Iowa Research Online
The University of Iowa's Institutional Repository

University of Iowa
Iowa Research Online

Theses and Dissertations

Fall 2016

Aerobic vinyl chloride degradation at the microbial community level

Xikun Liu
University of Iowa

Copyright © 2016 Xikun Liu

This dissertation is available at Iowa Research Online: https://ir.uiowa.edu/etd/2238

Recommended Citation
Liu, Xikun. "Aerobic vinyl chloride degradation at the microbial community level." PhD (Doctor of Philosophy) thesis, University of Iowa, 2016.
https://doi.org/10.17077/etd.ci5hq579.

Follow this and additional works at: https://ir.uiowa.edu/etd


Part of the Civil and Environmental Engineering Commons

AEROBIC VINYL CHLORIDE DEGRADATION AT THE MICROBIAL COMMUNITY LEVEL

by
Xikun Liu

A thesis submitted in partial fulfillment
of the requirements for the Doctor of Philosophy degree
in Civil and Environmental Engineering
in the Graduate College of
The University of Iowa

December 2016

Thesis Supervisor: Associate Professor Timothy E. Mattes

Copyright by

XIKUN LIU

2016

All Rights Reserved

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Xikun Liu

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy degree
in Civil and Environmental Engineering at the December 2016 graduation.

Thesis Committee: _____
Timothy E. Mattes, Thesis Supervisor

_____
Gene F. Parkin

_____
Jerald L. Schnoor

_____
Kevin L. Knudtson

_____
Alison M. Cupples

_____
Gregory H. LeFevre

ACKNOWLEDGEMENTS

Time flies when you are having fun. I could not believe five years and a half has passed by since I arrived in the United States.

First and foremost, I would like to thank my advisor, Professor Timothy E Mattes. When came to the University of Iowa, I was on a non-thesis track. Later on, I developed interest in microbial research, and Dr. Mattes kindly allowed me to work voluntarily in his lab, which later on turned into a funded research role. Dr. Mattes has been an excellent advisor, who has provided enormous guidance on experiments and especially writing, for a non-native English speaker like me. Nonetheless, the best things I appreciated about Dr. Mattes are his rigorous scientific attitude and trust in me to explore the research questions I am interested with new technologies.

I am grateful to be in such a small but nice lab group. Dr. Yang Oh Jin and Mengchen Lee handed me the initial labs skills for microbiology research. Carly and Tristan helped me constructed and monitored the enrichment cultures. Special thanks to my “secondary advisor”, collaborator and close friend, Dr. Yi Liang, for supporting me not only in my research career, but also my personal life. I also would like to thank other previous and present members, Adina Howe, Caitlin Krause, Sean Lee, Trisha Madhavan, Patrick Richard and Jessica Ewald, for supporting me and being such friendly co-workers.

The Civil and Environmental Engineering department in University of Iowa has a nourishing and collaborative spirit. I enjoyed seminars, lectures and parities with all my colleagues. The faculties and graduate students are some of the friendliest people I have ever met. I would like to pay special thanks to Professor Just’s group. As we were the only two groups doing microbiology research in the department, Ellen and Hunter shared much experience and thoughts with me. I also want to thank our lab managers and postdoc, John Durst, Eric Jetter, Deb Williard and Drew Latta for providing technical

assistances, and to Judy Holland, Jennifer Rumping and Kim Lebeck for their excellent administrative service.

Most of my projects were accomplished through collective efforts of people from different institutes. My top co-authors are Dr. Alison Cupples and Fernanda Paes Wilson from Michigan State University, with two paper published and another two in process. I would like to thank Dr. Cupples to provide me guidance on dissertation and science career, and Fernanda for conducting the laborious stable isotope probing (SIP) experiments, and teaching me about SIP. I also would like to thank Dr. Ke Yu and his student Yang Wu for the metagenomic analysis. Dr. Ke Yu was a complete stranger to me before Dr. Yujie Men's introduction, but he was so patient to teach me through the metagenomics analysis. I was highly motivated by his enthusiasm about science.

My time at Iowa was enjoyable, also largely due to my friends and family, both in the US and back in China. They have given me unconditional love and supports. Particularly, I want to thank my parents for understanding me and let their only child to travel to the other side of the planet to pursue her dream. I want to thank my husband, Cheng Chen. He has been a role model for me, who is always hard-working, composed, encouraging and sometimes funny. He is my best friend in almost every aspects of life.

Last but not least, I would like to extend my thanks to my committee members: Gene Parkin, Jerald Schnoor, Alison Cupples, Kevin Knudtson, Gregory LeFevre and Timothy Mattes. Thank you for reading this long but hopefully interesting thesis and providing insightful comments.

This research was funded by the National Science Foundation Grant No. 1233154, Graduate College Post-Comprehensive Research Award, Graduate College Summer Fellowship and Ballard Seashore Dissertation Fellowship from the University of Iowa.

Xikun Liu

The University of Iowa

October 2016

ABSTRACT

Vinyl chloride (VC) is a human carcinogen and common groundwater contaminant in the United States. Some of the indigenous bacteria can utilize VC for their growth, which is important for bioremediation. As previous studies have been majorly focused on VC-degrading bacteria in pure cultures, we initiated the study to investigate the microbial community structure and interactions in more complicated systems, such as mixed-pure cultures and groundwater enrichment cultures. Finally, we extended our study into the field by investigating the diversity and abundance of functional genes in VC-assimilating pathways at six contaminated sites.

In our first study, *Nocardioides* was found to be the most dominant genus in Carver groundwater enrichment cultures via stable-isotope probing and 16S rRNA gene amplicon Illumina sequencing. As cross-feeding was observed, in the second study, mixed-pure culture experiment was conducted to explore the potential effects of VC-assimilating *Nocardioides* on other bacteria, which showed VC cometabolizer *Mycobacterium* strain JS622 would take up carbon from VC to sustain their growth when mixed with VC-assimilating *Nocardioides* sp. strain JS614. The third study was conducted with a different groundwater source from Fairbanks, AK, which again showed *Nocardioides* is dominant in the microbial community. A novel VC-assimilating *Nocardioides* sp. bacteria was isolated, named XL1. The putative genome of XL1 extracted from enrichment culture metagenome was 99% to 100% identical to strain JS614, with a plasmid genome bin similar to strain JS614 plasmid pNOCA01, though the morphology of strain XL1 was distinct from strain JS614. About 90% of the plasmid contigs could be mapped onto *Nocardioides* sp. strain JS614 plasmid with 100% identity,

containing known aerobic ethene and VC degradation pathway genes encoding alkene monooxygenase and epoxyalkane: coenzyme M transferase (EaCoMT). Glutathione synthase and glutathione S-transferase genes, possibly involved in epoxide detoxification, were found in *Polaromonas*, *Mesorhizobium* and *Pseudomonas*-affiliated genome bins. The study also showed cultures adapted to VC faster after amended with ethene. The in-situ study (the fourth study) revealed 192 different EaCoMT T-RFs from six chlorinated ethene contamination sites via T-RFLP analysis, implicating higher EaCoMT diversity than previously known. Phylogenetic analysis revealed that a majority of the 139 cloned sequences (78.4%) grouped with EaCoMT genes found in VC- and ethene-assimilating *Mycobacterium* strains and *Nocardioides* sp. strain JS614. EaCoMT gene abundance was positively correlated with VC and ethene concentrations at the sites studied.

PUBLIC ABSTRACT

Vinyl chloride (VC) is a human carcinogen and common groundwater contaminant in the US. Bacteria in nature can utilize VC for their growth, which is important for bioremediation. As previous studies have been focused on pure VC-degrading bacteria, we initiated the study to investigate more complicated systems, such as mixed-pure cultures and groundwater enrichment cultures. Finally, we extended our study into the contaminated sites by investigating the diversity and abundance of functional genes related to VC assimilation.

In our study, microbial community could adapt to VC faster after feeding ethene. *Nocardioides* was found to be the most dominant genus in three groundwater enrichment cultures from contaminated sites in Carver, MA and Fairbanks, AK. A novel VC-assimilating *Nocardioides* bacteria was isolated, named XL1. It also contains a plasmid nearly identical to *Nocardioides* sp. strain JS614, which bears known genes encoding the key enzymes in VC-assimilating pathway: alkene monooxygenase and epoxyalkane: coenzyme M transferase (EaCoMT). There were other bacteria abundant in the cultures, including *Polaromonas*, *Mesorhizobium* and *Pseudomonas*, which have glutathione S-transferase that could also detoxify VC epoxide. The in-situ study revealed 192 different EaCoMT types from six chlorinated ethene contaminated sites via T-RFLP analysis, implicating higher EaCoMT diversity than previously known. A majority of the EaCoMT recovered from the environment are similar to those in *Mycobacterium* strains and *Nocardioides* sp. strain JS614. EaCoMT gene abundance was positively correlated with VC and ethene concentrations at the sites studied. Overall, these findings provides scientific basis for VC bioremediation practice.

TABLE OF CONTENTS

List of Tables xiii

List of Figures xv

Chapter I. Research overview and hypotheses 1

Chapter II. Background 4

2.1 Vinyl chloride 4

2.1.1 Physicochemical properties 4

2.1.2 Toxicity 4

2.1.3 Sources 4

2.2 Vinyl chloride-degrading bacteria 5

2.2.1 Anaerobic degradation of VC 5

2.2.2 Aerobic degradation of VC 6

2.3 Microbial ecology of aerobic VC-assimilating communities 7

2.4 Possible adaptation and evolutionary mechanisms of VC degraders 9

2.4.1 Selective enrichment of a population (clonal amplification) 9

2.4.2 Missense mutation and larger scale rearrangement 9

2.4.3 Horizontal gene transfer (HGT) 10

2.4.4 Other possible mechanisms in VC-assimilating culture 10

2.5 Major molecular analysis tools for this project 11

2.5.1 Shotgun metagenomic sequencing 11

2.5.2 DNA-stable isotope probing (DNA-SIP) 13

2.5.3 Quantitative polymerase chain reaction (qPCR) 14

2.5.4 Terminal restriction fragment length polymorphism (T-RFLP) 15

Chapter III. Identifying bacteria linked to carbon uptake from vinyl chloride in groundwater enrichment cultures using stable isotope probing (SIP) 16

3.1 Abstract 16

3.2 Introduction 17

3.3 Materials and Methods 20

3.3.1 Experimental design for SIP 20

3.3.2 Analytical methods 21

3.3.3 DNA extraction, ultracentrifugation and fractionation 21

3.3.4 MiSeq Illumina sequencing and SIP fraction analysis 22

3.3.5 Primer design, 16S rRNA gene clone libraries and qPCR 23

3.3.6 Functional gene (etnC and etnE) qPCR 25

3.3.7 Microbial community analysis during VC degradation 26

3.4 Results and Discussion 26

3.4.1 VC degradation and the identification of phylotypes responsible for label uptake 26

3.4.2 Primer design, specificity test and 16S rRNA abundance in SIP fractions 31

3.4.3 Functional gene (etnC and etnE) abundance in SIP fractions 34

3.4.4 Microbial community changes during VC degradation 36

3.5 Conclusions 38

Chapter IV. Carbon uptake of vinyl chloride (VC)-assimilators and cometabolizers from VC as revealed by stable isotope probing (SIP) on mixed pure cultures 39

4.1 Abstract 39

4.2 Introduction 40

4.3 Materials and Methods 42

4.3.1 Chemicals, media and analytical method 42

4.3.2 Preparation of bacterial strains 43

4.3.3 Purity check 44

4.3.4 Confirmation of assimilating and cometabolizing activities..... 46

4.3.5 Stable isotope probing experiment..... 46

4.3.6 DNA extraction, fractionation and purification 47

4.3.7 Quantitative PCR on functional gene etnE 49

4.3.8 T-RFLP on 16S rRNA gene and functional gene etnE 50

4.4 Results..... 51

4.4.1 VC biodegradation patterns in strain JS614 and JS622 pure cultures 51

4.4.2 Bacterial growth and VC biodegradation patterns in a defined mixed culture of strains JS614 and JS622 55

4.4.3 Evidence of VC assimilation in mixed culture from the SIP-qPCR data 57

4.4.4 Carbon up-take from VC and growth observed in JS622 59

4.5 Discussion..... 62

4.5.1 JS622 can take up ^{13}C from VC when mixed with VC-assimilator JS614 62

4.5.2 Faster degradation of VC in mixed culture..... 64

4.5.3 The cause for irregular pattern in SIP-qPCR 64

4.5.4 COM primer set tends to amplify etnE1 66

4.6 Conclusion 67

Chapter V. Microbial adaptation to vinyl chloride in groundwater microcosms as revealed by metagenomics and other molecular tools 68

5.1 Abstract 68

5.2 Introduction..... 69

5.3 Materials and Methods..... 71

5.3.1 Groundwater and environmental sample collection..... 71

5.3.2 DNA extraction from groundwater samples and functional gene qPCR 72

5.3.3 Construction of VC- and ethene-fed enrichment cultures 72

5.3.4 Analytical methods, DNA and RNA extraction and reverse transcription-qPCR of enrichment cultures 75

5.3.5 16S rRNA gene sequencing and data analysis 75

5.3.6 Metagenomic sequencing and initial data analysis 79

5.3.7 Assembly, genome binning and annotation of metagenomes 82

5.3.8 DNA Stable isotope probing (SIP) 83

5.3.9 Primer design, verification and qPCR on SIP fractions 88

5.3.10 Isolation and Identification of ethene and VC degraders 89

5.3.11 Sequence accession 90

5.4 Results 91

5.4.1 Initial characterization of the groundwater sample 91

5.4.2 Faster adaptation to VC after ethene enrichment 93

5.4.3 Shifts in microbial community taxonomic structure after ethene and VC enrichment 97

5.4.4 SIP reveals primary bacteria involved in carbon uptake from VC 101

5.4.5 Isolation and characterization of bacteria from enrichment cultures 103

5.4.6 Genome binning revealed multiple Nocardioides strains and one plasmid 108

5.4.7 Expression of etnC and etnE genes in the enrichment cultures 119

5.4.8 Analysis of other genome bins 120

5.4.9 Identifying Nocardioides in the original groundwater 120

5.5 Discussion 121

5.6 Conclusions 130

Chapter VI. “Into the wild”: Epoxyalkane: coenzyme M transferase gene diversity and abundance in groundwater samples from chlorinated ethene contaminated sites 131

6.1 Abstract 131

6.2 Introduction 132

6.3 Materials and Methods 135

6.3.1 Site information, environmental sample collection and DNA extraction 135

6.3.2 Quantitative PCR 141

6.3.3 PCR amplification of EaCoMT genes 141

6.3.4 Terminal-restriction fragment length polymorphism (T-RFLP) analysis 145

6.3.5 T-RFLP data analysis 145

6.3.6 VC attenuation rate estimation and correlation analysis 148

6.3.7 Cloning and sequencing 148

6.3.8 Sequence analyses 151

6.3.9 Genbank submissions 151

6.4 Results 154

6.4.1 T-RFLP analysis of EaCoMT gene diversity in environmental samples 154

6.4.2 Phylogenetic analysis of EaCoMT gene diversity in environmental samples 158

6.4.3 Relationships between EaCoMT gene diversity, EaCoMT gene abundance, and contaminated site conditions 163

6.4.4 PCR modifications affect EaCoMT gene diversity estimates 167

6.4 Discussion 168

6.5 Conclusion 173

Chapter VII. Engineering and scientific significance 176

7.1 Distribution of VC-assimilators in the environment 176

7.2 A novel VC-assimilating *Nocardioides* strain 177

7.3 VC-degradation in complex cultures 178

7.4 Application of metagenomic analysis in bioremediation 178

Appendix I: Supplemental material 180

Supplemental Material for Chapter III 180

Supplemental Material for Chapter IV 181

Supplemental Material for Chapter V 182

Supplemental Material for Chapter VI 201

Appendix II: Other works during PhD research 214

Elucidating carbon uptake from vinyl chloride using stable isotope probing and Illumina sequencing 214

Partial nitritation ANAMMOX in submerged attached growth bioreactors with smart aeration at 20°C 216

Abundance and activity of vinyl chloride (VC)-oxidizing bacteria in a dilute groundwater VC plume biostimulated with oxygen and ethene 218

Reconstructing ancient prokaryotic lake communities from Lake Leija paleosediments in the Atacama desert 220

References 222

LIST OF TABLES

Table 4.1 Amount of VC degraded in each culture. Time point 1: ~20 µmol VC degradation; time point 2: ~50 µmol VC degradation; time point 3: ~70 µmol VC degradation. 48

Table 5.1 Summary of 16S rRNA gene amplicon Illumina sequencing data: number of sequences, number of OTUs, Chao index and Shannon index. 78

Table 5.2 Metagenomes used in this study (post-quality-screening). 81

Table 5.3 Summary of genome bins from ethene and VC culture. Completeness and contamination were calculated with CheckM [119]. 86

Table 5.4 Contigs from VC-PLSD of <100% nucleotide identity to *Nocardioides* sp. strain JS614 plasmid pNOCA01 (Genbank accession no. CP000509.1). 111

Table 5.5 Completeness of selective metabolic pathways and existence of genes in alternative ethene/VC assimilating pathway among selected genome bins, based on annotation with KEGG and NCBI NR databases. 114

Table 6.1 Summary of groundwater samples information in this study. Statistically significant k values are based on p<0.1 136

Table 6.2 Field groundwater geochemical parameters collected from the monitoring wells in this study. These parameters and groundwater velocities were taken from the reports cited in Table 6.1. 140

Table 6.3 EaCoMT clone sequences (139 in total) retrieved as a result of different PCR protocols attempted to amplify EaCoMT genes from environmental samples. 144

Table 6.4 Summary of REPK (4) predicted T-RF lengths from the products of F131/R562 primers (expected size 447-453bp) when subjected to in silico digestion with the AcoI (EaeI) restriction enzyme. 146

Table 6.5 Genbank accession numbers of existing EaCoMT sequences used in phylogenetic and diversity analyses in this study. 152

Table 6.6 Genbank accession numbers of MetE sequences used as outgroup sequences in phylogenetic tree construction. 153

Table 6.7 In-vitro and in-silico AcoI (EaeI) digestion of selected EaCoMT gene clones (those have <99% identity with other clone sequences retrieved from the same sample) matched with similar EaCoMT gene sequences from isolated strains. 156

Table 6.8 Correlation analysis between geochemical parameters with EaCoMT gene abundance and diversity. 165

Table 6.9 Examples of EaCoMT sequences in VC-, ethene- and propene-assimilating bacteria that are incorrectly annotated as MetE sequences in Genbank. 175

Table AI.1 Detailed qPCR information for Chapter III. 180

Table AI.2 Detailed qPCR information for Chapter IV. 181

Table AI.3 Primers used in Chapter V. 182

Table AI.4 Classification of clones with RDP classifier. 183

Table AI.5 Detailed qPCR information for Chapter V. 185

Table AI.6 Composition of microbial communities as revealed by metagenomics sequencing and 16S rRNA gene amplicon Illumina sequencing in Chapter V. 186

Table AI.7 Oligonucleotide primers used in Chapter VI. 201

Table AI.8 Information of qPCR parameters in this study. 202

Table AI.9 Summary of EaCoMT sequences found in the MG-RAST metagenomics database by July 2015. MG-RAST metagenome numbers are provided, along with the sample source and location, the top BLAST hit, abundance of each EaCoMT gene type in the sample, % identity and average amino acid alignment length. 203

LIST OF FIGURES

Figure 2.1 Anaerobic reductive dechlorination of PCE and TCE 5

Figure 2.2 Proposed aerobic degradation pathway of VC and ETH in VC/ETH assimilators. 8

Figure 2.3 Shotgun metagenomic sequencing work flow. 12

Figure 2.4 Example of SIP with ethidium bromide staining in CsCl gradient. 13

Figure 2.5 Example of qPCR standard curves. 14

Figure 2.6 Example of T-RFLP electropherogram. The x-axis represents T-RF length, the y-axis represents peak height. Standards are orange-colored and T-RFs in the sample are blue-colored. 15

Figure 3.1 Percent VC remaining in cultures amended with labeled vinyl chloride (empty square), unlabeled vinyl chloride (solid square) and in the abiotic controls (solid diamond). Arrows indicate when DNA was extracted. (By Fernanda Paes Wilson) 28

Figure 3.2 DNA concentration (ng/µL) at day 3 (A) and day 7 (B) in fractions obtained from the labeled VC and unlabeled VC amended cultures. The complete and dashed lines represent DNA concentrations from the unlabeled and labeled VC amended cultures, respectively. (By Fernanda Paes Wilson) 29

Figure 3.3 Relative abundance of phylotypes in the four heavy fractions from labeled VC amended cultures. The error bars represent standard deviations from six samples submitted for sequencing. (By Fernanda Paes Wilson) 30

Figure 3.4 *Sediminibacterium*, *Aquabacterium* and *Variovorax* 16S rRNA gene copies over the buoyant density range in DNA extracted from labeled and unlabeled VC amended cultures at days 3 and 7. Error bars represent standard deviation from three qPCR measurements. (By Fernanda Paes Wilson) 32

Figure 3.5 The *etnE* and *etnC* gene copies over the buoyant density range after fractionation from labeled and unlabeled VC amended cultures at days 3 and 7. The data points represent the average of duplicates and the error bars depict the range detected in qPCR. (By Xikun Liu) 35

Figure 3.6 Percent vinyl chloride remaining in triplicate live cultures and an abiotic control. The arrows indicate when DNA was extracted from these samples for 16S rRNA gene amplicon Illumina sequencing. (By Fernanda Paes Wilson) 36

Figure 3.7 The relative abundance of *Pseudomonas* (A), *Sediminibacterium* (B), *Aquabacterium* (C) and *Nocardioides* (D) during vinyl chloride degradation. (By Fernanda Paes Wilson) 37

Figure 4.1 Reviving Nocardioides sp. strain JS614 and Mycobacterium strain JS622 on ethene as the sole carbon source. 45

Figure 4.2 Nocardioides sp. strain JS614 assimilates VC (A); Mycobacterium strain JS622 could not use VC as the sole carbon source (B). 52

Figure 4.3 Nocardioides sp. strain JS614 assimilates VC (A); Mycobacterium strain JS622 could not use VC as the sole carbon source (B). (Replicate) 53

Figure 4.4 Nocardioides sp. strain JS614 (A and B) and Mycobacterium strain JS622 (C and D) could degrade VC when ethene was present. 54

Figure 4.5 The growth (assessed by OD600) and degradation of VC in mixed cultures. Three were amended with ^{13}C labeled VC (upper panel), the other three were amended with unlabeled VC (lower panel). 56

Figure 4.6 Evidence of ^{13}C take-up: Total DNA concentration across BD after ~70 µmol of VC degradation (A); etnE gene abundance in the mixed cultures after ~20 µmol (B), ~50 µmol (C) and ~70 µmol (D) of VC degraded across BD range. Error bars indicate the differences in qPCR replicates. 58

Figure 4.7 T-RFLP on 16S rRNA gene and functional gene etnE for Nocardioides sp. strain JS614 (A, C, D) and Mycobacterium strain JS622 (B, E) across SIP fractions with BD from 1.68-1.78 g mL^{-1} in cultures after ~70 µmol of VC was degraded. 60

Figure 4.8 T-RFLP of 16S rRNA genes on labeled and unlabeled VC amended cultures after ~20 µmol (C13-VC-1, C12-VC-1), ~50 µmol (C13-VC-2, C12-VC-2) and ~70 µmol of VC (C13-VC-3, C12-VC-3) was degraded. Error bars indicate the differences in T-RFLP replicates. 61

Figure 5.1 Experimental design of this study. 74

Figure 5.2 Rarefaction curve of: A) 16S rRNA gene amplicon Illumina sequencing based on a 97% OTU cutoff. Data was subsampled with the size of the smallest sample (15,463 sequences); B) metagenomics sequencing. 77

Figure 5.3 Example of genome binning using differential coverage method[115] for A) ethene cultures and B) VC cultures. 85

Figure 5.4 Functional gene *etnC* and *etnE* abundances in groundwater at the Fairbanks, AK site. 92

Figure 5.5 Consumption of ethene and VC and growth of bacteria (as estimated with OD600 measurements) in enrichment cultures: A) Group 1 ethene cultures; B) Group 1 VC cultures; C) Group 2 ethene cultures; D) Group 2 VC cultures. 94

Figure 5.6 VC and ethene (ETH) degradation and bacteria growth in: A) Group 3 VC cultures (GW:MSM=1:1); B) Group 4 VC cultures (GW:MSM=1:3); C) Ethene(ETH) abiotic control; D) VC Abiotic control. 95

Figure 5.7 Consumption of ethene and VC and growth of bacteria (as estimated with OD600 measurements) in enrichment cultures at 4 °C. 96

Figure 5.8 Phylogenetic composition of microbial communities on A) phylum and B) genus level, based on the MG-RAST best hit analysis of shot-gun metagenomics sequencing data (inner circle) and Mothur analysis of 16S rRNA gene Illumina sequencing data (outer circle). "Others" represents the combination of the remaining minor taxa. 99

Figure 5.9 Replicates of metagenomics sequencing and 16S rRNA gene amplicon Illumina sequencing of group 2 time point 2 ethene and VC cultures (G2-TP2-ETH, G2-TP2-VC) showed the results are reproducible. 100

Figure 5.10 Example of SIP-qPCR data: 16S rRNA gene qPCR targeting genera *Nocardioides*, *Pseudomonas*, *Pedobacter* and *Sediminibacterium* on SIP fractions from Group B at day 7. A full list of graphs is provided in Appendix Figure AI.1. 102

Figure 5.11 The isolate *Nocardioides* sp. strain XL1 showed growth and degradation of ethene and VC. 105

Figure 5.12 Morphology of *Nocardioides* sp. strain XL1 on A) MSM plate in ethene environment and B) 1/TSAG plate in ambient air. 106

Figure 5.13 Morphology of *Nocardioides* sp. strain XL1 in VC-MSM culture (right), which contained clumping flocs in both ethene and VC cultures. The liquid culture of *Nocardioides* sp. strain JS614 in VC-MSM culture (left) is provided for comparison, which remained homogenous in both ethene and VC cultures. 107

Figure 5.14 Comparison of putative plasmid bin ETH-PLSD and VC-PLSD with the completed *Nocardioides* sp. strain JS614 plasmid pNOCA01 using BLAST and visualized with BRIG [123]. 110

Figure 5.15 The contig coverages from different genome bins in ethene and VC enrichment cultures visualized as heatmaps in R [47]. 118

Figure 5.16 RT-qPCR data of functional gene *etnC* and *etnE* transcripts abundance in ethene and VC enrichment cultures at the last time point (TP3, 242-256 days). 119

Figure 5.17 Phylogenetic tree of complete GST aa sequences from *Pseudomonas*-affiliated genome bin (VC-PRT1) (in bold) and their top blast hits and GST aa sequences from *Rhodococcus* sp. AD45 (red-colored). 128

Figure 6.1 Examples of bulk VC attenuation rate (k_{bulk}) (A) and point VC decay rate (k_{point}) (B) calculations. 150

Figure 6.2 Clustered heatmap of T-RFLP profiles generated by AcoI (EaeI) restriction-digested partial EaCoMT genes..... 155

Figure 6.3 NMDS ordination of EaCoMT gene T-RFLP profiles from different groundwater samples. Sample identifiers are formatted by site, well number, and replicate (A or B). VC concentrations estimated in each of the wells is provided for reference. The final stress was 0.1402. 157

Figure 6.4 A phylogenetic tree depicting the relationship of deduced EaCoMT sequences from environmental samples in this study (Table 6.3), enrichment cultures and isolates from Genbank (Table 6.5). 160

Figure 6.5 A phylogenetic tree depicting the relationship of deduced EtnE amino acid sequences from environmental samples, enrichment cultures and isolates (Table 6.5), along with MetE sequences from related strains (Table 6.6). 162

Figure 6.6 Linear regression of EaCoMT gene (*etnE*) abundance with vinyl chloride (VC) (A) and ethene (B) concentration. Each data point represents a sampling well from the sites surveyed in this study (Table 6.1). 166

Figure 6.7 Analysis of potential bias introduced by touch down and nested PCR. A composite DNA sample from Australia (AUS39) was used for T-RFLP analysis of EaCoMT genes..... 167

Figure AI.1 qPCR of 16S rRNA gene using specific primers on SIP fractions to target: A) *Nocardioides*; B) *Pseudomonas*; C) *Pedobacter*; D) *Sediminibacterium*. Two time points were selected for the SIP analysis: day 3 and day 7. 197

CHAPTER I. RESEARCH OVERVIEW AND HYPOTHESES

Many pollutants can be degraded through microbial activities at contaminated sites, the process of which is referred to as natural attenuation. It is a cost-effective and environmental-friendly approach for remediation. Cleaning-up of vinyl chloride (VC) in groundwater is one of the examples. From previous lab-scale studies, under aerobic condition, VC can be cometabolized and even assimilated by etheneotrophs, which use ethene, a co-occurring compound with VC at most contaminated sites, as a growth substrate. However, previous aerobic VC-degradation studies have only focused on pure etheneotroph cultures. The microbial community structure and interaction within complex enrichment culture has never been investigated. Therefore, the **main objective** of my PhD study is to **capture the taxonomy and functional changes of microbial community during adaption to VC and explore alternative VC degradation pathways.** Our central hypothesis is that as a complex microcosm adapting to ethene and finally to VC as the sole carbon source, microbial diversity will decrease to form a specialized system, in which bacteria responsible for VC-assimilation will dominate, and other bacteria playing roles as scavengers or partially degrading VC and its epoxides will also remain in the culture.

We tested our central hypothesis by the accomplishing the following specific aims:

1. **Design Stable Isotope Probing (SIP) experiment to identify VC-assimilating bacteria (VC-assimilator) in contaminated groundwater enrichment cultures (Chapter III).**

   We performed SIP experiment on groundwater enrichment cultures to

identify bacteria able to up-take ^{13}C from $^{13}C_2$-VC. We hypothesized that some previously found VC-assimilators would be seen in the enrichment cultures, e.g. *Nocardioides* and *Mycobacteria*, and they would be able to up-take ^{13}C. However, we also hypothesized that some bacteria enriched in the culture will not be able to take ^{13}C, as they might be scavengers.

2. Investigate the interaction between VC-cometabolizing bacteria (VC-cometabolizer) and VC-assimilator by using pure cultures of known bacteria (Chapter IV).

We used mixed pure cultures (VC-assimilators and cometabolizers) to test up-take of ^{13}C from $^{13}C_2$-VC. Our hypothesis was VC-assimilator could up-take ^{13}C, whereas VC-cometabolizers could not up-take ^{13}C within a short period of time.

3. Study taxonomy and functional changes of microbial community while adapting to ethene and VC using metagenomic sequencing (Chapter V).

We took a closer look at the ethene and VC-assimilating cultures by sequencing the whole metagenome of the cultures. We hypothesized that we could find the pathway of scavengers and potentially some novel VC-degradation pathways.

4. Isolate VC-assimilators (Chapter V).

We isolated bacteria that are able to grow on ethene and VC. We hypothesized that novel VC-assimilator could be discovered in our cultures. They may have the functional genes encoded on a plasmid similar to known VC-

assimilators.

As the genes related to VC-assimilation will be identified during achieving of the previous specific aims, we will further investigate whether these gene are prevalent at contaminated sites and try to correlate it with VC degradation rates in-situ. To do that, we conducted:

5. Investigation on the diversity and abundance of functional genes related to VC-assimilation at contaminated sites (Chapter VI).

Our hypothesis was that the functional genes involved in VC-assimilating process (e.g. *etnE*) exist at contaminated sites; the abundance of functional genes is positively correlated with ethene and VC concentration and attenuation rates at contaminated sites.

CHAPTER II. BACKGROUND

2.1 Vinyl chloride

2.1.1 Physicochemical properties

Under standard conditions, VC is a colorless and flammable gas with a sweet odor. It is very volatile (vapor pressure=2.66 × 103 mmHg, Henry's Constant 0.0278 $m^3 \cdot atm \cdot mol^{-1}$ at 25 °C)[1] and of low water solubility (0.11 g/100 mL at 25 °C). It is heavier than air (0.9106 g cm^{-3}) and is soluble in almost all organic solvents.

2.1.2 Toxicity

Vinyl chloride is a known human carcinogen and common groundwater contaminant. VC toxicity stems from its oxidized form, chlorooxirane, which is a highly reactive epoxide that can form DNA adducts and cause mutation [2]. Chlorooxirane can be readily formed in human body through the initial metabolism of VC by cytochrome P450 in the liver. It has a low maximum contaminants limit (MCL) in drinking water (2 ppb, compared to 5 ppb of PCE and TCE)[3].

2.1.3 Sources

VC is naturally occurring, however at very low concentration. It has been proposed as generated during the oxidative degradation of organic matter in soil, for example, in a reaction between humic substances, chloride ions and an oxidant (ferric ions or hydroxyl radicals)[4]. Most VC occurrences are considered as through anthropological processes. It is most commonly produced as monomer of polyvinyl chloride for plastic production.

The major source of VC in groundwater is from anaerobic dechlorination of tetrachloroethene (PCE) and trichloroethene (TCE) at contaminated sites[5]. PCE and TCE are chlorinated solvents widely used in civilian, industrial and military services. These chemicals were usually disposed in underground storage tanks, which might have erosion problems, and in some cases the chemicals were inappropriately disposed directly onto the ground. The chemicals travel into groundwater plume and form dense non-aqueous phase liquid (DNAPL) below groundwater, where PCE and TCE can be reduced to VC and ethene (ETH) by bacteria such as *Dehalococcoides* spp. [6, 7] through the following processes:

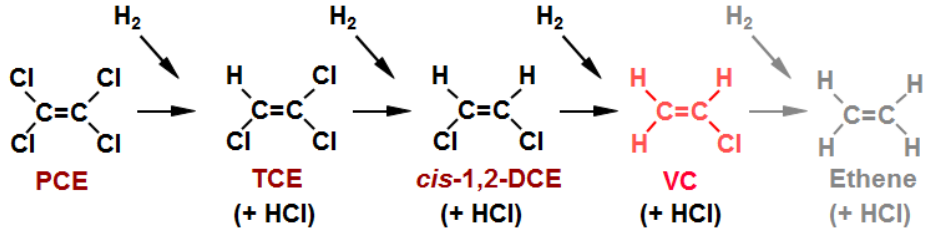


Figure 2.1 Anaerobic reductive dechlorination of PCE and TCE.

2.2 Vinyl chloride-degrading bacteria

2.2.1 Anaerobic degradation of VC

In anaerobic processes, VC can be reduced to ethene, as shown in Figure 2. However, the dechlorination process from VC to ethene can be difficult to proceed because in some cases VC is only cometabolized (not linked to growth)[6], which could cause accumulation of VC in groundwater. In many cases, VC and ethene coexist at contaminated sites.

2.2.2 Aerobic degradation of VC

VC can be degraded through the same aerobic pathway (Figure 2.2)[8] as ethene[9, 10], due to their structural similarity. The initial attack on VC by an alkene monooxygenase (AkMO, encoded by *etnABCD* genes) adds oxygen to the double bond, forming chlorooxirane. The epoxide can be conjugated to coenzyme M (CoM) by epoxyalkane: coenzyme M transferase (EaCoMT, encoded by *etnE* gene), and the product could enter the central metabolic pathways[9-12].

Some bacteria that use ethene as the main carbon source (etheneotrophs) can only cometabolize VC to chlorooxiranes[13, 14], which are termed as VC cometabolizers. Other than etheneotrophs, bacteria growing on methane[15], toluene[16], etc. can also fortuitously degrade VC via non-specific oxygenases.

When ethene is not present, some bacteria can grow on VC as their sole carbon source, which are termed as VC-assimilators. As they both have *etnABCD* and *etnE* genes, until now these two groups cannot be genetically distinguished.

Currently isolated ethene- and/or VC-assimilating bacteria include *Nocardioides* sp. JS614, *Mycobacterium* sp. strain JS60, JS61, JS615-621, TM1, TM2[10], JS622-625[11]; *Pseudomonas putida* strain AJ and *Ochrobactrum* sp. strain TD[17]; *Mycobacterium chubuense* strain NBB4 [18]and *Haliea* sp. ETY-M[19]. In particular, there is an *etnE* allele (named *etnE1*) found in JS614. This *etnE1* allele has a 7 bp-deletion compared to normal *etnE*, which may cause the gene to lose its function. However, JS614 demonstrates higher VC degradation rate compared to other strains, which also might be a result of this duplicated *etnE* [9].

There is evidence that the genes that participate in aerobic VC degradation,

etnABCD and *etnE* in *Nocardioides* sp. strain JS614 are located on linear plasmids (~308 kb) [9]. Linear plasmids have been found in other strains that have showed VC or ETH degradation ability via pulsed-field gel electrophoresis (PFGE), including ethene/propane degrader *Xanthobacter* Py2(320 kb) [20], *Rhodococcus*(~210 kb) [21-23], *Mycobacterium*(110 to 330 kb)[11], *Pseudomonas* (~260 kb) and *Ochrobactrum* (three plasmids, ranging from 90 kb to 320 kb)[24]. Therefore, it is highly possible that the functional genes related to VC degradation in these organisms are also on plasmids.

2.3 Microbial ecology of aerobic VC-assimilating communities

Before the research presented in this dissertation, there was no studies about the structure of VC-assimilating communities, except one about enrichment culture kinetics [25]. There was only one research about possible interactions between etheneotrophs and methanotrophs present in the same culture[26], which showed that methanotrophs stimulated etheneotroph destruction of VC.

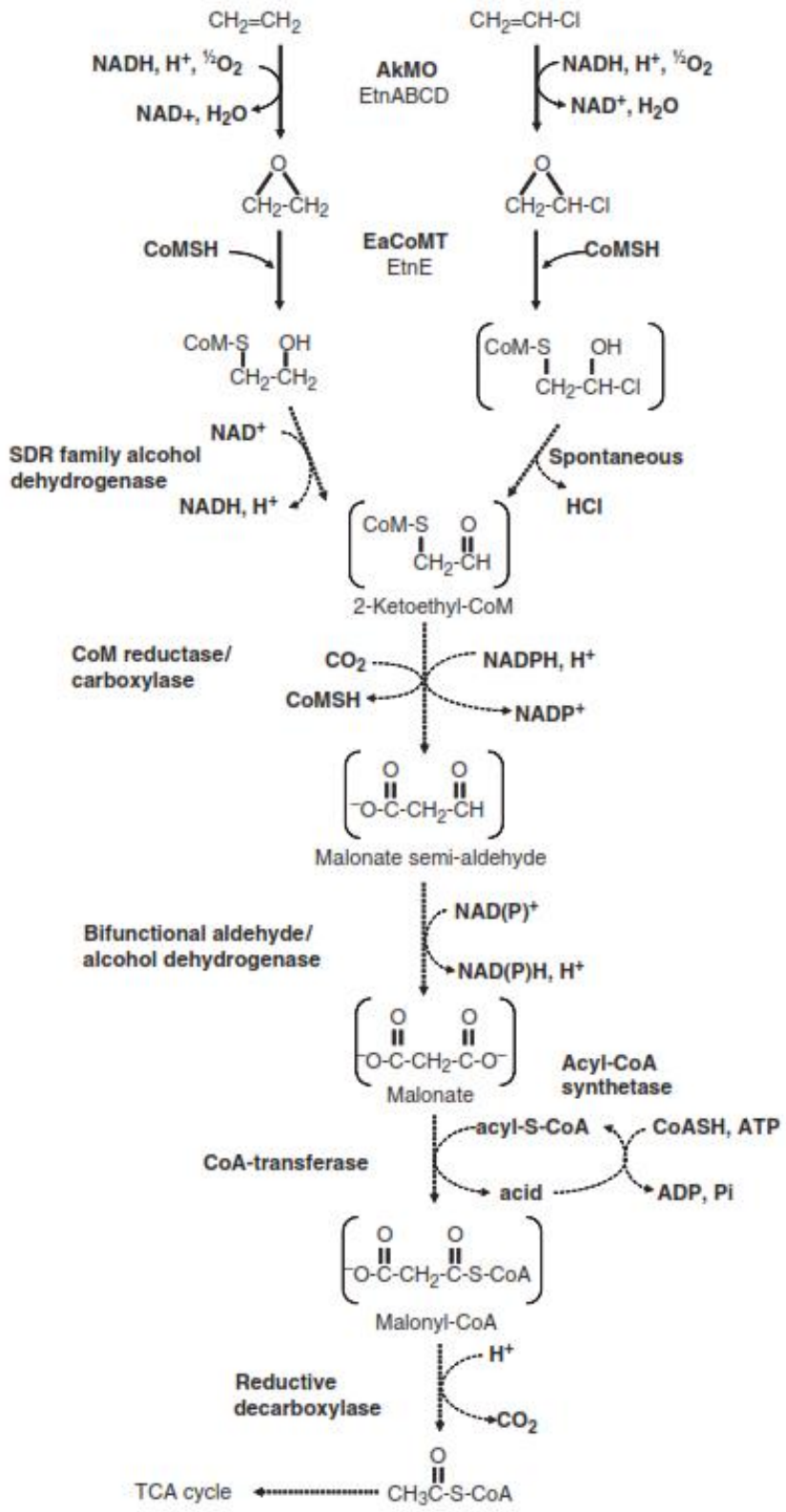


Figure 2.2 Proposed aerobic degradation pathway of VC and ETH in VC/ETH assimilators.

2.4 Possible adaptation and evolutionary mechanisms of VC degraders

2.4.1 Selective enrichment of a population (clonal amplification)

Microbial community can adapt to synthetic toxic compounds via selective enrichment of a certain population. One previous study on benzalkonium chlorides (BACs) degrading culture found that, the bacteria community adapted to BACs primarily through enrichment of a BAC-degrading P. *nitroreducens* (>50% in metagenomics analysis) in the bioreactor operated for three years[27]. There are other studies reporting microbial adaptation via selective enrichment, including but not limited to, BAC[28], tetrabromobisphenol A (TBBPA)[29] and phenol-containing sewage[30]. Therefore, it is possible that during the adaptation to VC, the groundwater bacteria community will be selected to be majorly composed of VC-assimilating bacteria, such as *Nocardioides* JS614.

2.4.2 Missense mutation and larger scale rearrangement

Missense mutation and larger scale rearrangement of genes can change correspondent protein products (enzymes) or related regulatory genes. Missense mutations have been observed in previous study when adapting ethene-assimilating *Mycobacterium* strain JS623 to VC as the sole carbon source[31]. In this case, two major mutations (W243G and R257L) of *etnE* gene developed after exposure to VC for about 86 days. The plasmids carrying these mutated *etnE* genes were electroporated into wild-type JS623 to yield recombinant strains, which displayed higher EaCoMT activity and decreased time to VC adaption compared to wild-type JS623.

2.4.3 Horizontal gene transfer (HGT)

There is evidence that genes related to VC degradation (*etnABCD* and *etnE*) are located on linear plasmids (See *Aerobic degradation of VC*), which is one type of mobile genetic elements (MGEs). Genes for the degradation of organic pollutants are often associated with MGEs such as plasmids and transposons [32]. Previous studies found evolutionarily related catabolic genes and gene clusters in various bacteria can originate from very distant locations [33]; the phylogeny of catabolic genes often does not match that of the 16S rRNA genes of the corresponding hosts[34], which are possibly due to HGT. Previous studies have found that EaCoMT genes (*etnE*) in VC-assimilating *Pseudomonas* and *Ochrobactrum* strains are very similar to those found in *Mycobacteria* [35], which supports this hypothesis.

2.4.4 Other possible mechanisms in VC-assimilating culture

There are several other mechanisms may facilitate the culture to adapt to VC as the main carbon source: duplication of genes (e.g. *etnE1* in JS614) and increased plasmid copy numbers could promote VC degradation activities[9, 36]; increased level of related chemicals in VC degradation pathway may also facilitate the adaptation, e.g., previous studies in our research group have shown increased CoM level contributed to degradation of VC[36]; there is evidence that epoxides are inducers of alkene monooxygenase[37], which means VC culture derived from ethene-assimilating culture may adapt more promptly.

2.5 Major molecular analysis tools for this project

2.5.1 Shotgun metagenomic sequencing

Shotgun metagenomic sequencing can comprehensively sample the entire genome of all the organisms in a complex culture, which is different from PCR-based approach, results in less mosaicism and biases [38]. Shotgun metagenomic sequencing has been applied in microbial evolution studies [27, 39] and investigation of complex microbial system, especially anaerobic systems, for example dechlorination environment[40] and activated sludge[41, 42]. The merit of shotgun metagenomic sequencing in studying adaption lies in its ability to look at both taxonomic and functional information. The general shotgun sequencing process includes generating DNA fragments of random length (~200-800 bp), adding adapter and barcodes, sequencing, genome assembly, genome binning and annotation (Figure 2.3). In this study we used the MiSeq Illumina sequencing system, which can generate up to 8.5 Gb of sequencing data of 2×250 bp sequencing length [43]. Illumina sequencing data can be processed using designed pipelines such as MetAMOS[44] and CLC Genomics Workbench (https://www.qiagenbioinformatics.com/) or a combination of different command-line tools, such as Trimmomatic (quality trimming and pre-processing)[45], IDBA (assembly)[46] and R (statistical analysis and graphing)[47].

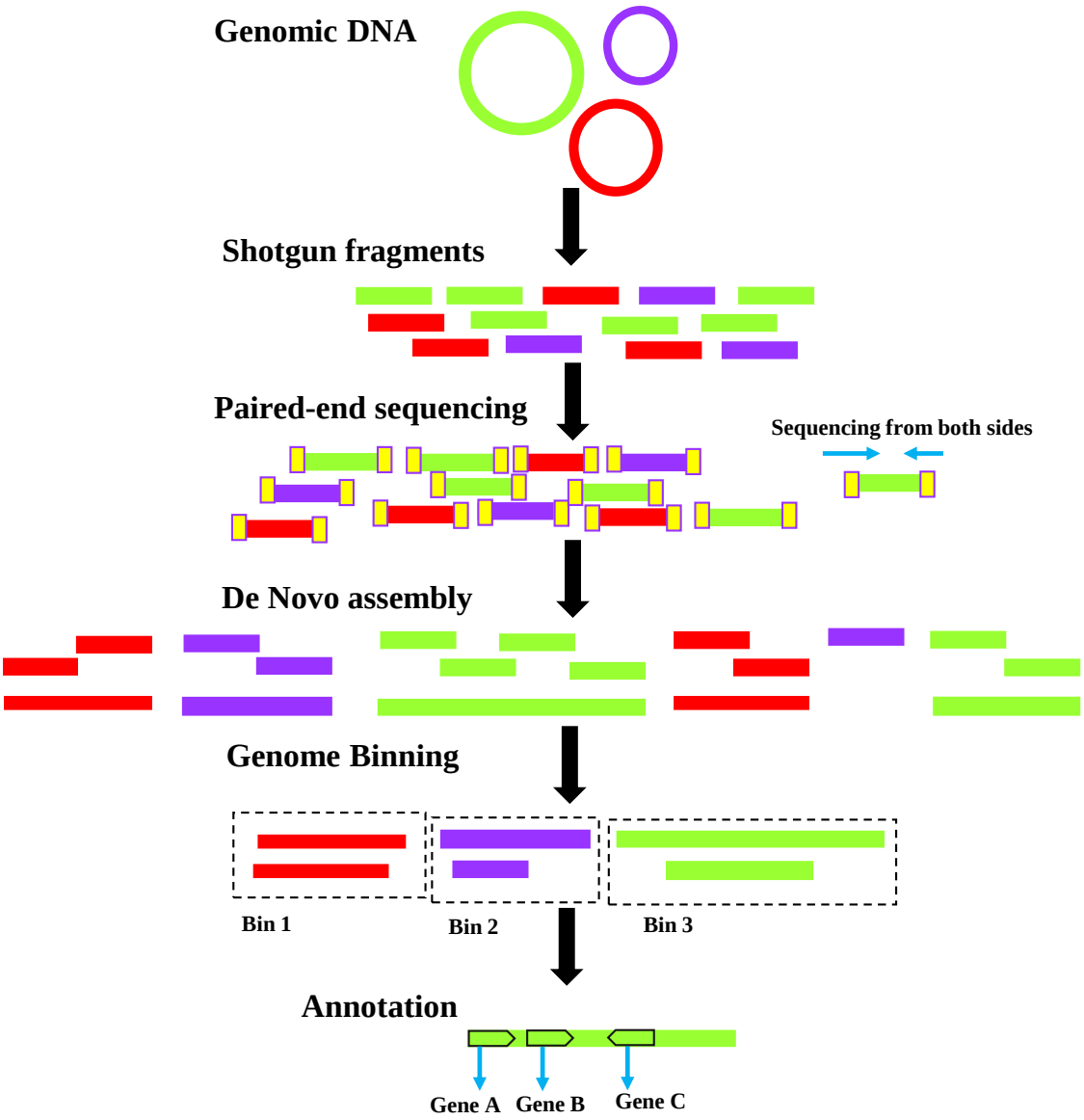


Figure 2.3 Shotgun metagenomic sequencing work flow.

2.5.2 DNA-stable isotope probing (DNA-SIP)

DNA-stable isotope probing (Figure 2.4) has been widely used in detecting bacteria responsible for growth-coupled degradation of contaminants (e.g. BTEX, 2,4-D)[48-54]. The SIP process usually includes culturing bacteria (incubation period) with ^{13}C or ^{15}N labelled molecules, during which bacteria performing growth-coupled degradation of labelled molecules would incorporate ^{13}C or ^{15}N into DNA, RNA and protein. For DNA-SIP, DNA will be extracted after incubation period and centrifuged in CsCl solution to form a gradient, yielding about 13-20 fractions of different buoyant density. The fractions are then subject to downstream molecular and genetic analysis. Detailed protocol on DNA-SIP has been published[55]. An unlabeled control group is often included to identify the label-enriched fractions. The incubation period is usually from 5~40 days [56-58].

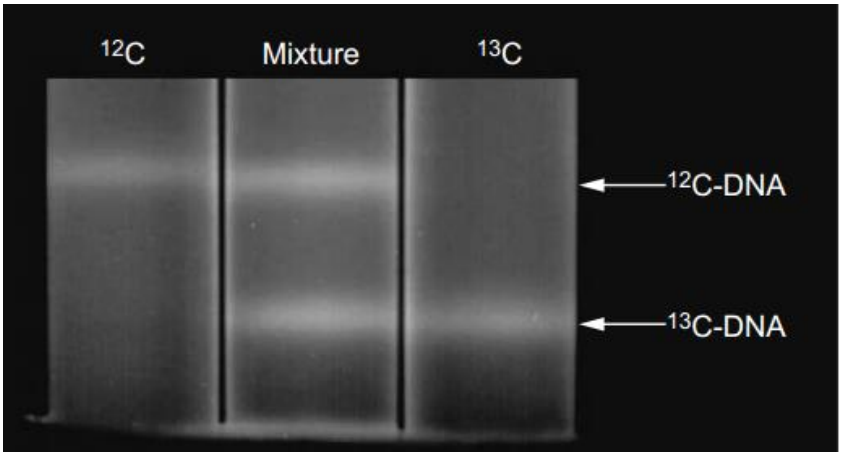


Figure 2.4 Example of SIP with ethidium bromide staining in CsCl gradient.

2.5.3 Quantitative polymerase chain reaction (qPCR)

Quantitative PCR [59] is a type of real-time PCR, which is based on the standard end-point PCR reactions, with the addition of fluorescent label to track the change of DNA abundance (Figure 2.5). To quantify the absolute abundance of target genes, a standard curve needs to be constructed using DNA templates with known gene copy numbers. The publication of qPCR data should follow the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines[60]. The United States Environmental Protection Agency has adopted qPCR as a monitoring method for water quality[61, 62].

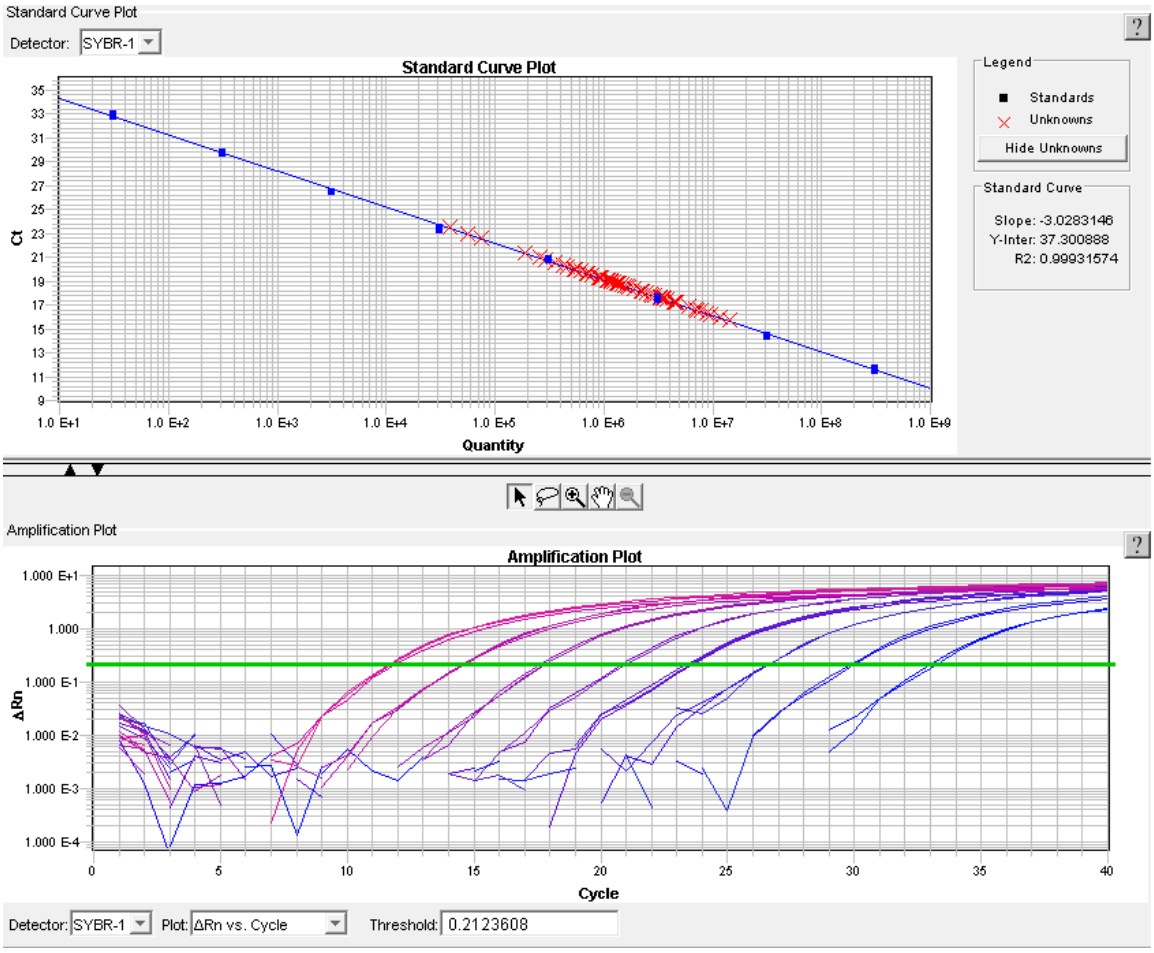


Figure 2.5 Example of qPCR standard curves.

2.5.4 Terminal restriction fragment length polymorphism (T-RFLP)

Terminal restriction fragment length polymorphism (T-RFLP)[63] is a culture-independent fingerprinting method for profiling the microbial community. In T-RFLP, the target gene is amplified with fluorescent-labeled PCR primers, usually at the 5' end of one of the primer, followed by restriction enzyme digestion. The digested product is mixed with a DNA size standard, and the fragments are then separated by capillary or gel electrophoresis using an automated sequencer. Only the terminal fluorescent-labeled restriction fragments (T-RFs) are detected. During the process, an electropherogram is generated, which shows the length of the T-RFs and the abundance (represented by the height and area of peaks). Since T-RFLP is based on end-point PCR product, it is semi-quantitative.

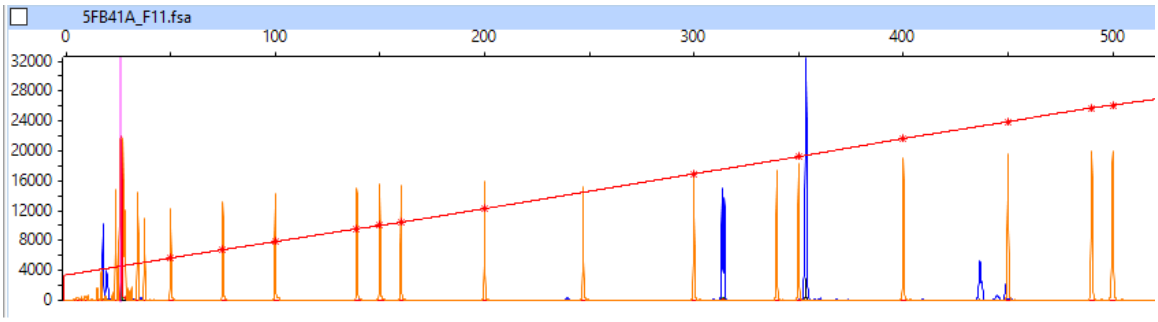


Figure 2.6 Example of T-RFLP electropherogram. The x-axis represents T-RF length, the y-axis represents peak height. Standards are orange-colored and T-RFs in the sample are blue-colored.

CHAPTER III. IDENTIFYING BACTERIA LINKED TO CARBON UPTAKE FROM VINYL CHLORIDE IN GROUNDWATER ENRICHMENT CULTURES USING STABLE ISOTOPE PROBING (SIP)

This is a collaborative work published [64]with Michigan State University. This work, together with another study [65], was focused on enrichment cultures from a contaminated groundwater plume at Carver, MA. Xikun Liu was in charge of qPCR and clone library on functional genes in these two studies. Other experiments were conducted by Fernanda Paes Wilson, who is also the first-author of these two papers. This work is presented here because the experimental methods and results are essential for the following research conducted by Xikun Liu. As the other study conveyed essentially the same methods and results, it is not presented in this chapter and only the abstract is included in the Appendix II.

3.1 Abstract

Vinyl chloride (VC) is a common groundwater contaminant and a known human carcinogen. Bioremediation is a potential cleanup strategy for contaminated sites, however, little is known about the microorganisms responsible for aerobic VC degradation in mixed microbial communities. In attempts to address this knowledge gap, the microorganisms able to assimilate ^{13}C from $^{13}C_2$-VC within a mixed culture capable of rapid VC degradation (120 μmol in seven days) were identified using stable isotope probing (SIP). At two time points during VC degradation (days 3 and 7), DNA was extracted from replicate cultures initially amended with labeled or unlabeled VC. The extracted DNA was fractioned via ultracentrifugation and the fractions of greater buoyant

density (heavy fractions) were subject to high throughput sequencing. Following this, specific primers were designed for the most abundant phylotypes in the heavy fractions. Then, qPCR was used across the buoyant density gradient to confirm label uptake by these phylotypes. From qPCR and/or sequencing data, five phylotypes were found to be dominant in the heavy fractions, including *Nocardioides, Sediminibacterium, Aquabacterium, Variovorax* and, to a lesser extent, *Pseudomonas*. The abundance of two functional genes (*etnC* and *etnE*) associated with VC degradation was also investigated in the SIP fractions. Peak shifts of *etnC* and *etnE* gene abundance towards heavier fractions were observed, indicating uptake of ^{13}C into the microorganisms harboring these genes. Analysis of the total microbial community indicated a significant dominance of *Nocardioides* over the other label-enriched phylotypes. Overall, the data indicate *Nocardioides* is primarily responsible for VC degradation in this mixed culture, with the other putative VC degraders generating a small growth benefit from VC degradation. The specific primers designed towards the putative VC degraders may be of use for investigating VC degradation potential at contaminated sites.

3.2 Introduction

Vinyl chloride (VC) is a common contaminant in surface water and groundwater, threatening both environmental and human health. This chemical is classified as a human carcinogen, with a Maximum Contamination Level (MCL) of 0.002 mg L^{-1} (ppb) and a MCL Goal of zero mg L^{-1}. VC in groundwater originates primarily from the transformation of tetrachloroethene (PCE) and trichloroethene (TCE). These chlorinated ethenes are common contaminants due to their widespread use and previous careless

disposal. In groundwater, natural biotic and abiotic processes degrade these solvents, leading to the accumulation of the metabolites, *cis*-dichloroethene and VC [5, 66]. Notably, VC is the most carcinogenic of the chlorinated ethenes and has the lowest regulatory limit in drinking water.

The use of microorganisms in groundwater bioremediation occurs through monitored natural attenuation, biostimulation or bioaugmentation. The biodegradation of VC can occur by either anaerobic or aerobic pathways and by co-metabolism or direct VC assimilation [8]. Anaerobic transformation of VC by *Dehalococcoides spp.* can occur either by co-metabolism or through energy generating VC reduction [66, 67]. For the aerobic metabolism of VC, several bacteria belonging to the phyla *Actinobacteria* and *Proteobacteria* have been isolated [10, 24, 31, 68-75]. While these microorganisms have greatly contributed to our understanding of VC metabolism, limitations associated with culture-based methods have likely resulted in an incomplete understanding of VC degrading microorganisms in mixed microbial communities or in groundwater at contaminated sites.

To address the limitations associated with culture-based methods, various molecular tools have been adopted. For instance, stable isotope probing (SIP), has the advantage of linking function (e.g. carbon uptake) to microorganism identity in mixed microbial communities [76]. This method was first introduced for investigating methanol-utilizing microorganisms in soil [56], but since then, has been used to identify microorganisms able to metabolize a variety of carbon and nitrogen sources. SIP involves sample exposure to a labeled compound (in this work, $^{13}C_2$-VC) and DNA extraction over time. The DNA is then subjected to ultracentrifugation, fractionation (to separate label

incorporated DNA from the unlabeled DNA) and DNA sequencing. The phylotypes that are present in the heaviest fractions are considered responsible for incorporating the labeled carbon from the target chemical.

Another molecular tool important for understanding the potential for contaminant biodegradation in mixed communities and in groundwater is quantitative PCR (qPCR). In the case of VC biodegradation, qPCR assays have been developed and applied to target *etnC* and *etnE* [77, 78]. The gene *etnC* encodes the alkene monooxygenase (*AkMO*) alpha subunit, which is responsible for the initial attack on VC, to convert it to VC epoxide. The gene *etnE* encodes the epoxyalkane coenzyme M transferase (*EaCoMT*), which conjugates the VC epoxide to CoM and introduces the metabolite to the central metabolic pathway. However, qPCR on functional genes alone cannot link activity to microorganism identity.

In a previous study, we investigated the microorganisms able to uptake carbon from VC in a mixed culture capable of slower VC degradation (~120 µmol in 45 days) using SIP and also examined *etnE* abundance in the heavy SIP fractions [65](Appendix II, first study). In the present study, the overall objective was to identify microorganisms associated with carbon uptake from VC in an enrichment culture (derived from different groundwater than the previous study) capable of more rapid VC degradation (~120 µmol in 7 days). Further, unlike previous research, qPCR primers were developed towards the identified enriched phylotypes, to provide more robust evidence of carbon uptake. In addition, the presence of two functional genes associated with VC degradation (*etnE* and *etnC*) in the heavy SIP fractions was investigated. Further, changes in the total microbial community during VC degradation were examined. The research identifies several novel

phylotypes as being responsible for carbon uptake from VC.

3.3 Materials and Methods

3.3.1 Experimental design for SIP

The VC enrichment culture was derived from groundwater from well 63-I (collected September 2009) at a VC contaminated site in Carver, MA. At the time of sampling, the groundwater was slightly anoxic (DO 0.61 mg L^{-1}). The culture was developed using a previously described approach[79], involving transferring 1 mL of well 63-I groundwater to 100 mL of mineral salts media (MSM)[68] in 160 mL serum bottles with air headspace and maintaining O_2 by injecting pure O_2 as needed after the onset of biodegradation. The culture was repeatedly fed VC for one year and then subcultured at 1:100 once, followed by subculturing at 1:1000 three times.

Microcosms for SIP experiment were subculture from VC mixed culture described above. These microcosms were prepared as previously described [79] in the 160 mL serum bottles and the final VC concentration is about 47 mg L^{-1}. Two abiotic control microcosms (obtained via autoclaving) and six live microcosms were amended with unlabeled VC (99%, Specialty Gases of America, Toledo, OH, USA). In addition, six live microcosms were amended with labeled VC ($^{13}C_2$-VC, 99%, Cambridge Isotope Laboratories, Xenia, OH, USA). The microcosms were protected from light and incubated at room temperature (21-23 °C) on a shaker (200-300 rpm). VC concentrations were monitored for seven days and DNA was extracted from labeled VC-amended and unlabeled VC-amended microcosms at days 3 and 7.

3.3.2 Analytical methods

Headspace samples (100 µL) were analyzed via gas chromatography (Perkin Elmer, Waltham, MA, USA) with flame ionization detection and a capillary column (DB-624, diameter 0.53 mm; J&W Scientific, Santa Clara, CA). Peak areas were compared to an external standard for VC quantification. Aqueous phase VC concentrations were calculated using previously reported Henry's Law constant[1].

3.3.3 DNA extraction, ultracentrifugation and fractionation

The UltraClean Microbial DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA) was used for total nucleic acid extraction according to the manufacturer's recommended procedure. Quantified DNA extracts (~10 µg) were loaded into Quick-Seal polyallomer tubes (13 by 51 mm, 5.1 mL; Beckman Coulter, Indianapolis, IN) along with a Tris-EDTA (pH 8.0)-CsCl solution for ultracentrifugation. Prior to sealing (cordless Quick-Seal tube topper; Beckman), the buoyant density (BD) was determined with a model AR200 digital refractometer (Leica Microsystems Inc., Buffalo Grove, IL, USA) and adjusted to a final BD of 1.73 g mL^{-1} by the addition of small volumes of CsCl solution or Tris-EDTA buffer. Samples were fractioned via ultracentrifugation at 178,000 × g (20 °C) for 48 h in a StepSaver 70 V6 vertical titanium rotor (8 by 5.1 mL capacity) within a Sorvall WX 80 Ultra Series centrifuge (Thermo Scientific, Waltham, MA, USA). Following ultracentrifugation, the tubes were placed onto a fraction recovery system (Beckman), and fractions (~20, 150 µL each) were collected. The BD of each fraction was measured, and CsCl was removed by glycogen-

assisted ethanol precipitation. The quantity of DNA in each fraction was determined by Qubit™ (Invitrogen, Carlsbad, CA, USA) to ascertain which fractions were enriched in labeled carbon.

3.3.4 MiSeq Illumina sequencing and SIP fraction analysis

The four heaviest fractions from each time point (days 3 and 7) from the labeled VC-amended microcosms were analyzed to determine which microorganisms were enriched in these fractions and were therefore responsible for carbon uptake from VC. The fractions were selected based on the DNA concentration in each fraction from the labeled VC amended microcosms compared to the unlabeled VC amended microcosms (see below). In all, 48 samples were subjected to high throughput sequencing (MiSeq Illumina Sequencing) at the Research Technology Support Facility at Michigan State University (RTSF). These samples included six replicates per fraction, four fractions per time point, and two time points.

PCR and Illumina sequencing were performed at RSTF using a previously described protocol [80]. This involved the amplification of the V4 region of the 16S rRNA gene using a set of multiplex indexed primers. Following amplification, individual reactions were quantified (Picogreen assay), a pool of equimolar amounts of each was made, and these were purified using Ampure XP beads. A final gel purification step was included to ensure non-specific products were eliminated. The combined library was loaded onto the Illumina MiSeq Platform using a standard MiSeq paired end (2x250 bp) flow cell and reagent cartridge.

The Mothur software [81] was used to analyze the data generated by MiSeq Illumina

using a SOP developed by Schloss (http://www.mothur.org/wiki/MiSeq_SOP). This involved the construction of contigs, error and chimera removal followed by sequence alignment for Operational Taxonomic Units (OTU) assignment based on the SILVA database [82]. Final data matrices were exported to Excel 2013 SR-1 (Microsoft Corporation, Redmond, WA). To identify which OTUs were responsible for the label uptake in each fraction at each time point, the relative abundance (%) of each OTU was calculated.

3.3.5 Primer design, 16S rRNA gene clone libraries and qPCR

Specific primers (*sedF* and *sedR*; *aquaF* and *aquaR*; and *varF* and *varR*) were designed to target the phylotypes responsible for the VC assimilation (identified above) using Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). The new primers were tested for specificity using 16S rRNA gene clone libraries and Sanger sequencing. Following this, the primers were used in the qPCR assays across the BD gradient to confirm label uptake.

Three 16S rRNA clone libraries were generated using total DNA extracted from the labeled microcosms from day 3. DNA was PCR-amplified using 3 primer sets: *sedF* (5'- CGG GCA GTT AAG TCA GTG GT-3') and *sedR* (5'- TGC CTT CGC AAT AGG TGT TCT-3'); *aquaF* (5'- CGT AGG GTG CGA GCG TTA AT-3') and *aquaR* (5' - CCA TCC CCC TCT ACC GTA CT-3'); and *varF* (5'-TCT GTG ACT GCA TTG CTG GA-3') and *varR* (5'-CGG TGT TCC TCC GCA TAT CT-3') (IDT, Integrated DNA Technologies, Coralville, IA, USA). The PCR program consisted of an initial denaturation (95°C, 5 min), 29 cycles of amplification (95°C, 30 s; 58°C, 30 s; 72°C 1:30

min), and a terminal extension step (72°C, 30 min). Agarose gel electrophoresis and the subsequent staining of the gels with ethidium bromide confirmed the presence of PCR products. The PCR products were purified with QIAquick PCR purification kit (Qiagen Inc., Alameda, CA, USA) and cloned into *Escherichia coli* TOP10 vector supplied with a TOPO TA cloning kit (Invitrogen). Clones of *E. coli* were grown on Luria-Bertani (LB) medium solidified with 15 g agar L^{-1} with 50 μg ampicillin L^{-1} for 16 h at 37 ºC. Colonies with inserts were verified by PCR with primers M13F (5'-TGT AAA ACG ACG GCC AGT-3') and M13R (5'-AAC AGC TAT GA CAT G-3'), plasmids were extracted from the positive clones with a QIAprep miniprep system (Qiagen, Inc.) and the insertions were sequenced by RTSF. The clone sequences for each library were aligned against the Illumina sequences using MEGA 6 to determine the specificity of each primer set.

Amplification and qPCR measurements were conducted in a Chromo 4 real-time PCR cycler (Bio-Rad, Philadelphia, PA, USA) using QuantiTect SYBR Green PCR Kit (Qiagen Inc.) and the primer sets *sedF-sedR*, *aquaF-aquaR* and *varF-varR*. DNA from both time points from the density gradient fractions from both the labeled VC and unlabeled VC amended microcosms were subject to qPCR in triplicate. Each 25 μL PCR mixture contained 12.5 μL QuantiTect SYBR Green PCR Master Mix solution, 1.25 μL each 10 μM primer, 9 μL DNA-free water and 1 μL DNA template. The thermal protocol consisted of an initial denaturation (95°C, 15 min), 40 cycles of amplification (95°C, 15 s; 58°C, 20 s; 72°C 20 s), and a terminal extension step (72°C, 2 min). Melting curves were constructed from 55°C to 95°C and read every 0.6°C for 2 s. Cloned plasmid DNA was used as a standard for quantification, and the numbers of gene copies were determined as previously described [83] (plasmid size was 3,931 bp, in addition to inserts of

148 bp by *sedF-sedR*, 139 bp by *aquaF-aquaR* and 86 bp by *varF-varR*). Detailed parameters for qPCR is in Appendix I Table AI.1) as required by MIQE.[60]

3.3.6 Functional gene (etnC and etnE) qPCR

Amplification and qPCR measurements were conducted on an ABI 7000 Sequence Detection System (Applied Biosystems, Foster City, CA, USA) using Power SYBR Green PCR Master Mix (Applied Biosystems). Primer sets for amplifying *etnC* genes were RTC_F (5'-ACCCTGGTCGGTGTKSTYTC-3') and RTC_R (5'-TCATGTAMGAGCCGACGAAGTC-3') and for *etnE* genes were RTE_F (5'-CAGAAYGGCTGYGACATYATCCA-3') and RTE_R (5'-CSGGYGTRCCCGAGTAGTTWCC-3')[77]. DNA from both time points from the density gradient fractions from both the labeled VC and unlabeled VC amended microcosms were subject to qPCR in duplicates. Each 25 µL reaction mixture contained 12.5 µL of Power SYBR Green PCR Master Mix solution, 750 nM of each qPCR primers and 1 µL of DNA template. A previously developed thermal protocol was used[77].

Serial dilutions of PCR products were used to generate the standard curves in the qPCR assays and involved 1138 bp *etnC* and 890 bp *etnE* from *Nocardioides* sp. strain JS614. Primer sets to produce the qPCR standards for *etnC* were JS614 EtnCF (5'-GCGATGGAGAATGAGAAGGA-3') and JS614 EtnCR (5'-TCCAGTCACAACCCTCACTG-3') [77] and CoMF1L (5'-AACTACCCSAAYCCSCGCTGGTACGAC-3') and CoMR2E (5'-GTCGGCAGTTTCGGTGATCGTGCTCTTGAC-3') for *etnE* [11]. The 25 µL reactions contained 12.5 µL Qiagen PCR Master Mix, 0.2 µM of each primer and 2 ng of JS614

genomic DNA. Genes per μL of PCR product were estimated as described previously[77]. ABI 7000 System SDS software (Applied Biosystems) was used to analyze real-time PCR fluorescence data using the auto baseline function.

3.3.7 Microbial community analysis during VC degradation

All extracted DNA from the SIP study was fractioned, and therefore it was not possible to examine the relative abundance of the identified VC assimilators in the total microbial community. Thus, an additional experiment was performed to determine the relative abundance of these microorganisms during VC degradation.

For this, frozen culture aliquots (5 mL) were incubated for 30 days in a sterile serum bottle (160 mL) MSM (as before) and VC (~20 mg L^{-1}). Following VC degradation, 5 mL of fresh culture was added to sterile serum bottles (160 mL) containing MSM and VC (~47 mg L^{-1}). These microcosms were prepared as previously described[79], and included one abiotic control microcosm (obtained via autoclaving) and three live microcosms. The microcosms were protected from light (room temperature, 21-23 °C) and were placed on a shaker (200-300 rpm). VC concentrations were monitored for 18 days. DNA was extracted as described above from six time points and was analyzed using MiSEQ Illumina and Mothur (as described above). The percent relative abundances of the identified phylotypes were determined at each time point.

3.4 Results and Discussion

3.4.1 VC degradation and the identification of phylotypes responsible for label uptake

Direct aerobic VC oxidation was first reported in groundwater samples [84]. Since then, aerobic VC degradation and the isolation of VC degrading microorganisms has been

reported in many studies [10, 24, 31, 68, 69, 71-75]. These authors observed degradation over a wide range of time period, from as little as 2-20 days [10, 24, 69, 73, 75] to as much as 55-476 days [10, 31]. However, as previously discussed, limited research has identified the microorganisms responsible for VC degradation within mixed communities.

In the current research, VC was degraded by the mixed culture in the SIP study in only seven days. VC degradation occurred in both the labeled and unlabeled VC amended microcosms but not in the abiotic controls, confirming biological removal (Figure 3.1). The DNA samples extracted on days 3 and 7 were subject to ultracentrifugation. Following this, the DNA concentration in each fraction was determined (Figure 3.2). The presence DNA in the lighter BD fractions from the unlabeled VC amended samples and DNA in the heavier fractions from the labeled VC amended samples confirms the uptake of labeled carbon by the microorganisms in the mixed community.

The DNA in these heavier fractions (indicated with arrows in Figure 3.2) were submitted for sequencing. After Mothur analysis, the final numbers of sequences obtained were 2,010,477 and 2,073,705 for days 3 and 7, respectively. The chimeric percentage was 1.32% and 1.93% on days 3 and 7, respectively. Among the final sequences, < 2% were unique at both time points.

Interestingly, at both time points, five genera were dominant in the heavy fractions, including *Nocardioides (Actinobacteria)*, *Sediminibacterium (Bacteroidetes)*, *Aquabacterium (Proteobacteria)*, *Variovorax (Proteobacteria)* and, to a lesser extent, *Pseudomonas* (*Proteobacteria*) (Figure 3.3). The level of enrichment was similar in all four heavy fractions at both time points. The data indicates these microorganisms are responsible for label uptake from VC. In particular, *Nocardioides, Sediminibacterium* and

Aquabacterium were the three most abundant at both time points. It is uncertain if the enriched phylotypes were responsible for the initial transformation of this chemical, or if they were assimilating VC metabolites.

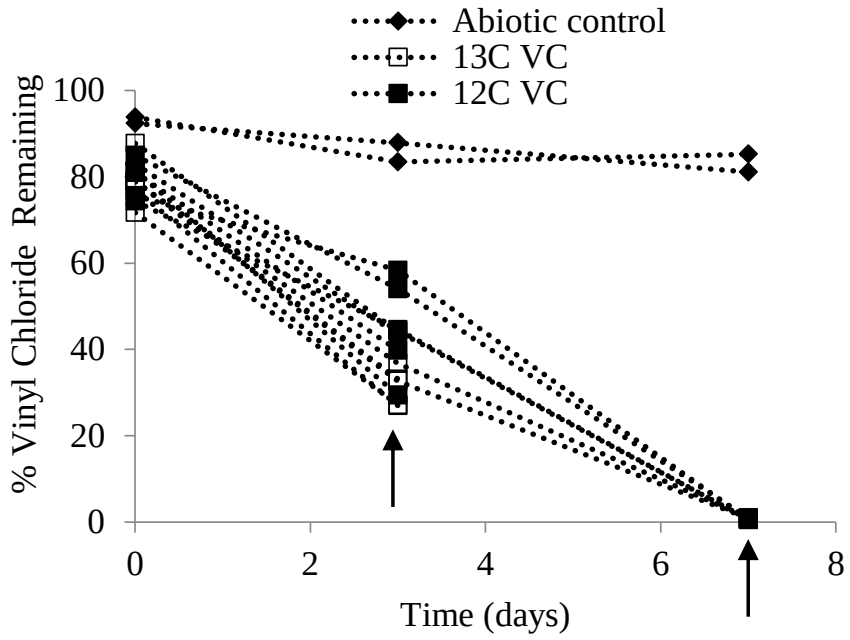


Figure 3.1 Percent VC remaining in cultures amended with labeled vinyl chloride (empty square), unlabeled vinyl chloride (solid square) and in the abiotic controls (solid diamond). Arrows indicate when DNA was extracted. (By Fernanda Paes Wilson)

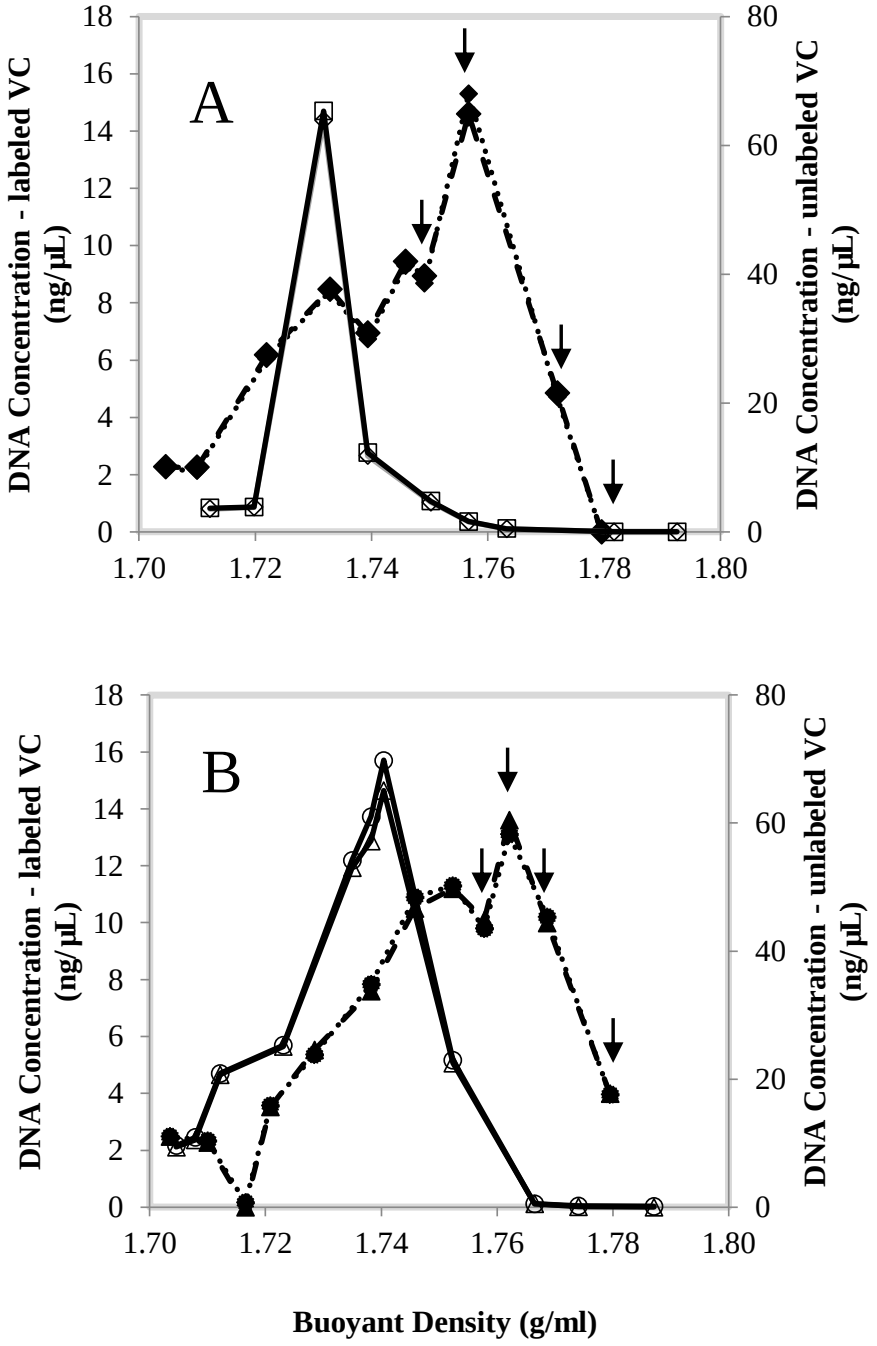


Figure 3.2 DNA concentration (ng/µL) at day 3 (A) and day 7 (B) in fractions obtained from the labeled VC and unlabeled VC amended cultures. The complete and dashed lines represent DNA concentrations from the unlabeled and labeled VC amended cultures, respectively. Arrows indicate samples selected for sequencing. (By Fernanda Paes Wilson)

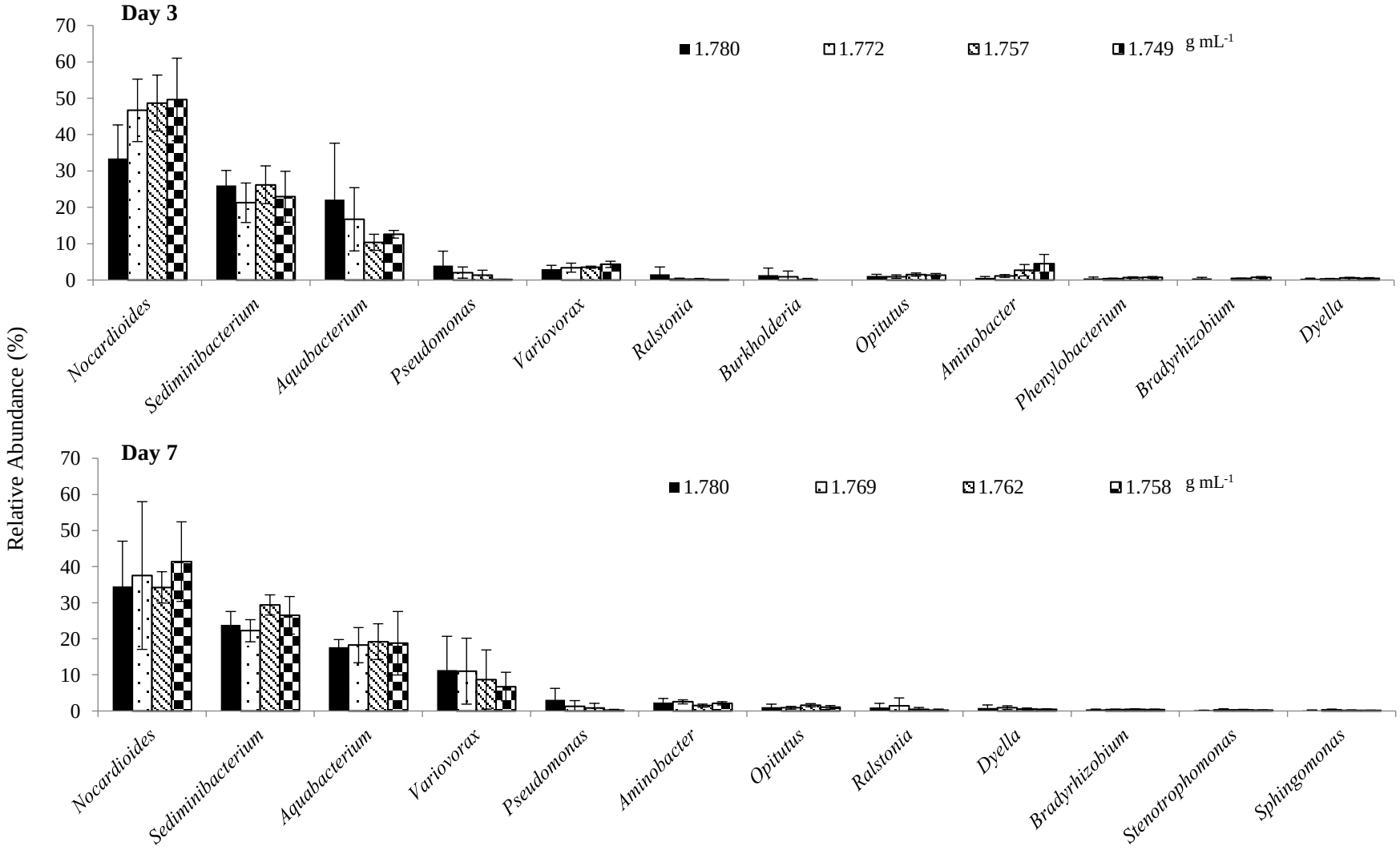


Figure 3.3 Relative abundance of phylotypes in the four heavy fractions from labeled VC amended cultures. The error bars represent standard deviations from six samples submitted for sequencing. (By Fernanda Paes Wilson)

3.4.2 Primer design, specificity test and 16S rRNA abundance in SIP fractions

To provide further evidence of label uptake, specific primers were designed towards three of the five dominant genera in the heavy fractions: *Sediminibacterium*, *Aquabacterium* and *Variovorax. Nocardioides* was not targeted as this genus has already been linked to VC metabolism. The primers were utilized to compare gene abundance across the BD profile between the samples amended with labeled VC and those amended with unlabeled VC.

First, the primer specificity was tested by amplifying DNA, creating 16S rRNA gene clone libraries, and sequencing these clones. For this, 16 clones were sequenced for each primer set. The primer sets illustrated a satisfactory level of specificity for their targets. For *Sediminibacterium*, 14 of 16 clones were 100% identical with the *Sediminibacterium* sequence, and 2 of 16 aligned at a 99% identity level. For *Aquabacterium*, 13 of 16 aligned at 100%, two of 16 aligned at 99% and 1 of 16 aligned at 98% with the *Aquabacterium* sequence. For *Variovorax*, 8 of 13 aligned at 100% and 5 of 13 aligned at 98.8% with the *Variovorax* sequence.

Following the specificity testing, qPCR assays for *Sediminibacterium*, *Aquabacterium* and *Variovorax* were performed on all of the fractions across the BD gradient at both time points. As expected, maximum gene copy values were at higher BD for the labeled VC amended samples compared to the unlabeled VC amended samples, again indicating label uptake by these phylotypes (Figure 3.4).

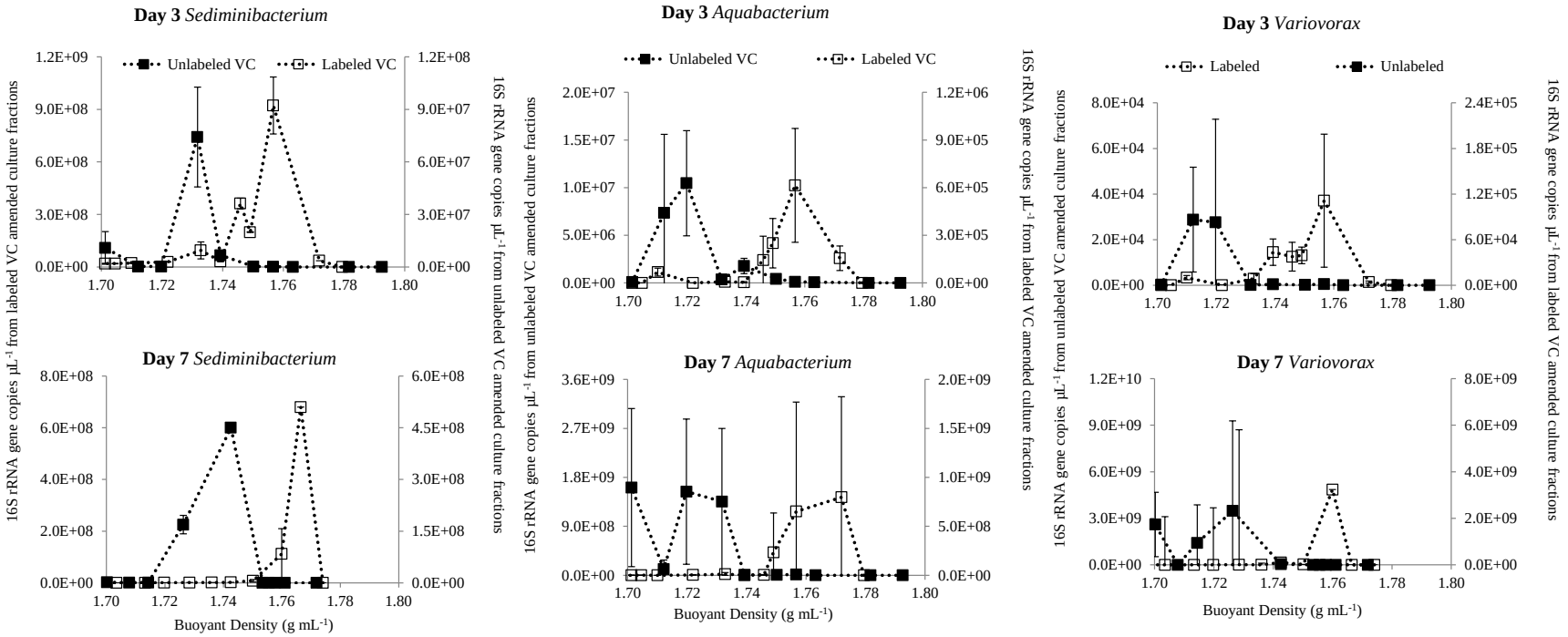


Figure 3.4 *Sediminibacterium*, *Aquabacterium* and *Variovorax* 16S rRNA gene copies over the buoyant density range in DNA extracted from labeled and unlabeled VC amended cultures at days 3 and 7. Error bars represent standard deviation from three qPCR measurements. (By Fernanda Paes Wilson)

Quantitative PCR has emerged as the method of choice for enumerating targeted genes. *Dehalococcoides* 16S rRNA primers have been tested and used for site assessment and bioremediation implementation for chlorinated pollutants under anaerobic conditions [85-87]. In the present study, the 16S rRNA primers were designed for the phylotypes associated with carbon uptake from VC (*Sediminibacterium*, *Aquabacterium* and *Variovorax)*. These were used to provide additional evidence these microorganisms were involved in label uptake. The shifts observed in the 16S rRNA gene abundance peaks toward heavier fractions during the VC degradation process indicate the accumulation of ^{13}C in the 16S rRNA genes of *Sediminibacterium*, *Aquabacterium* and *Variovorax*. Similar shifts have been observed by other authors who combined SIP and qPCR for toluene-degrading cultures with the gene *bssA*[52].

To date, known aerobic VC-degrading microorganisms belong to the phyla *Actinobacteria* and *Proteobacteria*. Within these two phyla, several isolates have been associated with aerobic VC degradation, including those in the genera *Mycobacterium* [10, 68, 69, 71, 72, 74, 88], *Pseudomonas* [11, 75], *Nocardioides*[10], *Ochrobactrum*[24], and *Ralstonia* [73].

In agreement with previous research[10], the current study indicates that the genus *Nocardioides* is associated with VC degradation. Interestingly, both *Pseudomonas* and *Ralstonia* were enriched to a lesser extent in the heavy fractions for the labeled VC amended samples (Figure 3.3) and both have previously been linked to aerobic VC removal [24, 73, 75]. Isolates within the genus *Sediminibacterium* have not previously been linked to aerobic VC degradation, although this genus has been found at groundwater sites contaminated with chlorinated aliphatic hydrocarbons [88]. The genera *Aquabacterium*

and *Variovorax* have not previously been associated with VC degradation. However, *Aquabacterium* has been observed at sites contaminated with chlorinated aliphatic hydrocarbons [88].

To our knowledge, only one other report exists on identifying the microorganisms able to assimilate carbon from vinyl chloride in a mixed microbial community [65]. In our previous study, *Nocardioides* was also a dominant VC assimilator and with other phylotypes (*Brevundimonas, Tissierella* and *Rhodoferax*) being associated with minor levels of label uptake. Taken collectively, *Nocardioides* is clearly an important phylotype for aerobic VC degradation in mixed communities. As both cultures were derived from contaminated site groundwater, it is possible that this genus is also important for *in situ* VC degradation.

3.4.3 Functional gene (etnC and etnE) abundance in SIP fractions

Quantitative PCR assays for the functional genes *etnC* and *etnE* were also performed for BD fractions from both the labeled and unlabeled VC-amended microcosms. Overall, the highest unlabeled peaks were at lighter BD fractions (BD = 1.720 - 1.760 g mL^{-1}) and the highest labeled peaks were at heavier BD fractions (BD = 1.740 - 1.780 g mL^{-1}) (Figure 3.5). The genes *etnC* and *etnE* are involved in VC metabolism and are therefore useful biomarkers for VC degradation potential. Peak shifts of *etnC* and *etnE* gene abundance towards heavier fractions were observed indicating uptake of ^{13}C into the microorganisms harboring these genes. The data generated here are in agreement with previous research, which has indicated the importance of *etnC* and *etnE* in *Nocardioides* sp. JS614 [9, 65].

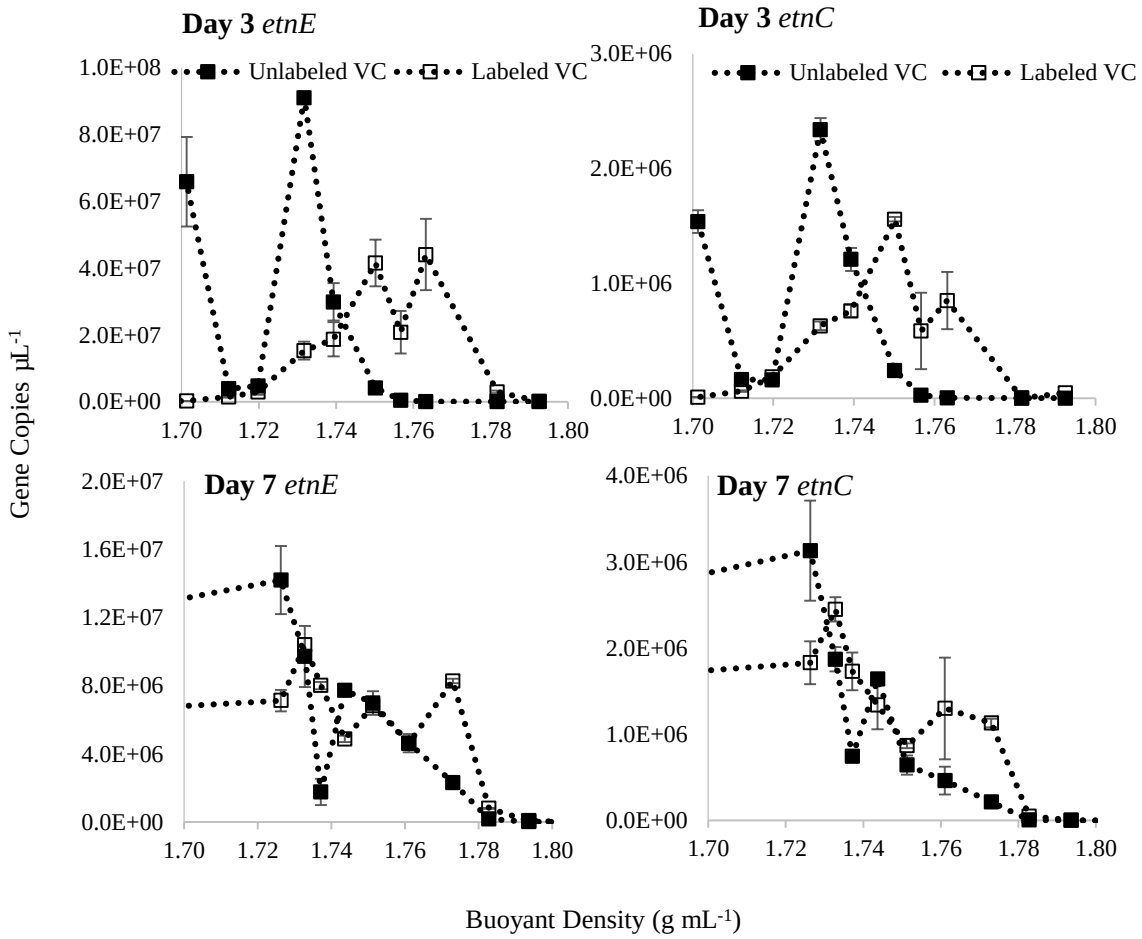


Figure 3.5 The *etnE* and *etnC* gene copies over the buoyant density range after fractionation from labeled and unlabeled VC amended cultures at days 3 and 7. The data points represent the average of duplicates and the error bars depict the range detected in qPCR. (By Xikun Liu)

3.4.4 Microbial community changes during VC degradation

As expected, in the experiment to examine microbial community changes over time, degradation of VC occurred in the live cultures but not in the abiotic control (Figure 3.6). DNA extracted during VC removal (days 2, 9, 10, 13, 14 and 16) was examined with high throughput sequencing to determine the relative abundance of the phylotypes identified as being enriched in the heavy fractions (Figures 3.3 and Figure 3.7). The data indicate that the relative abundance of *Nocardioides* was high initially and remained high during VC degradation (18-33 % relative abundance). Surprisingly, *Sediminibacterium, Aquabacterium* and *Pseudomonas* were present at less than 1% relative abundance and *Variovorax* was <0.0017% in these cultures. However, the abundance of the three phylotypes did increase with time, suggesting they experienced a growth benefit from VC degradation.

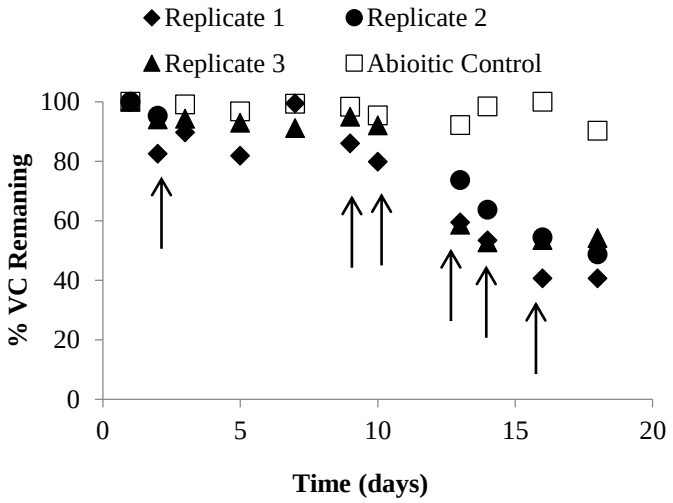


Figure 3.6 Percent vinyl chloride remaining in triplicate live cultures and an abiotic control. The arrows indicate when DNA was extracted from these samples for 16S rRNA gene amplicon Illumina sequencing. (By Fernanda Paes Wilson)

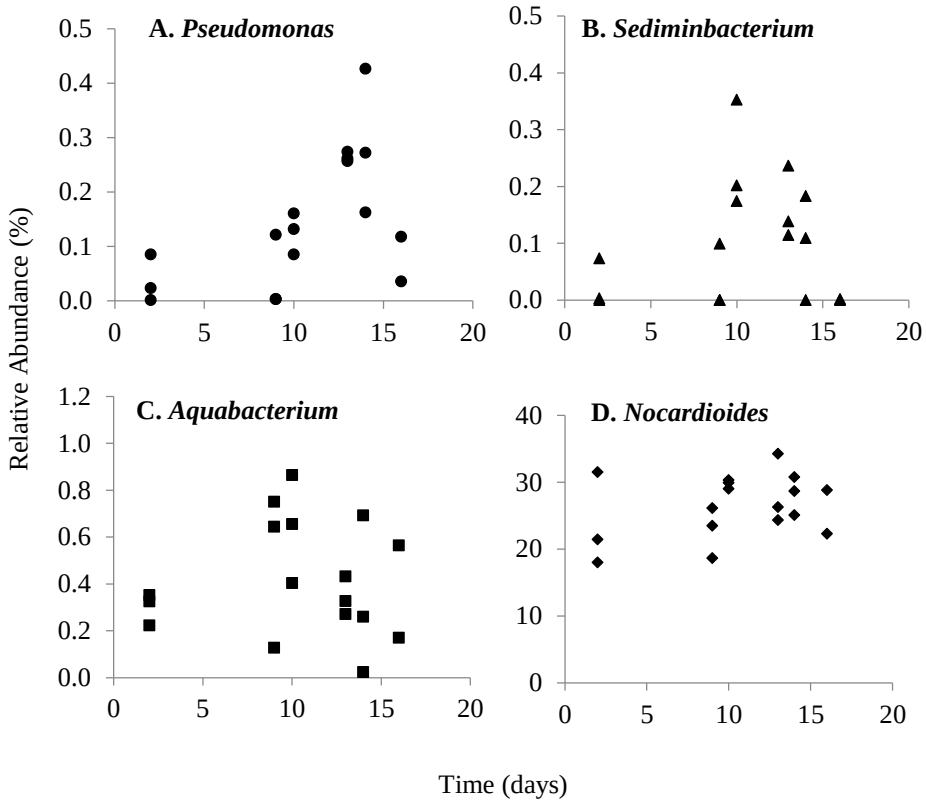


Figure 3.7 The relative abundance of *Pseudomonas* (A), *Sediminibacterium* (B), *Aquabacterium* (C) and *Nocardioides* (D) during vinyl chloride degradation. (By Fernanda Paes Wilson)

Note: the y-axis has different scales. The relative abundance of Variovorax was <0.0017% in these cultures.

3.5 Conclusions

The microorganisms responsible for assimilating ^{13}C from VC during aerobic VC degradation included *Nocardioides* (*Actinobacteria*), *Sediminibacterium* (*Proteobacteria*), *Aquabacterium* (*Proteobacteria*), *Variovorax* (*Proteobacteria*) and, to a lesser extent, *Pseudomonas* (*Proteobacteria*). Therefore, both previously identified VC assimilators (*Nocardioides, Pseudomonas*) and novel phylotypes (*Sediminibacterium, Aquabacterium, Variovorax*) were responsible for label uptake from VC. Additional research is needed to determine if these microorganisms are using carbon directly from VC or from VC metabolites. The overall microbial community analysis of this culture indicated *Nocardioides* was dominant, although, the relative abundance of the other phylotypes increased as VC was degraded, suggesting a growth benefit from VC degradation. SIP on two functional genes (*etnE* and *etnC*) indicated the microorganisms harboring these genes were involved in label uptake. The specific primers designed towards the putative VC degraders may be of use for investigating VC degradation potential at contaminated groundwater sites.

CHAPTER IV. CARBON UPTAKE OF VINYL CHLORIDE (VC)-ASSIMILATORS AND COMETABOLIZERS FROM VC AS REVEALED BY STABLE ISOTOPE PROBING (SIP) ON MIXED PURE CULTURES

This is a collaborative study with Michigan State University (MSU). Fernanda Paes from MSU was in charge of SIP fraction collection and purification. Xikun Liu was in charge of all the other works presented in this study.

4.1 Abstract

Nocardioides sp. strain JS614 is a known vinyl chloride (VC)-assimilating bacterium whereas *Mycobacterium* strain JS622 is a VC-cometabolizing bacterium. With an original goal to distinguish VC-assimilating from cometabolizing bacteria, pure *Nocardioides* sp. strain JS614 and *Mycobacterium* stains JS622 were mixed in 1:1 ratio and amended with VC. Stable isotope probing (SIP), terminal restriction fragment length polymorphism (T-RFLP) and qPCR were applied on the mixed culture to study the assimilation of VC and their growth within the VC enrichment culture. To minimize cross-feeding, the SIP experiment only lasted for about two days. The OD600 of mixed cultures showed a decrease from 0.04 to 0.02 followed by a steady increase to about 0.05 during the experiment, while VC was continuously being degraded (about 70 μmol consumed at the last data point). The T-RFLP of 16S rRNA gene and functional gene *etnE* on SIP fractions suggested that both JS614 and JS622 incorporated ^{13}C into their genomic DNA, although to a small extent. The T-RFLP of 16S rRNA gene also showed the relative abundance of JS622 in the mixed culture was growing from about 10% to 40% while JS614 was decreasing from about 80% to 50% during the course of the experiment. Before the SIP experiment, the VC-assimilating and cometabolizing

activities were confirmed for these two strains. This study showed that VC cometabolizer *Mycobacterium* strain JS622 took up carbon from VC to sustain their growth when mixed with VC-assimilating *Nocardioides* sp. strain JS614. The mechanism of their interaction needs further investigation.

4.2 Introduction

Bacteria degrade toxic organic compounds through a variety of metabolic pathways that can be described as two main categories: assimilation and cometabolism. Bacteria obtain carbon and energy source from the organic compound to sustain their growth can be termed assimilation processes. These bacteria are referred to here as assimilators. Cometabolism processes occur when bacteria fortuitously degrade one compound while using another compound as the primary substrate to sustain their growth. These bacteria are referred to here as cometabolizers.

In bioremediation practice, both assimilation and cometabolism can contribute to the degradation of toxic organic contaminants. One example is the biodegradation of vinyl chloride (VC), a known human carcinogen. *Nocardioides* sp. strain JS614 is model aerobic VC-assimilator, displaying the highest protein yield (10.3 g protein mol^{-1}) and maximum VC utilization rate (43.1 nmol min^{-1} mg^{-1} protein) at 20 °C among all known VC isolates[10]. Unlike other VC-assimilating isolates studied, *Nocardioides* sp. strain JS614 undergoes an extended lag period in response to VC starvation[10, 36]. Strain JS614 is also unusual among other VC-assimilators in that it contains an *etnE* allele with a 7-bp deletion, called *etnE1*.

The wild-type *Mycobacterium* sp. strain JS622 was isolated on ethene as the

carbon and energy source [11]. *Mycobacterium* sp. strain JS622 can degrade VC via cometbolism: it can degrade VC while ethene is present (this study, Figure 4.4), but it normally could not sustain on VC as the sole carbon source[11]. However, strain JS622 was shown to adapt to VC as a carbon and energy source after a prolonged incubation period (110-125 days)[31].

The ethene and VC aerobic degradation pathway currently known includes the following process: the first step is breaking up the double-carbon bond by alkene monooxygenase (AkMO, encoded by gene *etnABCD*), yielding epoxyethane from ethene and/or chlorooxirane from VC. Then an epoxyalkane-coenzyme M transferase (EaCoMT, encoded by gene *etnE*) will conjugate the CoM group onto these epoxides, converting them to hydroxyalkyl-CoM derivatives[9, 10, 12]. The gene *etnABCD* and *etnE* have been found in *Nocardioides* sp. strain JS614 and the enzyme activities of its AkMO and EaCoMT have been confirmed[9], whereas in *Mycobacterium* sp. strain JS622 only the presence of $etnC^{77}$ and $etnE^{11}$ have been validated. As *Mycobacterium* strain JS622 is ethene-assimilating, the reason why it can not utilize VC as the sole carbon source remains elusive.

For VC bioremediation strategies, the growth-coupled assimilation process represents a more sustainable and cost-effective approach than cometabolism. However, current monitoring methods (e.g. quantitative PCR) cannot distinguish between these two groups, as both of them have the functional gene *etnC* and *etnE*. Therefore, the initial goal of this study was to develop a diagnostic approach involving stable isotope probing (SIP) to distinguish between VC-cometabolizers and VC-assimilators. We were interested to see whether SIP could be used to distinguish between strain JS622 and strain JS614 in

a VC-assimilating mixed culture. Therefore, we studied culture growth and VC degradation patterns in defined mixed cultures of strain JS614 and strain JS622. We hypothesized that the buoyant density of VC-assimilating *Nocardioides* sp. strain JS614 would become heavier in SIP experiment, while *Mycobacterium* strain JS622 would not. Nonetheless, on the other hand, it is possible that *Nocardioides* sp. strain JS614 could have effects on *Mycobacterium* strain JS622, which means these two strains may have different behavior mixing together compared to being fed on VC individually. Our experiement results showed the later case, in which strain JS622 started to grow on VC after mixed with strain JS614. Therefore, this case is reported here to discuss the potential interaction between VC assimilators and cometabolizers.

4.3 Materials and Methods

4.3.1 Chemicals, media and analytical method

Ethene (99%) (Airgas, Cedar Rapids, IA, USA), unlabeled VC ($^{12}C_2$-VC, 99.5%) (Sigma Aldrich, St. Louis, MO, USA) and stable isotope labeled VC ($^{13}C_2$-VC, 99%) (Cambridge Isotope Laboratories, Tewksbury, MA, USA) were used in this study. All the other chemicals were either reagent or molecular biology grade. Minimal salts medium (MSM) was prepared as previously described [79, 89]. Trace metals solution (TMS)(per liter) was prepared to contain 60 g $MgSO_4$•$7H_2O$, 6.37 g EDTA ($Na_2(H_2O)_2$), 1 g $ZnSO_4$•$7H_2O$, 0.5 g $CaCl_2$•$2H_2O$, 2.5 g $FeSO_4$•$7H_2O$, 0.1 g $NaMoO_4$•$2H_2O$, 0.1 g $CuSO_4$•$6H_2O$, 0.2 g $CoCl_2$•$6H_2O$ and 0.52 g $MnSO_4$•H_2O in deionized water. The $1/10^{th}$ strength trypticase soy agar plus 1% glucose (TSAG) plates (per liter) was prepared to contain 3g Tryptic Soy Broth (TSB), 15g Bacto agar, 10g glucose(1%) in deionized water.

Ethene and VC concentrations in enrichment cultures were monitored via gas chromatography with flame-ionization detection and calculated as described previously [9, 72]. Optical density at 600 nm (OD_{600}) was measured with a Cary 50 Bio UV-Visible spectrophotometer to monitor bacterial growth.

4.3.2 Preparation of bacterial strains

Frozen stocks of *Nocardioides* sp. strain JS614 (ATCC # BAA-499) and *Mycobacterium* strain JS622[11] were revived using ethene as the sole carbon source to reach an active status for experiments with VC. Both bacteria stocks (1 mL of each, OD600=1.30-3.28) were thawed on ice and added separately into a modified 2-liter Erlenmeyer flasks with 500 mL sterile MSM, 1 mL of filter-sterilized (0.22 μm PVDF membrane from MILLEX, Sigma Aldrich, St. Louis, MO, USA) TMS and 100 ml of filter-sterilized ethene. The bottles were then capped with slotted butyl rubber septa (Wheaton, Millville, NJ, USA) and aluminum crimp caps. Both pure cultures were incubated aerobically on a circular shaker at 200 revolutions per minute (rpm) at room temperature (RT, ~22°C) in the dark. Headspace VC concentration and OD600 were measured along to monitor the growth condition of both cultures. After both cultures reached exponential phase (JS614 OD600=0.375 and JS622 OD600=0.383) (Figure 4.1), cultures were pelleted separately via centrifugation for 10 min at 8000 ×g, 4°C. Pellets were washed with 25 mL sterile MSM, added with 100 μL sterile tween 80 (Sigma Aldrich, diluted to 0.01%) and pelleted again. The cultures were resuspended into 10 mL MSM. Final OD600 of JS614=13.4 and OD600 of JS622 =3.19. Immediately following OD600 measurement these dense cultures were used for purity check and further mixed

culture experiment.

4.3.3 Purity check

Bacteria culture harvested from 500 mL flask in Bacteria Preparation were diluted and spread plated onto $1/10^{th}$ strength TSAG plates and incubated at 30°C for purity check. Within two weeks, only small milky-white colonies with diameter about 3mm showed up on *Nocardioides* sp. strain JS614 plates. On *Mycobacterium* strain JS622 plates, only pure white colonies with diameter about 1mm showed up, with agglomeration of pure white colonies at some sites. The morphology of JS614 and JS622 on the plates showed they were pure cultures and similar to previous documentation in our lab. The colonies were harvested from the plate and DNA were extracted separately for strain JS614 and JS622 using a previously published bead-beating method [31], followed by PCR on 16S rRNA gene. The Sanger sequencing of *Nocardioides* sp. strain JS614 and *Mycobacterium* strain JS622 16S rRNA gene PCR products showed clear peaks on the electropherogram, adding another line of evidence for their purity. The BLAST result against NCBI Genbank showed the PCR products sequenced were 100% to previously deposited 16S rRNA sequences from *Nocardioides* sp. strain JS614 and *Mycobacterium* strain JS622. Sanger-sequencing was performed at the Iowa Institute of Human Genetics at the University of Iowa.

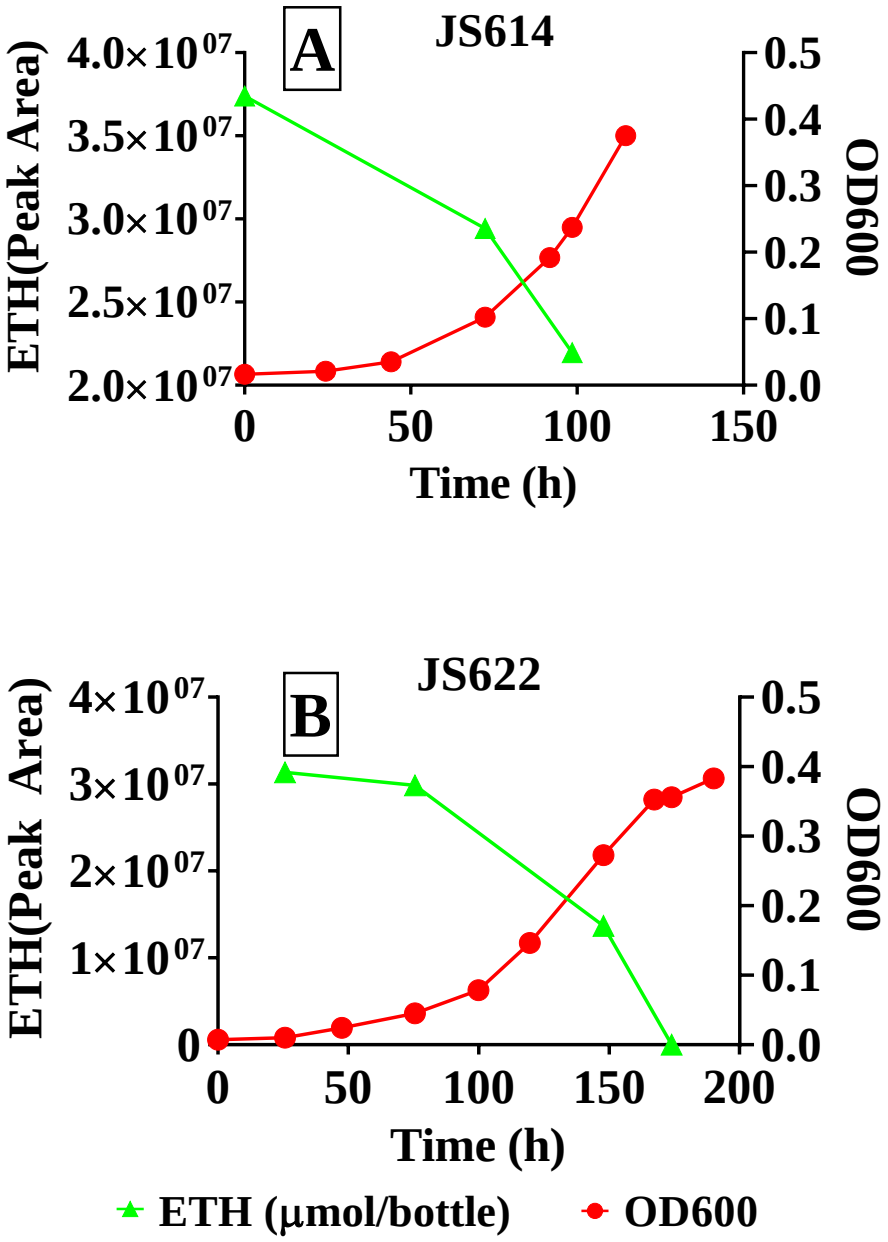


Figure 4.1 Reviving *Nocardioides* sp. strain JS614 and *Mycobacterium* strain JS622 on ethene as the sole carbon source.

Note: As standards were not constructed for 500 mL flasks, the raw GC reads (peak area) were presented to show the relative abundance of ethene in each culture. OD600 was measured to reflect the growth of bacteria.

4.3.4 Confirmation of assimilating and cometabolizing activities

Nocardioides sp. strain JS614 and *Mycobacterium* strain JS622 stock were revived as in Bacteria Preparation from the same batch of stock as for mixed culture experiment. After reaching the mid-exponential phase, cultures were harvested, washed and spiked in 160 mL serum bottles separately as stated above. Four bottles were prepared, two for JS614 and the other two for JS622. Both *Nocardioides* sp. strain JS614 and *Mycobacterium* strain JS622 were fed with filter-sterilized unlabeled VC (50-100 μmol) respectively. OD600 and headspace were measured along to monitor bacteria growth and degradation of ethene and VC. Another two bottles were prepared, one for strain JS614 and one for strain JS622, to show that they can degrade VC when ethene is present. Each culture bottle was fed with filter-sterilized ethene (~450 μmol) and unlabeled VC (20 μmol).

4.3.5 Stable isotope probing experiment

Six 160 mL serum bottles were prepared, to which calculated amount of JS614 and JS622 stocks, sterile MSM and filter-sterilized TMS were added to make the final volume 72 mL of liquid culture (initial OD_{600} =~0.04, JS614:JS622=1:1), leaving 88 ml of headspace. The calculation was based on the assumption that OD600 is proportional to the concentration of cells. The equations used are provided below:

0.02×72 mL=OD600 (JS614) × A mL of JS614 (e.q.1)

0.02×72 mL=OD600 (JS614) × B mL of JS622 (e.q.2)

$$\frac{\text{C mL TMS}}{\text{72 mL}} = \frac{\text{1 mL}}{\text{500 mL}} \text{ (e.q.3)}$$

Volume (mL) of MSM =72-A-B-C (e.q.4)

The bottles were then capped as stated above. Autoclaved controls were also prepared in parallel to test for abiotic losses. Three bottles were fed with filter-sterilized 1.5 mL (about 68 to 73 μmol) unlabeled VC and three were fed with 1.5 mL (about 58 to 96 μmol) labeled VC. All bottles were incubated as described above.

4.3.6 DNA extraction, fractionation and purification

Three time points (~20, ~50 and ~70 μmol VC degradation) were selected for DNA extraction in both labeled and unlabeled VC amended cultures. A list of the amount of VC degraded and the original amount of VC in each culture is provided in Table 4.1. At each time point, one bottle with labeled VC and one with unlabeled VC were sacrificed for analyses. Each culture was first split into two sterile 50 mL centrifuge bottles (Falcon, Corning, NY, USA), then pelleted by centrifugation and washed as described above. The pellets were finally suspended in 600 μL Sodium Chloride-Tris-EDTA (STE) buffer and extracted using a bead-beating method [31]. DNA concentration was measured using Qubit® dsDNA BR Assay Kits.

Table 4.1 Amount of VC degraded in each culture. Time point 1: ~20 µmol VC degradation; time point 2: ~50 µmol VC degradation; time point 3: ~70 µmol VC degradation.

Sample	Initial VC amount (µmol)	Time before DNA extraction (h)	Cumulative VC degraded (µmol)	VC degraded (%)	Average VC degradation rate* (µmol h^{-1})
12C-VC-1 (~20 µmol)	67.92	3.83	20.39	30.01	5.32
12C-VC-2 (~50 µmol)	68.40	45.0	46.84	68.47	1.04
12C-VC-3 (~70 µmol)	73.10	35.1	72.16	98.71	2.06
13C-VC-1 (~20 µmol)	57.79	4.08	23.84	41.25	5.84
13C-VC-2 (~50 µmol)	77.77	34.6	54.10	69.57	1.56
13C-VC-3 (~70 µmol)	96.09	34.7	65.41	68.07	1.89

Note: In sample names, C13-VC represents the labeled VC amended samples, C12-VC represents the unlabeled VC amended samples and 1, 2, 3 at the end represent the time point.

*Average VC degradation rates were calculated as VC degraded (µmol)(column 4) divided by time lasted before DNA extraction (h)(column 3).

Quantified DNA extracts (~10 μg) were processed and centrifuged as described previously [64, 65], which includes mixing with Tris-EDTA (pH 8.0)-CsCl solution, sealing of centrifuge tubes and centrifuge for about 48 hrs at 178,000 × g (20 °C). Following ultracentrifugation, the tubes were placed onto a fraction recovery system (Beckman Coulter, Indianapolis, IN, USA), and 16 fractions (a total of 150 μL) were collected for each culture. The BD of each fraction was measured, and CsCl was removed by glycogen-assisted ethanol precipitation[64]. Fractions were re-purified using QIAquick PCR Purification Kit as necessary.

4.3.7 Quantitative PCR on functional gene etnE

The *etnE* in all 96 (16 fractions × 3 time points × 2 experimental group for labeled and unlabeled VC respectively) DNA fractions (BD=1.370-1.781 g mL^{-1}) extracted from the three VC degradation points were quantified using three runs of qPCR. Each run contained duplicates of 32 DNA fractions from both labeled and unlabeled VC amended cultures (16 fractions each) from each time point.

An ABI 7000 Sequence Detection System (Applied Biosystems, Foster City, CA, USA) was used to conduct qPCR as described previously [77, 90]. Reaction mixtures (25 μL) contained 12.5 μL of Power SYBR Green PCR Master Mix (Applied Biosystems), 750 nM *etnE* qPCR primers (RTE_F: 5'-CAGAAYGGCTGYGACATYATCCA-3' and RTE_R: 5'-CSGGYGTRCCCGAGT-AGTTWCC-3') [77]and 2 μL of DNA extract.

Standard curves were developed in triplicate using *etnE* from *Nocardioides* sp. strain JS614[10] amplified using the CoMF1L (5'-AACTACCCSAAYCCSCGCTGGTACGAC-3') and CoMR2E (5'-GTCGGCAGTTTCGGTGATCGTGCTCTTGAC-3') primer set[11]. Reactions (25 μL)

contained 12.5 µL Qiagen PCR Master Mix, 0.2 µM of each primer and 2 ng of total DNA. Genes per µL of PCR product were estimated using a previously reported equation[77]. ABI 7000 System SDS software (Applied Biosystems) was used to analyze real-time PCR fluorescence data using the auto baseline function. The fluorescence threshold was set manually to optimize qPCR efficiency and obtain a linear fit of the standard curve, yielding PCR efficiency between 96%-114% and R^2=99.3-100%. Related information is provided in accordance with MIQE guidelines [60] in Table AI.2.

4.3.8 T-RFLP on 16S rRNA gene and functional gene etnE

We performed T-RFLP on both the 16S rRNA gene and functional gene *etnE* to differentiate *Nocardioides* sp. strain JS614 and *Mycobacterium* strain JS622 in the defined mixed culture. The restriction enzyme AluI was selected for 16S rRNA gene and AcoI (EaeI) for *etnE*, using the default settings of the software program REPK [91]. *In silico* analysis showed the restriction enzyme AluI generates a 198 bp (in vivo 197-199 bp)16S rRNA gene fragment of JS622 and a 234 bp (in vivo 233-235bp)16S rRNA gene fragment of JS614, based on the PCR products (525 bp) from 8F (5'-AGAGTTTGATCMTGGCTCAG-3')[92] and 533R (5'-TTACCGCGGCTGCTGGCAC -3')[93] universal primers. Restriction enzyme AcoI generates a 395 bp (in vivo 390, 391 and 395-396 bp) *etnE* fragment of JS622 and two *etnE* fragments of JS614, which are 319 bp (in vivo 319-321 bp)(*etnE*) and 433 bp (433-435 bp)(*etnE1*) respectively, based on the PCR products (891 bp) from CoM F1L and CoM R2E primers.

T-RFs were generated, purified and analyzed as published previously for 16S rRNA gene (with fluorescently-labeled 6-FAM 8F and unmodified 533R primers) [94, 95]

and *etnE* gene (with fluorescently-labeled 6-FAM CoMF1L and unmodified CoM R2E primers)[11] respectively. Fragment sizes <48 bp and >500 bp (maximum standard size) were excluded from further analysis.

Two rounds of T-RFLP were performed. T-RFLP was executed on the 16S rRNA gene amplified from total DNA of labeled and unlabeled VC amended cultures at the three VC degradation time points, which aimed to track the growth of strain JS614 and JS622 in the mixed culture. The other round was executed on both 16S rRNA gene and *etnE* gene for all the SIP fragments, using only the cultures from the last time point (1.5~2 days, ~70 µmol VC degraded), to collect evidence for VC-assimilation. Relative abundance (% peak area) of JS614 and JS622 16S rRNA gene and *etnE* fragments were calculated for each SIP fraction and plotted to show the distribution of these genes across SIP fraction gradient.

4.4 Results

4.4.1 VC biodegradation patterns in strain JS614 and JS622 pure cultures

Ethene-grown strain JS614 degraded a spike of VC (100 µmol) in about two days and concurrently displayed an increase in OD600 frm 0.0236 to 0.0547 (Figure 4.2 A). Ethene-grown strain JS622 could only degrade VC from about 100 µmol to about 50 µmol and OD600 dropped from 0.0165 to 0.0026. (Figure 4.2 B). This confirmed that *Nocardioides* sp. strain JS614 was able to utilize VC as the sole carbon source for growth, while *Mycobacterium* strain JS622 could only not. Replicate experiment showed consistent degradation pattern (Figure 4.3). Both strains degraded VC (~20 µmol) completely when ethene was present (Figure 4.4).

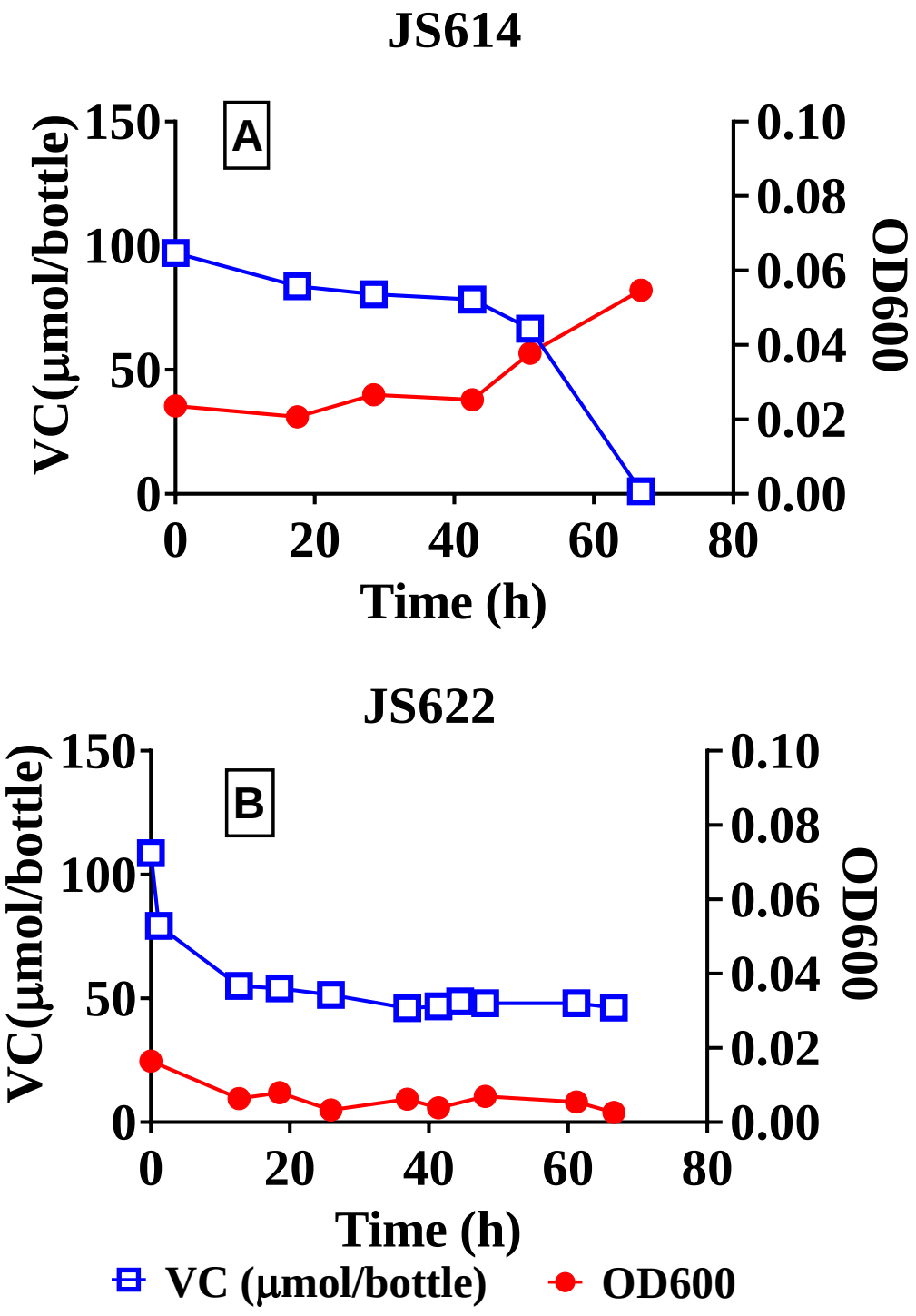


Figure 4.2 *Nocardioides* sp. strain JS614 assimilates VC (A); *Mycobacterium* strain JS622 could not use VC as the sole carbon source (B).

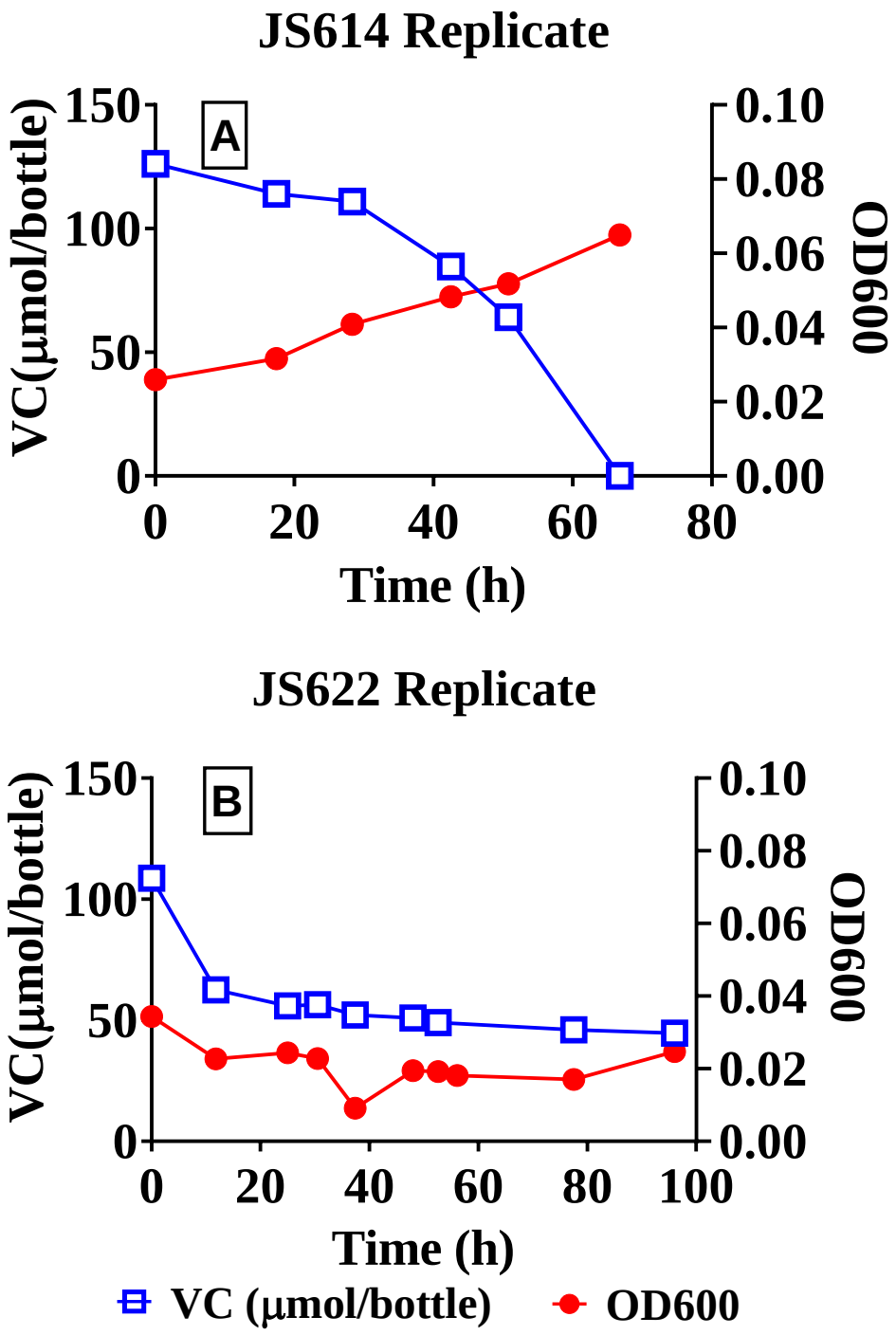


Figure 4.3 *Nocardioides* sp. strain JS614 assimilates VC (A); *Mycobacterium* strain JS622 could not use VC as the sole carbon source (B). (Replicate)

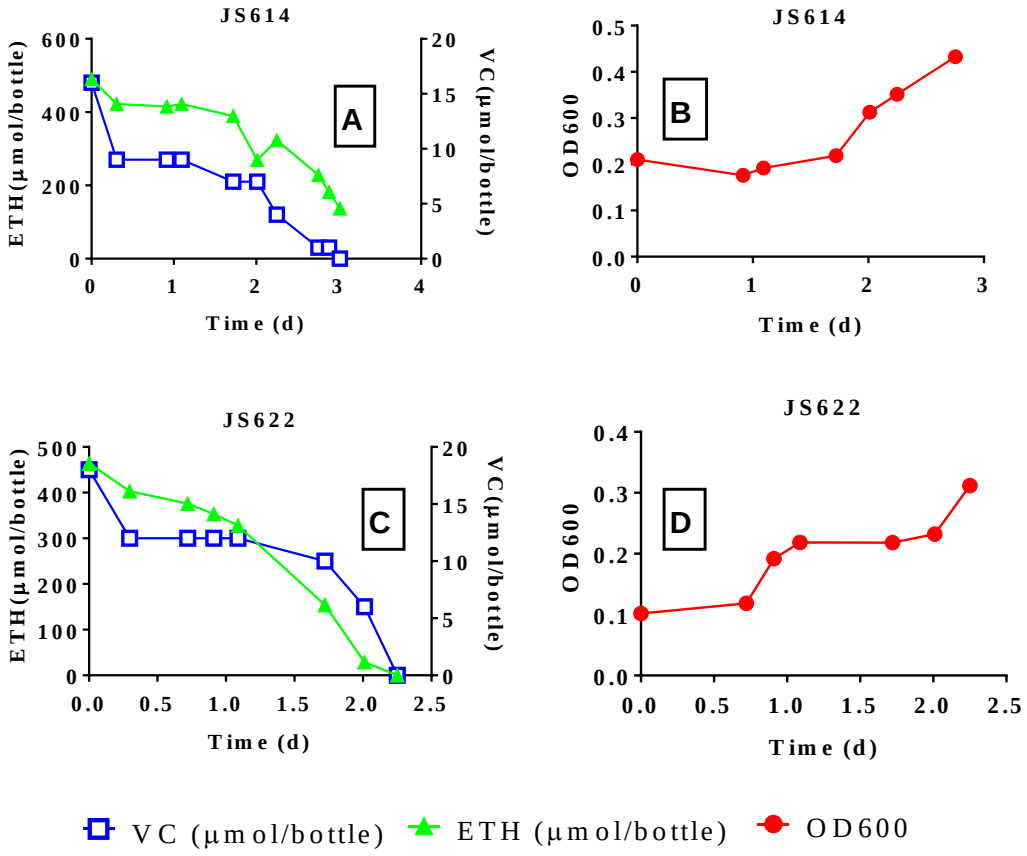


Figure 4.4 *Nocardioides* sp. strain JS614 (A and B) and *Mycobacterium* strain JS622 (C and D) could degrade VC when ethene was present.

4.4.2 Bacterial growth and VC biodegradation patterns in a defined mixed culture of strains JS614 and JS622

When ethene-grown *Nocardioides* sp. strain JS614 and *Mycobacterium* strain JS622 were mixed together at a 1:1 ratio, according to OD600 measurements and subsequently fed VC, VC degradation was first noted at 3 hr in all six cultures (15 to 57 µmol VC degraded) (Figure 4.5). After 48 hrs about 65 µmol of labeled VC and 73 µmol of unlabeled VC was degraded in cultures harvested at VC degradation time point 3 (Figure 4.5) (Table 4.1, ~ 70 µmol VC degradation point). As VC was degraded, there was a decrease in OD600 during the first 12 hours (from ~0.04 to ~0.02). Upon the completion of the experiment, the OD600 has steadily increased and exceeded the initial level (0.048 to 0.049). Overall, the pattern of VC biodegradation appeared cometabolic (decrease in VC concentration and no increase or even decrease in OD600) initially, then switched to a growth-coupled pattern (decrease in VC concentration and increase in OD600).

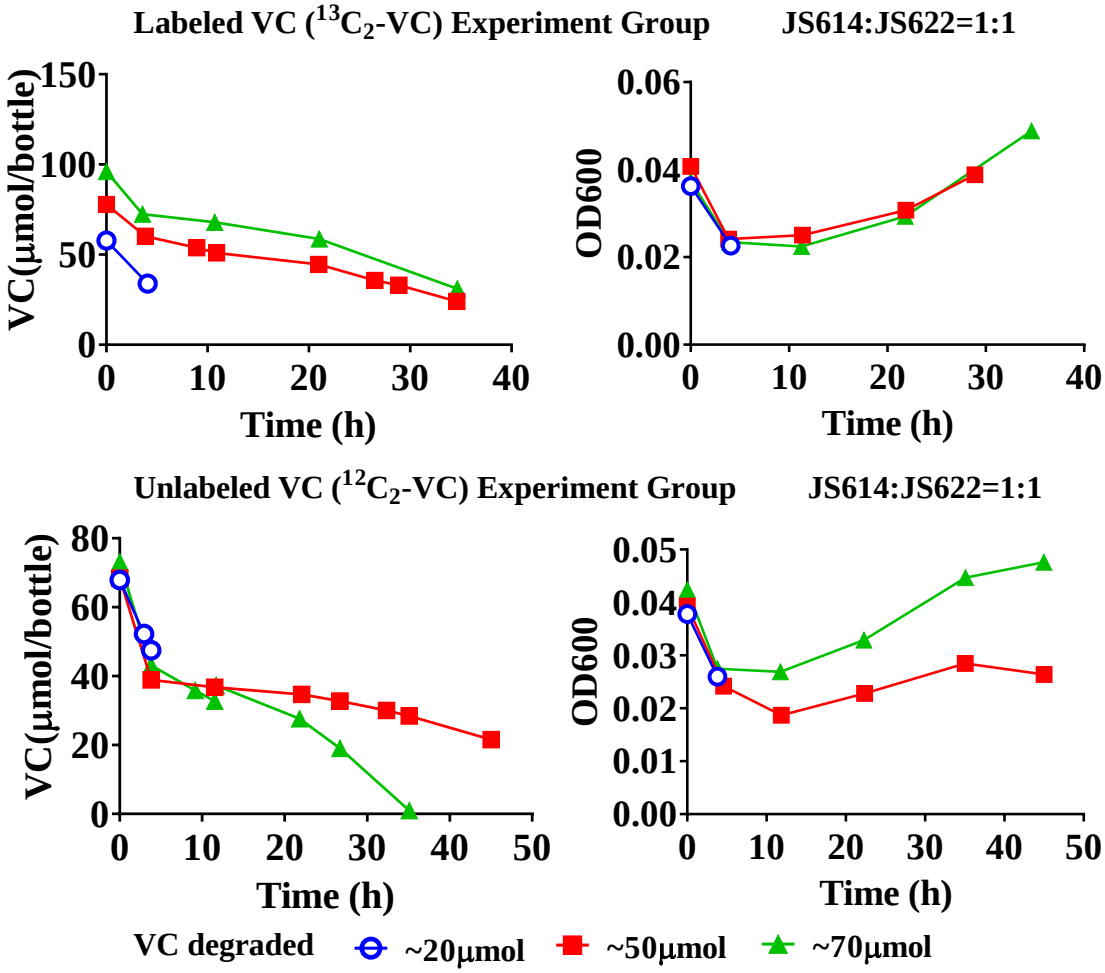


Figure 4.5 The growth (assessed by OD600) and degradation of VC in mixed cultures. Three were amended with ^{13}C labeled VC (upper panel), the other three were amended with unlabeled VC (lower panel).

4.4.3 Evidence of VC assimilation in mixed culture from the SIP-qPCR data

The DNA concentration in each fraction from the last time point (~70 μmol of VC degradation) was determined. In the labeled VC amended microcosm, most of the heavier fractions (BD=1.74-1.80 g mL^{-1}) have higher DNA concentrations than those in the unlabeled VC amended control group (Figure 4.6).

The abundance of functional gene *etnE* was also higher in heavier fractions (BD=1.74-1.80 g mL^{-1}) in the labeled VC amended microcosm compared to the unlabeled VC amended controls at each time point (Figure 4.6 B-D). As more VC was degraded in the mixed culture, the separation between *etnE* abundances in labeled and unlabeled VC amended cultures became more apparent, indicating the incorporation of ^{13}C into DNA. However, unexpectedly, the pattern of DNA concentration and *etnE* qPCR showed genes were spread all over the BD range, even in very light SIP fractions (BD=1.37-1.57 g mL^{-1}).

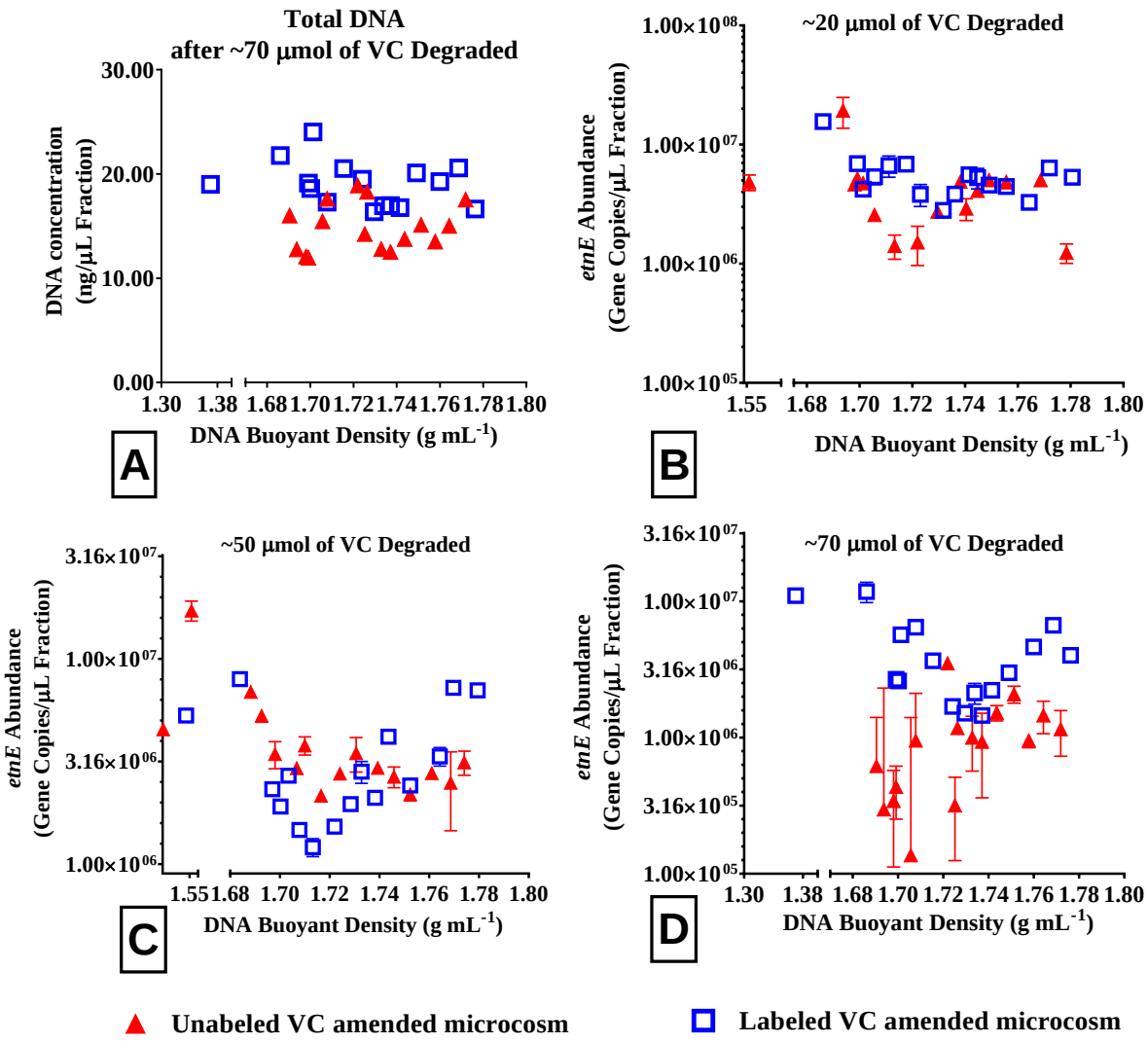


Figure 4.6 Evidence of ^{13}C take-up: Total DNA concentration across BD after ~70 µmol of VC degradation (A); *etnE* gene abundance in the mixed cultures after ~20 µmol (B), ~50 µmol (C) and ~70 µmol (D) of VC degraded across BD range. Error bars indicate the differences in qPCR replicates.

4.4.4 Carbon up-take from VC and growth observed in JS622

T-RFLP methods for both 16S rRNA gene and functional gene *etnE* were applied here to check the carbon up-take from VC in *Nocardioides* sp. strain JS614 and *Mycobacterium* strain JS622 respectively. Unexpectedly, the T-RFLP results showed that, in labeled VC amended cultures, both 16S rRNA gene and *etnE* gene from *Mycobacterium* strain JS622 were more abundant in heavier fractions (BD=1.74-1.78 g mL^{-1}) after ~70 μmol of VC was degraded (Figure 4.7 B and E), compared to the unlabeled VC amended cultures, indicating ^{13}C up-take from labeled VC. However, the 16S rRNA gene and the *etnE1* allele in *Nocardioides* sp. strain JS614 did not show higher abundance in the heavier fractions (Figure 4.8 A and D) in the labeled VC amended cultures. There was a slight enrichment observed with JS614 *etnE* (Figure 4.8 C) in heavier fractions from the labeled VC amended cultures.

To track the approximate growth of *Nocardioides* sp. strain JS614 and *Mycobacterium* strain JS622 in the mixed culture, the 16S rRNA gene T-RFLP method was also applied on the total DNA samples from the three VC degradation time point in both labeled and unlabeled VC amended cultures. The result showed that the relative abundance of 16S rRNA gene of *Nocardioides* JS614 decreased from about 80% to 50% in the mixed culture. On the contrary, the relative abundance of 16S rRNA gene of *Mycobacterium* JS622 increased from 11% to 40% in the mixed culture, which was unexpected, indicating the growth of this strain (Figure 4.8).

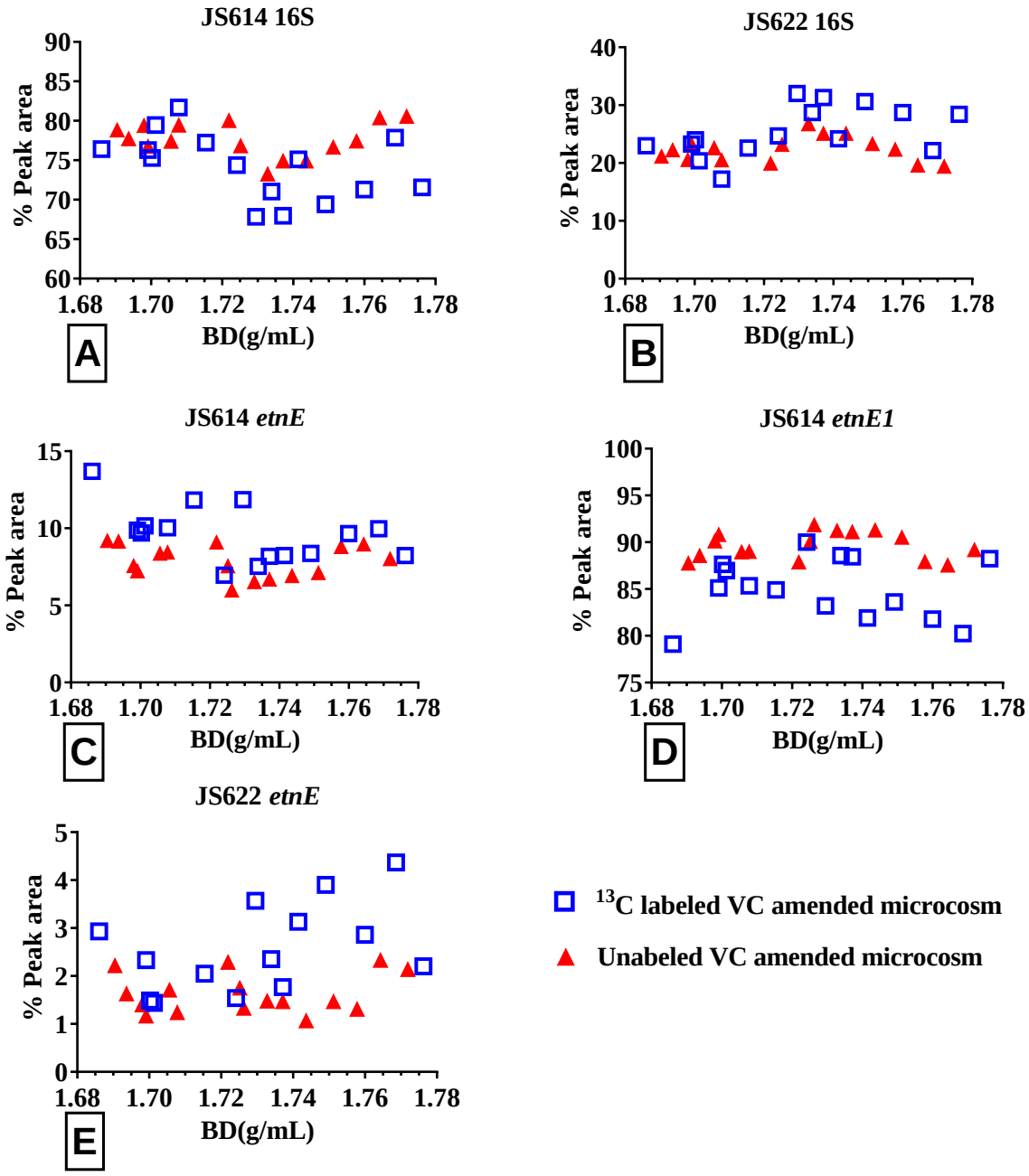


Figure 4.7 T-RFLP on 16S rRNA gene and functional gene *etnE* for *Nocardioides* sp. strain JS614 (A, C, D) and *Mycobacterium* strain JS622 (B, E) across SIP fractions with BD from 1.68-1.78 g mL^{-1} in cultures after ~70 µmol of VC was degraded.

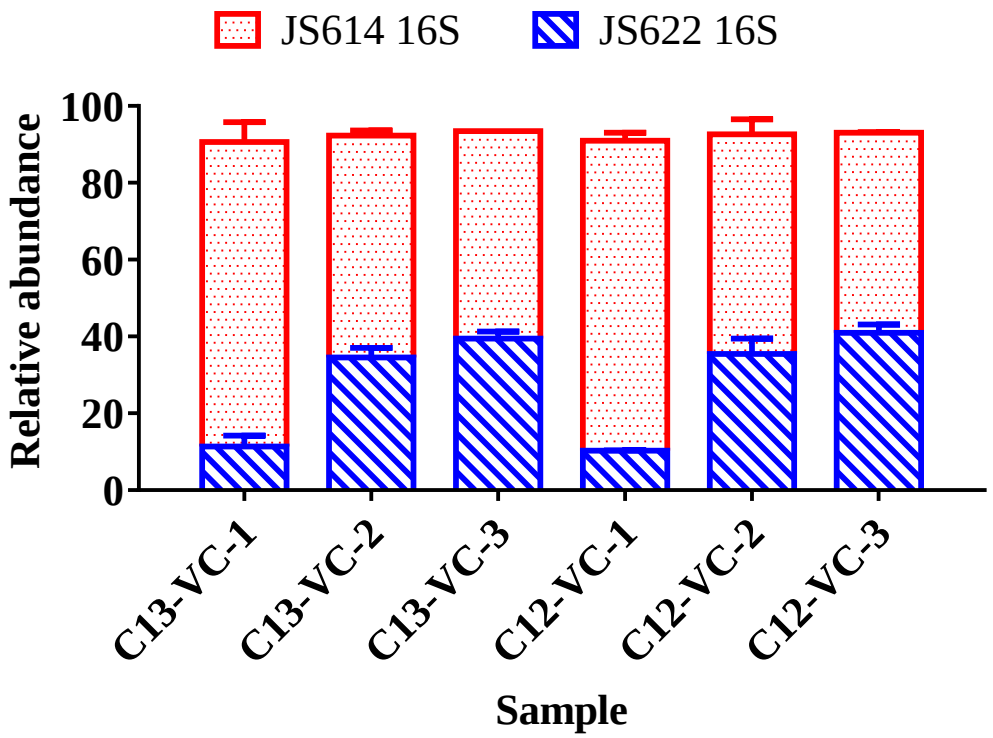


Figure 4.8 T-RFLP of 16S rRNA genes on labeled and unlabeled VC amended cultures after ~20 µmol (C13-VC-1, C12-VC-1), ~50 µmol (C13-VC-2, C12-VC-2) and ~70 µmol of VC (C13-VC-3, C12-VC-3) was degraded. Error bars indicate the differences in T-RFLP replicates.

Note: In sample names, C13-VC represents the labeled VC amended samples, C12-VC represents the unlabeled VC amended samples and 1, 2, 3 at the end represent the time point.

4.5 Discussion

4.5.1 JS622 can take up ^{13}C from VC when mixed with VC-assimilator JS614

Mycobacterium strain JS622 is known as an ethene-assimilating bacteria and can cometabolize VC after fed on ethene [31]. VC assimilation by strain JS622 was observed in one case after 110-125 days of VC-incubation [31]. As most of the ethene-assimilating bacteria can utilize the same pathway to assimilate VC, *Mycobacterium* strain JS622 has the unexpected behavior that could not readily grow on VC as the sole carbon source[11], which means it should not up-take ^{13}C from labeled VC. The pure culture experiment before culture-mixing also confirmed that *Mycobacterium* strain JS622 could not readily sustain on VC as the sole carbon source. However, after mixing *Mycobacterium* strain JS622 with the known VC-assimilator *Nocardioides* sp. JS614, it was observed via T-RFLP data that *Mycobacterium* strain JS622 began to grow in the VC amended mixed culture. Although the absolute amount of *Nocardioides* sp. strain JS614 and *Mycobacterium* strain JS622 could not be quantified accurately via T-RFLP, the relative abundance of *Mycobacterium* strain JS622 kept increasing during the course of the mixed-culture experiment. Moreover, the DNA buoyant density was observed to have shifted towards heavier fractions in both JS622 16S rRNA gene and functional gene *etnE*, indicating it has up-taken ^{13}C from labeled VC.

There were two possible explanations. One explanation might be that cross-feeding has happened in the culture, where *Mycobacterium* was either 1) consuming the cell debris of *Nocardioides* sp. JS614, or 2) benefiting from metabolites produced by *Nocardioides* sp. strain JS614, especially 2-Ketoethyl-CoM (the CoM-conjugated VC epoxides). In the first scenario, we should observe genes from both strain JS614 and

strain JS622 shifted towards the heavier fractions in the SIP experiment. However, the T-RFLP on SIP fractions showed that the enrichment of *Nocardioides* sp. JS614 DNA in heavier fractions was not obvious. Furthermore, the experiment only lasted for less than three days, which minimized the effects of cross-feeding on cell debris. Therefore, the second scenario is more plausible. *Mycobacterium* strain JS622 possesses the complete ethene- and VC-assimilation pathway, but the EaCoMT it produced was not as effective as those in *Nocardioides* sp. strain JS614 to convert VC epoxides. This is also supported by observation in previous studies, where *Mycobacterium* strain JS622 had lower EaCoMT activity (850-1000 nmol min^{-1} mg^{-1} of protein on ethene)[9, 11] and *Nocardioides* sp. strain JS614 demonstrated higher EaCoMT activity (3550-9170 nmol min^{-1} mg^{-1} of protein on VC and 1290-6790 nmol min^{-1} mg^{-1} of protein on ethene) [9]. When mixed with *Nocardioides* sp. strain JS614, *Mycobacterium* strain JS622 was able to take the intermediates, e.g. 2-Ketoethyl-CoM, produced in the VC-assimilation pathway of JS614 and completed the full pathway to support its growth, which resulted a competition for the intermediates. Therefore, we could see an increase of the relative abundance of *Mycobacterium* strain JS622 in the mixed culture whereas a slight decrease of the relative abundance of *Nocardioides* JS614 was observed.

Another explanation is that *Nocardioides* sp. strain JS614 could help improve the EaCoMT activity of *Mycobacterium* strain JS622. Previous study[8] showed that addition of exogenous Coenzyme M (CoM) could increase the activity of EaCoMT in cell-free assays. It is possible that the EaCoMT activity in *Mycobacterium* strain JS622 was promoted by more CoM produced by *Nocardioides* sp. strain JS614, therefore strain JS622 was able to continue the downstream reactions in the ethene/VC assimilation

pathway.

4.5.2 Faster degradation of VC in mixed culture

As *Nocardioides* sp. strain JS614 experienced lag-period after starvation in previous experiments [11, 36], it has been proposed that an ideal VC degrader for bioremediation would combine the high activity of *Nocardioides* sp. strain JS614 with the starvation-resistant character of *Mycobacterium* strains [8]. In this study, we see the possibility of combining *Nocardioides* sp. strain JS614 and *Mycobacterium* strain JS622 to faster degrade VC in mixed culture. When JS614 was cultured alone on VC as the sole carbon source (Figure 1E), it degraded ~50 μmol of VC in 44 hrs, which is an average degradation rate of 1.09 μmol h^{-1}, with OD600 increased from ~0.02-0.04. Surprisingly, when *Nocardioides* sp. strain JS614 was co-cultured with *Mycobacterium* strain JS622, the six mixed cultures showed average VC degradation rates of 1.04-5.83 μmol h^{-1}(Table 4.1), with OD600 changed from ~0.04 to ~0.02 then back to ~0.05, which is faster than pure *Nocardioides* sp. strain JS614 culture. Albeit the mixed cultures were not tested for starvation, their faster VC degradation rate is intriguing for bioremediation application.

4.5.3 The cause for irregular pattern in SIP-qPCR

Although in our study DNA was seen to be accumulated in the heavier fractions compared to the unlabeled VC control group, neither the qPCR-BD pattern nor the TRFLP-BD pattern looks ideal: the DNA spread all over the BD with high DNA concentration in light fractions (<1.60 g mL^{-1}), and there was no significant peak. Ideally, after enriched on labeled substrate for sufficient amount of time, the genome of a certain

bacteria should aggregate at a specific buoyant density after centrifugation, forming one distinct peak on the gene abundance-BD graph. In this study, since two different strains (JS614 and JS622) were used, they were expected to form about two peaks on the graph, if their genomic GC contents are not the same. Our irregular pattern is probably caused by problems during DNA extraction, since the only process changed in this study compared to previous studies is using the self-developed bead-beating method instead of MO BIO PowerWater DNA Isolation Kit for DNA extraction. It is possible that the bead-beating protocol we used could not yield high-purity DNA and has introduced some chemicals that interfere with centrifugation. In this study, we also encountered PCR inhibition problem after the sample were processed by bead-beating extraction and centrifugation, which indicates PCR inhibitor (e.g. proteins) exist in our samples. Therefore, we recommended using kits yielding high-quality DNA when performing SIP experiments.

Other factors also need to be considered. For example, the experiment lasted for less than three days, which could cause insufficient enrichment on labeled VC. In this case, since genomic DNA with different amount of ^{13}C will have different BD, the pattern will not be an ideal single peak for the labeled VC amended culture. Previous study found that for bacteria having genomic GC content =70%, DNA with ^{13}C might appear all over the range of BD=1.74-1.77 g mL^{-1} [90] if the enrichment is not sufficient. In our study, the strain JS614 has chromosome GC content=71.65% and its plasmid that bearing *etnE* has a GC content=68.01%, which should have a peak area around BD=1.74-1.77 g mL^{-1}, even if the enrichment was not sufficient. However, the qPCR-BD and T-RFLP-BD patterns are broader than that range, which suggests factors other than insufficient

enrichment have affected the pattern. The whole-genome of *Mycobacterium* strain JS622 has never been sequenced before so there is no GC content data available for this strain. Therefore, it is hard to predict around which BD range JS622 DNA will aggregate. Nonetheless, the intrinsic BD range of *Nocardioides* sp. strain JS614 and *Mycobacterium* strain JS622 genome could have been determined in the unlabeled VC amended cultures via centrifugation of the DNA, if the bead-beating method has not affected the centrifugation and following molecular analysis.

Last but not least, T-RFLP is not an absolute quantification method, which is not as accurate as qPCR to manifest the peak areas across BD. However, since both *Nocardioides* sp. strain JS614 and *Mycobacterium* strain JS622 are *Actinobacteria*, with similar 16S rRNA gene (89% identity based on Genbank accessions AF498652.1 and AY162027.1) and *etnE* gene (74% identity based on Genbank accessions CP000508.1 and AY243040.1), our several attempts to design primers to distinguish these two strains have failed, as our primers always amplified both of the strains.

4.5.4 COM primer set tends to amplify etnE1

The T-RFLP of functional gene *etnE* was based on the PCR using degenerate primer set COM F1L and COM R2E. This primer set has been used in amplifying *etnE* gene from pure and enrichment cultures and environmental samples [11, 77, 96-98]. In this study, we discovered that this primer sets tends to amplify more *etnE1* than *etnE* sequences (Figure 4.7 C and D): the T-RFLP data showed JS614 *etnE1* was of 80-90% relative abundance of all *etnE* gene in the culture, whereas *etnE* of JS614 was only accounted for 5-15%. Nonetheless, from the whole-genome sequencing data of

Nocardioides sp. strain JS614, only one copy of *etnE* and *etnE1* are observed, which means their ratio in JS614 should be 1:1. Taking a close look at the sequences, we found that the difference lies in R2E primer. The R2E primer has 4-bp mismatches with *etnE1* and 5-bp mismatches with *etnE*. The *etnE1* allele has a 7-bp-deletion, which could cause loss of function of this gene [9]. As COM F1L and COM R2E have bias against *etnE* when *etnE1* is present, it might not be an ideal primer sets for clone library or diversity study if *etnE1* or similar genes are present. New reverse primer should be designed to address this issue.

4.6 Conclusion

Mycobacterium strain JS622 could not readily grow on VC as the sole carbon source. However, when mixed with VC-assimilating *Nocardioides* sp. strain JS614, growth was observed for *Mycobacterium* strain JS622 and ^{13}C from labeled VC was seen accumulated in JS622. The change of behavior was possibly caused by the existence of *Nocardioides* sp. strain JS614 in the mixed culture, which could provide metabolites such as chlorooxirane, Coenzyme M-conjugated epoxide or more Coenzyme M. The mixed culture demonstrated faster VC degradation rates compared to pure *Nocardioides* sp. strain JS614 culture, which could be considered for bioremediation practice.

CHAPTER V. MICROBIAL ADAPTATION TO VINYL CHLORIDE IN GROUNDWATER MICROCOSMS AS REVEALED BY METAGENOMICS AND OTHER MOLECULAR TOOLS

This is a collaborative work with Michigan State University and Peking University. Fernanda Paes from Michigan State University was in charge of centrifugation of DNA from cultures in stable isotope probing (SIP) experiments, SIP fraction collection and purification. Ke Yu and Yang Wu from Peking University were in charge of analysis on ethene culture metagenome. Carly Lintner and Tristan Thomas from Mattes' Lab helped build the initial enrichment cultures and culture monitoring. Jim Fish from Alaska Department of Conservation provided the groundwater sample. Kevin Knudtson from the Iowa Institute of Human Genetics provided the metagenomic sequencing methods. All the other works presented in this study was done by Xikun Liu.

5.1 Abstract

Vinyl Chloride (VC), a known human carcinogen, often appears in groundwater as a result of incomplete reductive dechlorination of the higher chlorinated ethenes and ethanes at contaminated sites. Certain indigenous bacteria can adapt to aerobically degrade VC. In this study, microcosms were constructed using groundwater from a contaminated site and gradually adapted to ethene or VC as the sole carbon and energy source. The study showed that the microbial community adapted to VC faster after ethene enrichment. Metagenomics and 16S rRNA gene amplicon Illumina sequencing revealed that *Nocardioides*, *Pseudomonas*, *Polaromonas* and *Pedobacter* were the four most

dominant genera in ethene and VC-fed cultures. Stable-isotope probing demonstrated *Nocardioides* strains were the significant contributors to VC-uptake. Ethene and VC-assimilating *Nocardioides* sp. strain XL1 was isolated from the culture. Three *Nocardioides*-affiliated genome bins were extracted from metagenomes of VC cultures, with 70% to 99% similarity with the genome of VC-assimilating *Nocardioides* sp. JS614. One genome bin containing plasmid DNA was extracted from the VC metagenomes, which showed co-occurrence with two *Nocardioides*-affiliated genome bins, including one identified as from the genome of strain XL1. About 90% of the plasmid contigs could be mapped onto *Nocardioides* sp. strain JS614 plasmid with 100% identity, containing known aerobic ethene and VC degradation pathway genes encoding alkene monooxygenase and epoxyalkane: coenzyme M transferase. Glutathione synthase and glutathione S-transferase genes, possibly involved in epoxide detoxification, were found in *Polaromonas*, *Mesorhizobium* and *Pseudomonas*-affiliated genome bins.

5.2 Introduction

Vinyl chloride (VC) is a known human carcinogen with a US EPA Maximum Contaminant Level of 2 ppb and Maximum Contaminant Level Goal of 0 ppb in drinking water [3]. Direct contamination of groundwater by VC monomer, used in the plastic production industry is possible [99], but VC is more frequently generated during anaerobic reductive dechlorination of the widely used solvents tetrachloroethene (PCE) and trichloroethene (TCE) [5, 8, 100] at contaminated sites. The most commonly observed daughter products of PCE and TCE reductive dechlorination processes are *cis*-dichloroethene (DCE), VC and ethene [6, 101].

Biodegradation of VC can occur either anaerobically or aerobically. *Dehalococcoides spp.* are of the only known genus that can degrade VC through co-metabolism or energy generating VC reduction[66, 67] under strict anaerobic condition, which are applied in bioaugmentation to remediate VC-contaminated groundwater[102]. On the other hand, aerobic degradation of VC can happen at different oxygen levels and in several bacterial genera. Some enrichment cultures derived from aquifer materials from chlorinated ethene-contaminated sites could sustain VC oxidation even at very low oxygen concentrations [25, 103]. A variety of bacteria can aerobically cometabolize VC using methane [104, 105] or ethene [11, 13, 14] as the primary substrate, and even utilize VC as the sole carbon source (VC-assimilating bacteria). Known VC-assimilating bacteria include strains of *Mycobacterium* [10, 11, 71, 72], *Nocardioides* [9, 10], *Pseudomonas* [17, 106], *Ochrobactrum* [17] and *Ralstonia* [73]. Each of these isolates is also able to use ethene as a carbon and energy source. Several *Mycobacteria* and the one *Nocardioides* (*Nocardioides* sp. strain JS614) have been confirmed to use the following pathway for aerobic VC degradation: alkene monooxygenase (AkMO, encoded by genes *etnABCD*) initiates the attack on carbon double bond of ethene or VC to form an epoxide. Subsequently, epoxyalkane: coenzyme M transferase (EaCoMT, encoded by gene *etnE*) conjugates a CoM group to the epoxide [8, 12, 107, 108]. The downstream pathway enzymes are not characterized, but could include CoM reductase/carboxylase, aldehyde/alcohol dehydrogenase, acyl-CoA synthetase, CoA-transferase and a reductive decarboxylase [8], ultimately forming acetyl-CoA which enters central metabolism [109].

Previous studies indicated the genes encoding the ethene and VC pathway locate on a linear megaplasmid (200-300 kb) in some of the VC-assimilating bacteria above,

including *Mycobacterium*[11], *Pseudomonas* and *Ochrbactrum*[17], and conclusive experimental and sequencing evidence was provided for *Nocardioides* sp. strain JS614 [9, 110], which revealed a 307,814 bp plasmid pNOCA01 (Genbank accession no. CP000508.1). Similar plasmids have not yet been studied in mixed culture and environmental samples.

Previous studies of VC-assimilating bacteria have mostly focused on *Mycobacterium sp*. isolates. However, recent works in VC enrichment cultures have revealed the dominance of *Nocardioides* species [64, 65]. Here, we advance our understanding of VC-assimilating bacteria in mixed cultures by tracking microbial community changes in groundwater microcosms as they adapt to VC as the sole carbon and energy source with shot-gun metagenomic sequencing. We are also interested in isolating dominant VC-assimilating bacteria in the enrichment culture for further characterization.

5.3 Materials and Methods

5.3.1 Groundwater and environmental sample collection

The Fairbanks, AK site is contaminated with TCE, trans- and *cis*-1,2-DCE (VC precursors; data obtained via personal communication). At the time of sampling, no VC had been detected at the site and ethene was not measured.

The groundwater sample (~ 1 L) used for constructing enrichment cultures was collected in a sterile Nalgene bottle from monitoring well MW-11M at the Wendell Avenue site in Fairbanks, AK, USA in March 2014. Biomass for DNA extraction was also collected, by passing groundwater (275-3000 mL, duplicate sampling from Well

85R, 9M, 13M, 4M, 12M and 11M) through Sterivex-GP 0.22 μm membrane filter cartridges (Millipore Corporation, Billerica, MA, USA) in the field as described previously [77] until the filter was clogged by sediment. Groundwater and filters were shipped in a cooler on ice packs overnight to the lab and stored at 4 °C and -80 °C respectively until further use and extraction.

5.3.2 DNA extraction from groundwater samples and functional gene qPCR

DNA was extracted from filters using the PowerWater Sterivex DNA Isolation Kit (MO BIO, Carlsbad, CA, USA). The qPCR protocol, including standard curve construction procedures for detecting *etnC* and *etnE* in groundwater and calculation was conducted as previously described [77], using the primer sets RTC_F/R (for *etnC*) and RTE_F/R (for *etnE*) (Table AI.3). Detailed qPCR information (e.g., fluorescent threshold, standard curve y-intercept and PCR efficiency) is provided (Table AI.5) in accordance with MIQE guidelines [60].

5.3.3 Construction of VC- and ethene-fed enrichment cultures

Four groups of enrichment cultures were set up as shown in Figure 5.1. Enrichment cultures were constructed by adding a mixture of groundwater (GW) and autoclaved minimal salts medium (MSM)[79] into sterile serum bottles (160 mL). The trace metal solution (TMS) [79] component of MSM[79] (0.1 mL) was filter sterilized through 0.22μm Millex GV Durapore PVDF membrane units into the bottles. The final GW/MSM mixtures were in ratios of 1:1 (36 mL GW and 36 mL MSM, Group 1) and 1:3 (18 mL GW and 54 mL MSM, Group 2). Either filter sterilized ethene (99%, Airgas, Cedar

Rapids, IA, USA) (450 μmol) or $^{12}C_2$-VC (unlabeled) (99%, Customgas Solutions, Durham, NC, USA) (25-75 μmol) were injected into initial cultures. For the stable isotope probing experiment, $^{13}C_2$-VC (labeled VC) (99%, Cambridge Isotope Laboratories, Tewksbury, MA, USA) was also used. Abiotic controls were constructed identically to enrichment cultures except that no groundwater was added (used 72 mL sterile MSM instead). Subcultures were developed by transferring 5 mL of a culture into 67 mL sterile MSM and prepared as described above. After about 180-190 days, new VC subcultures were developed from ethene enrichment cultures in Groups 1 and 2 (Figure 5.1). The ethene enrichment cultures were maintained and transferred in parallel to the VC enrichment cultures as control groups. Cultures were incubated at both room temperature (21-25 °C) and at 4 °C, kept away from light and inverted (about 60°) on an orbital shaker (200 rpm).

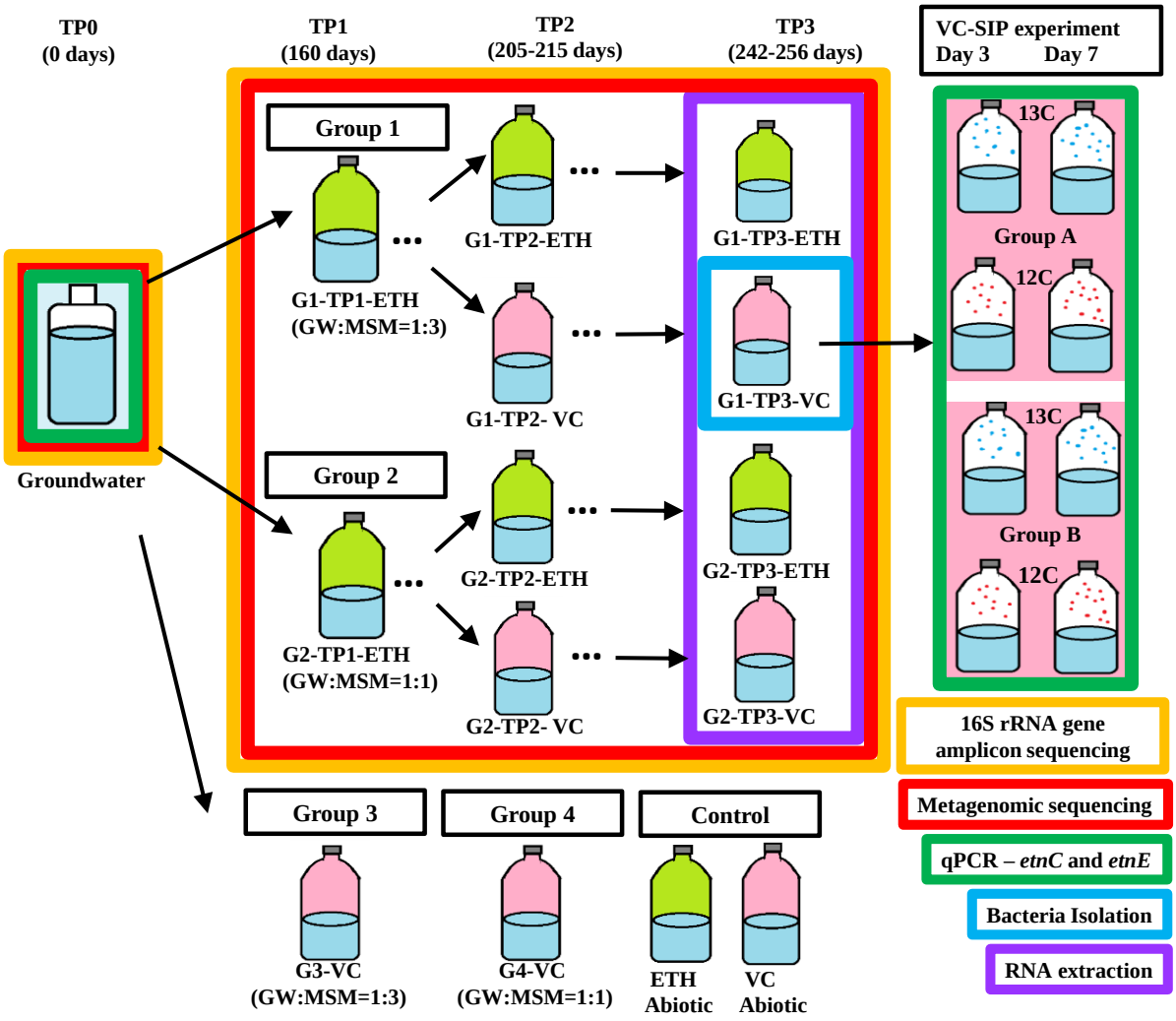


Figure 5.1 Experimental design of this study.

Only cultures incubated at room temperature are shown. The abbreviations are as follows: G1-Group 1, G2-Group 2; TP1-Time Point 1, TP2-Time Point 2; VC-vinyl chloride enrichment culture, ETH-ethene enrichment culture; GW-groundwater, MSM-minimum salts medium. Boxes with different colors illustrate the analyses applied on the cultures.

5.3.4 Analytical methods, DNA and RNA extraction and reverse transcription-qPCR of enrichment cultures

Ethene and VC concentrations in enrichment cultures were monitored via gas chromatography with flame-ionization detection and calculated as described previously using external standards consisting of a series of samples with known ethene or VC concentrations [9, 72]. Optical density at 600 nm (OD_{600}) was measured with a Cary 50 Bio UV-Visible spectrophotometer to monitor bacterial growth. Biomass was collected by sampling 20 mL of liquid culture and passing the sample through Sterivex-GP 0.22 μm membrane filter cartridges (Millipore Corporation, Billerica, MA) and DNA was extracted using the PowerWater Sterivex DNA Isolation Kit (MO BIO, Carlsbad, CA, USA).

RNA was extracted from ethene and VC enrichment cultures (5 mL) at the last time point with the UltraClean Microbial RNA Isolation Kit (MO BIO, Carlsbad, CA, USA). Luciferase mRNA (Promega Corporation, Madison, WI, USA; GenBank accession no. X65316) (1×10^9 gene copies) was added during the extraction to account for the efficiency of reverse transcription. Quantification of luciferase mRNA and construction of luciferase DNA standard curves was conducted as previously described[111]. Triplicate quantification for standards and duplicate quantification for samples were performed.

5.3.5 16S rRNA gene sequencing and data analysis

A total of 11 DNA samples (one pooled MW-M11 groundwater DNA extract from duplicate samples and five samples each from Group 1 and 2 enrichment cultures,

respectively, at three different time points: 160 days, 205-215 days and 242-256 days) (Figure 5.1) were subjected to 16S rRNA gene amplicon sequencing. Sequencing was conducted at Argonne National Laboratory (Argonne, IL, USA) on the Illumina MiSeq platform using 2×151 nt (paired-end) sequencing according to a published protocol [112], using primer set 515f and 806r (Table AI.3) with Illumina flow cell adapter sequences included [80] to target the V4 region of the16S rRNA gene. Replicate sequencing runs of ethene and VC enrichment cultures from the second time point in Group 2 (G2-TP2-VC and G2-TP2-ETH) were performed to check reproducibility.

Mothur 1.36.1 [81] was used to process 1,484,136 raw sequences, using a MiSeq standard operating procedure (SOP) (The Schloss Lab, http://www.mothur.org/wiki/MiSeq_SOP, accessed August, 2015) [113], resulting in 1,156,242 sequences and 30,097 operational taxonomic units (OTUs) based on the 97% similarity cutoff criterion. Briefly, the procedure included contig construction, quality control, alignment with the SILVA database [82, 114], chimera removal, OTU assignment, classification based on Ribosomal Database Project (RDP) classifier [115], construction of rarefaction curves (Figure 5.2) to assess the sequencing quality and calculation of Chao and Shannon diversity indices. Detailed calculation results of number of OTUs, Chao and Shannon indices are provided in Table 5.1 for reference.

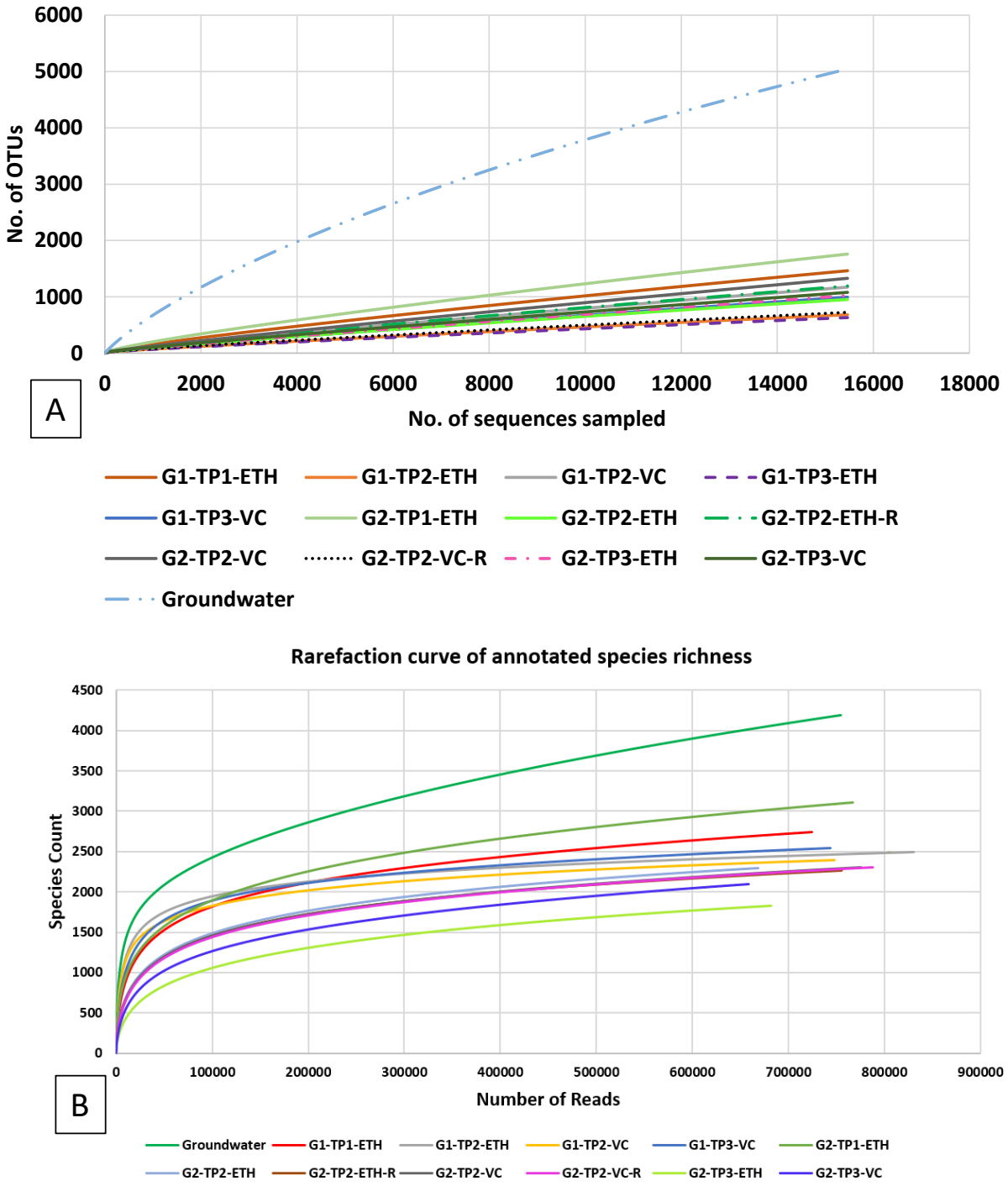


Figure 5.2 Rarefaction curve of: A) 16S rRNA gene amplicon Illumina sequencing based on a 97% OTU cutoff. Data was subsampled with the size of the smallest sample (15,463 sequences); B) metagenomics sequencing.

Note: Species count was calculated based on all the annotated genes in MG-RAST (Meyer et al 2008) and the graph was generated with MG-RAST.

Table 5.1 Summary of 16S rRNA gene amplicon Illumina sequencing data: number of sequences, number of OTUs, Chao index and Shannon index.

Sample	No. of seqs	No. of OTUs				Chao Index				Shannon Index			
		96%	97%	98%	99%	96%	97%	98%	99%	96%	97%	98%	99%
Groundwater	86,493	21119	30097	50397	105330	5565.63	7283.58	10225.46	18711.39	5.81	6.05	6.39	7.00
G1-TP1-ETH	82,175					2588.99	4195.25	8578.98	19263.42	2.15	2.35	2.56	3.26
G1-TP2-ETH	83,302					1739.91	2573.31	5591.34	12186.78	1.88	1.96	2.13	2.69
G1-TP2-VC	96,009					2445.43	4111.97	8316.91	17461.07	2.00	2.12	2.32	2.88
G1-TP3-ETH	15,463					3248.60	5895.03	9267.20	21607.88	2.35	2.49	2.81	3.64
G1-TP3-VC	109,469					2013.93	3289.01	6808.11	13524.46	1.07	1.15	1.29	1.72
G2-TP1-ETH	90,610					2694.32	4042.47	8214.00	19613.61	2.83	3.11	3.38	4.03
G2-TP2-ETH	84,874					1944.66	3893.59	8964.39	17458.25	1.92	2.12	2.37	2.98
G2-TP2-ETH-R	105,965					1861.87	3792.36	8256.01	17247.33	1.76	1.94	2.15	2.69
G2-TP2-VC	117,866					2268.57	3739.48	8256.76	15652.15	1.52	1.70	1.88	2.39
G2-TP2-VC-R	80,545					1799.82	3107.72	7775.65	14338.26	1.68	1.87	2.07	2.62
G2-TP3-ETH	98,786					3312.23	5191.06	10388.41	18243.01	1.75	1.87	2.06	2.56
G2-TP3-VC	104,685					2034.65	3457.93	9464.41	16710.42	1.88	1.98	2.15	2.64

Note: Calculation was based on a subsample of 15,463 sequences (the lowest number of sequences among all samples) from each sample at cutoff= 96%, 97%, 98% and 99% nucleotide similarity. Only 97% was used for further analysis. Other cutoff values are provided here for reference.

5.3.6 Metagenomic sequencing and initial data analysis

The 11 DNA samples described in the previous section were also subjected to shot-gun metagenomic sequencing, with replicate sequencing also done on G2-TP2-VC and G2-TP2-ETH to assess the reproducibility of the sequencing process. About 270 ng of genomic DNA from each sample were sheared using a Covaris E220 (Covaris, Inc., Woburn, MA, USA) to generate fragments with an average length of 600 bp. Indexed sequencing libraries were generated using the KAPA Hyper Prep kit for Illumina sequencing (Cat. no. KK8500, KAPA Biosystems, Inc., Wilmington, MA, USA). The indexed libraries were pooled and fragments in the size range of 450-700 bp were collected using the BluePippin targeted extraction system (Sage Scientific, Inc., Beverly, MA, USA). Libraries were sequenced using the Illumina 2x250 nt (paired-end) SBS v2 chemistry on an Illumina MiSeq (Illumina, Inc., San Diego, CA, USA). The samples were prepared and sequenced at the Iowa Institute of Human Genomics (IIHG) at the University of Iowa.

Meta Genome Rapid Annotation using Subsystem Technology (MG-RAST) v3.6 [116] was used for initial analysis of the metagenomes, which included quality-screening, merging of paired-end data, taxonomic classification of fragments based on the non-redundant database M5NR[117] and calculation of Alpha diversity (Shannon index). All procedures used default settings and standard pipeline on MG-RAST, which returned 3.8 Gbp (about 9.6 $\times 10^6$ paired sequences) of metagenomics data for 13 metagenomes (11 samples and replicate sequencing runs for G2-TP2-VC and G2-TP2-ETH)(Table 5.2). Organism abundance was calculated with the built-in best hit classification method on MG-RAST, which reports the functional and taxanomic annotation of the best hit in the

M5NR for each sequence[118].

Table 5.2 Metagenomes used in this study (post-quality-screening).

Sample Name	Notes	MG-RAST ID.	Sequences (bp)	G+C Content
Groundwater	Original groundwater	4618471.3	235,737,319	51 ±12 %
G1-TP1-ETH	Group 1 time point 1 ethene enrichment culture (before VC adaptation)	4618475.3	227,873,300	62 ±10 %
G1-TP2-ETH	Group 1 time point 2 ethene enrichment culture	4618477.3	260,330,443	59 ±9 %
G1-TP2-VC	Group 1 time point 2 VC enrichment culture	4618476.3	211,910,710	56 ±10 %
G1-TP3-ETH	Group 1 time point 3 ethene enrichment culture	4618481.3	230,618,911	60 ±10 %
G1-TP3-VC	Group 1 time point 3 VC enrichment culture	4618480.3	235,430,397	63 ±8 %
G2-TP1-ETH	Group 2 time point 1 ethene enrichment culture (before VC adaptation)	4618474.3	244,603,305	63 ±9 %
G2-TP2-ETH	Group 2 time point 2 ethene enrichment culture	4618479.3	209,954,096	64 ±11 %
G2-TP2-ETH-R	Group 2 time point 2 ethene enrichment culture (sequencing replicate)	4618473.3	217,228,157	64 ±11 %
G2-TP2-VC	Group 2 time point 2 VC enrichment culture	4618478.3	240,930,808	66 ±10 %
G2-TP2-VC-R	Group 2 time point 2 VC enrichment culture (sequencing replicate)	4618472.3	227,821,242	66 ±10 %
G2-TP3-ETH	Group 2 time point 3 ethene enrichment culture	4618470.3	208,077,808	68 ±8 %
G2-TP3-VC	Group 2 time point 3 VC enrichment culture	4618482.3	200,627,040	68 ±8 %

5.3.7 Assembly, genome binning and annotation of metagenomes

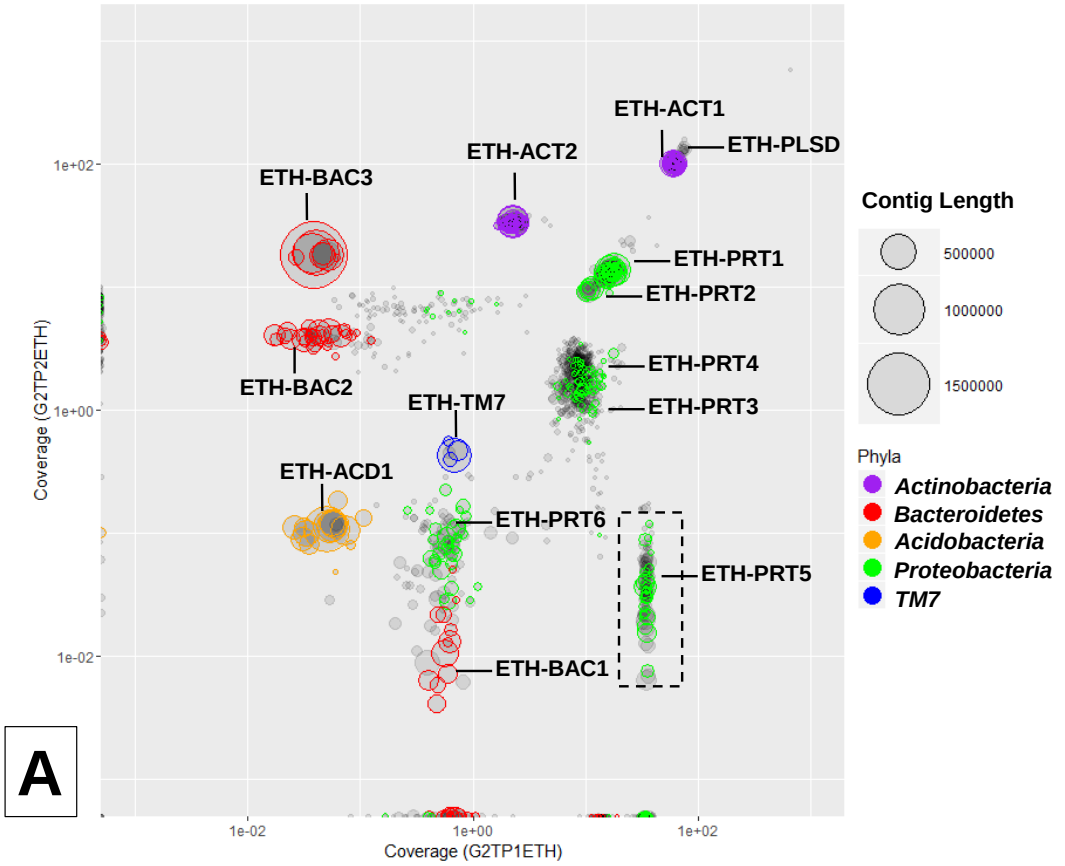
To obtain contigs and genome bins, the metagenomics sequencing data were processed according to the Kalamazoo metagenomic assembly protocol (https://khmer-protocols.readthedocs.io/en/v0.8.3/metagenomics/index.html) and the multi-metagenome protocol [119]. The process included data trimming with Trimmomatic [45], quality filter with khmer [120, 121], contig assembly using IDBA-UD [46], mapping reads back to contigs with bowtie [122], data format conversion with self-written scripts, genome binning using differential coverage, manually checking the consistency in contig coverages and paired-end read tracking [119]. Sequencing data from ethene and VC cultures were assembled separately to distinguish the two enrichment conditions. Sequencing reads (12,282,084 reads in ethene cultures and 8,788,540 reads in VC cultures) were assembled into 29,023 contigs for ethene cultures and 13,912 contigs for VC cultures. All contigs were >600 bp in length. Ethene and VC cultures were binned separately using various combinations of samples to optimize the binning results (e.g. G1-TP1-ETH vs. G2-TP1-ETH, G1-TP2-VC vs. G1-TP3-VC, G2-TP2-VC v.s.G2-TP3-VC, G1-TP2-VC vs. G2-TP2-VC, etc.). An example is provided in Figure 5.3. Finally, 14 genome bins were recovered from the ethene cultures and 11 genome bins were extracted from the VC cultures. Genome completeness and contamination from the binning process was evaluated with CheckM [123] and is reported in Table 5.3. Contigs in extracted genomes were translated with Prodigal [124] and annotated with Ghostkoala [125] based on the Kyoto Encyclopedia of Genes and Genomes (KEGG)[126] and local BLAST against NCBI non-redundant protein database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/). Genus level classifications of the genome bins were inferred by using the taxonomic classification of the top BLAST hits of the ten

longest contigs in the bin. The annotated plasmid comparison graph was generated with BRIG [127] using NCBI BLAST v.2.5.0 for Windows system.

5.3.8 DNA Stable isotope probing (SIP)

Eight subcultures were prepared from the G1-TP3-VC culture around day 240 and split into two groups (four bottles in each group) (Figure 5.1). Two cultures in each group were fed with 100 μmol unlabeled VC per bottle and two cultures were fed with 100 μmol ^{13}C labeled VC per bottle. VC concentrations and OD_{600} were monitored during the experiment. DNA was extracted from cultures at day 3 and day 7 using the method as described above.

The extracted DNA was quantified and ~10 μg was loaded into Quick-Seal polyallomer tubes (13 by 51 mm, 5.1 mL; Beckman Coulter, Indianapolis, IN, USA) along with a Tris-EDTA (pH 8.0)-CsCl solution for ultracentrifugation. Using a digital refractometer (model AR200, Leica Microsystems Inc., Buffalo Grove, IL, USA), the buoyant density (BD) was determined and adjusted to a final reading of 1.73 g mL^{-1} by adding CsCl solution or Tris-EDTA buffer. Tubes were sealed and ultracentrifuged at 178,000 × g (20 °C) for 48 h as described previously [65]. Following ultracentrifugation, the tubes were placed onto a fraction recovery system (Beckman, Bream, CA, USA), and 13 fractions from each centrifuge tube (150 μL for each fraction) were collected. The BD of each fraction was measured, and CsCl was removed by glycogen-assisted ethanol precipitation, giving a final volume of 30 μL. The quantity of DNA obtained in each fraction was determined by Qubit™ (Invitrogen, Carlsbad, CA, USA) with Qubit® dsDNA BR Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA).


ETH-ACT1
ETH-PLSD
ETH-ACT2
ETH-BAC3
ETH-PRT1
ETH-PRT2
ETH-PRT4
ETH-BAC2
ETH-TM7
ETH-PRT3
ETH-ACD1
ETH-PRT6
ETH-PRT5
ETH-BAC1
1e+02
1e+00
1e-02
Coverage (G2TP2ETH)
Coverage (G2TP1ETH)
Contig Length
500000
1000000
1500000
Phyla
Actinobacteria
Bacteroidetes
Acidobacteria
Proteobacteria
TM7
A

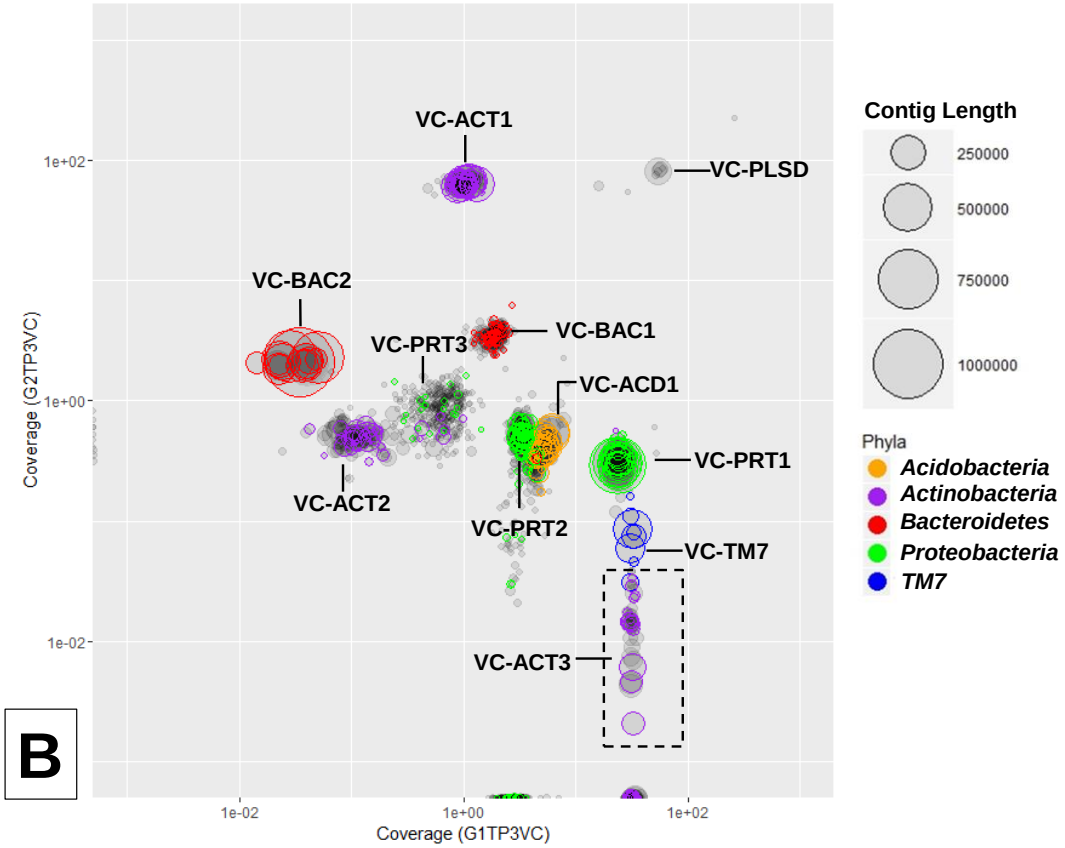

VC-ACT1
VC-PLSD
VC-BAC2
VC-PRT3
VC-BAC1
VC-ACD1
VC-PRT1
VC-ACT2
VC-PRT2
VC-TM7
VC-ACT3
1e+02
1e+00
1e-02
Coverage (G2TP3VC)
Coverage (G1TP3VC)
Contig Length
250000
500000
750000
1000000
Phyla
Acidobacteria
Actinobacteria
Bacteroidetes
Proteobacteria
TM7
B

Figure 5.3 Example of genome binning using differential coverage method[119] for A) ethene cultures and B) VC cultures.

Note: Contigs clustering together represent putative population genomes and were extracted as initial bins. As multiple species could be present in the same coverage-defined subset (e.g. ETH-PRT3 and ETH-PRT4), the binned genome was further refined by principle component analysis of tetranucleotide frequencies as described in the protocol[119]. Circle size is proportional to the contig length. Coloring is based on phyla level classification. The grey clusters on top-right corner of both graphs were also extracted, which were later noted containing *Nocardioides* sp. strain JS614 plasmid sequences.

Table 5.3 Summary of genome bins from ethene and VC culture. Completeness and contamination were calculated with CheckM [123].

Genome	Top BLAST hit inferred genus	Size(MB)	GC (%)	KEGG Annotated (%)	Number of contigs	Longest contig(bp)	Mean contig length(bp)	Number of predicted genes	Completeness (%)	Contamination (%)
Ethene cultures										
ETH-ACD1	*Chloracidobacterium*	4.1	54.13	38.7	17	858,493	242,124	1,425	88.95	1.81
ETH-ACT1	*Nocardioides*	4.8	71.86	41.2	73	312,513	65,816	1,932	99.16	2.52
ETH-ACT2	*Nocardioides*	4.2	71.01	40.8	46	433,173	90,239	1,665	99.16	0.84
ETH-BAC1	*Niastella*	3.3	45.06	42.3	48	314,892	77,818	1,275	99.23	0.00
ETH-BAC2	*Niastella*	3.3	44.84	41.2	40	253,891	82,314	1,197	92.31	0.00
ETH-BAC3	*Pedobacter*	4.4	37.43	37.0	7	1,862,975	630,555	1,405	98.97	0.51
ETH-PRT1	*Mesorhizobium*	6.9	62.76	45.1	119	466,295	57,598	3,042	100	0.00
ETH-PRT2	*Methylobacterium*	1.0	63.01	14.4	7	229,546	149,623	169	10.92	0.00
ETH-PRT3	*Polaromonas*	4.4	63.38	40.8	290	99,343	15,150	1,665	80.64	2.52
ETH-PRT4	*Bradyrhizobium*	4.2	63.91	41.6	499	29,984	8,508	1,816	43.56	0.08
ETH-PRT5	*Dokdonella*	4.1	66.22	41.6	70	209,138	58,099	1,357	67.23	0.00
ETH- PRT6	*Polaromonas*	4.0	61.04	55.0	99	154,317	40,575	2,164	95.8	0.00
ETH-TM7	*TM7*	0.8	49.44	41.5	4	478,860	199,196	357	67.99	0.00
ETH-PLSD	*Nocardioides* plasmid	0.3	68.21	10.7	11	89,694	25,822	290	92.12*	0.00

*compared with *Nocardioides* sp. strain JS614 plasmid pNOCA01 (Genbank accession no. CP000508.1) [110]

Table 5.3 Continued:

VC cultures										
Genome	Top BLAST hit inferred genus	Size(MB)	GC (%)	KEGG Annotated (%)	Number of contig	Longest contig(bp)	Mean contig length(bp)	Number of predicted genes	Completeness (%)	Contamination (%)
VC-ACD1	genus unknown	4.1	54.17	38.5	48	391,754	84,319	3,630	94.87	1.71
VC-ACT1	*Nocardioides*	4.7	70.72	39.8	142	182072	32921	4,669	97.08	1.17
VC-ACT2	*Nocardioides*	4.7	70.72	39.8	137	182072	34,034	4,662	96.05	1.12
VC-ACT3	*Nocardioides*	3.6	71.01	41.5	133	162192	26,813	3,447	88.25	1.12
VC-BAC1	*Pedobacter*	2.8	44.64	41.4	215	51,033	12,959	2,629	76.42	0.03
VC-BAC2	*Pedobacter*	4.8	37.74	35.7	18	1,041,262	265,658	4,113	97.61	1.20
VC-PRT1	*Pseudomonas*	6.3	64.60	50.7	61	713,109	103,827	5,822	100	1.62
VC-PRT2	*Mesorhizobium*	7.5	62.73	41.3	242	200,473	30,930	7,555	98.21	8.18
VC-PRT3	*Bradyrhizobium*	2.2	63.06	44.1	323	14,791	6,840	2,297	25.86	3.45
VC-TM7	*TM7*	0.8	49.43	41.5	10	331,633	84,674	895	67.75	0.00
VC-PLSD	*Nocardioides* plasmid	0.3	68.12	12.7	20	151,553	15,526	313	*97.68	0.00

*compared with *Nocardioides* sp. strain JS614 plasmid pNOCA01 (Genbank accession no. CP000508.1) [110]

5.3.9 Primer design, verification and qPCR on SIP fractions

The qPCR primer sets (Table AI.3) NocF/R(this study), Pse435F/686R [128], PedF/R [129] and SedF/R[64] were used to quantify 16S rRNA genes from *Nocardioides spp.*, *Pseudomonas spp.*, *Pedobacter spp.* and *Sediminibacterium spp*., respectively, in the SIP fractions. To target *Nocardioides* strains in the enrichment cultures, NocF/R primers were designed with Primer-BLAST [130](http://www.ncbi.nlm.nih.gov/tools/primer-blast/) using the V4 region of the most abundant 16S rRNA gene sequence classified as *Nocardioides* in the 16S rRNA gene Illumina sequencing dataset as the template. Default settings were used except that the primer specificity stringency was set to four total mismatches and four mismatches within the last seven base pairs at the 3′ end[77]. The specificity of these four primer sets was tested by cloning and sequencing amplicons from G1-TP3-VC (the same culture of group A day 3 in SIP experiment), using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA, USA) and the cloning procedures described previously [64]. All the sequenced clones (six to ten clones for each primer set) were the desired targeting genus based on RDP classifier[131] (Table AI.4), except *Sediminibacterium*. However, BLAST against NCBI Genbank showed the top blast hits (100% coverage and 98% nucleotide identity) of these clones were all uncultured *Sediminibacterium* (Genbank Accession No. JX512347).

All SIP fractions (two groups at day 3 and 7, 13 fractions each sample, a total of 52 fractions) were subject to duplicate qPCR analysis on the same 384-well plate for each primer set using an ABI 7900 Sequence Detection System (Applied Biosystems, Foster city, CA, USA). Each 20 µL reaction mixture contained 12 µL of Power SYBR Green PCR Master Mix solution (Applied Biosystems), 0.4µL of each primer (10µM) and 2 µL

of DNA template (0.3-6.7 ng/L). The thermocycler protocol consisted of an initial denaturation (95°C, 10 min), 40 cycles of amplification (95°C, 15 s; 60°C, 1 min) followed by an amplicon melting program from 55°C to 95°C. Plasmids (3,931 bp plus product size) containing PCR products of each primer set were used as quantification standards in triplicates. Gene abundance was determined as previously described [77].

5.3.10 Isolation and Identification of ethene and VC degraders

Pure cultures were isolated from ethene and VC enrichment cultures (Group 1) by spreading 20 μL dilutions (10^{-1} to 10^{-5}, 10^{-4} and 10^{-5} worked best) onto MSM-noble agar plates and incubating in a desiccator containing an ethene atmosphere (1-10% (vol/vol)) (ethene incubator). After incubation for about one month, colonies were selected and streaked on duplicate MSM-noble agar plates with one plate incubated in the ethene atmosphere and the other plate incubated in ambient air. Isolates showing significantly more growth in the ethene atmosphere compared to ambient air were investigated further. At least three streaks were performed before conducting purity confirmation experiments. Lines of evidence for culture purity included: 1) single colony morphology type on MSM and 1/10 strength trypticase soy agar-glucose (TSAG) plates, 2) single cell morphology type according to light microscopy with a 100× oil immersion objective and 3) distinct single peaks in Sanger sequencing electropherograms of 16S rRNA gene whole-cell PCR products from the plate. The purity of VC-assimilating isolate XL1 was further confirmed by Sanger sequencing the RNA polymerase beta subunit gene (*rpoB)*, *etnC*, *etnE* and *etnE1* PCR products amplified from liquid culture DNA extracts. The 16S rRNA gene, *rpoB*, *etnC*, *etnE* and *etnE1* genes were amplified with the 27F and 1492R primer set [132],

Nocardioides-specific primer set *rpoB*_F/R [36, 133], JS614etnC_F/R[77], CoMF3/R3[9] and CoMF1L/R2E primer sets [9, 11](Table AI.3) respectively. Utilization of ethene and VC was confirmed by reinoculation of the isolates into MSM-ethene (about 450 μmol ethene per bottle=15.6 mg/L) and MSM-VC (about 100 μmol VC per bottle=41.5 mg/L) cultures VC consumption and OD_{600} were monitored during the experiment. The nearly-complete 16S rRNA gene (1489 bp) was cloned and sequenced from isolates. Taxonomic classification was determined by RDP classifier and BLAST against NCBI Genbank. Routine spread plating on TSAG was conducted to check purity. In some cases, bacteria were isolated on R2A agar plate, which is described in the Result section.

5.3.11 Sequence accession

Metagenomic sequences that passed the quality control pipeline on MG-RAST were deposited under the accession numbers 4618470.3 to 4618482.3 as listed in Table 5.2. The Illumina amplicon sequencing of 16S rRNA genes were deposited in NCBI Sequence Read Archive under bioproject number PRJNA350263. The partial 16S rRNA gene, *rpoB*, *etnC*, *etnE* and *etnE1* gene of isolate *Nocardioides* sp. strain XL1 were deposited in NCBI Genbank under accession no. KY029028, KY042025, KY042023, KY094130 and KY042024.

5.4 Results

5.4.1 Initial characterization of the groundwater sample

Best hit analysis on MG-RAST of MW-11M groundwater metagenomic reads showed a very small number of archaeal (3.5%), eukaryotic (1.3%) and virus (0.5%) sequences, revealing the majority of the community were bacteria, in which *Proteobacteria* (35.9%), *Bacteroidetes* (18.5%), *Firmicutes* (11.0%), *Actinobacteria* (10.7%) dominated (Table AI.6). The microbial community structure as revealed by 16S rRNA gene amplicon sequencing was consistent with the metagenomics data (Table AI.6 and Figure 5.8).

Although VC has not been detected at the Wendell Ave. site, qPCR data (this study and [98]) revealed that both *etnC* and *etnE* present (10^3~10^6 genes/L of groundwater) in groundwater samples from the six wells sampled, including well MW-11M used for constructing enrichment culture (Figure 5.4). This indicates that the groundwater microbial community holds the metabolic potential for aerobic VC and ethene-oxidation.

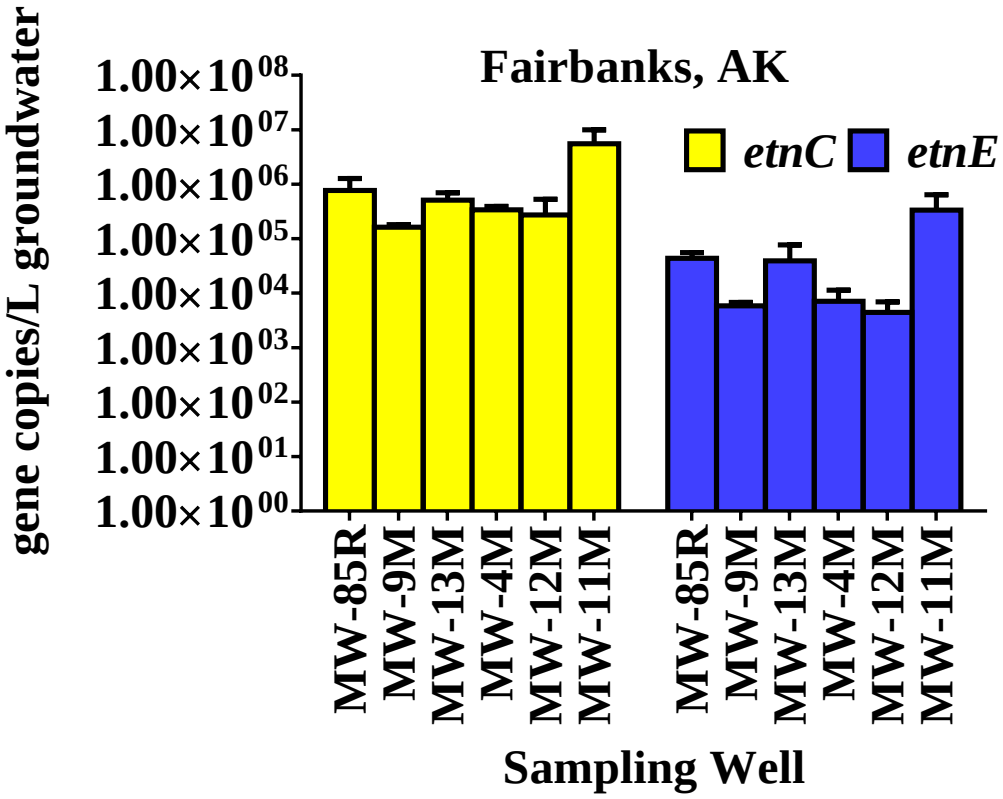


Figure 5.4 Functional gene *etnC* and *etnE* abundances in groundwater at the Fairbanks, AK site.

Note: Groundwater and DNA samples from Well MW-11M were used for enrichment culture construction and groundwater metagenomic sequencing in this study. Samples were presented as from the center of the plume (left, 85R) to the edge (right, 11M). The abundance of *etnE* gene in Well 4M and 13M has been reported previously[98].

5.4.2 Faster adaptation to VC after ethene enrichment

The presence of ethene- and VC-oxidizing functional genes in Fairbanks, AK GW suggests that viable ethene- and VC-oxidizing bacteria were present. We tested this hypothesis by preparing groups of ethene- and VC-fed microcosms (Fig. 5.1). Initial onset of ethene degradation in the ethene-fed microcosms was observed around day 50 from both Group 1 and 2 (Figure 5.5 A and C). After about 450 µmol of ethene was consumed, cell growth was observed in both microcosms (OD_{600} from 0.04 to 0.28 and 0.07 respectively, with both cultures displaying visible floc formation. However, no evidence of growth was noted in VC-fed microcosms (Group 3 and 4) (Figure 5.6 A and B). Microcosms incubated at 4 °C showed no growth on VC or ethene (Figure 5.7).

Ethene consumption in microcosms began to lag after the second spike of ethene (~450 µmol), probably because of oxygen depletion (data not collected). After 160 days, the ethene cultures were transferred into fresh aerobic media. Ethene consumption and cell growth resumed after another 20 days (around day 180 from the initial microcosm construction, Figure 5.5 A and C). Between days 180-190, new VC enrichment cultures (Figure 5.5 B and D) were constructed from the ethene enrichment cultures of Group 1 and 2. These VC enrichment cultures began degrading VC without an appreciable lag period, consuming about 100 µmol of VC in 5 days (Group 1, Figure 5.5 B) and 68 µmol of VC in 3 days (Group 2, Figure 5.5 D) respectively. Cell growth (OD_{600} =0.03±0.01 to 0.19±0.07) was also observed in both cultures.

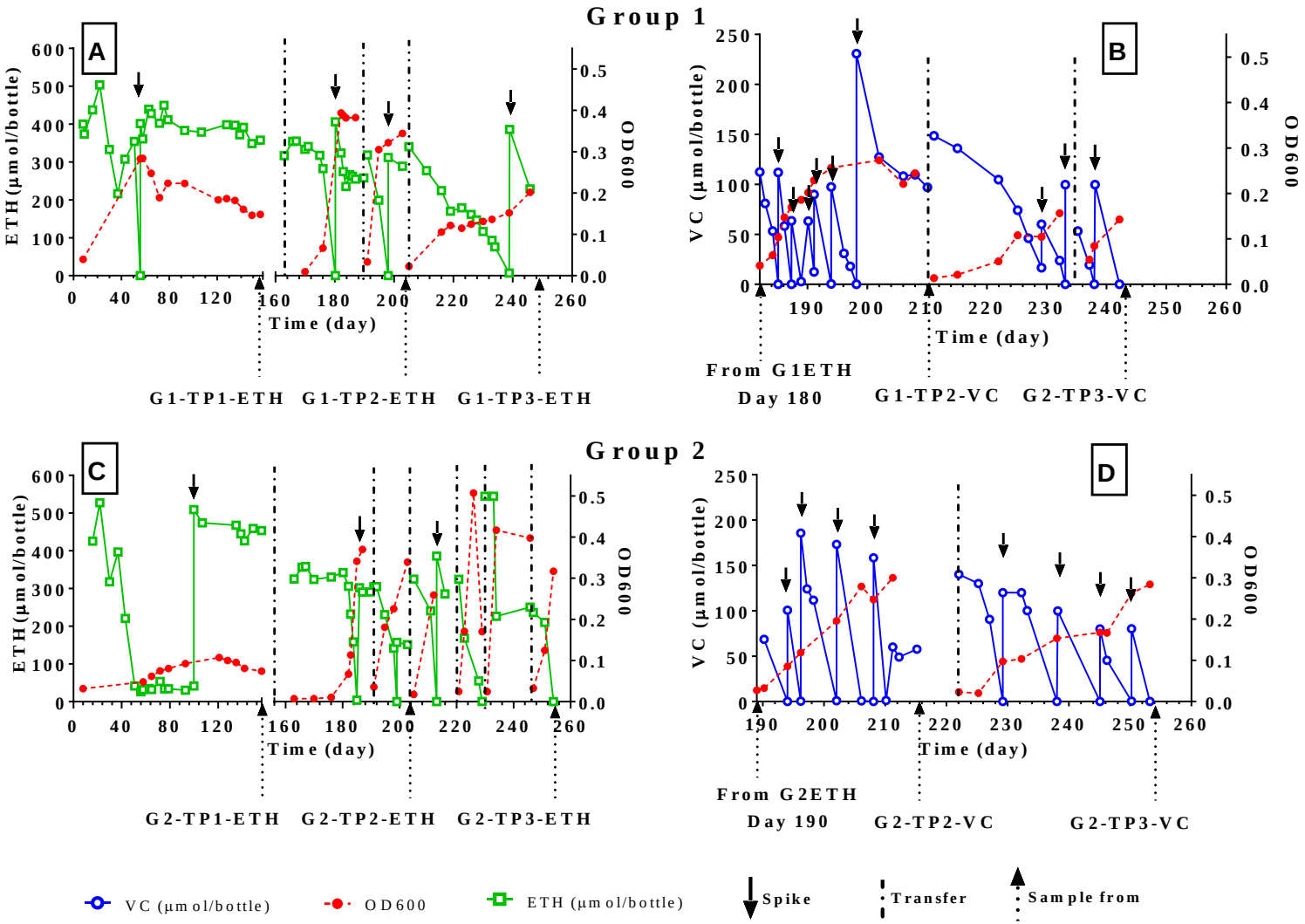


Figure 5.5 Consumption of ethene and VC and growth of bacteria (as estimated with OD600 measurements) in enrichment cultures: A) Group 1 ethene cultures; B) Group 1 VC cultures; C) Group 2 ethene cultures; D) Group 2 VC cultures.

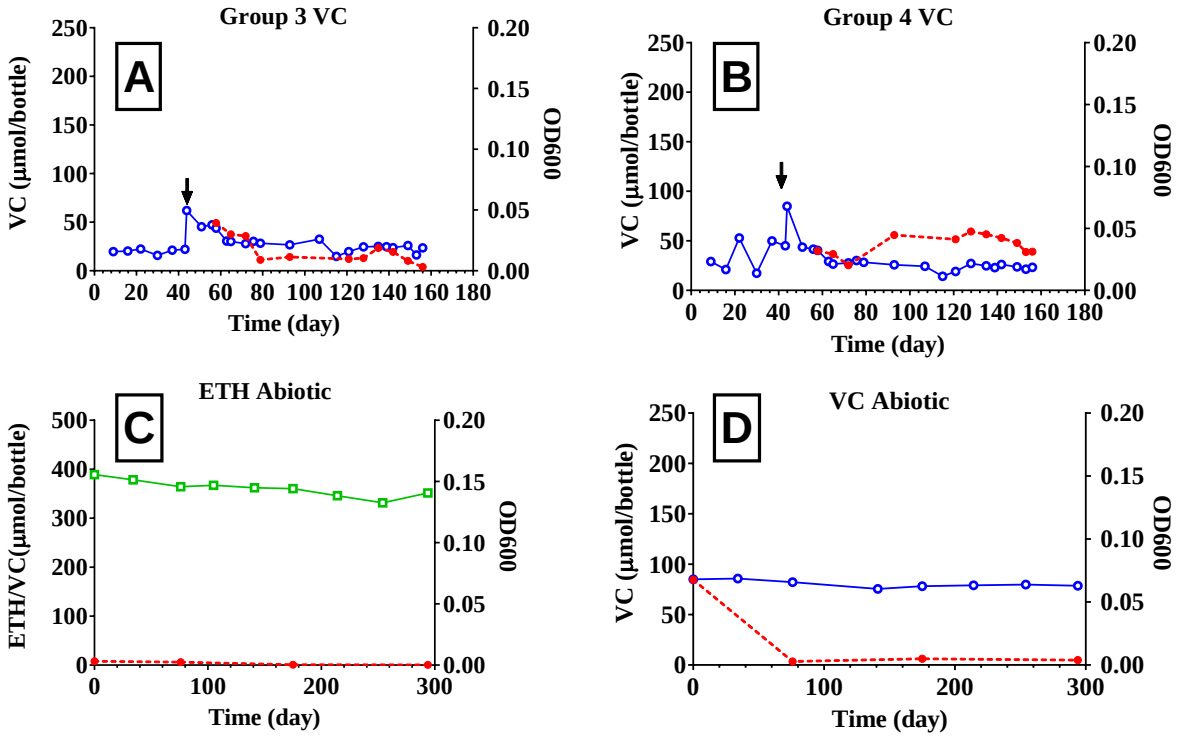


Figure 5.6 VC and ethene (ETH) degradation and bacteria growth in: A) Group 3 VC cultures (GW:MSM=1:1); B) Group 4 VC cultures (GW:MSM=1:3); C) ETH abiotic control; D) VC Abiotic control.

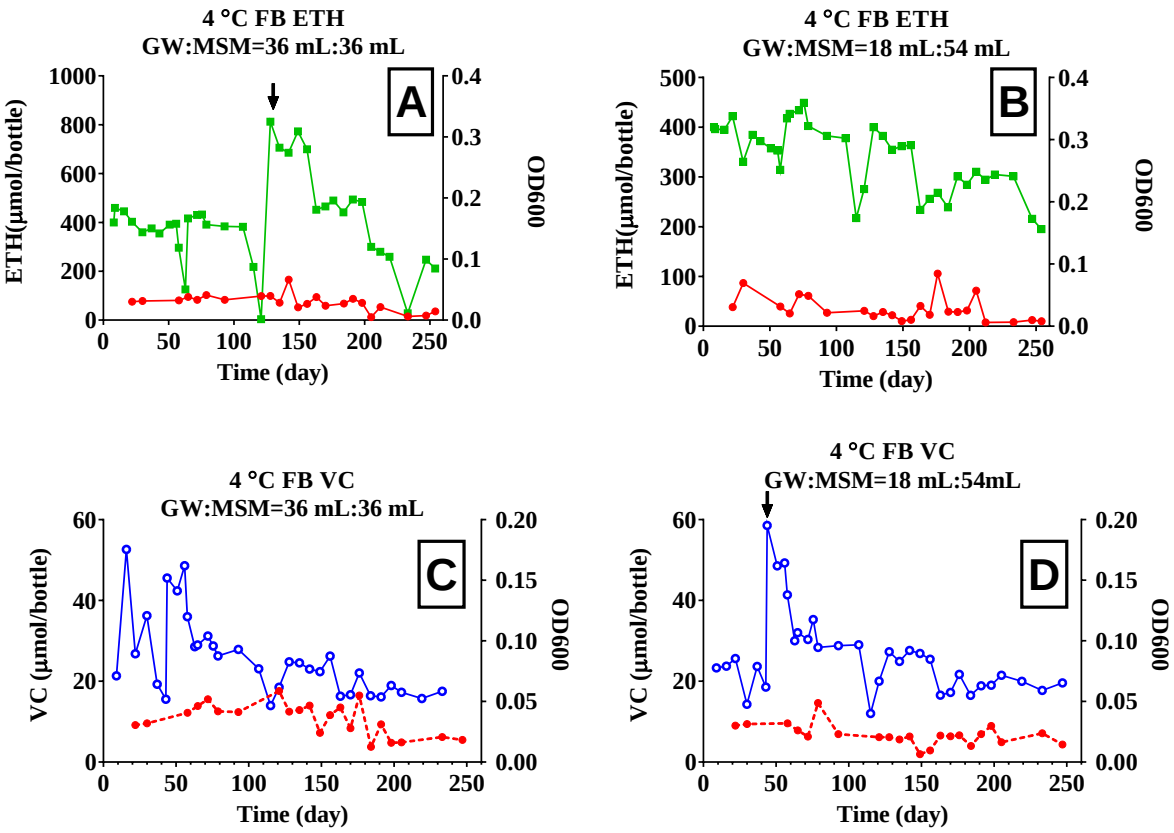


Figure 5.7 Consumption of ethene and VC and growth of bacteria (as estimated with OD600 measurements) in enrichment cultures at 4 °C. A) Ethene (ETH) culture, GW:MSM=1:1; B)ETH culture, GW:MSM=1:3; C)VC culture, GW:MSM=1:1; D)VC culture, GW:MSM=1:3.

5.4.3 Shifts in microbial community taxonomic structure after ethene and VC enrichment

Changes in microbial community taxonomic composition in response to ethene and VC enrichment were assessed by metagenomic and 16S rRNA gene amplicon Illumina sequencing (Figure 5.8, Table AI.5). Replicate sequencing runs with both sequencing methods showed good reproducibility (Figure 5.9). After 160 days, both experimental groups showed dramatic decrease in biological diversity. Alpha diversity (Shannon index based on metagenomic data) decreased from 780.61 (groundwater) to as low as 3.75 (G2-TP3-ETH). Both communities in Group 1 and 2 had few *Archaea*, *Eukaryota* and *Virus* sequences (<0.5%). *Actinobacteria* (46.4 and 25.2% in G1-TP1-ETH and G2-TP1-ETH, respectively) and *Proteobacteria* (37.0% and 60.6% in G1-TP1-ETH and G2-TP1-ETH, respectively) were more abundant after ethene enrichment in comparison to the groundwater metagenome. As enrichment on ethene and VC continued in two parallel experimental groups, *Actinobacteria (*47.4% and 35.4% in G1-TP3-ETH and G1-TP3-VC, respectively) and *Proteobacteria (*30.3% and 55.9% in G1-TP3-ETH and G1-TP3-VC, respectively) remained the most abundant phyla in Group 1 at day 242-256, whereas in Group 2, *Actinobacteria* (91.1% and 85.7% in G2-TP3-ETH and G2-TP3-VC, respectively) were the most dominant. The most abundant bacterial genus across all samples was *Nocardioides* (phylum *Actinobacteria*) (18.1-84.2% among nine cultures from two experimental groups). The exception was the VC culture at day 242-256 in Group 1(G1-TP3-VC), in which *Pseudomonas* (phylum *Proteobacteria*) (29.0%) was the most abundant genus.

The relative abundance of microbial community members as determined with 16S rRNA gene amplicon sequencing was consistent with the metagenomics data at the

Phylum level (Figure 5.8). *Nocardioides* was the dominant genus (3.4-35.1% in 9 cultures from 2 experimental groups). VC culture at day 242-256 in Group 1 (G1-TP3-VC) was dominant with *Pseudomonas* (71.0%). Different from metagenomics data, the 16S rRNA gene sequencing data showed in Group 2, *Pedobacter* (17.5-47.5% in all cultures from Group 2) and *Sediminibacterium* (24.9-31.3% in cultures from Group 2 timepoint 3) were also dominant in the community.

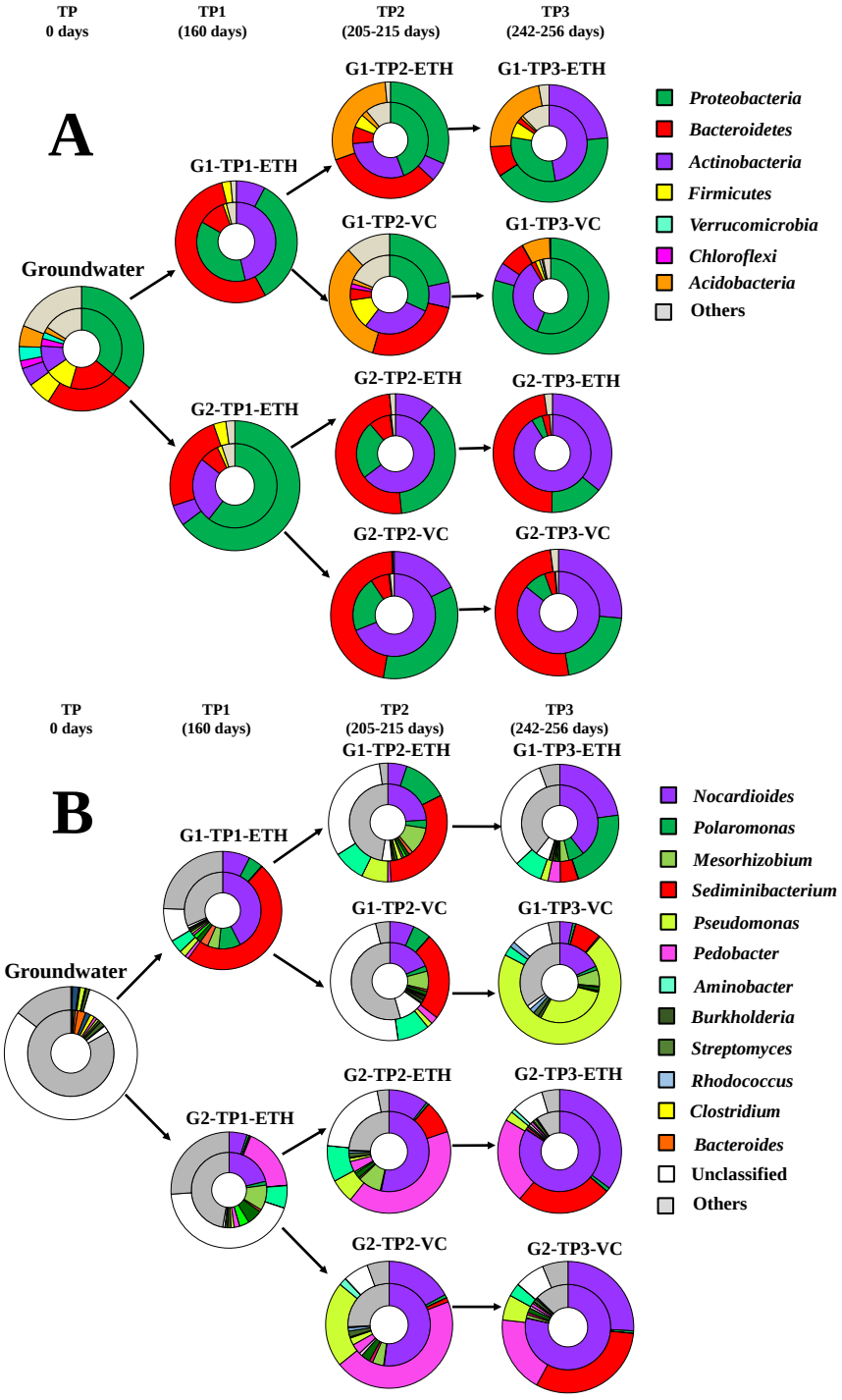


Figure 5.8 Phylogenetic composition of microbial communities on A) phylum and B) genus level, based on the MG-RAST best hit analysis of shot-gun metagenomics sequencing data (inner circle) and Mothur analysis of 16S rRNA gene Illumina sequencing data (outer circle). "Others" represents the combination of the remaining minor taxa.

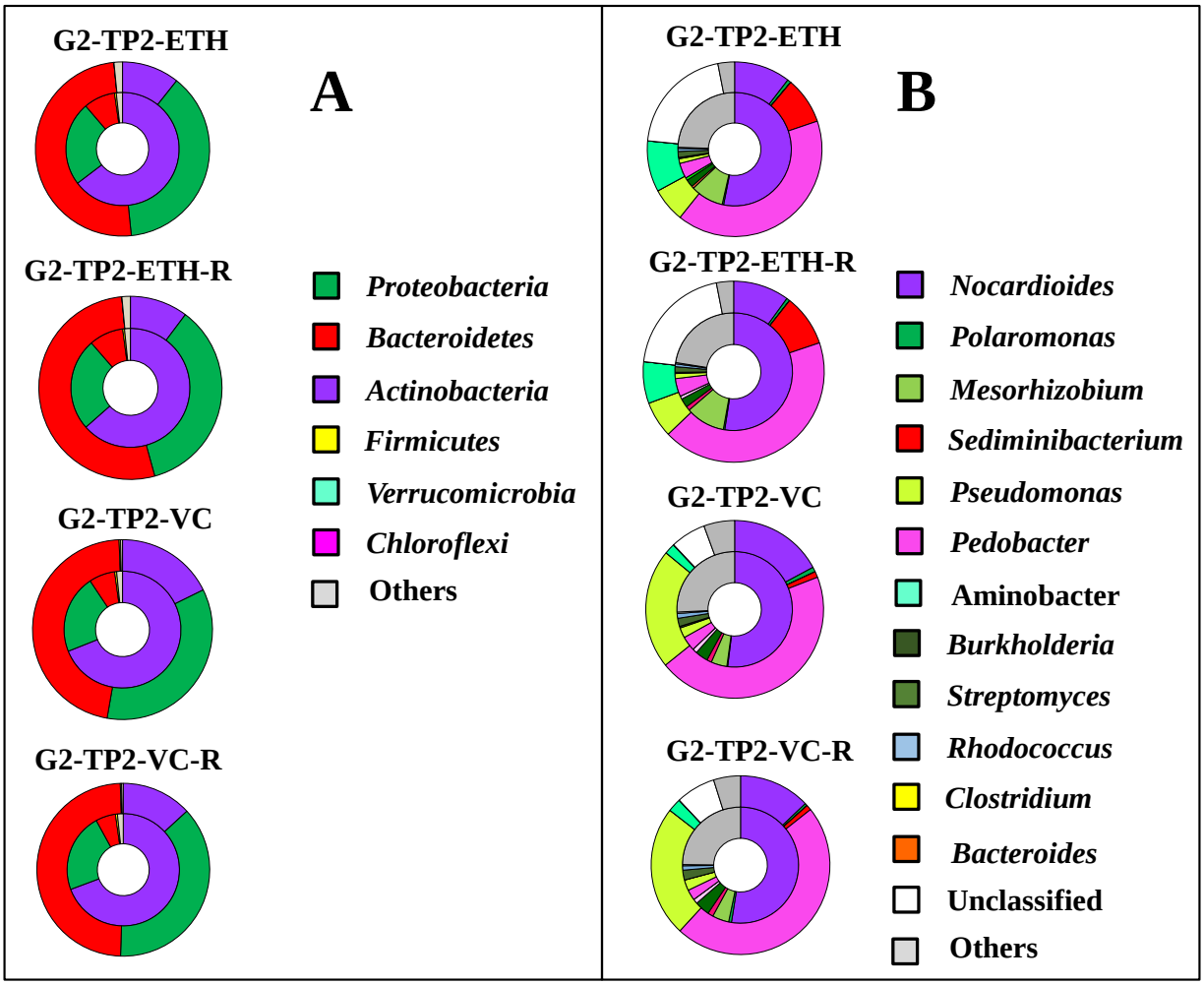


Figure 5.9 Replicates of metagenomics sequencing and 16S rRNA gene amplicon Illumina sequencing of group 2 time point 2 ethene and VC cultures (G2-TP2-ETH, G2-TP2-VC) showed the results are reproducible.

Note: R represents the replicate. Panel A: classified to phylum; Panel B: classified to genus.

5.4.4 SIP reveals primary bacteria involved in carbon uptake from VC

We hypothesized that the bacterial groups gaining dominance in the VC enrichment cultures were utilizing VC as a carbon and energy source. The qPCR of 16S rRNA genes on SIP fractions enabled us to test this hypothesis. Maximum 16S rRNA gene copy numbers of *Nocardioides sp.*, *Pedobacter sp.* and *Sediminibacterium* sp. in 13C-VC-amended cultures generally appeared in higher BD fractions compared to the unlabeled VC control group, indicating that these three incorporated carbon from VC into their genomic DNA (Figure 5.10 and Appendix Figure AI.1). In particular, DNA from the *Nocardioides* strain shifted more into the higher BD fractions in comparison to the controls. For example, in Group B at day 7, the highest peak shifted from ~1.73 g mL^{-1} to ~1.76 g mL^{-1}. In contrast, *Pseudomonas* did not show significant shifts towards heavier BD fractions in the labeled VC amended culture in comparison to the unlabeled VC amended controls, except only in Group A time at day 3 (Figure AI.1), from 1.72 g mL^{-1} to 1.74 g mL^{-1}.

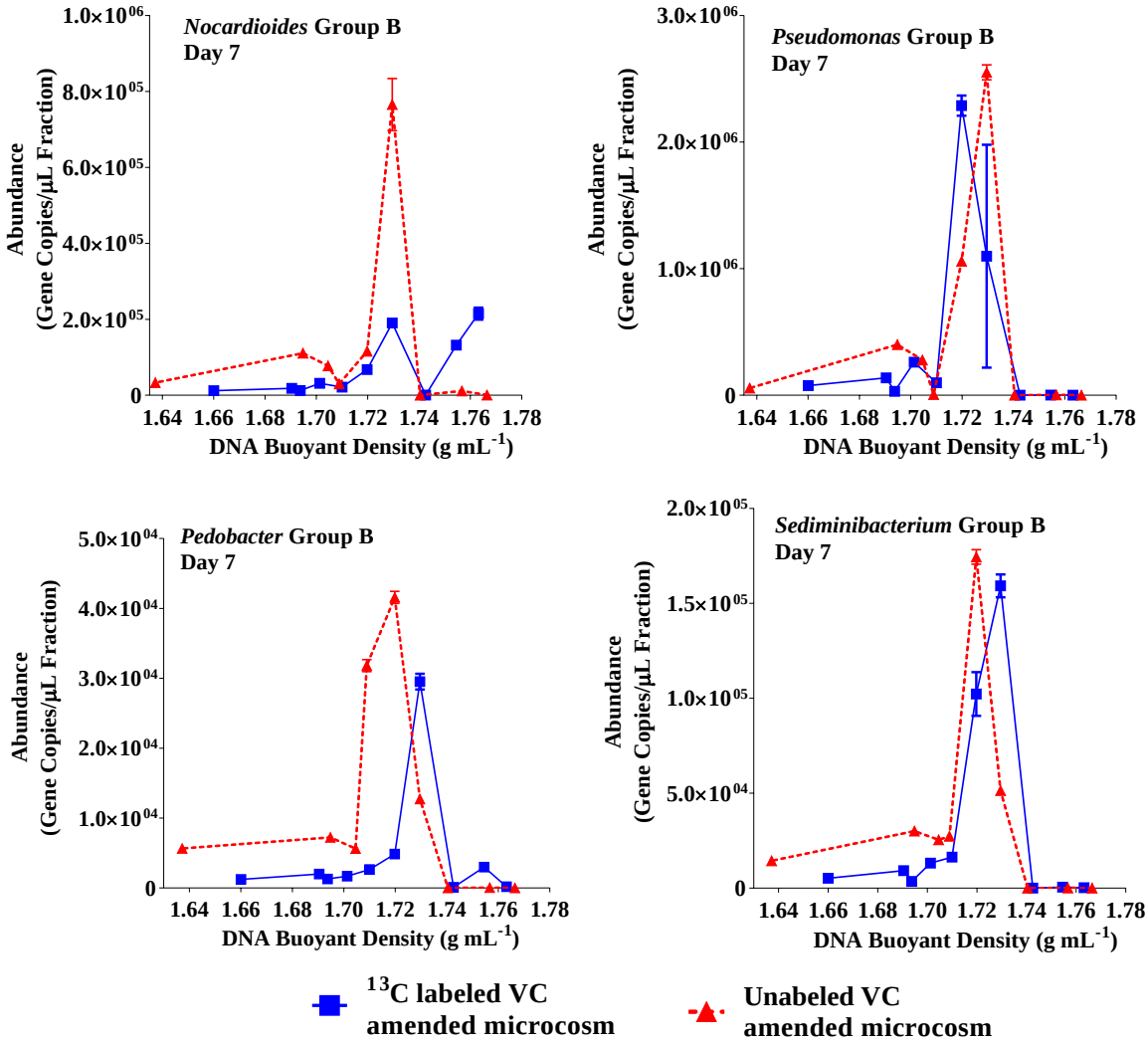


Figure 5.10 Example of SIP-qPCR data: 16S rRNA gene qPCR targeting genera *Nocardioides*, *Pseudomonas*, *Pedobacter* and *Sediminibacterium* on SIP fractions from Group B at day 7. A full list of graphs is provided in Appendix Figure AI.1.

5.4.5 Isolation and characterization of bacteria from enrichment cultures

As *Nocardioides* was the dominant bacterial genus in the VC enrichment cultures and showed carbon uptake from VC, we hypothesized that at least one strain of *Nocardioides* VC-assimilating bacteria was present in the culture and attempted to isolate the bacteria. One VC-assimilating strain (named XL1) was successfully isolated from the culture G1-TP3-VC. This isolate demonstrated growth on both VC and ethene as the sole carbon source (Figure 5.11). Sequencing of the XL1 16S rRNA gene (1489 bp) showed that it is 99.7% (3 bp different) identical to the VC-assimilating *Nocardioides* sp. JS614. Sequencing of *rpoB*, *etnC*, *etnE* and *etnE1* genes of strain XL1 showed they are all 100% identical to *Nocardioides* sp. JS614.

Strain XL1 was relatively slow-growing, requiring about one month to form colonies (~1 mm in diameter) on MSM agar plates in the ethene incubator and two to three weeks to form colonies (~1 mm in diameter) on 1/10 TSAG plates in ambient air (Figure 5.12). On MSM agar plates, strain XL1 colonies were small, clear, and round-shaped with a smooth surface. On 1/10 TSAG plates, strain XL1 formed creamy colonies, with small colonies scattered around large ones. In liquid MSM culture, strain XL1 showed hydrophobic behavior-clumping and adhering to the septum of the serum bottles (Figure 5.13), which is very different from its close relative *Nocardioides* sp. JS614 that forms homogenous liquid culture in MSM. Strain XL1 appeared to be rod-shaped under optical microscopy (100× oil immersion objective) with staining.

In the ethene incubator, one strain similar to *Mesorhizobium* sp. DR 6-01 (100% nucleotide identity in partial 16S rRNA gene (740 bp), sequencing of PCR products) was also isolated on MSM agar plate. Another strain was isolated on R2A plates, which

showed similarity with *Pedobacter* (98% nucleotide identity in partial 16S rRNA gene (780 bp), sequencing of PCR products). However, both strains showed no growth in liquid culture using ethene or VC as the sole carbon source. *Pseudomonas* was not found in any isolation attempts with MSM agar plates in the ethene incubator, R2A and 1/10 TSAG plates in ambient air.

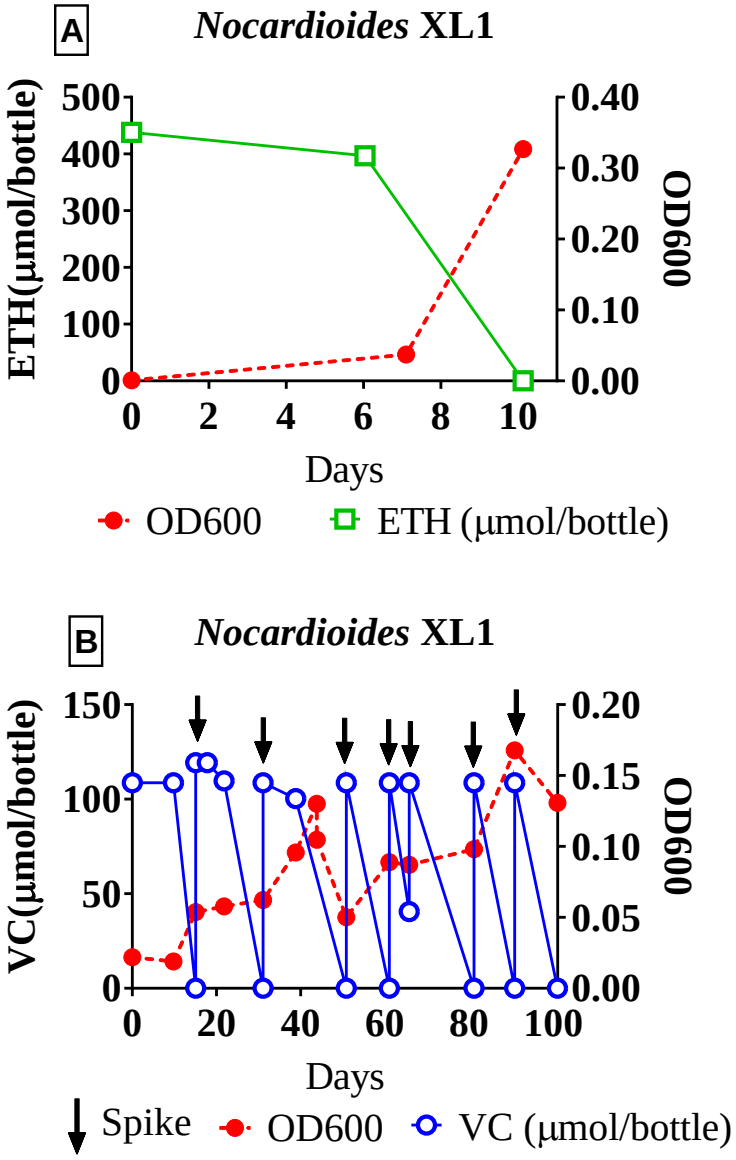


Figure 5.11 The isolate *Nocardioides* sp. strain XL1 showed growth and degradation of ethene and VC.

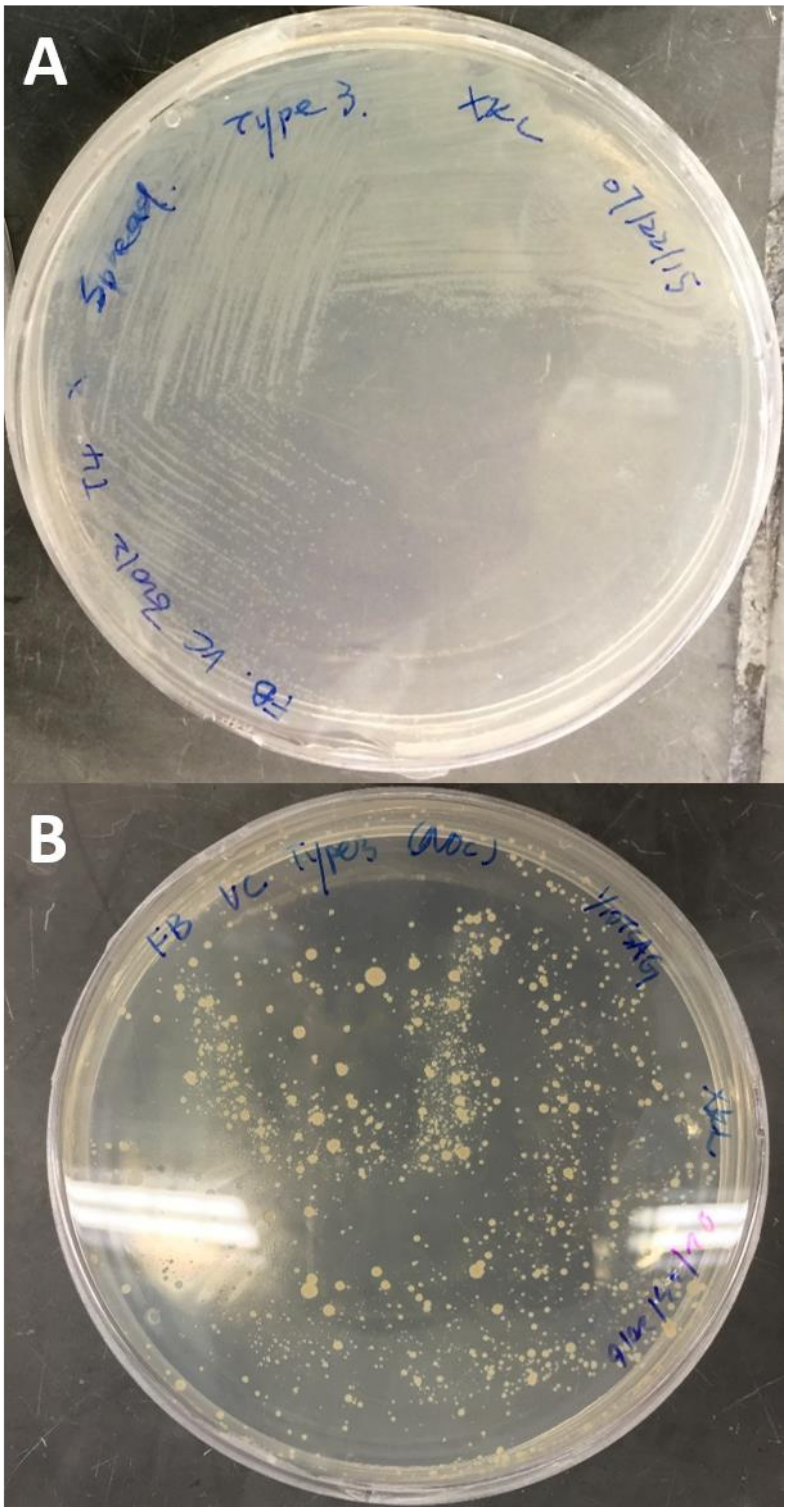


Figure 5.12 Morphology of *Nocardioides* sp. strain XL1 on A) MSM plate in ethene environment and B) 1/TSAG plate in ambient air.

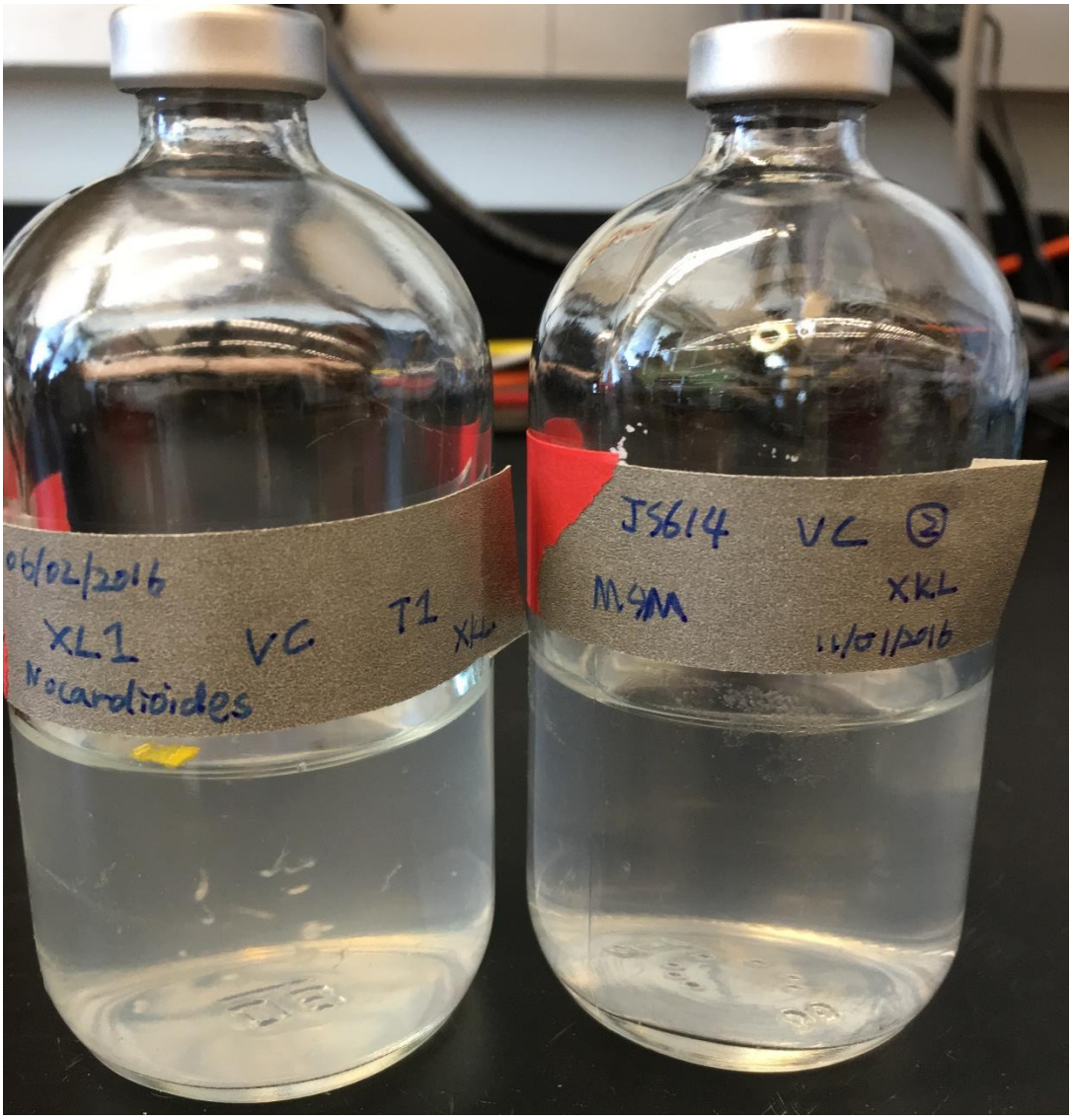


Figure 5.13 Morphology of *Nocardioides* sp. strain XL1 in VC-MSM culture (right), which contained clumping flocs in both ethene and VC cultures. The liquid culture of *Nocardioides* sp. strain JS614 in VC-MSM culture (left) is provided for comparison, which remained homogenous in both ethene and VC cultures.

5.4.6 Genome binning revealed multiple Nocardioides strains and one plasmid

Among the binned genomes from ethene and VC cultures metagenomes (Figure 5.3), five genomes (ETH-ACT1, ETH-ACT2, VC-ACT1, VC-ACT2 and VC-ACT3) were about 70% to 100% identical to the *Nocardioides* sp. strain JS614 genome (Genbank accession no. CP000508.1) [110] based on BLAST analysis of contigs and CheckM[123] estimated completeness ranged from 88.3% to 99.2% (Table 5.3). Among these five *Nocardioides*-affiliated genomes, contigs from ETH-ACT1 and VC-ACT1 displayed 99-100% coverage and were 99-100% identical to the *Nocardioides* sp. strain JS614 genome. Contigs from genome bin ETH-ACT2 and VC-ACT2 showed 99-100% nucleotide identity to each other and 70% to 100% nucleotide identity to the *Nocardioides* sp. strain JS614 genome. VC-ACT3 was not very similar to any of the other genome bins attributed to *Nocardioides*, with contigs (that can be aligned to contigs from VC-ACT2 and VC-ACT3) 73% to 100% identical to VC-ACT2 contigs and 75% to 95% identical to VC-ACT1 contigs.

No assembled 16S rRNA genes were found in any genome bins, therefore *rpoB* sequences were used to identify which genome corresponded to XL1 [36, 133, 134]. The partial *rpoB* gene (773 bp) from XL1 was 100% identical to *rpoB* genes retrieved from genome bins ETH-ACT1 (contig 6; 3501 bp) and VC-ACT1 (contig 19; 3465 bp), whereas it was 86% to 90% identical to the other *Nocardioides*-affiliated genome bins, from which only partial *rpoB* genes ranging from 270 bp to 3474 bp were successfully retrieved.

In genome bins designated ETH-PLSD and VC-PLSD (Figure 5.3), all of the contigs (except three from VC-PLSD) showed 100% coverage and nucleotide identity to the *Nocardioides* sp. JS614 plasmid pNOCA01 (Genbank accession no. CP000509.1)

[110](Figure 5.14), with completeness of 92.12% and 97.68% respectively compared to pNOCA01. Three VC-PLSD contigs showed 100% coverage and nucleotide identity to regions of the *Nocardioides* sp. strain JS614 chromosome, where integrase and IS3 family transposase were encoded, but only 50-70% coverage and 68-100% nucleotide identity to JS614 plasmid pNOCA01 (Table 5.4).

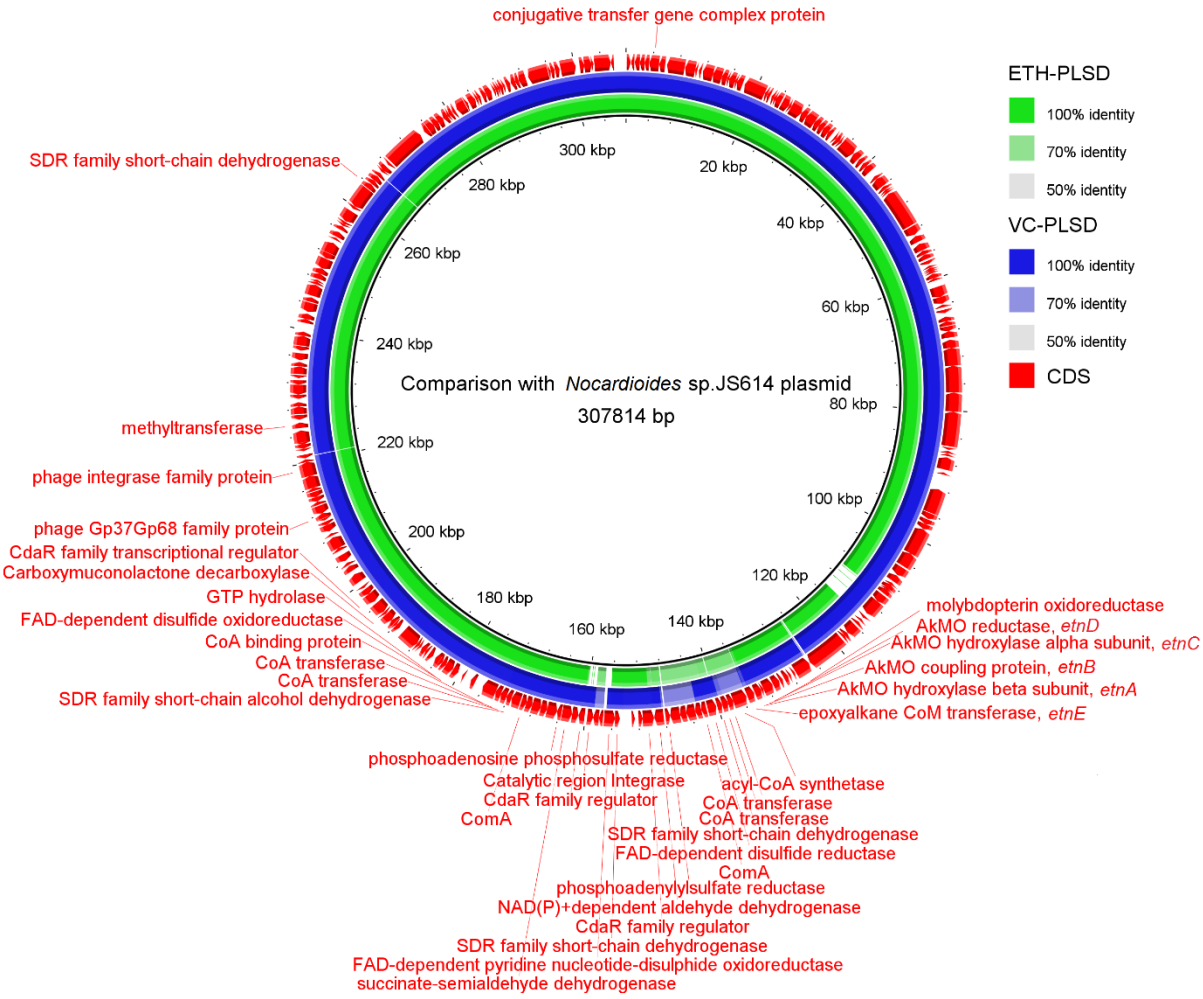


Figure 5.14 Comparison of putative plasmid bin ETH-PLSD and VC-PLSD with the completed *Nocardioides* sp. strain JS614 plasmid pNOCA01 using BLAST and visualized with BRIG [127].

Note: Since the plasmids have duplicated regions, when mapped with BRIG, they appeared as areas of <100% nucleotide identity.

Table 5.4 Contigs from VC-PLSD of <100% nucleotide identity to *Nocardioides* sp. strain JS614 plasmid pNOCA01 (Genbank accession no. CP000509.1).

Contig	Length (bp)	JS614 Plasmid		JS614 chromosome		Genes included (Genbank Accession)	Coverage in each VC sample			
		Coverage	ID	Coverage	ID		G1-TP2-VC	G1-TP3-VC	G2-TP2-VC	G2-TP3-VC
1500	5,105	55%	68%	100%	100%	JS614 integrase(WP_011751502.1)	15.09	29.11	40.27	54.48
7923	1,196	70%	70%	100%	100%	JS614 transposase IS3(ABL81662.1)	12.38	25.81	36.94	44.91
8105	1,175	62%	70%	100%	100%	JS614 peptidase M23B(WP_011755693.1)	11.32	15.81	41.66	55.54

In the ETH-ACT1, ETH-ACT2, VC-ACT1, VC-ACT3 genome bins, genes related to known ethene and VC degradation pathway genes were not detected. In VC-ACT2, partial *etnE* genes were found at the end of contig 239 with 98% (622 bp out of 635 bp) nucleotide identity to *Nocardioides* sp. strain JS614 *etnE*. In the genome bins containing plasmid DNA (ETH-PLSD and VC-PLSD), the gene cluster *etnABCD,* previously implicated as encoding AkMO and *etnE*, encoding EaCoMT (Figure 5.14) were found on the same contigs (9917 bp for both), with the same arrangement and orientation (*etnEABCD*) as in *Nocardioides* sp. strain JS614. The sequencing of partial *etnC* (1044 bp) and *etnE* (775 bp) from XL1 showed they are 100% identical to sequences in ETH-PLSD and VC-PLSD. Both ETH-PLSD and VC-PLSD have genes thought to participate in the downstream ethene and VC assimilation pathway, including genes encoding short-chain dehydrogenase (SDR) family alcohol dehydrogenase, acyl-CoA synthase and CoA transferase [8] (Figure 5.14). The putative plasmid bins also have genes encoding putative (2R)-phospho-3-sulfolactate synthase (ComA), which catalyzes the first step in CoM biosynthesis in methanogens [135], and conjugative transfer gene complex protein, catalytic region integrase and phage integrase (Figure 5.14). Central metabolic pathways were also checked (Table 5.5): all *Nocardioides*-affiliated genome bins (except the plasmid bins) contain a complete TCA pathway (KEGG pathway ko00020) and partial glyoxylate pathway (KEGG pathway ko00630), which is essential for completing VC assimilation [8]. Genes similar to *comB* gene in the Coenzyme M biosynthesis pathway in methanogens (KEGG pathway ko00680, enzyme EC 3.1.3.71) were detected in genome bins VC-ACT2, VC-ACT3 and ETH-ACT2, with 82.3% to 85.7% amino acid (aa) identity to the 2-phosphosulfolactate phosphatase (Genbank

accession no. WP_056909234.1) from *Nocardioides* sp. Root122.

Table 5.5 Completeness of selective metabolic pathways and existence of genes in alternative ethene/VC assimilating pathway among selected genome bins, based on annotation with KEGG and NCBI NR databases.

1(green)-pathway is complete/gene is present;

0.5(yellow)-only some genes in the pathway were found in the binned genome;

0(red)-no gene in the pathway was found in the genome bin/gene is not present.

Category	Pathway	VC-ACD1	VC-ACT1	VC-ACT2	VC-ACT3	VC-BAC1	VC-BAC2	VC-PRT1	VC-PRT2	ETH-ACT1	ETH-ACT2	ETH-PRT3	ETH-PRT6
Central metabolism	gluconeogenesis	1	1	1	0.5	1	1	1	1	1	1	1	1
	glycolysis	1	1	1	0.5	1	1	1	1	1	1	1	1
	TCA cycle	1	1	1	1	1	1	1	1	1	1	1	1
	glyoxylate cycle	0.5	0.5	0.5	0.5	0.5	1	0.5	0.5	0.5	0.5	0.5	0.5
	coenzyme M biosynthesis	0	0	0.5	0.5	0.5	0.5	0	0.5	0	0.5	0	0
Carbohydrate degradation	acetate to acetyl-CoA(C fixation)	1	1	1	1	1	1	1	1	1	1	1	1
	fatty acid degradation	1	1	1	1	1	1	1	1	1	1	1	1
	purine metabolism	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	methionine metabolism	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	glycerol degradation	0.5	0.5	1	1	0.5	1	1	1	0.5	1	1	1
Nitrogen	anammox	0.5	0.5	0.5	0.5	0.5	0	0.5	0.5	0.5	0.5	0.5	0.5
	nitrate reduction (denitrification)	0.5	0.5	0.5	0	0.5	0	0.5	0.5	0.5	0.5	0.5	0.5
	dissimilatory nitrate reduction	0.5	1	1	1	0	0	1	1	1	1	1	1
Sulfur	sulfate reduction (assimilatory)	0.5	0.5	1	1	1	0.5	1	1	0.5	1	1	0.5
	sulfide oxidation	0	0	0.5	0.5	0	0	0	0	0	0	0	1
	sulfite oxidation	0.5	0.5	0.5	0.5	0	0	0	0	0	0	0	0
Energy	oxidative phosphorylation	1	1	1	1	1	1	1	1	1	1	1	1

Table 5.5 Continued:

Category	Pathway	VC-ACD1	VC-ACT1	VC-ACT2	VC-ACT3	VC-BAC1	VC-BAC2	VC-PRT1	VC-PRT2	ETH-ACT1	ETH-ACT2	ETH-PRT3	ETH-PRT6
Mobility	flagella assembly	0	0.5	0.5	0	0	0.5	1	1	0.5	0	0	0
	bacterial chemotaxis	0	1	0.5	0.5	0.5	0.5	1	1	1	0.5	0	0
Category	Gene	VC-ACD1	VC-ACT1	VC-ACT2	VC-ACT3	VC-BAC1	VC-BAC2	VC-PRT1	VC-PRT2	ETH-ACT1	ETH-ACT2	ETH-PRT3	ETH-PRT6
Alternative Pathways	glutathione synthase, GSS	0	0	1	0	0	0	1	1	0	1	1	0
	glutathione reductase, GSR	0	1	0	0	0	0	1	1	1	0	1	1
	glutathione transferase, GST	0	0	0	0	0	0	1	1	0	0	1	1
	epoxide hydrolase gene	1	1	1	0	0	0	1	1	1	0	0	0

To explore the origin of the plasmid bins in the enrichment cultures, we examined contig coverage of each *Nocardioides*-affiliated genome bin (Figure 5.15). In ethene cultures, the contig coverages of ETH-PLSD showed variance with that of ETH-ACT1, with the ratio between contig coverages of ETH-ACT1 and ETH-PLSD (single-copy regions) ranging from about 1:1 to 1:2, but not with ETH-ACT2. In VC cultures, the contig coverage of VC-PLSD did not shift with the contig coverages of VC-ACT1, VC-ACT2 or VC-ACT3 individually, indicating the plasmid was not from a single *Nocardioides* strain in the VC culture. Particularly, in sample G1-TP2-VC and G2-TP3-VC, the genome bin VC-ACT2 and VC-ACT3 had low contig coverage (<2), whereas the contig coverage from genome bin VC-ACT1 and VC-PLSD were about 20 (single-copy regions) to 300 (multi-copy regions), with the ratio between VC-ACT1 and VC-PLSD coverages (single-copy regions) ranging from about 1:1 to 1:2. This indicates that VC-PLSD could have originated from the bacteria represented by genome bin VC-ACT1. However, in G1-TP3-VC, where VC-ACT1 was of low coverage (<5), genome bin VC-PLSD still had high coverage at about 50 (single-copy regions) to 150 (multi-copy regions), with VC-ACT3 also of high contig coverages from 30 (single-copy regions) to 200 (multi-copy regions), suggesting VC-PLSD also related to the bacteria represented by VC-ACT3. The contig coverages of other genome bins do not show any correlation with the contig coverages of the plasmid bins.

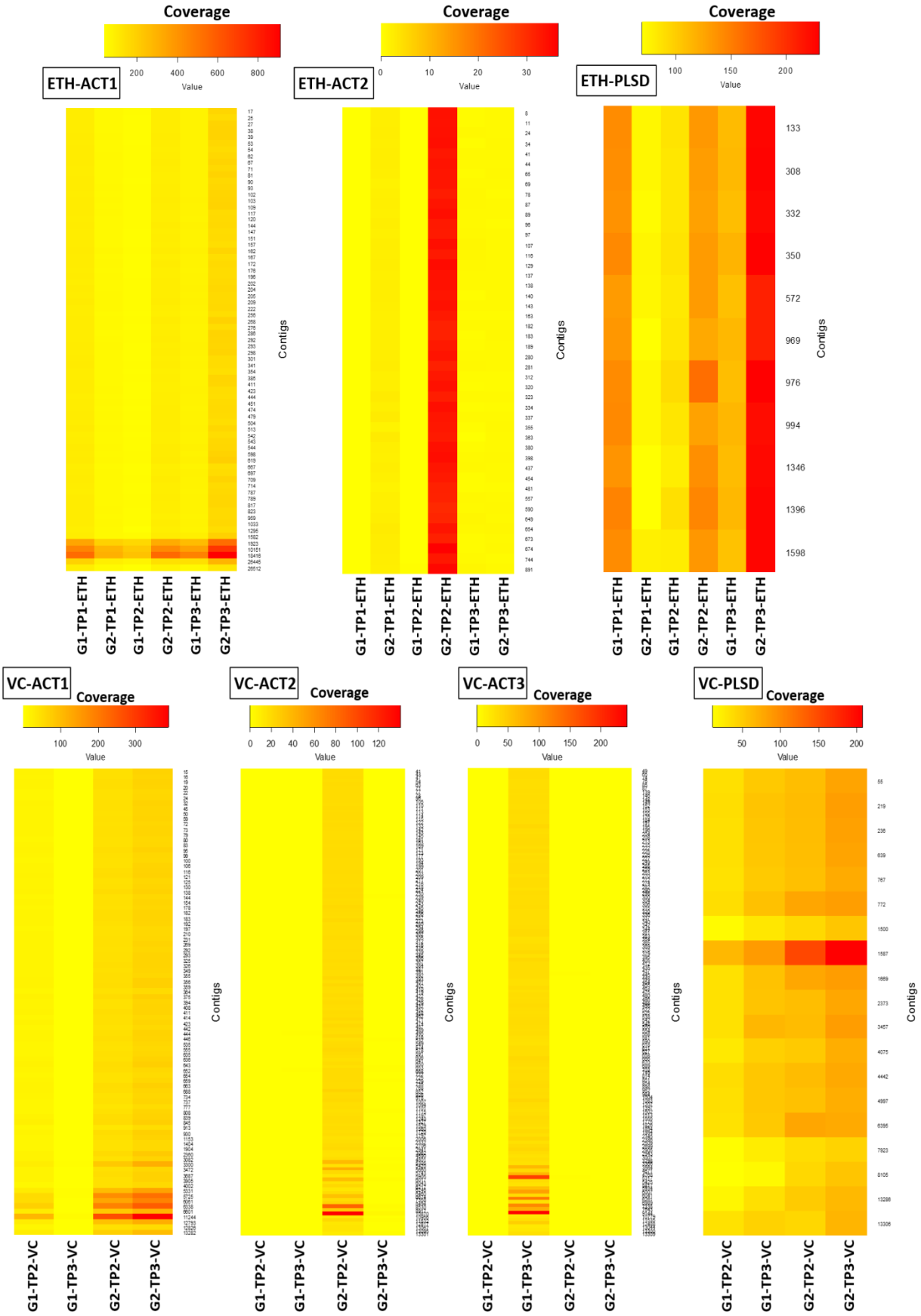

ETH-ACT1
Coverage
200 400 600 800
Value
Contigs
G1-TP1-ETH
G2-TP1-ETH
G1-TP2-ETH
G2-TP2-ETH
G1-TP3-ETH
G2-TP3-ETH
ETH-ACT2
Coverage
0 10 20 30
Value
Contigs
G1-TP1-ETH
G2-TP1-ETH
G1-TP2-ETH
G2-TP2-ETH
G1-TP3-ETH
G2-TP3-ETH
ETH-PLSD
Coverage
100 150 200
Value
133
308
332
350
572
969
976
994
1346
1396
1598
Contigs
G1-TP1-ETH
G2-TP1-ETH
G1-TP2-ETH
G2-TP2-ETH
G1-TP3-ETH
G2-TP3-ETH
VC-ACT1
Coverage
100 200 300
Value
Contigs
G1-TP2-VC
G1-TP3-VC
G2-TP2-VC
G2-TP3-VC
VC-ACT2
Coverage
0 20 40 60 80 100 120
Value
Contigs
G1-TP2-VC
G1-TP3-VC
G2-TP2-VC
G2-TP3-VC
VC-ACT3
Coverage
0 50 100 150 200
Value
Contigs
G1-TP2-VC
G1-TP3-VC
G2-TP2-VC
G2-TP3-VC
VC-PLSD
Coverage
50 100 150 200
Value
Contigs
G1-TP2-VC
G1-TP3-VC
G2-TP2-VC
G2-TP3-VC

Figure 5.15 The contig coverages from different genome bins in ethene and VC enrichment cultures visualized as heatmaps in R [47].

Note: Rows represent the contigs belonging to the genome and columns represent ethene and VC cultures. The coverage is illustrated by color (yellow to red as the coverage increases) for each genome.

5.4.7 Expression of etnC and etnE genes in the enrichment cultures

Since *etnC* and *etnE* are part of the gene cascade encoding enzymes for initial break-down of ethene and VC, we hypothesized that the gene expression of *etnC* and *etnE* were detectable. Their expressions in the liquid cultures were measured via RT-qPCR at the last time point (TP3) of the experiment, corrected with reference luciferase mRNA. The recovery efficiency of mRNA was 0.13% to 5.91%. Transcripts abundance of *etnC* varied from 9.7×10^{8} to 6.3×10^{10} transcripts/L of culture and *etnE* from 6.4×10^{6} to 1.6×10^{11} transcripts/L of cultures (Figure 5.16), respectively.

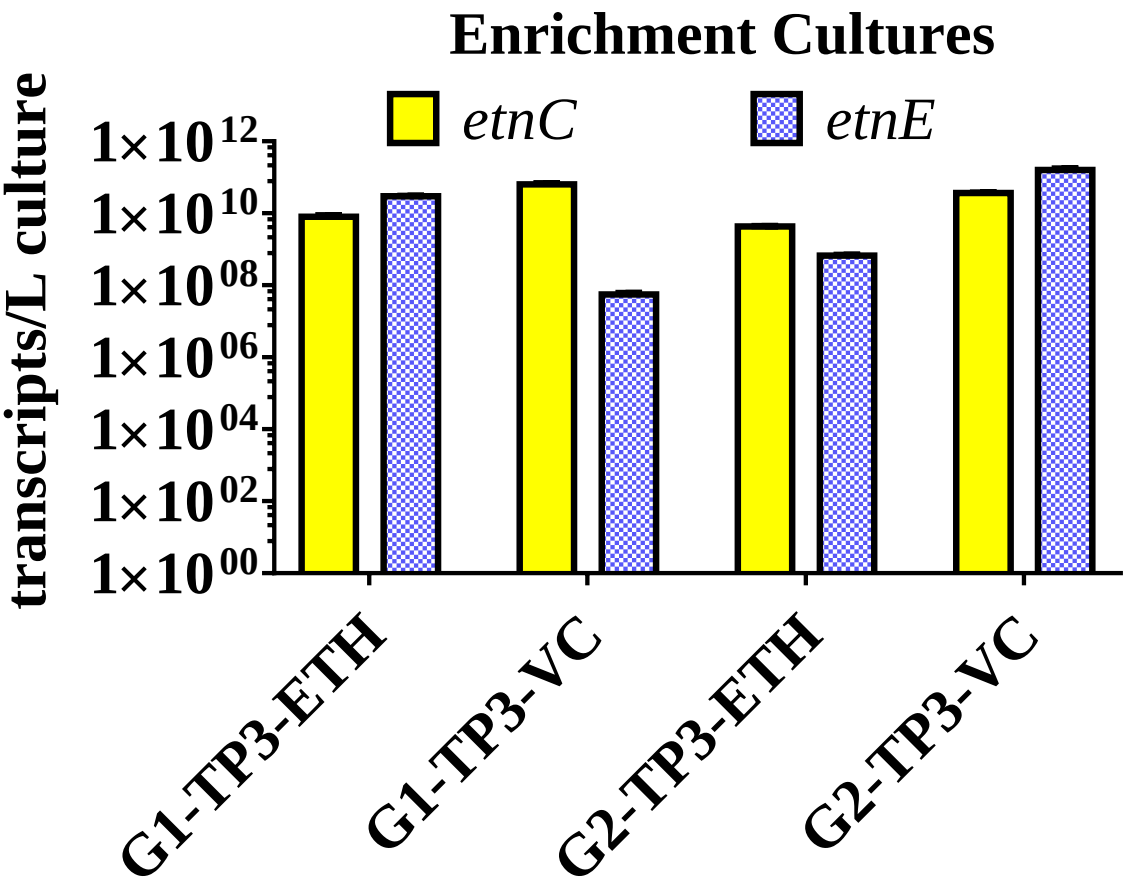


Figure 5.16 RT-qPCR data of functional gene *etnC* and *etnE* transcripts abundance in ethene and VC enrichment cultures at the last time point (TP3, 242-256 days). Error bars represent the differences between qPCR duplicates.

5.4.8 Analysis of other genome bins

Because a significant amount of sequences attributed to *Polaromonas* appeared in the ethene culture G1-TP3-ETH and *Pseudomonas* sequences were also dominant in the Group 1 VC culture sampled at time point 3 (G1-TP3-VC), we investigated their potential roles (ETH-PRT3, ETH-PRT6 and VC-PRT1) in the culture (Table 5.3) by analyzing genome annotation data. Although no *etnABCD* and *etnE* genes were found in these genomes, they all have complete pathways of central metabolism, fatty acids, amino acids and derivatives metabolism (Table 5.5). Notably, these *Proteobacteria* genome bins (*Polaromonas* and *Pseudomonas*) all have glutathione disulfide reductase (NADPH dependent) (KEGG pathway ko00480, EC 1.8.1.7), glutathione synthase (KEGG pathway ko00480, EC 6.3.2.3) and glutathione S-transferase (GST) (KEGG pathway ko00480, EC 2.5.1.18) (Table 5.5). The *Pseudomonas*-affiliated genome bins also have genes encoding epoxide hydrolase.

As the SIP data showed *Pedobacter* were taking up ^{13}C in labeled VC amended cultures, we also searched known and proposed VC assimilation pathways in *Pedobacter*-affiliated genome ETH-BAC3, VC-BAC1 and VC-BAC2 (Table 5.3). Neither genes in the known VC pathway (e.g. *etnABCD* and *etnE*) nor proposed alternative pathways were detected in these genome bins.

5.4.9 Identifying Nocardioides in the original groundwater

As *Nocardioides* sp. became more abundant during ethene and VC enrichment and was proved to be assimilating ethene and VC, we went back to investigate its presence in the original groundwater sample. The relative abundance of *Nocardioides*

was 0.71% in metagenome and 0.02% in 16S rRNA gene Illumina sequencing data. From the MG-RAST analysis, a total of 4066 sequences related to *Nocardioides* sp. JS614 were found in the groundwater (100% to 50% nucleotide identity, average alignment length 126 bp), among which 528 of the sequences (relative abundance=0.08%) were of 100% nucleotide identity to JS614. From 16S rRNA gene Illumina sequencing, four sequences (relative abundance=0.005%) were 100% identical with 16S rRNA gene sequence in *Nocardioides* sp. strain XL1. Previous study showed that the *etnE* cloned from groundwater at this Fairbanks, AK site were of 98-99% nucleotide identity to JS614 *etnE1* and *etnE* [98]. Taken together, the evidence showed that bacteria similar to *Nocardioides* sp. strain JS614 were present in the original groundwater sample used in this study.

5.5 Discussion

Aerobic VC-assimilating bacteria are thought to be widely distributed in the environment, especially at chlorinated ethene contaminated sites [10, 98]. In this study, ethene and VC-assimilating *Nocardioides* strains were significantly enriched after the groundwater was amended with ethene. The bacterial culture readily shifted from growth on ethene to growth on VC, despite the fact that the original groundwater consortium could not grow on VC. A similar phenomenon was observed in previous study [10], where ethene-assimilating *Mycobacterium* sp. strain K1 adapted to VC as the sole carbon and energy source, but its source consortium could not grow on VC [14]. As no alternative VC-assimilating pathway was found in the metagenomes, this suggests that a similar mechanism controlled adaptation to VC in our experiments. This is also reflected by the

similar bacterial community structure in ethene and VC enrichment cultures.

There are several possible explanations for why the consortium required enrichment on ethene in order to transition to growth on VC. During metabolism of ethene, bacteria generate epoxyethane, a known inducer of alkene monooxygenase [36, 37]. As more alkene monooxygenase is expressed in the culture, this provides more available enzyme to degrade VC. At chlorinated ethene contaminated sites, ethene is the desired end product of anaerobic dechlorination. Bioremediation strategies have applied cometabolism of VC, for example, at the Carver, MA site[136], where ethene was used as a primary substrate. However, our study showed the presence of ethene selects for bacteria that use their existing ethene-assimilation pathway for VC-assimilation. This suggests that it is necessary to provide ethene as the primary substrate initially, but it might not need ethene along the entire course of VC removal in-situ. The expression level of functional gene *etnC* and *etnE* in the ethene and VC cultures was about five to seven fold higher than reported for groundwater at VC-contaminated sites[111, 137], which means the genes were not only involved, but also highly expressed under ethene- and VC-assimilating conditions.

The *Nocardioides* sp. strain XL1 was isolated from the VC enrichment culture, which was linked to genome bins ETH-ACT1 and VC-ACT1 based on analysis of essential single copy gene *rpoB*. Therefore, these genome bins can be used to infer the potential metabolic functions of strain XL1. Although the 16S rRNA gene, *rpoB* and even the contigs from genome bins representing strain XL1 were 99% to 100% similar to previously identified VC-assimilating *Nocardioides* sp. strain JS614, there are still some obvious differences. There are 51 translated aa sequences from VC-ACT1 97.4% to

99.9% identical to strain JS614. For example, in VC-ACT1 (representing the strain XL1), the aa sequence of twitching motility protein (encoded by *pilT*) (Genbank accession no. WP_011754594.1) has an aa mutation (Leucine to Methionine, 1st aa) and the aa sequence of *FliA/WhiG* family RNA polymerase sigma factor (Genbank accession no. WP_011754582.1) has an aa mutation (Methionine to Valine, 3rd aa) compared to *Nocardioides* sp. strain JS614. These mutations are non-conservative, which may cause the loss of flagella function in strain XL1, forming clumpy liquid cultures (Figure 5.13). However, the clumpy appearance may also result from the hydrophobic cell walls that many *Actinobacteria* are known to have. The *Nocardioides* sp. strain XL1 appeared red-colored (gram-negative) during gram staining, contradicts to the expected purple-color (gram-positive, an attribute of *Nocardioides*), which may be caused by a hydrophobic cell wall. There are other differences in the genome, including three translated aa sequences in VC-ACT1 not found in *Nocardioides* sp. strain JS614: contig 45 aa 75 (96.4% identical to Genbank accession no. AJR18274.1), contig 73 aa 100 (70.2% identical to Genbank accession no. WP_022910218.1) and contig 192 aa 7 (60.4% identical to Genbank accession no. WP_058857804.1). However, since many of them are not annotated, it is hard to predict their related functions. Another difference between *Nocardioides* sp. strain XL1 and strain JS614 was that there was no lag period observed for *Nocardioides* sp. strain XL1 after VC was depleted in the culture (Figure 5.11)[36].

There were other *Nocardioides*-affiliated bacteria in the VC enrichment cultures, and one is represented by genome bin VC-ACT3. It was unique in VC cultures-no similar genome was binned in ETH cultures. It also showed co-occurrence with VC-PLSD at high coverage (abundance) level in the culture G1-TP3-VC. However, our attempt has

been unsuccessful to isolate this strain. We suspect that horizontal gene transfer might have happened between the different *Nocardioides* in the culture, during which the plasmids bearing ethene and VC-assimilation pathway in *Nocardioides* sp. strain XL1 were transferred into the bacteria represented by genome bin VC-ACT3.

The plasmid genome bins VC-PLSD and ETH-PLSD extracted from metagenome are highly identical to the plasmid pNOCA01 found in *Nocardioides* sp. strain JS614, despite three contigs in VC-PLSD are 100% identical to sequences in JS614 chromosome. The sizes of the plasmid bins are also around 300 kb, similar to the plasmid pNOCA01 (308 kb). The functional genes regarding ethene and VC assimilation are only found in these plasmid genome bins, which were always at the highest coverage level in all the samples (69-230 in six ethene cultures and 10-207 in four VC cultures).

As there is not much data about the copy numbers of the plasmid in ethene- and VC- assimilating bacteria, we attempted to estimate the copy numbers based on metagenome data, via comparing the coverage of contigs with single copy regions in the plasmid bins (ETH-PLSD and VC-PLSD) versus contigs carrying the essential single copy genes (ESS) in the chromosome genome bins (ETH-ACT1, VC-ACT1 and VC-ACT3) using the following equation and calculated for each culture (sample):

$$plasmid\ copy\ no. = \frac{Avg.coverage\ of\ single\ copy\ contigs\ in\ plasmid\ bin}{\sum Avg.coverage\ of\ ESS\ bearing\ contigs\ in\ chormosome\ bin}$$

In the six ethene cultures, we assumed that only the bacteria strain represented by ETH-ACT1 has the plasmids. From the calculation, the plasmid copy number varies from 1.2 to 1.8 in bacteria represented by ETH-ACT1. In the four VC cultures, we assumed that both VC-ACT1 and VC-ACT3 have the plasmid. The plasmid copy number in bacteria represented by VC-ACT1 and VC-ACT3 vary from 1.3 to 2.1. These numbers

indicate the ethene- and VC-assimilating bacteria harbor one or two copy of the plasmid. Nonetheless, it is also possible that there were more than one type of plasmid in the culture, with very high similarity. The de-novo assembly could not distinguish the origins of the similar plasmid fragments, therefore, they could be assembled on the same contig.

In this study, we found a contig in VC-PLSD covering the start and end of the *Nocardioides* sp. strain JS614 plasmid pNOCA01, which was proved to be linear with pulsed-field gel electrophoresis (PFGE) in previous study [9]. The contig extends from the 109913 bp to the 1st bp then from the 307814 bp (the end of pNOCA01) to 266175 bp of pNOCA01. If the plasmid is 100% identical with pNOCA01, this evidence showed that the plasmid is actually circular. However, linear plasmids were also seen via PFGE in other aerobic alkene-degrading bacteria, such as VC degrader *Pseudomonas putida* strain AJ (260 kb) and *Ochrobactrum* strain TD (three plasmids, ranging from 90 kb to 320 kb)[24] and ethene/propane degrader *Xanthobacter* Py2(320 kb) [20]. We speculated that the mega-plasmids were super-coiled circular plasmids that behave like linear plasmid during PFGE.

Other than *Nocardioides*, there was a significant amount of sequences associated with the genera *Polaromonas*, *Pseudomonas* and *Pedobacter* found in the ethene and VC cultures. These heterotrophic bacteria could participate in metabolite exchange through carbon, nitrogen and sulfur cycling, amino acid and lipid metabolism, and breaking down other microbial products, thereby enhancing overall VC removal.

Genes associated with the glutathione biosynthesis pathway were found in *Polaromonas*, *Mesorhizobium* and *Pseudomonas*-affiliated genome bins. In *cis*-DCE detoxification, glutathione was conjugated to *cis*-1,2-dichloroepoxyethane via GST in

Rhodococcus sp. strain AD45 [138, 139]. As *Pseudomonas* was the most abundant bacteria genus in one of the VC enrichment cultures, the GST genes recovered in the *Pseudomonas*-affiliated genome bin VC-PRT3 was analyzed in more detail. In phylogenetic analysis, two GST aa sequences from VC-PRT1 grouped with GST from *Rhodococcus* sp. strain AD45 (Figure 5.17), one of which (VC-PRT1_1_221) is of 98% coverage and 30 % similarity to *IsoILR1* (Genbank accession no. AAS66909). *IsoILR1* was proved to have activity towards *cis*-1,2-dichloroepoxyethane[138], whereas *IsoILR2* (Genbank accession no. AAS66910) in the same clad is nine amino acids different from *IsoILR1* and lacks activity towards *cis*-1,2-DCE epoxide. VC-PRT1_1_221 (100% identical to MULTISPECIES: hypothetical protein *Pseudomonas* (Genbank Accession No. WP_024762862)) not only remain identical to *IsoILR1* at these mutation sites (except one at the 15th aa near N terminal), but also shared all the functional sites and regions annotated in *IsoILR1*, indicating its potential activity towards epoxides. We speculated that the *Polaromonas*, *Mesorhizobium* and *Pseudomonas* in our cultures were also able to detoxify chlorooxirane (VC epoxide) with the same mechanism via conjugation to glutathione via GST, which contributed to their robustness in the culture. On the other hand, the potential glutathione detoxification mechanism in these *Proteobacteria* would be competing for epoxide with ethene- and VC-assimilating bacteria. This was reflected in Group 1, where *Nocardioides* is not as abundant as in cultures from Group 2 (with fewer *Proteobacteria*). GST are present in a large number of *Proteobacteria* genomes [140], including *Polaromonas*, *Mesorhizobium* and *Pseudomonas*. It would be interesting to further investigate whether these *Proteobacteria* have the ability to detoxify alkene epoxides through glutathione conjugation. Nonetheless, from the cultures in Group 2, we

can tell that these *Proteobacteria* are not essential for maintaining a robust ethene- or VC-assimilating bacterial consortium. The SIP data also supports this argument, showing *Pseudomonas* was not taking up ^{13}C from labeled VC.

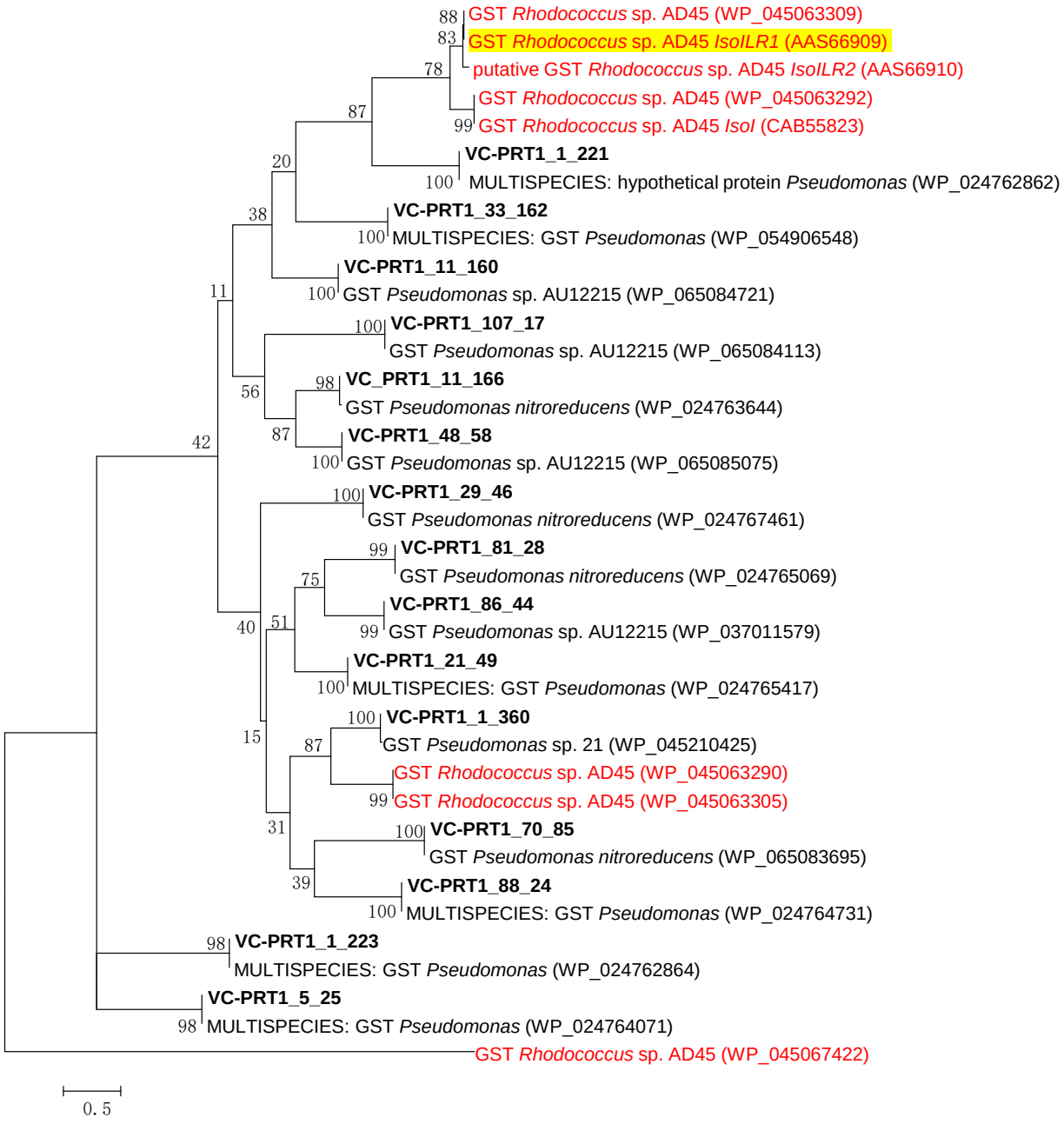


Figure 5.17 Phylogenetic tree of complete GST aa sequences from *Pseudomonas*-affiliated genome bin (VC-PRT1) (in bold) and their top blast hits and GST aa sequences from *Rhodococcus* sp. AD45 (red-colored).

Note: An alignment with 335 aa was generated (including gaps) in ClustalW (Thompson et al 2002) and the maximum likelihood (Tamura et al 2011) tree was constructed with 500 times bootstrap and visualized in MEGA6 (Tamura et al 2013). The bar represents a 50% difference. The highlighted sequence has been proved to conjugate GSH to cis-1, 2-dichloroepoxyethane.

Other than GST detoxification, there was other possible epoxide degradation pathway. For example, epichlorohydrin can be degraded via epoxide hydrolase [141, 142]. Epoxide hydrolase encoding genes were found in *Nocardioides*- and *Pseudomonas*-affiliated genome bins. Nonetheless, the activity of these enzymes needs to be tested in further experiments.

With respect to *Sediminibacterium*, the metagenomics data was not consistent with 16S rRNA gene Illumina sequencing data, the latter of which showed a significant amount of *Sediminibacterium* 16S rRNA V4 region sequences in cultures from Group 2. The different databases used in metagenomics (M5NR database) and 16S rRNA gene (RDP classifier) analysis may have an effect on the result. But this may also be a result of sequencing bias in 16S rRNA gene amplicon sequencing: either the primer sets bias towards *Sediminibacterium* or *Sediminibacterium* have more copies of 16S rRNA genes than other bacteria. Another reason might be the genome similarity between *Pedobacter* and *Sediminibacterium*. In our *Pedobacter*-affiliated genome bins, some of the coding sequences were annotated as *Sediminibacterium* (<0.5%). The problem of 16S rRNA gene amplicon sequencing has been discussed in previous literature [143], stating it is limited by the presence of variable copy numbers in bacterial genomes and sequence variation within closely related taxa or within a genome. However, the SIP data did show a slight trend of *Pedobacter* and *Sediminibacterium* taking up ^{13}C from labeled VC-fed cultures. Previous study [65] has also revealed *Sediminibacterium* was responsible for ^{13}C label-uptake from VC. But since no genes from the VC-assimilation pathway was found in these *Pedobacter/Sediminibacterium*-affiliated genome bins, they were probably scavengers in the culture.

5.6 Conclusions

This study showed the groundwater microbial community can adapt faster to VC after ethene enrichment. During the process, *Nocardioides* was selected as the major ethene and VC-assimilating bacteria. Three types of *Nocardioides* were found in the cultures, two of which were related to VC-assimilation and carrying the plasmid with functional genes in the ethene and VC assimilating pathway. A novel VC-assimilating bacteria, *Nocardioides* sp. strain XL1, was isolated from the VC culture. There were also a significant amount of *Polaromonas*, *Pseudomonas* and *Pedobacter* in the cultures, which were scavengers. There was another potential role of *Polaromonas* and *Pseudomonas*, which could use glutathione S-transferase to detoxify the epoxyalkane and chlorooxirane.

CHAPTER VI. "INTO THE WILD": EPOXYALKANE: COENZYME M TRANSFERASE GENE DIVERSITY AND ABUNDANCE IN GROUNDWATER SAMPLES FROM CHLORINATED ETHENE CONTAMINATED SITES

This study was majorly conducted by Xikun Liu. Sample source is listed in the Acknowledgement, coordinated by Professor Timothy E Mattes. Dr. Yang Oh Jin was responsible for extraction of the Carver, Oceana and Soldotna samples. This study has been published in Applied and Environmental Microbiology on March 25th, 2016.

6.1 Abstract

Epoxyalkane:coenzyme M transferase (EaCoMT) plays a critical role in aerobic biodegradation and assimilation of alkenes, including ethene, propene and the toxic chloroethene vinyl chloride (VC). To improve our understanding of the diversity and distribution of EaCoMT genes in the environment, novel EaCoMT-specific T-RFLP and nested PCR methods were developed and applied to groundwater samples from six different contaminated sites. T-RFLP analysis revealed 192 different EaCoMT T-RFs. By constructing clone libraries, we retrieved 139 EaCoMT gene sequences from these samples. Phylogenetic analysis revealed that a majority of the sequences (78.4%) grouped with EaCoMT genes found in VC- and ethene-assimilating *Mycobacterium* strains and *Nocardioides* sp. strain JS614. The four most abundant T-RFs were also matched with EaCoMT clone sequences related to *Mycobacterium* and *Nocardioides* strains. The remaining EaCoMT sequences clustered within two emergent EaCoMT gene subgroups represented by sequences found in propene-assimilating *Gordonia rubpertinctus* strain B-276 and *Xanthobacter autotrophicus* strain Py2. EaCoMT gene

abundance was positively correlated with VC and ethene concentrations at the sites studied.

6.2 Introduction

Short-chain alkenes (e.g. ethene, propene, and butene) are common hydrocarbons in the environment, primarily encountered as fossil fuel components, products of living organisms, or generated by the chemical industry [8]. For instance, ethene is generated by both plants [144] and bacteria [145].

Chlorinated alkenes (e.g. vinyl chloride (VC)) are also naturally occurring, albeit at very low levels[4]. VC is produced industrially as a monomer for polyvinyl chloride plastics. However, most environmental VC is generated by incomplete anaerobic dechlorination of the widely used solvents tetrachloroethene (PCE) and trichloroethene (TCE) in groundwater, where ethene can also be generated as a complete dechlorination product [8]. PCE, TCE, and VC are common groundwater contaminants [100], and sites contaminated with chloroethenes are widely distributed across the USA [146] and elsewhere [8, 147]. VC is of particular concern as a known human carcinogen [148].

In aerobic bacteria that utilize short-chain and chlorinated alkenes as carbon and energy sources, a monooxygenase enzyme typically catalyzes the initial attack [8, 107] forming aliphatic epoxides. Aliphatic epoxides are highly reactive molecules that covalently bind proteins and nucleic acids, leading to toxic and mutagenic effects [149, 150] in most organisms. Certain aerobic alkene-oxidizing bacteria metabolize and/or detoxify these epoxides by conjugation to coenzyme M (CoM) with the enzyme Epoxyalkane: Coenzyme M transferase (EaCoMT; encoded by the gene designated *etnE* in some

organisms) [12, 107, 108].

EaCoMT has been implicated in aerobic assimilation of propene [108, 151], ethene, and VC [8, 9, 12] . EaCoMT belongs to a subset of the alkyl transferase family, in which zinc catalyzes thiol activation for nucleophilic attack [151, 152]. A functionally analogous transferase enzyme is the cobalamin-independent methionine synthase MetE which transfers a methyl group to homocysteine during methionine synthesis [153]. Both EaCoMT and MetE contain a conserved His-X-Cys-X-Cys zinc binding motif [12, 154, 155], which is important in thiol group transfer. However, with the exception of the EaCoMT from *Xanthobacter* Py2 [107], MetE and EaCoMT do not share significant nucleotide sequence identity.

Bacteria expressing EaCoMT during growth on VC and/or ethene include strains of *Mycobacterium* [10, 11, 71, 72], *Nocardioides* [9, 10], *Pseudomonas* [17] and *Ochrobactrum* [17]. Propene-assimilating *Gordonia rubripertincta* strain B-276 (previously identified as *Nocardia corallina* and *Rhodococcus rhodochrous* strain B-276) [156, 157] and *Xanthobacter* strain Py2 [37] express homologous EaCoMT genes during growth on propene, but can also grow on ethene. EaCoMT genes are known to be carried on linear plasmids in several VC-, ethene-assimilating bacteria as well as the propene-assimilating *Gordonia* strain B-276 and *Xanthobacter* strain Py2 [9, 11, 20, 22, 24]. A recently isolated ethene-assimilating *Haliea* strain [19] also contains a putative EaCoMT gene.

Some ethene-assimilating bacteria degrade VC fortuitously (i.e. via cometabolism [13, 14]) while others can use VC as a carbon and energy source (VC-assimilators). Several ethene-assimilating isolates have successfully transitioned from cometabolic VC degradation to growth-coupled VC metabolism after an extended incubation period with

VC [31, 106]. Mutations in the EaCoMT gene were implicated in the evolution of an ethene-assimilating bacterium into a VC-assimilating bacterium in laboratory experiments [72].

The EaCoMT gene (often designated as *etnE*) could be a useful biomarker for aerobic VC biodegradation in the field. Quantitative PCR (qPCR) assays for *etnE* have been developed [77, 97, 111] and applied [65, 78, 111, 158] in an effort to understand VC biodegradation in microcosm studies as well as directly from environmental samples. However, because *etnE* sequences in ethene- and VC-assimilating bacteria are very similar, the EaCoMT gene qPCR assay can not distinguish VC-assimilating bacteria from ethene-assimilating bacteria that fortuitously degrade VC.

Despite the importance of EaCoMT in global carbon and halogen cycling and as a diagnostic biomarker for VC bioremediation, very little is known about its distribution and diversity in the environment. Therefore, the primary goals of this study were to retrieve EaCoMT genes from environmental samples and expand the available database of EaCoMT gene sequences, examine EaCoMT diversity and distribution patterns at contaminated sites, and investigate relationships between EaCoMT gene abundance and diversity and VC concentrations and attenuation rates at contaminated sites. Because of the burgeoning use of the EaCoMT gene as a biomarker for VC biodegradation in the environment, we focused our efforts on geographically diverse chloroethene-contaminated groundwater samples.

6.3 Materials and Methods

6.3.1 Site information, environmental sample collection and DNA extraction

Groundwater samples from six sites featuring varying VC concentrations (Table 6.1) were collected in collaboration with consulting firms and United States agencies. The Kotzebue, AK and Fairbanks, AK sites were contaminated with TCE and *cis*-dichloroethene (i.e. potential VC precursors) but VC was not detected. The Carver, MA, Soldotna, AK and Oceana Naval Air Station, VA sites contain relatively dilute groundwater VC plumes (i.e. less than 100 µg/L VC) at the time of sampling. Finally, the basalt site in Melbourne, Australia contains relatively high concentrations of VC (up to 72 mg/L).

Table 6.1 Summary of groundwater samples information in this study. Statistically significant k values are based on $p<0.1$

k_{point} and k_{bulk} are estimated values of the point VC decay rate and bulk VC attenuation rate, respectively. NA: not analyzed

Location	Well ID No.	Sample designation	Sampling Date	DNA (ng/μL)	VC (μg/L)	Ethene (μg/L)	DO (mg/L)	k_{point} (yr^{-1})	k_{bulk} (yr^{-1})	EaCoMT gene abundance (genes/L)*	Diversity (H' and 1/D)
Carver, MA[a] (CARV)	RB46D	CARV46	9/15/2010	3.25	1.5	NA	0.65	0.365 (p<0.001)	0.159 (p=0.001)	NA	NA
	RB63I	CARV63	9/15/2010	3.95	1.3	NA	0.29	0.183 (p<0.001)		NA	0.73±0.01 1.55±0.02
	RB46D	CARV46-1	9/29/2009	NA	1.8	3.9	0.31	0.365 (p<0.001)	0.119 (p=0.001)	1.2×10^{5}	NA
	RB64I	CARV64	09/29/2009	NA	<0.46	<0.1	0.45	0.1095 (p<0.001)		1.9×10^{4}	NA
Soldotna, AK[b] (SOLD)	MW6	SOLD6	05/12/2009	1.00	9.6	50	0.53	0.621 (p<0.001)	1.332 (p=0.050)	6.3×10^{5}	1.05±0.50 2.28±0.94
	MW40	SOLD40	09/22/2008	2.61	20.7	90	1.29	0.730 (p<0.001)	2.093 (p=0.059)	1.6×10^{5}	1.53±0.15 2.64±0.38
Oceana, VA[c] (OCEA)	MW18	OCEA18	08/06/2009	0.14	0.8	<1	1.89	0.730 (p<0.001)	0.008 (p=0.524)	4.3×10^{3}	1.29±0.22 3.48±0.64
	MW25	OCEA25	11/21/2008	8.96	19	<1	1.20	0.256 (p=0.132)	0.003 (p=0.379)	2.4×10^{4}	1.54±0.14 3.73±0.38
Melbourne, Australia[d] (AUS)	039IJ-1	AUS39-1	10/10/2011	1.24	72000	780	0.07	NA	0.368 (p=0.034)	8.6×10^{6}	1.47±0.00 3.93±0.01
	039IJ-3	AUS39-3	10/10/2011	0.32	4400	NA	0.28	NA		4.7×10^{4}	0.98±0.07 2.12±0.04
	039IJ-6	AUS39-6	10/17/2011	1.00	53000	230	1.07	NA		3.0×10^{6}	1.17±0.03 2.45±0.04
	039IJ-7	AUS39-7	10/17/2011	0.78	15000	40	2.2	NA		3.0×10^{5}	1.09±0.03 2.41±0.05
	039IJ-8	AUS39-8	10/18/2011	2.28	24000	NA	0.65	NA		1.3×10^{6}	1.08±0.07 2.36±0.08

Table 6.1 Continued.

Kotzebue, AK[e] (KOTZ)	MW10-01	KOTZ01	10/22/2013	10.4	<0.62	NA	1.46	NA	NA	6.6×10^5	0.95±0.09 2.11±0.33
	MW10-03	KOTZ03	10/22/2013	1.31	<0.62	NA	8.70	NA		9.8×10^4	0.91±0.08 2.15±0.05
Fairbanks, AK[e] (FAIR)	MW-4M	FAIR4	03/27/2014	15.7	<0.4	<0.06	1.36	NA	NA	7.0×10^3	1.39±0.42 2.87±1.13
	MW-13M	FAIR13	03/27/2014	14.3	<0.4	<0.06	0.88	NA		3.9×10^4	1.68±0.15 3.54±0.45

VC, ethene and geochemical data was provided by personal communication and/or publicly available reports as described below:
[a]James Begley of MT Environmental Restoration and Sam Fogel of Bioremediation Consulting, Inc.;
[b]May 2009 Groundwater Monitoring Report, River Terrace RV Park, Soldotna, AK, Alaska Department of Environmental Conservation via Tim McDougall of Oasis Environmental and James Fish of Alaska Department of Environmental Quality;
[c]Long-term Monitoring Report (2009) for SMWUs 2B, 2C, and 2E, Oceana NAS, Virginia Beach, VA, via Laura Cook of CH2MHill;
[d]Dora Ogles from Microbial Insights, Inc.;
[e]James Fish of Alaska Department of Environmental Conservation;
*Quantification of *etnE* genes at CARV, SOLD, and OCEA were published previously [77] and are provided here for reference and used in correlation analyses.

These VC-contaminated sites also feature conditions favorable for VC oxidation processes. Molecular oxygen and ethene were injected into the VC plume at the Carver site to promote VC oxidation, with apparent success [159]. Reverse-transcription qPCR (RT-qPCR) evidence of VC-oxidizer activity at Carver site groundwater has also been reported[111]. Biostimulation of the VC plume with an oxygen releasing compound was conducted at the Oceana site, where VC concentrations have been decreasing since 2004 [160]. VC oxidation in aerobic microcosms constructed with sediments from the Soldotna site has been reported [161]. Finally, evidence of aerobic VC degradation in the vadose zone was observed at the Australia site [78]. Groundwater geochemical parameters collected at these sites (i.e. dissolved oxygen (DO), pH, temperature, and oxidation reduction potential (ORP)) as well as VC and ethene concentrations in monitoring wells at the time of sampling are provided (Table 6.1, Table 6.2).

Biomass for DNA extraction was collected by passing groundwater (1-3 L) through Sterivex-GP 0.22 μm membrane filter cartridges (Millipore Corporation, Billerica, MA) in the field as described previously [77]. Filters (with the exception of Australia samples) were shipped overnight to the University of Iowa and stored at -80 °C until extraction. Sterivex filter samples from Australia were handled by Microbial Insights, Inc (10515 Research Drive, Knoxville, TN 37932).

Australia, Carver, Oceana and Soldotna samples were extracted using MoBio PowerSoil DNA Isolation Kit (MO BIO, Carlsbad, CA) as described previously [77], while the Kotzebue and Fairbanks samples were extracted using the MoBio PowerWater Sterivex DNA Isolation Kit. Elution buffer volumes were 100 μL (Australia, Carver, Soldotna, Kotzebue), 40 μL (Oceana) and 50 μL (Fairbanks). DNA concentrations were

estimated with Qubit™ fluorometer (Invitrogen, Waltham MA) using the Quant-iT dsDNA HS assay kit (Table 6.1).

Table 6.2 Field groundwater geochemical parameters collected from the monitoring wells in this study. These parameters and groundwater velocities were taken from the reports cited in Table 6.1.

Well	Sample Designation	pH	Temperature (℃)	ORP (millivolts)	Groundwater Flow Velocity (ft/yr)
Carver, MA RB46D* (09/29/2009)	CARV46-1	5.9	14.05	135	277.4
Carver, MA RB64I* (09/29/2009)	CARV64	5.28	12.04	152	
Carver, MA RB46D (09/15/2010)	CARV46	5.6	14.08	217	
Carver, MA RB63I (09/15/2010)	CARV63	5.91	13.36	91	
Oceana, VA MW18 (08/06/2009)	OCEA18	7.39	21.4	-53	10.0
Oceana, VA MW25 (11/21/2008)	OCEA25	6.37	20	-93	
Soldotna, AK MW6 (05/12/2009)	SOLD6	4.85	3.21	3.7	54.75
Soldotna, AK MW40 (09/22/2008)	SOLD40	6.11	6.16	-55	
Fairbanks, AK 4M (03/27/2014)	FAIR4	6.26	4.56	34.8	--
Fairbanks, AK 13M (03/27/2014)	FAIR13	6.68	3.8	-20.9	
Kotzebue, AK 10-01 (10/22/2013)	KOTZ01	7.2	0.4	-101.4	--
Kotzebue, AK 10-03 (10/22/2013)	KOTZ03	6.48^{+}	1.9^{+}	-125.1^{+}	
Australia 039IJ-1,3,6,7,8 (Oct.2011)	AUS39-1,3,6,7,8	NA	NA	NA	123.03

NA: data not provided.

--: VC attenuation rates were not calculated for these sites as VC was not detected, thus groundwater flow velocities were not used in the analysis

*These samples were used in the correlation analysis of *etnE* abundance with geochemical parameters, ethene concentrations, VC concentrations and attenuation rates, but was not included in the T-RFLP and clone library analyses.

$^{+}$Kotzebue, AK 10-03: The original data sheet marked as "Parameters measured in-situ - may not represent true groundwater conditions" therefore were excluded in correlation analysis of these parameters.

6.3.2 Quantitative PCR

Reaction mixtures for qPCR (25 µL) contained 12.5 µL of Power SYBR Green PCR Master Mix (Applied Biosystems, Life Technologies Corporation, Carlsbad, CA), 750 nM RTE primers (Table AI.7) and 2 µL of DNA extract. Bovine serum albumin (0.5 µg) was added to alleviate possible PCR inhibition [162]. All qPCRs were performed in at least triplicate for all samples with the ABI 7000 Sequence Detection System and analyzed by ABI 7000 System SDS Software (Applied Biosystems) using the auto baseline and auto Ct functions. Standards were prepared using *Nocardioides* sp. strain JS614 genomic DNA as the template as described previously [77] except that 0.2 µM of each CoMF1L and CoMR2E primer (Table AI.7) were used to minimize formation of primer-dimers. Other detailed qPCR information (e.g. fluorescent threshold and efficiency) are provided (Table AI.8) in accordance with MIQE guidelines [60]. In dissociation curve analysis, *etnE* amplicons generated with RTE primers displayed melting temperatures of 84.7-85.3 °C (JS614 standards), 82.8-84.1 °C (Australia samples), 82.6-84.4 °C (Kotzebue samples) and 84.7-85.6 °C (Fairbanks samples).

6.3.3 PCR amplification of EaCoMT genes

In general, conventional PCR amplification of EaCoMT genes from environmental DNA extracts did not yield visible PCR products on agarose gels for every site. Therefore, we investigated successive rounds of conventional PCR, touchdown PCR, and nested PCR approaches (Table 6.3). The nested PCR modification, which effectively amplified EaCoMT genes from all environmental samples, was performed as follows. The first round of PCR utilized the CoMF1L and CoMR2E primer set (0.2 µM each;

Table AI.7) as described previously [77] and 1 μL of DNA extract (containing 0.14 – 15.7 ng template) (Table 6.1). A subsequent round of nested PCR was performed with F131 and R562 primers (0.2 μM each; Table 6.1) [163] and 2 μL of the initial reaction mixture. The nested PCR thermocycler program consisted of an amplification phase (30 cycles of 94 °C for 20 s, 60 °C for 45 s, 72 °C for 30 s) and a final extension (72 °C for 15 min). Negative controls for the nested PCR, which used 2 μL of negative control reaction mixtures used in the first round of PCR, showed no amplification.

Amplification of EaCoMT genes from Carver DNA extracts required a touchdown phase during the first round of PCR. The thermocycling program consisted of an initial denaturation step (94°C, 5 min) followed by a touch-down phase (20 cycles of 94°C for 30 s, 65°C for 45 s (0.5°C decrease of each cycle), and 72°C for 1 min), a general amplification phase (10 cycles of 94°C for 30 s, 55°C for 45 s and 72°C for 1 min) and a final extension (72 °C for 15 min). Nested PCR was then performed as described above.

The potential that nested and touchdown PCR could introduce lower apparent gene diversity was investigated by constructing a pooled sample of DNA (referred as AUS39) extracted from five different monitoring wells at the Australia site (i.e. samples 39-1, 39-3, 39-6, 39-7 and 39-8). DNA from AUS39 was amplified with three methods: conventional PCR, nested PCR and a combined touchdown PCR-nested PCR. The amplicons generated were subjected to T-RFLP analysis, and diversity indices were calculated to compare these three methods.

Only nested PCR results were used for comparative analyses between sites. However, when constructing clone libraries both conventional and nested PCR amplicons

from Australia samples were used. We also performed combined touchdown-nested PCR on three additional samples (Soldotna MW6, Oceana MW18, and Oceana MW25) (Table 6.3) for PCR bias comparisons in clone libraries.

Table 6.3 EaCoMT clone sequences (139 in total) retrieved as a result of different PCR protocols attempted to amplify EaCoMT genes from environmental samples.

Site	Well No.	Direct PCR	Direct PCR + second amplification	Direct touchdown PCR	Nested PCR	Nested PCR with touchdown	Total clones sequenced
Carver, MA (CARV)	RB46D	-	-	-	-	TD CARV46 Clones 2-6, 8-10	8
	RB63I	-	-	-	-	TD CARV63 Clones 1-3,5-10	9
Soldotna, AK (SOLD)	MW6	-	-	-	SOLD6 Clones 1-7,9,10	TD SOLD6 Clones 1,2,5,6,8	14
	MW40	-	-	-	SOLD40 Clones 1,2,5,7,9,10	NA	6
NAS Oceana, VA (OCEA)	MW18	-	-	-	OCEA18 Clones 1-3, 8	TD OCEA18 Clone 2-7, 9,10	12
	MW25	-	-	-	OCEA25 Clones 1-8	TD OCEA25 Clone 1,2,4,5,7,8	14
Melbourne, Australia (AUS)	39IJ-1	COM AUS39-1 Clone 1-8	NA	NA	AUS39-1 Clone 1-8	NA	16
	39IJ-3	COM AUS39-3 Clone 1-4, 7-10	NA	NA	NA	NA	8
	39IJ-6	COM AUS39-6 Clone 1,2, 4-10	NA	NA	AUS39-6 Clone 1,3,5-8	NA	15
	39IJ-7	COM AUS39-7 Clone 1-6, 8-10	NA	NA	NA	NA	9
	39IJ-8	COM AUS39-8 Clone 1-6, 9	NA	NA	NA	NA	7
Kotzebue, AK (KOTZ)	MW10-01	Faint Band	-	NA	KOTZ01 Clones 1-5	NA	5
	MW10-03	Faint Band	-	NA	KOTZ03 Clones 1-4	NA	4
Fairbanks, AK (FAIR)	MW-4M	-	-	NA	FAIR4 Clones 1-6	NA	6
	MW-13M	-	-	NA	FAIR13 Clones 1-6	NA	6

Note: -: negative result; NA: not attempted.

6.3.4 Terminal-restriction fragment length polymorphism (T-RFLP) analysis

The restriction enzyme AcoI (EaeI) (New England BioLabs, Inc., Ipswich, MA) was selected for EaCoMT T-RFLP analysis using the default settings of the software program REPK [91]. This restriction enzyme maximized differentiation between EaCoMT sequences, based on analysis of a database populated with full-length EaCoMT sequences from VC- and ethene-assimilating isolates deposited in Genbank (Table 6.4).

All samples used in T-RFLP analysis were subjected to nested PCR with fluorescently-labeled 6-FAM F131 and unlabeled R562 primers. Digestions were performed in duplicate. PCR products (12 μL) were digested with AcoI (EaeI) and precipitated with glycogen, sodium acetate and ethanol following the manufacturer's protocol. The resulting terminal restriction fragments (T-RFs) were analyzed with an Applied Biosystems 3730 DNA analyzer with GeneScan 500 LIZ size standard at the Iowa Institute of Human Genetics at the University of Iowa. T-RF sizes were estimated using Peak Scanner Software (Applied Biosystems).

6.3.5 T-RFLP data analysis

T-RFLP data was further processed with T-REX software [164] using peak area to evaluate T-RF abundance, filter background noise and round fragment sizes to the nearest whole number. Fragment sizes <48 bp and >453 bp were excluded from further analysis. The final data matrix contained 28 samples (rows, duplicated for each well) from 14 groundwater sampling wells at six sites and 192 unique EaCoMT T-RFs (columns).

Table 6.4 Summary of REPK (4) predicted T-RF lengths from the products of F131/R562 primers (expected size 447-453bp) when subjected to in silico digestion with the AcoI (EaeI) restriction enzyme.

EaCoMT Group	EaCoMT gene source strain (Genbank accession no.)	Predicted T-RF (bp)
1	*Mycobacterium rhodesiae* strain JS60 (AY243034) *Mycobacterium* sp. strain JS621 (AY243039) *Ochrobactrum* sp. TD LBB Enrichment (DQ370436) *Ochrobactrum* sp. TD VC Enrichment (AY858985) *Pseudomonas putida* strain AJ LBB-grown (DQ370435) *Pseudomonas putida* strain AJ VC-grown (AY858984)	363
2	*Mycobacterium chubuense* strain NBB4 (GU174752)	453 (no cut)
3	*Mycobacterium gadium* strain JS616 (AY243036) *Mycobacterium mageritense* strain JS625 (AY243043) *Mycobacterium moriokaense* strain JS619 (AY243038) *Mycobacterium* sp. strain JS624 (AY243042) *Mycobacterium tusciae* strain JS617 (AY243037)	24
	Mycobacterium aichiense strain JS61 (AY243035)	48
	Mycobacterium rhodesiae strain JS622 (AY243040) *Mycobacterium* sp. strain JS623 (AY243041)	312
4	*Nocardioides* sp. strain JS614 (AY243042)	240
5	*Nocardioides* sp. strain JS614 *etnE*1 (CP000508)	356
6	*Gordonia rubripertincta* B-276(AF426826)	149
7	*Xanthobacter* Py2 2-hydroxypropyl-CoM lyase (CP000782)	73
8	*Haliea* sp. putative EaCoMT (AB691746)	151
9	*Pseudomonas putida* MetE (AF363277)	450

Note: Groups 1-8 were used to select the enzyme employed in T-RFLP analysis. The primary criterion for restriction enzyme selection was that it can distinguish EaCoMT genes from these different groups.

A T-RFLP profile clustered heatmap was generated using the gplots, vegan and RColorBrewer packages in R [47]. Briefly, this was achieved by calculating Bray-Curtis dissimilarities and using these values to execute the -hclust command in R, which performs the average linkage hierarchical clustering method. T-RFs with relative abundance <1% in at least one sample were removed from the heatmap to better visualize the results.

T-RFLP profiles were also analyzed by nonmetric multidimensional scaling (NMDS) in R with the vegan package [165], standardized with Wisconsin smoothing. Bray-Curtis dissimilarities were calculated with a random starting configuration and a two dimensional solution was reached. The final stress was 0.1402. T-RFLP profile composition differences between each sample were evaluated by multi-response permutation procedure (MRPP) using the vegan package in R, based on Euclidean dissimilarity and 999 permutations as default. The chance-corrected within group (i.e. each contaminated site) agreement was 0.5109, indicating that samples are homogenous within groups. The P-value was 0.001, indicating the differences in EaCoMT gene T-RFLP profiles among groups were statistically significant.

The Shannon-Wiener index (H') and inverse Simpson index (1/D) [166] were calculated with the vegan package in R for each sample using the T-RFLP results. The relative abundance of each T-RF was treated as the "number of species" in the analysis. Australia composite sample T-RFs amplified with conventional PCR and combined touchdown-nested PCR were also analyzed with diversity indices, with fragment sizes <48 bp and >453 bp excluded.

6.3.6 VC attenuation rate estimation and correlation analysis

Groundwater VC attenuation rates (i.e. the point decay rate (k_{point}) for each well and the bulk attenuation rate (k_{bulk}) over a plume transect) were calculated following a US EPA protocol [167], which is used under the assumption that VC attenuation follows a pseudo-first-order rate law [167] (examples in Figure. 6.1). Linear regressions of ln VC concentration vs. time and ln VC concentration vs. distance were performed in order to estimate the attenuation rate. The estimated rates were not corrected for dilution, dispersion, or sorption. Attenuation rates with a p-value less than 0.1 (two-tailed) were considered statistically significant and used for further analysis. Groundwater flow rates, VC concentrations and site maps used for the calculation were obtained as described in Table 6.1.

Linear regressions and Spearman's correlations were used to analyze the relationship between the abundance and diversity of EaCoMT genes and groundwater parameters (i.e. VC and ethene concentration, DO, temperature, pH and ORP) as well as estimated VC attenuation rates (k_{point} and k_{bulk}). The potential significance of each relationship was established based on a p-value<0.05 (two-tailed). Wells or sites without adequate information for rate estimation and correlation analysis were excluded from this analysis.

6.3.7 Cloning and sequencing

Purifed PCR products amplified from environmental samples with both conventional PCR and nested PCR were ligated overnight at 4°C into the pCR®2.1 vector (1:1 molar insert to vector ratio) using the Original TA Cloning Kit (Invitrogen).

Ligations were transformed into One Shot® TOP10 chemically competent *E. coli*. Transformants were analyzed according to the cloning kit instructions. Plasmids were extracted using QIAprep Spin Miniprep Kit. Clones were PCR-screened with M13F and M13R primers (Table AI.7) and those with the appropriately sized inserts were Sanger-sequenced at the Iowa Institute of Human Genetics with the M13F and/or M13R primers.

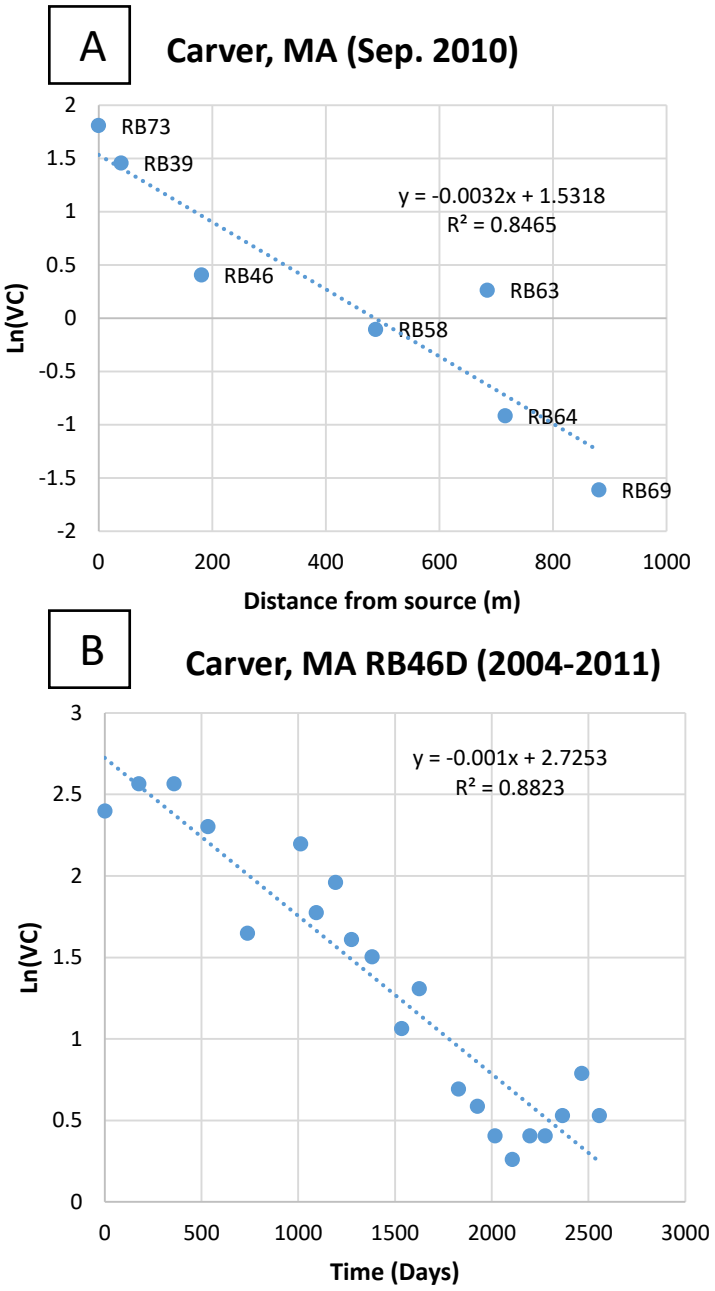


Figure 6.1 Examples of bulk VC attenuation rate (k_{bulk}) (A) and point VC decay rate (k_{point}) (B) calculations.

Note: To estimate the kbulk, a transect was developed across source zone and selected wells on the map, then the natural log of VC concentration from each well was plotted against the distance between each well and the source zone. Linear regression was performed and the kbulk is equal to slope times groundwater flow rate (Table 6.2). To estimate the kpoint, the ln VC concentration was plotted against time and a linear regression was performed. The kpoint is equal to the slope (unit conversions should be made if necessary). The retardation factor for VC was assumed to be 1.7.

6.3.8 Sequence analyses

Sequences with <99% identity to each other from the same sample were included in an amino acid sequence analysis as representatives. Deduced EtnE amino acid sequences (adjusted with ORF finder [168]) from partially sequenced clone were aligned with deduced EtnE (Table 6.5) and MetE (Table 6.6) sequences from Genbank using ClustalW [169] and trimmed to 166 aa (including gaps). MetE genes were included in the phylogenetic analysis to account for potential homology with EtnE. Phylogenetic trees were generated with MEGA5 using the maximum likelihood method [170]. The *Nocardioides* sp. URHA0032 MetE gene (Genbank accession WP_028637114) was used as the outgroup. A nucleotide phylogenetic tree was also constructed, with all 139 clone sequences obtained in this study plus existing EaCoMT gene sequences from Genbank. Nucleotide sequences were processed as described above for amino acid analysis except that the *Pseudomonas putida metE* [171] was used as the outgroup. The results were visualized with EvolView [172].

6.3.9 Genbank submissions

Representative unique sequences included in the phylogenetic trees were deposited in Genbank under accession numbers KR936138- KR936167.

Table 6.5 Genbank accession numbers of existing EaCoMT sequences used in phylogenetic and diversity analyses in this study.

Name	Accession Number
Carver MA Well RB63I-E01	GQ847819
Carver MA Well RB63I-E02	GQ847820
Carver MA Well RB63I-E03	GQ847821
Carver MA Well RB73-E05	GQ847813
Carver MA Well RB73-E06	GQ847814
Carver MA Well RB73-E10	GQ847817
Carver MA Well RB73-E11	GQ847818
Carver MA Well RB63I 75pc VC L4-1	KJ509931
Carver MA Well RB63I 75pc VC L4-4	KJ509932
Carver MA Well RB63I 75pc VC L4-10	KJ509933
Carver MA Well RB63I 75pc VC L9-1	KJ509934
Carver MA Well RB63I 75pc VC L9-4	KJ509935
Carver MA Well RB63I 75pc VC L9-8	KJ509936
Mycobacterium chubuense strain NBB4	GU174752
Nocardioides sp. strain JS614	AY772007
Nocardioides sp. strain JS614 *etnE*1	CP000508
Mycobacterium rhodesiae strain JS60	AY243034
Mycobacterium aichiense strain JS61	AY243035
Mycobacterium gadium strain JS616	AY243036
Mycobacterium tusciae strain JS617	AY243037
Mycobacterium moriokaense strain JS619	AY243038
Mycobacterium sp. strain JS621	AY243039
Mycobacterium rhodesiae strain JS622	AY243040
Mycobacterium sp. strain JS623	AY243041
Mycobacterium sp JS623 mutant 1	FJ602755
Mycobacterium sp. JS623 mutant 5	FJ602759
Mycobacterium sp. JS623 mutant 4	FJ602758
Mycobacterium sp. JS623 mutant 3	FJ602757
Mycobacterium sp. JS623 mutant 2	FJ602756
Mycobacterium sp.strain JS624	AY243042
Mycobacterium mageritense strain JS625	AY243043
Pseudomonas putida strain AJ VC Enrichment	AY858984
Pseudomonas putida strain AJ LBB Enrichment	DQ370435
Ochrobactrum sp. TD VC Enrichment	AY858985
Ochrobactrum sp. TD LBB Enrichment	DQ370436
Gordonia rubripertincta (*Rhodococcus rhodochrous*) B-276	AF426826
Haliea sp. ETY-M	AB691746
Xanthobacter Py2 2-hydroxypropyl-CoM lyase	CP000782

Table 6.6 Genbank accession numbers of MetE sequences used as outgroup sequences in phylogenetic tree construction.

Genus, Species and Strain	Genbank accession: genome location (Protein ID)	MetE Group
Mycobacterium tusciae JS617	NZ_KI912270.1:1983765-1984775 (WP_006245687.1)	1
Mycobacterium smegmatis JS623	CP003078.1:2183169-2184179 (AGB22566.1)	1
Mycobacterium chubuense NBB4	CP003053.1:1895804-1896820 (AFM16616.1)	1
Mycobacterium smegmatis str. mc^2-155	NC_018289.1:2441701-2442711 (WP_011728293.1)	1
Gordonia polyisoprenivorans VH2	CP003119.1:3444186-3445205 (AFA74057.1)	1
	CP003119.1:1951183-1952208 (AFA72765.1)	2
Pseudomonas putida	AF363277.1 (AAK29462.1)	2
Pseudomonas putida DOT-T1E	CP003734.1:2956795-2957880 (AFO48541.1)	2
Nocardioides sp. URHA0032	NZ_JIAV01000019.1:158602-159720 (WP_028637114.1)	2
Xanthomonas oryzae pv. oryzae PXO86	CP007166.1:476362-477390 (AJQ81599.1)	2
Xanthomonas citri subsp. citri strain UI7	CP008989.1:401372-402400 (AJY80461.1)	2

6.4 Results

6.4.1 T-RFLP analysis of EaCoMT gene diversity in environmental samples

A heatmap (Figure 6.2) displays EaCoMT gene T-RFLP profile patterns and clustering among different sites and monitoring wells. The heatmap reveals that several of the longer T-RFs were prominent among all groundwater samples, and some of them (e.g. the 314 bp, 354 bp, 364 bp and 448 bp T-RFs) matched T-RFs of EaCoMT clone sequences recovered from environmental samples (Table 6.7). However, none of these abundant T-RFs matched predicted T-RFs from *in silico* digestion of EaCoMT gene sequences previously deposited in Genbank (Table 6.4).

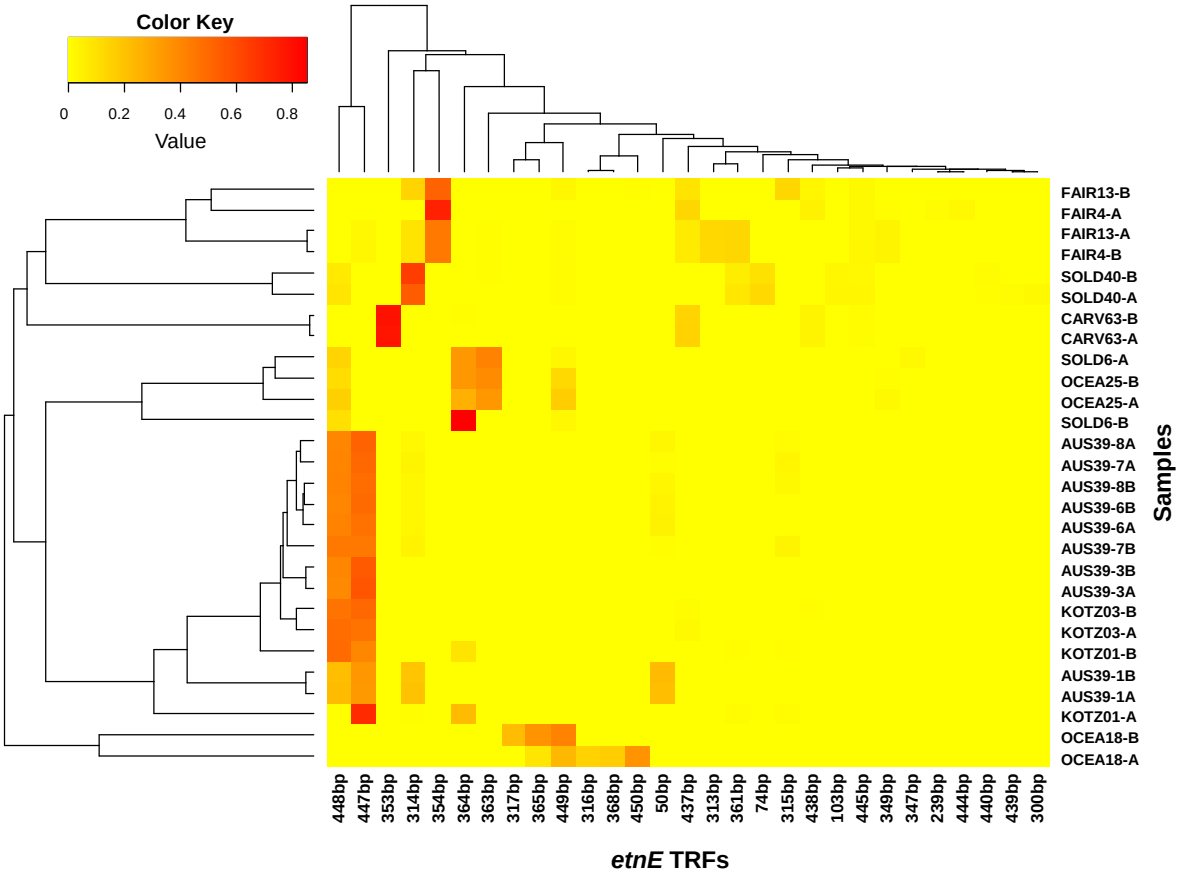


Figure 6.2 Clustered heatmap of T-RFLP profiles generated by AcoI (EaeI) restriction-digested partial EaCoMT genes.

Note: For clarity, T-RFs with relative abundance <1% were excluded from the graph. The higher the relative abundance of a particular T-RF in the sample, the warmer the coloring. Sample identifiers are formatted by site (using the first four characters of the name of each site), well number, and replicate (A or B). Please see Table 6.1 for specific site names.

Table 6.7 In-vitro and in-silico AcoI (EaeI) digestion of selected EaCoMT gene clones (those have <99% identity with other clone sequences retrieved from the same sample) matched with similar EaCoMT gene sequences from isolated strains.

Site	Well number	Clone ID	Observed T-RF (bp)	Predicted T-RF (bp)	Organism with closest BLAST hit (% identity)
Oceana NAS, VA	MW18	TD OCEA18 Clone 10	ND	24	*Mycobacterium tusciae* strain JS617 (99%)
Australia	039IJ-1	COM AUS39-1 Clone 4,7	50	48	*Mycobacterium gadium* strain JS616 (95%)
Australia	039IJ-6	AUS39-6 Clone 1	50	48	*Mycobacterium chubuense* NBB4 (92%)
Oceana NAS, VA	MW25	OCEA25 Clone 2	73	73	*Nocardioides* sp. JS614 (83%)
Oceana NAS, VA	MW25	OCEA25 Clone 4	104	103	*Nocardioides* sp. JS614 (83%)
Oceana NAS, VA	MW18	OCEA18 Clone 1,2	104	103	*Nocardioides* sp. JS614 (87%)
Oceana NAS, VA	MW25	OCEA25 Clone 1	313	312	*Nocardioides* sp. JS614 (85%)
Oceana NAS, VA	MW18	OCEA18 Clone 3	314	318	*Nocardioides* sp. JS614 (76%)
Kotzebue, AK	10-01	KOTZ01 Clone 1	314	312	*Mycobacterium* sp. JS624 (92%)
Australia	039IJ-1	COM AUS39-1 Clone 1,3	314	312	*Mycobacterium smegmatis* JS623 (94%)
Australia	039IJ-1	AUS39-1 Clone 4, AUS39-6 Clone 3	314	312	*Mycobacterium* sp. JS624 (93%-94%)
Fairbanks, AK	MW-4M	FAIR4 Clone 2	354	356	*Nocardioides* sp. JS614 (98%-99%)
Fairbanks, AK	MW-13M	FAIR13 Clone 3			
Carver, MA	RB46D	TD CARV46 Clone 3,10	364	363	*Mycobacterium rhodesiae* strain JS60 (99%-100%)
	RB63I	TD CARV 63 Clone 1,10			
Oceana NAS, VA	MW18	OCEA18 Clone 2			
	MW25	OCEA25 Clone 8			
Soldotna, AK	MW6	SOLD6 Clone 2,8			
Australia	039IJ-1	COM AUS39-1 Clone 2	448	453*	*Mycobacterium chubuense* NBB4 (99%)
Kotzebue, AK	10-01	KOTZ01 Clone 3			
	10-03	KOTZ03 Clone 2			
Soldotna, AK	MW6	SOLD6 Clone 1,3,6	440	447*	*Haliea* sp. ETY-M (76%-79%)
	MW40	SOLD40 Clone 1			

Note: The maximum T-RF size is 453 bp as PCR was performed with F131/R562 primers.

*No predicted AcoI (EaeI) restriction site. ND: not determined.

A NMDS analysis was also performed on the T-RFLP profiles (Figure 6.3). The NMDS analysis showed that EaCoMT genes from geographically distinct areas in some cases were similar (e.g. Australia and Kotzebue samples), but that EaCoMT genes recovered from geographically close locations (e.g. Alaska sites), did not necessarily cluster together. However, EaCoMT genes from Soldotna, AK well MW40 did group closely with EaCoMT genes from Fairbanks, AK.

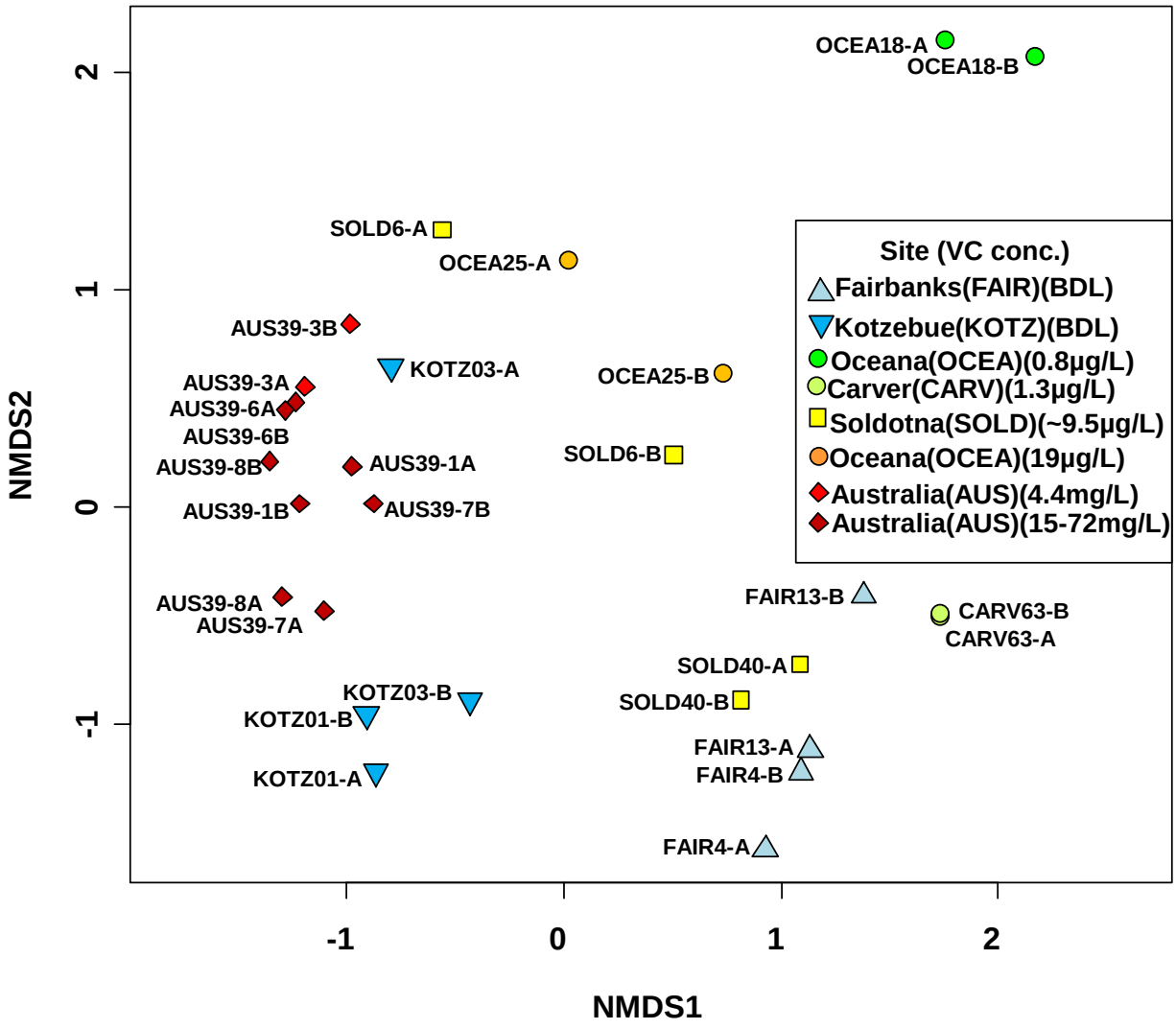


Figure 6.3 NMDS ordination of EaCoMT gene T-RFLP profiles from different groundwater samples. Sample identifiers are formatted by site, well number, and replicate (A or B). VC concentrations estimated in each of the wells is provided for reference. The final stress was 0.1402.

6.4.2 Phylogenetic analysis of EaCoMT gene diversity in environmental samples

A total of 139 sequences, 121 of which were unique (i.e. contained at least one bp difference from other sequences), were retrieved from clone libraries (Table 6.3). The nucleotide sequence identity of each clone to previously documented EaCoMT gene sequences found in isolates varied from 76%-99%. A nucleotide phylogenetic tree depicting the complete current EaCoMT gene database (Figure 6.4) revealed four potential EaCoMT gene subgroups (named according to the genus name of the first cultured representive in that group): *Mycobacterium, Nocardioides, Gordonia*, and *Xanthobacter*. Of the 121 unique EaCoMT sequences, 82 were related to EaCoMT genes found in isolated VC- and ethene-assimilating *Mycobacterium* strains (i.e. the *Mycobacterium* group). Notably, 11 of 12 EaCoMT sequences retrieved from the Fairbanks samples contained a 7-bp deletion at the 318 bp location of the 447 bp F131/R562 PCR product. These Fairbanks sequences were 98-99% identical to *etnE* allele in *Nocardioides* sp. strain JS614 designated *etnE1* [9]. The remaining EaCoMT sequence from Fairbanks was 95% identical to *Nocardioides* sp. strain JS614 *etnE*. All EaCoMT gene sequences from Soldotna MW40 contained a 1-bp deletion at the 12 bp location of the 447 bp PCR product. These Soldotna MW40 sequences were 76-78% identical to the putative EaCoMT gene sequence from the ethene-assimilating *Haliea* sp. ETY-M.

The maximum likelihood amino acid phylogenetic tree (Figure 6.5), constructed with 30 representative EaCoMT gene sequences (<99% identity with other clones from the same well) that did not contain frameshift deletions or internal stop codons and 19 EaCoMT sequences from isolated strains, was consistent with the nucleotide

phylogenetic tree. Of these 30 sequences, 40% formed a phylogenetic subgroup with *etnE* found in *Mycobacterium* strains and 6% grouped with the *etnE* from *Nocardioides* sp. strain JS614.

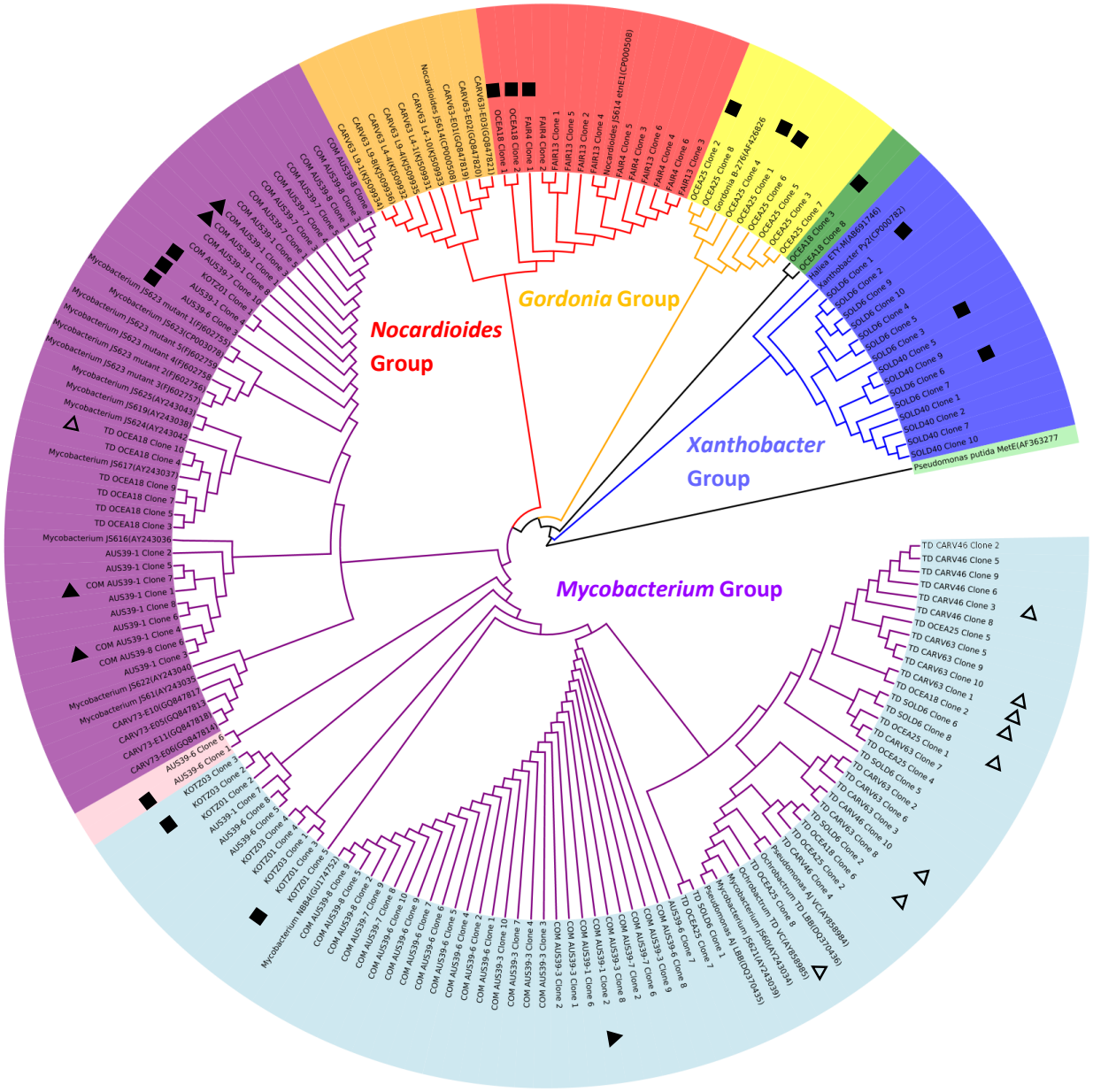


Figure 6.4 A phylogenetic tree depicting the relationship of deduced EaCoMT sequences from environmental samples in this study (Table 6.3), enrichment cultures and isolates from Genbank (Table 6.5).

Note: Genes are colored by group. The clones used for amino acid tree (Figure 6.5) were marked with symbols. The symbols refer to the PCR amplification method used: ▲ Direct PCR; ■ Nested PCR; △ Nested PCR with a touchdown modification. A total of 178 sequences were aligned with ClustalW [169] and trimmed to 464 bp (including gaps) in MEGA5 [170, 173]. The circle tree was constructed using the maximum likelihood [170] method with *Pseudomonas* putida MetE gene [171] as the outgroup. The circular tree was visualized using EvolView [172].

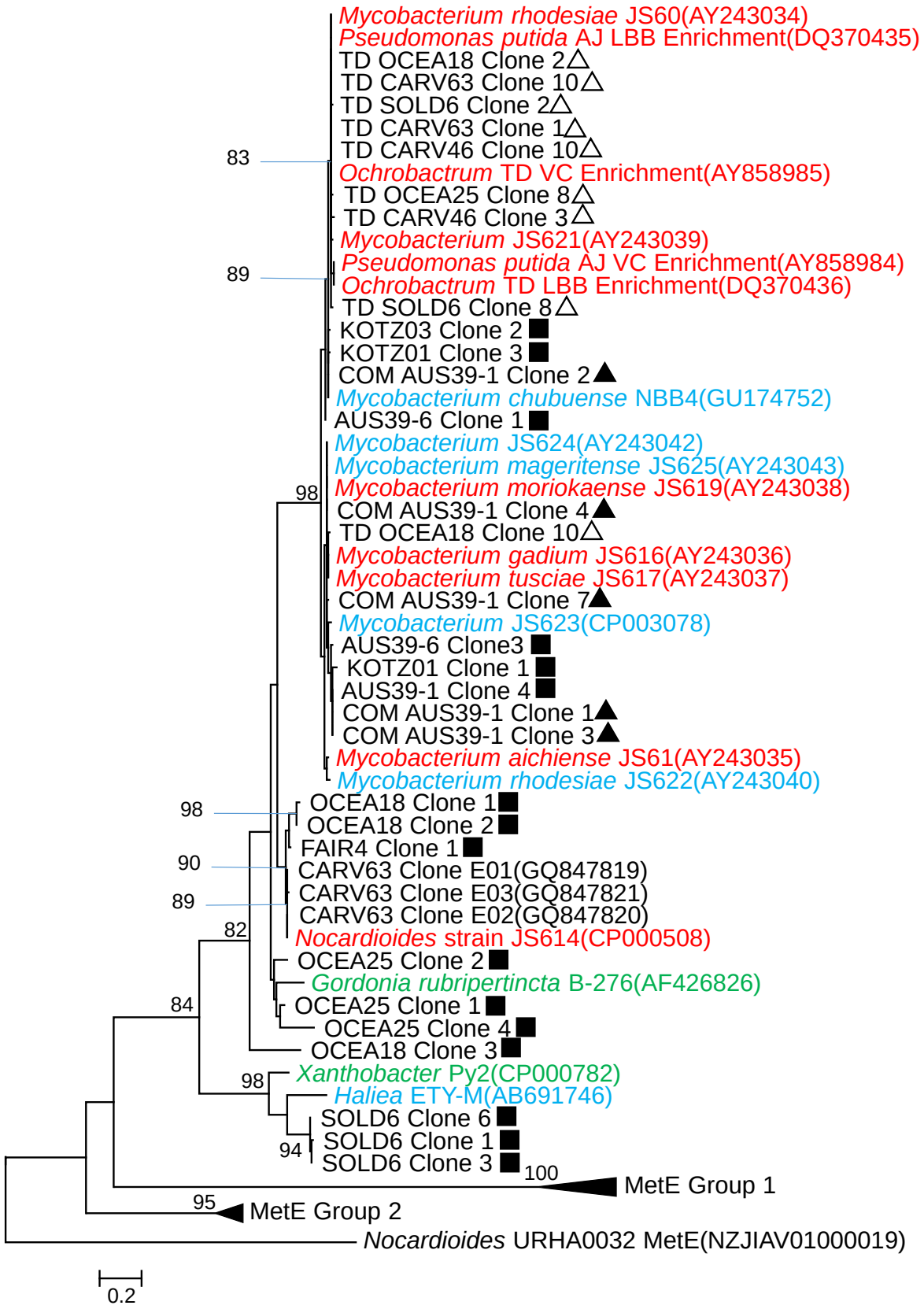

Mycobacterium rhodesiae JS60(AY243034)
Pseudomonas putida AJ LBB Enrichment(DQ370435)
TD OCEA18 Clone 2
TD CARV63 Clone 10
TD SOLD6 Clone 2
TD CARV63 Clone 1
TD CARV46 Clone 10
83
Ochrobactrum TD VC Enrichment(AY858985)
TD OCEA25 Clone 8
TD CARV46 Clone 3
Mycobacterium JS621(AY243039)
89
Pseudomonas putida AJ VC Enrichment(AY858984)
Ochrobactrum TD LBB Enrichment(DQ370436)
TD SOLD6 Clone 8
KOTZ03 Clone 2
KOTZ01 Clone 3
COM AUS39-1 Clone 2
Mycobacterium chubuense NBB4(GU174752)
AUS39-6 Clone 1
Mycobacterium JS624(AY243042)
Mycobacterium mageritense JS625(AY243043)
98
Mycobacterium moriokaense JS619(AY243038)
COM AUS39-1 Clone 4
TD OCEA18 Clone 10
Mycobacterium gadium JS616(AY243036)
Mycobacterium tusciae JS617(AY243037)
COM AUS39-1 Clone 7
Mycobacterium JS623(CP003078)
AUS39-6 Clone3
KOTZ01 Clone 1
AUS39-1 Clone 4
COM AUS39-1 Clone 1
COM AUS39-1 Clone 3
Mycobacterium aichiense JS61(AY243035)
Mycobacterium rhodesiae JS622(AY243040)
98
OCEA18 Clone 1
OCEA18 Clone 2
FAIR4 Clone 1
90
CARV63 Clone E01(GQ847819)
89
CARV63 Clone E03(GQ847821)
CARV63 Clone E02(GQ847820)
82
Nocardioides strain JS614(CP000508)
OCEA25 Clone 2
Gordonia rubripertincta B-276(AF426826)
84
OCEA25 Clone 1
OCEA25 Clone 4
OCEA18 Clone 3
98
Xanthobacter Py2(CP000782)
Haliea ETY-M(AB691746)
SOLD6 Clone 6
94
SOLD6 Clone 1
SOLD6 Clone 3
100
MetE Group 1
95
MetE Group 2
Nocardioides URHA0032 MetE(NZJIAV01000019)
0.2

Figure 6.5 A phylogenetic tree depicting the relationship of deduced EtnE amino acid sequences from environmental samples, enrichment cultures and isolates (Table 6.5), along with MetE sequences from related strains (Table 6.6).

Note: Isolates are color-coded: red = VC-assimilators; blue = etheneotrophs; green = propene-oxidizers. Carver sequences from Genbank were from ethene enrichment cultures [35]. Environmental samples are identified as described in Table 6.1. The symbols refer to the PCR amplification method used: ▲ Conventional PCR; ■ Nested PCR; △ Nested PCR with a touchdown modification. Amino acid sequences were deduced from partial EaCoMT sequences (excluding gaps). An alignment of 166 aa (including gaps) was generated in ClustalW [169] and the tree was constructed and visualized in MEGA5 using the Maximum Likelihood method [170] with *Nocardioides* URHA0032 MetE gene as the outgroup. The bar represents a 20% sequence difference. For environmental samples, only sequences <99% identity with other sequences from the same samples were included on the tree.

6.4.3 Relationships between EaCoMT gene diversity, EaCoMT gene abundance, and contaminated site conditions

Using the T-RFLP data, we quantified EaCoMT gene diversity in each groundwater sample by calculating Shannon-Wiener (H') and Inverse Simpson (1/D) diversity indices (Table 6.1). When considering T-RFs amplified with nested PCR only, Soldotna, Carver and Kotzebue samples showed the least EaCoMT gene diversity compared to other sites, with Carver well RB63I displaying the lowest EaCoMT gene diversity (H'=0.73±0.01, 1/D=1.55±0.02). Interestingly, the highest diversity was observed in Australia well 39-1 (H'=1.48±0.00, 1/D=3.93±0.01), which also contained the highest VC concentration (72 mg/L) across all the wells investigated in this study.

Quantification of EaCoMT gene abundance with qPCR confirmed the presence of EaCoMT genes in all samples included in this study (Table 6.1), ranging from 10^3-10^6 genes/L of groundwater. The highest EaCoMT gene abundances (4.7×10^4 - 8.7×10^6 genes/L of groundwater) were observed in the Australia samples, which were collected from a groundwater plume with high VC concentrations. Correlation analysis (Table 6.8) showed little evidence that the EaCoMT gene diversity has any significant relationship with VC attenuation rates, dissolved oxygen (DO), or other geochemical parameters. However, there were significant positive associations between VC concentration and EaCoMT gene abundance (Figure 6.6), both quantatively (linear correlation: $p<0.001$) and qualitatively (Spearman's rank correlation: $p<0.001$). There was also a siginificant linear correlation between ethene concentration and EaCoMT gene abundance ($p<0.001$) (Figure 6.6). Although there was no significant linear relationship between the bulk VC attenuation rate and EaCoMT gene abundance, we did notice a significant rank

correlation between these two variables (Spearman's correlation p=0.01).

Table 6.8 Correlation analysis between geochemical parameters with EaCoMT gene abundance and diversity.

Parameters	Correlation with EaCoMT gene Abundance (no. of genes/liter of GW)						Correlation with *etnE* Diversity						
		Linear Regression			Spearman				Linear Regression			Spearman	
	n	Slope	R^2	p-value	Rho	p-value	n	Index	Slope	R^2	p-value	Rho	p-value
Vinyl chloride concn (µg/L)	15	9.36×10^{1}	0.851	**<0.001**	0.699	**0.005**	14	H'	1.69×10^{-6}	0.019	0.641	0.107	0.715
								1/D	9.19×10^{-6}	0.086	0.309	0.189	0.514
Ethene concn (µg/L)	11	1.12×10^{4}	0.987	**<0.001**	0.859	**0.001**	9	H'	6.60×10^{-5}	0.006	0.850	-0.319	0.386
								1/D	9.00×10^{-4}	0.130	0.341	-0.151	0.682
Bulk VC attenuation rate (yr^{-1})	11	-2.57×10^{5}	0.004	0.853	0.636	**0.039**	10	H'	9.59×10^{-2}	0.058	0.503	-0.110	0.636
								1/D	-2.32×10^{-1}	0.040	0.580	-0.200	0.465
Point VC decay rate (yr^{-1})	6	3.33×10^{5}	0.134	0.475	0.203	0.700	5	H'	5.06×10^{-1}	0.152	0.516	0.359	0.567
								1/D	8.17×10^{-2}	0.059	0.693	0.359	0.567
DO concn (mg/liter)	14	-1.46×10^{6}	0.158	0.160	-0.257	0.374	13	H'	7.20×10^{-2}	0.027	0.590	0.170	0.579
								1/D	1.24×10^{-1}	0.012	0.721	0.104	0.737
ORP (millivolts)	9	-8.55×10^{2}	0.087	0.441	-0.350	0.359	8	H'	-2.10×10^{-3}	0.167	0.315	-0.262	0.536
								1/D	-5.90×10^{-3}	0.250	0.207	-0.286	0.501
pH	9	-4.98×10^{4}	0.024	0.693	-0.250	0.521	8	H'	8.73×10^{-2}	0.044	0.619	0.238	0.582
								1/D	3.94×10^{-1}	0.163	0.321	0.452	0.268
Temperature(°C)	9	-2.11×10^{4}	0.367	0.084	-0.683	**0.050**	8	H'	3.25×10^{-3}	0.006	0.852	0.167	0.703
								1/D	4.04×10^{-2}	0.177	0.300	0.405	0.327

Note: Bold values indicate relationships that are significant (two-tailed; p-value <0.05); n represents the number of wells available for the analysis.

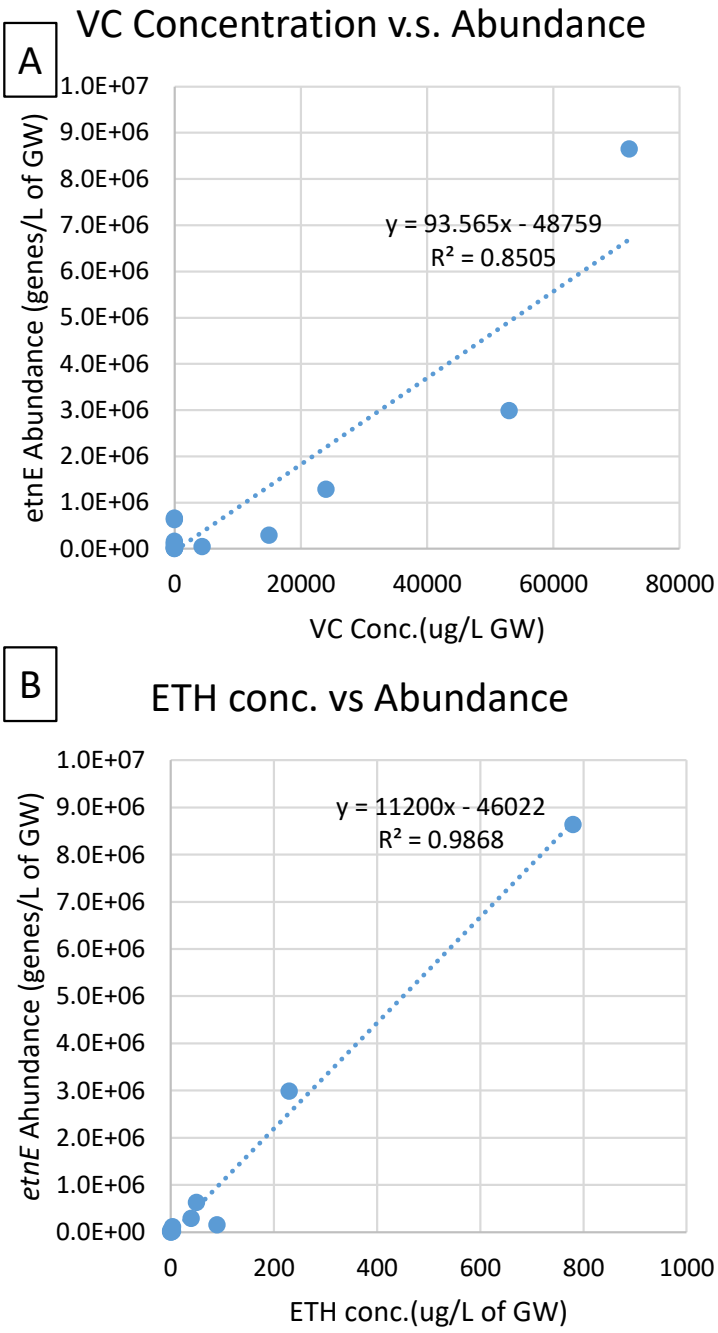


Figure 6.6 Linear regression of EaCoMT gene (*etnE*) abundance with vinyl chloride (VC) (A) and ethene (B) concentration. Each data point represents a sampling well from the sites surveyed in this study (Table 6.1).

6.4.4 PCR modifications affect EaCoMT gene diversity estimates

Because we could amplify EaCoMT genes from the Australia samples with conventationl PCR, we used the composite Australia DNA sample (AUS39) to assess potential bias introduced by nested and touchdown PCR modifications (Figure 6.7). The AUS39 sample, when amplified directly with COM primers, showed greater EaCoMT gene diversity (H'=2.04±0.00, 1/D=5.36±0.04) compared to when it was amplified by nested PCR (H'=1.21±0.41, 1/D=2.70±1.02). The touchdown-nested PCR-amplified AUS39 sample showed less EaCoMT gene diversity (H'=0.47±0.09, 1/D=1.41±0.13) compared to amplification by nested PCR alone.

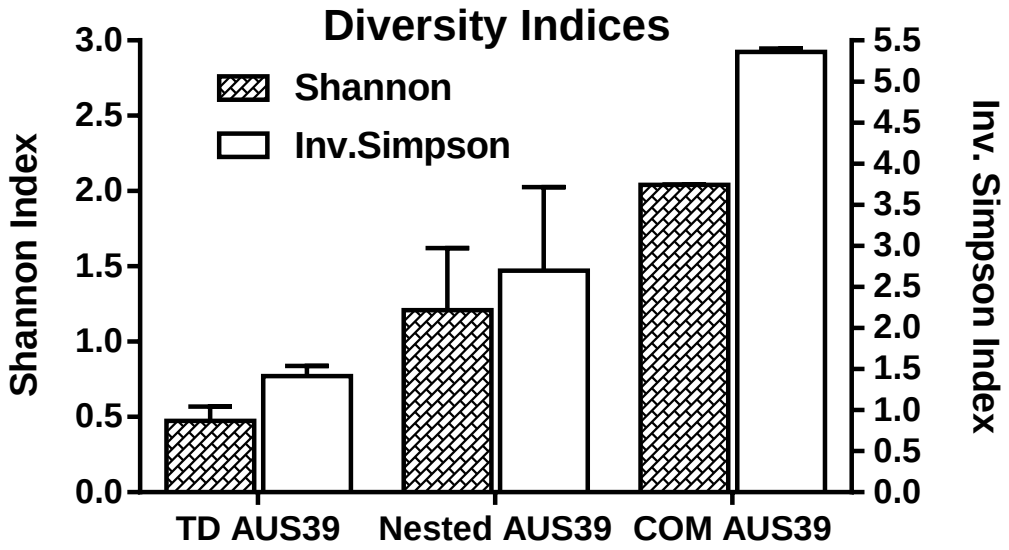


Figure 6.7 Analysis of potential bias introduced by touch down and nested PCR. A composite DNA sample from Australia (AUS39) was used for T-RFLP analysis of EaCoMT genes.

Note: Shannon-Wiener and inverse Simpson diversity indices [166] were calculated based on the T-RFLP profiles. Bar heights represent the average of duplicate T-RFLP analyses and the error bars represent the range. All the T-RFs <48 bp and >453 bp were excluded. The COM, Nested and TD designation indicates where the COM primer set, nested PCR and touchdown-nested PCR methods were used for T-RFLP.

To further assess the potential bias introduced by touchdown PCR, clone libraries were constructed with the touchdown-nested PCR-amplified SOLD6, OCEA18, and OCEA25 samples (Table 6.3). A total of 19 clones were sequenced from these clone libraries, and representative sequences were included in the phylogenetic analysis (Figure 6.4 and 6.5). All EaCoMT sequences retrieved by the touchdown modification grouped with those found in *Mycobacterium* isolates.

6.4 Discussion

This study has expanded our view of EaCoMT gene diversity. Previously, our understanding of EaCoMT gene diversity was based on 16 sequences from VC-, ethene-, and propene- assimilating isolates and a few EaCoMT genes sequenced from VC and ethene enrichment cultures [35, 65]. T-RFLP analysis has now revealed 192 different EaCoMT T-RFs. The 139 partial EaCoMT genes sequenced from groundwater samples were 76-99% identical to EaCoMT sequences found in isolates. These observations suggest that we have uncovered several novel EaCoMT sequences.

EaCoMT genes were present at all six sites surveyed in this study, even at sites where VC and ethene were not detected (Kotzebue and Fairbanks). The frequent occurrence of EaCoMT sequences (i.e. found in 112 metagenomes from 37 separate and geographically diverse sites; Table AI.9) in the MG-RAST metagenomics database [116] further supports the notion that EaCoMT genes are widespread and globally distributed among a variety of environments. Most of the EaCoMT sequences in these metagenomes were similar to EaCoMT genes found in *Mycobacterium* and *Nocardioides* strains (Table AI.9).

VC- and ethene-assimilating strains isolated to date are primarily members of the genus *Mycobacterium* [10], and the majority of environmental EaCoMT sequences recovered so far are similar to those found in *Mycobacterium* strains. This suggests that *Mycobacterium* strains are significant contributors to EaCoMT gene diversity in the environment. However, as the EaCoMT genes found in Mycobacteria are known to be plasmid-borne, it is possible that they are transferred to or have originated from other bacteria. Observing EaCoMT genes in VC-assimilating *Pseudomonas* and *Ochrobactrum* strains grouping with those in Mycobacteria supports this hypothesis (Figure 6.5).

The emerging clade of *Nocardioides* sp. JS614-like EaCoMT genes in the environment is notable. Although *Nocardioides* sp. JS614 is the sole isolated VC- and ethene-assimilating representative of the *Nocardioides* genus [10], a related *Nocardioides* strain was implicated as a dominant microbial community member of a VC-degrading enrichment culture likely using VC as a carbon and energy source [65]. Taken together, this data suggests that *Nocardioides* sp. could play a significant, yet currently underappreciated, role in VC- and ethene-assimilation in the environment.

Interestingly, 11 EaCoMT sequences obtained from the Fairbanks site which contained a 7-bp deletion grouped closely with the *etnE1* allele in *Nocardioides* sp. strain JS614 (Figure 6.4). The JS614 *etnE1* allele also contains a 7-bp deletion that would yield a frameshift mutation in the gene product and an expected loss of activity [133]. The JS614 *etnE1* allele appears to have resulted from a gene duplication event involving the functional EaCoMT gene found on the plasmid harbored by JS614 [133]. Inspection of Figure 6.2 reveals that additional EaCoMT sequences, presumably without deletions, were amplified from Fairbanks samples. It is possible that a related plasmid containing

two EaCoMT alleles was present in Fairbanks groundwater and the PCR primers used (COM-F1L/COM-R2E) preferentially amplify the EaCoMT allele with the 7-bp deletion. Further work is required to confirm this hypothesis.

Another emerging EaCoMT clade contains the 2-hydroxylpropyl CoM lyase from propene-assimilating *Xanthobacter* strain Py2. Until recently, strain Py2 was the only cultured representative harboring an EaCoMT gene from this group. However, the recently isolated marine ethene-assimilating strain *Haliea* ETY-M [19], also contains a putative EaCoMT. The newly discovered phylogenetic relationship between EaCoMT genes from *Xanthobacter* Py2 and *Haliea* ETY-M suggests that this group of EaCoMT genes could participate in ethene and VC biodegradation in contaminated groundwater, contrary to what has been proposed previously [174].

Environmental conditions that facilitate the apparently widespread occurrence of EaCoMT genes in environmental samples are not yet understood. The only known function for EaCoMT is the transfer of CoM onto epoxides of VC, ethene, and propene [12, 108]. The presence of VC and ethene in groundwater at relatively high concentrations in groundwater (e.g. VC>1 mg/L and/or ethene >230 µg/L at the Australia site) should promote plasmid-encoded EaCoMT gene maintenance as well as elevated EaCoMT gene abundance. This is supported by the strong positive correlations between alkene concentration (VC, ethene) and EaCoMT gene abundance in this study (Table 6.8). The presence of EaCoMT genes in groundwater could also facilitate natural attenuation of VC as long as other parameters aren't limiting (e.g. dissolved oxygen). The Spearman's correlation between EaCoMT gene abundance and the VC bulk attenuation rate supports this hypothesis (Table 6.8). Although there is a positive correlation between ethene and

VC concentration and EaCoMT gene abundance in the various monitoring wells included in this study, EaCoMT genes were found at the Fairbanks and Kotzebue sites (where VC and ethene were not detected). Additional research is required to determine why EaCoMT genes are maintained at sites such as these.

Alkene oxidation is an obligately aerobic process, however we observed no significant relationship between EaCoMT gene abundance and dissolved oxygen (DO). Aerobic VC oxidation can occur at very low DO levels (below 0.02 mg/L) [103] and VC-assimilating bacteria have been isolated from anaerobic groundwater[70]. The fact that we detected EaCoMT genes in low DO groundwater indicates that molecular oxygen either continuously enters the system or has been present in the groundwater in the past.

We did not observe any relationship between VC or ethene concentrations and EaCoMT gene diversity indices. This suggests that EaCoMT gene diversity patterns are currently not useful in site assessment for VC bioremediation. It is possible that the concentration of VC and/or ethene at a contaminated site could facilitate changes in EaCoMT gene diversity over time as microorganisms use these compounds as carbon and energy sources. In addition, we currently cannot differentiate between a VC-assimilator and VC-cometabolizer by analyzing EaCoMT sequences in environmental samples. Furthermore, the presence of EaCoMT genes in the environment that cannot produce an active EaCoMT (i.e. they contain 7-bp deletions or internal stop codons) complicates interpretation of EaCoMT sequence data from the environment. These and many other variables (e.g. CoM availability) could affect EaCoMT gene patterns in the environment. Future studies that address activities of different types of EaCoMT genes in the environment with respect to VC biodegradation rates could shed new light on this issue.

The PCR modifications employed in this study (nested and touchdown PCR) successfully amplified EaCoMT genes from environmental samples, but the results should be interpreted carefully with respect to PCR bias. The F131/R562 primers used in nested PCR were developed with *etnE* sequences from VC-assimilating *Mycobacterium* isolates [163], and are less degenerate than the CoMF1L/R2E primers, which were based on EaCoMT sequences from *Mycobacterium* strain JS60, *Gordonia* strain B-276 and *Xanthobacter* strain Py2. This was confirmed by diversity analysis, which showed lower EaCoMT diversity indices in the AUS39 composite sample T-RFLP data generated with nested PCR. However, since the nested PCR method was used on all the samples for TRFLP analysis, the results allow a logical comparison between sites.

Touchdown (TD) PCR will target genes that bind primers more specifically [175]. TD PCR with F131/R562 primers appeared to amplify more *Mycobacterium*-like *etnE*, as reflected by clone library analysis. TD PCR amplified *etnE* sequences also did not group with sequences obtained by nested PCR from the same samples (Figure 6.5), an observation which further supports the bias introduced by TD PCR. Our T-RFLP comparison experiment also showed that using TD PCR will underestimate gene diversity (Figure 6.7). However, a merit of TD PCR lies in its potential to reveal underrepresented genes in a sample.

Although the F131/R562 primers were successful when performing nested PCR, these primers do not amplify some important regions of EaCoMT genes. This includes the region where missense mutations (W243R, R257L) associated with adaptation to VC as a carbon and energy source observed in *Mycobacterium* sp. strain JS623 [72] and the His-X-Cys-X*n*-Cys zinc binding motif found in EaCoMTs and homologs [10, 154, 155].

Because EaCoMT and cobalamin (vitamin B12)-independent methionine synthase (MetE) are members of the same alkyl transferase family, many EaCoMT genes found in VC- and ethene-assimilating bacterial genome sequences (notably *Nocardioides* sp. strain JS614[110]) are incorrectly annotated as MetE genes in Genbank. Examples of inaccurate annotation are compiled and presented in Table 6.9. MetE shares functional homology but does not share significant nucleotide or amino acid with most EaCoMT sequences. In our phylogenetic analysis (Figure 6.5), the MetE sequences (Table 6.6) clearly grouped separately from translated EaCoMT sequences (bootstrap support = 62%). The analysis also indicates that the misannotated MetE sequences clearly group with other EaCoMT sequences.

6.5 Conclusion

The EaCoMT gene plays a critical role in assimilation of short-chain alkenes, such as ethene, VC, and propene. This is the first study that reports amplification, diversity analysis, correlation analysis and sequencing of EaCoMT genes from environmental samples. The EaCoMT gene database was significantly expanded, and potentially novel EaCoMT genes were discovered. Incorrectly annotated EaCoMT genes currently deposited in the database were also noted. These new sequences and insights will be useful in further developing environmental molecular diagnostic tools such as qPCR primers and probes, and will aid in the development and application of high-throughput sequencing approaches in future bioremediation studies.

The presence of VC and ethene in groundwater could help sustain EaCoMT gene pools in the environment. *Mycobacterium* and *Nocardioides*-like EaCoMT sequences

were the most widely distributed among the six sites investigated. The expanded clades of *Nocardioides*-like EaCoMT sequences and the discovery of *Haliea*-like EaCoMT sequences at the Soldotna site further suggest that potentially significant aerobic VC-degraders in the environment are not well represented in pure culture. Future research should aim to isolate and characterize aerobic VC-degrading bacteria from these under-represented groups.

Table 6.9 Examples of EaCoMT sequences in VC-, ethene- and propene-assimilating bacteria that are incorrectly annotated as MetE sequences in Genbank.

Genus, Species and Strain	Genbank accession: genome location (Protein ID)	Genbank annotation	Uniprot annotation (accession no.; % identity)	% identity (ID) with top EaCoMT BLAST hit
Mycobacterium sp. JS617	NZ_KI912270.1: 4898170-4899279 (WP_006247398.1)	methionine synthase (cobalamin-independent)	Epoxyalkane: coenzyme M transferase (B9VTP0; 97.6% to *Mycobacterium* sp. JS623)*	99% ID with JS617 EaCoMT gene (AY243037.1)
Mycobacterium sp. JS623	CP003078.1: 1868236-1869345 (AGB22266.1)		Epoxyalkane: coenzyme M transferase (B9VTP0; 100%)*	100% ID with JS623 EaCoMT gene (AY243041)
Mycobacterium chubuense strain NBB4	YP_006442905 (WP_014805821.1)		Epoxyalkane: coenzyme M transferase (D2K2D7; 100%)	94% ID with JS623 EaCoMT gene (ACM61851)
Nocardioides sp. JS614	CP000508.1: 134268-135380 (ABL79394.1)	methionine synthase (B12-independent) (plasmid)	Epoxyalkane: coenzyme M transferase (Q5U9J8; 100%)	100% ID with JS614 EaCoMT gene (AY772007.1)
	NC_008697.1: 134268-135380 (WP_011751519.1)	methionine synthase	Epoxyalkane: coenzyme M transferase (Q5U9J8; 100%)	100% ID with JS614 EaCoMT gene (AY772007.1)
Xanthobacter autotrophicus Py2 plasmid pXAUT01	CP000782.1: 81145-82275 (ABS70076.1)	methionine synthase (vitamin-B12 independent)	2-hydroxypropyl-CoM lyase (Q56837; 100%)	72% ID with putative *Haliea* EaCoMT gene (AB691746.2)
	CP000782.1: 266692-267822 (ABS70245.1)		2-hydroxypropyl-CoM lyase (Q56837l; 100%)	Apparent duplication of CP000782.1:81145-82275

***** These genes also had Uniprot hits to Methionine synthase II (Cobalamin independent) from *Mycobacterium* sp. strain JS623 (L0IWC5)

CHAPTER VII. ENGINEERING AND SCIENTIFIC SIGNIFICANCE

The works presented in this dissertation are the first studies to investigate ethene and VC-assimilating at the microbial community level, especially assisted by metagenomics. It is also the first time that genes involved in VC-assimilating are retrieved from environmental samples.

7.1 Distribution of VC-assimilators in the environment

In previous studies, *Nocardioides* sp. strain JS614 has been the only *Nocardioides* ethene- and VC-assimilating bacteria discovered. As most of the ethene- and VC-assimilating bacteria were *Mycobacterium*, they were thought to be predominantly distributed at sites contaminated with chlorinated solvents and are likely to be responsible for the natural attenuation of VC. The prevalence of *Mycobacterium*-like *etnE* gene sequences was observed in our study (Chapter VI). However, in enrichment culture studies (Chapter III and V), *Nocardioides* dominated in three separate enrichment cultures built with groundwater from two geographically distinct chlorinated ethene contaminated sites: Carver, MA and Fairbanks, AK, USA. Assisted by stable isotope probing (SIP), these *Nocardioides* were proved to be responsible for primary carbon up-take from VC. Further, in Chapter VI, the functional gene diversity survey across six contaminated sites found *Nocardioides*-like *etnE* sequences present at Carver, MA, Oceana, VA, and Fairbanks, AK (Chapter VI). The functional gene *etnE* appeared in 37 MG-RAST metagenomes, among which 21 have sequences similar to *etnE* gene in *Nocardioides* sp. strain JS614, representing samples from Antarctica, Asia, North and South America. Taken together, these evidences implied wide distribution of aerobic

ethene- and VC-degradation bacteria around the globe, not only at the chlorinated ethene contamination sites. *Nocardioides* is more ubiquitous in the environment than previously thought. From an engineering perspective, these indigenous bacteria are significant for in-situ biostimulation in remediation efforts.

7.2 A novel VC-assimilating *Nocardioides* strain

A novel ethene- and VC-assimilating bacterium, *Nocardioides* sp. strain XL1 was isolated (Chapter V) from the Fairbanks, AK groundwater enrichment culture. It is the second *Nocardioides* that encompasses the aerobic ethene- and VC-degradation pathway, with the functional genes also locating on megaplasmids (implicated from metagenome analysis). The genome and plasmid of strain XL1 is of 99% to 100% nucleotide similarity to previously found *Nocardioides* sp. strain JS614. Other than *Nocardioides* sp. strain XL1, the plasmid genome bin was also correlated with another *Nocardioides*-affiliated genome from the same culture, indicating horizontal gene transfer (HGT) might be involved. Collectively, it showed, again, that megaplasmids carry the genes in ethene- and VC-assimilating pathway, which could be passed around via HGT. However, more direct evidence is desirable, for example, isolation of the plasmid from *Nocardioides* sp. strain XL1 and other *Nocardioides*. In terms of engineering application, *Nocardioides* sp. strain XL1 could be a good candidate for bioaugmentation in VC-removal. In previous studies, *Nocardioides* sp. strain JS614 has demonstrated higher protein yield and faster degradation of VC, compared to other VC-assimilating *Mycobacterium*. However, it has an extended lag period after VC starvation. The newly isolated strain XL1 showed no obvious lag period after VC depletion in the culture. Further studies on the kinetics of

Nocardioides sp. strain XL1 degrading VC are needed.

7.3 VC-degradation in complex cultures

The real ecosystem is far more complicated than pure cultures. The mixed pure-culture experiment (Chapter IV) showed the VC-cometabolizing *Mycobacterium* strain JS622 grew on VC after mixed with VC-assimilating *Nocardioides* sp. strain JS614, indicating the interaction between bacteria could alternate their survival in mixed culture. The enrichment culture studies (Chapter III and V) have seen various bacteria present other than the primary VC-assimilators, which could contribute to the overall carbon, nitrogen and sulfur cycling. Particularly, *Proteobacteria*, especially *Pseudomonas* were found abundant in VC-assimilating cultures, however, it is not assimilating VC, as showed by SIP data. These proteobacteria possess glutathione S-transferase, which may also detoxify the VC-epoxides. It would be an interesting direction to further study their roles in VC-degradation. Overall, the removal of VC is due to the collective efforts of bacteria in a complex ecosystem.

7.4 Application of metagenomic analysis in bioremediation

Metagenomic sequencing is a powerful tool for investigation of the functional potential of complex ecosystems. We have made the first attempt to apply metagenomics analysis on VC-assimilating microbial community (Chapter V), which revealed alternatives for VC epoxide detoxification and deepened our understanding for VC degradation. However, in order to define a biological process, metagenomics must be complemented with in vivo experiments. The application of metagenomics in

bioremediation is still in a premature stage, largely due to insufficient annotation of genes and experimental data. The future of metagenomics lies in the improvement of the database, which requires collective efforts of microbiologists, bioinformatists and environmental scientists. Hopefully, metagenomics will eventually help to predict degradation potential at contaminated sites.

APPENDIX I: SUPPLEMENTAL MATERIAL

Supplemental Material for Chapter III

Table AI.1 Detailed qPCR information for Chapter III.

Primers and Samples	Range	Slope (Average ± St dev)	R^2 (Average ± St dev)	Amplification Efficiency (%) (Average ± St dev)
sedF-sedR Day 3	10^2-10^9	-2.77±0.39	0.98±0.02	129.62±27.5
sedF-sedR Day 7	10^2-10^9	-3.30±0.03	0.97±0.002	100.90±1.1
aquaF-aquaR Day 3	10^2-10^9	-3.16±0.005	0.99±0.002	107.36±0.2
aquaF-aquaR Day 7	10^2-10^9	-3.49±0.28	0.98±0.01	94.06±10.3
varF-varR Day 3	10^2-10^9	-3.16±0.31	0.96±0.03	108.55±15.3
varF-varR Day 7	10^2-10^9	-2.66±0.32	0.93±0.06	140.55±26.5
etnE Day 3	10-10^9	-3.29±0.23	0.99±0.001	101.95±10.1
etnE Day 7	10-10^9	-3.30±0.04	0.99±0.0001	100.91±1.4
etnC Day 3	10-10^8	-3.3	0.99	100.86
etnC Day 7	10-10^8	-2.95	0.99	117.72

Table AI.2 Detailed qPCR information for Chapter IV.

Degradation Point	Target Gene	Primer	Standard	Genbank Accession No. of Templates	Template	Fluorescence threshold (Auto Ct threshold)	Slope	PCR Efficiency	R^2	Y-intercept	Gene copies per ng template
~20 µmol	*etnE*	RTE	JS614 *etnE* PCR product	CP000508	JS614 gDNA	0.023	-3.41	96.61%	0.994	37.01	1.024×10^9
~50 µmol						0.082	-3.23	103.86%	0.998	37.41	1.024×10^9
~70 µmol						0.212	-3.03	113.90%	0.999	37.30	1.024×10^9

Note: qPCR of *etnE* on SIP fractions from ~20 µmol, ~50 µmol and ~70 µmol VC degradation time point genomic DNA. ABI 7000 System SDS Software (Applied Biosystems) was used to analyze the qPCR data. The "auto baseline" function was used in all situations. All fluorescent threshold numbers were adjusted manually. Optimized PCR efficiency was calculated by $E=(1-10^{(-1/slope)})$. All standards were between $10^3\times10^7$ gene copies.

Table AI.3 Primers used in Chapter V.

Name	Sequence (5'-3')	Product Size (bp)	Reference
16S Illumina sequencing primer			
515f	GTGYCAGCMGCC GCGGTAA	250-293	80
806r	GGACTACHVGGGTWTCTAAT		
qPCR primers			
RTE_F	CAGAAYGGCTGYGACATYATCCA	151	77
RTE_R	CSGGYGTRCCCGAGTAGTTWCC		
RTC_F	ACCCTGGTCGGTGTKSTYTC	106	
RTC_R	TCATGTAMGAGCCGACGAAGTC		
NocF	ATACGGGCAGACTAGAGGCA	113	This Study
NocR	CAGGACATGCCCAGAGAACC		
Pse435F	ACTTTAAGTTGGGAGGAAGGG	211-251	128
Pse686R	ACACAGGAAATTCCACCACCC		
PedF	AATGGAGGCAACTCTGAACCA	77	129
PedR	TCCCGAATAAAAGCAGTTTACGA		
SedF	CGGGCAGTTAAGTCAGTGGT	148	64
SedR	TGCCTTCGCAATAGGTGTTCT		
PCR primers			
27F	AGAGTTTGATCMTGGCTCAG	~1484	132
1492R	CGGTTACCTTGTTACGACTT		
rpoB_F	GCTTCGGGTTGAAGTAGTAGTTGT	802	36, 133
rpoB_R	GCAAAGATCACCGAACCTCTC		
JS614 etnC_F	GCGATGGAGAATGAGAAGGA	1138	77
JS614 etnC_R	TCCAGTCACAACCCTCACTG		
CoM_F1L	AACTACCCSAAYCCSCGCTGGTACGAC	891	9, 11
CoM_R2E	GTCGGCAGTTTCGGTGATCGTGCTCTTGAC		

Table AI.4 Classification of clones with RDP classifier.

Clone Sequence	Family	% Confidence of Family level	Genus	% Confidence of Genus level
Noc1M13F	*Nocardioidaceae*	100%	*Nocardioides*	99%
Noc2M13F	*Nocardioidaceae*	100%	*Nocardioides*	99%
Noc3M13F	*Nocardioidaceae*	100%	*Nocardioides*	99%
Noc4M13F	*Nocardioidaceae*	100%	*Nocardioides*	99%
Noc5M13F	*Nocardioidaceae*	100%	*Nocardioides*	99%
Noc6M13F	*Nocardioidaceae*	100%	*Nocardioides*	99%
Noc7M13F	*Nocardioidaceae*	100%	*Nocardioides*	99%
Noc8M13F	*Nocardioidaceae*	100%	*Nocardioides*	99%
Noc9M13F	*Nocardioidaceae*	100%	*Nocardioides*	99%
Pse61stM13F	*Pseudomonadaceae*	100%	*Pseudomonas*	97%
Pse91stM13F	*Pseudomonadaceae*	100%	*Pseudomonas*	97%
Pse1M13F	*Pseudomonadaceae*	100%	*Pseudomonas*	97%
Pse2M13F	*Pseudomonadaceae*	100%	*Pseudomonas*	97%
Pse3M13F	*Pseudomonadaceae*	100%	*Pseudomonas*	97%
Pse4M13F	*Pseudomonadaceae*	100%	*Pseudomonas*	92%
Pse5M13F	*Pseudomonadaceae*	100%	*Pseudomonas*	97%
Pse6M13F	*Pseudomonadaceae*	100%	*Pseudomonas*	97%
Ped1M13F	*Sphingobacteriaceae*	100%	*Pedobacter*	100%
Ped2M13F	*Sphingobacteriaceae*	100%	*Pedobacter*	100%
Ped3M13F	*Sphingobacteriaceae*	100%	*Pedobacter*	100%

Table AI.4 Continued.

Clone Sequence	Family	% Confidence of Family level	Genus	% Confidence of Genus level
Ped4M13F	*Sphingobacteriaceae*	100%	*Pedobacter*	100%
Ped5M13F	*Sphingobacteriaceae*	100%	*Pedobacter*	100%
Ped6M13F	*Sphingobacteriaceae*	100%	*Pedobacter*	100%
Ped7M13F	*Sphingobacteriaceae*	100%	*Pedobacter*	100%
Ped8M13F	*Sphingobacteriaceae*	100%	*Pedobacter*	100%
Ped9M13F	*Sphingobacteriaceae*	100%	*Pedobacter*	100%
Ped10M13F	*Sphingobacteriaceae*	100%	*Pedobacter*	100%
Sed1M13F	*Chitinophagaceae*	100%	*Asinibacterium*	43%
Sed2M13F	*Chitinophagaceae*	100%	*Asinibacterium*	43%
Sed3M13F	*Chitinophagaceae*	100%	*Asinibacterium*	44%
Sed4M13F	*Chitinophagaceae*	100%	*Asinibacterium*	43%
Sed5M13F	*Chitinophagaceae*	100%	*Asinibacterium*	43%
Sed6M13F	*Chitinophagaceae*	100%	*Asinibacterium*	43%
Sed7M13F	*Chitinophagaceae*	100%	*Asinibacterium*	43%
Sed8M13F	*Chitinophagaceae*	100%	*Asinibacterium*	44%
Sed9M13F	*Chitinophagaceae*	100%	*Asinibacterium*	44%
Sed10M13F	*Chitinophagaceae*	100%	*Asinibacterium*	43%

Table AI.5 Detailed qPCR information for Chapter V.

Sample	Target Gene	Fluorescence threshold	Slope	PCR Efficiency (%)	R^2	Y-intercept
Fairbanks Groundwater	*etnC*	0.09	-3.38	97.70	0.998	40.88
	etnE	0.12	3.28	101.60	0.996	36.79
Stable Isotope Probing cultures	16S rRNA gene of *Nocardioides*	0.25	-3.46	94.40	0.997	36.75
	16S rRNA gene of *Pseudomonas*	0.25	-3.40	97.03	0.998	35.51
	16S rRNA gene of *Pedobacter*	0.30	-3.50	93.16	0.993	37.95
	16S rRNA gene of *Sediminibacterium*	0.29	-3.35	98.88	0.986	36.26
RT-qPCR	*etnC*	0.58	-3.31	100.33	0.999	40.84
	etnE	0.58	-3.42	95.92	0.999	38.58
	Luciferase DNA	0.69	-3.23	104.16	0.999	36.06

Table AI.6 Composition of microbial communities as revealed by metagenomics sequencing and 16S rRNA gene amplicon Illumina sequencing in Chapter V.

Note: Numbers of sequences in metagenomes were calculated using MG-RAST and 16S rRNA gene were calculated using Mothur.

Groundwater				
	Metagenomics		16S	
Phyla	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Proteobacteria	236127	35.94%	31513	35.87%
Bacteroidetes	121647	18.51%	20283	23.09%
Firmicutes	72268	11.00%	5419	6.17%
Actinobacteria	70009	10.65%	4137	4.71%
Chloroflexi	20072	3.05%	1808	2.06%
Verrucomicrobia	16663	2.54%	3251	3.70%
Acidobacteria	14763	2.25%	4768	5.43%
Others	105509	16.06%	16674	18.98%
SUM	657058	100.00%	87853	100.00%
G1-TP1-ETH				
	Metagenomics		16S	
Phyla	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Actinobacteria	378723	46.37%	6347	7.72%
Proteobacteria	302134	36.99%	28263	34.39%
Bacteroidetes	91351	11.18%	44404	54.03%
Firmicutes	11560	1.42%	1958	2.38%
Others	32976	4.04%	1207	1.47%
SUM	816744	100.00%	82179	100.00%
G1-TP2-ETH				
	Metagenomics		16S	
Phyla	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Proteobacteria	364080	44.26%	26337	31.62%
Actinobacteria	241055	29.30%	4422	5.31%
Bacteroidetes	59622	7.25%	27156	32.60%
Firmicutes	47042	5.72%	3	0.00%
Acidobacteria	20690	2.52%	24168	29.01%
Others	90139	10.96%	1216	1.46%
SUM	822628	100.00%	83302	100.00%

Table AI.6 Continued.

G1-TP3-ETH				
	Metagenomics		16S	
Phyla	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Actinobacteria	373089	47.39%	3634	23.50%
Proteobacteria	238383	30.28%	6529	42.22%
Firmicutes	52731	6.70%	2	0.01%
Bacteroidetes	18161	2.31%	1297	8.39%
Acidobacteria	8859	1.13%	3562	23.03%
Others	96103	12.21%	441	2.85%
SUM	787326	100.00%	15465	100.00%
G1-TP2-VC				
	Metagenomics		16S	
Phyla	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Proteobacteria	212544	31.88%	20821	21.69%
Actinobacteria	190252	28.54%	6559	6.83%
Firmicutes	81869	12.28%	5	0.01%
Bacteroidetes	32759	4.91%	24979	26.02%
Chloroflexi	13084	1.96%	2	0.00%
Acidobacteria	12714	1.91%	32274	33.61%
Others	123452	18.52%	11373	11.85%
SUM	666674	100.00%	96013	100.00%
G1-TP3-VC				
	Metagenomics		16S	
Phyla	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Proteobacteria	472340	55.91%	86955	79.43%
Actinobacteria	298733	35.36%	5587	5.10%
Bacteroidetes	16950	2.01%	7998	7.31%
Firmicutes	16034	1.90%	13	0.01%
Acidobacteria	5070	0.60%	8521	7.78%
Others	35666	3.15%	395	0.36%
SUM	844793	100.00%	109469	100.00%

Table AI.6 Continued.

G2-TP1-ETH				
	Metagenomics		16S	
Phyla	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Proteobacteria	497577	60.55%	58707	64.76%
Actinobacteria	206760	25.16%	4759	5.25%
Bacteroidetes	61494	7.48%	22344	24.65%
Firmicutes	14605	1.78%	2816	3.11%
Others	41309	5.03%	2026	2.23%
SUM	821745	100.00%	90652	100.00%
G2-TP2-ETH				
	Metagenomics		16S	
Phyla	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Actinobacteria	496002	64.65%	9054	10.67%
Proteobacteria	184775	24.09%	32011	37.72%
Bacteroidetes	68985	8.99%	42462	50.03%
Firmicutes	3829	0.50%	13	0.02%
Others	13586	1.77%	1334	1.57%
SUM	767177	100.00%	84874	100.00%
G2-TP3-ETH				
	Metagenomics		16S	
Phyla	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Actinobacteria	768272	91.13%	35352	35.79%
Proteobacteria	38692	4.59%	14191	14.37%
Bacteroidetes	26273	3.12%	47069	47.65%
Others	9842	0.87%	2176	2.20%
SUM	843079	100.00%	98788	100.00%

Table AI.6 Continued.

G2-TP2-VC				
	Metagenomics		16S	
Phyla	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Actinobacteria	544662	68.97%	20893	17.73%
Proteobacteria	171393	21.70%	41285	35.03%
Bacteroidetes	56663	7.17%	54972	46.64%
Firmicutes	4239	0.54%	305	0.26%
Others	12777	1.62%	415	0.35%
SUM	789734	100.00%	117870	100.00%
G2-TP3-VC				
	Metagenomics		16S	
Phyla	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Actinobacteria	692349	85.73%	27677	26.44%
Proteobacteria	70919	8.78%	21941	20.96%
Bacteroidetes	32291	4.00%	52919	50.55%
Firmicutes	2687	0.33%	16	0.02%
Others	9315	1.15%	2134	2.04%
SUM	807561	100.00%	104687	100.00%
G2-TP2-ETH-Replicate				
	Metagenomics		16S	
Phyla	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Actinobacteria	532073	63.50%	10854	10.24%
Proteobacteria	210385	25.11%	37586	35.47%
Bacteroidetes	78021	9.31%	55928	52.78%
Firmicutes	4078	0.49%	16	0.02%
Others	13334	1.59%	1581	1.49%
SUM	837891	100.00%	105965	100.00%
G2-TP2-VC-R				
	Metagenomics		16S	
Phyla	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Actinobacteria	522866	69.31%	10601	13.16%
Proteobacteria	171274	22.70%	30079	37.34%
Bacteroidetes	43451	5.76%	39426	48.95%
Firmicutes	4192	0.56%	190	0.24%
Others	12606	1.67%	255	0.32%
SUM	754389	100.00%	80551	100.00%

Table AI.6 Continued.

Groundwater				
	Metagenomics		16S	
Genera	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Nocardioides	4682	0.71%	19	0.02%
Polaromonas	4293	0.65%	217	0.25%
Mesorhizobium	4733	0.72%	87	0.10%
Chitinophaga	8141	1.23%	23	0.03%
Bacteroides	24319	3.68%	0	0.00%
Geobacter	15486	2.34%	1202	1.39%
Clostridium	13770	2.08%	200	0.23%
Pedobacter	8996	1.36%	108	0.12%
Pseudomonas	7203	1.09%	994	1.15%
Burkholderia	7938	1.20%	160	0.18%
Aminobacter	1	0.00%	90	0.10%
Streptomyces	14198	2.15%	677	0.78%
Rhodococcus	3036	0.46%	28	0.03%
Unclassified	20121	3.04%	66371	76.74%
Others	524603	79.30%	16317	18.87%
SUM	661520	100.00%	86493	100.00%
G1-TP1-ETH				
	Metagenomics		16S	
Genera	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Nocardioides	347660	42.46%	6185	7.53%
Polaromonas	75127	9.18%	3217	3.91%
Mesorhizobium	39395	4.81%	144	0.18%
Chitinophaga	28164	3.44%	0	0.00%
Bradyrhizobium	17073	2.09%	5	0.01%
Acidovorax	13919	1.70%	0	0.00%
Sediminibacterium	0	0.00%	39697	48.31%
Pedobacter	8433	1.03%	829	1.01%
Pseudomonas	7702	0.94%	1503	1.83%
Burkholderia	6973	0.85%	0	0.00%
Aminobacter	11	0.00%	2972	3.62%
Streptomyces	5404	0.66%	0	0.00%
Rhodococcus	2319	0.28%	0	0.00%
Unclassified	8479	1.04%	7510	9.14%
Others	258158	31.53%	20113	24.48%
SUM	818817	100.00%	82175	100.00%

Table AI.6 Continued.

G1-TP2-ETH				
	Metagenomics		16S	
Genera	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Nocardioides	197915	23.99%	4343	5.21%
Polaromonas	28463	3.45%	10251	12.31%
Mesorhizobium	95239	11.54%	4	0.00%
Chitinophaga	13506	1.64%	0	0.00%
Bradyrhizobium	11356	1.38%	2	0.00%
Acidovorax	10170	1.23%	0	0.00%
Sediminibacterium	0	0.00%	26371	31.66%
Pedobacter	6923	0.84%	740	0.89%
Pseudomonas	15894	1.93%	5958	7.15%
Burkholderia	10710	1.30%	0	0.00%
Aminobacter	10	0.00%	7271	8.73%
Streptomyces	6440	0.78%	0	0.00%
Rhodococcus	2240	0.27%	3	0.00%
Unclassified	34420	4.17%	26447	31.75%
Others	391819	47.49%	1912	2.30%
SUM	825105	100.00%	83302	100.00%
G1-TP3-ETH				
	Metagenomics		16S	
Genera	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Nocardioides	311814	39.53%	3527	22.81%
Polaromonas	52359	6.64%	3405	22.02%
Mesorhizobium	29519	3.74%	19	0.12%
Chitinophaga	1738	0.22%	0	0.00%
Bradyrhizobium	5947	0.75%	1	0.01%
Acidovorax	6024	0.76%	0	0.00%
Sediminibacterium	0	0.00%	761	4.92%
Pedobacter	2080	0.26%	519	3.36%
Pseudomonas	4960	0.63%	321	2.08%
Burkholderia	6105	0.77%	0	0.00%
Aminobacter	2	0.00%	1189	7.69%
Streptomyces	7980	1.01%	0	0.00%
Rhodococcus	3275	0.42%	0	0.00%
Unclassified	45858	5.81%	4855	31.40%
Others	311190	39.45%	866	5.60%
SUM	788851	100.00%	15463	100.00%

Table AI.6 Continued.

G1-TP2-VC				
	Metagenomics		16S	
Genera	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Nocardioides	124106	18.57%	6367	6.63%
Polaromonas	15161	2.27%	4754	4.95%
Mesorhizobium	51840	7.76%	52	0.05%
Chitinophaga	5333	0.80%	0	0.00%
Bradyrhizobium	10864	1.63%	7	0.01%
Acidovorax	2694	0.40%	0	0.00%
Sediminibacterium	0	0.00%	22925	23.88%
Pedobacter	3078	0.46%	1949	2.03%
Pseudomonas	4269	0.64%	1428	1.49%
Burkholderia	5102	0.76%	0	0.00%
Aminobacter	9	0.00%	8419	8.77%
Streptomyces	7879	1.18%	0	0.00%
Rhodococcus	3253	0.49%	9	0.01%
Unclassified	70999	10.62%	46399	48.33%
Others	363810	54.43%	3700	3.85%
SUM	668397	100.00%	96009	100.00%
G1-TP3-VC				
	Metagenomics		16S	
Genera	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Nocardioides	153414	18.12%	3760	3.43%
Polaromonas	12767	1.51%	906	0.83%
Mesorhizobium	56905	6.72%	12	0.01%
Chitinophaga	3702	0.44%	0	0.00%
Bradyrhizobium	12195	1.44%	2	0.00%
Acidovorax	3702	0.44%	0	0.00%
Sediminibacterium	0	0.00%	7461	6.82%
Pedobacter	1994	0.24%	392	0.36%
Pseudomonas	245763	29.02%	77820	71.09%
Burkholderia	11424	1.35%	0	0.00%
Aminobacter	5	0.00%	2830	2.59%
Streptomyces	18707	2.21%	0	0.00%
Rhodococcus	17162	2.03%	1584	1.45%
Unclassified	14263	1.68%	11389	10.40%
Others	294877	34.82%	3313	3.03%
SUM	846880	100.00%	109469	100.00%

Table AI.6 Continued.

G2-TP1-ETH				
	Metagenomics		16S	
Genera	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Nocardioides	175253	21.26%	4559	5.03%
Polaromonas	13468	1.63%	564	0.62%
Mesorhizobium	90066	10.92%	68	0.08%
Chitinophaga	7660	0.93%	0	0.00%
Bradyrhizobium	53671	6.51%	40	0.04%
Acidovorax	34831	4.22%	1	0.00%
Sediminibacterium	0	0.00%	332	0.37%
Pedobacter	19815	2.40%	15891	17.54%
Pseudomonas	10672	1.29%	97	0.11%
Burkholderia	13138	1.59%	0	0.00%
Aminobacter	24	0.00%	5672	6.26%
Streptomyces	7223	0.88%	0	0.00%
Rhodococcus	1694	0.21%	34	0.04%
Unclassified	7438	0.90%	39674	43.79%
Others	389552	47.25%	23678	26.13%
SUM	824505	100.00%	90610	100.00%
G2-TP2-ETH				
	Metagenomics		16S	
Genera	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Nocardioides	407980	53.11%	8853	10.43%
Polaromonas	3200	0.42%	574	0.68%
Mesorhizobium	72432	9.43%	1	0.00%
Chitinophaga	5104	0.66%	0	0.00%
Bradyrhizobium	16736	2.18%	10	0.01%
Acidovorax	5840	0.76%	1	0.00%
Sediminibacterium	0	0.00%	7358	8.67%
Pedobacter	33726	4.39%	34762	40.96%
Pseudomonas	10313	1.34%	5383	6.34%
Burkholderia	2782	0.36%	0	0.00%
Aminobacter	5	0.00%	8023	9.45%
Streptomyces	13199	1.72%	0	0.00%
Rhodococcus	6443	0.84%	2	0.00%
Unclassified	2705	0.35%	17295	20.38%
Others	187650	24.43%	2612	3.08%
SUM	768115	100.00%	84874	100.00%

Table AI.6 Continued.

G2-TP3-ETH				
	Metagenomics		16S	
Genera	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Nocardioides	709961	84.20%	34709	35.14%
Polaromonas	2251	0.27%	1030	1.04%
Mycobacterium	7901	0.94%	0	0.00%
Mesorhizobium	3994	0.47%	8	0.01%
Bradyrhizobium	6902	0.82%	6	0.01%
Kribbella	8062	0.96%	1	0.00%
Sediminibacterium	0	0.00%	24634	24.94%
Pedobacter	8654	1.03%	22053	22.32%
Pseudomonas	2638	0.31%	2474	2.50%
Burkholderia	966	0.11%	0	0.00%
Aminobacter	0	0.00%	1092	1.11%
Streptomyces	10953	1.30%	0	0.00%
Rhodococcus	3880	0.46%	0	0.00%
Unclassified	1117	0.13%	8215	8.32%
Others	75943	9.01%	4564	4.62%
SUM	843222	100.00%	98786	100.00%
G2-TP2-VC				
	Metagenomics		16S	
Genera	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Nocardioides	410540	51.93%	20233	17.17%
Polaromonas	1916	0.24%	1026	0.87%
Mesorhizobium	35244	4.46%	1	0.00%
Mycobacterium	10588	1.34%	0	0.00%
Bradyrhizobium	28164	3.56%	15	0.01%
Kribbella	10124	1.28%	2	0.00%
Sediminibacterium	0	0.00%	1255	1.06%
Pedobacter	32260	4.08%	53152	45.10%
Pseudomonas	22602	2.86%	25641	21.75%
Burkholderia	3628	0.46%	0	0.00%
Aminobacter	6	0.00%	2268	1.92%
Streptomyces	18069	2.29%	0	0.00%
Rhodococcus	10585	1.34%	91	0.08%
Unclassified	3262	0.41%	7573	6.43%
Others	203513	25.74%	6609	5.61%
SUM	790501	100.00%	117866	100.00%

Table AI.6 Continued.

G2-TP3-VC				
	Metagenomics		16S	
Genera	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Nocardioides	633181	78.38%	27115	25.90%
Polaromonas	1339	0.17%	662	0.63%
Mesorhizobium	9628	1.19%	6	0.01%
Mycobacterium	6927	0.86%	0	0.00%
Bradyrhizobium	14882	1.84%	17	0.02%
Kribbella	7260	0.90%	0	0.00%
Sediminibacterium	0	0.00%	32806	31.34%
Pedobacter	9343	1.16%	19767	18.88%
Pseudomonas	4805	0.59%	6533	6.24%
Burkholderia	1479	0.18%	0	0.00%
Aminobacter	0	0.00%	3403	3.25%
Streptomyces	9655	1.20%	0	0.00%
Rhodococcus	4186	0.52%	49	0.05%
Unclassified	1827	0.23%	7753	7.41%
Others	103355	12.79%	6574	6.28%
SUM	807867	100.00%	104685	100.00%
G2-TP2-ETH-Replicate				
	Metagenomics		16S	
Genera	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Nocardioides	439875	52.44%	10608	10.01%
Polaromonas	3295	0.39%	626	0.59%
Mesorhizobium	91103	10.86%	0	0.00%
Mycobacterium	9721	1.16%	0	0.00%
Bradyrhizobium	20390	2.43%	9	0.01%
Kribbella	7266	0.87%	0	0.00%
Sediminibacterium	0	0.00%	9775	9.22%
Pedobacter	41663	4.97%	45606	43.04%
Pseudomonas	11636	1.39%	6856	6.47%
Burkholderia	2860	0.34%	0	0.00%
Aminobacter	15	0.00%	7881	7.44%
Streptomyces	12924	1.54%	0	0.00%
Rhodococcus	6531	0.78%	4	0.00%
Unclassified	2919	0.35%	21348	20.15%
Others	188646	22.49%	3252	3.07%
SUM	838844	100.00%	105965	100.00%

Table AI.6 Continued.

G2-TP2-VC-Replicate				
	Metagenomics		16S	
Genera	No. of Sequences	Relative Abundance	No. of Sequences	Relative Abundance
Nocardioides	396512	52.51%	10403	12.92%
Polaromonas	5956	0.79%	365	0.45%
Mesorhizobium	35009	4.64%	0	0.00%
Mycobacterium	11562	1.53%	0	0.00%
Bradyrhizobium	31382	4.16%	13	0.02%
Kribbella	9828	1.30%	0	0.00%
Sediminibacterium	0	0.00%	749	0.93%
Pedobacter	22385	2.96%	38297	47.55%
Pseudomonas	22536	2.98%	19047	23.65%
Burkholderia	0	0.00%	0	0.00%
Aminobacter	2	0.00%	1954	2.43%
Streptomyces	20735	2.75%	0	0.00%
Rhodococcus	9156	1.21%	34	0.04%
Unclassified	2537	0.34%	5762	7.15%
Others	187560	24.84%	3921	4.87%
SUM	755160	100.00%	80545	100.00%

Figure AI.1 qPCR of 16S rRNA gene using specific primers on SIP fractions to target: A) *Nocardioides*; B) *Pseudomonas*; C) *Pedobacter*; D) *Sediminibacterium*. Two time points were selected for the SIP analysis: day 3 and day 7.

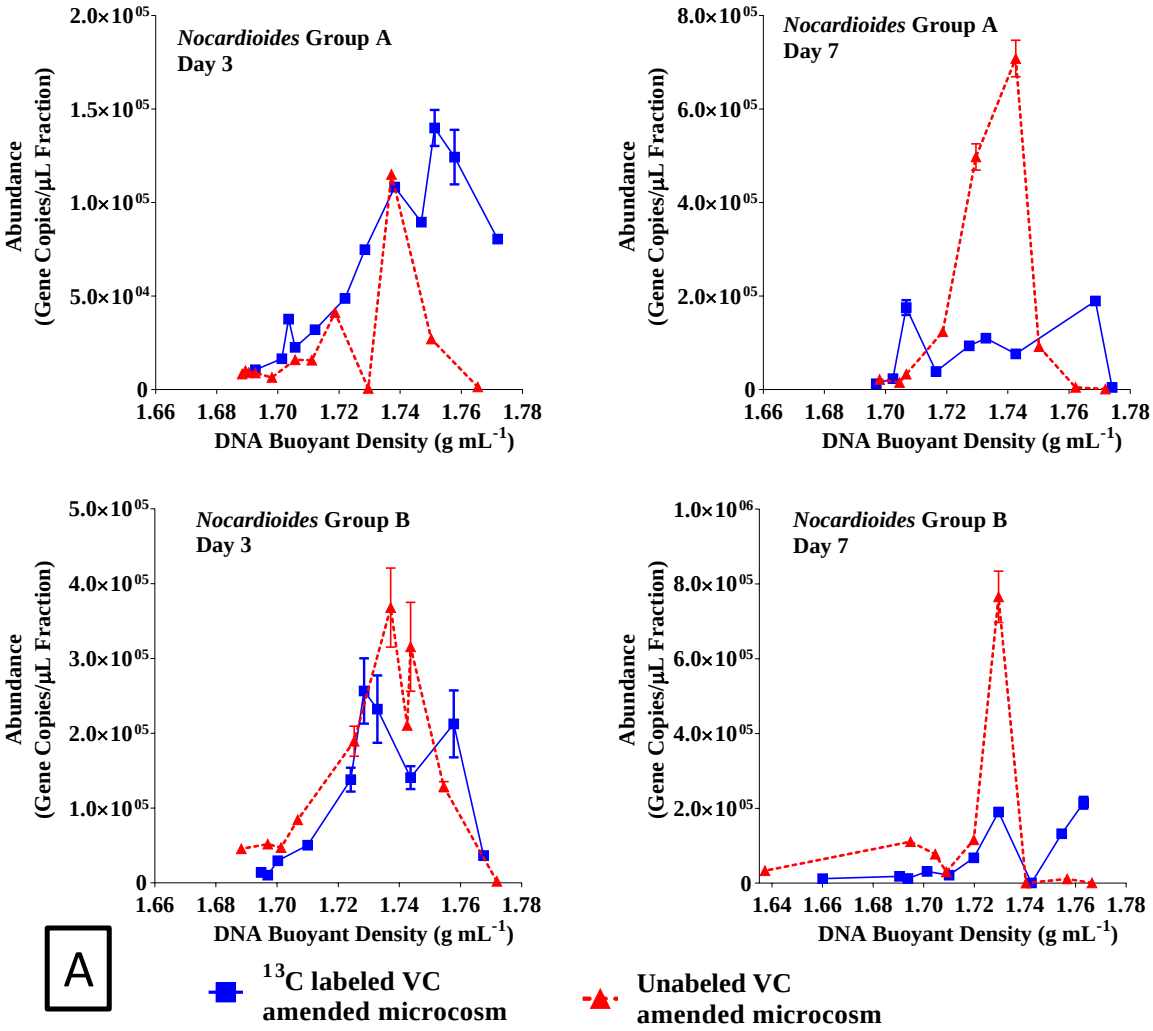


Figure AI.1 Continued.

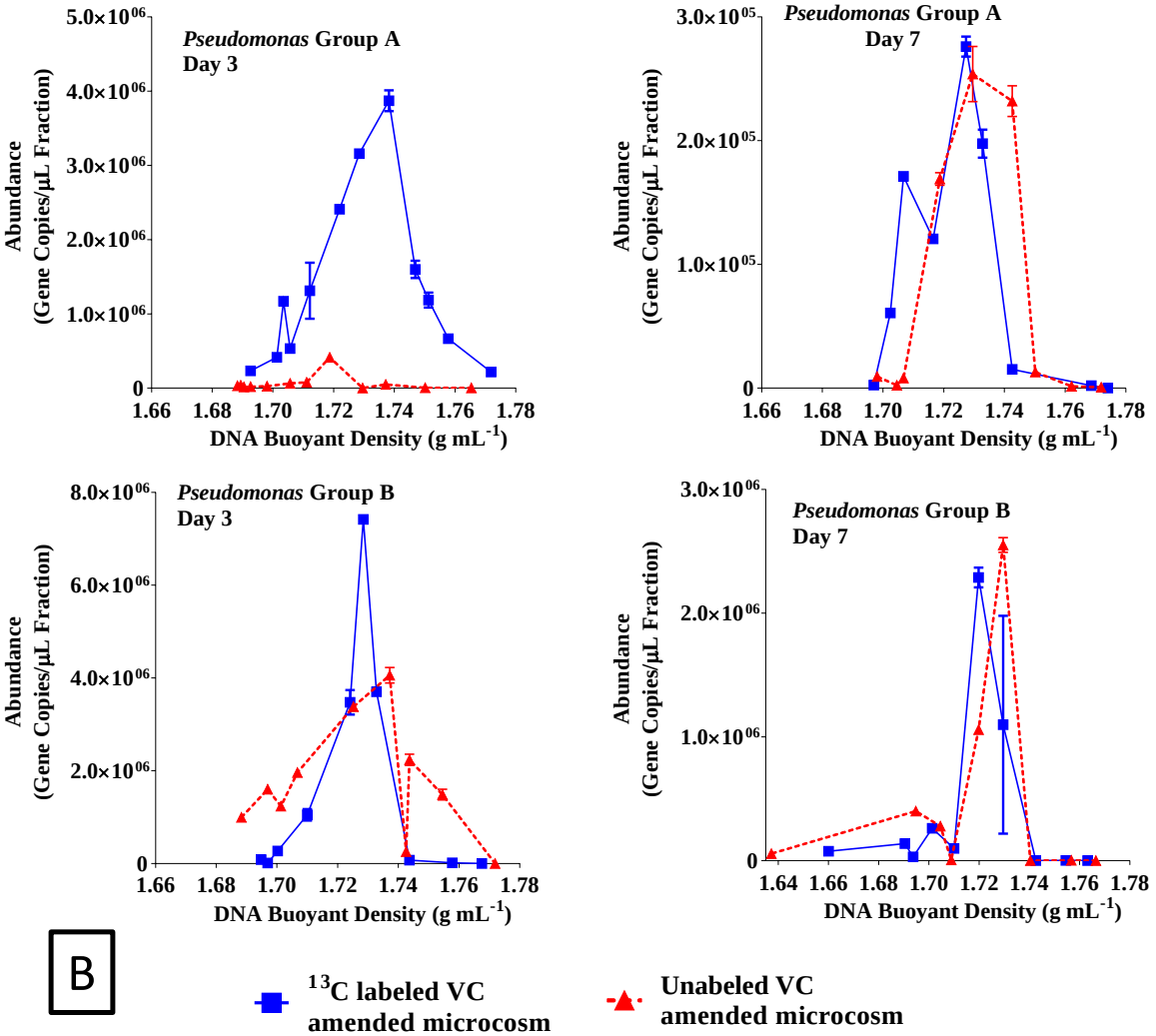


Figure AI.1 Continued.

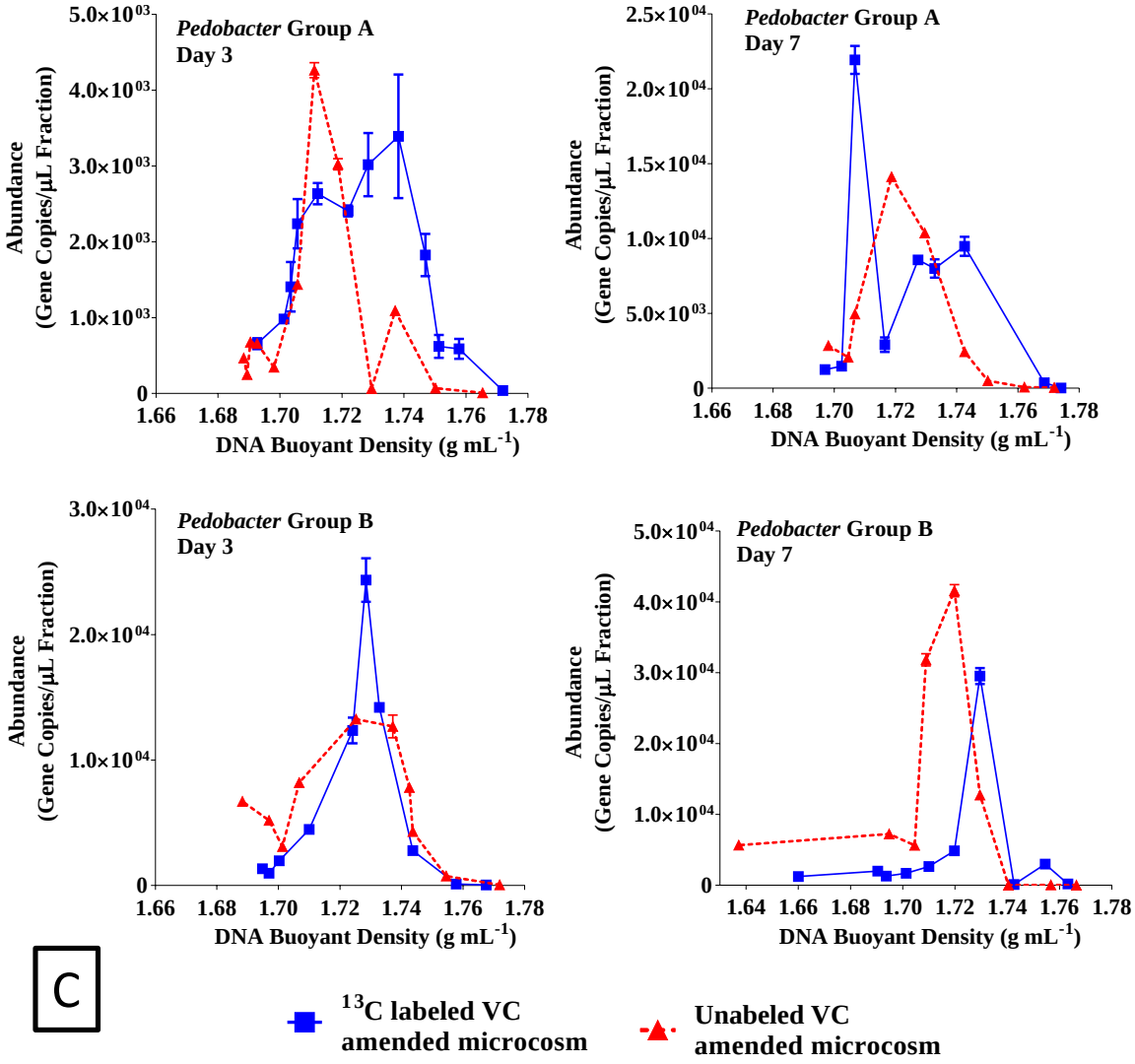


Figure AI.1 Continued.

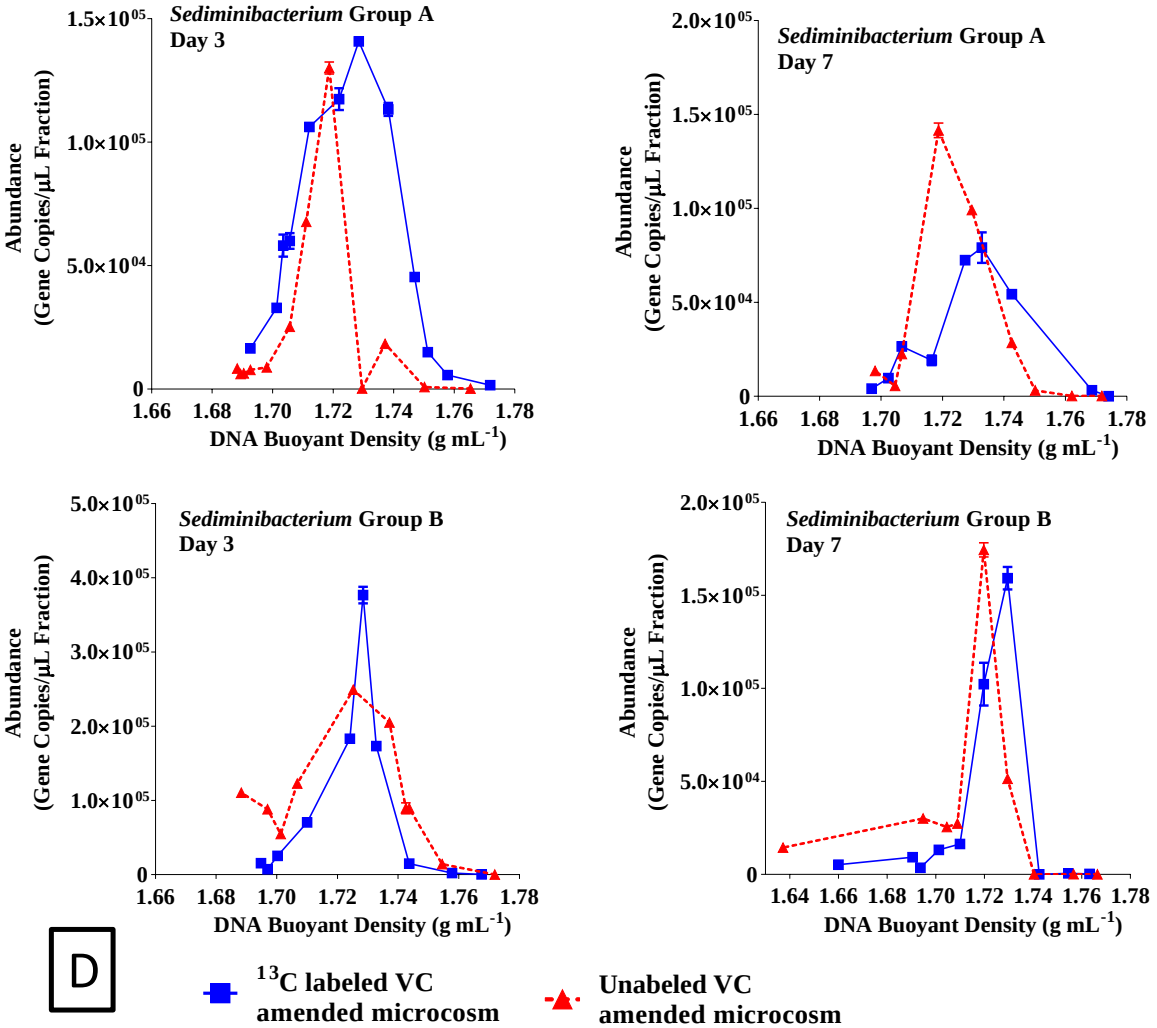


Table AI.7 Oligonucleotide primers used in Chapter VI.

Primer name	Sequence (5'-3')	Product size (bp)	Reference
COM-F1L	AACTACCCSAAYCCSCGCTGGTACGAC	834-891	[11]
COM-R2E	GTCGGCAGTTTCGGTGATCGTGCTCTTGAC		
RTE F	CAGAAYGGCTGYGACATYATCCA	151	[77]
RTE R	CSGGYGTRCCCGAGTAGTTWCC		
F131	GGAAGCGTTCGARGAYGCSGT	447-453	[163]
R562	TAGGGCCAGACRAACTCGTCGA		
M13F	GTAAAACGACGGCCAG	169-199	Invitrogen
M13R	CAGGAAACAGCTATGAC		

Table AI.8 Information of qPCR parameters in this study.

Sample	Target Gene	Fluorescence threshold	Slope	PCR Efficiency (%)	R^2	Y-intercept
Australia (AUS)	*etnE*	0.089	-3.31	100.43	0.993	38.72
Kotzebue (KOTZ)	*etnE*	0.123	-3.28	101.6	0.996	36.79
Fairbanks (FAIR)	*etnE*	0.123	-3.28	101.6	0.996	36.79

Note: All *etnE* genes were amplified with RTE primers. *Nocardioides* sp. strain JS614 (Genbank Accession No. CP000508) *etnE* PCR products amplified with CoM primers were used as standards. Standards were amplified from *Nocardioides* sp. strain JS614 genomic DNA. There are 1.025×10^9 genes per *ng etnE* template. Kotzebue and Fairbanks *etnC* and *etnE* were run on the sample plate.

Table AI.9 Summary of EaCoMT sequences found in the MG-RAST metagenomics database by July 2015. MG-RAST metagenome numbers are provided, along with the sample source and location, the top BLAST hit, abundance of each EaCoMT gene type in the sample, % identity and average amino acid alignment length.

Metagenome	Sample Source	Location	Top Hit in BLAST	Abundance	% ID	Avg. Align Length (aa)
4508942.3	Agricultural soil	Richmond, Indiana, USA	epoxyalkane coenzyme M transferase mutant 1 *Mycobacterium* sp. JS623	1	85.19	54
			epoxyalkane coenzyme M transferase, uncultured bacteria	1	74.14	58
4516651.3	Air	Beijing, China	epoxyalkane: coenzyme M transferase *Nocardioides* sp. JS614	124	84.47	26.483
			epoxyalkane: coenzyme M transferase *Rhodococcus rhodochrous*	51	86.245	27.72
			epoxyalkane: coenzyme M transferase *Mycobacterium chubuense* NBB4	14	83.062	22.462
			epoxyalkane coenzyme M transferase, uncultured bacteria (JS614-like)	10	84.458	23
			epoxyalkane coenzyme M transferase, uncultured bacteria (JS614-like)	9	71.18	24.625
			epoxyalkane coenzyme M transferase, uncultured bacteria (JS614-like)	8	83.664	27.231
			epoxyalkane coenzyme M transferase, uncultured bacteria (JS623-like)	5	77.66	23.556
			epoxyalkane coenzyme M transferase mutant 1 *Mycobacterium* sp. JS623	7	84.549	23.714
			putative epoxyalkane: coenzyme M transferase *Taylorella equigenitalis* MCE9	75	78.742	25.232
			putative epoxyalkane: coenzyme M transferase *Corynebacterium aurimucosum* ATCC 700975	70	82.491	26.886

Table AI.9 Continued.

4497384.3	Bulk Soil	Anazonia, Brazil	2-hydroxypropyl-CoM lyase (EC 4.4.1.23) (Epoxyalkane: CoM transferase) (EaCoMT) (Aliphatic epoxide carboxylation component I), *Bacillus tusciae* DSM 2912	1	80	15
4537093.3	Marine Sediment	Santa Barbara, CA, USA	epoxyalkane coenzyme M transferase, uncultured bacteria (JS623 and JS614-like)	14	66.61	32.17
4520078.3	Livestock body fluid	Ellinbank, Victoria, Australia	2-hydroxypropyl-CoM lyase (EC 4.4.1.23) (Epoxyalkane: CoM transferase) (EaCoMT) (Aliphatic epoxide carboxylation component I), *Bacillus tusciae* DSM 2912	1	84.62	13
4470378.3	Permafrost Soil (DNA)	Active Layer	Epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623 and JS614-like)	13	72.9	31.97
			Epoxyalkane: coenzyme M transferase mutant 3 , *Mycobacterium* sp. JS623	1	60	30
			Epoxyalkane: coenzyme M transferase mutant 1 *Mycobacterium* sp. JS623	5	75.51	33.8
			Epoxyalkane: coenzyme M transferase mutant 5 , *Mycobacterium* sp. JS623	1	72.97	37
4491382.3	Permafrost Soil (RNA)	Active Layer	Epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	4	73.29	37.25
			Epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	2	69.4	42
			Epoxyalkane: coenzyme M transferase mutant 1 *Mycobacterium* sp. JS623	1	78.26	46
4470381.3	Permafrost Soil	Thermokarst Bog	Epoxyalkane: coenzyme M transferase, *Mycobacterium* sp. JS623	2	66.115	32
			Epoxyalkane: coenzyme M transferase mutant 1 *Mycobacterium* sp. JS623	1	63.89	39

Table AI.9 Continued.

4477875.3	Bulk Soil	Misiones, Argentina	Epoxyalkane: coenzyme M transferase mutant 1 *Mycobacterium* sp. JS623	1	62.5	24
4524575.3	Rhizosphere Soil	Golm, Germany	Epoxyalkane: coenzyme M transferase mutant 1 *Mycobacterium* sp. JS623	1	82.98	47
			Epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	2	71.7	50.5
4480863.3	Anaerobic digester sludge	Ithaca, New York, USA	epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	2	66.03	20.5
4481985.3	River Sediment	Southeastern Montana, USA	epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	1	84.62	13
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	3	80.77	16.33
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	1	71.43	21
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	1	66.67	21
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	2	73.43	23
4541641.3	Bulk Soil	Cedar Creek, MN, USA	epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	9	74.577	41.334
			epoxyalkane: coenzyme M transferase, *Mycobacterium* sp. JS623	8	76.636	42.875
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	3	78.367	42.667
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	5	69.074	37.4
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	2	69.416	37.5

Table AI.9 Continued.

4480704.3	Dust	Chad Desert, Chad	epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	1	82.76	29
			epoxyalkane: coenzyme M transferase, *Mycobacterium* sp. JS623	3	75.107	27.333
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	5	83.036	23.4
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	1	92.86	14
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	1	72.22	18
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	6	80.638	27.5
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	2	77.78	27
			epoxyalkane: coenzyme M transferase mutant 1 *Mycobacterium* sp. JS623	2	72.425	29
4523235.3	Paddy Field Soil	Changshu, China	epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	1	75.86	29
4477900.3	Bulk Soil	Garwood Valley, Antarctica	epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	1	84.62	13
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	3	84.843	22
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	1	62.5	24
4564114.3	Sea Sediment	Elba, Mediteranian, Italy	putative epoxyalkane:coenzyme M transferase, *Nitrococcus mobilis* Nb-231	3	67.65	73.667
			epoxyalkane:coenzyme M transferase, *Mycobacterium chubuense* NBB4	1	52.46	61

Table AI.9 Continued.

4512589.3	Oil Contaminated Soil	Varennes, QC, Canada	epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	2	77.27	22
4511145.3	Soil	Loma Ridge, CA, USA	putative epoxyalkane:coenzyme M transferase, *Nitrococcus mobilis* Nb-231	1	73.33	30
			putative epoxyalkane:coenzyme M transferase, *Taylorella equigenitalis* MCE9	1	78.26	23
4529836.3	Soil	Koeln, Germany	epoxyalkane: coenzyme M transferase, *Nocardioides* sp. JS614	26	74.377	33.122
			epoxyalkane:coenzyme M transferase, *Mycobacterium tusciae* JS617	1	82.5	40
			epoxyalkane:coenzyme M transferase, *Mycobacterium chubuense* NBB4	6	63.101	36.636
			epoxyalkane:coenzyme M transferase, *Rhodococcus rhodochrous*	15	72.168	36.379
4554870.3	Aquatic microbial mat	Kowary, Poland	putative epoxyalkane:coenzyme M transferase, *Nitrococcus mobilis* Nb-231	1	84.38	32
			putative epoxyalkane:coenzyme M transferase, *Rhodobacterales* bacterium HTCC2654	1	61.11	54
4484670.3	Chyme	Salt Lake City, UT, USA	putative epoxyalkane:coenzyme M transferase, *Taylorella equigenitalis* MCE9	3	80.888	23.2
			putative epoxyalkane:coenzyme M transferase, *Corynebacterium aurimucosum* ATCC 700975	1	84	25
4530091.3	Forest organic material	Johnsbach, Austria	epoxyalkane: coenzyme M transferase, *Nocardioides* sp. JS614	8	72.674	32.732
			epoxyalkane: coenzyme M transferase, *Mycobacterium rhodesiae* JS60	1	61.54	39

Table AI.9 Continued.

4530091.3	Forest organic material	Johnsbach, Austria	epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	1	89.29	28
			putative epoxyalkane: coenzyme M transferase, *Rhodobacterales* bacterium HTCC2654	35	69.874	37.47
			putative epoxyalkane: coenzyme M transferase, *Nitrococcus mobilis* Nb-231	3	74.857	34.333
4482593.3	Tundra Soil	Daring Lake, NWT, Canada	epoxyalkane: coenzyme M transferase, *Nocardioides* sp. JS614	142	80.226	29.636
			epoxyalkane: coenzyme M transferase, *Rhodococcus rhodochrous*	41	79.474	28.875
			epoxyalkane: coenzyme M transferase, *Mycobacterium chubuense* NBB4	40	79.054	27.405
			epoxyalkane: coenzyme M transferase, *Mycobacterium rhodesiae* JS60	20	76.425	29.2
			epoxyalkane: coenzyme M transferase, *Mycobacterium rhodesiae*	10	76.455	27
			epoxyalkane: coenzyme M transferase, *Mycobacterium tusciae* JS617	6	71.198	31.833
			epoxyalkane: coenzyme M transferase, *Mycobacterium* sp. JS621	5	79.76	27.2
			epoxyalkane: coenzyme M transferase, *Ochrobactrum* sp. TD	4	74.803	24.75
			epoxyalkane: coenzyme M transferase, *Mycobacterium mageritense*	4	76.825	25
			epoxyalkane: coenzyme M transferase, *Mycobacterium gadium*	2	72.857	27

Table AI.9 Continued.

4482593.3	Tundra Soil	Daring Lake, NWT, Canada	epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	23	75.773	28.739
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	12	71.595	30.304
			epoxyalkane: coenzyme M transferase mutant 1 *Mycobacterium* sp. JS623	17	77.505	30.273
			epoxyalkane: coenzyme M transferase mutant 5 *Mycobacterium* sp. JS623	5	81.377	25.444
			putative epoxyalkane: coenzyme M transferase, *Nitrococcus mobilis* Nb-231	55	76.949	22.399
			putative epoxyalkane: coenzyme M transferase, *Rhodobacterales* bacterium HTCC2654	52	71.176	25.935
			putative epoxyalkane: coenzyme M transferase, *Taylorella equigenitalis* MCE9	45	73.13	23.891
			putative epoxyalkane: coenzyme M transferase, *Corynebacterium aurimucosum* ATCC 700975	37	74.491	22.331
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	23	75.773	28.739
4494863.3	Activated Sludge	Xiangcheng, Henan, China	epoxyalkane: coenzyme M transferase, *Nocardioides* sp. JS614	1	84.62	13
			epoxyalkane: coenzyme M transferase, *Mycobacterium rhodesiae* JS60	1	66.67	21
4508039.3	Marine Sediment	Gulf of Mexico, USA	putative epoxyalkane: coenzyme M transferase, *Corynebacterium aurimucosum* ATCC 700975	1	84.62	13
4502924.3	Arable Soil	Auburn, IL, USA	epoxyalkane: coenzyme M transferase, *Mycobacterium chubuense* NBB4	33	67.86	23.365
			epoxyalkane: coenzyme M transferase, *Rhodococcus rhodochrous*	21	75.322	23.256

Table AI.9 Continued.

4502924.3	Arable Soil	Auburn, IL, USA	epoxyalkane: coenzyme M transferase, *Mycobacterium tusciae* JS617	3	62.145	26.5
			epoxyalkane: coenzyme M transferase, *Mycobacterium rhodesiae* JS60	3	68.965	24.75
			epoxyalkane: coenzyme M transferase, *Nocardioides* sp. JS614	59	71.111	23.065
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	4	68.733	29
			epoxyalkane: coenzyme M transferase mutant 1 *Mycobacterium* sp. JS623	2	79.15	16
4502926.3	Arable Soil	Mansfield, IL, USA	epoxyalkane: coenzyme M transferase, *Nocardioides* sp. JS614	97	75.81	21.961
			epoxyalkane: coenzyme M transferase, *Mycobacterium chubuense* NBB4	34	70.866	22.079
			epoxyalkane: coenzyme M transferase, *Rhodococcus rhodochrous*	21	73.414	23.897
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	5	76.389	22
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	5	83.666	24.8
			epoxyalkane: coenzyme M transferase mutant 1 *Mycobacterium* sp. JS623	7	82.629	23.429
			epoxyalkane: coenzyme M transferase mutant 4 *Mycobacterium* sp. JS623	3	75.657	30.667
4502927.3	Arable Soil	Urbana, IL, USA	epoxyalkane: coenzyme M transferase, *Mycobacterium chubuense* NBB4	46	72.736	22.416
			epoxyalkane: coenzyme M transferase, *Rhodococcus rhodochrous*	44	70.166	23.228

Table AI.9 Continued.

4502927.3	Arable Soil	Urbana, IL, USA	epoxyalkane: coenzyme M transferase, *Mycobacterium tusciae* JS617	3	84.207	29
			epoxyalkane: coenzyme M transferase *Nocardioides* sp. JS614	99	71.358	22.421
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	5	78.918	22
			epoxyalkane: coenzyme M transferase mutant 5 *Mycobacterium* sp. JS623	3	85.12	15.333
			putative epoxyalkane: coenzyme M transferase, *Nitrococcus mobilis* Nb-231	252	74.985	23.502
			putative epoxyalkane: coenzyme M transferase, *Rhodobacterales* bacterium HTCC2654	174	74.938	22.701
			putative epoxyalkane: coenzyme M transferase, *Taylorella equigenitalis* MCE9	167	73.661	21.975
			putative epoxyalkane: coenzyme M transferase, *Corynebacterium aurimucosum* ATCC 700975	109	72.434	23.424
4469540.3	Permafrost Soil	USA	epoxyalkane: coenzyme M transferase, *Mycobacterium chubuense* NBB4	4	77.81	27
			epoxyalkane: coenzyme M transferase, *Rhodococcus rhodochrous*	4	77.235	27.5
			epoxyalkane: coenzyme M transferase, *Mycobacterium rhodesiae* JS60	2	75.313	29.667
			epoxyalkane: coenzyme M transferase mutant 1 *Mycobacterium* sp. JS623	1	76.47	17
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	3	81.96	24
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	3	86.49	37

Table AI.9 Continued.

4469540.3	Permafrost Soil	USA	epoxyalkane: coenzyme M transferase mutant 5 *Mycobacterium* sp. JS623	1	80.65	31
			putative epoxyalkane: coenzyme M transferase, *Rhodobacterales* bacterium HTCC2654	44	75.44	20.278
			putative epoxyalkane: coenzyme M transferase, *Nitrococcus mobilis* Nb-231	23	73.278	25.726
			putative epoxyalkane: coenzyme M transferase, *Taylorella equigenitalis* MCE9	16	71.455	22.304
4520035.3	Intertidal Sediment	Plum Island, USA	epoxyalkane: coenzyme M transferase *Nocardioides* sp. JS614	1	84.54	17.778
			epoxyalkane: coenzyme M transferase, *Mycobacterium chubuense* NBB4	1	64	25
			putative epoxyalkane: coenzyme M transferase, *Rhodobacterales bacterium* HTCC2654	5	77.142	41.5
4508984.3	Sea water	Kalvhage ȭrden, Sweden	2-hydroxypropyl-CoM lyase (EC 4.4.1.23) (Epoxyalkane: CoM transferase) (EaCoMT) (Aliphatic epoxide carboxylation component I) *Bacillus tusciae* DSM 2912	1	100	10
4472705.3	Mucus, Human Microbe	USA	2-hydroxypropyl-CoM lyase (EC 4.4.1.23) (Epoxyalkane: CoM transferase) (EaCoMT) (Aliphatic epoxide carboxylation component I) *Bacillus tusciae* DSM 2912	1	84.62	13
4519770.3	Oil contaminated soil	Houston, TX, USA	2-hydroxypropyl-CoM lyase (EC 4.4.1.23) (Epoxyalkane: CoM transferase) (EaCoMT) (Aliphatic epoxide carboxylation component I) *Bacillus tusciae* DSM 2912	1	84.62	13
4460182.3	Estuarine bulk water	Puget Sound, WA, USA	putative epoxyalkane: coenzyme M transferase, *Kocuria rhizophila* DC2201	1	55	80

Table AI.9 Continued.

4582806.3	Soil	Dumpas West, Malaysia	epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	8	75.597	81.875
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS614-like)	4	78.925	38.5
			epoxyalkane: coenzyme M transferase, uncultured bacteria (JS623-like)	3	78.52	44
			epoxyalkane: coenzyme M transferase *Nocardioides* sp. JS614	30	81.419	54.1
			epoxyalkane: coenzyme M transferase *Rhodococcus rhodochrous*	10	78.844	53
			epoxyalkane: coenzyme M transferase *Mycobacterium tusciae* JS617	1	78.57	42
			epoxyalkane: coenzyme M transferase mutant 1 *Mycobacterium* sp. JS623	1	57.89	57
			epoxyalkane: coenzyme M transferase mutant 2 *Mycobacterium* sp. JS623	2	57.451	47
			epoxyalkane: coenzyme M transferase mutant 3 *Mycobacterium* sp. JS623	2	71.155	32.5
			putative epoxyalkane: coenzyme M transferase, *Nitrococcus mobilis* Nb-231	24	65.98	44.417
			putative epoxyalkane: coenzyme M transferase, *Rhodobacterales* bacterium HTCC2654	17	68.826	43.177
			putative epoxyalkane: coenzyme M transferase, *Taylorella equigenitalis* MCE9	5	66.18	40.4
4554871.3	Gold Mine Aquatic Microbial Mat	Zolty Stok, Poland	epoxyalkane: coenzyme M transferase, *Mycobacterium chubuense* NBB4	7	65.253	37.429
			epoxyalkane: coenzyme M transferase *Rhodococcus rhodochrous*	2	59.795	42.5

APPENDIX II: OTHER WORKS DURING PHD RESEARCH

Elucidating carbon uptake from vinyl chloride using stable isotope probing and Illumina sequencing

Fernanda Paes, **Xikun Liu**, Timothy E. Mattes, Alison M. Cupples

Abstract

Vinyl chloride (VC), a known human carcinogen, is a common and persistent groundwater pollutant at many chlorinated solvent contaminated sites. The remediation of such sites is challenging because of the lack of knowledge on the microorganisms responsible for in situ VC degradation. To address this, the microorganisms involved in carbon assimilation from VC were investigated in a culture enriched from contaminated site groundwater using stable isotope probing (SIP) and high-throughput sequencing. The mixed culture was added to aerobic media, and these were amended with labeled (13C-VC) or unlabeled VC (12C-VC). The cultures were sacrificed on days 15, 32, and 45 for DNA extraction. DNA extracts and SIP ultracentrifugation fractions were subject to sequencing as well as quantitative PCR (qPCR) for a functional gene linked to VC-assimilation (etnE). The gene etnE encodes for epoxyalkane: coenzyme M transferase, a critical enzyme in the pathway for VC degradation. The relative abundance of phylotypes was compared across ultracentrifugation fractions obtained from the 13C-VC- and 12C-VC-amended cultures. Four phylotypes were more abundant in the heavy fractions (those of greater buoyant density) from the 13C-VC-amended cultures compared to those from the 12C-VC-amended cultures, including *Nocardioides*, *Brevundimonas*, *Tissierella*, and

Rhodoferax. Therefore, both a previously identified VC-assimilating genus (*Nocardioides*) and novel microorganisms were responsible for carbon uptake. Enrichment of *etnE* with time was observed in the heavy fractions, and *etnE* sequences illustrated that VC-assimilators harbor similar *Nocardioides*-like *etnE*. This research provides novel data on the microorganisms able to assimilate carbon from VC.

Contribution to the paper

qPCR analysis on functional genes; clone library of functional genes.

Partial nitritation ANAMMOX in submerged attached growth bioreactors with smart aeration at 20°C

James M. Shannon, Lee W. Hauser, **Xikun Liu**, Gene F. Parkin, Timothy E. Mattes and Craig L. Just

Published in Environmental Sciences: Processes and Impacts 17(1) in November 2014

Abstract

Submerged attached growth bioreactors (SAGBs) were operated at 20°C for 30 weeks in smart-aerated, partial nitritation ANAMMOX mode and in a timer-controlled, cyclic aeration mode. The smart-aerated SAGBs removed 48-53% of total nitrogen (TN) compared to 45% for SAGBs with timed aeration. Low dissolved oxygen concentrations and cyclic pH patterns in the smart-aerated SAGBs suggested conditions favorable to partial nitritation ANAMMOX and stoichiometrically-derived and numerically modeled estimations attributed 63-68% and 14-44% of TN removal to partial nitritation ANAMMOX in these bioreactors, respectively. Ammonia removals of 36-67% in the smart-aerated SAGBs, with measured oxygen and organic carbon limitations, further suggest partial nitritation ANAMMOX. The smart-aerated SAGBs required substantially less aeration to achieve TN removals similar to SAGBs with timer-controlled aeration. Genomic DNA testing confirmed that the dominant ANAMMOX seed bacteria, received from a treatment plant utilizing the DEMON® sidestream deammonification process, was a *Candidatus Brocadia* sp. (of the *Planctomycetales* order). The DNA from these bacteria

was also present in the SAGBs at the conclusion of the study providing evidence for attached growth and limited biomass washout.

Contribution to the paper

Performed DNA extraction, clone library analysis and wrote correspondent part of the manuscript.

Abundance and activity of vinyl chloride (VC)-oxidizing bacteria in a dilute groundwater VC plume biostimulated with oxygen and ethene

Mattes TE, Jin YO, Livermore J, Pearl M, **Liu X**

Abstract

Clean-up of vinyl chloride (VC)-contaminated groundwater could be enhanced by stimulating aerobic VC-oxidizing bacterial populations (e.g., methanotrophs) with amendments such as molecular oxygen. In addition, ethene gas injection could further stimulate a different group of aerobic ethene- and VC-oxidizing bacteria called "etheneotrophs." We estimated the abundance and activity of these different VC-oxidizing bacteria in portions of a dilute groundwater VC plume subjected to oxygen and ethene biostimulation. Pyrosequencing of 16S rRNA genes, amplified from community DNA extracted from five groundwater monitoring wells, revealed that Proteobacteria dominated the microbial community. Among the Proteobacteria, methanotroph relative abundance was 6.00 % (well RB52I), 2.81 % (well RB46D), 56.3 % (well RB58I), 23.8 % (well RB63I), and 2.57 % (well RB64I). Reverse transcription qPCR (RT-qPCR) analysis was used to determined methanotroph and etheneotroph functional gene expression from selected monitoring wells. Resulting transcript per gene ratios for methanotroph functional genes (pmoA and mmoX) were 0.013 (RB46D), 0.017 (RB63I), 0.112 (RB64I), and 0.004 (RB46D), 0.239 (RB63I), and 0.199 (RB64I), respectively. Transcript per gene ratios for etheneotroph functional genes (etnC and etnE) were 0.37 (RB46D), 0.81 (RB63I), 5.85 (RB64I), and 0.38 (RB46D), 0.67 (RB63I), and 2.28

(RB64I), respectively. When considered along with geochemical and contaminant data from these wells, our RT-qPCR results suggest that methanotrophs and etheneotrophs were participating in VC cometabolism. We conclude that these molecular diagnostic techniques could be helpful to site managers interested in documenting the effectiveness of VC bioremediation strategies.

Contribution to the study

16S rRNA gene amplicon 454 pyro-sequencing data analysis

Reconstructing ancient prokaryotic lake communities from Lake Leija paleosediments in the Atacama desert

Liu X, Peate IU, Spak SN, Liang Y and Mattes TE

Research in Progress

Abstract

Lake Lejia, located in the high Altiplano of the Atacama Desert, is a region affected by rapid climatic changes (e.g. the Younger Dryas, a severe cold spell between 12,800 and 11,500 years before present). This purpose of this study was to reconstruct ancient microbial communities in Lake Leija over a time period that encompasses the Younger Dryas, and examine the possible effects climate change on microbial community structure. To accomplish this 25 Lake Leija paleo sediment samples, deposited between 14,000 and 4,300 years ago, were collected and subjected to duplicate DNA extraction, followed by 2×151 nt (paired-end) MiSeq Illumina sequencing targeting the V4 region of the 16S rRNA gene. Across all samples the majority (97.8-100%) of the sequences were classified as *Bacteria* with the dominant phyla being *Actinobacteria*, *Bacteroidetes*, *Firmicutes* and *Proteobacteria*. A dramatic shift from a bacterial community dominated by *Bacteroidetes* and *Proteobacteria* to one dominated by *Actinobacteria* was observed between samples 9 and 10. This shift appeared to correspond to geochemical changes in the sediment layers. Bacterial community diversity in the older organic-rich sediment layers was greater than the community diversity observed in more recent sediment layers. Dominant bacterial genera in the community include *Salinimicrobium, Aciditerrimonas*,

Gaiella, *Marinobacter* and *Planomicrobium,* representatives of which have been found in saline aquatic and other extreme environments. Bacterial and archaeal sequences associated with methane, sulfur, iron and nitrogen cycling, as well as those expected to carry out photosynthesis and/or light-mediated ATP synthesis were found among all sediment samples.

Contribution to the study

DNA extraction and 16S rRNA gene Illumina sequencing data analysis

REFERENCES

1. Gossett, J. M., Measurement of Henry's Law constants for C1 and C2 chlorinated hydrocarbons. *Environ Sci Technol* **1987,** *21* (2), 202-208.
2. Guengerich, F. P.; Langouet, S.; Mican, A. N.; Akasaka, S.; Muller, M.; Persmark, M., Formation of etheno adducts and their effects on DNA polymerases. *IARC Sci Publ* **1999,** (150), 137-45.
3. United States Environmental Protection Agency, National primary drinking water regulations. https://www.epa.gov/ground-water-and-drinking-water/table-regulated-drinking-water-contaminants, 2016.
4. Keppler, F.; Borchers, R.; Pracht, J.; Rheinberger, S.; Scholer, H. F., Natural formation of vinyl chloride in the terrestrial environment. *Environ Sci Technol* **2002,** *36* (11), 2479-2483.
5. Bradley, P. M., History and ecology of chloroethene biodegradation: a review. *Biorem J* **2003,** *7*, 81-109.
6. Maymó-Gatell, X.; Chien, Y. T.; Gossett, J. M.; Zinder, S. H., Isolation of a bacterium that reductively dechlorinates tetrachloroethene to ethene. *Science* **1997,** *276* (5318), 1568-1571.
7. Maymó-Gatell, X.; Anguish, T.; Zinder, S. H., Reductive dechlorination of chlorinated ethenes and 1,2-dichloroethane by "*Dehalococcoides* ethenogenes" 195. *Appl Environ Microbiol* **1999,** *65* (7), 3108-3113.
8. Mattes, T. E.; Alexander, A. K.; Coleman, N. V., Aerobic biodegradation of the chloroethenes: pathways, enzymes, ecology, and evolution. *FEMS Microbiol Rev* **2010,** *34* (4), 445-475.
9. Mattes, T. E.; Coleman, N. V.; Spain, J. C.; Gossett, J. M., Physiological and molecular genetic analyses of vinyl chloride and ethene biodegradation in *Nocardioides* sp. strain JS614. *Arch Microbiol* **2005,** *183* (2), 95-106.
10. Coleman, N. V.; Mattes, T. E.; Gossett, J. M.; Spain, J. C., Phylogenetic and kinetic diversity of aerobic vinyl chloride-assimilating bacteria from contaminated sites. *Appl Environ Microbiol* **2002,** *68* (12), 6162-6171.
11. Coleman, N. V.; Spain, J. C., Distribution of the coenzyme M pathway of epoxide metabolism among ethene- and vinyl chloride-degrading *Mycobacterium* strains. *Appl Environ Microbiol* **2003,** *69* (10), 6041-6046.
12. Coleman, N. V.; Spain, J. C., Epoxyalkane: coenzyme M transferase in the ethene and vinyl chloride biodegradation pathways of *Mycobacterium* strain JS60. *J Bacteriol* **2003,** *185* (18), 5536-5545.
13. Freedman, D. L.; Herz, S. D., Use of ethylene and ethane as primary substrates for aerobic cometabolism of vinyl chloride. *Water Environ Res* **1996,** *68* (3), 320-328.
14. Koziollek, P.; Bryniok, D.; Knackmuss, H. J., Ethene as an auxiliary substrate for the cooxidation of *cis*-1,2-dichloroethene and vinyl chloride. *Arch Microbiol* **1999,** *172* (4), 240-246.
15. Dolan, M. E.; McCarty, P. L., Small column microcosm for assessing methane-stimulated vinyl chloride transformation in aquifer samples. *Environ Sci Technol* **1995,** *29* (8), 1892-1897.

16. Schafer, A.; Bouwer, E. J., Toluene induced cometabolism of *cis*-1,2-dichloroethylene and vinyl chloride under conditions expected downgradient of a permeable Fe(0) barrier. *Water Res* **2000,** *34* (13), 3391-3399.
17. Danko, A. S.; Saski, C. A.; TomkinS, J. P.; Freedman, D. L., Involvement of coenzyme M during aerobic biodegradation of vinyl chloride and ethene by *Pseudomonas putida* strain AJ and *Ochrobactrum* sp. strain TD. *Appl Environ Microbiol* **2006,** *72* (5), 3756-3758.
18. Coleman, N. V.; Yau, S.; Wilson, N. L.; Nolan, L. M.; Migocki, M. D.; Ly, M. A.; Crossett, B.; Holmes, A. J., Untangling the multiple monooxygenases of *Mycobacterium chubuense* strain NBB4, a versatile hydrocarbon degrader. *Environ Microbiol Rep* **2011,** *3* (3), 297-307.
19. Suzuki, T.; Nakamura, T.; Fuse, H., Isolation of two novel marine ethylene-assimilating bacteria, *Haliea* species ETY-M and ETY-NAG, containing particulate methane monooxygenase-like genes. *Microbes Environ* **2012,** *27* (1), 54-60.
20. Krum, J. G.; Ensign, S. A., Evidence that a linear megaplasmid encodes enzymes of aliphatic alkene and epoxide metabolism and coenzyme M (2-mercaptoethanesulfonate) biosynthesis in *Xanthobacter* strain Py2. *J Bacteriol* **2001,** *183* (7), 2172-7.
21. Meinhardt, F.; Schaffrath, R.; Larsen, M., Microbial linear plasmids. *Appl Microbiol Biotechnol* **1997,** *47* (4), 329-336.
22. Saeki, H.; Akira, M.; Furuhashi, K.; Averhoff, B.; Gottschalk, G., Degradation of trichloroethene by a linear-plasmid-encoded alkene monooxygenase in *Rhodococcus corallinus* (*Nocardia corallina*) B-276. *Microbiology-Uk* **1999,** *145*, 1721-1730.
23. Stecker, C.; Johann, A.; Herzberg, C.; Averhoff, B.; Gottschalk, G., Complete nucleotide sequence and genetic organization of the 210-kilobase linear plasmid of *Rhodococcus erythropolis* BD2. *J Bacteriol* **2003,** *185* (17), 5269-5274.
24. Danko, A. S.; Luo, M. Z.; Bagwell, C. E.; Brigmon, R. L.; Freedman, D. L., Involvement of linear plasmids in aerobic biodegradation of vinyl chloride. *Appl Environ Microbiol* **2004,** *70* (10), 6092-6097.
25. Singh, H.; Loffler, F. E.; Fathepure, B. Z., Aerobic biodegradation of vinyl chloride by a highly enriched mixed culture. *Biodegradation* **2004,** *15* (3), 197-204.
26. Findlay, M.; Smoler, D. F.; Fogel, S.; Mattes, T. E., Aerobic vinyl chloride metabolism in groundwater microcosms by methanotrophic and etheneotrophic bacteria. *Environ Sci Tech* **2016,** *50* (7), 3617-3625.
27. Oh, S.; Tandukar, M.; Pavlostathis, S. G.; Chain, P. S. G.; Konstantinidis, K. T., Microbial community adaptation to quaternary ammonium biocides as revealed by metagenomics. *Environ Microbiol* **2013,** *15* (10), 2850-2864.
28. Tandukar, M.; Oh, S.; Tezel, U.; Konstantinidis, K. T.; Pavlostathis, S. G., Long-term exposure to benzalkonium chloride disinfectants results in change of microbial community structure and increased antimicrobial resistance. *Environ Sci Technol* **2013,** *47* (17), 9730-9738.
29. Lefevre, E.; Cooper, E.; Stapleton, H. M.; Gunsch, C. K., Characterization and Adaptation of Anaerobic Sludge Microbial Communities Exposed to Tetrabromobisphenol A. *PLoS One* **2016,** *11* (7), 20.

30. Na, J. G.; Lee, M. K.; Yun, Y. M.; Moon, C.; Kim, M. S.; Kim, D. H., Microbial community analysis of anaerobic granules in phenol-degrading UASB by next generation sequencing. *Biochem Eng J* **2016,** *112*, 241-248.
31. Jin, Y. O.; Mattes, T. E., Adaptation of aerobic, ethene-assimilating *Mycobacterium* strains to vinyl chloride as a growth substrate. *Environ Sci Technol* **2008,** *42* (13), 4784-4789.
32. Top, E. M.; Springael, D.; Boon, N., Catabolic mobile genetic elements and their potential use in bioaugmentation of polluted soils and waters. *FEMS Microbiol Ecol* **2002,** *42* (2), 199-208.
33. Whyte, L. G.; Smits, T. H. M.; Labbe, D.; Witholt, B.; Greer, C. W.; van Beilen, J. B., Gene cloning and characterization of multiple alkane hydroxylase systems in *Rhodococcus* strains Q15 and NRRL B-16531Gene cloning and characterization of multiple alkane hydroxylase systems in Rhodococcus strains Q15 and NRRL B-16531. *Appl Environ Microbiol* **2002,** *68* (12), 5933-5942.
34. McGowan, C.; Fulthorpe, R.; Wright, A.; Tiedje, J. M., Evidence for interspecies gene transfer in the evolution of 2,4-dichlorophenoxyacetic acid degraders. *Appl Environ Microbiol* **1998,** *64* (10), 4089-4092.
35. Chuang, A. S.; Jin, Y. O.; Schmidt, L. S.; Li, Y. L.; Fogel, S.; Smoler, D.; Mattes, T. E., Proteomic analysis of ethene-enriched groundwater microcosms from a vinyl chloride-contaminated site. *Environ Sci Technol* **2010,** *44* (5), 1594-1601.
36. Mattes, T. E.; Coleman, N. V.; Chuang, A. S.; Rogers, A. J.; Spain, J. C.; Gossett, J. M., Mechanism controlling the extended lag period associated with vinyl chloride starvation in *Nocardioides* sp. strain JS614. *Arch Microbiol* **2007,** *187* (3), 217-226.
37. Ensign, S. A., Aliphatic and chlorinated alkenes and epoxides as inducers of alkene monooxygenase and epoxidase activities in *Xanthobacter* strain Py2. *Appl Environ Microbiol* **1996,** *62* (1), 61-66.
38. Harris, S. R.; Clarke, I. N.; Seth-Smith, H. M. B.; Solomon, A. W.; Cutcliffe, L. T.; Marsh, P.; Skilton, R. J.; Holland, M. J.; Mabey, D.; Peeling, R. W.; Lewis, D. A.; Spratt, B. G.; Unemo, M.; Persson, K.; Bjartling, C.; Brunham, R.; de Vries, H. J. C.; Morre, S. A.; Speksnijder, A.; Bebear, C. M.; Clerc, M.; de Barbeyrac, B.; Parkhill, J.; Thomson, N. R., Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. *Nat Genet* **2012,** *44* (4), 413-U221.
39. Oh, S.; Caro-Quintero, A.; Tsementzi, D.; DeLeon-Rodriguez, N.; Luo, C. W.; Poretsky, R.; Konstantinidis, K. T., Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. *Appl Environ Microbiol* **2011,** *77* (17), 6000-6011.
40. Brisson, V. L.; West, K. A.; Lee, P. K. H.; Tringe, S. G.; Brodie, E. L.; Alvarez-Cohen, L., Metagenomic analysis of a stable trichloroethene-degrading microbial community. *ISME J* **2012,** *6* (9), 1702-1714.
41. Guo, F.; Wang, Z. P.; Yu, K.; Zhang, T., Detailed investigation of the microbial community in foaming activated sludge reveals novel foam formers. *Sci Rep* **2015,** *5*, 9.
42. Cai, M. W.; Wilkins, D.; Chen, J. P.; Ng, S. K.; Lu, H. Y.; Jia, Y. Y.; Lee, P. K. H., Metagenomic reconstruction of key anaerobic digestion pathways in municipal sludge and industrial wastewater biogas-producing systems. *Front Microbiol* **2016,** *7*, 12.

43. Loman, N. J.; Misra, R. V.; Dallman, T. J.; Constantinidou, C.; Gharbia, S. E.; Wain, J.; Pallen, M. J., Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* **2012,** *30* (5), 434-439.
44. Treangen, T. J.; Koren, S.; Sommer, D. D.; Liu, B.; Astrovskaya, I.; Ondov, B.; Darling, A. E.; Phillippy, A. M.; Pop, M., MetAMOS: a modular and open source metagenomic assembly and analysis pipeline. *Genome Biol* **2013,** *14* (1), 20.
45. Bolger, A. M.; Lohse, M.; Usadel, B., Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014,** *30* (15), 2114-2120.
46. Peng, Y.; Leung, H. C. M.; Yiu, S. M.; Chin, F. Y. L., IDBA - A practical iterative de Bruijn graph de novo assembler. *RECOMB. Lisbon.* **2010,** *6044*, 426-440.
47. R Development Core Team, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria, ISBN 3-900051-07-0, http://www.R-project.org, 2008.
48. Cupples, A. M.; Shaffer, E. A.; Chee-Sanford, J. C.; Sims, G. K., DNA buoyant density shifts during N-15-DNA stable isotope probing. *Microbiol Res* **2007,** *162* (4), 328-334.
49. Cupples, A. M., Real-time PCR quantification of *Dehalococcoides* populations: methods and applications. *J Microbiol Met* **2008,** *72* (1), 1-11.
50. Luo, C. L.; Xie, S. G.; Sun, W. M.; Li, X. D.; Cupples, A. M., Identification of a novel toluene-degrading bacterium from the Candidate Phylum TM7, as determined by DNA stable isotope probing. *Appl Environ Microbiol* **2009,** *75* (13), 4644-4647.
51. Sun, W. M.; Xie, S. G.; Luo, C. L.; Cupples, A. M., Direct link between toluene degradation in contaminated-site microcosms and a *Polaromonas* strain. *Appl Environ Microbiol* **2010,** *76* (3), 956-959.
52. Sun, W. M.; Cupples, A. M., Diversity of five anaerobic toluene-degrading microbial communities investigated using stable isotope probing. *Appl Environ Microbiol* **2012,** *78* (4), 972-980.
53. Xie, S. G.; Sun, W. M.; Luo, C. L.; Cupples, A. M., Stable isotope probing identifies novel m-xylene degraders in soil microcosms from contaminated and uncontaminated Sites. *Water Air Soil Pollut* **2010,** *212* (1-4), 113-122.
54. Xie, S. G.; Sun, W. M.; Luo, C. L.; Cupples, A. M., Novel aerobic benzene degrading microorganisms identified in three soils by stable isotope probing. *Biodegradation* **2011,** *22* (1), 71-81.
55. Neufeld, J. D.; Vohra, J.; Dumont, M. G.; Lueders, T.; Manefield, M.; Friedrich, M. W.; Murrell, J. C., DNA stable-isotope probing. *Nat Protoc* **2007,** *2* (4), 860-866.
56. Radajewski, S.; Ineson, P.; Parekh, N. R.; Murrell, J. C., Stable-isotope probing as a tool in microbial ecology. *Nature* **2000,** *403* (6770), 646-649.
57. Radajewski, S.; Webster, G.; Reay, D. S.; Morris, S. A.; Ineson, P.; Nedwell, D. B.; Prosser, J. I.; Murrell, J. C., Identification of active methylotroph populations in an acidic forest soil by stableisotope probing. *Microbiol SGM* **2002,** *148*, 2331-2342.
58. Radajewski, S.; McDonald, I. R.; Murrell, J. C., Stable-isotope probing of nucleic acids: a window to the function of uncultured microorganisms. *Curr Opin Biotechnol* **2003,** *14* (3), 296-302.
59. Higuchi, R.; Fockler, C.; Dollinger, G.; Watson, R., Kinetic PCR analysis: Real-time monitoring of DNA amplification reactions. *Nature Biotechnol* **1993,** *11* (9), 1026-1030.

60. Bustin, S. A.; Benes, V.; Garson, J. A.; Hellemans, J.; Huggett, J.; Kubista, M.; Mueller, R.; Nolan, T.; Pfaffl, M. W.; Shipley, G. L.; Vandesompele, J.; Wittwer, C. T., The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **2009,** *55* (4), 611-622.
61. United States Environmental Protection Agency (US EPA), Method 1611: Enterococci in water by TaqMan quantitative polymerase chain reaction (qPCR) assay. Washington, D.C., 2012.
62. Fout, G. S.; Cashdollar, J. L.; Griffin, S. M.; Brinkman, N. E.; Varughese, E. A.; Parshionikar, S. U., EPA Method 1615. Measurement of enterovirus and norovirus occurrence in water by culture and RT-qPCR. Part III. Virus detection by RT-qPCR. *J Vis Exp* **2016,** (107), 13.
63. Liu, W. T.; Marsh, T. L.; Cheng, H.; Forney, L. J., Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol* **1997,** *63* (11), 4516-4522.
64. Wilson, F. P.; Liu, X. K.; Mattes, T. E.; Cupples, A. M., *Nocardioides*, *Sediminibacterium*, *Aquabacterium*, *Variovorax*, and *Pseudomonas* linked to carbon uptake during aerobic vinyl chloride biodegradation. *Environ Sci Pollut Res* **2016,** *23* (19), 19062–19070.
65. Paes, F.; Liu, X. K.; Mattes, T. E.; Cupples, A. M., Elucidating carbon uptake from vinyl chloride using stable isotope probing and Illumina sequencing. *Appl Microbiol Biotechnol* **2015,** *99* (18), 7735-7743.
66. He, J.; Kirsti M. Ritalahti; Kun-Lin Yang; Stephen S. Koenigsberg; Löffler, F. E., Detoxification of vinyl chloride to ethene coupled to growth of an anaerobic bacterium. *Nature* **2003,** *424* (6944), 62-65.
67. Cupples, A. M.; Spormann, A. M.; McCarty, P. L., Growth of a *Dehalococcoides*-like microorganism on vinyl chloride and cis-dichloroethene as electron acceptors as determined by competitive PCR. *Appl Environ Microb* **2003,** *69* (2), 953-959.
68. Hartmans, S.; Debont, J. A. M.; Tramper, J.; Luyben, K., Bacterial degradation of vinyl chloride. *Biotechnol. Lett.* **1985,** *7* (6), 383-388.
69. Fathepure, B. Z.; Elango, V. K.; Singh, H.; Bruner, M. A., Bioaugmentation potential of a vinyl chloride-assimilating *Mycobacterium* sp., isolated from a chloroethene-contaminated aquifer. *Fems Microbiology Letters* **2005,** *248* (2), 227-234.
70. Fullerton, H.; Rogers, R.; Freedman, D. L.; Zinder, S. H., Isolation of an aerobic vinyl chloride oxidizer from anaerobic groundwater. *Biodegradation* **2014,** *25* (6), 893-901.
71. Hartmans, S.; Debont, J. A. M., Aerobic vinyl-chloride metabolism im *Mycobacterium aurum* L1. *Appl Environ Microbiol* **1992,** *58* (4), 1220-1226.
72. Jin, Y. O.; Cheung, S.; Coleman, N. V.; Mattes, T. E., Association of missense mutations in epoxyalkane coenzyme M transferase with adaptation of *Mycobacterium* sp. strain JS623 to growth on vinyl chloride. *Appl Environ Microbiol* **2010,** *76* (11), 3413-3419.
73. Elango, V. K.; Liggenstoffer, A. S.; Fathepure, B. Z., Biodegradation of vinyl chloride and cis-dichloroethene by a *Ralstonia* sp. strain TRW-1. *Appl Microbiol Biotechnol* **2006,** *72* (6), 1270-1275.

74. Taylor, A. E.; Dolan, M. E.; Bottomley, P. J.; Semprini, L., Utilization of fluoroethene as a surrogate for aerobic vinyl chloride transformation. *Environ Sci Technol* **2007,** *41* (18), 6378-6383.
75. Verce, M. F.; Ulrich, R. L.; Freedman, D. L., Characterization of an isolate that uses vinyl chloride as a growth substrate under aerobic conditions. *Appl Environ Microbiol* **2000,** *66* (8), 3535-3542.
76. Cupples, A. M., The use of nucleic acid based stable isotope probing to identify the microorganisms responsible for anaerobic benzene and toluene biodegradation. *J Microbiol Methods* **2011,** *85* (2), 83-91.
77. Jin, Y. O.; Mattes, T. E., A quantitative PCR assay for aerobic, vinyl chloride- and ethene-assimilating microorganisms in groundwater. *Environ Sci Technol* **2010,** *44* (23), 9036-9041.
78. Patterson, B. M.; Aravena, R.; Davis, G. B.; Furness, A. J.; Bastow, T. P.; Bouchard, D., Multiple lines of evidence to demonstrate vinyl chloride aerobic biodegradation in the vadose zone, and factors controlling rates. *J Contam Hydrol* **2013,** *153*, 69-77.
79. Coleman, N. V.; Mattes, T. E.; Gossett, J. M.; Spain, J. C., Biodegradation of *cis*-dichloroethene as the sole carbon source by a beta-proteobacterium. *Appl Environ Microbiol* **2002,** *68* (6), 2726-2730.
80. Caporaso, J. G.; Lauber, C. L.; Walters, W. A.; Berg-Lyons, D.; Lozupone, C. A.; Turnbaugh, P. J.; Fierer, N.; Knight, R., Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci USA* **2011,** *108*, 4516-4522.
81. Schloss, P. D.; Westcott, S. L.; Ryabin, T.; Hall, J. R.; Hartmann, M.; Hollister, E. B.; Lesniewski, R. A.; Oakley, B. B.; Parks, D. H.; Robinson, C. J.; Sahl, J. W.; Stres, B.; Thallinger, G. G.; Van Horn, D. J.; Weber, C. F., Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **2009,** *75* (23), 7537-7541.
82. Pruesse, E.; Quast, C.; Knittel, K.; Fuchs, B. M.; Ludwig, W. G.; Peplies, J.; Glockner, F. O., SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **2007,** *35* (21), 7188-7196.
83. Ritalahti, K. M.; Amos, B. K.; Sung, Y.; Wu, Q. Z.; Koenigsberg, S. S.; Loffler, F. E., Quantitative PCR targeting 16S rRNA and reductive dehalogenase genes simultaneously monitors multiple *Dehalococcoides* strains. *Appl Environ Microbiol* **2006,** *72* (4), 2765-2774.
84. Davis, J.; Carpenter, C., Aerobic degradation of vinyl chloride in groundwater samples. *Appl Environ Microb* **1990,** *56*, 3878-3880.
85. Hatt, J. K.; Loffler, F. E., Quantitative real-time PCR (qPCR) detection chemistries affect enumeration of the *Dehalococcoides* 16S rRNA gene in groundwater. *J Microbiol Methods* **2012,** *88* (2), 263-270.
86. Kanitkar, Y. H.; Stedtfeld, R. D.; Steffan, R. J.; Hashsham, S. A.; Cupples, A. M., Loop-mediated isothermal amplification (LAMP) for rapid detection and quantification of *Dehalococcoides* biomarker genes in commercial reductive dechlorinating cultures KB-1 and SDC-9. *Appl Environ Microbiol* **2016,** *82* (6), 1799-1806.

87. Stedtfeld, R. D.; Stedtfeld, T. M.; Kronlein, M.; Seyrig, G.; Steffan, R. J.; Cupples, A. M.; Hashsham, S. A., DNA extraction-free quantification of *Dehalococcoides* spp. in groundwater using a hand-held deviceDNA extraction-free quantification of Dehalococcoides spp. in groundwater using a hand-held device. *Environ Sci Technol* **2014,** *48* (23), 13855-13863.
88. Guan, X. Y.; Liu, F.; Xie, Y. X.; Zhu, L. L.; Han, B., Microbiota associated with the migration and transformation of chlorinated aliphatic hydrocarbons in groundwater. *Environ. Geochem. Health* **2013,** *35* (4), 535-549.
89. Hartmans, S.; Kaptein, A.; Tramper, J.; Debont, J. A. M., Characterization of a *Mycobacterium* sp. and a *Xanthobacter* sp. for the removal of vinyl chloride and 1,2-dichloroethane from waste gases. *Appl Microbiol Biotechnol* **1992,** *37* (6), 796-801.
90. Buckley, D. H.; Huangyutitham, V.; Hsu, S. F.; Nelson, T. A., Stable isotope probing with 15N achieved by disentangling the effects of genome G+C content and isotope enrichment on DNA density. *Appl Environ Microbiol* **2007,** *73* (10), 3189-3195.
91. Collins, R. E.; Rocap, G., REPK: an analytical web server to select restriction endonucleases for terminal restriction fragment length polymorphism analysis. *Nucleic Acids Res* **2007,** *35*, W58-W62.
92. Weisburg, W. G.; Barns, S. M.; Pelletier, D. A.; Lane, D. J., 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* **1991,** *173* (2), 697-703.
93. Schutte, U. M. E.; Abdo, Z.; Bent, S. J.; Shyu, C.; Williams, C. J.; Pierson, J. D.; Forney, L. J., Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. *Appl Microbiol Biotechnol* **2008,** *80* (3), 365-380.
94. Liang, Y.; Martinez, A.; Hornbuckle, K. C.; Mattes, T. E., Potential for polychlorinated biphenyl biodegradation in sediments from Indiana Harbor and Ship Canal. *Int Biodeterior Biodegradation* **2014,** *89*, 50-57.
95. Watanabe, K.; Nagao, N.; Toda, T.; Kurosawa, N., The dominant bacteria shifted from the order *Lactobacillales* to *Bacillales* and *Actinomycetales* during a start-up period of large-scale, completely-mixed composting reactor using plastic bottle flakes as bulking agent. *World J Microbiol Biotechnol* **2009,** *25* (5), 803-811.
96. Abe, Y.; Aravena, R.; Zopfi, J.; Parker, B.; Hunkeler, D., Evaluating the fate of chlorinated ethenes in streambed sediments by combining stable isotope, geochemical and microbial methods. *J Contam Hydrol* **2009,** *107* (1-2), 10-21.
97. Jin, Y. O.; Mattes, T. E., Assessment and modification of degenerate qPCR primers that amplify functional genes from etheneotrophs and vinyl chloride-assimilators. *Lett Appl Microbiol* **2011,** *53* (5), 576-580.
98. Liu, X. K.; Mattes, T. E., Epoxyalkane:Coenzyme M transferase gene diversity and distribution in groundwater samples from chlorinated-ethene-contaminated sites. *Appl Environ Microbiol* **2016,** *82* (11), 3269-3279.
99. Agency for Toxic Substances and Disease Registry (ATSDR) *Toxicological Profile for Vinyl Chloride (Update)*; Public Health Service, U.S. Department of Health and Human Services: Atlanta, GA., 1997.
100. Squillace, P. J.; Moran, M. J.; Lapham, W. W.; Price, C. V.; Clawges, R. M.; Zogorski, J. S., Volatile organic compounds in untreated ambient groundwater of the United States, 1985-1995. *Environ Sci Technol* **1999,** *33*, 4176-4187.

101. Freedman, D. L.; Gossett, J. M., Biological reductive dechlorination of tetrachloroethylene and trichloroethylene to ethylene under methanogenic conditions. *Appl Environ Microbiol* **1989,** *55* (9), 2144-2151.
102. Steffan, R. J.; Vainberg, S., *Bioaugmentation for groundwater remediation*. Springer: New York, USA, 2013.
103. Gossett, J. M., Sustained aerobic oxidation of vinyl chloride at low oxygen concentrations. *Environ Sci Technol* **2010,** *44* (4), 1405-1411.
104. Fogel, M. M.; Taddeo, A. R.; Fogel, S., Biodegradation of chlorinated ethenes by a methane-utilizing mixed culture. *Appl Environ Microbiol* **1986,** *51* (4), 720-724.
105. Tsien, H. C.; Brusseau, G. A.; Hanson, R. S.; Wackett, L. P., Biodegradation of trichloroethylene by *Methylosinus trichosporium* OB3b. *Appl Environ Microbiol* **1989,** *55* (12), 3155-3161.
106. Verce, M. F.; Ulrich, R. L.; Freedman, D. L., Transition from cometabolic to growth-linked biodegradation of vinyl chloride by a *Pseudomonas* sp. isolated on ethene. *Environ Sci Technol* **2001,** *35*, 4242-4251.
107. Ensign, S. A., Microbial metabolism of aliphatic alkenes. *Biochemistry* **2001,** *40* (20), 5845-5853.
108. Allen, J. R.; Clark, D. D.; Krum, J. G.; Ensign, S. A., A role for coenzyme M (2-mercaptoethanesulfonic acid) in a bacterial pathway of aliphatic epoxide carboxylation. *Proc Natl Acad Sci USA* **1999,** *96* (15), 8432-8437.
109. Debont, J. A. M.; Harder, W., Metabolism of ethylene by *Mycobacterium* E20. *FEMS Microbiol Lett* **1978,** *3* (2), 89-93.
110. Coleman, N. V.; Wilson, N. L.; Barry, K.; Brettin, T. S.; Bruce, D. C.; Copeland, A.; Dalin, E.; Detter, J. C.; Glavina del Rio, T.; Goodwin, L. A.; Hammon, N. M.; Han, S.; Hauser, L. J.; Israni, S.; Kim, E.; Kyrpides, N.; Land, M. L.; Lapidus, A.; Larimer, F. W.; Lucas, S.; Pitluck, S.; Richardson, P.; Schmutz, J.; Tapia, R.; Thompson, S.; Tice, H. N.; Spain, J. C.; Gossett, J. G.; Mattes, T. E., Genome sequence of the ethene- and vinyl chloride-oxidizing actinomycete *Nocardioides* sp. strain JS614. *J Bacteriol* **2011,** *193* (13), 3399-3400.
111. Mattes, T. E.; Jin, Y. O.; Livermore, J.; Pearl, M.; Liu, X. K., Abundance and activity of vinyl chloride (VC)-oxidizing bacteria in a dilute groundwater VC plume biostimulated with oxygen and ethene. *Appl Microbiol Biotechnol* **2015,** *99* (21), 9267-9276.
112. Caporaso, J. G.; Lauber, C. L.; Walters, W. A.; Berg-Lyons, D.; Huntley, J.; Fierer, N.; Owens, S. M.; Betley, J.; Fraser, L.; Bauer, M.; Gormley, N.; Gilbert, J. A.; Smith, G.; Knight, R., Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **2012,** *6* (8), 1621-1624.
113. Kozich, J. J.; Westcott, S. L.; Baxter, N. T.; Highlander, S. K.; Schloss, P. D., Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* **2013,** *79* (17), 5112-5120.
114. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glockner, F. O., The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **2013,** *41* (D1), D590-D596.

115. Cole, J. R.; Wang, Q.; Fish, J. A.; Chai, B. L.; McGarrell, D. M.; Sun, Y. N.; Brown, C. T.; Porras-Alfaro, A.; Kuske, C. R.; Tiedje, J. M., Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nuc Acids Res* **2014,** *42* (D1), D633-D642.
116. Meyer, F.; Paarmann, D.; D'Souza, M.; Olson, R.; Glass, E. M.; Kubal, M.; Paczian, T.; Rodriguez, A.; Stevens, R.; Wilke, A.; Wilkening, J.; Edwards, R. A., The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **2008,** *9*, 386.
117. Wilke, A.; Harrison, T.; Wilkening, J.; Field, D.; Glass, E. M.; Kyrpides, N.; Mavrommatis, K.; Meyer, F., The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *BMC Bioinformatics* **2012,** *13*, 5.
118. Wilke, A.; Glass, E. M.; Bischof, J.; Braithwaite, D.; DSouza, M.; Gerlach, W.; Harrison, T.; Keegan, K.; Matthews, H.; Paczian, T.; Tang, W.; Trimble, W. L.; Wilkening, J.; Desai, N.; Meyer, F. *MG-RAST technical report and manual for version 3.3.6*; Argonne National Laboratory: Chicago, IL, USA., 2013.
119. Albertsen, M.; Hugenholtz, P.; Skarshewski, A.; Nielsen, K. L.; Tyson, G. W.; Nielsen, P. H., Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnol* **2013,** *31* (6), 533-538.
120. Crusoe, M. R.; Alameldin, H. F.; Awad, S.; Boucher, E.; Caldwell, A.; Cartwright, R.; Charbonneau, A.; Constantinides, B.; Edvenson, G.; Fay, S.; Fenton, J.; Fenzl, T.; Fish, J.; Garcia-Gutierrez, L.; Garland, P.; Gluck, J.; González, I.; Guermond, S.; Guo, J.; Gupta, A.; Herr, J. R.; Howe, A.; Hyer, A.; Härpfer, A.; Irber, L.; Kidd, R.; Lin, D.; Lippi, J.; Mansour, T.; McA'Nulty, P.; McDonald, E.; Mizzi, J.; Murray, K. D.; Nahum, J. R.; Nanlohy, K.; Nederbragt, A. J.; Ortiz-Zuazaga, H.; Ory, J.; Pell, J.; Pepe-Ranney, C.; Russ, Z. N.; Schwarz, E.; Scott, C.; Seaman, J.; Sievert, S.; Simpson, J.; Skennerton, C. T.; Spencer, J.; Srinivasan, R.; Standage, D.; Stapleton, J. A.; Steinman, S. R.; Stein, J.; Taylor, B.; Trimble, W.; Wiencko, H. L.; Wright, M.; Wyss, B.; Zhang, Q.; Zyme, E. n.; Brown, C. T. The khmer software package: enabling efficient nucleotide sequence analysis.
121. Zhang, Q. P.; Pell, J.; Canino-Koning, R.; Howe, A. C.; Brown, C. T., These are not the k-mers you are looking for: efficient online k-mer counting using a probabilistic data structure. *PLoS One* **2014,** *9* (7), 13.
122. Langmead, B.; Salzberg, S. L., Fast gapped-read alignment with Bowtie 2. *Nat Methods* **2012,** *9* (4), 357-U54.
123. Parks, D. H.; Imelfort, M.; Skennerton, C. T.; Hugenholtz, P.; Tyson, G. W., CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **2015,** *25* (7), 1043-1055.
124. Hyatt, D.; Chen, G. L.; LoCascio, P. F.; Land, M. L.; Larimer, F. W.; Hauser, L. J., Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **2010,** *11*, 11.
125. Kanehisa, M.; Sato, Y.; Morishima, K., BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* **2016,** *428* (4), 726-731.
126. Ogata, H.; Goto, S.; Sato, K.; Fujibuchi, W.; Bono, H.; Kanehisa, M., KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **1999,** *27* (1), 29-34.

127. Alikhan, N. F.; Petty, N. K.; Ben Zakour, N. L.; Beatson, S. A., BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **2011,** *12*, 10.
128. Bergmark, L.; Poulsen, P. H. B.; Abu Al-Soud, W.; Norman, A.; Hansen, L. H.; Sorensen, S. J., Assessment of the specificity of *Burkholderia* and *Pseudomonas* qPCR assays for detection of these genera in soil using 454 pyrosequencing. *FEMS Microbiol Lett* **2012,** *333* (1), 77-84.
129. Charan, S. S.; Pawar, K. D.; Severson, D. W.; Patole, M. S.; Shouche, Y. S., Comparative analysis of midgut bacterial communities of *Aedes aegypti mosquito* strains varying in vector competence to dengue virus. *Parasitol Res* **2013,** *112* (7), 2627-2637.
130. Ye, J.; Coulouris, G.; Zaretskaya, I.; Cutcutache, I.; Rozen, S.; Madden, T. L., Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **2012,** *13*, 11.
131. Wang, Q.; Garrity, G. M.; Tiedje, J. M.; Cole, J. R., Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **2007,** *73* (16), 5261-5267.
132. Grabowski, A.; Nercessian, O.; Fayolle, F.; Blanchet, D.; Jeanthon, C., Microbial diversity in production waters of a low-temperature biodegraded oil reservoir. *FEMS Microbiol Ecol* **2005,** *54* (3), 427-443.
133. Chuang, A. S.; Mattes, T. E., Identification of polypeptides expressed in response to vinyl chloride, ethene, and epoxyethane in *Nocardioides* sp. strain JS614 by using peptide mass fingerprinting. *Appl Environ Microbiol* **2007,** *73* (13), 4368-4372.
134. Case, R. J.; Boucher, Y.; Dahllof, I.; Holmstrom, C.; Doolittle, W. F.; Kjelleberg, S., Use of 16S rRNA and *rpoB* genes as molecular markers for microbial ecology studies. *Appl Environ Microbiol* **2007,** *73* (1), 278-288.
135. Graham, D. E.; Xu, H. M.; White, R. H., Identification of coenzyme M biosynthetic phosphosulfolactate synthase - A new family of sulfonate-biosynthesizing enzymes. *J Biol Chem* **2002,** *277* (16), 13421-13429.
136. Begley, J. F.; Czarnecki, M.; Kemen, S.; Verardo, A.; Robb, A. K.; Fogel, S.; Begley, G. S., Oxygen and ethene biostimulation for a persistent dilute vinyl chloride plume. *Groundwater Monit Rem* **2012,** *32* (1), 99-105.
137. Liang, Y.; Cook, L. J.; Mattes, T. E., Temporal abundance and activity trends of vinyl chloride (VC)-degrading bacteria in a dilute anoxic VC plume at Naval Air Station Oceana. *Submitted* **2016**.
138. Rui, L. Y.; Kwon, Y. M.; Reardon, K. F.; Wood, T. K., Metabolic pathway engineering to enhance aerobic degradation of chlorinated ethenes and to reduce their toxicity by cloning a novel glutathione S-transferase, an evolved toluene o-monooxygenase, and gamma-glutamylcysteine synthetase. *Environ Microbiol* **2004,** *6* (5), 491-500.
139. Vlieg, J.; Kingma, J.; van den Wijngaard, A. J.; Janssen, D. B., A glutathione S-transferase with activity towards *cis*-1,2-dichloroepoxyethane is involved in isoprene utilization by *Rhodococcus* sp. strain AD45. *Appl Environ Microbiol* **1998,** *64* (8), 2800-2805.
140. Vuilleumier, S.; Pagni, M., The elusive roles of bacterial glutathione S-transferases: new lessons from genomes. *Appl Microbiol Biotechnol* **2002,** *58* (2), 138-146.

141. Vandenwijngaard, A. J.; Janssen, D. B.; Witholt, B., Degradation of epichlorohydrin and halohydrins by bacterial dultures isolated from fresh-water sediment. *J Gen Microbiol* **1989,** *135*, 2199-2208.
142. Jacobs, M. H. J.; Vandenwijngaard, A. J.; Pentenga, M.; Janssen, D. B., Characterization of the epoxide hydrolase from an epichlorohydrin-degrading *Pseudomonas* sp. *Eur J Biochem* **1991,** *202* (3), 1217-1222.
143. Vetrovsky, T.; Baldrian, P., The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* **2013,** *8* (2), 10.
144. Wang, K. L. C.; Li, H.; Ecker, J. R., Ethylene biosynthesis and signaling networks. *Plant Cell* **2002,** *14*, S131-S151.
145. Mansouri, S.; Bunch, A. W., Bacterial ethylene synthesis from 2-oxo-4-thiobutyric acid and from methionine. *J Gen Microbiol* **1989,** *135*, 2819-2827.
146. Agency for Toxic Substances and Disease Registry (ATSDR) HazDat database: ATSDR's hazardous substance release and health effects database. http://www.atsdr.cdc.gov/ (accessed March 8, 2015).
147. Swartjes, F. A., *Dealing with contaminated sites: from theory towards practical application*. Springer Science & Business Media: 2011; p 264.
148. Bucher, J. R.; Cooper, G.; Haseman, J. K.; Jameson, C. W.; Longnecker, M.; Kamel, F.; Maronpot, R.; Matthews, H. B.; Melnick, R.; Newbold, R.; Tennant, R. W.; Thompson, C.; Waalkes, M., *Report on Carcinogens, Thirteenth Edition* U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program http://ntp.niehs.nih.gov/pubhealth/roc/roc13/: 2014.
149. Barbin, A.; Bresil, H.; Croisy, A.; Jacquignon, P.; Malaveille, C.; Montesano, R.; Bartsch, H., Liver microsome mediated formation of alkylating agents from vinyl bromide and vinyl chloride. *Biochem Biophys Res Comm* **1975,** *67* (2), 596-603.
150. Henschler, D., Toxicity of chlorinated organic compounds-effects of the introduction of chlorine in organic molecules. *Angew Chem Int Ed Engl* **1994,** *33* (19), 1920-1935.
151. Ensign, S. A.; Allen, J. R., Aliphatic epoxide carboxylation. *Annu Rev Biochem* **2003,** *72*, 55-76.
152. Matthews, R. G.; Goulding, C. W., Enzyme-catalyzed methyl transfers to thiols: the role of zinc. *Curr Opini Chem Biol* **1997,** *1* (3), 332-339.
153. Goulding, C. W.; Matthews, R. G., Cobalamin-dependent methionine synthase from *Escherichia coli*: Involvement of zinc in homocysteine activation. *Biochemistry* **1997,** *36* (50), 15749-15757.
154. Zhou, Z. H. S.; Peariso, K.; Penner-Hahn, J. E.; Matthews, R. G., Identification of the zinc ligands in cobalamin-independent methionine synthase (*MetE*) from *Escherichia coli*. *Biochemistry* **1999,** *38* (48), 15915-26.
155. Krum, J. G.; Ellsworth, H.; Sargeant, R. R.; Rich, G.; Ensign, S. A., Kinetic and microcalorimetric analysis of substrate and cofactor interactions in epoxyalkane : CoM transferase, a zinc-dependent epoxidase. *Biochemistry* **2002,** *41* (15), 5005-5014.
156. Furuhashi, K.; Taoka, A.; Uchida, S.; Karube, I.; Suzuki, S., Production of 1,2-epoxyalkane from 1-alkenes by *Nocardia corallina* B-276. *Appl Microbiol Biotechnol* **1981,** *12* (1), 39-45.

157. Clark, D. D.; Ensign, S. A., Evidence for an inducible nucleotide-dependent acetone carboxylase in *Rhodococcus rhodochrous* B276. *J Bacteriol* **1999,** *181* (9), 2752-2758.
158. Atashgahi, S.; Maphosa, F.; Dogan, E.; Smidt, H.; Springael, D.; Dejonghe, W., Small-scale oxygen distribution determines the vinyl chloride biodegradation pathway in surficial sediments of riverbed hyporheic zones. *FEMS Microbiol Ecol* **2013,** *84* (1), 133-142.
159. Begley, J. F.; Czarnecki, M.; Kemen, S.; Verardo, A.; Robb, A. K.; Fogel, S.; Begley, G. S., Oxygen and ethene biostimulation for a persistent dilute vinyl chloride plume. *Ground Water Monit Remediat* **2012,** *32* (1), 99-105.
160. Cook, L. J.; Hickman, G.; Chang, A.; Landin, P.; Reisch, T., Comparison of aerobic and anaerobic biotreatments of low-level vinyl chloride. In *Fifth International Conference on Remediation of Chlorinated and Recalcitrant Compounds*, Battelle Memorial Institute: Monterey, CA, May 22–25, 2006.
161. Bradley, P. M.; Chapelle, F. H. *Chloroethene biodegradation potential in the "lower" contaminant plume, River Terrace RV Park, Soldotna, Alaska*; Open-File Report 2004-1427; United States Gelogical Survey: Reston, VA, 2004.
162. Kreader, C. A., Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. *Appl Environ Microbiol* **1996,** *62* (3), 1102-1106.
163. Begley, J. F.; Hansen, E.; Wells, A. K.; Fogel, S.; Begley, G. S., Assessment and monitoring tools for aerobic bioremediation of vinyl chloride in groundwater. *Remediat J* **2009,** *20* (1), 107-117.
164. Culman, S. W.; Bukowski, R.; Gauch, H. G.; Cadillo-Quiroz, H.; Buckley, D. H., T-REX: software for the processing and analysis of T-RFLP data. *BMC Bioinformatics* **2009,** *10*, 171.
165. Jesus, E. D.; Marsh, T. L.; Tiedje, J. M.; Moreira, F. M. D., Changes in land use alter the structure of bacterial communities in Western Amazon soils. *ISME J* **2009,** *3* (9), 1004-1011.
166. Magurran, A. E., *Ecological diversity and its measurement*. Princeton, N.J. Princeton University Press., 1988.
167. Newell, C. J.; Rifai, H. S.; Wilson, J. T.; Connor, J. A.; Aziz, J. A.; Suarez, M. P., *Calculation and use of first-order rate constants for monitored natural attenuation studies*. United States Environmental Protection Agency, National Risk Management Research Laboratory: 2002.
168. Stothard, P., The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques* **2000,** *28* (6), 1102-1104.
169. Thompson, J. D.; Gibson, T. J.; Higgins, D. G., Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics* **2002,** *Chapter 2*, Unit 2.3.
170. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S., MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molec Biol Evol* **2011,** *28* (10), 2731-2739.
171. Alaminos, M.; Ramos, J. L., The methionine biosynthetic pathway from homoserine in Pseudomonas putida involves the *metW*, *metX*, *metZ*, *metH* and *metE* gene products. *Arch. Microbiol.* **2001,** *176* (1-2), 151-154.

172. Zhang, H. K.; Gao, S. H.; Lercher, M. J.; Hu, S. N.; Chen, W. H., EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res* **2012,** *40* (W1), W569-W572.
173. Tamura, K.; Dudley, J.; Nei, M.; Kumar, S., MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molec Biol Evol* **2007,** *24* (8), 1596-1599.
174. Krishnakumar, A. M.; Sliwa, D.; Endrizzi, J. A.; Boyd, E. S.; Ensign, S. A.; Peters, J. W., Getting a handle on the role of Coenzyme M in alkene metabolism. *Microbiol Molec Biol Rev* **2008,** *72* (3), 445-456.
175. Don, R. H.; Cox, P. T.; Wainwright, B. J.; Baker, K.; Mattick, J. S., Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res* **1991,** *19* (14), 4008-4008.