4-dioxane contamination is often accompanied by other chlorinated solvents, it would be useful to examine the biodegradation of multiple contaminants simultaneously including trichloroethylene and trichloroethene