THESIS

BIOGEOCHEMICAL CHARACTERIZATION OF A LNAPL BODY IN SUPPORT OF STELA

Submitted by

Maria Irianni Renno

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2013

Master's Committee:

Advisor: Susan K. De Long
Co-advisor: Tom Sale

Thomas Borch
Fred Payne

ABSTRACT

BIOGEOCHEMICAL CHARACTERIZATION OF A LNAPL BODY IN SUPPORT OF STELA

Microbially-mediated depletion of light non-aqueous phase liquids (LNAPL) has gained regulatory acceptance as a method for managing impacted sites. However, the fundamental microbiology of anaerobic hydrocarbon degradation, in source zones, remains poorly understood. Two site-specific studies (Zeman, 2012 & McCoy, 2012) performed at the Center for Contaminant Hydrology (CCH), Colorado State University (CSU) demonstrated that LNAPL biodegradation increases drastically when temperatures are maintained between 18°C and 30 °C as compared to lower or higher temperatures. These results have supported the design of a Sustainable Thermally Enhanced LNAPL Attenuation (STELA) technology that is currently being tested at field scale at a former refinery in Wyoming. The focus of the present study was to perform a depth-resolved characterization of the mixed microbial communities present in LNAPL-impacted soils, as well as to characterize the site's geochemical parameters in order to establish a baseline data set to evaluate the STELA system performance.

Seventeen soil cores were collected from the impacted site, frozen on dry ice and subsampled at 6-inch intervals for analysis of biogeochemical parameters. Multi-level sampling systems were installed at the core sites to monitor aqueous and gas phases. Diesel and gasoline range organics and benzene, toluene, ethylbenzene and xylenes (BTEX) present in the cores and in water samples were analyzed. Temperature, inorganic dissolved ions, pH, and oxidation reduction potential (ORP) were also measured. DNA was extracted in triplicate from each subsample corresponding to the study's center core (21 samples). Total Eubacteria and Archaea

were quantified via 16S rRNA gene-targeted qPCR. Microorganisms present at selected depth intervals were identified via 454 pyrosequencing of both eubacterial and archaeal 16S rRNA genes.

Results indicate that at the study site, the majority of the hydrocarbon contamination is found between 5 and 12 feet below ground surface (bgs). The average of the maximum total petroleum hydrocarbon (TPH) soil concentrations within each core was 17,800 mg/kg with a standard deviation of 8,280 mg/kg. The presence of methane in the vadose zone and depleted sulfate concentrations in water samples suggest that both methanogenesis and sulfate reduction are likely driving LNAPL depletion processes. Four distinct biogeochemical zones where identified within the surveyed aquifer region. Interestingly, the quantity of eubacterial 16S rRNA genes dominate the quantity of archaeal 16S rRNA genes at sampled depths within the aerobic aquifer region. In the strictly anaerobic aquifer regions, these quantities are approximately equal. The latter can be interpreted as evidence of syntrophism, which has been reported in other hydrocarbon biodegradation studies. Pyrosequencing results support these findings as well and contribute to further elucidating the spatial correlation between microbial communities and geochemical parameters.

In- situ biodegradation rates are largely controlled by the quantity and activity of key microbes capable of mediating conversion of specific hydrocarbon constituents. Furthermore, it is anticipated that biodegradation rates are governed by complex interactions of diverse microbial communities that vary both in space and time. The overall vision of this initiative is that advancing a better understanding of processes controlling biologically mediated losses of

LNAPL will support the development of more efficient treatment technologies for LNAPL releases. In particular, the site specific analysis produced through this study will support the development of STELA.

ACKNOWLEDGMENTS

I would like to thank all who in one way or another contributed to my research.

Thank you, Dr. Susan K. De Long and Dr. Tom Sale for this opportunity. I truly feel your combined unique skills and perspectives brought the best out of me during my time at CSU. Tom and Susan, I really appreciate your patience and the time and effort you invested in me. Your commitment and dedication to our research field have been and continue to be an inspiration.

Thank you, Dr. Payne and Dr. Borch for agreeing to serve on my committee. I hold you both in very high regards and truly appreciate the bragging rights. I am looking forward to your feedback as I know it will be very valuable to my work.

Thank you, Adam Byrne and Daria Ahkbari for being my project buddies. You made it a lot of fun! I learned a lot from both of you; without your contributions this work would be incomplete. Adam, special thanks to you for generating all of the MVS images that are presented in this body of work.

Thank you, Paige Griffin Wilson, Allison Hawkins, Jazzy Jeramy Jasmann, Dr. Emilie Lefevre, Anna Skinner, Kristen Wiles, and Natalie Zeman for being excellent coworkers, great friends and even better sounding boards.

Mitch Olson and Dr. Julio Zimbron, thank you for your mentoring advice and your technical insights.

Thank you, CCH people: Zoe Bezold, Brett Ledecker, Dr. Jens Blotevogel, Calista Campbell, Mark Chalfant, Jeremy Chignel, Missy Tracy, Wess Tuli, Scott Williams, Ellen Daugherty, Gary Dick, Sonja Koldewyn, Saeed Kiaalohosseini, Jack Martin , Rene Santin, Kevin Saller, Emily Stockwell, Jennifer Wahlberg, for being great co-workers and for your valuable inputs to my work.

Thank you, Trihydro for the excellent field Support. Special thanks Alysha Anderson and Tom Gardner for your hospitality, the educational experience in the field and for the quality of samples you helped us obtain. Thank you, Ben Mc Alexander, Stephanie Whitfield, Kurt Toggle for the hospitality and the additional technical support.

Thank you, Mark Lyverse and Chevron team; and, thank you, Harley Hopkins and Exxon Mobil team for the technical and financial support , for engaging us in your discussions and providing us with field sites for our research.

Thank you, GTRG people (Meg, Andy, Shane, Jess and Patty) for being a phone call away when needed.

Drs: Ballare, Barreto, Battista, Csavina, Hanzlik, Komisar, Lefevre, Rossi, Blanchard and committee members thank you for being great academic examples and people of excellence.

Dani, Sim, Paul and Kate thank you for being such incredible role models and for always having my back! Baby, Barbie, Ceci, Tata, Agus, Paulie, Rubia, Mechi, Caro, Pia, Jose and Flor thank you for being the best friends in the world and always being there for me!

Cotos, Rennos, mama, papa, Chris and Tete thank you for your unconditional love and support. I love you all so much! Dion, mama y papa special thanks for always picking up my slack and creating the space I need to find my path. Dion, thank you for your love and partnership, I can do nothing without you!

Thank you, God for all of these opportunities and for putting all these people in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 Introduction

Petroleum liquids have been a cornerstone of modern industry for the last century (Sale, 2003). Unfortunately, standard practices of the last century have resulted in releases of petroleum liquids to shallow soils and ground water. After decades of active remediation, past petroleum releases remain as a social liability.

An emerging need is new technologies for petroleum liquids, herein referred to as LNAPL, that are more effective, lower cost and more sustainable than current options. A new technology, potentially meeting this need is STELA. STELA is based on the premise that low levels of heating can dramatically reduce the longevity of LNAPLs through biologically mediated processes. This hypothesis is being advanced through a field demonstration at a former refinery in Wyoming. The research presented herein provides a baseline for evaluating performance of the STELA field demonstration.

1.1 Research Objectives

This study performed a detailed depth-resolved biogeochemical characterization of a LNAPL body. The objectives of this investigation were 1) to elucidate the biogeochemical characteristics of the impacted site prior to treatment, 2) to develop an approach to determine the on-going system's performance, and 3) to develop an understanding of how complex interactions between contaminants and environmental conditions influence the structure of the microbial ecology of an impacted site, and thus, impact key microbial processes.

1.2 Thesis Overview

Chapter 2 provides background information on topics relevant to this thesis work. Chapter 3 is presented in manuscript form and corresponds to the baseline biogeochemical characterization of the impacted site prior to thermal treatment. Chapter 4 describes ongoing and future work. The final section presents the study's overall conclusions.

2.0 Literature Review

2.1 Petroleum Releases in the U.S.

LNAPL contamination in the environment is caused mainly by releases from petroleum production, storage, transmission and refining infrastructure. LNAPL releases have been a common occurrence in surface and ground water environments. An estimated hundred million gallons of oil (325,000 tons of oil), are accidentally spilled worldwide, on a yearly basis (Fingas, 2010). According to a report prepared for the American Petroleum Institute (API)(Etkin, 2009) , for the period between 1998 and 2007, approximately 4,770 tons of crude oil were released into the environment by offshore exploration and oil platforms in the U.S and 94,009 tons of oil were released from pipelines associated with offshore oil and inland production in U.S. waters. This study also reported that 4,026 tons of petroleum liquids are released to terrestrial environments and ground water sources in the U.S. annually. The main sources of petroleum associated with inland contaminant releases are storage facilities, refineries, and pipelines (Etkin, 2001).

Although the drastic scale of offshore releases such as BP's Deepwater Horizon release in the Gulf Coast in 2010 (666,400 tons of oil, (Peterson et al., 2012)) and Exxon Valdez's release in 1989 (34,900 tons of oil (Sylves & Comfort, 2012)) attract major publicity, in the U.S, inland spills occur more often than coast line or Deepwater Horizon's spills. A recent publication cited an American database where 88% of petroleum spills over 10,000 gallons corresponded to inland spills. Furthermore, petroleum hydrocarbons are deemed to be the most predominant soil contaminants in the U.S. There are between 400,000 to 500,000 contaminated sites in the

U.S. that could potentially translate into 12 billion U.S. dollars for the remediation industry (Zvomuya & Murata, 2012). Close to a half of these sites are hydrocarbon impacted sites (EPA, 2013a).

2.2 Chemical Composition of LNAPL

Non-aqueous phase liquids (NAPL) are a common source of contaminants in ground water, soil and gas. NAPL can be denser than water (DNAPL) or lighter than water (LNAPL). LNAPLs are commonly composed of petroleum liquids, which often contain a wide range of hydrocarbons (Nadim et al., 2000).

Alkanes, the main compounds in crude oil, are composed of hydrogen-saturated carbon chains. Alkenes are also present in petroleum and are carbon molecules that contain one or more carbon-carbon bonds that are unsaturated (i.e., alkenes contain at least one carbon double bond). Unsaturated carbon chains with triple carbon-carbon bonds are called alkynes. Acetylene is an example of this type of hydrocarbon and can be present in petroleum crudes. Alkanes and alkenes can be linear or branched and present chains of varying lengths. Cycloalkanes are hydrocarbons that contain one or more carbon ring structures. Aromatic hydrocarbons also can be present in petroleum liquids and are composed of one or more substituted or unsubstituted benzene rings. Benzene rings are six-carbon cyclic hydrocarbons that contain alternating carbon-carbon double bonds and carbon-carbon single bonds. Benzene, toluene, ethylbenzene and xylenes (BTEX) compounds are all aromatic hydrocarbons. Hydrocarbons containing multiple aromatic rings are called polycyclic aromatic compounds

(PAHS)(Schwarzenbach et al., 2005). BTEX and PAHs are the compounds that establish risk and thus determine remedial action due to their recalcitrance and high toxicity.

A less technically-rigorous, industry-adopted classification of petroleum-based compounds separates these into paraffinic and asphaltic/naphthenic oils. Paraffinic-based crudes are rich in straight-chain carbon compounds, whereas asphaltic crudes are rich in cyclic crude compounds (School, 1954).

Total petroleum hydrocarbon (TPH) is a term used to express petroleum-derived hydrocarbon concentrations in gas, water and soil samples and includes gasoline and diesel range organics (GRO and DRO, respectively). These two main hydrocarbon classification groups separate compounds based on their boiling points. GRO compounds have boiling points close to those of gasoline (60ºC – 170 ºC) and include C_{4}-$_{10}$ alkanes, C_{4}-$_{7}$ alkenes, and aromatics (BTEX are considered GRO compounds). DRO compounds have boiling points close to that of diesel fuel (170ºC -400 ºC) and include C_{8}-$_{12}$ to C_{24}-$_{26}$ alkanes and PAHs (Zvomuya & Murata, 2012).

2.3 Transport and Environmental Fate of LNAPLs

Petroleum is a mixture of naturally occurring hydrocarbons that can be present in the solid, liquid or gaseous phase. Temperature and pressure affect the state at which petroleum components are found in the environment (Uren, 1956). The fate of hydrocarbon contaminants released to the environment is dictated by the chemical composition of the petroleum released (i.e., whether it is crude or refined, and whether it is composed mainly of heavy or light hydrocarbon chains) and the location where the release occurred (marine waters, inland waters, soil environment, wetlands, industrial areas, or rural or urban locations)(Etkin,

2009). The aforementioned factors will also influence the remedial action taken to address the spill or leak.

When a leak or spill occurs, LNAPL will migrate downward due to gravity until it reaches the water table. Some LNAPLs can become trapped in pore spaces during this downward movement through the unsaturated zone (subsurface region existing between the soil surface and the water table). At the capillary fringe (zone immediately above the water table that is fully saturated with water due to capillary pressure), lateral movement of the LNAPL pool occurs. The extent of the lateral spread depends largely on the amount of LNAPL present (Charles J. Newell et al.,1995). Vertical smearing of LNAPL in the unsaturated zone may also occur if the water table fluctuates.

LNAPL trapped in the pore space along the vadose zone can dissolve into flowing water. Water flow in the vadose zone can occur due to infiltration caused by precipitation. When LNAPL constituents become mobile and reach the water table, contaminant plumes can form. Contaminants can also partition into the gas phase forming vapor plumes. Chemical constituents in LNAPL can also sorb to solid surfaces. Partition coefficients define the tendency of a compound to exist in a given phase (LaGrega et al., 1994). LNAPL exists as a free phase, the contaminant can partition between four different phases (Figure 2.1).

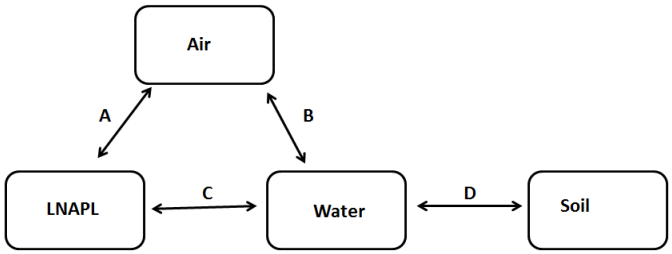


Figure 2.1: Four-phase system in LNAPL zones. System represents a water wet porous media, adapted from (DiGiulio, 1992). Letters and arrows represent contaminant partitioning between different phases.

Partitioning between LNAPL and the gas phase (Figure 2.1-A) can be described by Equation 1 that combines the ideal gas law with Raoult's law. Partitioning between aqueous and gas phase (Figure 2.1-B) can be described by Henry's law (Equation 2). Partitioning from LNAPL into aqueous phase (Figure 2.1-C) can be described by Raoult's Law (Equation 3), and partitioning from the aqueous phase to the solid phase (Figure 2.1-D) can be described by the organic carbon partition coefficient KOC (Equation 4).

Equation1: $C_{gas} = \chi(P_i/(RT)$

Where:

C_{gas} = concentration of the LNAPL compound in the gas phase (M/L^3)

χ = the mole fraction of the compound in the LNAPL phase (dimensionless)

P_i = the pure compound vapor pressure (F/L^2)

R = the universal gas constant ($(L^3.F/L^2)/(T.mole)$)

T = temperature

Equation 2: $H = C_{gas}/C_{water}$

Where:

C_{gas} = concentration of the LNAPL compound in the gas phase (M/L^3)

C_{water} = concentration of the LNAPL compound in aqueous phase (M/L^3)

H = Henry's constant (dimensionless)

Equation3: $C_J=\chi C_{jNAPL}$

Where:

C_{JNAPL} = concentration of LNAPL compound in LNAPL phase (M/L^3)

C_j = concentration of LNAPL compound in aqueous phase (M/L^3)

χ = mole fraction of compound in LNAPL (dimensionless)

Equation 4 : $w_{sorbed} = f_{oc} K_{oc} C_{water}$

Where:

w_{sorbed} = mass sorbed (M/M)

C_{water} = concentration of the LNAPL compound in the aqueous phase (M/L^3)

f_{oc} = fraction of organic carbon present in the soil (dimensionless)

K_{oc} = organic carbon partitioning coefficient (L^3/M)

Properties of the porous media that contains LNAPL also dictate its environmental fate. For example, as can be inferred from equation 4, the amount of contaminant mass that is sorbed to the soil phase increases as the carbon content of the soil increases. Aquifer

characteristics also influence LNAPL migration patterns. An EPA study (Charles J. Newell et al., 1995) based on J.W. Mercer's and R. Cohen's research presented in 1990 (Mercer & Cohen, 1990) identified the following factors as influential in LNAPL migration at the field scale: volume of LNAPL released, type of release (continuous leak or one time large volume spill), LNAPL properties such as density, fugacity and viscosity, permeability, pore size and size distribution.

Another important parameter to consider when predicting LNAPL migration through porous media is the LNAPL residual saturation (Sr); this value describes the fraction of pore space occupied by non-aqueous fluid at which a rapid increase with capillary pressure does not translate into fluid decreased saturation (Corey, 1994). Due to the existence of residual saturation, an amount of LNAPL remains trapped in the pore space, establishing a secondary contaminant source that can gradually dissolve into groundwater systems, when environmental changes such as infiltration of precipitation or seasonal water table fluctuations, occur.

In addition to these subsurface transport phenomena, biodegradation processes also can take place, which can change LNAPL composition over time further complicating the prediction of its environmental fate. Hydrocarbons have different susceptibilities to microbial degradation; for example, short linear alkanes are easier to biodegrade than complex PAHs. As the more readily biodegradable compounds get depleted LNAPL phases are enriched in asphaltic compounds and become less mobile and more recalcitrant. Understanding subsurface contaminant distribution and biogeochemistry can help predict contaminant's environmental fate. The latter is important when deciding on a remedial action.

2.4 Remediation technologies

The EPA accounts for over 200,000 hydrocarbon-contaminated sites in the U.S and in the year 2002, allocated over 23 million dollars to the development of hydrocarbon clean up technologies (EPA, 2013b). There are two main remediation approaches to treating organic contaminants in soil: extractive methods and destructive approaches. The extractive methods involve removal of the contaminant from the ground. Destructive approaches are often implemented in-situ. Waste disposal and post-removal treatment, which are required for extractive methods, are often more expensive alternatives (Hua & Hopf, 2006).

Some popular in-situ remediation technologies implemented to address hydrocarbon contamination include: hydraulic recovery, physical barriers, soil vapor extraction (SVE), air sparging in combination with SVE, enhanced bioremediation and monitored natural attenuation (MNA) (LaGrega et al., 1994). Barriers or containment walls are permanent or replaceable units that are set to interrupt the flow path of a contaminant plume. Some physical barriers are designed to contain the contaminant within a controlled space. Barriers can also be reactive and have the primary objective of treating the contaminant as it flows through the barrier (LaGrega et al., 1994).

SVE is another in-situ treatment option for hydrocarbon impacted soils SVE is typically utilized to remove volatile organic contaminants (VOCs) from soils in the vadose zone. This technology is based on a flushing air flow using wells in the vicinity of the contaminated source. Contaminants evaporate from the soil matrix into the soil air space and the extracted vapors are directed from an extraction well into a treatment system. A variation on this technology is

known as venting. In this SVE variation, an air stream is introduced into the system. The air sparged into the system causes the VOCs sorbed (soil) or dissolved (water) phase to move into the air stream. The air stream is then directed to an extraction well and into a storage or treatment unit (Kitanidis & McCarty, 2010). Introducing air into a system (bioventing) has the added advantage of introducing oxygen and potentially enhancing biodegradation rates. However, pumping air into the subsurface can be expensive and in high contaminant zones oxygen is readily depleted; therefore, large volumes of air must be pumped in order to achieve the desired results.

When feasible, MNA is a desirable remedial approach, given that it is a relatively inexpensive technology, and it is minimally invasive in comparison with the other in-situ treatment technologies (Wiedemeier et al., 1999). MNA is an *in-situ* remediation technology based on proving that naturally occurring processes are reducing contaminant mass in soil and groundwater (Jørgensen et al., 2010). MNA, according to the U.S. EPA, is a clean approach that relies on long-term performance monitoring to determine whether or not well established remedy goals are being met (EPA, 2004). There are several chemical, physical and geological factors that have an influence on natural attenuation (NA) rates. Given the complexity of the interrelationships between the biogeochemical factors affecting NA rates, NA performance monitoring and NA monitoring system design varies site to site. Naturally occurring processes that drive attenuation rates include dilution, dispersion, volatilization, sorption, physical and chemical transformations and biodegradation (Röling & van Verseveld, 2002). Non-destructive processes are those that reduce the contaminant concentration without achieving total mass reduction, such as contaminant dissolution into the aqueous phase, or contaminant sorption

from a plume to a solid surface. Destructive processes are those that destroy the contaminant such as bioremediation, photolysis or hydrolysis (abiotic). Biodegradation is the more significant process in hydrocarbon natural attenuation, given that it effectively reduces contaminant mass and can occur under several environmental conditions (Lee & Lee, 2003).

Three lines of evidence are commonly considered to prove that NA is a viable treatment option for a particular site (Wiedemeier et al., 1999), and they include:

1- A historical database proving that the site's contaminant concentration has diminished over time.
2- Geochemical data from the site demonstrating that electron donor and/or elector acceptor have been depleted and concentrations of metabolic byproducts and/ or biodegradation intermediates have increased, and
3- Microbiological data that provides information regarding in-situ biodegradation rates.

Remedy designs often combine MNA with another technology to increase biodegradation rates. Common NA enhancement technologies include: the addition of specific microorganisms (bioaugmentation),or the addition of nutrients or/and electron donors (bio-stimulation). Molasses and vegetable oil are two common electron donor sources used in the remediation industry for reductive dehalogenation (Antizar-Ladislao, 2010). Common electron acceptors added as biostimulants, for hydrocarbon oxidation, include: oxygen (air sparging), iron (ZVI clay or PBRs), and sulfate as gypsum (solid) or dissolved in recirculating water.

The geologic complexity of a site in combination with the complexity of the factors discussed in the previous sections demand site-specific solutions. The first step towards tailored remedial actions is to quantitatively determine the extent of the problem and to set quantifiable goals (Sale, 2003). The latter requires the development of a thorough site conceptual model prior to the remedial design. An understanding of site-specific microbiology and geochemistry is important to guide the selection of optimal remedies and system design.

2.5 Hydrocarbon Biodegradation

Under appropriate environmental conditions, LNAPL can be completely mineralized to carbon dioxide and water (Sihota et al., 2010). Commonly identified hydrocarbon biodegradation processes include: aerobic respiration, nitrate reduction, iron reduction, sulfate reduction and methanogenesis. However, depending on a site's geochemistry and the age of the spill, different electron accepting processes can occur. Additionally, the dominant biodegradation/biotransformation processes vary spatially, at a given site, as illustrated in a conceptual model of an LNAPL-impacted site presented in Figure 2.2.

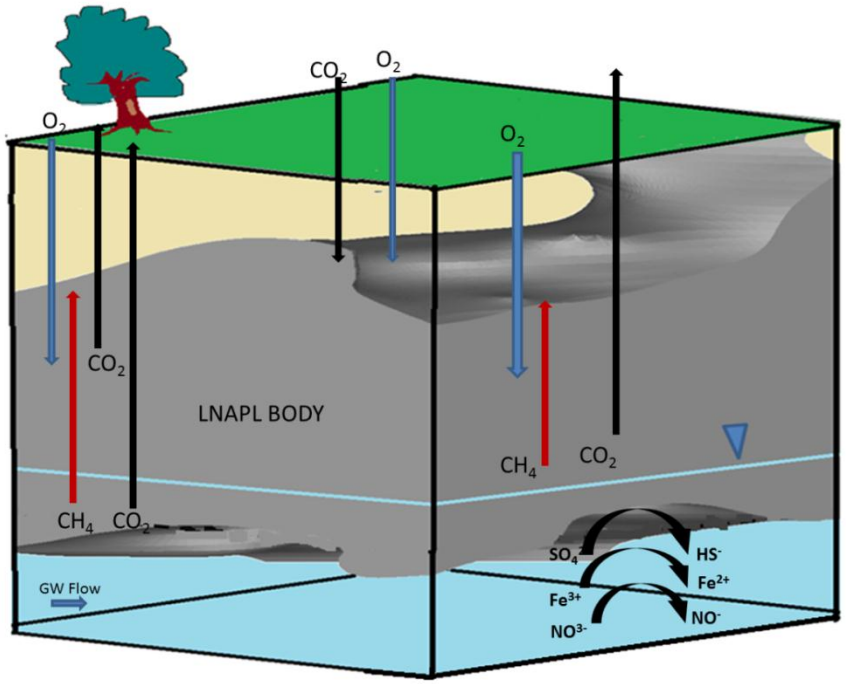


Figure 2.2: Conceptual model of a hydrocarbon impacted site. Drafted based on (Sihota et al., 2010). Water is represented in blue. LNAPL is represented in gray, and soil is represented in yellow. Water table is indicated in the Figure. Carbon dioxide that diffuses in from the atmosphere and that is produced by hydrocarbon biodegradation under different metabolic processes is represented by black arrows. Oxygen diffusion into the subsurface is represented by blue arrows. Red arrows represent methane flux created by methanogenic LNAPL degradation.

Close to the surface, critical aerobic processes associated with the biodegradation of hydrocarbons can be observed. Depending on the nature of the water body (its depth and proximity to surface waters) oxygen dependent processes can also take place at the edge of the contaminant plume. Aerobic oxidation yields the highest amount of energy for cellular metabolism and growth, while methanogenesis yields the least amount of energy. Through the aerobic route, various hydrocarbon compounds, including PAHs are readily degraded (Lahvis et

al., 1999). Methane oxidation, which produces carbon dioxide, is another important process that can occur where methane and oxygen fronts meet (Ma et al., 2012; Johnson et al., 2006).

LNAPL-impacted zones are electrochemically diverse. Oxygen is depleted rapidly at the core of LNAPL zones and at the center of contaminant plumes, due to high contaminant levels present. Thus, Anoxic zones are often established deeper in the subsurface and in high contaminant regions like the core of an LNAPL body (gray regions in Figure 2.2). The biogeochemical characteristics of the site will dictate the biodegradation process taking place. It is common to observe different processes occurring in the proximity of the same source zone (Naidu et al., 2012). Pathways of different contaminant oxidation processes can often be interrelated given that they are occurring within proximity of each other.

Generally, alkanes are aerobically degraded by a common route that includes a stepwise oxidation (HC → alcohol →aldehyde →acid) and produces intermediates that will be metabolized via the Krebs cycle (Savage et al., 2010). Aerobic degradation pathways for BTEX and PAHs can differ among different organisms, but all of these reaction pathways funnel into a central pathway that leads to the Krebs cycle. The initial step in cellular metabolism of PAHs is ring cleavage. This step is often catalyzed by oxygenases, which add a hydroxyl group to destabilize the aromatic ring. There exist a wide variety of monooxygenases and dioxygenases that catalyze this crucial step. Catechols, protocachetuates, gentisates and bensoquinols are common intermediates found in the aerobic degradation of BTEX and more complex molecules as PAHs (Pérez-Pantoja et al., 2010). New intermediates are being discovered in the aerobic degradation of hydrocarbons, and some of these can be produced by oxygen-independent

pathways suggesting these intermediates might be the link between aerobic and anaerobic hydrocarbon degradation pathways.

In contrast with aerobic hydrocarbon degradation pathways, where oxygen is utilized both as an electron acceptor and as an activation reactant, anaerobic hydrocarbon biodegradation pathways present unique biochemical activation steps (Widdel & Rabus, 2001). BTEX compounds, in particular benzene, have become a focus of anaerobic biodegradation research in the past two decades. BTEX are often regulatory drivers since they are very mobile compounds, given their high solubility values; benzene is also a known carcinogen. Furthermore, BTEX compounds, especially benzene, can be very recalcitrant in anoxic environments (van Agteren et al., 1998). The electronic configuration of aromatic rings makes them highly stable molecules; thus, ring cleavage requires a significant energy input.

There are four well-researched, anaerobic hydrocarbon degradation enzymatic pathways. The initial steps in these pathways are: 1) addition of fumarate yielding aromatic substituted succinates, 2) methylation of unsubstituted aromatics, 3) hydroxylation of an alkyl substituent via dehydrogenases, and 4) direct carboxylation of the aromatic compound (Weelink, 2008; Chakraborty & Coates, 2005; Foght, 2008). These four activation reactions occur in ring cleavage or substrates that can be β-oxidized. Central intermediates (e.g., benzoyl-co-A) are produced by the previously described reactions. Such metabolites can be further oxidized or used to produce biomass (Tierney & Young, 2010).

Despite BTEX and PAHs being relatively thermodynamically stable compounds under anaerobic conditions, several studies have been able to demonstrate their biodegradation using

a range of electron acceptors (Chakraborty & Coates, 2005; Jahn et al., 2005; Beller et al., 1992; Evans & Fuchs, 1988; Sakai et al., 2009). A comprehensive table describing these processes, the thermodynamic value for the reactions, terminal electron acceptors (TEAs) involved and biodegraded contaminants can be found in a recently published review regarding hydrocarbon biodegradation under anaerobic conditions (Foght, 2008).

2.6 Syntrophic Interactions in Hydrocarbon-Degrading Communities

As discussed in Section 2.5, in LNAPL bodies, methanogenesis is a prevalent process because other electron acceptors are rapidly depleted (Sihota & Mayer, 2012). Studies in which methanogenesis has been established as the predominant biodegradation process highlight the need for syntrophic microbial interactions to occur in order to overcome thermodynamic and kinetic limitations (extremely low energy yields and metabolite buildup) in these habitats (Jones et al., 2007; Dojka et al., 1998; Head et al., 2010). Syntrophic microbial interactions can add complexity to elucidating in-situ biodegradation pathways at LNAPL-impacted sites.

Syntrophy is a metabolic process that involves at least two organisms. Syntrophic processes have been defined as “tightly coupled mutualistic interactions” (Sieber et al., 2012) where the pool size of exchanged intermediates between involved organisms has to be kept low for the cellular cooperation to continue. Syntrophy has also been defined as a “survival strategy” (Morris et al., 2013) adopted by ecological systems where thermodynamically favorable processes cannot occur as isolated processes. In these types of systems, energy can be obtained by microbial communities that have the ability to couple unfavorable energetic processes together in order to be able to net an overall exergonic reaction.

Syntrophic processes have been found to occur in combination with a range of electron accepting processes, including nitrate reduction, iron reduction, sulfate reduction and methanogenesis (Kleinsteuber et al., 2012). An example of key processes found in sulfate-dependent syntrophic hydrocarbon-degrading communities is presented in Figure 2.3. Key processes include fermentation (Figure 2.3- A and D), hydrogenothrophic methanogenesis (Figure 2-3; B), acetoclastic methanogenesis (Figure 2.3-C), syntrophic fermentation (done by secondary fermenters) (Figure 2.3- D), homo-acetogenesis (Figure 2.3-E), lithotrophic sulfate reduction (Figure 2.3- F and G), and acetate oxidation via syntrophic sulfate reduction (Figure 2.3-H).

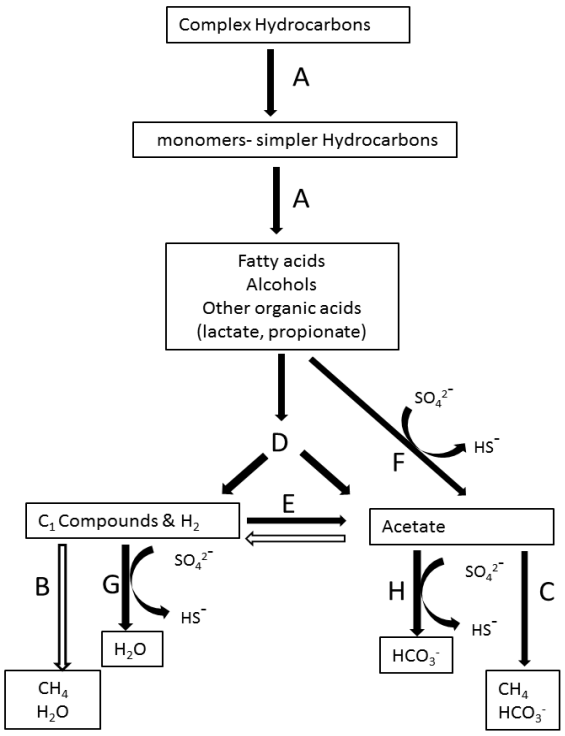


Figure 2.3: Key syntrophic processes in a sulfate-reducing/methanogenic community. White arrows indicate reactions that are less likely to occur. Adapted from (Payne et al., 2006) and (Mbadinga et al., 2011).

In a recent review, the syntrophic processes in a microbial community were summarized into three sequential processes (Sieber et al., 2012). Initially, primary fermenters hydrolyze complex polymers (such as cellulose, proteins or complex hydrocarbons). Simpler molecules can then be fermented to acetate or longer fatty acid chains, such as butyrate and propionate, and to carbon dioxide and hydrogen. The second step involves secondary, syntrophic fermentative metabolism of the molecules originated from the first step, such as benzoate, other complex fatty acids, and other complex molecules such as alcohols and PAHs. Lastly the molecular hydrogen, the acetate, and the carbonic acid produced during the earlier steps are metabolized to carbon dioxide and methane via acetoclastic and hydrogenothrophic methanogens. Although, most of the downstream degradation processes in syntrophic anaerobic degradation of hydrocarbons are known, the initial activation steps of recalcitrant molecules such as benzene, is not yet understood.

2.7 Hydrocarbon-Degrading Microorganisms

While high concentrations of petroleum liquids have been proven to be detrimental to many bacterial species due to the toxic effect these contaminants have on bacterial cell walls (Bell et al., 2013), other species can thrive in the presence of hydrocarbons and can utilize these contaminants as an energy source. Enrichment studies and pure culture isolation studies have identified microorganisms capable of growing on a range of hydrocarbons.

Under aerobic conditions *Rhodococcocus spp.*, *Mycobacterium spp., and Sphingomonas spp*. have been shown to be able to degrade longer hydrocarbon chains as well as PAHs (Kanaly & Harayama, 2000). Other organisms isolated from aerobic hydrocarbon-degrading

communities include *Pseudomonas aeuroginosa*, *Pseudomonas putida* and *Flavobacterium sp.* (Trzesicka-Mlynarz & Ward, 1995).

Under denitrifying and micro-aerophilic conditions, *Pseudomonas aeuroginosa* also has been shown to degrade aromatic compounds (Chayabutra & Ju, 2000). Other identified microorganisms associated with nitrate-reducing processes involving hydrocarbon degradation include: *Pseudomonas balearica, Vibrio pelagius,* Thauera *aromatica, Azoarcus toluovorans, Magnetospirillum sp., Nitrospira sp.*, *Nitrosomonas sp.*, *and Chlorobium sp.* (Zedelius et al., 2011), (Rockne et al., 2000), (Kuntze et al., 2011), (Yagi et al., 2010). *Dechloromonas spp.* are nitrate-reducing bacteria that can grow in pure cultures utilizing benzene as their sole carbon source (Kuntze et al., 2011). *Dechloromonas spp*. were originally isolated under chlorate-reducing conditions. *Geobacter metallireducens* has the ability of degrade complex hydrocarbons under nitrate-reducing conditions as well as under iron-reducing conditions (Lovley et al., 1993).

Other iron-reducing, hydrocarbon-degrading microorganisms have been found in enrichment culture studies. These microorganisms have been phylogenetically affiliated with the *Deltaproteobacteria* family (Botton & Parsons, 2007; Kunapuli et al., 2007) and the Firmicutes phylum (Kunapuli et al., 2007). Recently, a strictly iron-reducing, thermophilic, benzene-degrading pure culture has been isolated and identified as *Ferroglobus placidus* by (Holmes et al., 2011). *Desulfuromonas palmitati* is another strict anaerobe that can reduce iron to oxidize long complex hydrocarbon derived chains (Coates et al., 1995).

Several studies have also identified hydrocarbon-degrading microorganisms under sulfate-reducing conditions. Sulfate reducers identified include: *Desulfobacula toluolica, Desulfobacula phenolica* (Widdel & Rabus, 2001)*, Desulfotomaculum sp., Desulfosarcina sp., Desulfobacter sp.* (Kniemeyer et al., 2007), *Desulfatibacillum alkenivorans* (Cravo-Laureau et al., 2004) and *Desulfovibrio* sp. (Voordouw et al., 1996). Many of the sulfate-reducing genera contain metabolically diverse organisms that have the ability of obtaining energy via classic lithotrophic sulfate reduction processes or can act as secondary fermenters under methanogenic conditions. Some of these organisms include members of the genus *Desulfovibrio*. Of particular note are *Pelotomaculum sp*., which are often grouped with sulfate reducers because they carry genes that encode for two key enzymes involved in sulfate reduction (adenylyl sulfate reductase and dissimilatory sulfite reductase (Plugge C.M., 2011)); however, *Pelotomaculum spp*. are fermenters that grow via syntrophic interactions. *Pelotomaculum spp.* have been associated with the anaerobic degradation of benzene (Kleinsteuber et al., 2008; Kleinsteuber et al., 2012; Vogt et al., 2011) .

Other fermenting organisms associated with syntrophic hydrocarbon degradation pathways include members of the genus *Syntrophus* (Siddique et al., 2011; McInerney et al., 2007). Secondary fermenters that scavenge simpler acids in syntrophic degrading communities include *Anaerolinea sp., Bacteroides sp.,* and *Synergistes sp*. among others (Kleinsteuber et al., 2012). Secondary fermenters are often found living in association with methanogenic Archaea. There exist two main types of methanogenic Archaea: acetoclastic and hydrogenothrophic methanogens. These organisms have the important task of keeping hydrogen and acetate levels low so the syntrophic hydrocarbon-degrading processes remain energetically favorable.

Methane produced via acetoclastic methanogenesis is derived from acetate. Acetoclastic methanogens include *Methanosaeta spp.* and *Methanosarcina spp.* among others. Under high acetate concentrations, *Methanosarcina spp.* have been shown to outcompete *Methanosaeta spp.* (Galand et al., 2005). Hydrogenothrophic methanogens include members of the genera *Methanobrevibacter* (Asakawa et al., 1993) and *Methanoculleus* (Blotevogel et al., 1991). They are both within the order of *Methanosarcinales* and produce methane from H_2 or propionate but cannot metabolize acetate.

2.8 Molecular Biology Tools

There are two aspects of the utilization of molecular microbiological methods that serve the remediation industry. The first one is that of providing evidence for biodegradation processes occurring at a site, and thus satisfying regulators and site owners regarding the performance of a bio-based technology. The second one is that of discovering new processes and organisms with the potential to deplete or attenuate contaminants, with the objective of applying relevant findings to the development of new remediation technologies.

Relevant questions to the remediation industry that molecular biology tools can address were compiled by an expert panel assembled by SERDP and ESTCP (Alleman et al., 2005). These questions included:

- What is the potential for degradation to occur based on the presence or absence of genes or microorganisms of interest?

- What is the link between presence/absence of target genes or microorganisms and the activity of interest?
- Is the spatial and temporal distribution of organisms appropriate to meet goals?
- Can the biological process be limited by an environmental constraint?
- Can the desired process be enhanced in order to improve degradation rates?
- What controls the metabolism of key organisms in a given environment?
- Which microorganisms the key players?
- What are the environmental factors that structure a successful biodegrading community?

Molecular biology tools, as defined by the aforementioned expert panel, include tools that target biomarkers to provide information regarding microorganisms and processes related to biodegradation in natural or engineered systems (Alleman et al., 2005). These biomarkers can include DNA sequences, RNA sequences, peptides, proteins and lipids. Tools that target DNA are the most often applied in the field because DNA is the most stable biomolecule, and it is easier to isolate from environmental samples. DNA gives information regarding the potential for a certain process to occur; however, it does not prove that a biodegradation process has occurred or is occurring. For this reason, molecular biology tools that target DNA sequences have to be complimented by tools that provide other lines of evidence. Tools that target RNA molecules, proteins, or measure metabolic activity are useful to the remediation field as they provide direct evidence of gene expression or cellular function.

One of the most targeted biomarkers in environmental molecular biology studies, including those pertaining to bioremediation, is the 16S small subunit ribosomal RNA gene (16S rRNA gene) (Whitby & Lund, 2009). 16S rRNA genes are an ideal biomarker to survey the microbial composition of environmental samples because these genes are found in every prokaryote (Bacteria and Archaea), they have largely conserved structure, function, and genetic sequence, and they contain both constant and variable genetic regions (Rastogi & Sani, 2011). The constant regions in this gene allow for groups of organisms (e.g., Bacteria) to be targeted with one single tool or assay. In contrast, the variable regions within the 16s rRNA gene allow for different types of organisms within the same kingdom to be further classified according to their phylogeny.

Polymerase chain reaction (PCR) based methods are the key technologies used to investigate genetic biomarkers (e.g., 16s rRNA genes). PCR is a method that exponentially amplifies short DNA sequences (~50-1000 base pairs). A set of primers complimentary to the interrogated genetic sequence is needed. End-point PCR allows for determination of the presence or absence of a certain gene. Quantitative PCR (qPCR) is a variant on PCR that allows for detection and relative quantification of genes within a sample. qPCR has become an accepted screening and monitoring tool for tracking growth and distribution of organisms in a given environment (Alleman et al., 2005). Terminal restriction fragment length polymorphism (T-RFLP), denaturing gradient gel electrophoresis (DGGE), and 16s rRNA gene clone libraries are other PCR-based technologies that are utilized in the investigation of microbial environmental samples. T-RFLP is a fingerprinting method that identifies operational taxonomic units (OTUs), but does not provide genetic sequences or phylogenetic identities of microorganisms present in

a given sample. 16s rRNA gene clone libraries can be used to provide an indication of microbial diversity. Sequencing of clones is possible from the constructed libraries so novel organisms can be identified. Community structure can also be established through sequencing OTUs identified via DGGE. However, all of these methods are labor intensive and are not widely commercially available.

In contrast, recent advances in gene sequencing technologies have made high-throughput sequencing (i.e., pyrosequencing) of DNA from environmental samples commercially available. This is bringing about great progress in the understanding of microbial communities in a range of environments and engineered systems (Fierer et al., 2012). Cost-effectiveness, automation, increases in both the achievable read sequence length and the accuracy of the sequencing process (in non-Sanger-based technologies) are important technology advancements that are contributing to the fast progress of environmental molecular biology (Shendure et al., 2008). Cyclic Array Sequencing is one of the recently released gene sequencing technologies, and 454 FLX (Roche^{TM} , Branford, CT) and Solexa (Illumina^{TM}, San Diego, CA) systems are included in this category. The sequencing features are clonal in nature in the sense that each resolvable unit contains DNA from only one species (Shendure et al., 2008). The DNA is first prepared by creating blunt ends and attaching universal adaptors primers (short DNA sequences) to the ends. Each individual DNA fragment (that has the same adaptor on its end) is then attached to a bead, which is isolated in a water droplet immersed in an oil solution where each individual PCR reaction takes place. This form of PCR (emulsion PCR) ensures that each reaction amplifies a 16S rRNA gene from a single species. After the amplification step, each bead is captured in a reaction well where DNA sequencing

reactions occur (Lorato, 2010). Detection of each nucleotide is based on real-time luciferase-mediated pyrophosphate release. Generated sequences are then compared against databases containing known sequences of 16S rRNA genes to obtain detailed information on the phylogenetic identity of the microorganisms present in the surveyed sample. More specifically, the microbial identities can be established based on the non-conserved regions of the 16s rRNA genes.

A system's biology approach to understanding critical processes and key players involved in *in situ* contaminant degradation is becoming an important component in the design of bio-based remedies (Chakraborty et al., 2012). Molecular microbiology tools help inform this holistic approach which is based on understanding all components of a system. In a microbial hydrocarbon – degrading community, important interrelated parameters to characterize in order to understand key processes are: electron donors (contaminants), electron acceptors, soil types, type and number of microorganisms present as well as their community role (biological function).

2.9 Summary

Industrial petroleum exploitation has resulted in increased environmental hydrocarbon contamination since the beginnings of the previous century. Once petroleum liquids enter the environment, the contaminant's fate is influenced by aquifer structure, climate conditions, land use, and aquifer ecology. Over the past two decades, non-invasive remediation technologies such as MNA or enhanced NA have gained momentum. Developing appropriate technologies to support remedy designs, based on exploiting natural contaminant depletion processes, has

become a key focus point in the engineering research community. Furthermore, recent advancements in molecular biology and microbiology have highlighted the importance of understanding the complex interactions between environment, microbial community and contaminant present when implementing treatment systems that rely on biodegradation. There is a need to elucidate the spatial correlation between microbial communities and geochemical parameters present at impacted sites. Understanding the effects these relationships might have on LNAPL depletion rates will contribute to the design of more efficient remediation technologies.

3.0 Biogeochemical Characterization of a LNAPL Body at an Impacted Site

3.1 Introduction

Subsurface hydrocarbon contamination has resulted from accidental petroleum spills and leaks, which have been a common occurrence in soils and water over the last century (Sale, 2003). An estimated 400,000 to 800,000 metric tons of crude oil seep into the environment on a yearly basis, worldwide (Das & Chandran, 2010). Petroleum releases into soil environments can result in the formation of LNAPL source zones, which originate vapor and liquid plumes containing contaminants including benzene, toluene, ethylbenzene and xylenes (BTEX) (Newman et al., 1991; Landmeyer et al., 1998). Benzene is a known carcinogen, and thus, a regulatory driver for remediation technologies (Wai, 1995; Walden & Spence, 1997; Chen & Taylor, 1997). Although the processes governing LNAPL depletion are, as yet, not well understood, LNAPLs have been observed to persist at contaminated sites for years to decades (Huntley & Beckett, 2002). Thus, source zone management is a key aspect of eliminating groundwater, soil and gas threats posed by hydrocarbon contamination (Chadalavada et al., 2012).

Monitored natural attenuation (MNA) and enhanced natural attenuation have become popular management strategies for hydrocarbon-impacted sites given that these strategies are relatively inexpensive and involve minimally invasive technologies (Jørgensen et al., 2010) , (Declercq et al., 2012; Sihota et al., 2010). Biodegradation is one of the main processes driving natural LNAPL depletion and occurs both under aerobic (McNally et al., 1998) and anaerobic conditions (Boopathy et al., 2012; Weelink et al., 2010). However, persistent LNAPL source

zones are often anaerobic (Coates et al., 1997), and the fundamental microbiological processes responsible for the degradation of recalcitrant contaminants (e.g. benzene) at field sites are still not well- understood. Thus, there is a need to elucidate how site geochemistry and indigenous microbiology affect degradation rates in LNAPL zones in order to drive development and implementation of sound remedies (Illman & Alvarez, 2009).

Recent research has indicated an increased focus on resolving anaerobic biological depletion reactions in hydrocarbon-contaminated regions. To this end, some studies have used cultured-based and microcosm-based approaches to investigate the relationships existing between key microbial functions, microbial ecology, electron donors (i.e., contaminants) present, electron acceptors present, and other physical and chemical factors (Simarro et al., 2013), (Wu et al., 2008), (Viñas et al., 2005; Morris et al., 2012). Although controlled laboratory studies represent a powerful approach for conducting hypothesis-driven research, the findings of such studies can have limited relevance to field sites because the laboratory conditions may select for different microbial communities than those encountered under field conditions (i.e., often times the enriched microbial communities differ markedly from the communities present at contaminated sites) (Simarro et al., 2013). In contrast, studies that focus on analysis of *in situ* microbial communities at hydrocarbon-impacted sites have the potential to lead to more field-relevant findings; however, to date the number of such studies remains limited (Sutton et al., 2013; Acosta-González et al., 2013). Furthermore, the widespread applicability of studies conducted to date remains unknown because indigenous microorganisms vary among impacted sites due to climate, hydrogeology, contaminant type and composition. Additionally, available studies have reported microbial community composition at the phylum level at best (Sutton et

al., 2013; Acosta-González et al., 2013), which has limited utility for predicting subsurface microbial activities. Thus, studies that characterize geochemical conditions and microbial communities at the genus level are needed at a broader range of field sites. Such studies will contribute to our fundamental understanding and drive technology development and remedy decisions.

The study presented herein was performed at a decommissioned refinery in Wyoming. While active (1923-1982), the facility processed local petroleum crudes into fuels including diesel and gasoline. Based on a study performed by the current site consultant (Trihydro, 2002), we hypothesized the existence of three distinct biogeochemical zones: 1) an aerobic low-contaminant mass zone, 2) an anaerobic high-contaminant mass zone, and 3) an anaerobic sulfidogenic low-contaminant mass zone. The overarching goal of this research was to inform the development and implementation of passive remedy strategies at LNAPL-impacted sites. The study objectives were: 1) to gain understanding regarding how site geochemistry, including contaminant distribution, influences microbial community structures in the subsurface, and 2) to identify indigenous microorganisms present at the impacted site with the potential of degrading hydrocarbons in LNAPL zones. A depth-resolved characterization of the geochemical parameters and the microbial communities present at the site was performed.

3.2 Materials and Methods

3.2.1 Site Description

This study was performed at the site of a former refinery in Wyoming (Figure 3.1 A). Operations at the refinery ceased in 1982. Former operating units were located in an area of

the property underlain by the South Property Aquifer. The site is adjacent to the North Platte River and underlain by fluvial sands. The water table fluctuates seasonally, between 8.2-9.6 feet below the ground surface (bgs). Besides refining petroleum liquids into diesel and gasoline, other goods like asphalt and coke were also produced. The contaminated area was disconnected from the river by a Waterloo™ sheet pile wall that extends from ground surface to bedrock. A former site assessment revealed that the smear zone extends from approximately 2 ft to 13 ft below the ground surface (bgs) (Trihydro, 2002). A sulfate vertical gradient within the saturated zone also was measured at the site, and sulfate concentrations were found to increase with depth (Trihydro, 2002). Sulfate concentrations appeared to be controlled by the solubility of gypsum.

3.2.2 Soil Sampling

Seventeen soil cores were collected at the site during the installation of multilevel sampling systems (MLS), which were used to monitor chemical constituents in the aqueous and gas phases (see section 3.2.3). The well installation pattern is presented in Figure 1. Soil cores were collected by direct push drilling using a Geoprobe® rig and stored in acetate sleeves (AMS Power Probes™, American Falls, ID). Immediately after collection, soil cores were cut into 2.5-ft sections and placed on dry ice to preserve geochemical conditions during transportation to the laboratory. Once received at the laboratory, cores were stored at -20℃ until sampling, which was completed in less than three days. Subsamples of each frozen core were taken at 6-in intervals using a circular saw, and samples were analyzed for soil types, hydrocarbons, and microbial ecology. Soil contaminant concentration and water and gas analyses were performed

at all sample points within the study area, while the microbial analysis was performed only on core C3 (Figure 3.1 B).

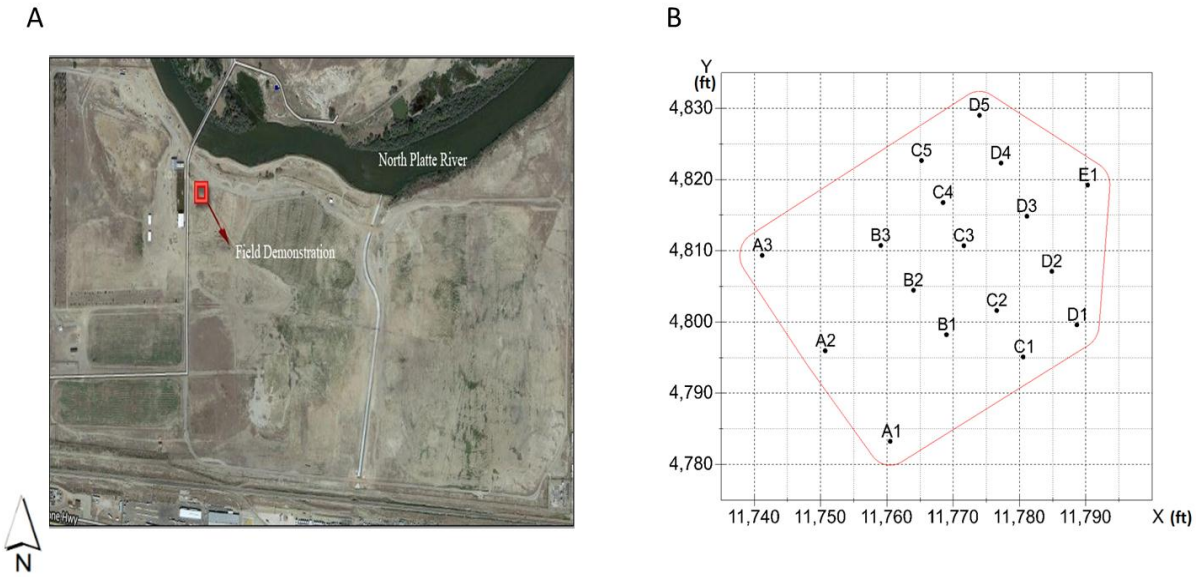


Figure 3.1: Field site images: A) Aerial view of site B) Plan view schematic of field demonstration area. Figure displays soil-core collection and MLS locations.

3.2.3 MLS Description and Gas and Water Sampling

The MLS were installed in the borings created during soil core collection. The MLS consisted of six sections of 1/8-inch O.D. Teflon tubing bundled around a section of PVC pipe (0.5-inch ID). Each MLS had a total of six sampling ports that were located at 2-ft depth intervals with three ports located in the vadose zone and three ports located below the water table (Figure 3.2-A). Sample ports consisted of Nitex™ (HD3-10, Tetko, Inc., Elmsford, N.Y.) cloth wrapped around each piece of Teflon tubing.

For water sampling, a peristaltic pump (Cole-Parmer, Chicago, IL) was connected to each of the MLS ports that reach below the water table via Master Flex ^{TM} tubing (Cole-Parmer), and water was pumped to a Multi-Probe Flow Monitoring System^{TM} (Geotech, Denver, CO). A pH probe (Symphony^{TM}, VWR, Radnor, PA), an ORP probe (model: NCL-100, ORION^{TM}, Thermoscientific, Waktham, MA) and an ORION^{TM} Five Star Plus meter (Thermoscientific, Waltham, MA) were utilized to collect pH and ORP readings in the field.

Water samples for total petroleum hydrocarbon (TPH) analysis were collected in 10-ml glass vials (C4020-10, VWR) sealed with Teflon-lined septa and aluminum crimp-caps (Figure 3.2-C). Water samples (10 ml) for cation and anion analysis were filtered through 0.45-µm Acrodisc^{TM} syringe filters (PSF, Life Sciences Advanced Technologies, St. Petersburg, FL) and collected in 10-ml vials. For cation analysis samples were preserved with 50 µl of a 70% HNO_3 solution and sealed with teflon lined septa and aluminum crimp caps. To preserve the anion samples, care was taken to prevent water sample exposure to oxygen. To achieve this, a needle (connected to the acrodisc filter) was used to inject the water through chlorobutyl-septum stoppers into pre-prepared, anaerobic 10-ml serum vials. The serum vials used for anion sample collection were pre-prepared inside an anaerobic glove box (95% N_2 and 5% H_2 atmosphere); the vials were sealed in the anaerobic chamber and a syringe was used to draw a vacuum through the septa, so samples could be injected through the septa without exposing them to air. All aqueous samples were immediately placed on ice and transported to the laboratory for further analysis.

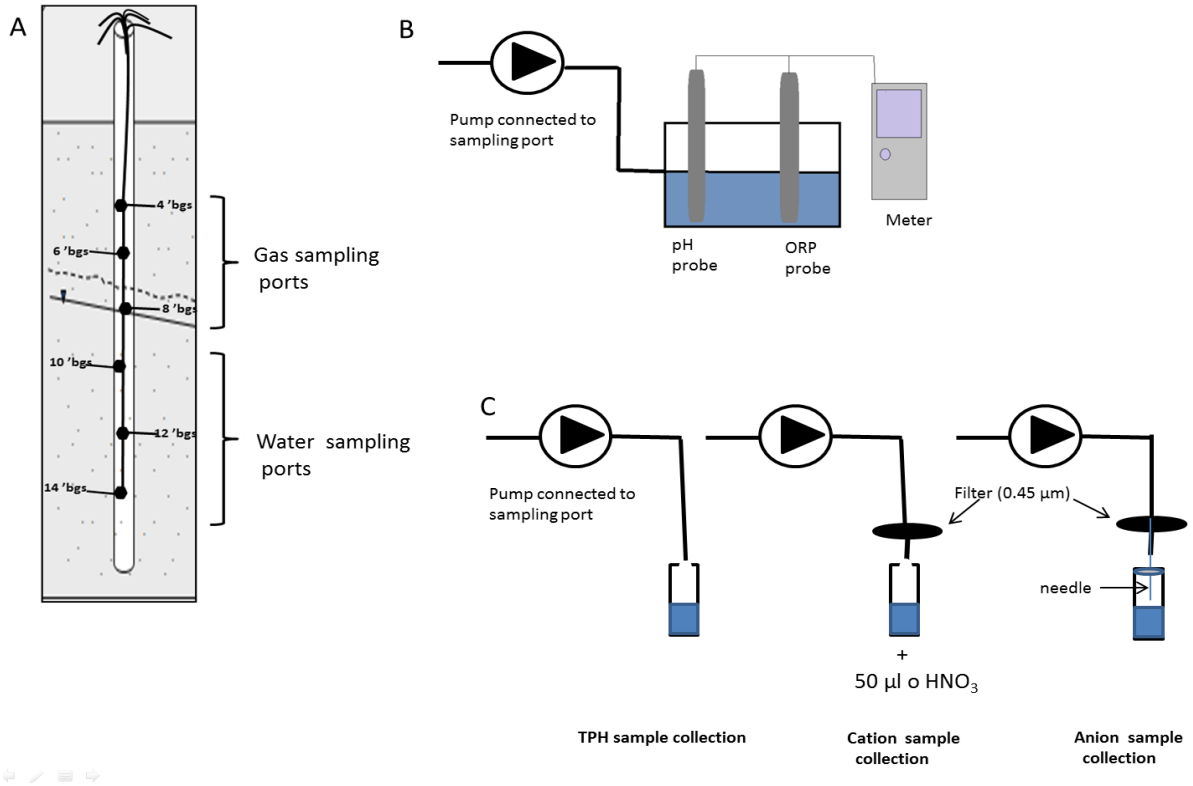


Figure 3-2: MLS and sampling procedure schematics. A) Schematic of a MLS. B) Schematic of pH and ORP data collection system. C) Schematic of water sample collection for TPH, anion and cation analysis.

Gas samples were collected and analyzed on site for carbon dioxide, methane and oxygen (CO_2, CH_4, and O_2) using an Eagle-2^{TM}, portable multi- gas analyzer (RKI, Union City, CA). The outlets of the sampling ports with inlets located in the vadose zone were connected to the instrument's measuring probe to collect the readings. A carbon filter (Landtec, San Bernardino, CA) was utilized in the measurements to avoid oversaturating the photoionization detector.

3.2.4 Chemical Analyses

3.2.4.1 Hydrocarbon Analysis in Soil Samples

Core subsamples (~10 g-30 g) were placed in 125-mL air-tight vials containing 50 ml of high-purity (ACS/HPLC certified) methanol (Honeywell Burdick & Jackson, Muskegon, MI) for hydrocarbon extraction (Figure 3.3). The samples were stored in the dark at 4˚C until analysis via gas chromatography (GC). Prior to analysis the samples were shaken vigorously on an SMI multitube vortex, (SMI, Midland, ON, Canada) for one hour. For analysis, 1 µl of the extract was injected into a Hewlett Packard Model 5890 Series II GC equipped with an automatic sample injector (Hewlett Packard Model 7673), a Restek (Bellefonte, PA) RTX-5^{TM} column (30 m length x 0.32 mm I.D. x 0.25 µm film thickness), and a flame ionization detector (FID). The GC temperature program was as follows: 45˚C for 3 minutes, 12˚C/min to 120 ˚C, 20˚C/ min to 300 ˚C, and held at 300 ˚C for 3 min. The injection port and detector temperatures were 250˚C and 300˚C, respectively. The supply rate for the carrier gas (helium) was 3 ml/min. All compounds were reported in mg/Kg of soil. Hydrocarbon quantities were reported for total petroleum hydrocarbons (TPH), diesel range organics (DRO), gasoline range organics (GRO), and BTEX compounds. Commercially available calibration standards for GRO, DRO and the individual BTEX components were utilized. A 9-component GRO EPA/Wisconsin mix (Restek, Bellefonte, PA) was used for GRO compounds; a 10-component DRO EPA/Wisconsin mix (Restek, Bellefonte, PA) was used for the DRO components. All calibration curves were characterized by an R^2>0.99 (n≥4). Two calibration standards were measured with each GC run to verify accuracy.

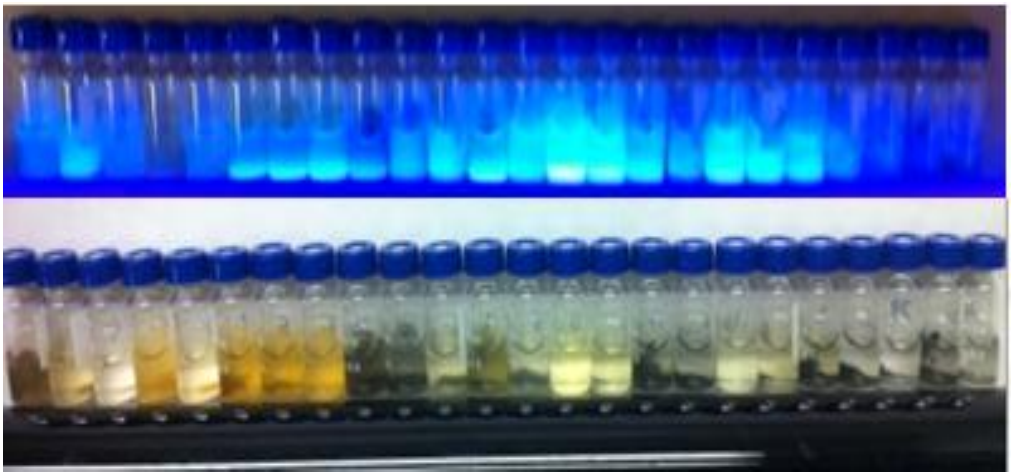


Figure 3.3: Core C3 methanol extracts in GC vials. Vials are lined up according to sample's depth, observed under UV light (top) and white light (bottom).

3.2.4.2 Hydrocarbon Analysis in Aqueous Samples

Aqueous samples were extracted with ACS/HPLC grade *n*-hexane (Sigma-Aldrich). For extraction, 4 ml of each sample were placed in 4-ml glass vials containing 400 µl of *n*-hexane. After the samples were shaken for 30 minutes on a vortex shaker (as described in section 3.2.4.1), the hexane and water phases were allowed to separate for 15 minutes. Next, 300 µl of the hexane phase were pipetted into a GC vial and stored at -20 ˚C until GC analysis. Aqueous-phase hydrocarbon analysis was conducted as described for soils, with the exception that the sample inlet gas flow was split by a 12:1 ratio, and the initial temperature ramp was decreased from 12 ˚C/min to 10 ˚C/min.

3.2.4.3 Analysis of Anions and Cations in Aqueous Samples

Samples for cation and anion analysis in water were submitted to the Soil Testing Lab at Colorado State University (CSU) (Fort Collins, CO) for analysis. Samples were analyzed by high

resolution inductive coupled plasma atomic emission spectroscopy (ICP), IRIS Advantage Dual View (TJA-Solutions, Winsford, England). 10 ml of each sample were analyzed for the following cations: K, Ca, Fe, Mg, and Mn. Ten ml of each sample were analyzed for the following anions: SO_4^{2-}, Cl^-, NO_3^-, CO_3^{2-} and HCO_3^-.

3.2.5 Microbial Ecology Characterization

3.2.5.1 Sample Pretreatment

The core subsamples collected for microbial characterization were stored at -20 ˚C until they were processed for DNA extraction. In preparation for DNA extraction, the samples were pretreated to remove hydrocarbons and other compounds such as humic substances (Figure 3.4) that were shown to affect the yield reproducibility of the DNA extraction procedure (see Appendix A-1); a washing pretreatment step was adapted from a previously published method (Whitby & Lund, 2009). In detail, 5 g of soil were placed in 15-ml centrifuge tubes. Next, 80 ng of dehydrated skimmed milk (VWR) and 10 µg of polydeoxinocinic-deoxycytidilic-acid (pdIdC) (Sigma-Aldrich, St. Louis, MO) were added to each sample, and the mixtures were vortexed with a Gennie-II vortex (Mo Bio, Carlsbad, CA) for one minute. The samples were then washed three times. For the first wash step, 10 ml of DNA-free, sterile, DI water was added to the mixture followed by the addition of: 500 µl of 50 mM tris-HCl (pH=8.3)(Sigma-Aldrich), 400 µl of 200 mM NaCl (VWR), 100 µl of 5 mM Na_2EDTA (Sigma-Aldrich) and 5 µl of Triton X-100 (5% V/V)(Sigma-Aldrich). The sample solutions were vortexed vigorously for 3 minutes and centrifuged at 13,000 rpm in a Sorval Legend XTR ^{TM} centrifuge (Thermoscientific, Ashville, NC)

for 5 minutes to pellet soils and biomass, and the supernatant was discarded. For the second wash step, 10 ml of DNA-free water were added followed by the addition of: 500 µl of 50 mM tris-HCl (pH=8.3), 400 µl of 200 mM NaCl, and 100 µl of 5 mM Na_2EDTA. The sample solution again was vortexed and centrifuged, and the supernatant was discarded. A final washing solution containing 10 ml of DNA-free water, 500 µl of 50 mM tris-HCl (pH=8.3) and 100 µl of 5 mM Na_2EDTA was added to the sample solution prior to vortexing, centrifuging the sample, and discarding the supernatant for a final time.

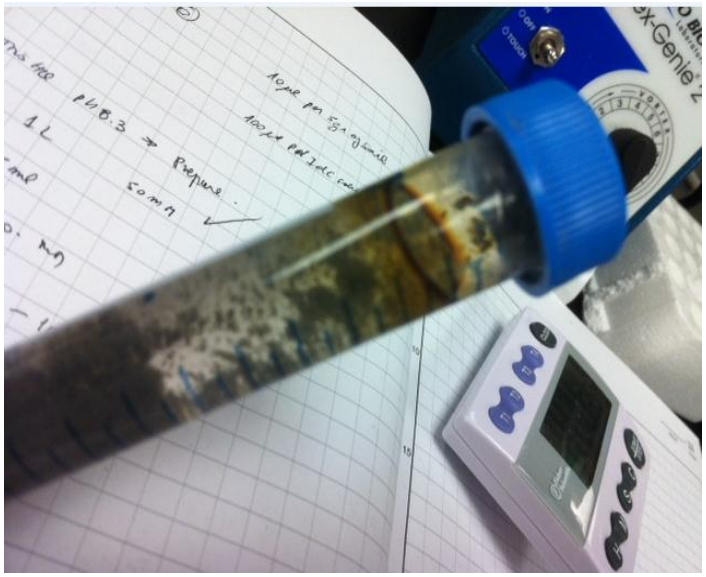


Figure 3.4: Image of the soil sample after initial pretreatment step.

3.2.5.2 DNA Extraction

DNA was extracted from the pretreated soil samples with the Powerlyser™ Powersoil® DNA Isolation Kit (MoBio, Carlbstad, CA) according to the manufacturer's instructions with modifications to maximize DNA yield. Approximately 0.5 g of soil were used for each extraction,

instead of 0.25 g as recommended by the manufacturer. Additionally, duplicate DNA extractions for each soil samples were pooled and processed with a single Powersoil® spin filter. Finally, the samples were eluted with 50 to 60 µl of elution buffer instead of 100 µl. DNA concentrations were quantified via optical density at 260 nm with a Gen5™ Biotek microplate reader using a Take 3™ microplate (Biotek, Winoosky, VT). DNA was extracted in triplicate from each core subsamples. DNA was stored at -20 °C prior to quantitative PCR (qPCR) and pyrosequencing analysis.

3.2.5.3 qPCR Assays

SYBR green™ or Taqman™ (Life technologies, Grand Island, NY) assays were used to quantify targeted genes. Genomic DNA was used as a standard to generate calibration curves for every target assayed. Table 3.1 displays the type of assay used, primer and probe names and sequences, target genes, and the names of the microorganisms used as controls. All assays were performed using an ABI 7300 real-time PCR system (Applied Biosystems, Foster City, CA).

Each 25-µl SYBR green ™ qPCR reaction contained the following: Power SYBR green™ (final concentration= 1X)(Life technologies, Grand Island, NY), forward and reverse primers (final concentration=2.5 µM), magnesium acetate (final concentration=10 µM), PCR grade water and 1 ng of DNA template. Thermocycling conditions were as follows: 95˚C for 10 min, followed by 40 cycles of 95˚C for 45 s, 56˚C for 30 s, and 60˚C for 30 s; fluorescence data was collected at the end of the elongation phase for every cycle. Dissociation curve analysis was conducted to confirm amplicon specificity.

Each 25-µl Taqman™ reaction contained the following: Taqman Universal PCR master mix™ (final concentration: 1X) (Life technologies, Long Island NY), forward and reverse primers and probe (final concentration: 5 µM primer, and 2.5 µM probe), and 1ng of DNA template and PCR grade water. PCR conditions were as follows: initial activation at 50˚C for 2 minutes, denaturation at 95˚C for 10 minutes, followed by 40 cycles of 95˚C for 15 seconds and 60˚C for 1 minute. Fluorescence data was collected at the end of the elongation phase for every cycle.

Table 3.1: Sequences, assay descriptions and strains used for calibration of qPCR assays.

	Primer name	5'-3' primer sequence	Targeted gene	Assay type	Citation	Calibration strains
Forward primer	27f	AGAGTTTGATCCTGGCTCAG	Eubacteria 16s rRNA gene	SYBR-green	(Ulrich & Edwards, 2003)	*Thauera aromatica* (ATCC # 700265D)
Reverse primer	1492r	TTCCGGTTGATCCYGCCGGA				
Probe	NA					
Forward primer	1Af	TCYGKTTGATCCYGSCRGAFG	Archaeal 16s rRNA gene	SYBR-green	(Ulrich & Edwards, 2003)	*Methanosarcina acetivorans* (ATCC # 35395)
Reverse primer	1100Ar	TGGGTCTCGCTCGTTG				
Probe	NA					
Forward primer	MLASf	GGTGGTGTMGGDTTCACMCARTA	Methyl coenzyme M-reductase α-subunit gene	SYBR-green	(Steinberg & Regan, 2009)	*Methanosarcina acetivorans* (ATCC # 35395)
Reverse primer	MCRAr	CGTTCATBCGTAGTTVGGRTAGT				
Probe	NA					
Forward primer	Dsr1f	ASCACTGGAAGCACGGCGG	α-subunit of DSR gene	SYBR-green	(Bourne et al., 2010)	*Desulfovibrio desulfuricans* (ATCC # 27774D-5)
Reverse primer	Dsr1r	CGTTCATBCGTAGTTVGGRTAGT				
Probe	Dsrap	(HEX)-CCGATAACRCYGCCGCCGTAACCGA-(TAMRA)				

3.2.5.4 454 Pyrosequencing

DNA samples from thirteen depth points within core C3 were selected to compare microbial populations in zones of varying geochemical and hydrocarbon composition. The microbial community composition of each sample was determined by 454 pyrosequencing of both eubacterial and archaeal 16S rRNA genes. Sample triplicates were pooled and submitted to Research and Testing Laboratory, LLC (Lubbock, TX) for analysis. Primers 939f and 1492r were used for the eubacterial 16s rRNA gene-targeted assay, and primers 341f and 958r were used for the archaeal assay. The sequencing platform used was the Genome Sequencer FLX plus 454^{TM} Pyrosequencer (Roche, Indianapolis, IN). Further details on the 454 pyrosequencing data analysis protocol was provided by Research And Testing Laboratory (Lubbock, TX) and is presented as Appendix E.

3.2.6 Data Analysis

3.2.6.1 Geostatistics

Hydrogeochemical data was represented utilizing MVS^{TM} (Mining Visualization System) software (CTech, Henderson, NV). Depth-discrete locations of measured soil and aqueous-phase concentrations of hydrocarbons are represented as the nodes along the vertical lines in the figures. Predicted constituent distributions were generated by kriging of the measured concentration data. Indicator kriging was used to create the geology subsurface model.

3.2.6.2 454 Pyrosequencing data analysis

Genus-level relative abundance (%) data were obtained from Research and Testing. Details on bioinformatics performed by the testing facility on each of the analysed samples can be found in Appendix B. The Shannon diversity index (H') was calculated for each sample as: $H' = -\sum p_i x Ln(p_i)$, where p_i corresponds to the fraction of the total sample that was represented by the genus *i*. Further, to represent this data visually, in general, when the relative abundance of a given genus was greater than or equal to 5%, the genus was assigned a color in pie charts. To highlight groups of microorganisms that have been shown to share functional capabilities, four grouped categories also were reported: putative nitrate reducers, potential iron reducers, potential sulfate reducers and potential methane oxidizers. Genera that individually represented less than 5% of the microbial communities were combined and reported as "other".

The following genera were included in the potential nitrate reducers group: *Nitrospina* spp., *Rhodanobacter* spp., *Acidovorax* spp., *Nitrospira* spp. *Thiolkalivibrio* spp., *Brevibacillus* spp., *Shingobacterium* spp., *Burkholderia* spp., *Parachlamydia* spp. *Caldisericum* spp., *Nitrosospira* spp. *Spiroplasma* spp., *Haemophilus* spp., *Azoarcus* spp.

The following genera were included in the potential iron reducers group*: Geobacter* spp., *Rhodoferax s*pp., *Desulfitobacterium* spp., *Desulfuromonas* spp., *Geothrix* spp., *Desulfuromusa* spp., *Thermolithobacter* spp., *Hydrogenophilus* spp.

The following genera were included in the potential sulfate reducers group: *Desulfovibrio spp., Desulfotomaculum* spp*., Desulfobacterium spp., Desulfonema spp.* (Hao,

1996), *Desulfobulbus spp., Desulfobotulus spp., Desulfomonile spp., Desulfothermus spp., Desulfoporosinus spp., Thermodesulfobacterium spp.* (Hurst C.J., 2007), *Desulfosarcina spp., Desulfatibacillum spp., Thermodesulfovibrio spp., Syntrophobacter spp., Desulfonatronum spp., Desulfoglaeba spp,, Desulfobacca spp., Desulfomicrobium spp., Desulfotalea spp., Desulfococcus spp.* (Stams, 2008), *Desulfatirhabdium spp.* (M. Balk, 2008)*, Thermodesulforhabdus spp,, Desulforhabdus spp. (Beeder et al., 1995), Desulfonauticus spp.* (S. Mayilrajj*,* 2009)*, Desulfurispora spp.* (Kaksonen A.H., 2007)*, Desulfotignum spp.* (Torsvik, 2007), *Desulfocella spp. (Garrity G.M. , 2005) Desulfacinum spp.* (Kuever, 2000) *, Desulfofustis spp.* (Friederich, 1996) *Desulforegula spp.* (Rees & Patel, 2001)*, Desulforhopalus spp.* (Isaksen & Teske, 1996) and *Pelotomaculum spp.* All of the bacteria included in this group have the enzymatic capability to reduce sulfate, but most of them also are capable of growing via fermentation. *Pelotomaculum spp.* have been shown to carry genes that code for two key enzymes involved in the sulfate reduction pathway (adenylyl sulfate reductase and dissimilatory sulfite reductase) (Plugge C.M., 2011) so this genus was included in the potential sulfate reducers group.

The following genera were included in the methane oxidizers group: *Methylophaga spp.* (J. Auclair, 2010), *Methylobacter spp., Methylosinus spp.* (T. Ren, 2000), and *Methylacidiphilum spp.* (Khadem A.F, 2012), *Methylocapsa spp.* (Ricke et al., 2005)*, Methylohalobius spp., Methylocysti spp., Methylococcus spp., Methylohalomonas spp., Methylocella spp., Methylophilus spp., Methylobacterium spp, Methylovorus spp., Methylocaldum spp., Methylothermus spp., Methylosoma spp.*

3.3 Results

This study presents a detailed biogeochemical survey of an LNAPL body (Figure 3.5). A comprehensive analysis of the TPH distribution in the subsurface revealed that, at this site, the contamination is distributed above and below the water table, which is located approximately between 8.5 and 9.5 ft below ground surface (bgs) (elevation 5,076 to 5,078 ft). The average TPH concentration within the LNAPL body (approximately between 0.5 ft bgs -12 ft bgs) was 8,100 with a standard deviation of 6,300 mg/kg. The highest TPH concentration (36,810 mg/kg) was recovered in a soil sample collected from core C1 at 9.5 ft bgs (elev. 5,076 ft)(Figure B.2). The second highest TPH value (31,000 mg/kg) was recorded in a soil sample collected from core A2 at 9 ft bgs (elev. 5076.7 ft). The smallest maximum TPH concentration within a single core (5810 mg/kg) was collected along core D5 at 5.5 ft bgs (elev. 5,079 ft)(Figure 3.1). The average of the maximum TPH concentrations within each core was 17,800 mg/Kg (±8,200 mg/kg). The average elevation value where each maximum concentration was recorded was approximately 5,078 ft (±1.6 ft), which corresponds to approximately 7.5 ft bgs. For most cores, the highest TPH values were recorded above the water table.

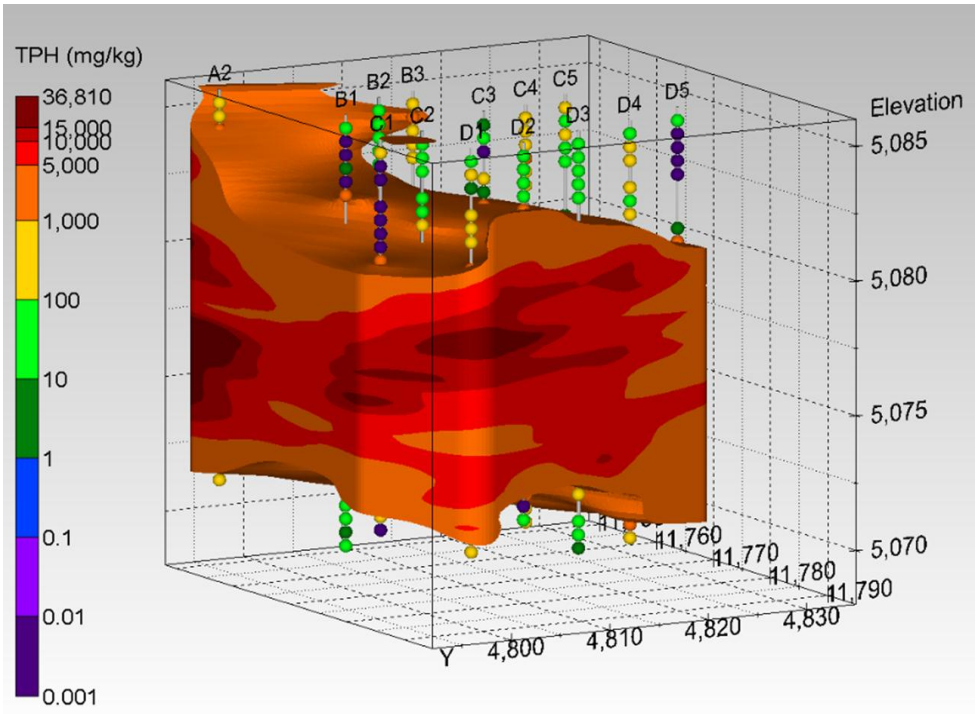


Figure 3.5: TPH distribution in the surveyed area. The nodes along the vertical core lines indicate the sample collection points, and the colors of these nodes indicate measured concentrations. Concentrations above 1,000 mg/kg were considered part of the LNAPL and the spatial distribution of the predicted continuous LNAPL region was determined via Kriging. Units along the x, y and z axes are ft.

3.3.1 Geochemical Characterization along the Transect C.

Transect C was selected as a representative transect due to its central location and general geochemical characteristics that are similar to those measured in the rest of the surveyed region (Figure 3.6). For example, the highest TPH value recorded for C3 was 18,095 mg/kg (Figure 3.6-A), which is within 2% of the average of the maximum TPH concentrations reported for all cores. Thus, detailed geochemical analysis is presented for transect C.

Sands and silts are the main type of soils present in the surveyed region (Figure 3.6-A). The top part of the vadose zone, approximately 0 to 3.5 ft bgs (elev 5,086.0 to 5,082.5 ft), is

composed mainly of silt and silty sand, but soils were found to be heterogeneous across the C transect. For example, in the top part of the vadose zone, core C2 was composed mainly of silty-sand, and core C3 was composed mainly of silt. Core C5 contained fine sand in this region as well. Soils in the lower portion of the vadose zone (between 3.5 and 9.5 ft bgs) were composed of medium, fine and silty sands, with the exception of two samples along C1 and C2 at around 4.5 ft bgs (5,081.5 elev. ft) that were composed of coarse sand. Below the water table, the soils were found to be mainly composed of coarse and medium sands along the center transect.

The majority of the contaminant mass present in soil, at this site, was DRO (80%<) (Figure 3.6-C). The highest DRO concentrations were found approximately between 4 and 11.5 ft bgs (elev. 5,081.5 to 5,074.5 ft). There was little or no DRO found closer to the surface, between approximately 0 and 3 ft bgs (elev. 5,086-5,083 ft). Between these depths, DRO concentrations were below 200 mg/kg with the exception of two samples collected along C1 that did not exceed concentrations of 500 mg/kg.

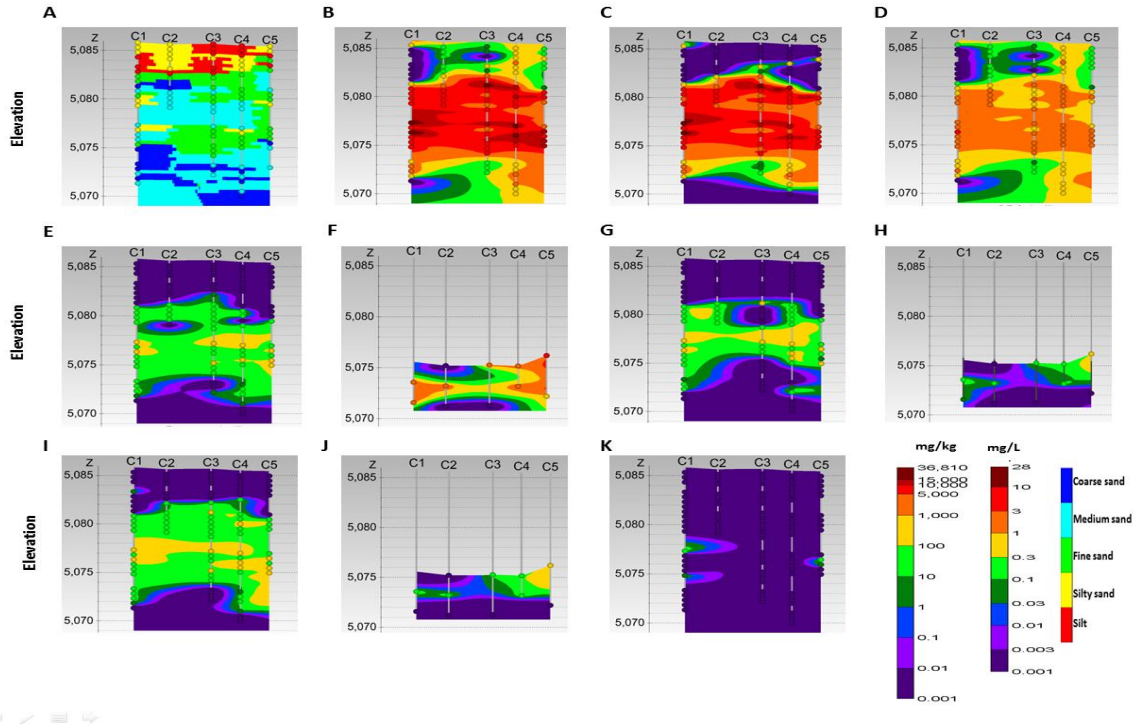


Figure 3.6: Contaminant depth distribution and soil type along the transect C. Soil contaminant concentrations are presented in mg/kg of wet soil. Aqueous concentrations are presented in mg/ml. Elevation is in ft. A) Soil type of the surveyed region. B) TPH concentrations in soil. C) DRO distribution in soil. D) GRO distribution in soil. E) Benzene distribution in soil. F) Benzene aqueous concentrations. G) Ethylbenzene soil distribution. H) Ethylbenzene aqueous concentrations. I) *m & p*-xylene concentrations in soils. J) *m & p*-xylene aqueous concentrations. K) *o*-xylene soil concentrations. Water samples could not be recovered from the MLS for C1 at 10.5 ft bgs, for C2 at 14.5 ft bgs, for C3 at 12.7 bgs, for C4 at 14.5 ft bgs, or for C5 at 13 ft bgs. *O*-xylene was not detected in aqueous samples. Toluene was not detected in soil or aqueous samples. Full page figures are available for each contaminant as supplemental information (Appendix B.1-B.11).

The maximum GRO concentration (6,900 mg/kg) was measured in a sample collected along core C1 at 9.5 ft bgs (Figure 3.6-D). Although GRO was detected in significantly smaller concentrations than DRO, the depth distribution of GRO was similar to that observed for DRO. However, in contrast to DRO, GRO was present in some samples collected above 3 ft bgs (e.g., in C4). In general, at similar depths GRO concentrations were higher for more transmisive sediments. This can be observed, for example, in the top of the vadose zone closer to the surface where GRO is depleted or found in very low concentrations in samples that were composed mainly of silt. In general more transmisive aquifer regions, present less amount of contaminants (Figure 3.6-A, B, C and D).

Along transect C, the highest benzene concentration (264 mg/Kg) was detected around 8.5 ft bgs (elev 5,077 ft) at C4 (Figure 3.6-E). Core C3 contained two locations with high soil benzene concentrations, one (189 mg/kg) at 8.5 ft bgs (elev 5,077 ft) and another (107.8 mg/kg) at 10.5 ft bgs (elev 5,075 ft). No benzene, ethylbenzene (Figure 3.6-G) or *o*-xylene (Figure 3.6-K) were detected in soil at depths of less than approximately 3 ft bgs (elev 5,082 ft) along transect C. High benzene concentrations were detected in water samples. The highest aqueous benzene concentration (4.42 mg/ml) measured at the site was detected in MLS B1, 10 ft bgs (elev 5,070 ft) (data not shown). The highest aqueous benzene concentration (3.13 mg/L) along transect C was detected in C5 at 10 ft bgs (elev 5,075 ft) (Figure 3.6-F). The soil depth distribution of ethylbenzene (Figure 3.6-G) was similar to that of benzene, but less contamination was present below 10.5 ft bgs (elev 5,075 ft). Ethylbenzene aqueous concentrations were below 1 mg/ml for all sampled points including those collected along transect C (Figure 3.6-H). O-xylene soil

concentration depth distribution (Figure 3.6-I) also was similar to that observed for benzene, along transect C, but there was contaminant present at shallower depths and more contaminant mass found around 4ft bgs (elev 5,081 ft), when compared to benzene. The aqueous concentration distribution of o-xylene (Figure 3.6-J) was very similar to that observed for ethylbenzene. *O-xylene* was detected only in very low concentrations (<34 mg/L) in few soil samples (Figure 3.6-K), and it was not detected in water samples. No toluene was found in soil or water (302 samples assayed).

3.3.2 Gas Samples in the Vadose Zone along Transect C

Oxygen, carbon dioxide and methane concentrations also were measured along transect C. The highest oxygen concentrations (2.3 to 13.0 % vol/vol) were recorded along C1 (Figure 3.7-A). In C2, oxygen levels were between 0.2 to 3.5 % vol/vol. The highest concentration measured within C2 was at 4.5 ft bgs (elev 5,081ft), and the lowest was at 8.5 ft bgs (elev 5,078 ft). Oxygen levels were low (0.2 to 1.1 % vol/vol) in all samples along C3 and C4. Gas levels in C5 were similar to C3, with the exception of the shallowest port (4 ft bgs or elev 5,082 ft), where oxygen levels were measured to be 1.9% vol/vol. Carbon dioxide and methane gas concentrations (Figure 3.7-B) followed an opposite trend; at locations where high levels of oxygen were observed, carbon dioxide and methane levels were low. Carbon dioxide levels along transect C ranged from 4.7 to 15.4 % vol/vol. The highest carbon dioxide concentration was measured along C5 at 14 ft bgs (elev 5,078.5ft). The lowest carbon dioxide concentration was measured at the shallowest port along C1 at 4.5 ft bgs. Generally methane concentrations were found to increase with depth (Figure 3.7-D). Values ranged from 0.5 to 5% vol/vol in all samples, but no methane was detected in some of the shallowest ports (e.g., C1 port at 4.5 ft bgs).

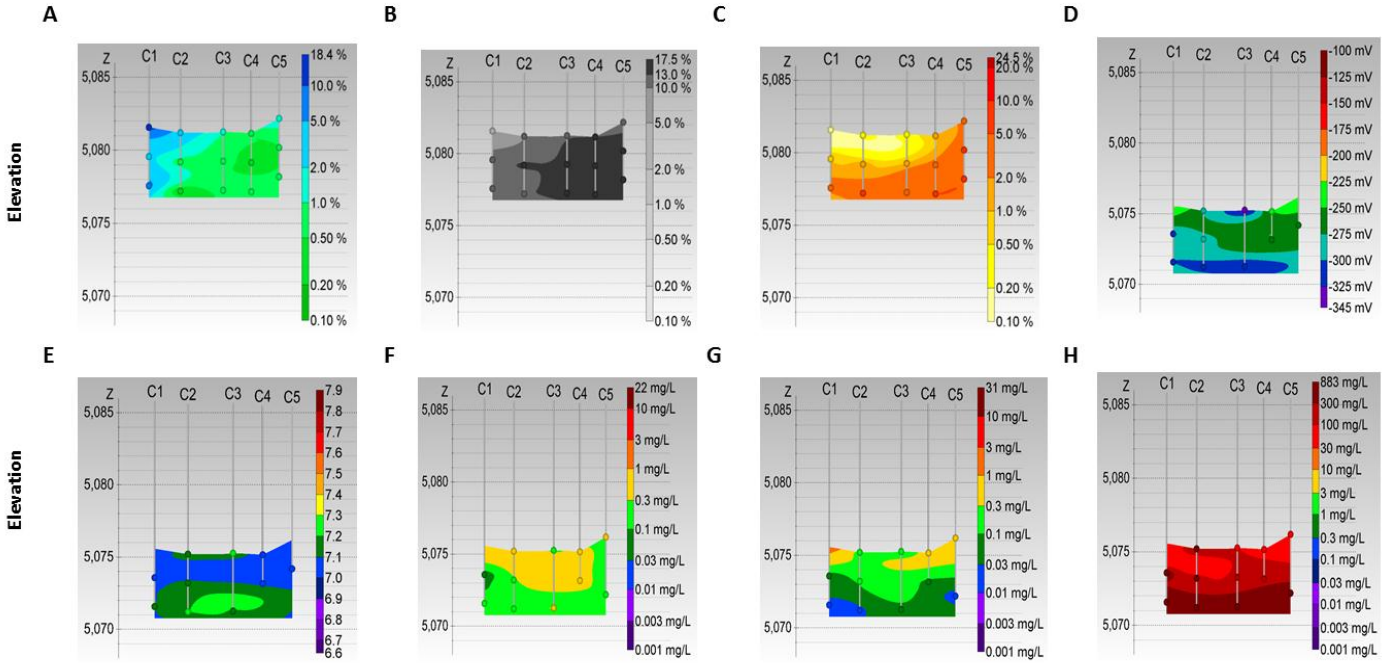


Figure 3.7: Gases in the vadose zone and water inorganics along transect C. Partial pressures were measured in gas sampled from the 3 MLS ports located above the water table. Results are reported in % vol/vol. Elevation is reported in ft. Aqueous concentrations were measured in water sampled from the 3 MLS ports located below the water table. Aqueous concentrations are reported in mg/L. A) Oxygen. B) Carbon dioxide. C) Methane. D) ORP. E) pH. F) Total aqueous nitrate. G) Total aqueous iron. H) Aqueous sulfate. Full page figures are available for each contaminant as supplemental information (appendix C.1-C.8)

3.3.3 Water Inorganic Chemistry

The saturated region of the aquifer was found to be a reducing environment with a neutral pH. ORP values ranged from -225 mV to -325 mV (Ag-AgCl) (Figure 3.7-D), and pH values ranged between 6.95 and 7.30 (Figure 3.7-E). Very low levels of nitrate and iron were detected in the aqueous samples (<1 mg/L for C transect) (Figures 3.7-F and G). In contrast, relatively high levels of sulfate were detected. Along transect C, sulfate values ranged from 33mg/L to 880 mg/L (Figure 3.7-H). The highest sulfate values were recorded in the samples collected at the deeper elevations, around 14.7 ft bgs (elev 5,071 ft).

3.3.4 Depth-Resolved Characterization of Biogeochemical Zones along the Central Core

In contrast to the initially proposed conceptual model, which hypothesized the existence of three biogeochemical zones across the LNAPL body, quantities of eubacterial and archaeal 16S rRNA genes coupled to the depth-resolved geochemical data, revealed the existence of four distinct biogeochemical zones along the center core (C3) of the surveyed region (Figure 3.8).

Zone I, located between 0 and 3 ft bgs (elev 5,081.4 – 5,078.4ft), is entirely above the water table. This zone contains low hydrocarbon mass (0.001 to 100 mg/kg TPH) and is presumably aerobic, given the low oxygen levels recorded immediately below at 4.7 ft bgs (elev 5,081.2 ft). In this region, quantities of eubacterial 16s rRNA genes were on average approximately 2 orders of magnitude larger than quantities of archaeal 16s rRNA genes (Figure 3.8).

Zone II, located between 3 ft bgs (elev 5,082.9 ft) and 6 ft bgs (5,080.4 elev ft), is also entirely above the water table. This is a high contaminant mass zone with TPH values that ranged from 3,000 mg/kg to 18,000 mg/kg. Gas data collected within this zone suggest that methane and oxygen fronts likely converge at the top of this zone, at approximately 3.5 ft bgs (Figure 3.7). Quantities of eubacterial 16s rRNA genes were on average approximately 2 orders of magnitude larger than quantities of archaeal 16s rRNA genes.

Zone III is located between 6.0 and 11.5 ft bgs. The depth to the water table along the central core is approximately 9 ft bgs (5,076.7 elev ft), and thus zone III contains regions that are both above and below the water table. Zone III is a high contaminant mass zone with TPH values that ranged from 1,000 mg/kg to 14,000 mg/kg. High benzene concentrations in soil were found at the samples collected at 8.5 ft bgs (elev 5, 077.2 ft) and at 10.5 ft bgs (elev 5,075.2 ft). Aqueous samples collected within this zone revealed a substantial amount of dissolved benzene (e.g., > 2 mg/L at 10.7 ft bgs [5,075.2 elev ft]). Methane and carbon dioxide were present in the gas collected within this zone (5% vol/vol of methane for both analyzed samples). Low levels of aqueous sulfate measured (95.5 mg/L). Quantities of eubacterial and archaeal 16S rRNA genes are of the same order of magnitude in this zone.

Zone IV, located between 12 ft bgs (elev 5,073.7 ft) and 14 ft bgs (5,072.7 elev ft), is entirely below the water table. Low soil TPH values were detected here; values ranged from 24 mg/kg to 90 mg/kg. Sulfate was detected in substantial concentrations. Only one water sample was collected along C3 within this zone (14.7 ft bgs); this sample contained 748 mg/L of sulfate.

As was observed for Zone III, quantities of eubacterial and archaeal 16S rRNA genes were of the same order of magnitude in this zone.

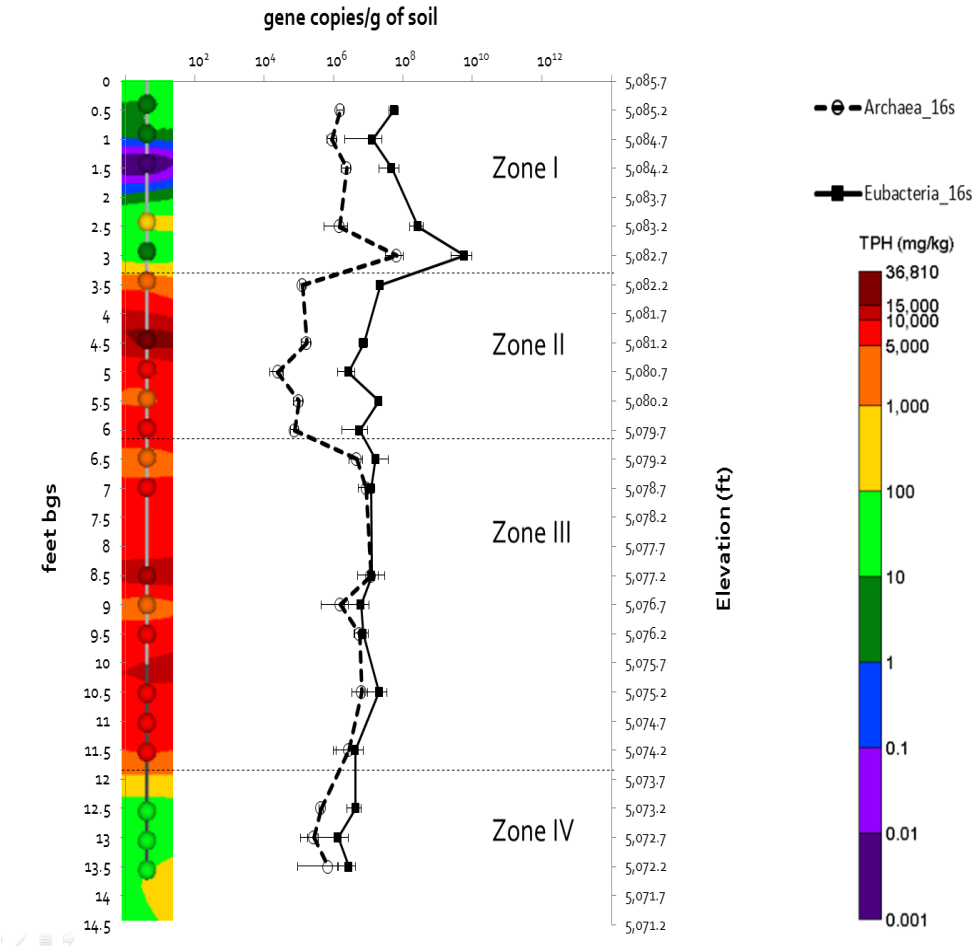


Figure 3.8: Quantities of 16s rRNA genes along the central core. Archaea (dashed line) and Eubacteria (solid line) 16s rRNA genes were quantified along the central core (C3). The error bars correspond to standard deviations for qPCR reactions run on DNA extracted from triplicate soil samples. TPH concentrations for core C3 are included on the left of the graph for easy referencing.

3.3.5 Depth-Resolved Characterization of Microbial Ecology along C3

Results of pyrosequencing analysis revealed that the structure of the bacterial and archaeal communities varied with depth (Figure 3.9). Shannon diversity indexes (H') (Hill, 1973) calculated for each analyzed sample showed that the microbial communities surveyed within zone II were the most diverse, while the microbial communities surveyed within zone III were the least diverse. Community analyses results further support the classification of 4 distinct biogeochemical zones with depth.

As expected, and predicted by the conceptual model, Zone I was mainly populated by aerobic organisms. *Acidobacterium* spp. dominated the bacterial communities sampled at 1 ft bgs and 1.5 ft bgs (34.5 and 63.5 % respectively). The microbial community at 3 ft bgs was primarily composed of members of the genera *Lysobacter* (39.9%), *Hydrogenophilus* (26.4%) and *Thiobacillus* (12.3%). This sample (3 ft bgs) also contained low numbers of putative iron and sulfate reducers (2.7 and 2.7 %, respectively). Collectively, *Nithrososphaera* spp. and *Candidatus Nithrososphaera* spp. represented over 70% of the of Archaea present in the soil samples collected at 1 ft bgs and 1.5 ft bgs. Additionally, members of the methanogenic genera *Methanobrevibacter (30.4%), Methanocella(24.9 %), Methanosaeta (17.6 %), Methanolinea(7.8%)* and *Methermicoccus (7.7%)* were present in the soil sample collected at 3 ft bgs.

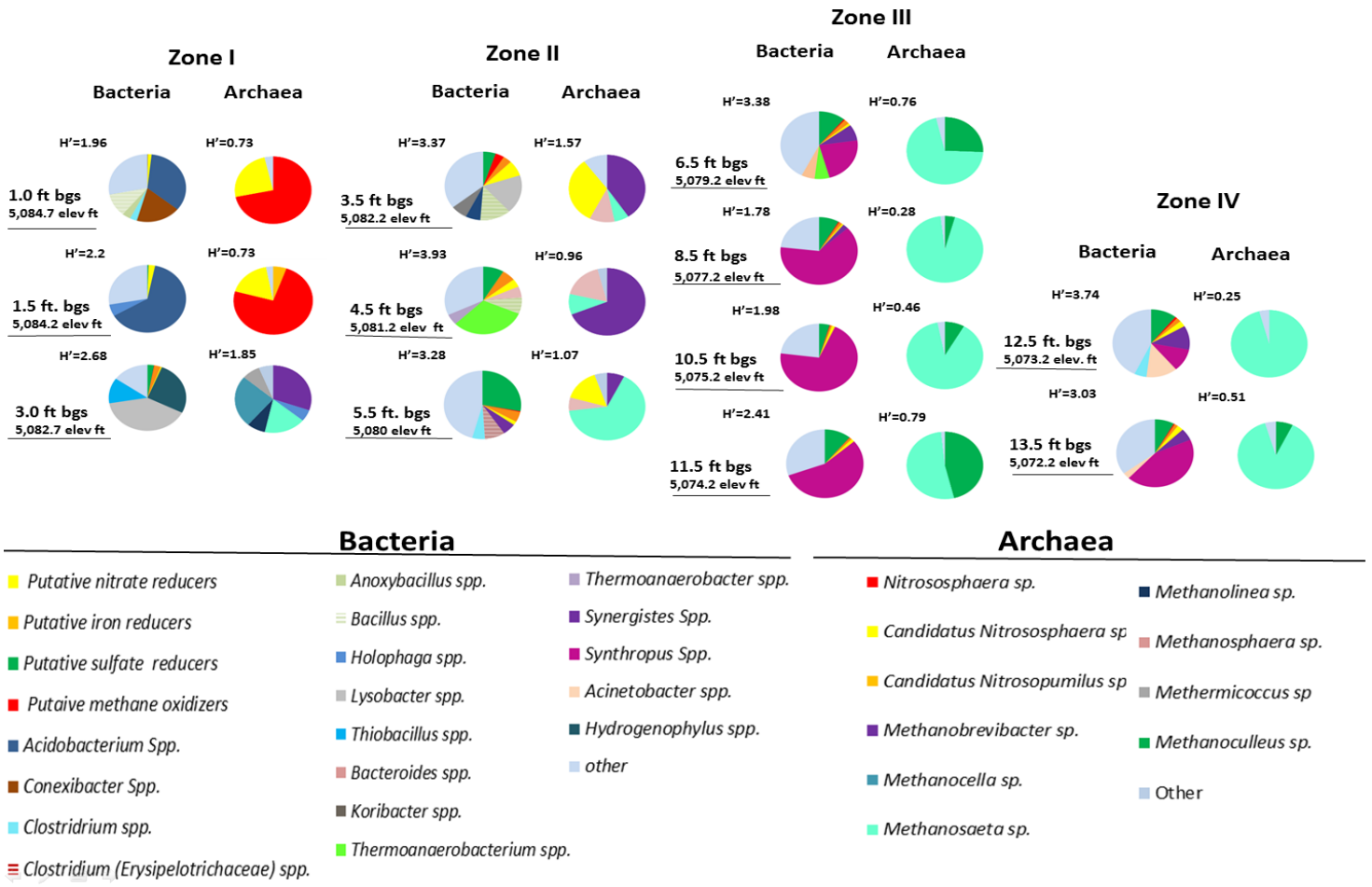


Figure 3.9: Microbial community analysis by pyrosequencing. A table with relative % of each sample's represented OTU's is is available as supplemental information (Appendix D).

Calculated Shannon indexes for Zone II's eubacterial communities ranged from 3.28 to 3.93 for the samples surveyed within the zone. In the sample collected at 3.5 ft bgs, putative iron reducers, nitrate reducers and sulfate reducers were present. This sample also contained the largest relative percent of methane oxidizers (3.7%) when compared to all other samples analyzed. The top of this zone (3.5 ft bgs) was characterized by communities rich in facultative, fermenting organisms such as those belonging to the genera *Lysobacter* and *Bacillus,* which represented 18% and 13%, respectively of the surveyed communities at this depth. Thiosulfate reducers associated with sulfur and iron cycles, such as T*hermoanaerobacterium* spp. (31.4%) and *Thermoanaerobacter* spp.(5.4%) were found at the center of this zone (4.5 ft bgs). The deepest sample analyzed within zone II (5.5 ft bgs) contained high percentages of putative sulfate-reducing organisms (27.9%), most of which (90.0 %) were identified as *Desulfovibrio* spp*., Clostirdium* spp. (5.3%), *Clostridium* (*Erysipelotrichaceae*) spp. (8.4 %) and *Synergistes* spp. (5.65%). Additionally, three different methanogenic genera where identified in Zone II: *Methanosphaera* spp. (0.7% to 17.2%), *Methanobrevibacter (*7.3% to 68.8%), and *Methanosaeta spp. (*6.2% to 65.7%).

Zone III was markedly less diverse than zone II; however, the sample collected at 6.5 ft bgs contained the most diverse community within zone III. This sample's bacterial community contained *Synergistes* spp. (37.0%), *Thermoanaerobacterium* spp (6.3%), *Acinetobacter* spp. (5.58%), and *Syntrophus* spp. (22.9%). A general loss of diversity with depth was observed as indicated by the decreasing Shannon indexes (e.g., 3.88 for 6.5 ft bgs and 1.78 for 10.5 ft bgs). Interestingly, the least diverse communities corresponded to the samples that contained the highest soil benzene concentrations (8.5 ft bgs and 10.5 ft bgs, respectively), and furthermore,

these communities contained the largest percentages of *Synthropus* spp. (63.6% for 8.5 ft bgs and 70.2% for 10.5 ft bgs). Putative sulfate-reducing bacteria also were found in every sampled community within zone III, with the highest percentages detected in the upper and the lower zone boundaries (11.4 % for 6.5 ft bgs and 11.1% for 11.5 ft bgs). Members of the sulfate-reducing group identified in this zone included *Pelotomaculum* spp., which represented 1.3 % and 4.6% of the total bacterial community sampled at 8.5 ft bgs and 11.5 ft bgs, respectively. The archaeal communities identified within zone III mainly were composed of two methanogenic genera: the acetoclastic genus *Methanosaeta* (70.7 % to 93.9 %) and the hydrogenothrophic genus *Methanoculleus* (6.0 % to 46.1 %).

Zone IV displayed generally increased diversity with respect to zone III. Within zone IV, the sample collected at 12.5 ft bgs contained the lowest hydrocarbon concentration (24 mg/kg of DRO and < 1 mg/kg of GRO), and this sample also contained the most diverse bacterial community among the anaerobic zones. Only 10.88% of the community corresponded to *Syntrophus* spp., and 11.49% of the community was *Synergistes. Clostridium spp.* (5.4%) and *Acinetobacter* spp. (13%) also were found in this sample. Putative sulfate reducers, including *Pelotomaculum* (2.34%), represented 10.8 % of the community. The sample collected at the deepest location (13.5 ft bgs) contained 90 mg/kg TPH, and all of it corresponded to DRO. The bacterial community analysis of this sample revealed that 43.4% of the bacterial community corresponded to *Syntrophus* spp., and only 5.74% of the community was *Synergistes* spp. *Pelotomaculum spp.* was present in both bacterial communities sampled within this region. It represented 2.4% of the total bacterial community sampled at 12.5 ft bgs and 1.4% of the total bacterial community sampled at 13.5 ft bgs. The archaeal communities characterized in zone IV

were composed mainly of *Methanosaeta spp. Methanoculleus* spp. (7.2%) were only found in the sample corresponding to 13.5 ft bgs.

3.4 Discussion

Understanding how site conditions influence hydrocarbon degradation processes is crucial when making remedy decisions. The site investigation reported herein provided detailed biogeochemical insights regarding natural processes occurring at various depths relative to an LNAPL body. High-resolution data indicated the existence of four distinct biogeochemical zones within the surveyed LNAPL impacted region. The findings are consistent with other studies that have reported an association between contaminant mass present and microbial community structure (Margesin et al., 2007), (Sutton et al., 2013; Thavamani et al., 2012). The results also suggest that other parameters such as soil type, electron acceptor availability, and location with respect to the water table may have an effect on the quantity and type of microorganisms present, and thus, are likely impacting on-going microbial processes. The effects of these other parameters on microbial community structure have been studied by others with interesting results. Some have reported that geochemical parameters besides hydrocarbon levels (e.g., redox gradients, soilt tye and metals present in soil), influence microbial ecology (Choi & Lee, 2011; Inceoğlu et al., 2013; Sun et al., 2013), while others observed no significant effects of geochemical parameters besides hydrocarbon levels (Sutton et al., 2013).

Zone I was established as a low-contaminant mass aerobic zone. Generally, microbial quantities were higher in this zone than in any other zone; this finding is likely due to Zone I being an aerobic environment. However, soil type could also be affecting the quantity of

microorganisms present. Studies have found that microbial counts are higher in silts, which were the main soil type present in C3 within zone I, than in sands (Kristensen et al., 2010). This phenomenon may be explained by the fact that silt provides higher organic content, and thus more hydrocarbon storage than sand making, more contaminant mass available to microorganisms, thus promoting their growth. By contrast, sand or silty sand may be a more mass-transfer limited type of substrate. A recent study showed that pyrene biodegradation rates are enhanced in silt and clay aggregates when compared to those in sandy sediments (Cui et al., 2011). Evidence suggests that such phenomena could be affecting biodegradation processes at the study site. While DRO contamination was not found within zone I, GRO was present and found to be unevenly distributed. GRO was present in a highly-transmisive (silty sand) region (C4 core) and absent in silty regions (C3 core and lower part of C1 core), possibly suggesting that contaminants previously present in the aerobic silty soils have been degraded. The observed differences in contaminant distribution in silty sand (C2) and silt aquifer regions (C3) also could be due to contaminant transport-related phenomena, as noted by another study (Österreicher-Cunha et al., 2009). Smaller hydrocarbon compounds have higher solubilities; water and the dissolved compounds can enter the pore space (in silty sands and perhaps not in silts), and as water evaporates the contaminant is left behind. This phenomenon could also explain the absence of DRO in this region and the presence of GRO in more transmisive layers. The location and composition of the released LNAPL could also explain the observed distribution.

Zone II was determined to be a high-contaminant-mass zone with the most diverse metabolic potential within the surveyed transect, as evidenced by the high Shannon indexes

calculated for archaeal and eubacterial communities. Zone II appeared to be a transition zone between the aerobic and anaerobic regions of the aquifer. Bacteria previously associated with hydrocarbon degradation were detected in this zone including gram-positive facultative, fermenters such as *Bacilli* (Ghumro, 2012), and members of the genera T*hermoanaerobacter* and *Thermoanaerobacterium.* Members of the latter two genera have been shown to produce acetate via fermentation (Fardeau, 2004) , and members of their associated family *Thermoanaerobacteraceae* have been linked with the production of hydrogen gas and carbon dioxide (Ueno et al., 2001). These findings provide a line of evidence that hydrocarbon degradation is likely taking place within this LNAPL-containing region. Further, detection of high levels of bacterial 16S rRNA genes from genera associated with sulfate reduction, such as *Desulfovibrio spp.*, at the interface between Zone II and Zone III (Fig. 3.9, 5.5 ft bgs) suggest that sulfate reduction may have played a role in hydrocarbon degradation in these zones. Alternatively, species denoted as sulfate-reducing bacteria may be involved in organic-acid fermentation with the generation of hydrogen, acetate and carbon dioxide, as it has been shown that sulfate-reducing bacteria may be more aptly considered facultative fermenters. This metabolic pathway is only possible in synthrophic association with hydrogenothrophic methanogens, which act as an electron sink for the produced H_2 (Plugge C.M., 2011) . Given that this study used DNA-based methods only, it is not possible to determine which metabolic processes are taking place; however, the presence of both hydrogenothrophic and acetoclastic methanogens also support the proposed syntrophic processes. The predominant metabolic pathways for *Desulfovibrio spp*. organisms in hydrocarbon-impacted sites are, as mentioned previously, oxidation of organic compounds via sulfate reduction or via syntrophic interactions

with methanogens by glucose fermentation (Fardeau, 2004). Additionally, zone II was the only zone containing substantial numbers of methane-oxidizing bacteria (Figure 3.9, 3.5 ft bgs) suggesting that methane and oxygen fronts meet at the top of this zone. Methane and oxygen gas data for this zone were consistent with this finding (Fig. 3.7-A and C).

Despite containing similar TPH levels as Zone II, Zone III was otherwise markedly distinct. Relatively low Shannon diversity indexes were calculated for archaeal and eubacterial communities within this zone suggesting that high contaminant levels coupled to highly-reducing conditions (Fig. 3.7-D) have promoted a loss of ecological diversity here. This phenomenon, which has been observed previously (Hemme et al., 2010, Qu et al., 2011), suggests that geochemical conditions in zone III have selected for highly specialized microbial communities. The microorganisms, methane and carbon dioxide levels, and aqueous sulfate levels detected indicate that both sulfate reduction and methanogenesis are likely driving hydrocarbon degradation processes within this zone. The microbial community analyses revealed the presence of phylotypes previously associated with hydrocarbon degradation such as *Syntrophus* spp. and *Pelotomaculum* spp. Interestingly, the highest level of benzene detected in the C3 core soil (Figure 3.8, 8.5 ft bgs) coincided with the highest detected percent of *Syntrophus* spp. (Figure 3.9, 8.5 bgs), which may suggest that *Syntrophus* spp. play a key role in anaerobic benzene degradation at field sites. Members of this genus syntrophically degrade benzoate (Mountfort et al., 1984), which has been postulated to be produced via the first step in anaerobic benzene degradation (Vogt et al., 2011), and other organic acids to acetate and hydrogen. *Syntrophus* spp. also have been associated with anaerobic degradation of petroleum liquids, such as cyclohexane (Siddique et al., 2011), (McInerney et al., 2007). *Pelotomaculum*

spp., which were found both above and below the water table (Figure 3.9, sulfate-reducing bacteria), also have been associated specifically with anaerobic benzene degradation (Kleinsteuber et al., 2008; Kleinsteuber et al., 2012; Vogt et al., 2011). The highest community percentages of *Pelotomaculum* spp. *(4.60%)* were found in the sample collected below the water table (11.5 ft bgs) at the boundary between zones III and IV. *Pelotomaculum* spp. carry genes that code for two key enzymes involved in the sulfate reduction pathway, adenylyl sulfate reductase and dissimilatory sulfite reductase (Plugge C.M., 2011), and thus were grouped with sulfate reducers; however, it has been shown that these microorganisms can only grow syntrophically through fermentation and not by sulfate reduction. Other metabolically diverse sulfate reducers, such as *Desulfosarcina, Desulfotomaculum and Desulfoporosinus,* were found within zone III though, and these genera may be involved in fermentation or sulfate reduction. Collectively *Methanosaeta* spp. (acetoclastic) and *Methanoculleus* spp. (hydrogenotrophs) account for over 90% of the Archaea detected within zone III. These two methanogens have been reported to predominate in environments where syntrophic hydrocarbon degradation occurs (Kleikemper et al., 2005; Liu et al., 2009). In areas where sulfate reduction is a dominant process, hydrogenotrophic organisms like *Methanoculleus spp.* are outcompeted by sulfate-reducing organisms who act as hydrogen scavengers. The fact that over forty percent of the Archaea identified in the deepest sample collected within zone III are members of the genus *Methanoculleus* provide a line of evidence suggesting that methanogenesis is a dominant process in this zone.

Zone IV is an anaerobic low-contaminant mass zone located entirely below the water table. The presence of *Syntrophus* spp., as well as other microorganisms that have been found

in diesel-impacted soils (e.g., *Acinetobacter* spp.)(Tan & Ji, 2010; Galperin & Kaplan, 2011; Harner et al., 2011), and low observed TPH concentrations (<91 mg/Kg) suggest that hydrocarbon degradation is occurring in this zone. Microbial data from this zone also supports the hypothesis that the presence of high levels of TPH selects for high levels of *Syntrophus* spp.; interestingly, within zone IV, the relative percentage of *Syntrophus* spp. was found to be much higher in the sample collected at 13.5 ft bgs (43.4%) where TPH levels were higher (100 mg/Kg) than in the sample collected at 12.5 ft bgs (10.6%) where TPH levels were lower (14 mg/kg). As for zone III, biogeochemical data suggest that both sulfate reduction and methanogenesis are driving hydrocarbon degradation. Sulfate reducers were detected in significant numbers within this zone. Furthermore, the observed sulfate gradient along C3 is consistent with microbially-mediated sulfate depletion. Although no gas samples could be recovered within zone IV (because it is below the water table), methanogenesis is also likely taking place given the percentages of *Methanosaeta spp*. detected within this zone.

3.5 Conclusions

The holistic, interdisciplinary site analysis approach developed during this study contributed to determining the existence of four distinct biogeochemical zones within the surveyed LNAPL body. High levels of bacteria and Archaea were detected in presence of LNAPL, which provides support for microbially-mediated LNAPL losses occurring at the site. Furthermore, sulfate reduction and methanogenesis were determined to be significant hydrocarbon depletion processes in strictly anaerobic zones. Syntrophic pathways are likely occurring as a well. Understanding complex interactions between contaminant distribution in

the subsurface, microbial ecology and other geochemical parameter are will inform the development of more efficient remediation technologies.

4.0 System Performance Monitoring of STELA's Pilot Study

4.1 Introduction

STELA (Sustainable Thermally Enhanced LNAPL Attenuation) is a potential treatment technology designed based on the results of laboratory and field studies performed by the Center for Contaminant Hydrology (CCH-CSU) in collaboration with Chevron, Exxon Mobil and Trihydro. Observations from carbon dioxide flux measurements at field sites, in combination with temperature data, demonstrated that loss rates in LNAPL zones where measured temperatures were ~20 ºC had LNAPL depletion rates that were approximately an order of magnitude larger than those observed when temperatures were below 18ºC (McCoy Thesis, 2012). Furthermore, soils from one of these field sites (former refinery in Wyoming) were collected and utilized in a thermal microcosm study. Results for the thermal microcosm study indicated that methanogenic rates were critically enhanced in those microcosms incubated at temperatures between 22ºC and 30 ºC (Zeman Thesis, 2012) as compared to lower temperatures (9ºC).

Thus, based on these previous findings, a field site pilot study was designed to test the effect of passive thermal management strategies on the reduction of an LNAPL body in a former refinery in Wyoming. The heating system was designed by Dr. Tom Sale and Daria Akhbari (Akhbari Thesis, 2013) in collaboration with Trihydro.

4.2 Pilot Status Update and System Performance Monitoring

The pilot thermal heating system is currently being tested at the site. Details of the field site, the monitoring system, and the baseline data collection are presented in Chapter 3. In addition to 17 monitoring wells (described in Chapter 3) that allow for depth discrete temperature, water and gas samples to be collected, 10 heating wells were installed. Figure 4.1 shows a conceptual plan view of the pilot area.

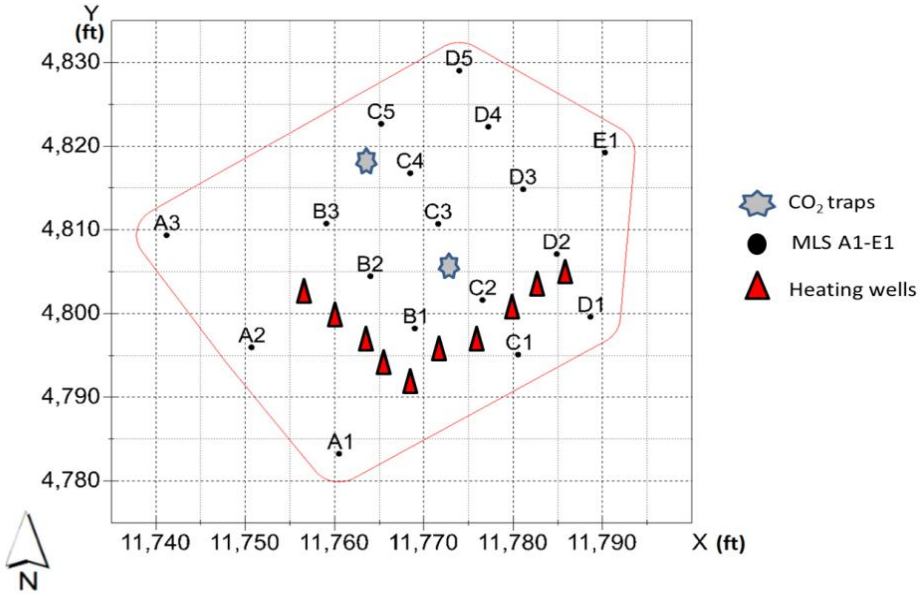


Figure 4.1: Conceptual image of pilot installation pattern. See (Akhbari, 2012) for exact location of heating wells.

The system performance monitoring plan includes: 1) water and gas sample analysis as described in the Materials and Methods section of Chapter 3, 2) temperature measurements through continuous data loggers in select locations and approximately bi-monthly discrete aqueous and gas-phase measurements recorded at each MLS port (data collected and analyzed by Daria Akhbari as part of his Master's Thesis (Akhbari, 2013), and 3) at grade CO_2 efflux

measurements, utilizing carbon dioxide traps (Eflux^{TM}, Fort Collins, CO). Figure 4.2 presents the monitoring schedule and the status of each task.

Data collected is currently being analyzed. Additionally, core samples will be taken in August 2013 to analyze for hydrocarbon concentrations in soil post-heating. The team will review the results and make informed recommendations on system design after the core collection and analysis.

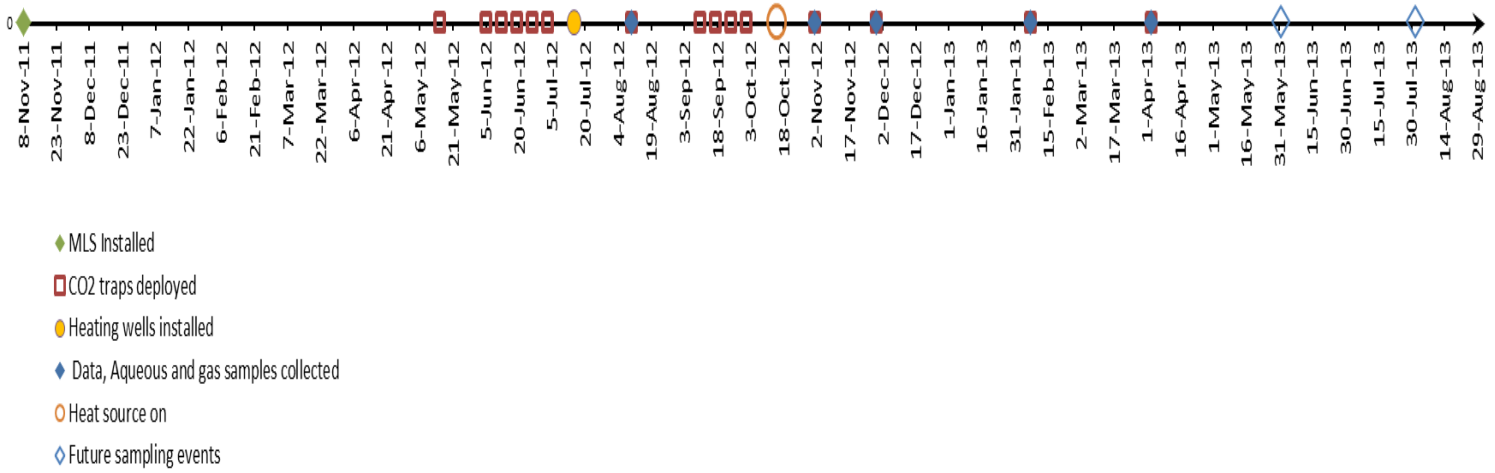


Figure 4.2: STELA,Project timeline.System performance monitoring schedule.

5.0 Overall Conclusions

The site investigation presented herein provided valuable insights regarding biogeochemical processes occuring above, in, and below a LNAPL body at a hydrocarbon-impacted site in Wyoming. The data aquired through this study will serve as a baseline comparison point to evaluate STELA's performance after a year of running the pilot. Furthermore, through the course of this study, a monitoring system to acquire real-time feedback on the system's performance was developed. Data analysis for the aforementioned system is ongoing. Understanding how the measured parameters fluctuate with time will help develop the monitoring plan that needs to accompany the technology currently under development (STELA).

The hollistic, interdisciplinary site analysis approach developed during this study contributed to determining the existence of four distinct biogeochemical zones within the surveyed LNAPL body. High levels of Bacteria and Archaea were detected in the presence of LNAPL, which provides support for microbially-mediated LNAPL losses ocurring at the site. Furthermore, evidence suggests that sulfate reduction and methanogenesis drive significant hydrocarbon depletion processes in strictly anaeerobic zones. Syntrophic pathways are likely ocurring as a well.

The baseline characterization presented herein was perforemed in support of STELA; however, it also provides relevant information for the potential development of other treatment technologies. At this particular site, the presence of metabollically diverse sulfate reducers and the established sufate gradient in combination with the absence of hydrocarbons

in sulfate rich aquifer regions, suggest that recirculating the sulfate rich subsurface waters through the LNAPL body could be an alternative effective treatment option.

References

Acosta-González, A., Rosselló-Móra, R., Marqués, S. 2013. Characterization of the anaerobic microbial community in oil-polluted subtidal sediments: aromatic biodegradation potential after the Prestige oil spill. *Environmental Microbiology*, **15**(1), 77-92.

Alleman, B., Chandler, D., Cole, J., Edwards, E., Fields, M., Haas, P., Halden, R., Hashsham, S., Hazan, T., Johnson, P. 2005. SERDP and ESTCP Expert Panel Workshop on Research and Development Needs for the Environmental Remediation Application of Molecular Biological Tools. DTIC Document.

Antizar-Ladislao, B. 2010. Bioremediation: working with Bacteria. *Elements*, **6**(6), 389-394.

Asakawa, S., Morii, H., Akagawa-Matsuchita, M., Koga, Y., Hayano, K. 1993. Characterization of Methanobrevibacter arboriphilicus SA isolated from a paddy field soil and DNA-DNA hybridization among M. arboriphilicus strains. *International Journal of Systematic Bacteriology*, **43**(4), 683-686.

Beeder, J., Torsvik, T., Lien, T. 1995. Thermodesulforhabdus norvegicus gen. nov., sp. nov., a novel thermophilic sulfate-reducing bacterium from oil field water. *Archives of Microbiology*, **164**(5), 331-336.

Bell, T.H., Yergeau, E., Maynard, C., Juck, D., Whyte, L.G., Greer, C.W. 2013. Predictable bacterial composition and hydrocarbon degradation in Arctic soils following diesel and nutrient disturbance. *The ISME Journal*.

Beller, H., Reinhard, M., Grbić-Galić, D. 1992. Metabolic by-products of anaerobic toluene degradation by sulfate-reducing enrichment cultures. *Applied and Environmental Microbiology*, **58**(9), 3192-3195.

Blotevogel, K.-H., Gahl-Janßen, R., Jannsen, S., Fischer, U., Pilz, F., Auling, G., Macario, A., Tindall, B.J. 1991. Isolation and characterization of a novel mesophilic, fresh-water methanogen from river sediment Methanoculleus oldenburgensis sp. nov. *Archives of Microbiology*, **157**(1), 54-59.

Boopathy, R., Shields, S., Nunna, S. 2012. Biodegradation of crude oil from the BP oil spill in the marsh sediments of Southeast Louisiana, USA. *Applied Biochemistry and Biotechnology*, **167**(6), 1560-1568.

Botton, S., Parsons, J.R. 2007. Degradation of BTX by dissimilatory iron-reducing cultures. *Biodegradation*, **18**(3), 371-381.

Bourne, D.G., Muirhead, A., Sato, Y. 2010. Changes in sulfate-reducing bacterial populations during the onset of black band disease. *The ISME Journal*, **5**(3), 559-564.

Chadalavada, S., Datta, B., Naidu, R. 2012. Optimal Identification of Groundwater Pollution Sources Using Feedback Monitoring Information: A Case Study. *Environmental Forensics*, **13**(2), 140-153.

Chakraborty, R., Coates, J.D. 2005. Hydroxylation and carboxylation—two crucial steps of anaerobic benzene degradation by Dechloromonas strain RCB. *Applied and Environmental Microbiology*, **71**(9), 5427-5432.

Chakraborty, R., Wu, C.H., Hazen, T.C. 2012. Systems biology approach to bioremediation. *Current Opinion in Biotechnology*, **23**(3), 483-490.

Charles J. Newell, Acree, S.D., Ross, R.R., Huling, S.G. 95. Light Nonaqueous Phase Liquids. EPA *Ground Water Issue*.

Chayabutra, C., Ju, L.-K. 2000. Degradation of n-hexadecane and its metabolites by Pseudomonas aeruginosa under microaerobic and anaerobic denitrifying conditions. *Applied and environmental microbiology*, **66**(2), 493-498.

Chen, C.-I., Taylor, R. 1997. Thermophilic biodegradation of BTEX by two consortia of anaerobic bacteria. *Applied microbiology and biotechnology*, **48**(1), 121-128.

Choi, H.-M., Lee, J.-Y. 2011. Groundwater contamination and natural attenuation capacity at a petroleum spilled facility in Korea. *Journal of Environmental Sciences*, **23**(10), 1650-1659.

Coates, J.D., Lonergan, D.J., Philips, E.J., Jenter, H., Lovley, D.R. 1995. Desulfuromonas palmitatis sp. nov., a marine dissimilatory Fe (III) reducer that can oxidize long-chain fatty acids. *Archives of Microbiology*, **164**(6), 406-413.

Coates, J.D., Woodward, J., Allen, J., Philp, P., Lovley, D.R. 1997. Anaerobic degradation of polycyclic aromatic hydrocarbons and alkanes in petroleum-contaminated marine harbor sediments. *Applied and Environmental Microbiology*, **63**(9), 3589-3593.

Corey, A.T. 1994. *Mechanics of immiscible fluids in porous media*. Water Resources Publication.

Cravo-Laureau, C., Matheron, R., Joulian, C., Cayol, J.-L., Hirschler-Réa, A. 2004. Desulfatibacillum alkenivorans sp. nov., a novel n-alkene-degrading, sulfate-reducing bacterium, and emended description of the genus Desulfatibacillum. *International Journal of Systematic and Evolutionary Microbiology*, **54**(5), 1639-1642.

Cui, X., Hunter, W., Yang, Y., Chen, Y., Gan, J. 2011. Biodegradation of pyrene in sand, silt and clay fractions of sediment. *Biodegradation*, **22**(2), 297-307.

Das, N., Chandran, P. 2010. Microbial degradation of petroleum hydrocarbon contaminants: an overview. *Biotechnology Research International*, **2011**.

Declercq, I., Cappuyns, V., Duclos, Y. 2012. Monitored natural attenuation (MNA) of contaminated soils: State of the art in Europe—A critical evaluation. *Science of the Total Environment*.

DiGiulio, D.C. 1992. Evaluation of soil venting application. *Journal of Hazardous Materials*, **32**(2), 279-291.

Dojka, M.A., Hugenholtz, P., Haack, S.K., Pace, N.R. 1998. Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. *Applied and Environmental Microbiology*, **64**(10), 3869-3877.

EPA. 2004. Chapter IX-Monitored Natural Attenuation.

EPA, U.S. 2013a. Frequent questions regarding petroleum brownfields, Vol. 2013.

EPA, U.S. 2013b. Frequent questions regarding petroleum brownfields.

Etkin, D.S. 2001. Analysis of Oil Spill Trends in the United States and Worldwide. *International Oil Spill Conference*.

Etkin, D.S. 2009. Analysis of U.S. Oil spillage. *API Publication* **356**.

Evans, W.C., Fuchs, G. 1988. Anaerobic degradation of aromatic compounds. *Annual Reviews in Microbiology*, **42**(1), 289-317.

Fardeau, M.L., C. Jeanthon, J.L. Garcia, M. Bonilla Salinas, S. L'Haridon, F. Verhe´, J.L. Cayol, B.K.C Patel and B. Ollivier. 2004. Isolation from oil reservoirs of novel thermophilic anaerobes phylogenetically related to Thermoanaerobacter subterraneus: reassignment of T. subterraneus, Thermoanaerobacter yonseiensis, Thermoanaerobacter tengcongensis and Carboxydibrachium pacificum to Caldanaerobacter subterraneus

gen. nov., sp. nov., comb. nov. as four novel subspecies. *International Journal of Systematic and Evolutionary Microbiology*, **54**, 467-474.

Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L., Owens, S., Gilbert, J.A., Wall, D.H., Caporaso, J.G. 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences*, **109**(52), 21390-21395.

Fingas, M. 2010. *Oil spill science and technology*. Gulf Professional Publishing.

Foght, J. 2008. Anaerobic biodegradation of aromatic hydrocarbons: pathways and prospects. *Journal of Molecular Microbiology and Biotechnology*, **15**(2-3), 93-120.

Friederich, M., N.Springer, W. Ludwig and B. Schink 1996. Phylogenetic positions of *Desulfofustis glycolicus* gen. nov., sp. nov., and *Syntrophobotulus glycolicus* gen. nov., sp. nov., two new strict anaerobes growing with glycolic acid. *International Journal of Systematic and Evolutionary Microbiology*, **46**(4), 1065-1069.

Galand, P., Fritze, H., Conrad, R., Yrjälä, K. 2005. Pathways for methanogenesis and diversity of methanogenic archaea in three boreal peatland ecosystems. *Applied and Environmental Microbiology*, **71**(4), 2195-2198.

Galperin, Y., Kaplan, I.R. 2011. Review of microbial processes in the near-surface environment and their implications for the chemical fingerprinting of hydrocarbon fuels. *Environmental Forensics*, **12**(3), 236-252.

Garrity G.M. , B.J.A.a.L.T. 2005. Bergey's Manual of Systematic Bacteriology, (Ed.) G. Garrity, Vol. 2, Williams & Wilkins.

Ghumro, P.B., M. Shafique, M.I. Ali, I. Javed, B. Ahmad, A. Jamal, N. Ali and A. Hameed. 2012. Isolation and screening of protease producing thermophilic Bacillus strains from different soil types of Pakistan. *African Journal of Microbiology Research*, **6**(8), 1663-1668.

Hao, O.J., J. M. Chen, L. Huang and R. L Buglass. 1996. Sulfate-Reducing Bacteria. *Critical Reviews in Environmental Science and Technology*, **26**(1), 155-187.

Harner, N., Richardson, T., Thompson, K., Best, R., Best, A., Trevors, J. 2011. Microbial processes in the Athabasca Oil Sands and their potential applications in microbial enhanced oil recovery. *Journal of industrial microbiology & biotechnology*, **38**(11), 1761-1775.

Head, I., Gray, N., Aitken, C., Sherry, A., Jones, M., Larter, S. 2010. Biogeochemistry of anaerobic crude oil biodegradation. *EGU General Assembly Conference Abstracts*. pp. 14882.

Hemme, C.L., Deng, Y., Gentry, T.J., Fields, M.W., Wu, L., Barua, S., Barry, K., Tringe, S.G., Watson, D.B., He, Z. 2010. Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. *The ISME journal*, **4**(5), 660-672.

Hill, M.O. 1973. Diversity and evenness: a unifying notation and its consequences. *Ecology*, **54**(2), 427-432.

Holmes, D.E., Risso, C., Smith, J.A., Lovley, D.R. 2011. Anaerobic oxidation of benzene by the hyperthermophilic archaeon Ferroglobus placidus. *Applied and Environmental Microbiology*, **77**(17), 5926-5933.

Hua, I., Hopf, A. 2006. Remediation of Aromatic Hydrocarbons in Low Permeability Soils: Updating the Remediation Decision Tree (Synthesis Study). *Joint Transportation Research Program*, 282.

Huntley, D., Beckett, G. 2002. Persistence of LNAPL sources: Relationship between risk reduction and LNAPL recovery. *Journal of Contaminant Hydrology*, **59**(1), 3-26.

Hurst C.J., R.L.c., J.L Garland, D.A. Lipson A.L Mills, L.D. stetzenbach. 2007. *Manual of Environmental Microbiology Third Edition*. ASM Press, Washington DC.

Illman, W.A., Alvarez, P.J. 2009. Performance assessment of bioremediation and natural attenuation. *Critical Reviews in Environmental Science and Technology*, **39**(4), 209-270.

İnceoğlu, Ö., Sablayrolles, C., van Elsas, J.D., Falcão Salles, J. 2013. Shifts in soil bacterial communities associated with the potato rhizosphere in response to aromatic sulfonate amendments. *Applied Soil Ecology*, **63**, 78-87.

Isaksen, M.F., Teske, A. 1996. Desulforhopalus vacuolatus gen nov, sp nov, a new moderately psychrophilic sulfate-reducing bacterium with gas vacuoles isolated from a temperate estuary. *Archives of Microbiology*, **166**(3), 160-168.

J. Auclair, F.L.p., S. Parent, and R. Villemur. 2010. Dissimilatory reduction of nitrate in seawater by a Methylophaga strain containing two highly divergent narG sequences. *The ISME Journal* **4**, 1304-1313.

Jahn, M.K., Haderlein, S.B., Meckenstock, R.U. 2005. Anaerobic degradation of benzene, toluene, ethylbenzene, and o-xylene in sediment-free iron-reducing enrichment cultures. *Applied and Environmental Microbiology*, **71**(6), 3355-3358.

Johnson, P., Lundegard, P., Liu, Z. 2006. Source Zone Natural Attenuation at Petroleum Hydrocarbon Spill Sites—I: Site-Specific Assessment Approach. *Ground Water Monitoring & Remediation*, **26**(4), 82-92.

Jones, D., Head, I., Gray, N., Adams, J., Rowan, A., Aitken, C., Bennett, B., Huang, H., Brown, A., Bowler, B. 2007. Crude-oil biodegradation via methanogenesis in subsurface petroleum reservoirs. *Nature*, **451**(7175), 176-180.

Jørgensen, K.S., Salminen, J.M., Björklöf, K. 2010. Monitored natural attenuation. in: *Bioremediation*, Springer, pp. 217-233.

Kaksonen A.H., S.S., Kroppenstedt R.M., Schumann P. and Puhakka J.A. 2007. Desulfurispora thermophila gen. nov., sp. nov., a thermophilic, spore-forming sulfate-reducer isolated from a sulfidogenic fluidized-bed reactor. *International Journal of Systematic and Evolutionary Microbiology*, **57**, 1089-1094.

Kanaly, R.A., Harayama, S. 2000. Biodegradation of high-molecular-weight polycyclic aromatic hydrocarbons by bacteria. *Journal of Bacteriology*, **182**(8), 2059-2067.

Khadem A.F, P.A.W.S., Jetten M. S. and Op den Camp H.J.M. 2012. Metabolic regulation of "Ca. Methylacidiphilum fumariolicum" SolV cells grown under different nitrogen and oxygen limitations. *Frontiers in Microbiology*, **3**, 1-15.

Kitanidis, P., McCarty, P. 2010. Delivery and Mixing in the Subsurface: Processes and Design Principles for. *Situ*.

Kleikemper, J., Pombo, S.A., Schroth, M.H., Sigler, W.V., Pesaro, M., Zeyer, J. 2005. Activity and diversity of methanogens in a petroleum hydrocarbon-contaminated aquifer. *Applied and Environmental Microbiology*, **71**(1), 149-158.

Kleinsteuber, S., Schleinitz, K.M., Breitfeld, J., Harms, H., Richnow, H.H., Vogt, C. 2008. Molecular characterization of bacterial communities mineralizing benzene under sulfate-reducing conditions. *Fems Microbiology Ecology*, **66**(1), 143-157.

Kleinsteuber, S., Schleinitz, K.M., Vogt, C. 2012. Key players and team play: anaerobic microbial communities in hydrocarbon-contaminated aquifers. *Applied Microbiology and Biotechnology*, **94**(4), 851-873.

Kniemeyer, O., Musat, F., Sievert, S.M., Knittel, K., Wilkes, H., Blumenberg, M., Michaelis, W., Classen, A., Bolm, C., Joye, S.B. 2007. Anaerobic oxidation of short-chain hydrocarbons by marine sulphate-reducing bacteria. *Nature*, **449**(7164), 898-901.

Kristensen, A.H., Henriksen, K., Mortensen, L., Scow, K.M., Moldrup, P. 2010. Soil physical constraints on intrinsic biodegradation of petroleum vapors in a layered subsurface. *Vadose Zone Journal: VZJ*, **9**(1), 137.

Kuever, S.M.S.a.J. 2000. Desulfacinum hydrothermale sp. nov., a thermophilic, sulfate-reducing bacterium from geothermally heated sediments near MilosIsland (Greece). *International Journal of Systematic and Evolutionary Microbiology*, **50**, 1239–1246.

Kunapuli, U., Lueders, T., Meckenstock, R.U. 2007. The use of stable isotope probing to identify key iron-reducing microorganisms involved in anaerobic benzene degradation. *The ISME Journal*, **1**(7), 643-653.

Kuntze, K., Vogt, C., Richnow, H.-H., Boll, M. 2011. Combined application of PCR-based functional assays for the detection of aromatic-compound-degrading anaerobes. *Applied and Environmental Microbiology*, **77**(14), 5056-5061.

LaGrega, M.D., Buckingham, P.L., Evans, J.C. 1994. Hazardous waste management.

Lahvis, M.A., Baehr, A.L., Baker, R.J. 1999. Quantification of aerobic biodegradation and volatilization rates of gasoline hydrocarbons near the water table under natural attenuation conditions. *Water Resources Research*, **35**(3), 753-765.

Landmeyer, J., Chapelle, F., Petkewich, M., Bradley, P. 1998. Assessment of natural attenuation of aromatic hydrocarbons in groundwater near a former manufactured-gas plant, South Carolina, USA. *Environmental Geology*, **34**(4), 279-292.

Lee, J.-Y., Lee, K.-K. 2003. Viability of natural attenuation in a petroleum-contaminated shallow sandy aquifer. *Environmental Pollution*, **126**(2), 201-212.

Liu, R., Zhang, Y., Ding, R., Li, D., Gao, Y., Yang, M. 2009. Comparison of archaeal and bacterial community structures in heavily oil-contaminated and pristine soils. *Journal of Bioscience and Bioengineering*, **108**(5), 400-407.

Lorato, L. 2010. Next Generation Sequencing Technology: 454 Pyrosequencing. *Biotech Articles*.

Lovley, D.R., Giovannoni, S.J., White, D.C., Champine, J.E., Phillips, E., Gorby, Y.A., Goodwin, S. 1993. Geobacter metallireducens gen. nov. sp. nov., a microorganism capable of coupling the complete oxidation of organic compounds to the reduction of iron and other metals. *Archives of Microbiology*, **159**(4), 336-344.

M. Balk, M.A., J. S. Sinninghe Damste, W. Irene, C. Rijpstra and A. J. M. Stams. 2008. Desulfatirhabdium butyrativorans gen. nov., sp. nov., a butyrate-oxidizing, sulfate-reducing bacterium isolated from an anaerobic bioreactor. *International Journal of Systematic and Evolutionary Microbiology*, **58**, 110-115.

Ma, J., Rixey, W.G., DeVaull, G.E., Stafford, B.P., Alvarez, P.J. 2012. Methane bioattenuation and implications for explosion risk reduction along the groundwater to soil surface pathway above a plume of dissolved ethanol. *Environmental Science & Technology*, **46**(11), 6013-6019.

Margesin, R., Hämmerle, M., Tscherko, D. 2007. Microbial activity and community composition during bioremediation of diesel-oil-contaminated soil: effects of hydrocarbon concentration, fertilizers, and incubation time. *Microbial Ecology*, **53**(2), 259-269.

Mbadinga, S.M., Wang, L.-Y., Zhou, L., Liu, J.-F., Gu, J.-D., Mu, B.-Z. 2011. Microbial communities involved in anaerobic degradation of alkanes. *International Biodeterioration & Biodegradation*, **65**(1), 1-13.

McInerney, M.J., Rohlin, L., Mouttaki, H., Kim, U., Krupp, R.S., Rios-Hernandez, L., Sieber, J., Struchtemeyer, C.G., Bhattacharyya, A., Campbell, J.W. 2007. The genome of Syntrophus aciditrophicus: life at the thermodynamic limit of microbial growth. *Proceedings of the National Academy of Sciences*, **104**(18), 7600-7605.

McNally, D.L., Mihelcic, J.R., Lueking, D.R. 1998. Biodegradation of three-and four-ring polycyclic aromatic hydrocarbons under aerobic and denitrifying conditions. *Environmental Science & Technology*, **32**(17), 2633-2639.

Mercer, J.W., Cohen, R.M. 1990. A review of immiscible fluids in the subsurface: Properties, models, characterization and remediation. *Journal of Contaminant Hydrology*, **6**(2), 107-163.

Morris, B.E., Henneberger, R., Huber, H., Moissl-Eichinger, C. 2013. Microbial syntrophy: interaction for the common good. *FEMS Microbiology Reviews*.

Morris, B.E., Herbst, F.A., Bastida, F., Seifert, J., von Bergen, M., Richnow, H.H., Suflita, J.M. 2012. Microbial interactions during residual oil and n-fatty acid metabolism by a methanogenic consortium. *Environmental Microbiology Reports*, **4**(3), 297-306.

Mountfort, D., Brulla, W., Krumholz, L.R., Bryant, M. 1984. Syntrophus buswellii gen. nov., sp. nov.: a benzoate catabolizer from methanogenic ecosystems. *International Journal of Systematic Bacteriology*, **34**(2), 216-217.

Nadim, F., Hoag, G.E., Liu, S., Carley, R.J., Zack, P. 2000. Detection and remediation of soil and aquifer systems contaminated with petroleum products: an overview. *Journal of Petroleum Science and Engineering*, **26**(1), 169-178.

Naidu, R., Nandy, S., Megharaj, M., Kumar, R., Chadalavada, S., Chen, Z., Bowman, M. 2012. Monitored natural attenuation of a long-term petroleum hydrocarbon contaminated sites: a case study. *Biodegradation*, **23**(6), 881-895.

Newman, W.A., Kimball, G., Consultants, D.E. 1991. Dissolved oxygen mapping: a powerful tool for site assessments and ground water monitoring. *Proceedings of the Fifth National Outdoor Action Conference on Aquifer Restoration, Ground Water Monitoring, and Geophysical Methods*.

Österreicher-Cunha, P., Vargas, E.d.A., Guimarães, J.R.D., Lago, G.P., Antunes, F.d.S., da Silva, M.I.P. 2009. Effect of ethanol on the biodegradation of gasoline in an unsaturated tropical soil. *International Biodeterioration & Biodegradation*, **63**(2), 208-216.

Payne, F.C., Suthersan, S.S., Nelson, D.K., Suarez, G., Tasker, I., Akladiss, N. 2006. Enhanced reductive dechlorination of PCE in unconsolidated soils. *Remediation Journal*, **17**(1), 5-21.

Pérez-Pantoja, D., González, B., Pieper, D. 2010. Aerobic degradation of aromatic hydrocarbons. in: *Handbook of Hydrocarbon and Lipid Microbiology*, Springer, pp. 799-837.

Peterson, C.H., Anderson, S.S., Cherr, G.N., Ambrose, R.F., Anghera, S., Bay, S., Blum, M., Condon, R., Dean, T.A., Graham, M. 2012. A tale of two spills: novel science and policy implications of an emerging new oil spill model. *BioScience*, **62**(5), 461-469.

Plugge C.M., Z.W., Scholten J.C.M. and Stams A.J.M. 2011. Metabolic flexibility of sulfate-reducing bacteria. *Frontiers in Microbiology*, **2**, 1-8.

Qu, J., Ren, G., Chen, B., Fan, J., Yong, E. 2011. Effects of lead and zinc mining contamination on bacterial community diversity and enzyme activities of vicinal cropland. *Environmental Monitoring and Assessment*, **182**(1-4), 597-606.

Rastogi, G., Sani, R.K. 2011. Molecular techniques to assess microbial community structure, function, and dynamics in the environment. in: *Microbes and Microbial Technology*, Springer, pp. 29-57.

Rees, G.N., Patel, B.K.C. 2001. Desulforegula conservatrix gen. nov., sp nov., a long-chain fatty acid-oxidizing, sulfate-reducing bacterium isolated from sediments of a freshwater lake. *International Journal of Systematic and Evolutionary Microbiology*, **51**, 1911-1916.

Ricke, P., Kube, M., Nakagawa, S., Erkel, C., Reinhardt, R., Liesack, W. 2005. First genome data from uncultured upland soil cluster alpha methanotrophs provide further evidence for a close phylogenetic relationship to Methylocapsa acidiphila B2 and for high-affinity methanotrophy involving particulate methane monooxygenase. *Applied and Environmental Microbiology*, **71**(11), 7472-7482.

Rockne, K.J., Chee-Sanford, J.C., Sanford, R.A., Hedlund, B.P., Staley, J.T., Strand, S.E. 2000. Anaerobic naphthalene degradation by microbial pure cultures under nitrate-reducing conditions. *Applied and Environmental Microbiology*, **66**(4), 1595-1601.

Röling, W.F., van Verseveld, H.W. 2002. Natural attenuation: What does the subsurface have in store? *Biodegradation*, **13**(1), 53-64.

S. Mayilrajj, A.H.K., R. Cord-Ruwisch , P. Schumann, C. Spro, B.J. Tindall and S. Spring. 2009. *Desulfonauticus autotrophicus* sp. nov., a novel thermophilic sulfate-reducing bacterium isolated from oil-production water and emended description of the genus *Desulfonauticus*. *Extremophiles*, **13**, 247-255.

Sakai, N., Kurisu, F., Yagi, O., Nakajima, F., Yamamoto, K. 2009. Identification of putative benzene-degrading bacteria in methanogenic enrichment cultures. *Journal of Bioscience and Bioengineering*, **108**(6), 501-507.

Sale, T. 2003. Answers to Frequently Asked Questions About Managing Risk at LNAPL Sites. *API Soil and Water Research Bulletin*(18), 1-20.

Savage, K.N., Krumholz, L.R., Gieg, L.M., Parisi, V.A., Suflita, J.M., Allen, J., Philp, R.P., Elshahed, M.S. 2010. Biodegradation of low-molecular-weight alkanes under mesophilic, sulfate-reducing conditions: metabolic intermediates and community patterns. *FEMS Microbiology Ecology*, **72**(3), 485-495.

School, S.T. 1954. Petroleum refining and related operations.

Schwarzenbach, R.P., Gschwend, P.M., Imboden, D.M. 2005. *Environmental Organic Chemistry*. Wiley-Interscience.

Shendure, J.A., Porreca, G.J., Church, G.M. 2008. Overview of DNA sequencing strategies. *Current Protocols in Molecular Biology*, 7.1. 1-7.1. 11.

Siddique, T., Penner, T., Semple, K., Foght, J.M. 2011. Anaerobic biodegradation of longer-chain n-alkanes coupled to methane production in oil sands tailings. *Environmental Science & Technology*, **45**(13), 5892-5899.

Sieber, J.R., McInerney, M.J., Gunsalus, R.P. 2012. Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. *Annual Review of Microbiology*, **66**, 429-452.

Sihota, N.J., Mayer, K.U. 2012. Characterizing vadose zone hydrocarbon biodegradation using carbon dioxide effluxes, isotopes, and reactive transport modeling. *Vadose Zone Journal*, **11**(4).

Sihota, N.J., Singurindy, O., Mayer, K.U. 2010. CO2-Efflux Measurements for Evaluating Source Zone Natural Attenuation Rates in a Petroleum Hydrocarbon Contaminated Aquifer. *Environmental Science & Technology*, **45**(2), 482-488.

Simarro, R., González, N., Bautista, L.F., Molina, M.C. 2013. Biodegradation of high-molecular-weight polycyclic aromatic hydrocarbons by a wood-degrading consortium at low temperatures. *FEMS microbiology ecology*, **83**(2), 438-449.

Stams, G.M.a.A.J.M. 2008. The ecology and biotechnology of sulphate-reducing bacteria. *Nature Reviews*, **6**.

Steinberg, L.M., Regan, J.M. 2009. mcrA-targeted real-time quantitative PCR method to examine methanogen communities. *Applied and Environmental Microbiology*, **75**(13), 4435-4442.

Sun, M.Y., Dafforn, K.A., Johnston, E.L., Brown, M.V. 2013. Core sediment bacteria drive community response to anthropogenic contamination over multiple environmental gradients. *Environmental Microbiology*.

Sutton, N.B., Maphosa, F., Morillo, J.A., Al-Soud, W.A., Langenhoff, A.A., Grotenhuis, T., Rijnaarts, H.H., Smidt, H. 2013. Impact of Long-Term Diesel Contamination on Soil Microbial Community Structure. *Applied and environmental microbiology*, **79**(2), 619-630.

Sylves, R.T., Comfort, L.K. 2012. The Exxon Valdez and BP Deepwater Horizon Oil Spills Reducing Risk in Socio-Technical Systems. *American Behavioral Scientist*, **56**(1), 76-103.

T. Ren, L.R., AND R. Knowles. 2000. Production and consumption of nitiric oxide by three methanothrophic bacteria. *Applied and Environmental Microbiology*, **66**(9), 3891-3897.

Tan, Y., Ji, G. 2010. Bacterial community structure and dominant bacteria in activated sludge from a 70 C ultrasound-enhanced anaerobic reactor for treating carbazole-containing wastewater. *Bioresource Technology*, **101**(1), 174-180.

Thavamani, P., Malik, S., Beer, M., Megharaj, M., Naidu, R. 2012. Microbial activity and diversity in long-term mixed contaminated soils with respect to polyaromatic hydrocarbons and heavy metals. *Journal of environmental management*, **99**, 10-17.

Tierney, M., Young, L. 2010. Anaerobic degradation of aromatic hydrocarbons. in: *Handbook of Hydrocarbon and Lipid Microbiology*, Springer, pp. 925-934.

Torsvik, H.O.a.T. 2007. Desulfotignum toluenicum sp. nov., a novel toluene-degrading, sulphate-reducing bacterium isolated from an oil-reservoir model column. *International Journal of Systematic and Evolutionary Microbiology*, **57**, 2865–2869.

Trzesicka-Mlynarz, D., Ward, O. 1995. Degradation of polycyclic aromatic hydrocarbons (PAHs) by a mixed culture and its component pure cultures, obtained from PAH-contaminated soil. *Canadian Journal of Microbiology*, **41**(6), 470-476.

Ueno, Y., Haruta, S., Ishii, M., Igarashi, Y. 2001. Characterization of a microorganism isolated from the effluent of hydrogen fermentation by microflora. *Journal of Bioscience and Bioengineering*, **92**(4), 397-400.

Ulrich, A.C., Edwards, E.A. 2003. Physiological and molecular characterization of anaerobic benzene-degrading mixed cultures. *Environmental microbiology*, **5**(2), 92-102.

Uren, L.C. 1956. *Petroleum production engineering*. McGraw-Hill.

van Agteren, M.H., Keuning, S., Janssen, D.B. 1998. Aromatic compounds. in: *Handbook on Biodegradation and Biological Treatment of Hazardous Organic Compounds*, Springer, pp. 189-286.

Viñas, M., Sabaté, J., Espuny, M.J., Solanas, A.M. 2005. Bacterial community dynamics and polycyclic aromatic hydrocarbon degradation during bioremediation of heavily creosote-contaminated soil. *Applied and Environmental Microbiology*, **71**(11), 7008-7018.

Vogt, C., Kleinsteuber, S., Richnow, H.H. 2011. Anaerobic benzene degradation by bacteria. *Microbial Biotechnology*, **4**(6), 710-724.

Voordouw, G., Armstrong, S.M., Reimer, M.F., Fouts, B., Telang, A.J., Shen, Y., Gevertz, D. 1996. Characterization of 16S rRNA genes from oil field microbial communities indicates the presence of a variety of sulfate-reducing, fermentative, and sulfide-oxidizing bacteria. *Applied and Environmental Microbiology*, **62**(5), 1623-1629.

Wai, L. 1995. Locating US national standards for drinking water. *Journal of government information*, **22**(2), 101-117.

Walden, T., Spence, L. 1997. Risk-based BTEX screening criteria for a groundwater irrigation scenario. *Human and Ecological Risk Assessment*, **3**(4), 699-722.

Weelink, S.A. 2008. *Degradation of benzene and other aromatic hydrocarbons by anaerobic bacteria*. Wageningen Universiteit.

Weelink, S.A., van Eekert, M.H., Stams, A.J. 2010. Degradation of BTEX by anaerobic bacteria: physiology and application. *Reviews in Environmental Science and Bio/Technology*, **9**(4), 359-385.

Whitby, C., Lund, S.T. 2009. *Applied Microbiology and Molecular Biology in Oil Field Systems*. Springer Science+ Business Media.

Widdel, F., Rabus, R. 2001. Anaerobic biodegradation of saturated and aromatic hydrocarbons. *Current Opinion in Biotechnology*, **12**(3), 259-276.

Wiedemeier, T.H., Rifai, H.S., Newell, C.J., Wilson, J.T. 1999. *Natural attenuation of fuels and chlorinated solvents in the subsurface*. Wiley.

Wu, Y., Luo, Y., Zou, D., Ni, J., Liu, W., Teng, Y., Li, Z. 2008. Bioremediation of polycyclic aromatic hydrocarbons contaminated soil with Monilinia sp.: degradation and microbial community analysis. *Biodegradation*, **19**(2), 247-257.

Yagi, J.M., Suflita, J.M., Gieg, L.M., DeRito, C.M., Jeon, C.-O., Madsen, E.L. 2010. Subsurface cycling of nitrogen and anaerobic aromatic hydrocarbon biodegradation revealed by nucleic acid and metabolic biomarkers. *Applied and Environmental Microbiology*, **76**(10), 3124-3134.

Zedelius, J., Rabus, R., Grundmann, O., Werner, I., Brodkorb, D., Schreiber, F., Ehrenreich, P., Behrends, A., Wilkes, H., Kube, M. 2011. Alkane degradation under anoxic conditions by a nitrate-reducing bacterium with possible involvement of the electron acceptor in substrate activation. *Environmental Microbiology Reports*, **3**(1), 125-135.

Zvomuya, F., Murata, A.P. 2012.*Soil Contamination and Remediation. Encyclopedia of Environmetrics.*

Appendix A: Diesel Affects Reproducibility of DNA Extraction Assay

The first step towards characterizing the microbial ecology of the system is extracting DNA from oily soils. Figure A.1 shows the DNA yields obtained for a soil sample, collected at 0.5 ft bgs. 5 g of this sample was spiked with different volumes of diesel. The experiment was done to study the impact of hydrocarbon content on the DNA extraction protocol. While the presence of diesel does not affect the average extraction yield, it does affect the reproducibility of the results obtained with the extraction method. It was also observed that hydrocarbons and humic substances present in the soil interfere with the qPCR assays chosen to quantify the system's microbial population (data not shown).

A soil wash protocol was adapted to remove hydrocarbons and other PCR inhibitors from the samples. After pretreating the samples, clean DNA material was obtained (Figure A.2). This new protocol also enables performance of downstream genetic analysis by qPCR. Figure A.3 illustrates the linear relationship between the quantity of the 16S rRNA genes reported by the qPCR assay and the amount of DNA (ng) assayed.

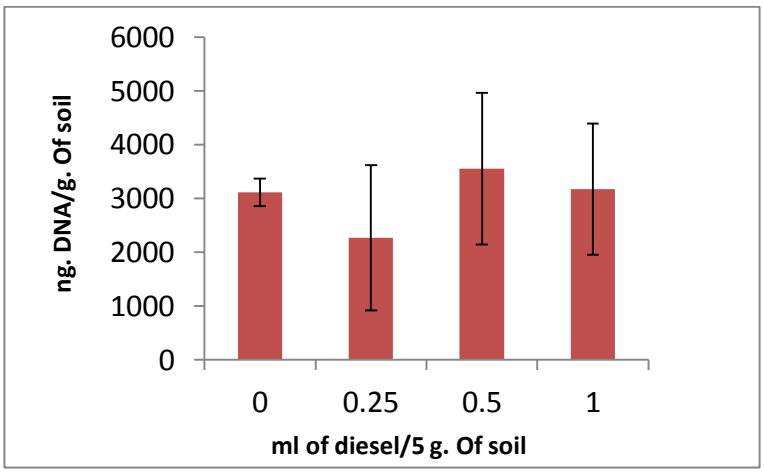


Figure A.1: Soil sample collected at 0.5 ft below ground surface spiked with different amounts of diesel.

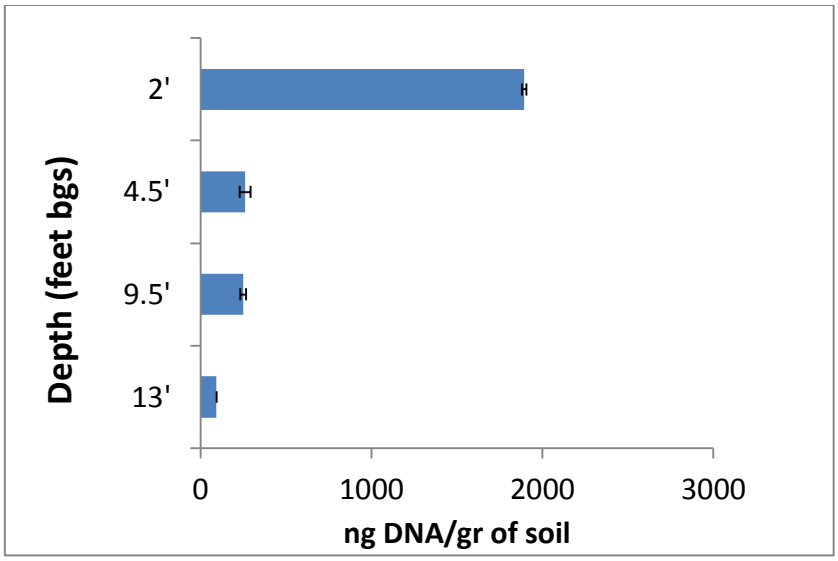


Figure A.2: DNA extraction yields after sample pretreatment.

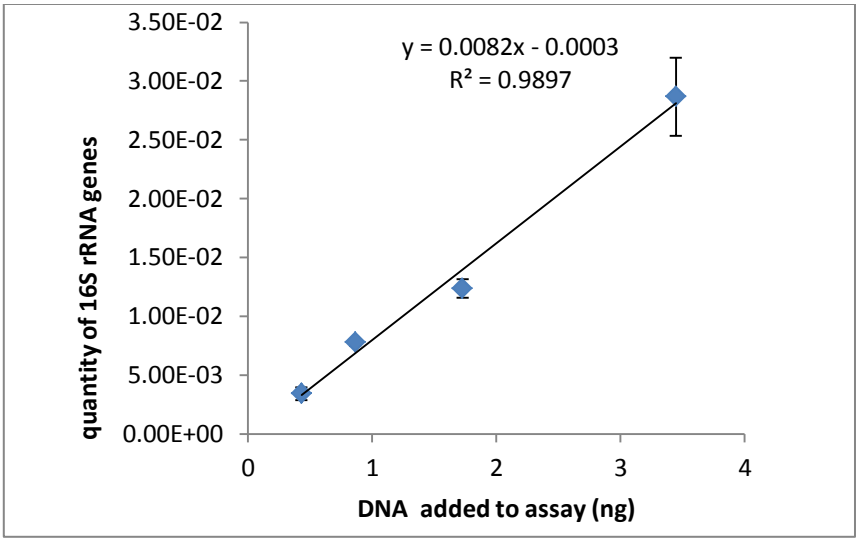


Figure A.3: Linear behavior of qPCR assay, after sample pretreatment. Linear relationship between the quantity of the 16S rRNA genes reported by the qPCR assay and the amount of DNA (ng) assayed, when DNA is extracted applying the sample pretreatment protocol.

Appendix B: Images of the Geochemical Characterization Analysis of Transect C

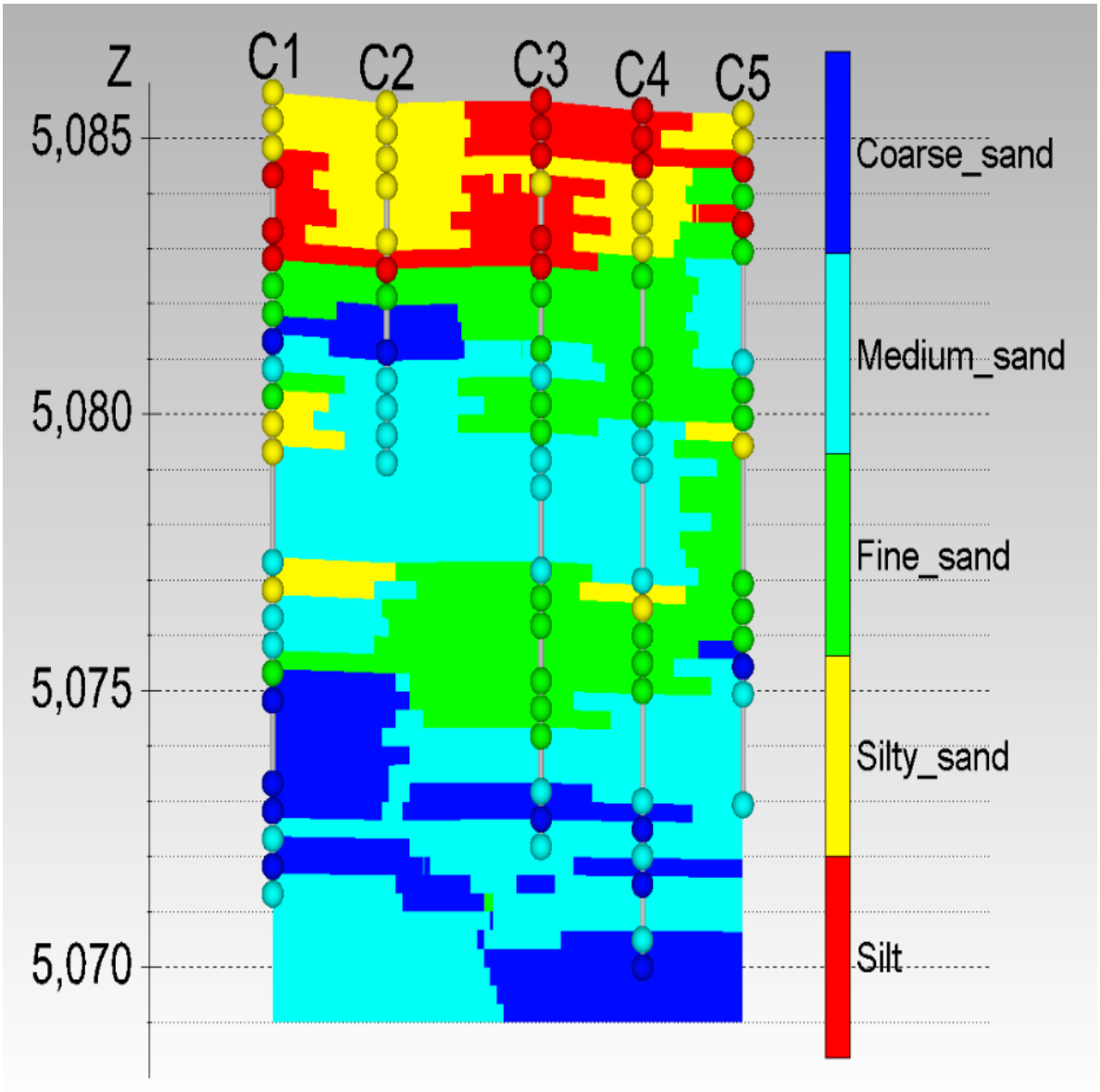


Figure B.1: Soil type sampled with depth along transect C.

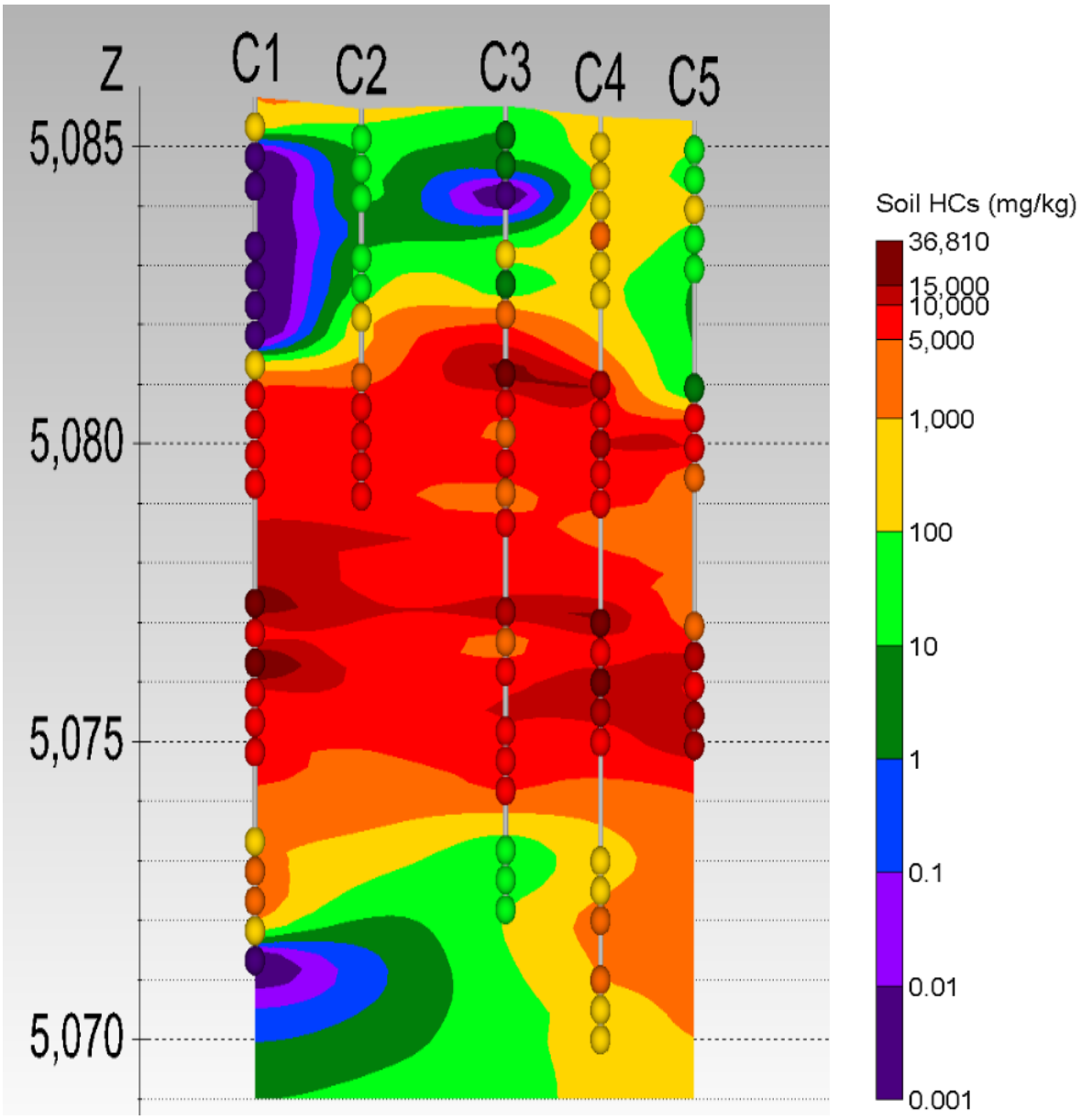


Figure B.2: TPH (mg/kg) distribution with depth along transect C.

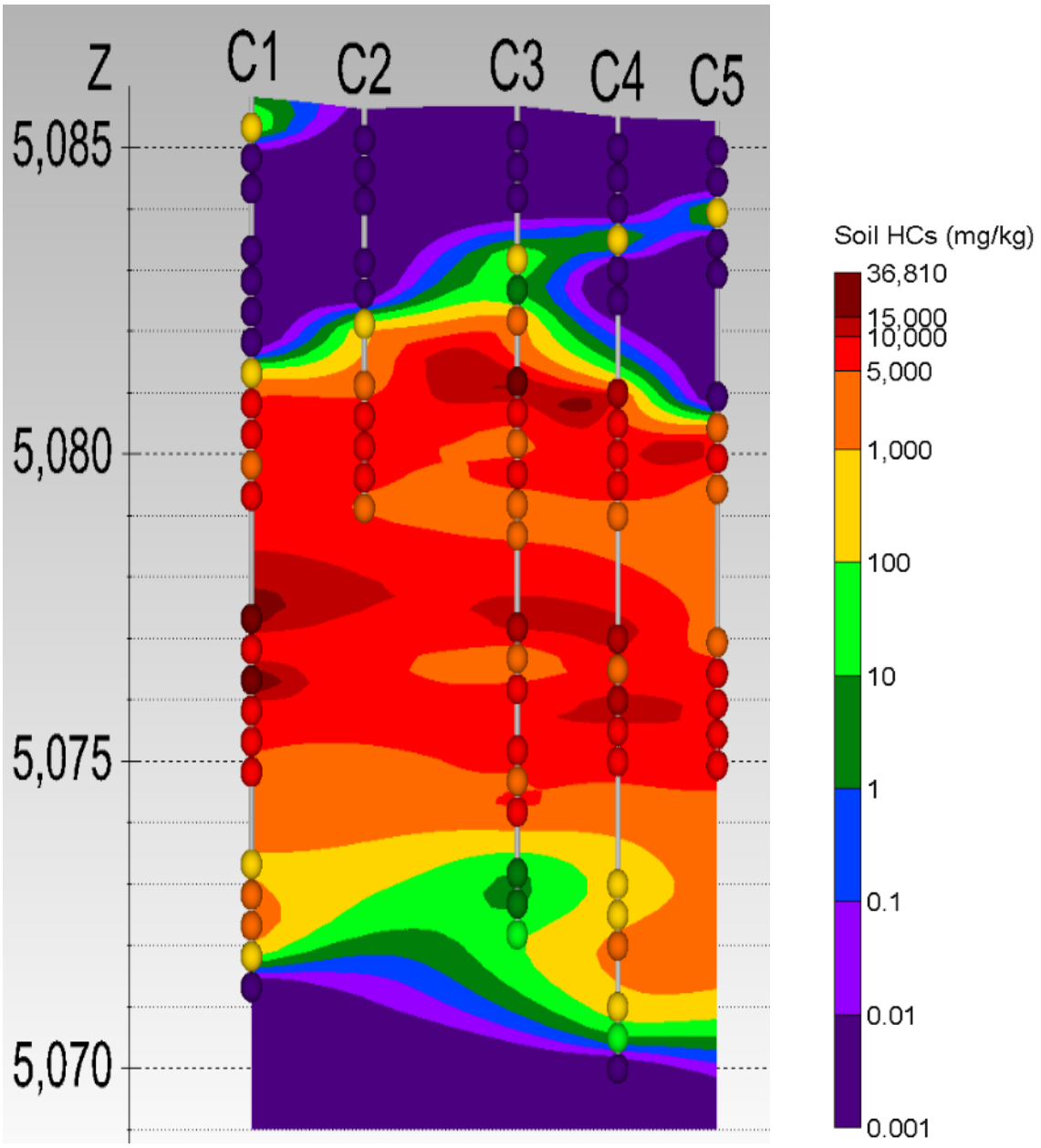


Figure B.3: DRO (mg/kg) distribution with depth along transect C.

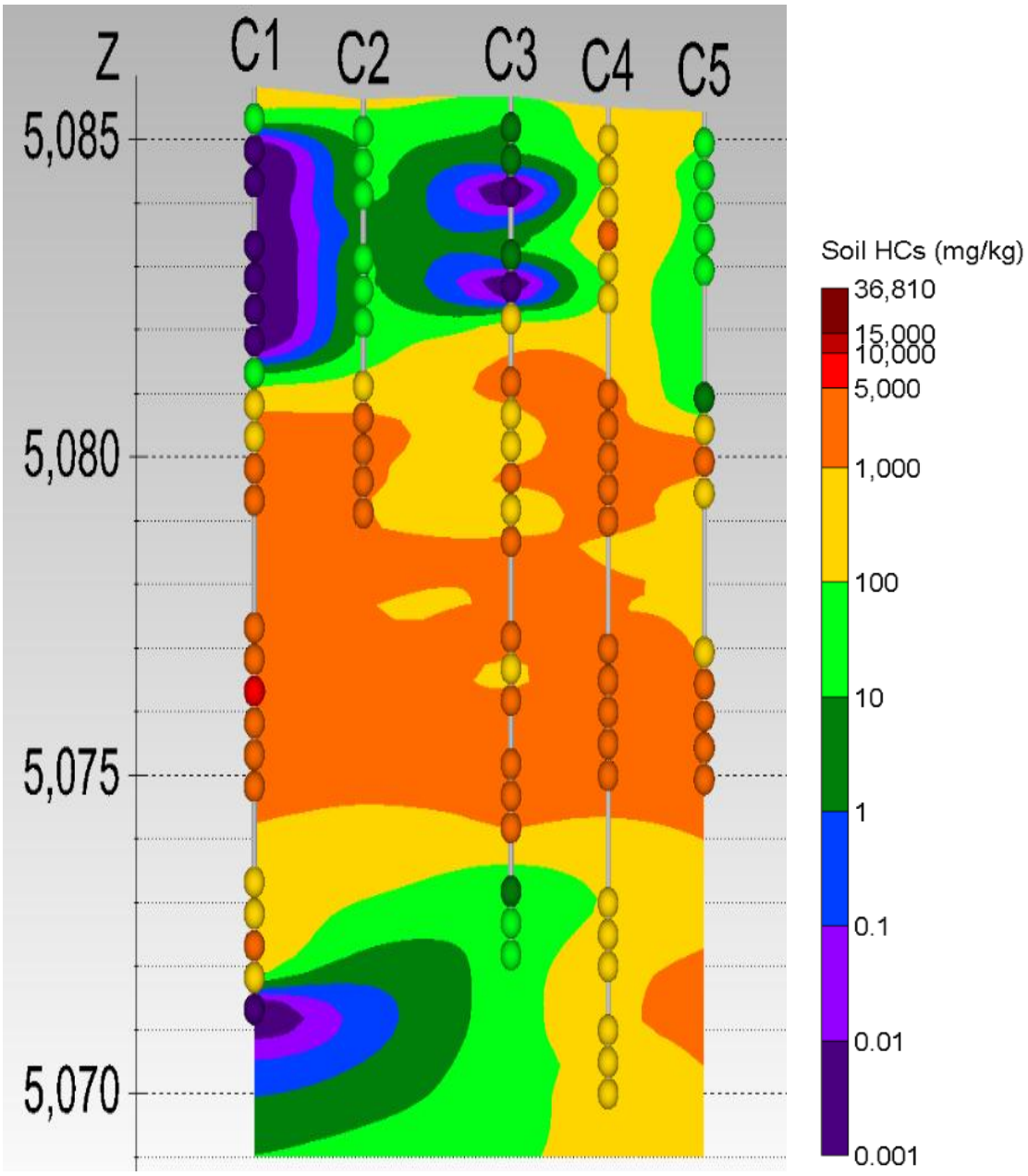


Figure B.4: GRO (mg/kg) distribution with depth along transect C.

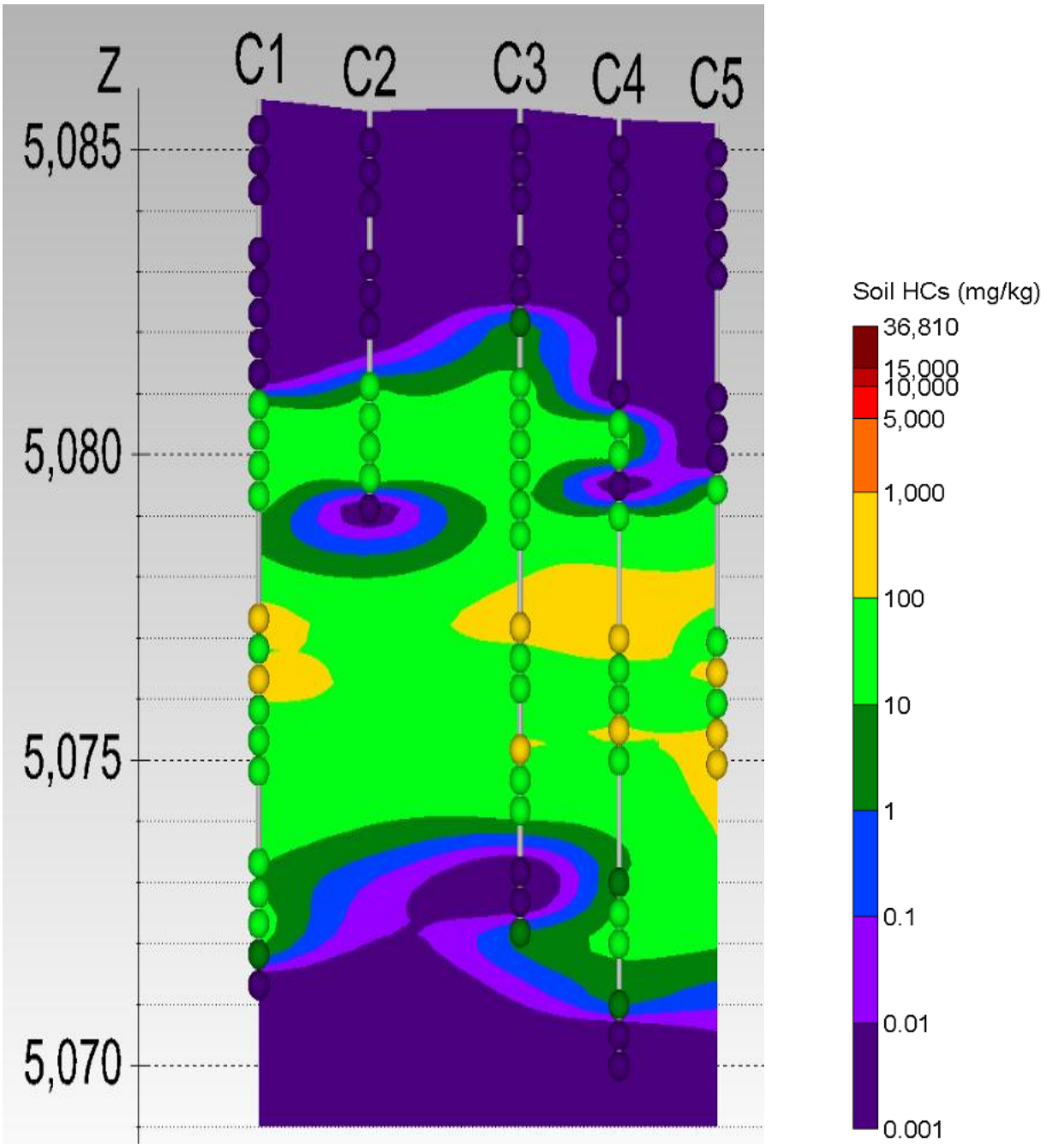


Figure B.5: Benzene (mg/kg) distribution with depth along transect C.

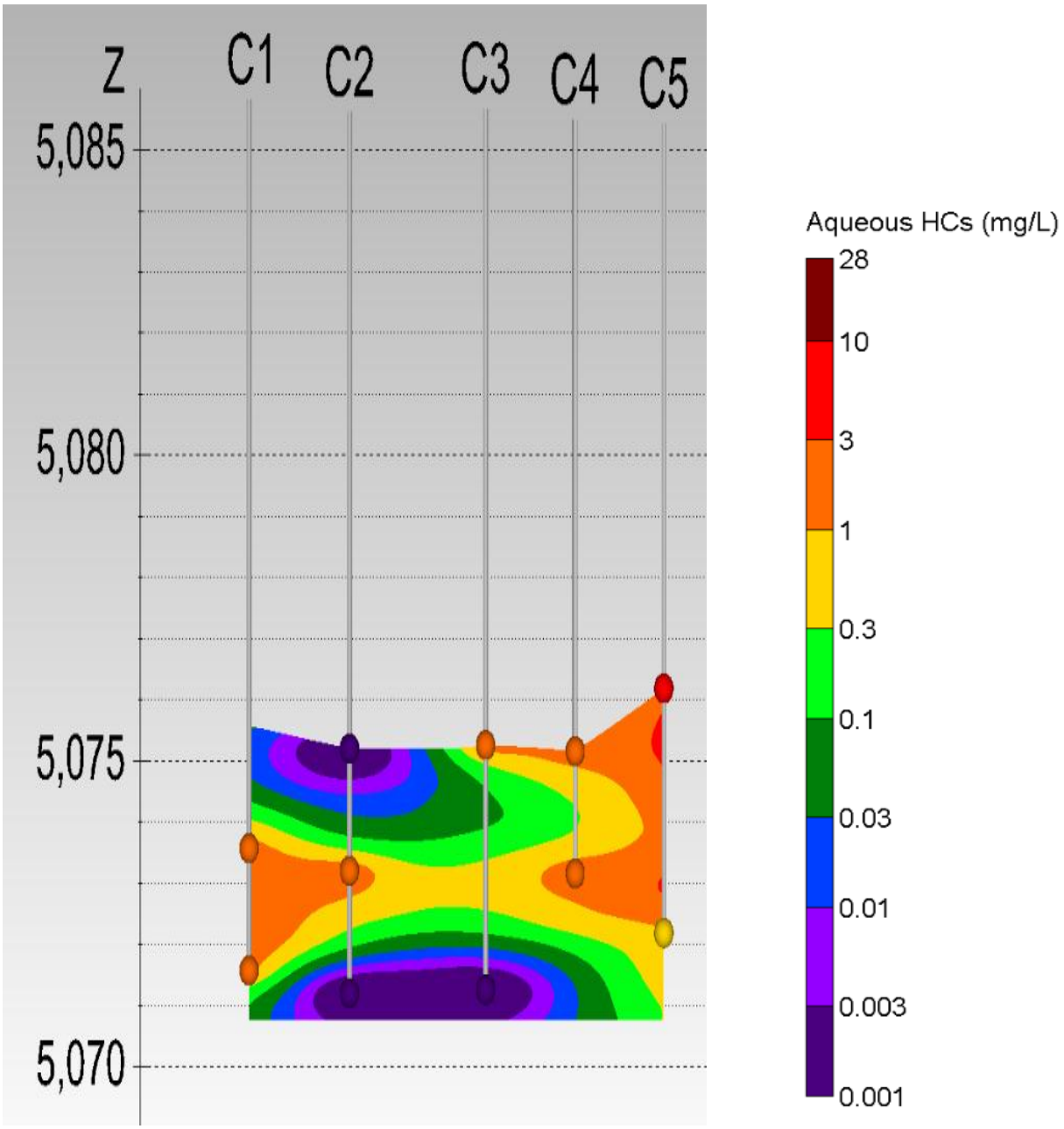


Figure B.6: Benzene (mg/L) aqueous distribution with depth along transect C.

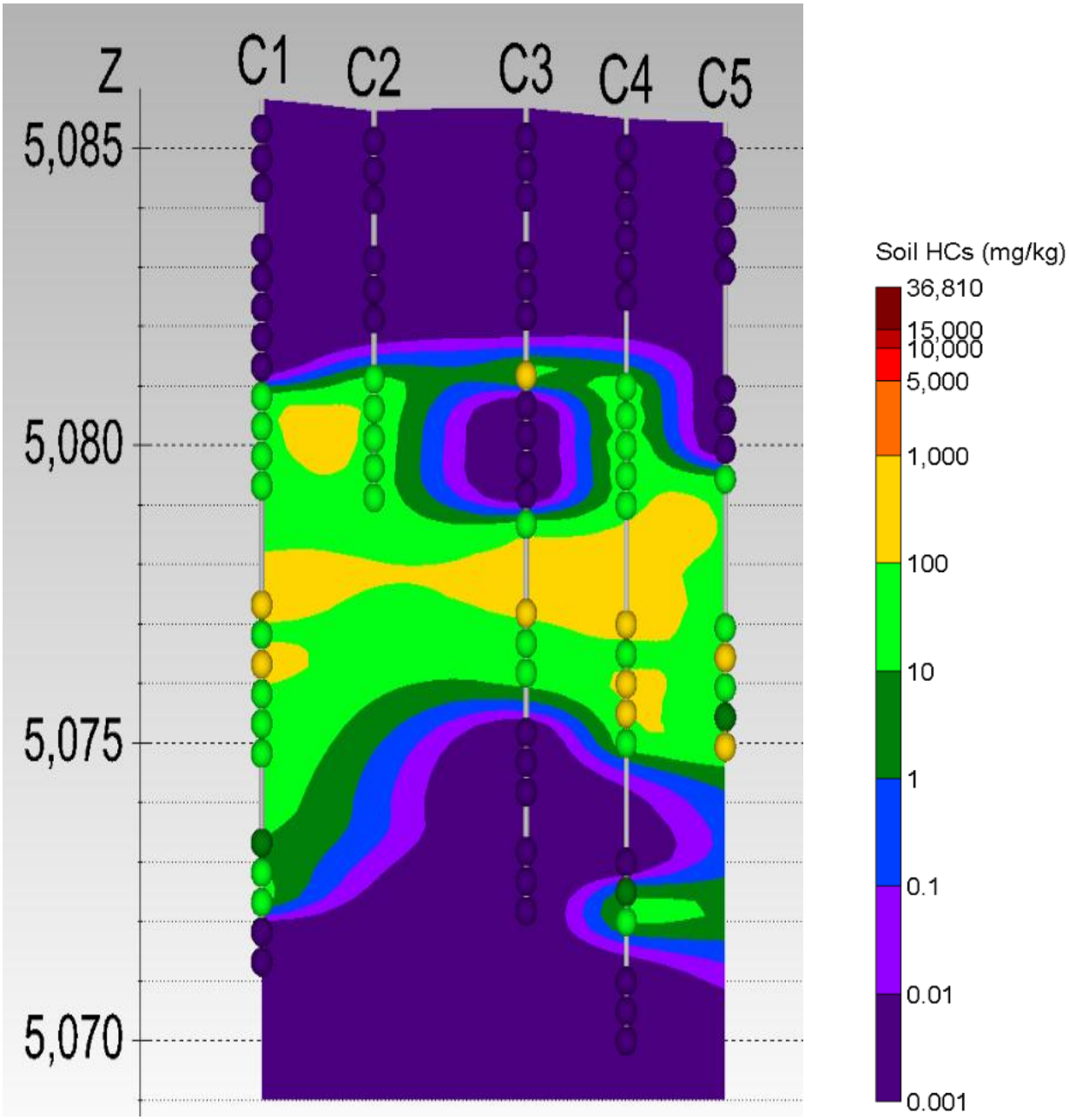


Figure B.7: Ethylbenzene (mg/kg) distribution with depth along transect C.

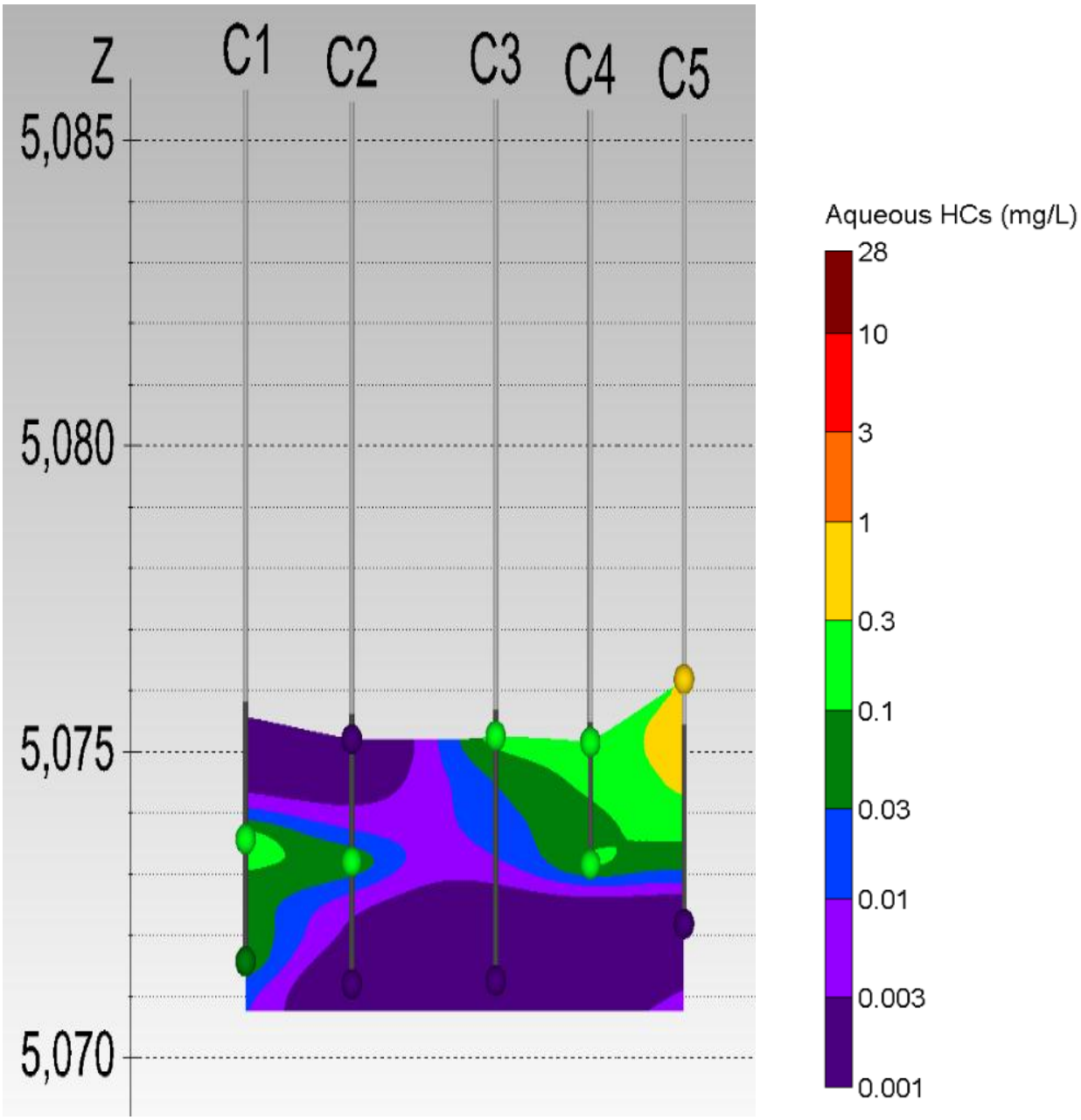


Figure B.8: Ehylbenzene (mg/L) aqueous distribution with depth along transect C.

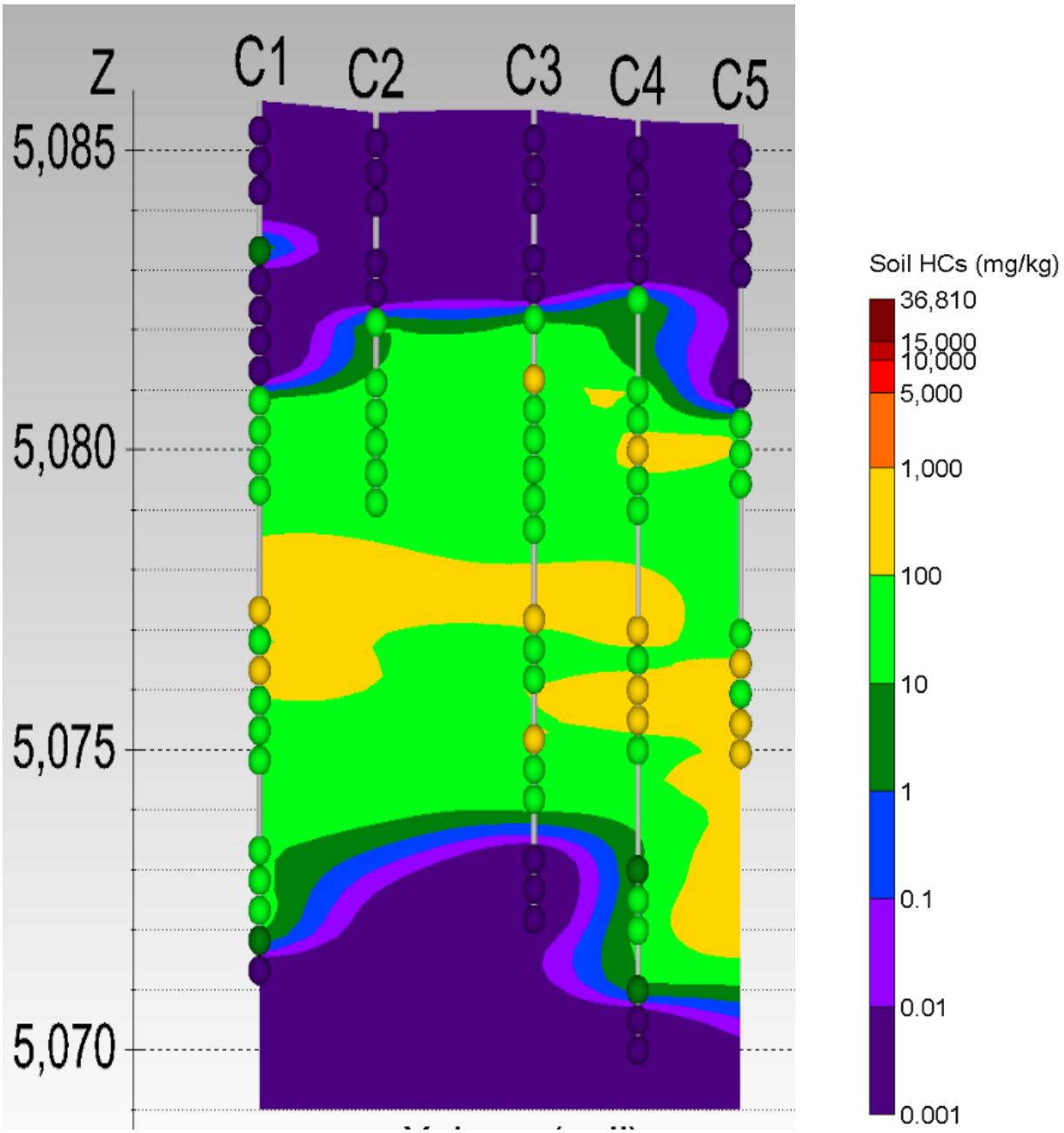


Figure B.9: *m*&*p*-xylenes (mg/kg) distribution with depth along transect C

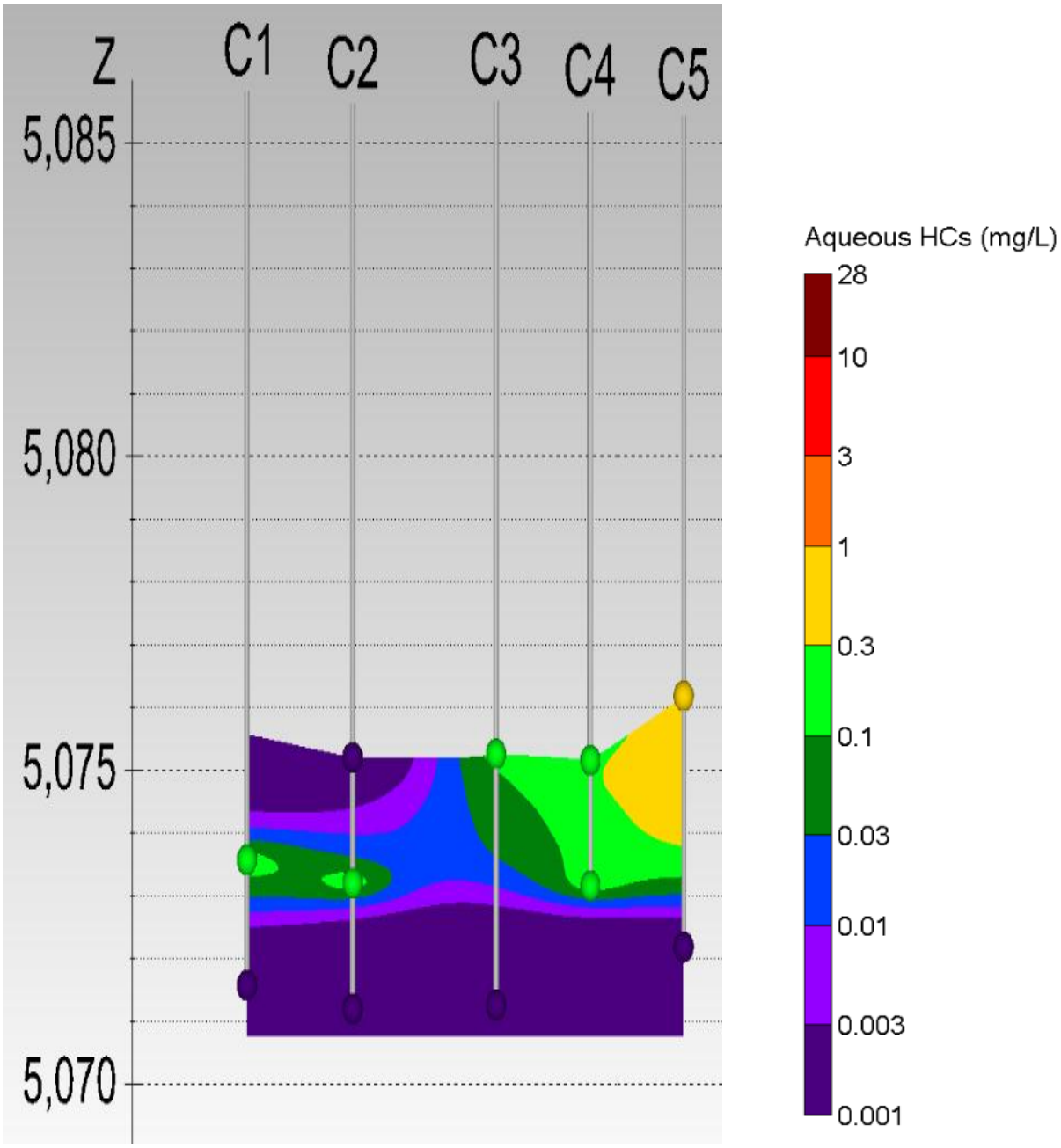


Figure B.10: *m*&*p*-xylenes (mg/L) aqueous distribution with depth along transect C.

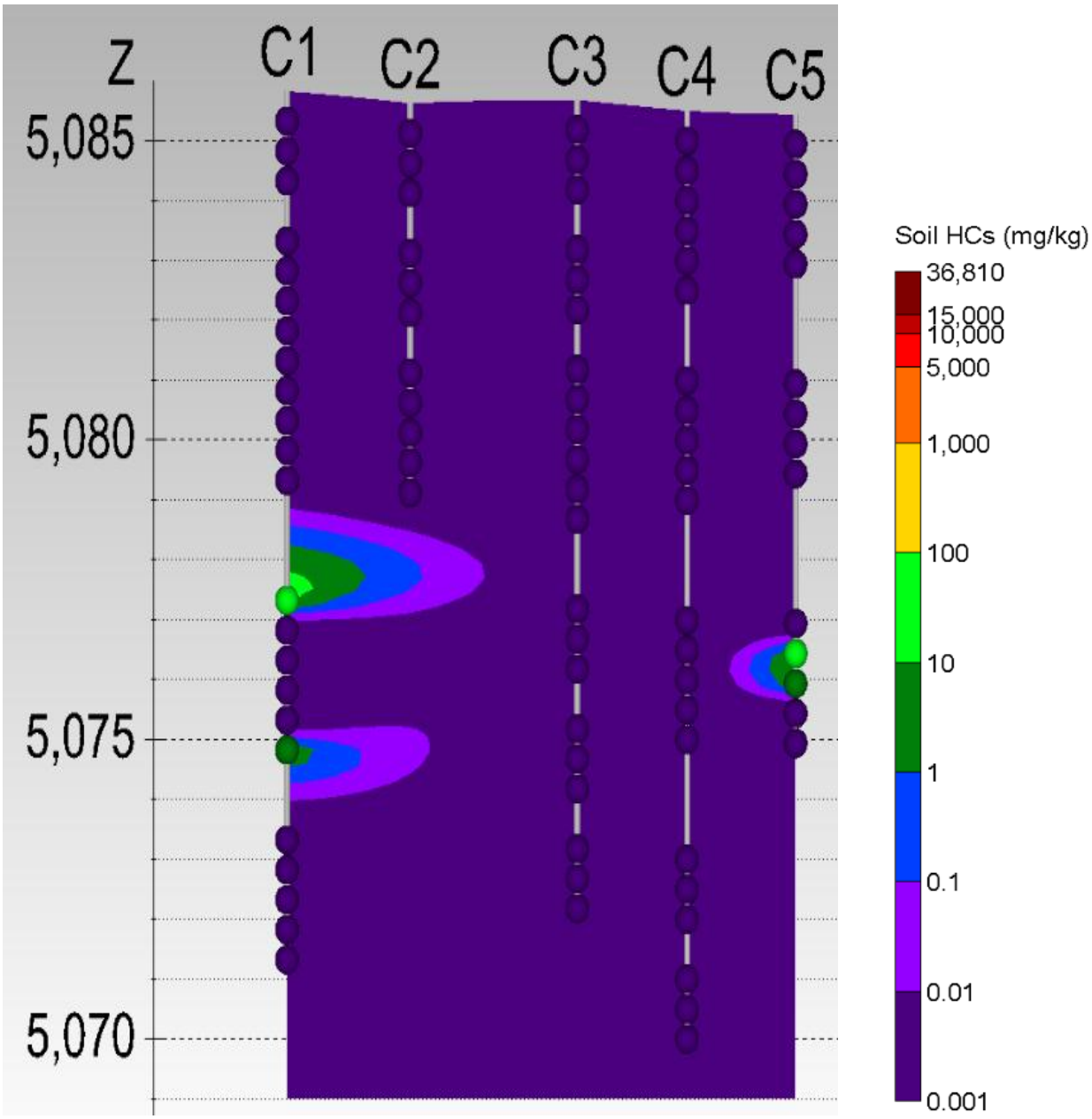


Figure B.11: *o*-xylenes (mg/kg) distribution with depth along transect C.

Appendix C: Images of the Gas Analysis and Inorganic Water Chemistry Analysis of Transect C

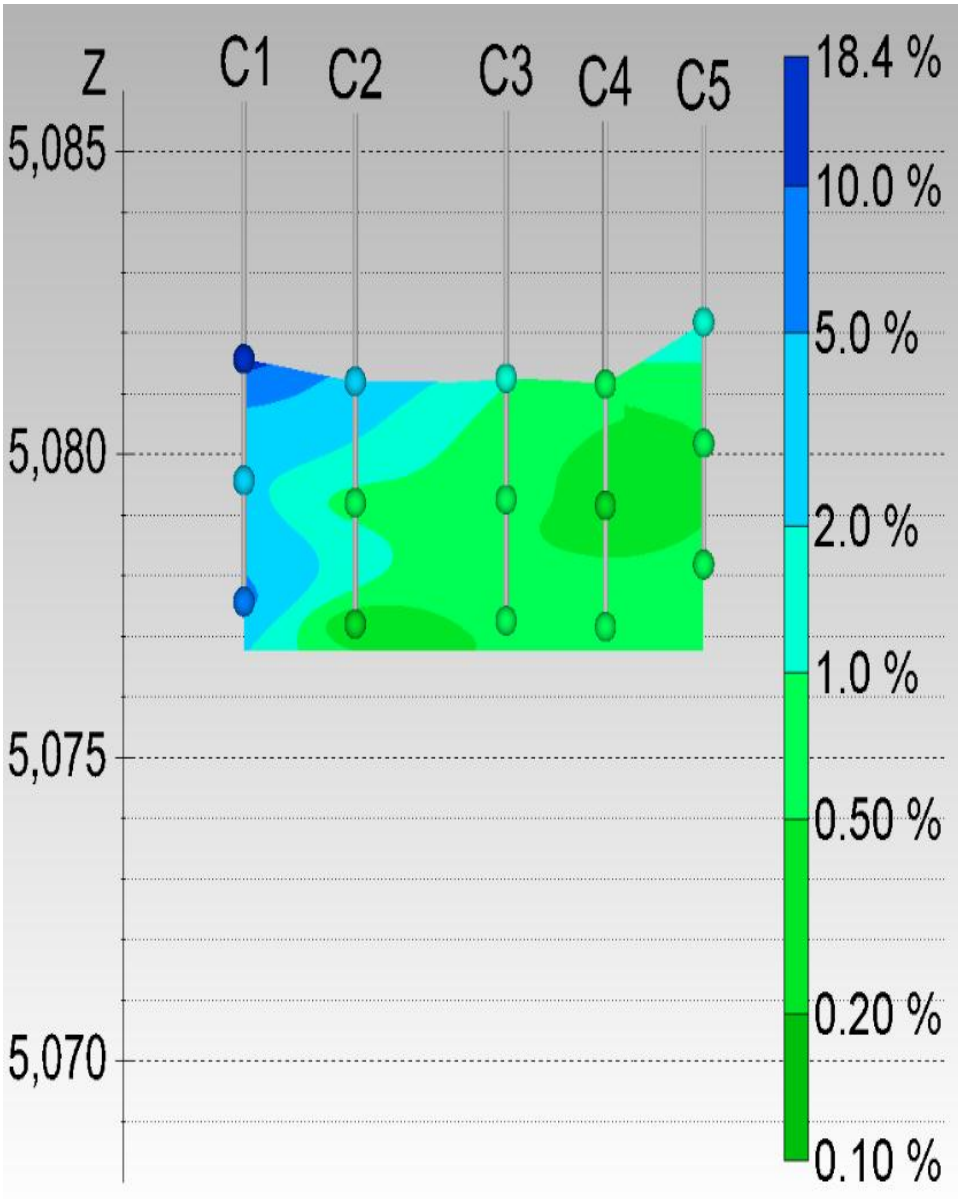
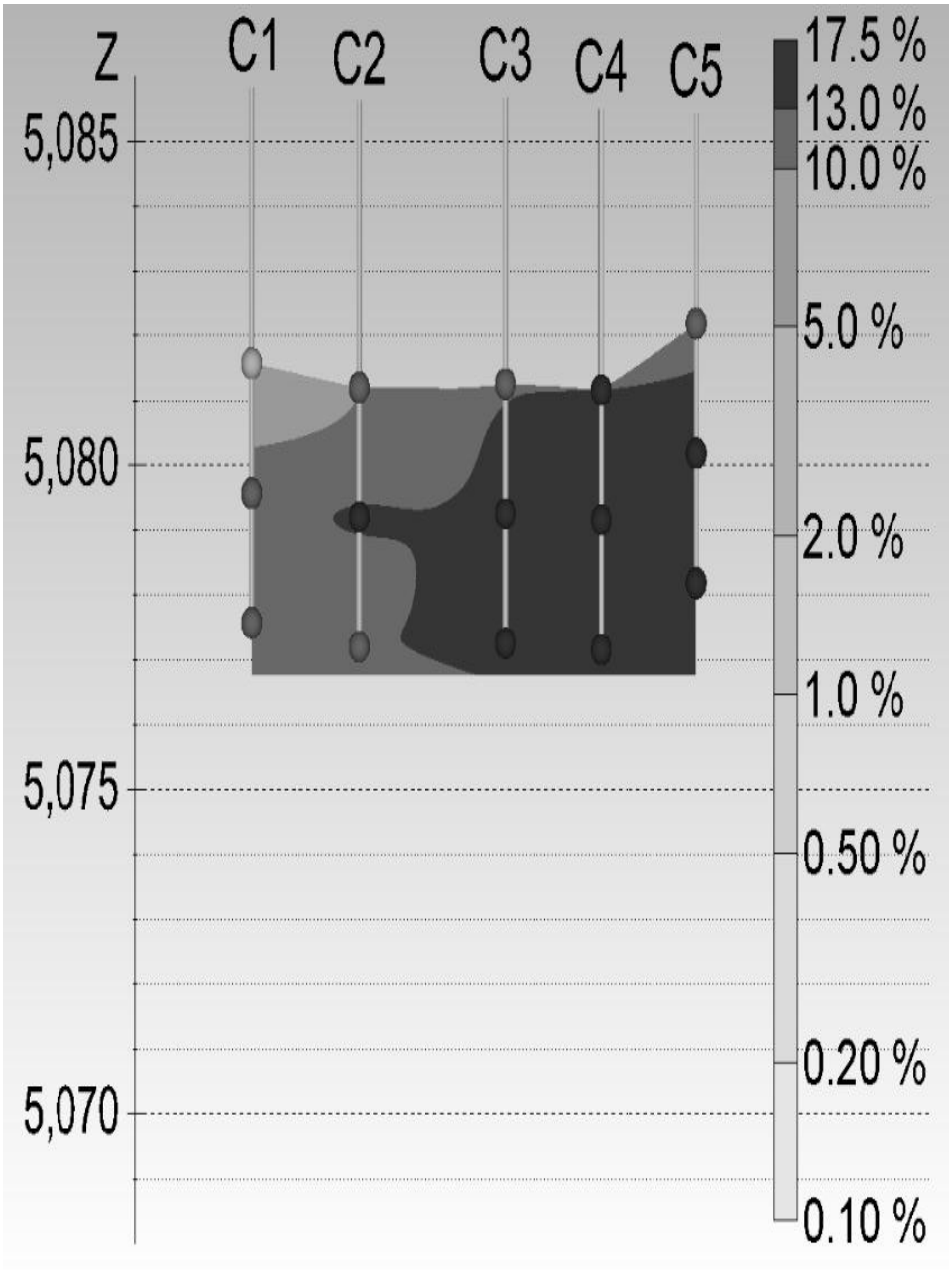


Figure C.1: Oxygen levels (%vol/vol) in the vadose zone measured along transect C.



FigureC.2: Carbon dioxide levels (%vol/vol) in the vadose zone measured along transect C.

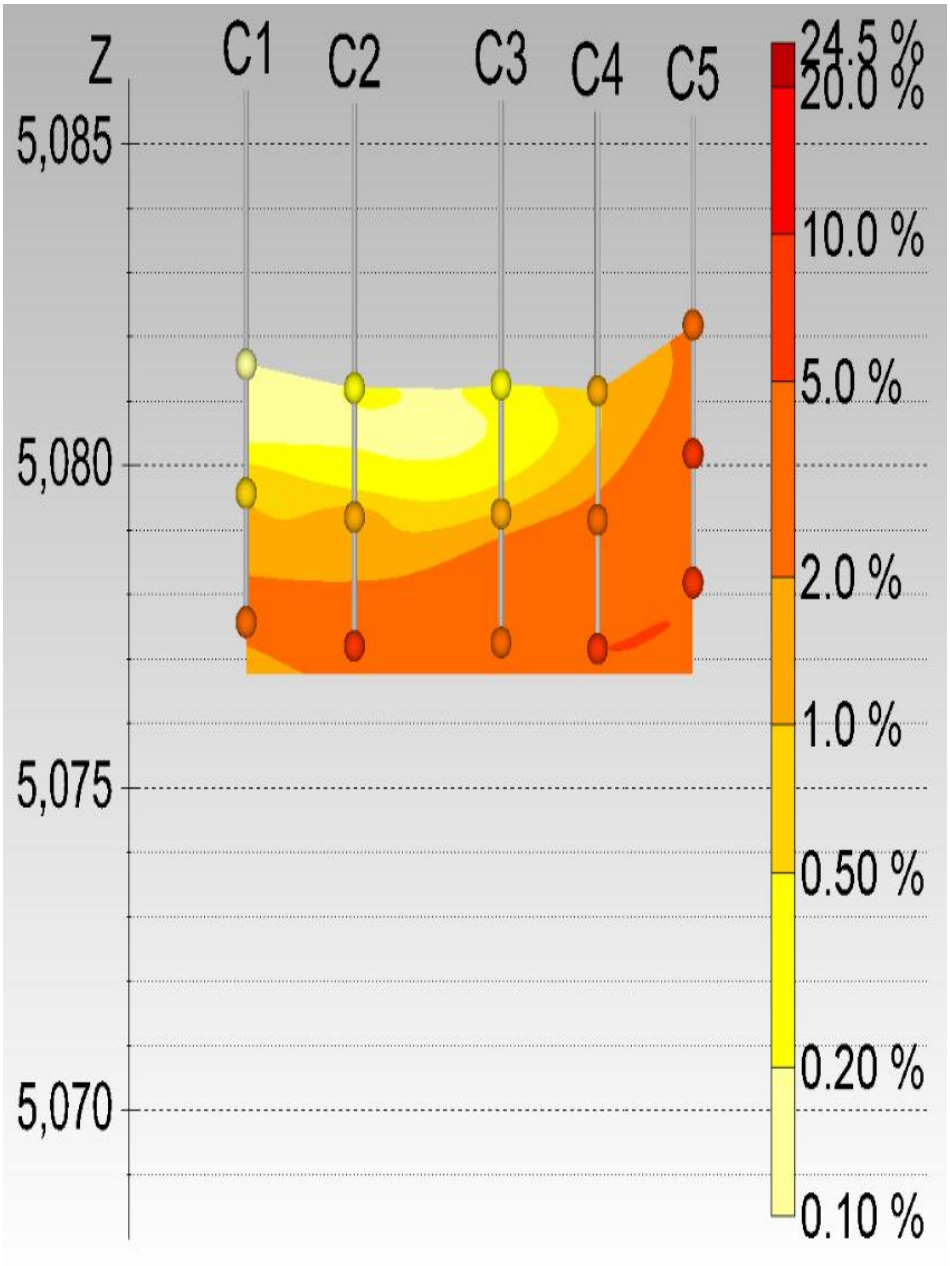


Figure C.3: Methane levels (%vol/vol) in the vadose zone measured along transect C.

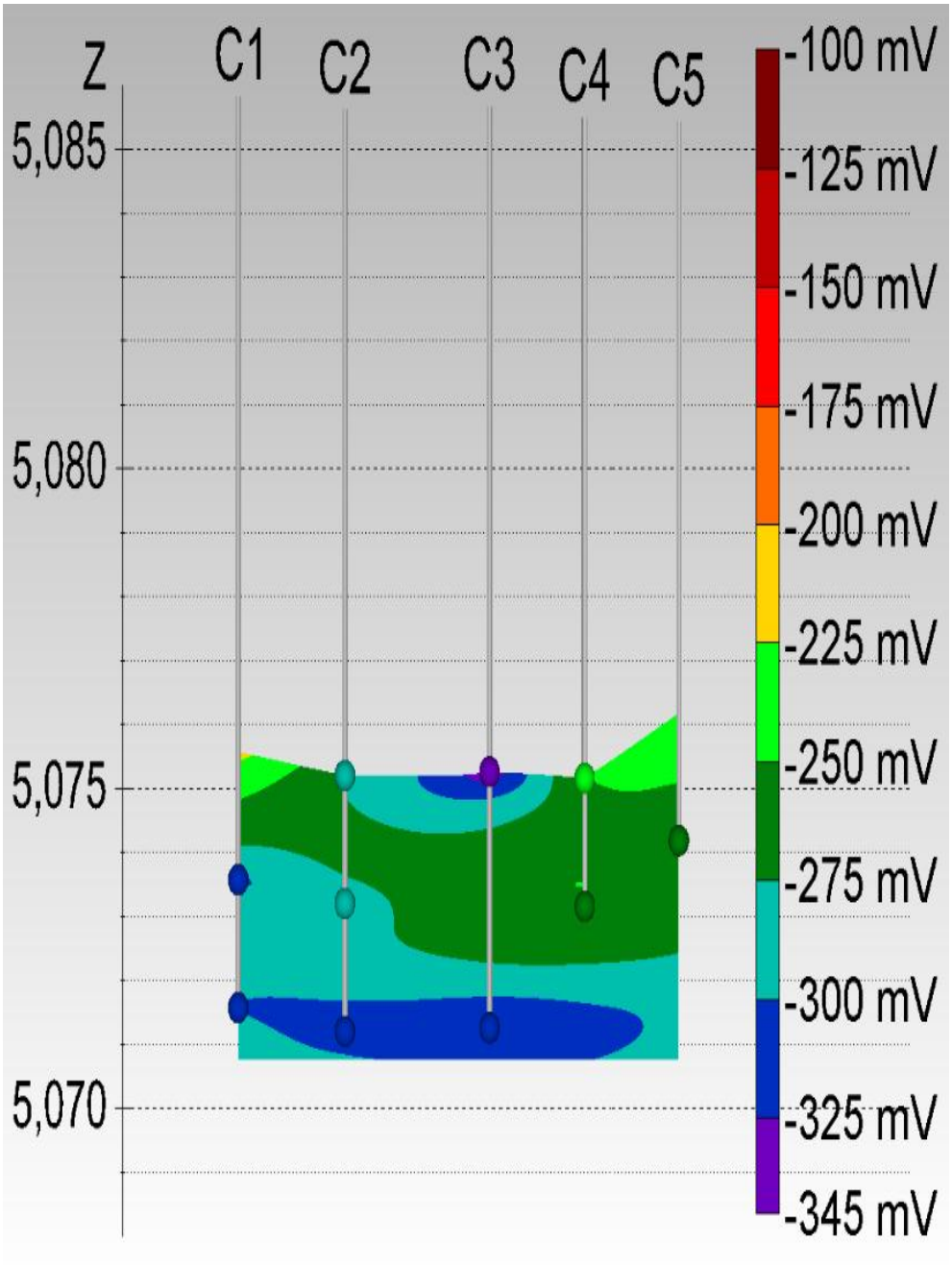


Figure C.4:ORP values (mV against 3M KCl) measured along transect C.

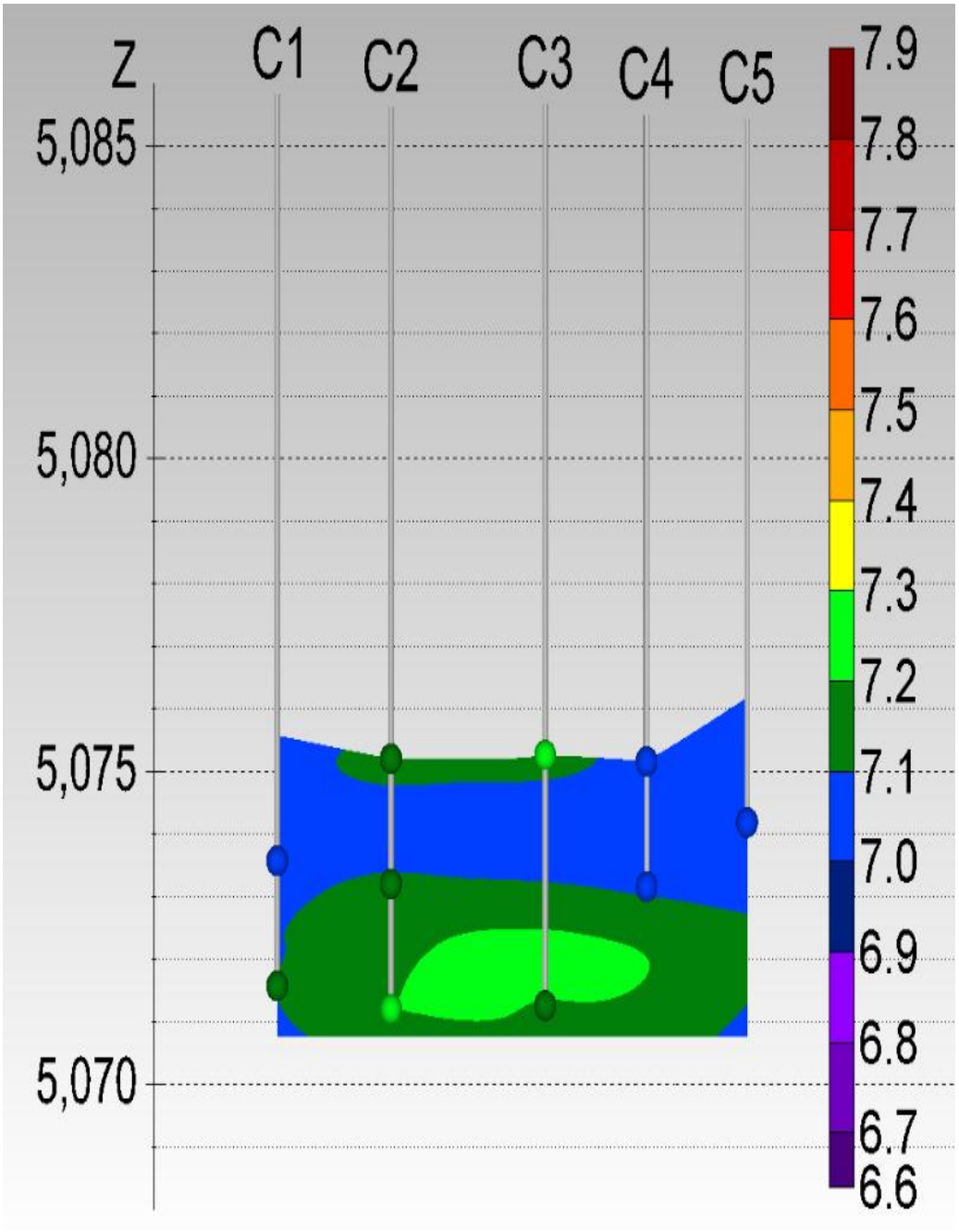


Figure C.5: pH values measured along transect C.

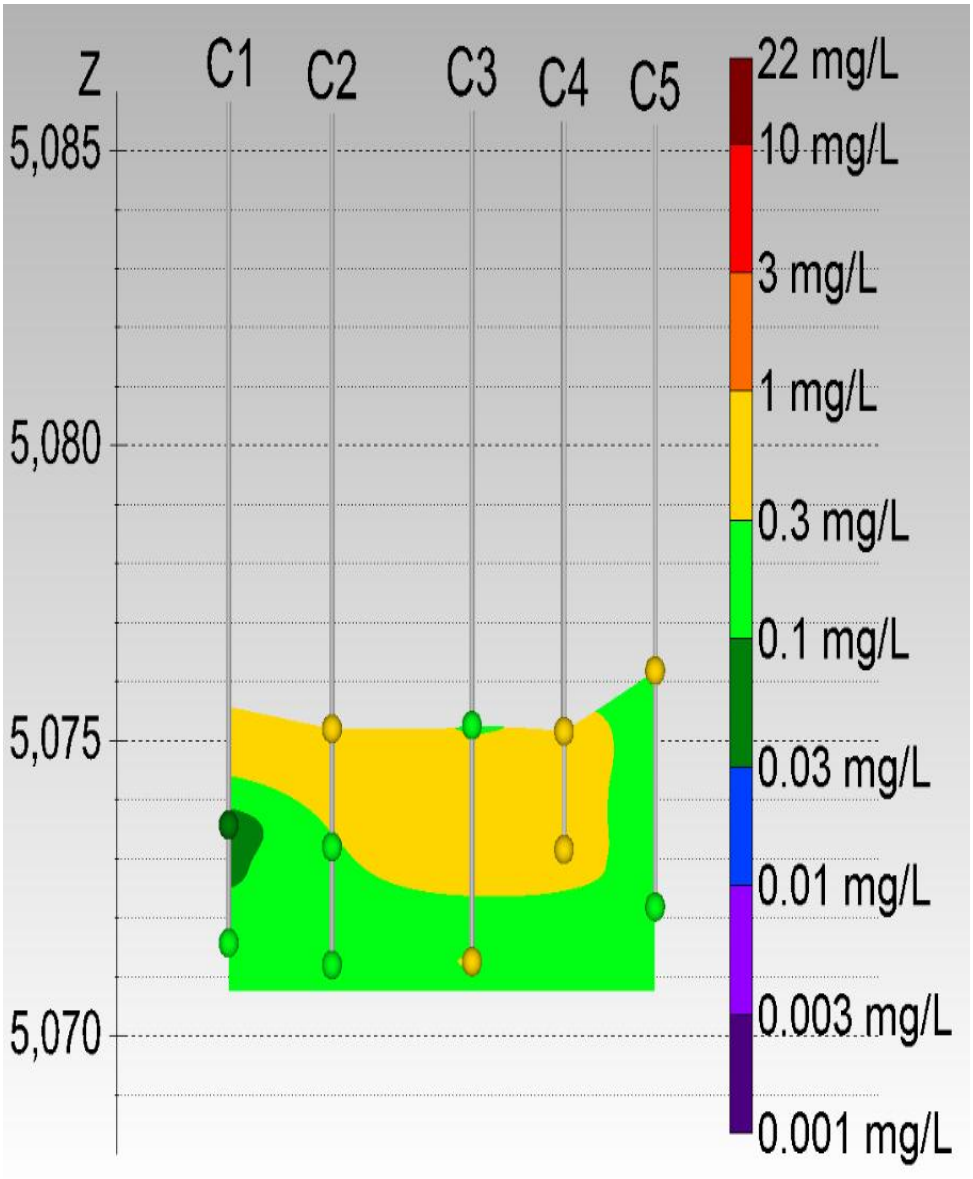


Figure C.6: Nitrate (mg/L) aqueous concentrations measured along transect C.

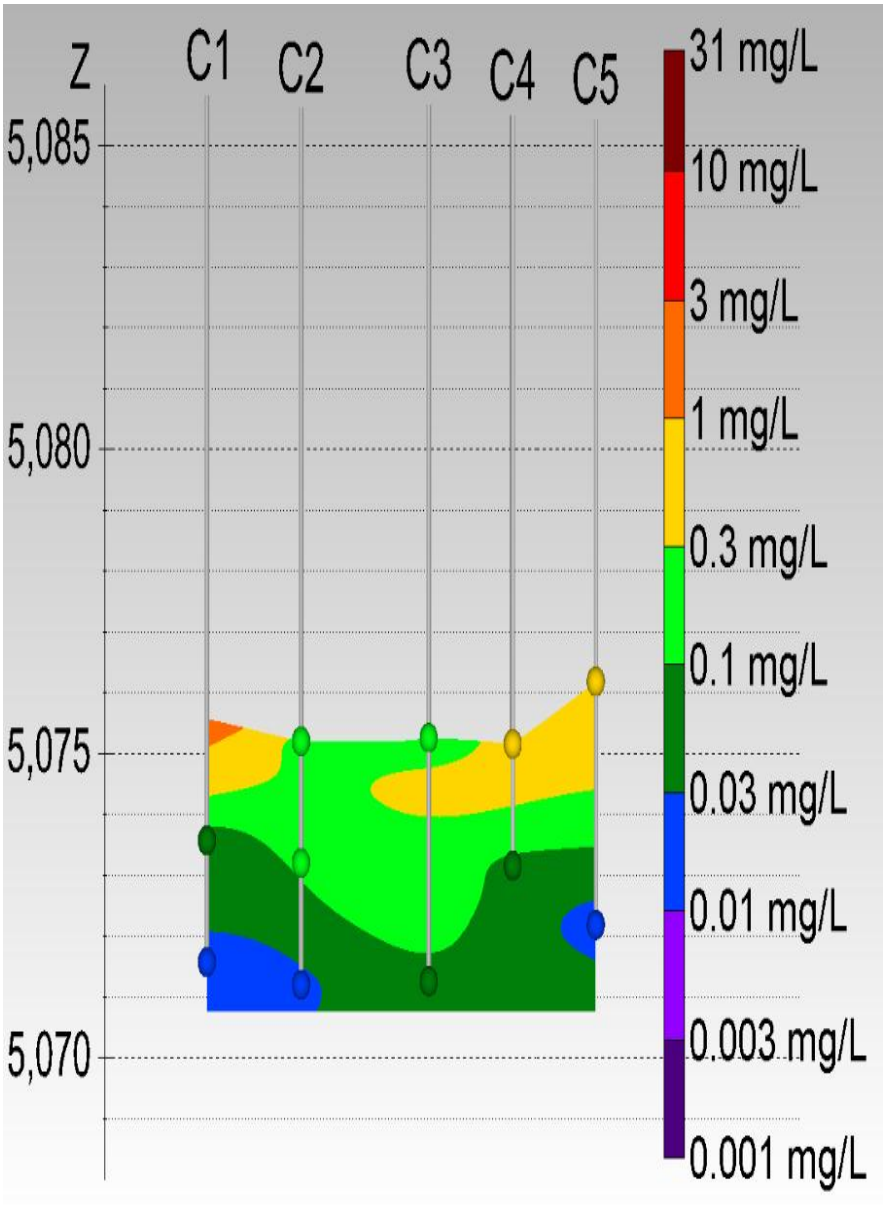


Figure C.7: Total iron (mg/L) aqueous concentrations measured along transect C.

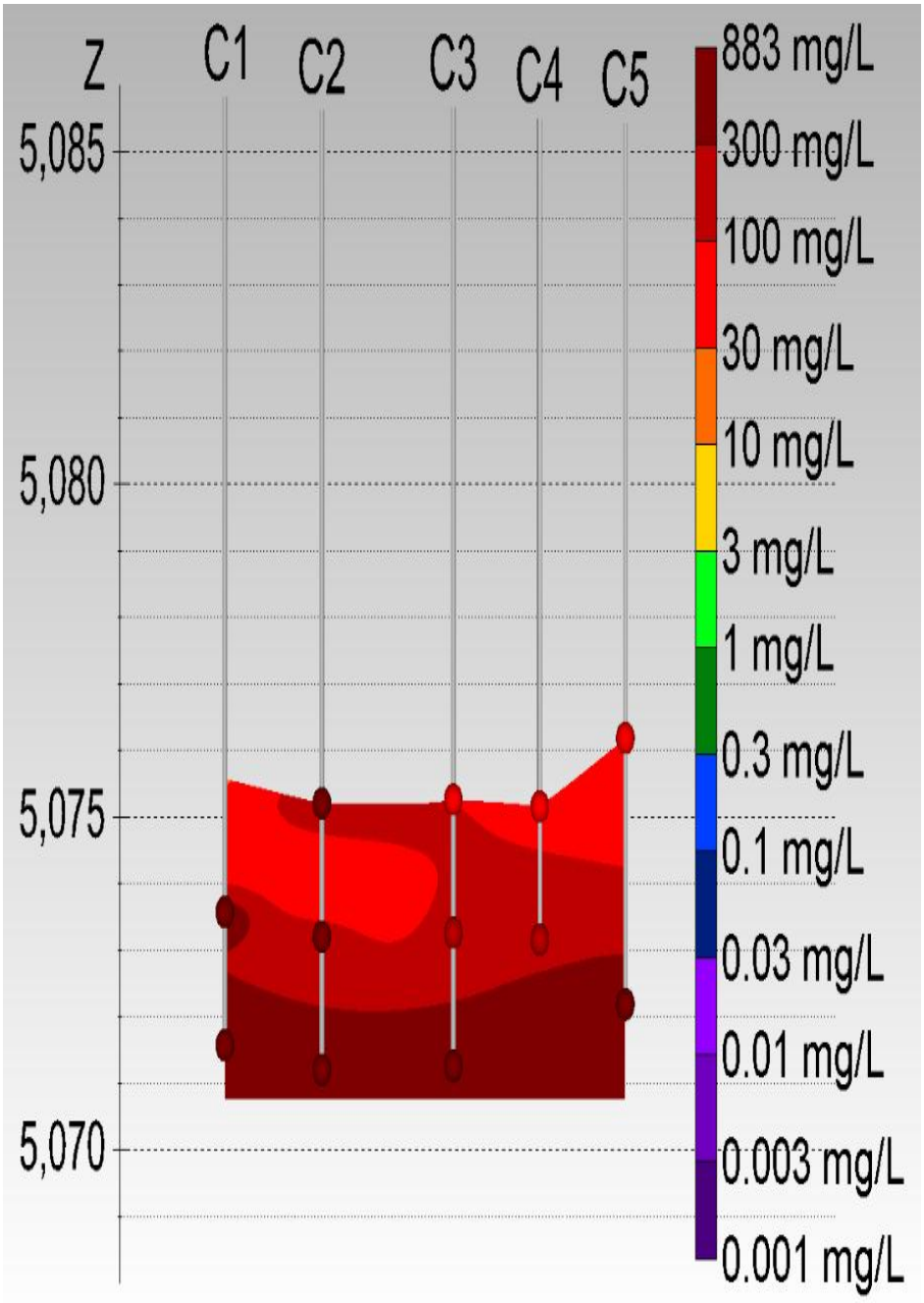


Figure C.8: Sulfate (mg/L) aqueous concentrations measured along transect C.

Appendix D: 454 Pyrosequencing Analysis Results

Table D.1: Results of 454 pyrosequencing analysis for Zone I samples.

	Eubacteria			Archaea		
Sample depth	**Identified OTUs (Genus) <5%**	**Relative abundance in Sample (%)**	**Shannon diversity index: H'**	**Identified OTUs (Genus) <5%**	**Relative abundance in Sample (%)**	**Shannon diversity index: H'**
1 ft bgs	Sulfate reducers	0.11	1.96	*Nitrososphaera sp*	71.78	0.73
5,084.7 elev. ft	Methane oxidizers	0.09		*Candidatus Nitrososphaera sp*	24.57	
	Iron Reducers	0.13		Other	3.66	
	Nitrate Reducers	1.43				
	Acidobacterium Sp.	34.05				
	Conexibacter Sp.	18.61				
	Cholstridrium sp	2.94				
	Anoxybacillus sp.	3.62				
	Bacillus sp.	10.90				
	Other	28.12				
1.5 ft bgs	Sulfate reducers	0.47	2.68	*Candidatus Nitrosopumilus sp.*	5.40	0.78
5,084.2 elev ft	Methane oxidizers	0.03		*Nitrososphaera sp.*	73.97	
	Iron Reducers	0.16		*Candidatus Nitrososphaera sp.*	17.70	
	Nitrate Reducers	2.58		Other	2.93	
	Acidobacterium Sp.	63.35				
	Holophaga sp.	5.63				
	Other	27.78				
3ft bgs	Sulfate reducers	2.73	2.2	*Methanobrevibacter sp.*	30.39	1.85
5,082.7 elev ft	Methane oxidizers	0.17		*Methanocella sp.*	5.54	
	Iron Reducers	2.70		*Methanosaeta sp.*	17.59	
	Nitrate Reducers	0.49		*Methanolinea sp.*	7.79	
	Hydrogenophilus sp.	26.36		*Methanosphaera sp.*	24.91	
	Lysobacter sp.	39.86		*Methermicoccus sp.*	7.73	
	Thiobacillus sp.	12.51		Other	6.06	
	Other	15.17				

Table D.2: Results of 454 pyrosequencing analysis for Zone II samples.

	Eubacteria			Archaea		
Sample depth	**Identified OTUs (Genus) <5%**	**Relative abundance in Sample (%)**	**Shannon diversity index: H'**	**Identified OTUs (Genus) <5%**	**Relative abundance in Sample (%)**	**Shannon diversity index: H'**
3.5 ft bgs	Sulfate reducers	5.25	3.37	*Methanobrevibacter sp.*	40.64	1.57
5,082.2 elev ft	Methane oxidizers	3.71		*Methanosaeta sp.*	6.20	
	Iron Reducers	3.65		*Methanosphaera sp.*	10.65	
	Nitrate Reducers	7.34		*Candidatus Nitrososphaera sp.*	32.32	
	Bacteroides sp	18.32		Other	10.19	
	Bacillus sp.	12.94				
	Acidobacterium Sp.	6.22				
	Koribacter sp	6.90				
	Other	35.67				
4.5 ft bgs	Sulfate reducers	8.69	3.93	*Methanobrevibacter sp.*	68.82	0.96
5,081.2 elev ft	Methane oxidizers	0.00		*Methanosaeta sp.*	9.95	
	Iron Reducers	5.68		*Methanosphaera sp.*	17.20	
	Nitrate Reducers	3.89		Other	4.03	
	Bacteroides Sp.	5.36				
	Chlostridium Sp.	7.80				
	Thermoanaerobacterium sp.	31.39				
	Thermoanaerobacter sp	5.15				
	Other	32.05				
5.5 ft bgs	Sulfate reducers	27.95	3.28	*Methanobrevibacter sp.*	7.32	1.17
5,081.2 elev ft	Methane oxidizers	0.37		*Methanosaeta sp.*	65.73	
	Iron Reducers	4.88		*Methanosphaera sp.*	6.36	
	Nitrate Reducers	1.71		*Candidatus Nitrososphaera sp.*	15.27	
	Synergistes Sp.	5.65		Other	5.32	
	Clostridium (Erysipelotrichaceae)	8.44				
	Clostridium sp	5.34				
	Other	45.67				

Table D.3: Results of 454 pyrosequencing analysis for Zone III samples.

	Eubacteria			Archaea		
Sample depth	**Identified OTUs (Genus) <5%**	**Relative abundance in Sample (%)**	**Shannon diversity index: H'**	**Identified OTUs (Genus) <5%**	**Relative abundance in Sample (%)**	**Shannon diversity index: H'**
6.5 ft bgs	Sulfate reducers	11.06	3.38	*Methanoculleus sp.*	25.65	0.76
5,080.2 elev ft	Methane oxidizers	0.54		*Methanosaeta sp.*	70.68	
	Iron Reducers	2.23		Other	3.67	
	Nitrate Reducers	1.09				
	Synergistes Sp.	7.77				
	Synthropus Sp.	22.91				
	Thermoanaerobacter sp	6.28				
	Acinetobacter sp.	5.58				
	Other	42.53				
8.5 ft bgs	Sulfate reducers	8.39	1.98	Methanoculleus sp.	4.27	0.28
5,077.2 elev ft	Methane oxidizers	0.42		Methanosaeta sp.	93.96	
	Iron Reducers	1.32		Other	1.78	
	Nitrate Reducers	0.71				
	Synergistes Sp.	2.35				
	Synthropus Sp.	63.57				
	other	23.23				
10.5 ft bgs	Sulfate reducers	4.43	1.78	*Methanoculleus sp.*	8.30	0.46
5,075.2 elev ft	Methane oxidizers	0.26		*Methanosaeta sp.*	88.61	
	Iron Reducers	0.81		Other	3.09	
	Nitrate Reducers	1.28				
	Synthropus Sp.	70.23				
	other	23.00				
11.5 ft bgs	Sulfate reducers	11.06	2.41	*Methanoculleus sp.*	46.13	0.79
5,074.2 elev ft	Methane oxidizers	0.16		*Methanosaeta sp.*	52.30	
	Iron Reducers	1.03		Other	1.57	
	Nitrate Reducers	1.42				
	Synthropus Sp.	55.79				
	other	30.54				

Table D.4: Results of 454 pyrosequencing analysis for Zone IV samples.

	Eubacteria			Archaea		
Sample depth	**Identified OTUs (Genus) <5%**	**Relative abundance in Sample (%)**	**Shannon diversity index: H'**	**Identified OTUs (Genus) <5%**	**Relative abundance in Sample (%)**	**Shannon diversity index: H'**
	Sulfate reducers	10.78	3.74	*Methanosaeta sp.*	95.78	0.25
12.5 ft bgs	Methane oxidizers	0.92		Other	4.22	
5,073.2 elev ft	Iron Reducers	2.02				
	Nitrate Reducers	2.81				
	Synergistes Sp.	11.49				
	Synthropus Sp.	10.88				
	Acinetobacter sp.	13.03				
	Clsotridium	5.18				
	other	42.89				
13.5 ft bgs	Sulfate reducers	8.074	3.03	*Methanoculleus sp.*	7.18	0.51
5,072.2 elev ft	Methane oxidizers	0.491		*Methanosaeta sp.*	88.22	
	Iron Reducers	1.497		Other	4.60	
	Nitrate Reducers	2.650				
	Synergistes Sp.	5.742				
	Synthropus Sp.	43.436				
	Acetinobacter sp.	2.969				
	other	35.141				

Appendix E: 454 Pyrosequencing data analysis protocol provided by Research and Testing Laboratory (Lubbock, TX)

Term Definitions

Terms used within this guide are defined as follows:

Tag

o The term tag refers to the 8-10 bp sequence at the 5' end of the sequence read.

o The tag is also known as the barcode in some programs.

Identity Percentage

o Identity percentage for a read is defined as the length of the HSP Identity divided by the length of the hit HSP.

o BLAST is run to return 5 hits and 1 HSP per hit.

HSP – High-Score Pair

o The highest scoring local alignment between a sequence read and the database sequence such that the score cannot be improved by extension or trimming of the alignment.

HSP coverage

o The HSP length divided by the query length, giving the percentage of the query sequence covered by the HSP.

Data Analysis Methodology

Overview of the Data Analysis Process

Once sequencing has completed, the data analysis pipeline will begin processing the data. The data analysis process consists of two major stages, the quality checking and reads denoising stage and the diversity analysis stage. During the read quality checking and denoising stage, denoising and chimera checking is performed on all the reads for each region of data. Then each remaining read is quality scanned to remove poor reads from each sample. The primary output of this stage is a quality checked and denoised FASTA formatted sequence, quality, and mapping file. This stage is performed for all customers whose data we know the encoded tags for. During the diversity analysis stage, each sample is run through our analysis pipeline to determine the taxanomic information for each read and then analyzed to provide the sample's microbial diversity. The output for this stage is a set of files detailing the taxanomic information for each read as well as the number and percentage of each species found within each sample. This stage is performed for all customers whose data is sequenced using primers based within the 16S, 18S, 23S, ITS and SSU regions.

The data analysis pipeline is broken down into the following steps, each of which is discussed more thoroughly in the sections below:

Quality Checking and Denoising

1. Denoising and Chimera Checking

2. SFF File Generation

3. Quality Checking and FASTA Formatted Sequence/Quality File Generation

- Microbial Diversity Analysis

1. Taxanomic Identification

2. Data Analysis

Denoising and Chimera Checking

Denoising The process of denoising is used to correct errors in reads from next-generation sequencing technologies including the Roche 454 technologies. According to the paper “Accuracy and quality of massively parallel DNA pyrosequencing” by Susan Huse, et al. and “Removing noise from pyrosequenced amplicons” by Christopher Quince, et al. the per base error rates from 454 pyrosequencing attain an accuracy rate of 99.5% [1] [2]. However, the large read numbers that the machine can generate mean that the total number of noisy reads can be substantial. In order to determine true diversity it becomes critical to determine which reads are good and which reads contain noise introduced by the experimental procedure. The Research and Testing Laboratory analysis pipeline attempts to correct this issue by denoising entire regions of data prior to performing any other steps of the pipeline.

The Research and Testing analysis pipeline performs denoising by performing the following steps on each region:

- Reads within the data set are sorted from longest read to shortest read.

- Using USEARCH [3], reads are then dereplicated meaning they are clustered together into groups such that each sequence is an exact match to portion of the seed sequence for the cluster. Each cluster is marked with the total number of member sequences.
- The seed sequence from each cluster is then sorted by abundance, largest cluster to smallest cluster. Keep in mind that no minimum size restrictions exist on the clusters, thus single member clusters will exist.
- Clustering at a 1% divergence using the USEARCH [3] application is performed on the seed sequences in order to determine similar clusters. The result of this stage is the consensus sequence from each new cluster, each tagged to show their total number of member sequences (dereplicated + clustered).
- The consensus sequences are re-sorted based upon their abundance, largest cluster to smallest cluster. However at this point a minimum size restriction if put into place and any cluster that does not contain at least two member sequences is removed from consideration.
- Clustering at a 5% divergence is once again performed using USEARCH [3] on the consensus sequences in order to determine if the consensus sequences have created additional clusters. The result of this stage is the seed sequence from each new cluster,

Once this process has completed we have removed all reads that failed to have a similar or exact match elsewhere on the region and through the use of consensus sequences we have helped correct base pair errors generated during sequencing.

Chimera Checking

As discussed in the paper “Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons” by Brian Haas, et al. the formation of chimeric sequences occurs when an aborted sequence extension is misidentified as a primer and is extended upon incorrectly in subsequent PCR cycles.Because amplification produces chimeric sequences that stem from the combination of two or more original sequences [4], we will perform chimera detection using the *de novo* method built into UCHIIME.

The Research and Testing analysis pipeline performs chimera detection and removal by executing UCHIIME in *de novo* mode on the clustered data that was output by our denoising methods. By using this method we can determine chimeras across entire region of data even after accounting for noise and removing low quality sequences.

SFF File Generation

SFF files are a binary file containing many data about a read in a single file. For each read, the sff contains a flowgram, quality score and sequence with defined lengths from QC measures performed by the machine. The sff represents the raw data and includes many reads that may have been excluded due to length or chimera detection or any other filter requested for custom processing. Since the files are binary, they cannot be opened with standard text editors. Special programs like Mothur [5] or BioPython [6] are able to load their data into human readable formats and output fasta, qual, flowgram or text (sff.txt) versions. Sff files or their derivatives can then be used for further processing of the data. Sff files provided may be of two forms. In the case of an entire region containing a single investigator’s samples, the entire region plus mapping file is provided. In cases where multiple investigators had samples on a single region, each sample is demultiplexed from the sff file using the Roche sffinfo tool by

providing its barcode, effectively eliminating it from any read extracted. The split sff can then be used for raw data or submitted directly to archives like the NCBI's SRA. In cases where a single sff for all samples is desired but an entire quadrant is not used, an investigator may request a single sff for a nominal charge. Alternatively, it is possible to use the provided split sff files for denoising/chimera removal by modifying the mapping files. Additional instructions are available if you wish to do so.

Quality Checking and FASTA Formatted Sequence/Quality File Generation The denoised and chimera checked reads generated during sequencing are condensed into a single FASTA formatted file such that each read contains a one line descriptor and one to many lines of sequence/quality scores. The Research and Testing Laboratory analysis pipeline takes the FASTA formatted sequence and quality files and removes all sequences that meet the following quality control requirements:

1. Failed sequence reads,
2. Sequences that have low quality tags, primers, or ends and
3. Sequences that fail to be at least 250 bp in length.

Sequences that pass the quality control screening are condensed into a single FASTA formatted sequence and quality file such that each read has a one line descriptor followed by a single line of sequence/quality data. The descriptor line in both files has been altered to contain the samples name followed by the original descriptor line, separated with a unique delimiter (::).

This stage of the pipeline creates the FASTA reads archive which contains the following files:

1. The sequence reads from all samples concatenated into a single sequence file. The original tags have been removed from each sequence and an “artificial tag” has been added in its place. The title of the file will be <name>.fas.

2. The quality scores from all samples concatenated into a single quality file. The scores are labeled with the corresponding sample name and will have a matching line in the .fas file. Since the original tags were removed from the sequence and an “artificial tag” was put into its place, the quality scores have been similarly altered such that the original scores for the tag have been removed and an “artificial quality tag” has been added in its place. The artificial quality tag consists of Q30s for the length of the tag. This file will be labeled <name>.qual.

3. A mapping file consisting of sample names included in the analysis. This file contains the information for each sample such that each line has the sample name, tag and primer used for the sample. This file will be labeled as: <name>.txt

Taxonomic Identification

In order to determine the identity of each remaining sequence, the sequences will first be sorted such that the FASTA formatted file will contain reads from longest to shortest. These sequences are then clustered into OTU clusters with 100% identity (0% divergence) using USEARCH [3]. For each cluster the seed sequence will be put into a FASTA formatted sequence file. This file is then queried against a database of high quality sequences derived from NCBI using a distributed .NET algorithm that utilizes BLASTN+ (KrakenBLAST www.krakenblast.com).

Using a .NET and C# analysis pipeline the resulting BLASTN+ outputs were compiled and data reduction analysis performed as described previously .

Based upon the above BLASTn+ derived sequence identity percentage the sequences were classified at the appropriate taxonomic levels based upon the following criteria. Sequences with identity scores, to well characterized 16S rRNA gene sequences, greater than 97% identity (<3% divergence) were resolved at the species level, between 95% and 97% at the genus level, between 90% and 95% at the family and between 85% and 90% at the order level , 80 and 85% at the class and 77% to 80% at phyla. Any match below this percent identity is discarded. In addition, the HSP must be at least 75% of the query sequence or it will be discarded, regardless of identity.

After resolving based upon these parameters, the percentage of each organism will be individually analyzed for each sample providing relative abundance information within and among the individual samples based upon relative numbers of reads within each. Evaluations presented at each taxonomic level, including percentage compilations represent all sequences resolved to their primary identification or their closest relative

Analysis description

These folders contain the actual result files from analysis. The folder contains the results for blasting the seed reads against the appropriate Research and Testing Laboratory database. It also contains a seqs_otu_table.txt file showing the predicted OTUS (defined by unique species) with the confidence interval used to give the level of classification for it. Also present is the seqs_otus.txt file which gives the clusters formed by USEARCH. The seed read used for a cluster is the first listed and is the longest sequence of the cluster.

The .csv files are in a comma separated format and can be opened with Excel or another text editor program. Each file can be dragged and dropped into Excel or you may choose to right click on the file name, select “Open With” and choose Excel as the program. Each file contains information about all the samples. Sample names span the first row with the bacterial/fungal designations at each respective taxonomic level are listed in the first column. Counts (or Percentages- if looking at the …Percent.csv file) of each of the respective taxonomic levels found within the sample are listed below the sample name.

Generally, the most relevant files are the Percent composition files, although the PercentTraceback can be used when confidence intervals are desired. The files include composition information forced to the top blast hit at a specific taxonomic level. For example, in the species files, the nearest well described species for each sequence is listed; similarly, the genus files contain the genus of the species as catalogued in the database, and so on for each taxonomic level. Research and Testing Laboratory uses 7 taxonomic levels for each organism: Kingdom, Phylum, Class, Order, Family, Genus, Species.

The information is organized by taxonomic level (each file specifies for which taxonomic level the information included is for). Files names include:

<name>Kingdom

<name>BelowMinimum

<name>Excluded

<name><taxa level>Counts

<name><taxa level>Percent

<name>SpeciesOptions

<name>SpeciesFullTaxa

<name>SpeciesTraceback

<name>SpeciesPercentTraceback

File descriptions:

<name>Kingdom.csv with the following columns followed by samples:

1. Query sequence name (with sample label) and its cluster information

2. The hit name with "-I" followed by the identity percentage and "Q" followed by the query length,

3. Identity column, indicating the identity percentage of the query to the hit sequence along the HSP (High Scoring Pair) region.

4. The count for the hit, based on the number of cluster members. If clustering is not performed, this is always 1.

<name>BelowMinimum.csv

Contains all hits from the blast results that fell below 77% identity OR had HSP coverage below 70%. These are given a classification at the closest species, but do not appear in any other file. These hits did not have sufficient similarity to any reference sample to have confidence in assigning to an organism. If a sample had all its reads fall into this file, all values in the Counts file will show as 0 and all the values in the Percent files will show as NaN for the sample as a result of division by zero.

<name>Excluded.csv

Contains hits against organisms that have been requested to be excluded from results by the sender. By default, this is always empty. Some items may be dropped from the blastout and will not appear in the excluded file. This currently includes plastid and mitochondrial sequences that may be present in the database.

Additionally, some reads generated may be present in none of the above files if they were not sufficiently similar to any of the reference sequences in the database. This may occur due to a spurious amplification or poor quality read. These reads are considered to be further noise in the data.

Each of the taxa levels contains 2 files

<name><taxa level>Counts.csv

Counts are merged on the organism term, with separation given to those with top hits among different species, condensing all identity scores and query lengths. A set of hits by a read against multiple organisms but with identical similarity have each organism listed, separated by "::". A read may have up to 5 hits, but only the best are used for determination. If a non-specific species is found as the top hit, a similar quality hit on a full species is set as the top hit for the Percent computations. Each unique set of terms has a single line with summed counts for all of its samples.

<name><taxa level>Percent.csv

Converts the previous file of raw read counts into percentages of composition per sample per organism, for the top hit only. Multiple organism hits are converted to top hit only for merging with other reads with similar top hits, except those with generic species as the top hit, which are changed to the next hit with a specific name. Some entries here may appear as NaN,

indicating that the count was zero as a result of discarding all hits for the sample due to percent identity or filtering.

Additional species files <name>SpeciesOptions.csv

Shows the percent file without discarding any top hits and calculating percents based on this. This shows what percent of the reads had only a single similar species based on sequence or multiple similarities.

<name>SpeciesFullTaxa.csv

Shows the Percent file with the full taxa of every organism listed. This way, multiple files do not need to be examined to determine the catalogued taxonomy used in the groupings

Due to incomplete taxonomic data for some organisms, the top hits may have unusual naming conventions. Some entries have no specific name, so they are abbreviated <Genus> sp. If a sample from the database had missing taxonomic data at the class level e.g. *Cyanobacteria*, then the class of the organism would be *Cyanobacteria (class)*. Others may be named with an "unclassified" e.g. *Clostridiaceae unclassified.* This occurs in many places, but the naming convention remains the same for any taxonomic level. Those without at least a defined class are not added to the database for bacteria and fungi. Also seen are species like *Pseudomonas sp Rhizobiaceae* or a genus like *Pseudomonas (Rhizobaceae)*. This is a generic Pseudomas species belonging to the Rhizobiaceae family. Since there are multiple entries for Pseudomonas sp in the NCBI Taxonomy, the suffix has been added to distinguish at each level where there may be confusion over which species it is, if only the usual nomenclature had been used.

Also included in the **Percent** folder is a "traceback" file, *SpeciesPercentTraceback*. This file reflects taxonomic information for organisms found in the sample based on the goodness of

the alignment against reference sequences. This file is similar to the previously described percent file, however, the nearest neighbor isn't forced, but instead based on the percent identity of the query sequence to the reference sequence; the "most certain" taxonomic level is listed. A summary table of the level which is used for identification based on identity score is provided below.

Identity to reference sequence	**Traceback Designation**
I > 97%	
97% ≥ I > 95%	(unk species)
95% ≥ I > 90%	(unk genus)
90% ≥ I > 85%	(unk family)
85% ≥ I > 80%	(unk order)
80% ≥ I > 77%	(unk class)