MICROBIAL ECOLOGY IN ENVIRONMENTS IMPACTED BY HYDRAULIC FRACTURING FLOWBACK AND PRODUCED WATER

By

Cheng Zhong

A thesis summitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Earth and Atmospheric Sciences

University of Alberta

Abstract

The increased development of unconventional resources recovery has dramatically changed the global energy landscape over the past two decades. Hydraulic fracturing combined with horizontal drilling, a key technology in recovering these resources, requires large freshwater volumes to fracture the low-permeability shale formations that host the hydrocarbons. Biofouling, biocorrosion, and biodegradation, caused by persistent microbial communities in the fractured subsurface, may influence the efficiency and costs of oil and gas recovery. Following fracturing, return fluid, referred to here as flowback and produced water (FPW), returns to the surface. FPW may cause contamination of the environment when surface spills occur. Microbial communities are basic units of ecosystems and drive virtually all biogeochemical cycling in the environment. However, knowledge of the effects of FPW on freshwater and soil microbial communities is limited. This thesis aims to answer scientific questions about the influence of FPW to soil and water microbial ecology, in order to enhance our understanding of the impacts of microbial communities to downhole production and microbial ecology in environments impacted by FPW.

Chapter One, the introduction, gives a general background of the hydraulic fracturing water cycle of unconventional oil and gas recovery. I introduce the major environmental and production issues related to FPW, the importance of understanding the microbial ecology for optimizing these environmental and production issues, and highlights the current knowledge gaps.

Chapter Two couples 16S rRNA gene sequencing and live/dead cell viability to assess freshwater community changes in simulated FPW spills by volume percent from 0.05% to 50%. In this study, three distinct patterns of microbial community shifts were identified: (1) indigenous freshwater genera remained dominant, (2) potential degraders of organic compounds in FPW

became dominant, and (3) no significant change in the relative abundance of taxa was observed. Live cells were quickly killed by exposure to 10% FPW and above, but cell counts recovered in the following days. The findings demonstrated that microbial taxa are effective fingerprints for detecting FPW derived pollution sources and severity.

In **Chapter Three**, I conducted metagenomics and respiration analyses on luvisol and chernozem soil microbial communities exposed to a series of diluted FPW samples. Our study showed that the luvisol soil was more vulnerable than the chernozem soil in terms of loss of biodiversity and lower respiration activity. However, with increasing FPW exposure concentrations, differences in incubation trajectories (e.g., respiration activity) converged as the microbial communities increasingly shifted towards *Marinobacter* spp. and had increased gene abundance related to degradation and tolerance of FPW-derived chemicals in the luvisol. In contrast, shifts in community composition and functional capacity were not evident in the chernozem soil. This study provides the first evidence that soil microbiota can shift their community structure and functional capacity to maintain soil functions under high FPW-induced stress.

Chapter Four is an investigation of genome consistency in FPW across unconventional oil and gas formations in Canada, China, and the United States. This study revealed two taxonomically and functionally distinct fractured shale microbial communities: a low diversity community in higher salinity produced fluids of the North American Basins dominated by halophiles, and a higher diversity community in the lower salinity produced fluids of the Sichuan Basin of China. Fermentation, sulfidogenesis, and methanogenesis are core functions that are conserved across all basins, while microbial communities in the FPW derived from the Sichuan Basin were shown to have more diverse metabolic pathways that potentially lead to sulfide

generation and methanogenesis. Our study suggests that despite dramatic salinity differences and microbial communities between different shales, microbial processes important for elemental cycling in the subsurface are similar globally.

Chapter Five summarizes the main findings of my dissertation of three independent studies into concise conclusions, identifies research limitations and gaps, and suggests future directions for research.

Preface

This dissertation consists of three research articles and a review article, is the cumulative efforts of my Ph. D. research between 2017 to 2020 focused on understanding the microbial ecology of energy recovery and environmental questions closely related to unconventional hydraulic fracturing FPW. I was involved in all aspects of the research from designing of the project to data collection, analysis and interpretation, and manuscript drafting. All the research projects were supervised by Dr. Daniel S. Alessi and Dr. Brian D. Lanoil.

Chapter One has been modified from a critical review manuscript, which has been submitted to Environmental Science & Technology upon invitation after submitting a successful proposal to the journal editors, with coauthors Deyi Hou, Ashkan Zolfaghari, Greg G. Goss, Brian D. Lanoil, Joel Gehman, Daniel C. W. Tsang, Daniel S. Alessi. I wrote the paper and the coauthors provided comprehensive support for the editing of the article.

Chapter Two has been published in FEMS Microbiology Ecology, Volume 96, Issue 5, May 2020, fiaa068, titled "Response of aquatic microbial communities and bioindicator modelling of hydraulic fracturing flowback and produced water" with coauthors Camilla L. Nesbø, Greg G. Goss, Brian D. Lanoil, and Daniel S. Alessi. Camilla L. Nesbø provided support on the bioinformatics data processing and interpretation. Greg G. Goss, Brian D. Lanoil, and Daniel S. Alessi supported the experimental costs, design of experiments, facilitated equipment use, and assisted with data interpretation and editing of the manuscript.

Chapter Three is a manuscript in preparation, with coauthors, Konstantin von Gunten, Camilla L. Nesbø, Yifeng Zhang, Xiaoqing Shao, Rong Jin, Kurt O. Konhauser, Greg G. Goss,

Brian D. Lanoil, Daniel S. Alessi. Konstantin von Gunten helped with the field sampling of soils in Fox Creek and Grand Prairie, Alberta. Camilla L. Nesbø provided support on bioinformatics data processing and interpretation. Yifeng Zhang supported the chemical analyses, and Rong Jin and Xiaoqing Shao helped with DNA extraction and polymerase chain reaction. Kurt O. Konhauser, Greg G. Goss, Brian D. Lanoil, and Daniel S. Alessi supported the experimental costs, design of experiments, facilitated equipment use, and assisted with data interpretation and editing of the manuscript.

Chapter Four is a manuscript in preparation with coauthors, Mikayla A. Borton, Camilla L. Nesbø, Fu Chen, Malcolm D. Forster, Liaozi Han, Greg G. Goss, Kelly C. Wrighton, Brian D. Lanoil, Daniel S. Alessi. Fu Chen and Liaozi Han provided sequencing and chemistry data for hydraulic fracturing FPW from China. Mikayla A. Borton and Kelly C. Wrighton provided sequencing and chemistry data for hydraulic fracturing FPW from the United States. Camilla L. Nesbø provided support on bioinformatics data processing and interpretation. Malcolm Forster helped to refine the genome bins assembled by sequencing reads. Greg G. Goss, Brian D. Lanoil, and Daniel S. Alessi provided supported the experimental costs, design of experiments, facilitated equipment use, and assisted with data interpretation and editing of the manuscript.

Acknowledgment

I am sincerely grateful to my supervisor Dr. Daniel S. Alessi, for his help in my Ph.D. study. During my Ph.D. study, Dan always provided his patient and professional guidance on my research, particularly for the geochemistry and writing strategies. Besides, he always encouraged and supported me to attend international conferences and activities worldwide. Additionally, his optimistic attitudes and unbelievable drinking capacity promoted the interests of my Ph.D. life.

Also, I would thank my co-supervisor, Dr. Brian D. Lanoil, for improving my knowledge of environmental microbiology and providing smart ideas and instructions on my research. I would especially like to acknowledge Dr. Camilla L. Nesbø for her kindly supports in bioinformatics throughout my Ph.D. study and Dr. Greg G. Goss for offering me his lab space and equipment and his useful advice on my research. I would thank Dr. Maya Bhatia and Dr. Suzanne Tank for their instruction on my research as my committee members. Besides, I appreciate all the supports and inspiration received by Dr. Karlis Muehlenbachs and Dr. Kurt O. Konhauser to pursue my Ph. D. degrees. I thank Dr. Deyi Hou and his environmental science group members for their kind help and support during my three-month visit to Tsinghua University.

Here, I would like to thank my friends Dr. Zhentao Shen, Dr. Konstantin von Gunten, Dr. Weiduo Hao, and Dr. Yunhui Zhang, Jin Rong, and all my friends, groupmates, and professors who have helped my research during my Ph.D. study, and accompanied me to go through the COVID-19 unusual time in 2020. Finally, but most importantly, I am grateful for having a great family who provides all their support to help me to pursue my Ph.D. degree. My instructors, friends, and family are the most important parts of my three-year Ph.D. study's unforgettable time. I cannot obtain the honor without the unwavering support from them.

Table of Contents

List of Tables

List of Figures

1. Chapter One: General Introduction [1]

1.1. Overview

Advancements in directional drilling and multi-stage hydraulic fracturing (HF) has unlocked vast hydrocarbon resources from unconventional reservoirs (Figure 1.1) (U.S. Environmental Protection Agency 2016). While the replacement of coal with natural gas (NG) from unconventional resources may improve air quality and lead to reductions in greenhouse gas emissions (Newell and Raimi 2014; Song *et al.* 2015), HF poses environmental challenges. For example, HF operations in the U.S. typically require 280 m^3 (10^{th} percentile) to 23,000 m^3 (90^{th} percentile) of fracturing fluid, consisting mostly of water sourced from surface water or groundwater, to fracture low permeability formations such as shales (U.S. Environmental Protection Agency 2016). After HF, 10% to 100% of the total injected volume (TIV) of fracturing fluid returns to the surface as flowback and produced water (FPW) (Gregory, Vidic and Dzombak 2011; Rivard *et al.* 2014; Alessi *et al.* 2017), which contains constituents of the fracturing fluid and their degradation products (Sun *et al.* 2019a), hydrocarbons indigenous to the formation, reservoir rock constituents that are dissolved/reacted with the fracturing fluid, and formation water that often has elevated salinity (U.S. Environmental Protection Agency 2016).

---

[1] This Chapter is revised from a submitted critical review manuscript to Environmental Science & Technology (following a successful peer-reviewed proposal to write the paper): Deyi Hou, Ashkan Zolfaghari, Greg G. Goss, Brian D. Lanoil, Joel Gehman, Daniel C. W. Tsang, Daniel S. Alessi

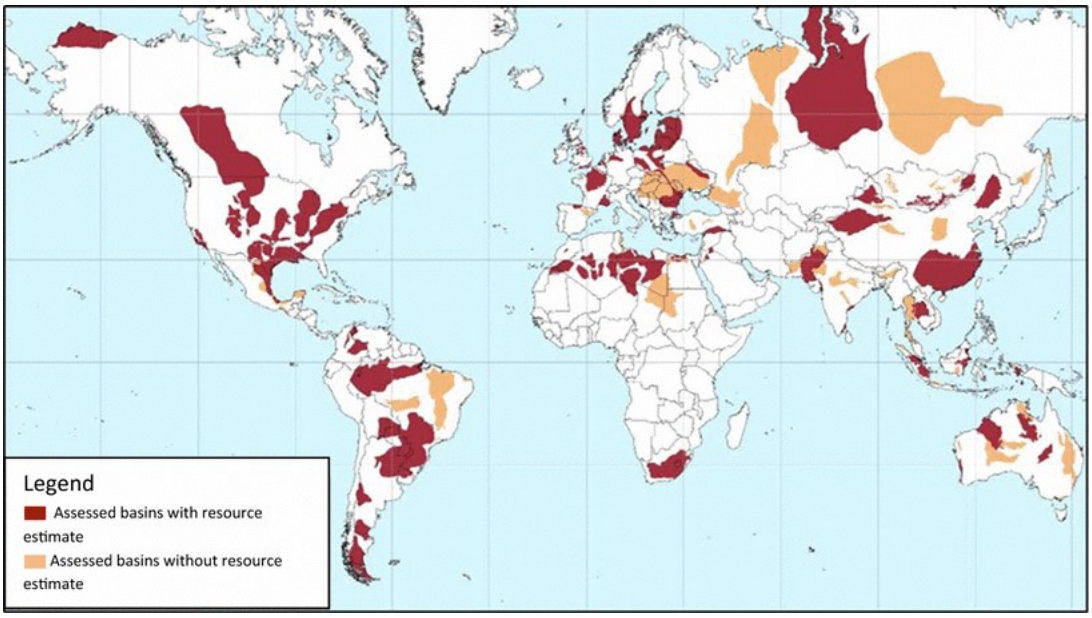

Figure 1.1 Global map of major shale basins, adopted from a previous report (U.S. Energy Information Administration 2013).

A previous report in 2016 notes that the HF water cycle includes: (1) water withdrawals to make fracturing fluids, (2) the mixing of water with proppant and chemical additives, and injection of fracturing fluids to fracture target low permeability oil and gas formations, and (3) the collection and disposal or reuse of FPW (U.S. Environmental Protection Agency 2016). Concerns related to the sustainability of the HF water cycle include over-exploitation of water resources (Hou, Luo and Al-Tabbaa 2012; Vidic *et al.* 2013; Vengosh *et al.* 2014), contamination of regional water resources and soil (Vidic *et al.* 2013; Vengosh *et al.* 2014; Folkerts, Blewett and Goss 2020), and other environmental problems such as land use (Vidic *et al.* 2013; Guo *et al.* 2020), air pollution and noise during HF operations (Allshouse *et al.* 2019), increased traffic accident rates (Graham *et al.* 2015; Allshouse *et al.* 2019), and induced-seismicity (Figure 1.2) (Atkinson and Eaton 2020). In North America, concerns related to HF have given rise to movies such as *Gasland* and *The Promised Land*, as well as periods of activism and protest (Vasi *et al.* 2015; Mazur 2016). Some

North America jurisdictions have imposed bans or moratoria on HF (Dokshin 2016; Arnold and Long 2019), and there have been concerns more generally about the industry's social license to operate (Gehman *et al.* 2016; Gehman, Lefsrud and Fast 2017). As one response, the industry introduced FracFocus, first as a voluntary effort, but now mandatory in many North America jurisdictions (Konschnik and Dayalu 2016; Avidan, Etzion and Gehman 2019).

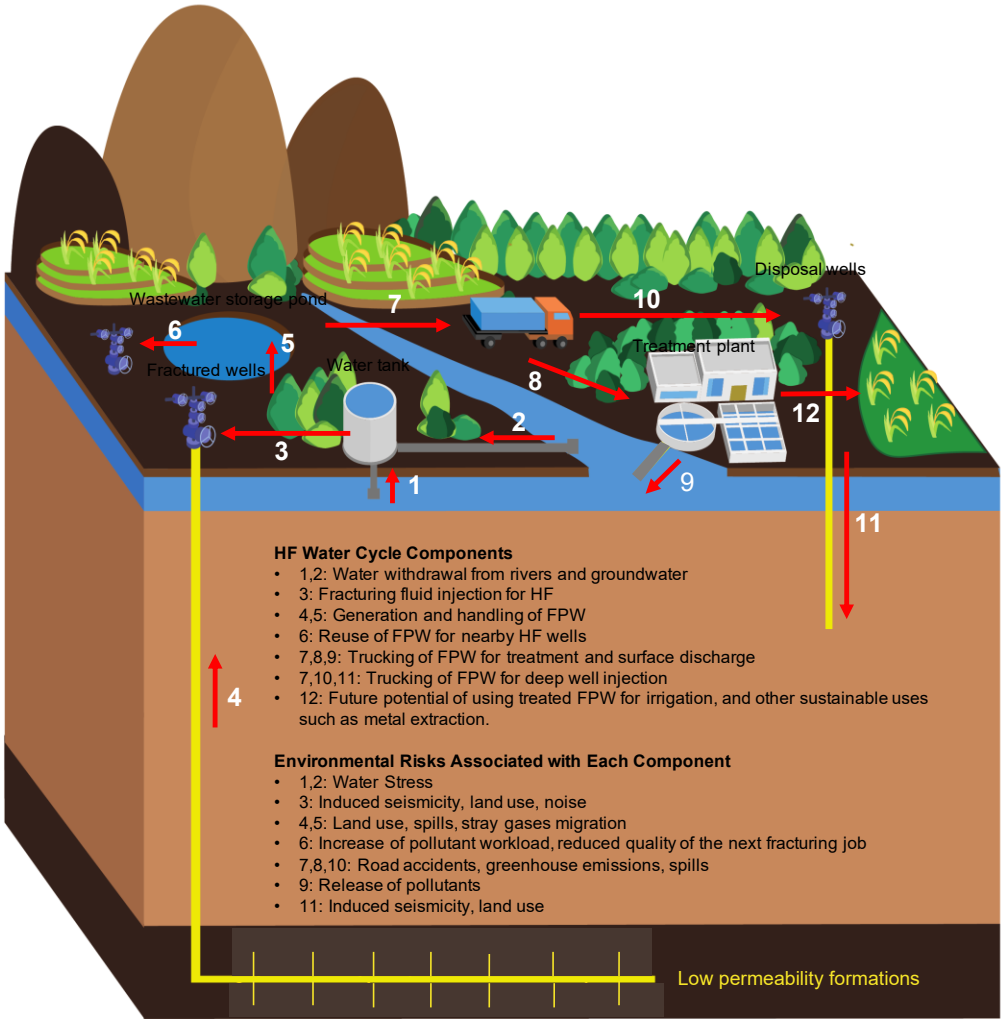

Figure 1.2 Core components of the hydraulic fracturing (HF) water cycle, including water withdrawal (1,2), chemical mixing and fracturing fluid injection (3), flowback and produced water (FPW) generation and handling (4,5), common FPW management methods (6-11), and potential

future directions (12). Major environmental risks associated with each component identified in this review are also presented.

In the research for my dissertation, I studied the microbial ecology and geochemistry of soils and water impacted by the HF water cycle, with focuses on the *in situ* microbial ecology and geochemistry in the fractured subsurface and soil and aquatic microbial ecology under FPW-induced changes in chemistry. To provide context for my dissertation research, studies on hydraulic fracturing water withdrawal and FPW generation are reviewed; chemical and microbiological characterization of FPW are reviewed; and FPW contamination and toxicity to the environment are reviewed, concluding with a discussion of FPW effects to microbial communities.

1.2. Water Consumption

1.2.1. HF Water Use Per Well

The water use per HF well in major unconventional plays in North America is between 1,300 - 23,800 m^3 (median value), including the Barnett, Eagle Ford, Fayetteville, Haynesville, Marcellus, Woodford, Niobrara, Bakken, Permian Formations (Kondash and Vengosh 2015), and between 4,000 m^3 - 10,400 m^3 (average value in Alberta and British Columbia, respectively) in the Montney Formation (Alessi *et al.* 2017). Recent evidence shows that water use per well in emerging shale gas plays such as the Weiyuan and Fuling in China (median: 30,300 - 34,000 m^3, average: 30,200 - 34,800 m^3) are higher than for North America HF wells (Zou *et al.* 2018; Shi *et al.* 2020). Consistently, higher water use per HF has been reported for other unconventional plays in in China, including the Junggar Basin (70,400 m^3), Ordos Basin (37,300 m^3), and Bohai Bay Basin (34,300 m^3) (Song *et al.* 2020).

1.2.2. Regional Water Quantity Impacts

Challenges in sourcing water for HF operations in water-short regions coupled with large water consumption per HF well may limit the pace of unconventional development (Mauter *et al.* 2014). For example, water use for shale gas production in Texas is concentrated in several counties, accounting for ~80% of total water consumed in 2008 (Nicot and Scanlon 2012). In China, basins with unconventional resources in northern China (e.g., the Junggar, Tuha, and Tarim Basins) are located in arid regions (precipitation <200 mm per year) (Piao *et al.* 2010). Implementing HF operations is projected to cause considerable water stress in certain regions of the Tarim and Junggar Basins (Guo *et al.* 2016). Although the Sichuan Basin has relatively more water resources than other unconventional basins in China and is more developed at present, sourcing water for HF operations is challenging in counties such as the Yubei, Beibei, and Suining districts (Krupnick, Wang and Wang 2014; Yu *et al.* 2016). In arid regions, regulatory and policy decisions such as restricting the allocations of water resources for water users or banning of the use of municipal water for HF operations are essential measures to overcome the challenges in distributing limited water resources among different sectors (Nicot and Scanlon 2012; Rivard *et al.* 2014; Cook and Webber 2016). Similar regulatory practices may be helpful for areas projected to have water shortages in China with active HF operations (Guo *et al.* 2016; Yu *et al.* 2016). For instance, in North America, jurisdictions have implemented both voluntary and mandatory measures in order to reduce freshwater consumption and spur water recycling and reuse (Notte *et al.* 2017; Hill *et al.* 2019).

Current studies suggest that HF will not be a major water user when compared to other users (Nicot and Scanlon 2012). In an early study published in 2012, water use for shale gas production was <1%, whereas irrigation and municipal consisted of 56% and 26%, respectively in

Texas (Nicot and Scanlon 2012). More broadly, annual water use for HF operations in the U.S. contributed to 0.04% of the total freshwater use between 2012 and 2014 (Kondash and Vengosh 2015). A more recent report showed that that approximately 11-12% of the total water (1.12 billion m^3) was allocated to the energy sector in 2018, and 14% of that water was specifically allocated to HF; actual consumption was far less (<20%) than the assigned allocation (Alberta Energy Regulator 2019). Projections using current HF water use data from the Sichuan Basin (20,000 m^3 – 30,000 m^3 per well) show that shale gas development may require 20-30 million m^3 per year for HF operations in China from 2016-2026, which is relatively small compared with the projected 36 billion m^3 domestic annual water use (Yu *et al.* 2016). Nevertheless, monitoring HF water use for HF is important because water use for HF often leads to water that is permanently removed from the surface hydrologic cycle compared to other water users such as irrigation (U.S. Environmental Protection Agency 2016). Within North America, several overlapping approaches provide such data. For instance, many jurisdictions in both the U.S. and Canada require operators to report water source and usage data via FracFocus (Buono *et al.* 2019). Other regulators, such as AER and the Pennsylvania Department of Environmental Protection additionally require such information to be directly reported (Hill *et al.* 2019). In turn, these data are compiled and made available via a number of data providers (Scanlon *et al.* 2020).

1.3. FPW Management

1.3.1. Volumetric Analysis of FPW

Estimating the volumes of FPW that return to the surface is essential for evaluation of the required capacities for wastewater storage, treatment, disposal and recycling. The storage of FPW in certain regions could be challenging as due to limited land availability (e.g., the Sichuan Basin,

China) where it is difficult to properly construct storage and treatment facilities (Chen *et al.* 2011; Tonglou and Hanrong 2014). Compared to water use estimates, regional-scale estimates of FPW volumes are also challenging as the volumes of FPW reported vary significantly by TIV, shut-in time, flowback time, and reservoir lithology (Shi *et al.* 2020). This is evident in the large variances in FPW volumes per HF well reported for major U.S. unconventional plays of 8,000 - 25,900 m^3 (Kondash and Vengosh 2015), elsewhere estimated to range from 1,700-14,300 m^3 (Kondash, Albright and Vengosh 2017). In Canada, an estimated 10,000 - 25,000 m^3 FPW per HF well is recovered from the Montney play (Alessi *et al.* 2017), and approximately 50,000 m^3 per HF well from the Duvernay play (Goss *et al.* 2015). The reported FPW volumes generated from shale gas wells in the Sichuan Basin are within a similar range (5,200 - 26,000 m^3) as shale gas plays in the U.S. (Kondash and Vengosh 2015; Zou *et al.* 2018).

1.3.2. Chemical Characterization of FPW

Understanding the varying compositions of FPW is key to assessing FPW risks to the environment and the design proper wastewater treatment strategies (Ferrer and Thurman 2015a; Sun *et al.* 2019b). The general inorganic chemical constituents of FPW produced from current shale formations are similar around the world, including elevated formation-derived total dissolved solids relative to freshwater (Barbot *et al.* 2013; Haluszczak, Rose and Kump 2013; Cluff *et al.* 2014; Wu *et al.* 2017; Flynn *et al.* 2019; Zhong *et al.* 2019). However, the overall salinity of FPW from shale gas plays in China (e.g., Weiyuan, Changning) is considerably lower (e.g., maximum total dissolved solids (TDS) of reported samples are less than 50,000 mg/L) than the Marcellus and Duvernay plays, in which salinity can be well over 100,000 ppm (Cluff *et al.* 2014; Dai *et al.* 2015; Guo *et al.* 2018; Zhong *et al.* 2019; Wang *et al.* 2020a). Metals such as vanadium, selenium, and uranium are commonly reported in the compositional analyses of FPW from North America

unconventional plays (Chermak and Schreiber 2014; Lauer, Harkness and Vengosh 2016). However, these elements have not been detected or reported for FPW from active unconventional plays in China, despite recent evidence that shows, for example, uranium enrichment in rock samples collected from shale gas wells drilled into the Wufeng-Longmaxi Formation in the Sichuan Basin (Wang *et al.* 2020b). These earth rare elements may not be enriched in the fluids due to short exposure times of the injected fluid to the formation (e.g., Qaidam FPW, < 7 days of flowback).

In general, FPW has elevated organic content compared to freshwater, which primarily consists of a wide range of chemical additives (e.g., surfactants and biocides) and hydrocarbons (e.g., alkanes, aromatic hydrocarbons) (Orem *et al.* 2014; Thurman *et al.* 2014; Ferrer and Thurman 2015b; Lester *et al.* 2015; U.S. Environmental Protection Agency 2016; Rosenblum *et al.* 2017; Wang *et al.* 2020a). TOC changes rapidly as well flowback proceeds, such that TOC concentrations may be far higher in early flowback than in later produced water, after chemicals are diluted by formation water (Zhong *et al.* 2019). The dissolved organic fraction of FPW contributes significant toxicity and it is analytically challenging to identify the organics profile (Kahrilas *et al.* 2015; Kekacs *et al.* 2015; He *et al.* 2017a, 2018a). There are potential organic tracers detected in North America unconventional plays, such as biocides (e.g., glutaraldehyde and alkyl dimethyl benzyl ammonium chloride), emulsifiers, stabilizers, and surfactants (e.g., triisopropanolamine) (Kahrilas *et al.* 2016; Zhong *et al.* 2019), that have not yet been reported in FPW from emerging shale gas plays in China, though these compounds may not be present due to the use of different chemical additives. Furthermore, chemical additives (e.g., polyacrylamide) can be degraded and transformed at downhole conditions (Xiong *et al.* 2018, 2020), which can result in a series of organic transformation products in the FPW (He *et al.* 2017a). These secondary

products are difficult to identify and may pose risks of increased toxicity as compared to the injected compounds (Kahrilas *et al.* 2015; He *et al.* 2017a). Additionally, it is analytically challenging to identify chemical additives and track potential transformation byproducts in highly saline FPW. Recent advances in nontarget profiling allow for measurement of organic compounds at higher resolution, providing the opportunity to identify many of these transformation products (Figure 1.3) (Sun *et al.* 2019a).

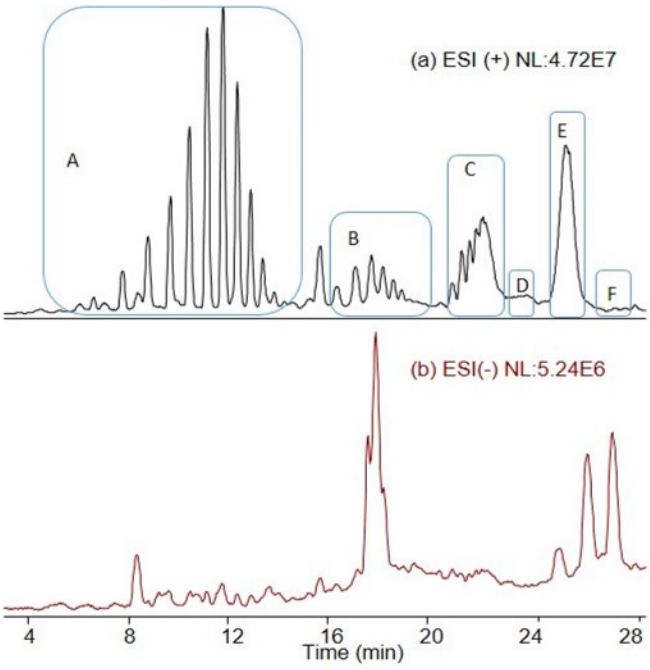

Figure 1.3 HPLC-Orbitrap MS total ion chromatograms of FPW-Day1 generated from Duvernay Formation, Canada using (a) positive ionization, and (b) negative ionization. NL is the normalized total ion abundance. Box A: Polyethylene glycol (PEG), Boxes B and C: AEO-carboxylates, Boxes D and E, Alkyl ethoxylates (AEOs). Box F: Octylphenol ethoxylates (OPEs). Adopted from a previous study (Sun *et al.* 2019a).

1.3.3. Microbiology

Analysis of FPW microbiology is critical to designing FPW treatment, reuse and recycling strategies, and to understanding biogeochemical processes that occur downhole and in water storage and treatment facilities. Early studies of the microbiology and microbial ecology of HF-FPW, primarily from plays in the U.S., are recent, with most appearing since 2010. Microbes within the taxa Proteobacteria, Clostridia, Synergistetes, Thermotogae, Spirochetes, and Bacteroidetes, and Methanomicrobia within Archaea were detected from wastewater storage impoundments from Marcellus shale gas plays (Murali Mohan *et al.* 2013b). Sulfidogenic bacteria (e.g., *Halanaerobium*) and methanogens (e.g., *Methanohalophilus*) of lower abundance were detected in FPW collected from separators in unconventional plays such as the Barnett, Marcellus, and Duvernay (Davis, Struchtemeyer and Elshahed 2012; Struchtemeyer and Elshahed 2012; Strong *et al.* 2013; Wuchter *et al.* 2013; Akob *et al.* 2015; Mouser *et al.* 2016). Enhancing our understanding of microbial ecology in the subsurface, previous studies showed diverse freshwater microbial communities shifted to low diverse community of sulfidogenic bacteria and methanogens from early flowback to late stages of produced water under increasing FPW salinity (Murali Mohan *et al.* 2013a; Cluff *et al.* 2014). A previous study showed similar but relatively fast changes in microbial community composition in FPW samples collected from the Duvernay play in Canada and suggested that FPW recycling may stimulate the enrichment of the sulfidogenic bacteria such as *Halanaerobium* (Zhong *et al.* 2019). However, high formation temperature, such as in the Bakken Formation may naturally limit the growth of microbes (Gaspar *et al.* 2016). To date, the limited literature on microbiology from China HF sites shows that sulfidogenic microorganisms are slightly enriched from the separator to the storage tank in Sichuan Basin

unconventional plays (Zhang *et al.* 2017). Additionally, sulfate-reducing and iron bacteria were detected in FPW from the Tarim Basin through a culture-based method (Liu *et al.* 2017).

Two previous reviews summarized the microbial communities that are prevalent in FPW from U.S. unconventional plays (Figure 1.4), and noted their potential to cause well souring, the clogging of pump systems, and degradation of chemical additives (Gaspar *et al.* 2014; Mouser *et al.* 2016). To maximize well productivity and reduce environmental impacts and financial costs, the reported growth of microbes in pipelines and water storage tanks suggests the need to re-assess the efficiency of biocides (Gaspar *et al.* 2014; Zhang *et al.* 2017). The dilution of HF fluid by formation water and degradation of biocides at downhole conditions may limit the performance of biocides in restricting the activity of undesirable microorganisms (Kahrilas *et al.* 2015, 2016; Zhong *et al.* 2019). Several past studies have focused on functional and culture-based metabolite analyses of the key NA FPW microbes *Halanaerobium* and *Methanohalophilus*, which further confirms their functional capacities for sulfide and methane production (Mohan *et al.* 2014; Liang *et al.* 2016; Vikram, Lipus and Bibby 2016; Lipus *et al.* 2017), and degradation of chemical additives (Evans *et al.* 2019a). These microbes are halotolerant in FPW, for example, members of *Halanaerobium*, that have optimal growth at salinities of 2.5 M (as NaCl) (An, Shen and Voordouw 2017; Booker *et al.* 2017). Glycine betaine pathways are central pathways for the *Halanaerobium* and *Methanohalophilus* to survive in the fractured subsurface of high salinity (Daly *et al.* 2016; Borton *et al.* 2018b, 2018a). Besides, these microbes can cause the *in situ* degradation of ethoxylate and glycol surfactants (Evans *et al.* 2019a), and build up adaptive immunity to survive in FPW (Vikram, Lipus and Bibby 2014; Nixon *et al.* 2019). A more recent study showed that the growth of *Halanaerobium* is also closely associated with viruses in the fractured subsurface (Daly *et al.* 2019). As noted, the number of studies beyond the U.S. is limited,

which has impeded a comprehensive understanding of the microbiology and microbial ecology in the fractured subsurface.

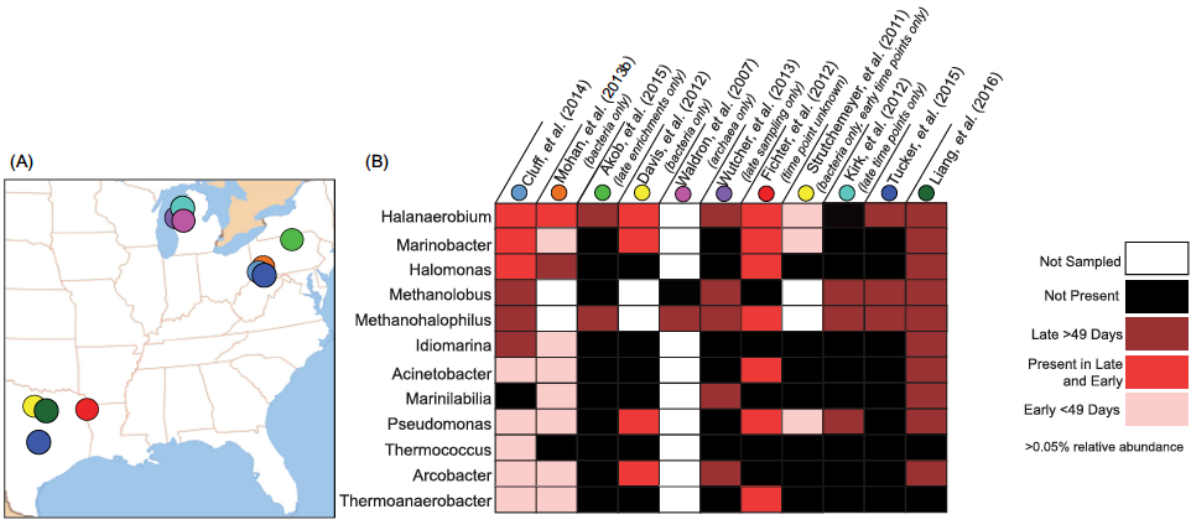

Figure 1.4 Overview of the early studies on the microbial community compositions in FPW generated from U.S. unconventional hydrocarbon plays, adopted from a previous study (Mouser *et al.* 2016).

1.4. Environmental Contamination

1.4.1. Surface Water, Groundwater, and Soil Contamination

FPW and other HF related materials such as drilling mud and fugitive gases may contaminate environments. FPW spills are a well-documented issue that influences the long-term sustainability of the HF water cycle in North America (Vidic *et al.* 2013; Vengosh *et al.* 2014; United States Environmental Protection Agency 2016). In the states of Pennsylvania, Colorado, North Dakota, and New Mexico, 6,622 of the total 21,300 HF wells reported spills to waterways and soils, and 2-16% of 31,481 HF wells in the four states had a spill each year between 2005 and

2014 (Figure 1.5) (Maloney *et al.* 2017; Patterson *et al.* 2017). FPW contamination has also occurred in groundwater near to relatively new shale gas plays in China (Huang *et al.* 2020).

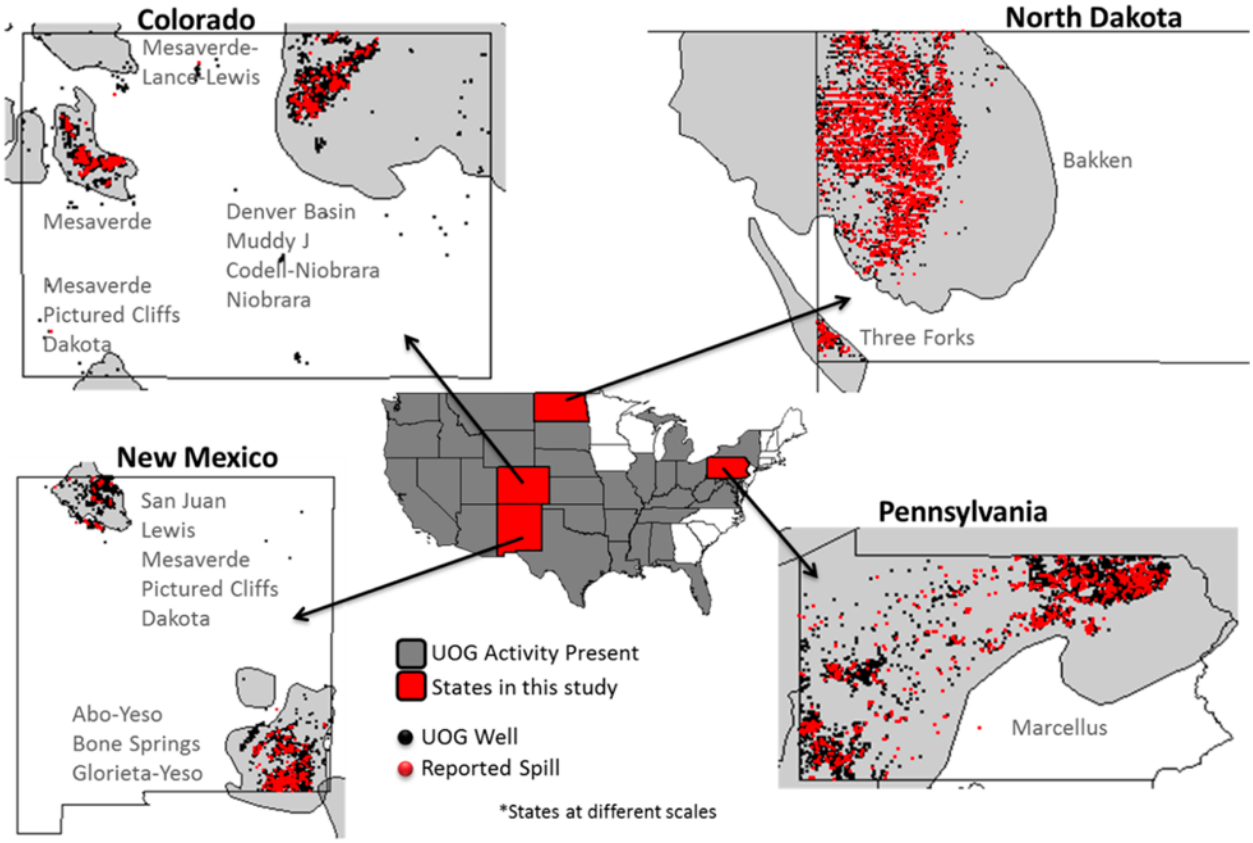

Figure 1.5 Location of unconventional oil and gas wells and reported spills an early study in the U.S., adopted from a previous study (Patterson *et al.* 2017).

It is challenging to identify the sources, pathways, and soil/water contamination caused by shale development. In the U.S., 37% of all HF wells simulated during 2014 are installed within 2 km of at least one recent (2000-2014) domestic groundwater well, resulting in drinking water sources at risk of being affected by HF operations (Jasechko and Perrone 2017). Wellbore failures and the application of HF to formations with pre-existing natural fractures and faults may provide pathways for contaminant transport from wellbores and reservoirs to shallow aquifers (Jackson *et*

al. 2013). For instance, a survey of 141 drinking water wells in the Appalachian Basin in the U.S. showed that some wells were contaminated by thermally postmature stray gases that match the isotopic signature of gases from the Marcellus Formation (Jackson *et al.* 2013). Another study showed that Marcellus shale gas development caused NG and other hydrocarbon contaminants to migrate laterally through kilometers of rock at shallow to intermediate depths to impact a drinking water source in Pennsylvania (Llewellyn *et al.* 2015). Similar cases of groundwater contamination also occurred in Colorado (Sherwood *et al.* 2016), and possibly in Quebec, Canada (Moritz *et al.* 2015). Notably, stray gas contamination of groundwater may not be associated with shale development all of the time (Barth-Naftilan, Sohng and Saiers 2018). Chemical and isotopic compositions could be influenced by mixing, migration, and oxidation processes, making it difficult to pinpoint the origin of the contamination (Moritz *et al.* 2015). Compared to volatile organic compounds, the probability of FPW contaminating shallow groundwater from the subsurface is low (Clark *et al.* 2015). Routine monitoring and developing effective indicators such as the measurement of carbon isotopes of multiple gas hydrocarbons (Jackson *et al.* 2013; Townsend-Small *et al.* 2015), and advanced detection instruments (Llewellyn *et al.* 2015), are essential to mitigate these risks.

1.4.2. FPW Toxicity and Environmental Impacts

Constituents of FPW may pose toxicity toward organisms in groundwater, surface water, and soil, and ultimately to human health (Tasker *et al.* 2018). The high salinity levels in FPW may suppress the activity of soil microbes, in contrast to organics such as polyacrylamide; soil sorption may reduce the mobility of heavy metals such as As(V) and Se(VI) (Chen *et al.* 2016, 2017). A wide range of metal ions, light hydrocarbons, organic matter, and radioactive materials have been found in FPW-contaminated water and sediments (Haluszczak, Rose and Kump 2013; Warner *et*

al. 2013; Lauer, Harkness and Vengosh 2016; Cozzarelli *et al.* 2017; Orem *et al.* 2017; Shrestha *et al.* 2017; Tasker *et al.* 2018; Preston *et al.* 2019). However, the environmental impacts may be attenuated naturally in aquatic environments with sufficient water flux. A recent two-year field survey in the U.S. showed that shale development has had limited effects on stream biology and geochemistry, highlighting the importance of conducting studies that control for regional and temporal variability (Mumford *et al.* 2020).

Contaminants in FPW have been shown to pose risks to ecosystems even at large FPW dilution factors. For example, a case study showed that heavy metals such as vanadium (4.7-218 μg/L) and selenium (1.1-172 μg/L) were enriched in surface water contaminated by FPW from the Bakken play (Lauer, Harkness and Vengosh 2016). Aquatic animals in frequent contact with such sediments may ingest metals and hydrocarbons derived from FPW that bind to sediment particle surfaces (Smalling *et al.* 2019). Organic compounds are also an important source of toxicity in FPW and have complicated fates in nature. For example, polycyclic aromatic hydrocarbons (PAHs) were observed to be bound to suspended sediments in FPW collected from the Duvernay play, and these sediments were shown to be toxic to zebrafish embryos (He *et al.* 2017a, 2018a). Therefore, sediments in FPW, for example those collected during sedimentation for FPW recycling, need to be adequately treated to reduce potential toxicity in the event of a surface release. Other compounds such as diphenyl phosphate, also detected in FPW from the Duvernay, are mobile in the soil matrix and may pose environmental risks to aquatic ecosystems (e.g., 96 h-LC50 of diphenyl phosphate on zebrafish embryos to be 50.0 ± 7.1 mg/L) (Funk *et al.* 2019).

Laboratory-based experiments showed that exposure to FPW leads to reductions in the growth, survival, and abundance of aquatic animals (Hossack *et al.* 2018; Folkerts *et al.* 2019; Wang *et al.* 2019; Folkerts, Blewett and Goss 2020). For example, depending on the FPW

chemistry, raw samples have been shown to cause acute mortality (as measured by the lethal concentration 50%, or LC50) at concentrations of 0.6%-3.9% (to zebrafish embryo) (He *et al.* 2017a) and 11.6% (to juvenile rainbow trout) (Delompré *et al.* 2019) in vertebrate aquatic species, and approximately 4% in invertebrate aquatic species (to Lumbriculus variegatus) (Mehler *et al.* 2020). Furthermore, exposure to FPW concentrations as low as 0.04% has been shown to decrease the reproductive potential of aquatic species (e.g., Daphnia magna) (Blewett *et al.* 2017a) and also inhibit other functional performance metrics (e.g., zebrafish) (Folkerts *et al.* 2017a). Morphological studies in fish have also determined that FPW dilutions as low as 2.5% may induce certain developmental deformities such as pericardial edema, yolk-sac edema, and tail/spine curvatures (in larval zebrafish) (Folkerts *et al.* 2017b) with gill remodeling occurring in juvenile and adult fish (in rainbow trout) exposed to 2.5%-3% solutions of FPW (Blewett *et al.* 2017b; Delompré *et al.* 2019). Altered expression of a myriad of genes, spanning key development genes such as *atp2a2a*, *tnnt2a*, and *nkx2.5* to specific detoxification genes such as *cyp1a*, *udpgt*, and *gst*, have also been recorded in numerous species exposed to FPW (Folkerts *et al.* 2017b; He *et al.* 2017b, 2018b), reflecting the complex toxicological nature of FPW. Many of these changes in gene expression are believed to be associated to observations of decreased performance and metabolism (e.g., organic toxicant metabolism), along with other previously mentioned sublethal toxicities (Blewett *et al.* 2017a; Folkerts *et al.* 2017b; He *et al.* 2017b, 2018b).

Microorganisms are effective fingerprints of the impacts of FPW on the environment (Zhong *et al.* 2020). Exposure to FPW has also been shown to lead to compositional shifts and growth reduction in soil and aquatic microbial communities (Akob *et al.* 2015; Kekacs *et al.* 2015; McLaughlin, Borch and Blotevogel 2016; Ulrich *et al.* 2018; Zhong *et al.* 2020). A previous study showed reduced diversity and shifts in compositions of microbial communities in the downstream

sediments near a wastewater disposal facility from an unconventional hydrocarbon development region (Akob *et al.* 2016). A broader survey showed shifts in composition of microbial communities in streams nearby HF sites in the Marcellus shale gas production region (Ulrich *et al.* 2018). Laboratory-based studies have also shown that microbes are able to degrade HF-FPW-derived chemicals, and high salinity may inhibit the biodegradation efficiency (Kekacs *et al.* 2015; McLaughlin, Borch and Blotevogel 2016; Zhong *et al.* 2020). These studies imply that FPW contamination may pose risks to ecosystem functions, as microorganisms are the basic units of an ecosystem and drive essential biogeochemical cycling. While microbes are key to the remediation of surfactants derived from FPW in natural attenuation processes (Heyob *et al.* 2017), few studies have investigated the functional potentials of these microbes once exposed to FPW. Such studies are needed to optimize bioremediation efforts and to understand natural attenuation in FPW contaminated water and soil. Due to challenges in accessing field-collected FPW, many studies have used synthetic brines that are chemically simple when compared to real-world FPW, and this simplicity may impede an accurate understanding of the FPW effects to organisms (Zhong *et al.* 2020).

1.5. Objectives

Understanding the microbiology and microbial ecology associated with FPW geochemistry is vitally important as microbes may adversely impact operations through detrimental effects such as biocorrosion, biofouling, and degradation of HF chemicals. Conversely, microbes may also generate methane in geologic formations, and remediate pollutants introduced into the environment by FPW spills. However, our knowledge of the microbial ecology and geochemistry in the hydraulically fractured subsurface, especially beyond the U.S., is rather limited. This knowledge gap impedes a full understanding of the biological effects (e.g., biofouling,

biocorrosion, or methane production) of *in situ* energy recovery. Additionally, soil and aquatic microbial ecology under real-world FPW induced stress are poorly understood. This gap poses an obstacle for assessing the risks of FPW to the surface environment in the case of a spill.

To address these gaps, this thesis primarily focused on investigating (1) the microbial ecology and geochemistry in the deep hydraulically fractured subsurface from both NA and Chinese basins, and (2) the impacts to surface aquatic and soil microbial communities under FPW-induced changes in geochemistry.

The following hypothesis was made with regards to microbial ecology in surface aquatic and soil environments under FPW-induced stress (Chapters 2 and 3):

- Microbial community diversity, composition, and functional capacities in water and soil will be shifted upon exposure to FPW, these shifts are sensitive to FPW exposure concentrations, and the underlying processes are linked to organic compounds, such as chemical additives, and salinity derived from FPW.

For the comparative analyses of microbial ecology and geochemistry in the fractured subsurface between China and North America (Chapter 4), the following hypotheses were made due to the considerable differences in salinity between the two regions:

- The microbial community diversity, composition, and functional capacities in FPW produced from China's unconventional hydrocarbon plays are distinct from those in North America.

- These differences are related to the different environmental constraints (e.g., changes in geochemistry), that are inherent to the depositional history between the two regions.

2. Chapter Two: Response of Aquatic Microbial Communities and Bioindicator Modelling of Hydraulic Fracturing Flowback And Produced Water [2]

2.1. Summary

The response of microbial communities to releases of hydraulic fracturing flowback and produced water (PW) may influence ecosystem functionalities. However, knowledge of the effects of PW spills on freshwater microbiota is limited. Here, we conducted two separate experiments: 16S rRNA gene sequencing combined with random forests modelling was used to assess freshwater community changes in simulated PW spills by volume from 0.05% to 50%. In a separate experiment, live/dead cell viability in freshwater community was tested during exposure to 10% volume PW. Three distinct patterns of microbial community shifts were identified: (i) indigenous freshwater genera remained dominant in <2.5% PW 1, (ii) from 2.5% to 5% PW 1, potential PW organic degraders such as *Pseudomonas*, *Rheinheimera*, and *Brevundimonas* became dominant, and (iii) no significant change in the relative abundance of taxa was observed in >5% PW 1. Microbial taxa including less abundant species such as *Cellvibrio* were potential bioindicators for the degree of contamination with PW. Additionally, live cells were quickly damaged by adding 10% PW, but cell counts recovered in the following days. Our study shows

---

[2]This Chapter has been published in FEMS Microbiology Ecology, Volume 96, Issue 5, May 2020, fiaa068: **Cheng Zhong**, Camilla L. Nesbø, Greg G. Goss, Brian D. Lanoil, Daniel S. Alessi

that the responses of freshwater microbiota vary by spill size, and these responses show promise as effective fingerprints for PW spills in aquatic environments.

2.2. Introduction

The use of hydraulic fracturing to extract oil and gas from impermeable shale formations has changed the global energy landscape and secured the energy-independence of countries that have historically imported a large fraction of their fuel consumption to meet energy demands (Vidic *et al.* 2013). Implicit in the rapid expansion of hydraulic fracturing is considerable water use and disposal footprints (Barbot *et al.* 2013; Vidic *et al.* 2013; Gagnon *et al.* 2016). A single shale oil and gas well can consume 13.7 - 23.8 million liters of freshwater during the fracturing process and subsequently may produce 5 - 50 million liters of flowback and produced water (PW) (Goss *et al.* 2015; Kondash and Vengosh 2015; Alessi *et al.* 2017). Estimates of spill frequencies and volumes vary widely depending on source; for example, a previous study reported the total volume spilled was about 7,600 m^3 with a median spill of 3.7 m^3 between 2005 and 2014 in the USA (U.S. Environmental Protection Agency 2015). In a study of 21,300 unconventional wells in Pennsylvania, Colorado, North Dakota, and New Mexico, 6,622 reported spills from 2005 to 2014 (Maloney *et al.* 2017). Another study reported 2-16% of 31,481 shale oil and gas wells in Colorado, New Mexico, North Dakota, and Pennsylvania had a spill each year between 2005 and 2014, and the largest spills exceeded 100 m^3 (Patterson *et al.* 2017). In any case, spills are not uncommon and include a high frequency of small incidents along with several large spills per year (Brantley *et al.* 2014). Given the frequency of PW surface releases to near-surface environments, it is important to understand potential impacts to surface water bodies, soils and aquifers.

As shale oil and gas extraction by hydraulic fracturing involves both injected fluids and components indigenous to the target geologic formation, the geochemical composition of PW is often complex, consisting of inorganic elements, petroleum compounds and residual chemical additives (Engle, Cozzarelli and Smith 2014; Akob *et al.* 2015; Ferrer and Thurman 2015b; Flynn *et al.* 2019). Furthermore, release of PW to surface water bodies and shallow aquifers may cause detrimental effects to aquatic animals and drinking water supplies (Parker *et al.* 2014; Blewett *et al.* 2017b, 2017a; Cozzarelli *et al.* 2017; Folkerts *et al.* 2017a; He *et al.* 2017b, 2018a; Orem *et al.* 2017; Hossack *et al.* 2018; Smalling *et al.* 2019; Wang *et al.* 2019; Preston *et al.* 2019). Understanding the effects of PW spills on the ecosystem microbiota is of environmental and economic importance. Microorganisms are basic units of the food web and drive biogeochemical processes in an ecosystem (Prosser *et al.* 2007). For example, microbes degrade organic matter, stabilise metals and facilitate greenhouse gas emissions to the atmosphere (Siddique *et al.* 2012). Previous studies have shown that fracturing chemicals such as polyethylene glycols (PEGs), isopropanol and petroleum hydrocarbons can be utilized by microorganisms in surface environments based on the laboratory cultivation conditions (Ulrich *et al.* 2009; Kekacs *et al.* 2015; McLaughlin, Borch and Blotevogel 2016). However, PW also contains constituents such as salts and biocides, which may restrict the growth of microorganisms (Murali Mohan *et al.* 2013a; Cluff *et al.* 2014; Akob *et al.* 2015; Kahrilas *et al.* 2015; Daly *et al.* 2016). Since the per PW spill volume can vary from small (0.15 m^3; fifth percentile) to larger scales (53 m^3; 95th percentile), we hypothesise the PW spill volume size is an essential factor in determining the spill impact on microbial communities at PW-contaminated sites, and may ultimately influence natural biodegradation pathways. Although microbial community shifts are critical to assessing the

impacts of PW spills, there is currently limited knowledge of fundamental changes in microbial community structures following PW spills into waterways.

In this study, we conducted laboratory-simulated late-stage PW spills into field-collected river water from a region of shale oil and gas development. The goals for this study were to (i) determine the impact of differing concentrations of PW on composition and diversity in aquatic microbiota and identify taxa that might be bioindicators of a PW spill, (ii) investigate cell viability kinetics under the influence of PW, and (iii) estimate the biodegradation potential of natural aquatic communities toward organic constituents of the PW. Our study advances our understanding of aquatic microbial community responses in a wide range of spill sizes, while providing fundamental knowledge to assess the fate of PW contaminants in surface releases of variable sizes.

2.3 Methods

2.3.1 Sample Preparation

The PW (PW 1) used to determine the impact of differing concentrations of PW on composition and diversity was collected in November 2016. The PW 1 sample returned to the surface at 53 days after initial flowback commenced. The sampling locationwas fromthewater/gas separator of a horizontally fractured well (Well ID: 100/12–30-063–21W5) in the Duvernay Formation located near Fox Creek, Alberta, Canada (Appendix 1 Figure S1). The details of the site information and PW_1 collection and transportation methods were described in a previous study (Zhong *et al.* 2019). The PW 1 sample was stored in sealed pails for 217 days until the experiments began. The freshwater river sample was collected in June 2018 from the Smoky River, which flows through the Duvernay shale oil and gas region (Appendix 1 Figure S1) and is a

significant source of water for the makeup of fracturing fluids. Smoky River freshwater was collected in four 5 L sterile glass bottles without headspace, stored in an opaque box with ice, and transported to the University of Alberta within 24 h. Experiments were conducted within 24 h of the arrival of the freshwater sample.

Prior to this study, the threshold for observing changes in community composition due to exposure to PW was unclear. The primary goal of this research was to capture dynamics in microbial community composition in freshwater under the effects of PW. To do so, we mixed aliquots of PW_1 and Smoky River freshwater in sterile flasks to total volumes of 100 mL, such that PW_1 consisted of 0.05%, 0.25%, 0.5%, 2.5%, 5%, 25% and 50% of the total volume in the sample series. Notably, we included mixing ratios such as 25% and 50%, which may be less likely to occur in the case of a spill because we aimed to investigate a broad range of concentrations to enhance our understanding of the effects of PW on microbial communities. We also aimed to explore potentially extreme conditions. Pure Smoky River freshwater and pure PW_1 were used as control groups. Sterile controls were prepared by autoclaving the mixed samples (5% PW_1 and 25% PW_1) twice for 45 min at 121ºC and 15 psi. Experiments were conducted in duplicate and all samples were loosely covered with aluminum foil and shaken at 70 rpm at room temperature for 7 days. Here, we used separate flasks for incubations corresponding to each target sampling day, in order to conduct sampling perturbing future samples. Each sample was filtered through 0.22 μm pore size hydrophilic polypropylene membranes (GE Healthcare Life Sciences, Ontario, Canada) at day 0, day 3 and day 7. The sampling scheme references a previous study that studied aerobic biodegradation of synthetic hydraulic fracturing fluids mixed in the laboratory (Kekacs *et al.* 2015), which itself conformed to the OECD 301 methods (OECD 1992) for studying

biodegradation. The filtered membranes were stored at -20ºC until DNA extractions. The filtered fluids were stored at 4 ºC for dissolved organic carbon (DOC) measurements.

2.3.2 Cell Viability Tests

Cell viability tests were conducted to determine how cells react to PW as a function of time. The live and dead cells were counted using the Live/Dead BacLight Viability kit (Life Technologies, Ontario, Canada). Due to the field sampling limitations, we were not able to obtain sufficient sample volumes for both the molecular analysis and cell viability tests. To address this issue, we used a PW (PW_2) for the cell viability tests that was as geochemically similar to the PW sample for molecular analysis (PW_1). PW_2 was collected from the same wellpad as PW_1 in September 2016. Cell counting was conducted within 24 h of the sample arrival. We mixed 10% by volume of PW_2 to two additional sources of freshwater (one sample is from a water storage impoundment near the fractured well and the other is from the North Saskatchewan River) and monitored cell viability for 1 month (0, 1 h, 6 h, day 1, day 3, day 7 and day 25). The incubation conditions of the cell viability tests were the same as those used for the molecular experiments. The mixing ratio and temporal schedule aimed at capturing the changes in cell health under the effect of PW_2, as well as cell recovery time afterwards. We aim to provide information about the status of cells when they are exposed to a medium-high range of PW. The results are not intended to be directly comparable with the 16S rRNA gene-based analyses. The detailed methods for live/dead cell counting were presented in our previous study (Zhong *et al.* 2019). Briefly, 15 randomly selected fields of "live" cells and "dead" cells were counted at 358 × magnification on a Leica DMRXA epifluorescence microscope equipped with FITC and rhodamine fluorescence filters. The live cell proportions were calculated by the live cells relative to the total cells per microscope field from the 15 observed microscope fields.

The live cells per mL was calculated from the live cells per microscope field by multiplying the conversion factor 1670 from the 15 observed microscope fields. A Student's t-test was used to analyse the statistical difference of the live cell numbers and live cell proportions between treatment groups and their control groups at each observed time point. ANOVA analysis was used to test if mixing 10% PW_2 had significant effects on the numbers and ratios of live cells compared with their controls.

2.3.3 Chemical Analyses

All samples were stored at 4°C until chemical analyses. The chemistry of the PW and Smoky River freshwater samples was characterized, including pH, total dissolved solids (TDS), DOC, total nitrogen (TN), major cations, and anions. The TDS was determined by weighing the residual solids after evaporating 10 mL of fluid at 200ºC in triplicate. Briefly, cations were measured using an Agilent 8800 inductively coupled plasma mass spectrometer (Agilent Technologies, California, USA) (Zhong *et al.* 2019). Anions were measured using a DX-600 ion chromatography (ThermoFisher Scientific, Massachusetts, USA) and a SmartChem Discrete Wet Chemistry Analyzer, Model 200 (Westco Scientific, Connecticut, USA) (Tabatabai and Frankenberger 1996; Westco Scientific 2007). The measurement of DOC and TN were achieved using a combustion catalyst method with a TOC-V CHS/CSN Total Organic Carbon Analyzer (Shimadzu Corporation, Kyoto, Japan) (Shimadzu Corporation 2001).

2.3.4 Sequencing of 16S rRNA Genes

DNA was extracted from the cells concentrated on filter membranes using the FastDNA Spin Kit for Soil (MP Biomedicals, Solon, USA). DNA extracts in duplicate were pooled together before PCR. The PCR primers were F515 (5′-GTGCCAGCMGCCGCGGTAA-3′) and R806 (5′-

GGACTACHVGGGTWTCTAAT-3′), which cover the V4 region of the 16S rRNA gene for bacteria and archaea. The PCR reaction using KAPA HiFi HotStart Ready Mix (Fisher Scientific, Ontario, Canada) began with a 3 min initial denaturation (95°C) followed by 35 cycles of 30 s denaturation (95°C), 30 s primer annealing (55°C), and 30 s extension (72°C) and a final 5 min extension (72°C). The PCR amplicons were submitted to The Applied Genomics Core Sequencing Facility at the University of Alberta for Illumina MiSeq paired-end sequencing. Of note, PCR amplicons were not successfully obtained from abiotic controls templates and the pure PW_1 templates, likely due to low DNA concentrations.

2.3.5 Bioinformatics and Statistics

Raw data were processed following the QIIME2 version 2018.6 standard operating procedure (https://qiime2.org/). Briefly, non-chimeric sequences were passed through quality filtering using the DADA2 pipeline implemented in QIIME2 (Appendix 1 Table S1). The filtered sequences were aligned to amplicon sequence variants (ASVs) features and further assigned to different taxonomic levels using the q2-feature-classifier, which was trained on Greengenes (version 13.8) at a 99% similarity threshold. The quality-controlled sequences were then manipulated and visualised using R (version 3.5.1) (Wickham 2009; R Core Team 2018). The datasets were rarefied to an even depth of 51,308 sequences in R in order to conduct alpha- and beta-diversity analyses of the microbial communities (Appendix 1 Table S1). Firstly, we conducted beta-diversity analysis and taxonomic analysis to group microbial community shifts by PW mixing ratios. For beta-diversity analyses, a non-metric multidimensional scaling (NMDS) ordination was performed based on the Bray-Curtis distance. The 95% confident intervals of distinguished clusters were identified using the vegan package in R (Oksanen *et al.* 2018). The envfit function implemented in the vegan package was used to correlate the 10 most abundant

genera (the relative abundance of sequences) of the entire datasets to the sample dissimilarity on the NMDS ordination (Oksanen *et al.* 2018). PERMANOVA was used to investigate the significance ($P < 0.05$) of the PW_1 proportion on beta-diversity. The alpha-diversity of each sample was assessed by using the observed numbers of ASVs, the Chao1 Richness Index, the Inverse Simpson's Index and the Shannon Diversity Index implemented in Phyloseq in R (McMurdie and Holmes 2013). The samples at each sampling day were grouped by mixing ratios, which showed similar microbial community responses based on beta-diversity analysis (similarity in ordination space) and taxonomic analysis. ANOVA analysis was used to test whether the PW_1 proportion effect was significant ($P < 0.05$) on alpha-diversity. Tukey's test was used for post-hoc analysis after the ANOVA analysis, regarding whether the differences of alpha-diversity indices between different PW_1 groupings (grouped by similarity in ordination space) were significant ($P < 0.05$). Similarly, changes of DOC concentrations were represented by the three mixing ratio groupings, and the results of different groupings were compared using ANOVA analysis combined with Tukey's test. The reads have been submitted to the National Center for Biotechnology Information Sequence Read Archive (BioProject: PRJNA593077).

2.3.6 Random Forest Modelling

Random forest modelling consists of a large number of decision trees that operate as an ensemble. This approach prevents overfitting of a classification model by averaging multiple classification models together. Here, the genera produced by sequencing were used as variables in the random forest. Important bioindicators in the NMDS clusters were identified by the randomForest package implemented in R (Breiman *et al.* 2018). For the random forest modeling, samples at day 0 (start points) were excluded from the total dataset and were used to as reference points for changes in subsequent samples. As a default, randomForest used 2/3 of the data to

construct modelling trees and the remaining 1/3 of the data to test model error. Each random forest model was set to create 500 decision trees. The model was repeated 1000 times from random sampling to model execution. The 20 most important predictors were determined based on the average values of the mean decrease in the Gini Index after 1000 runs of the random forest models were completed. The Gini Index is a measure of how each variable contributes to homogeneity of the nodes and leaves in the resulting random forests, from 0 (homogeneous) to 1 (heterogeneous). The genera with larger Gini Index values were more likely to be variables that separate the targeted groups.

2.4 Results and Discussion

2.4.1 Chemistry Characterization of PWs and Smoky River freshwater

The chemical analyses indicate lower pH and higher concentrations of solutes and nutrients in PW than for Smoky River freshwater (Table 2.1); therefore, even small PW releases may lead to large changes in freshwater chemistry. The pH of Smoky River freshwater was 7.2, while the pH of the PW_1 and PW_2 were 4.1 and 4.7, respectively. Reduced pH was a major driver of microbial community changes in hydraulic fracturing impacted streams (Ulrich *et al.* 2018). Additionally, the TDS, DOC, and TN of PWs were all significantly higher than the Smoky River freshwater. TDS concentrations were 219,037 mg L^{-1} in PW_1 and 216,637 mg L^{-1} in PW_2. TDS was dominated primarily by sodium and chloride, which were derived from the shale formations (Appendix 1 Table S2). The results suggest that the inorganic components of the two PW samples were similar. Salt may be one of the important limiting factors on cell biomass, diversity, and the degradation potential of the microbial communities (Davis, Struchtemeyer and Elshahed 2012; Murali Mohan *et al.* 2013a; Cluff *et al.* 2014; Kekacs *et al.* 2015; Mouser *et al.* 2016). DOC

concentrations were 85.3 mg L^{-1} in PW_1 and 200.8 mg L^{-1} in PW_2, while DOC concentrations were 13.9 mg L^{-1} in Smoky River freshwater. TN concentrations between the two PW samples were also similar. TN concentrations were 427.6 mg/L and 471.0 mg/L in PW_1 and PW_2, respectively, which were three orders of magnitudes higher than for Smoky River freshwater. DOC and TN could originate from several sources, including fracturing chemicals, reservoir hydrocarbons, and the injected surface water. In previous untargeted organic analyses of these PW samples at earlier flowback time from the same wellpad, we demonstrated that the major organic species were fracturing chemicals that included polyethylene glycols (PEGs) with 5-25 ethylene oxide units, the biocide alkyldimethylbenzylammonium chloride (ADBAC), and a series of petrogenic compounds such as fluorene and phenanthrene (He *et al.* 2018a). Biocides are one of the major chemical additives used in hydraulic fracturing, and the presence of biocide in hydraulic fracturing fluids may affect the overall performance of biodegradation efforts when a spill occurs (McLaughlin, Borch and Blotevogel 2016). In our samples, ADBAC was below the detection limit. The decrease in biocide concentrations is likely caused by dilution from the formation water or their chemical decomposition and transformation (Kahrilas *et al.* 2015, 2016). Ammonium was the dominant species in the TN, and the ammonium concentrations were similar to those found in PW produced by hydraulic fracturing in the Marcellus and Fayetteville Formations (up to 420 mg/L) (Harkness *et al.* 2015). The sources of ammonium remain to be further studied. Based on the previous studies, ammonium in the PW could be associated with the fracturing chemicals such as the breakers (e.g., ammonium persulfate) (Luek *et al.* 2018) and clay stabilisers (e.g., tetramethyl ammonium chloride) (Butkovskyi *et al.* 2017). The breakers allow a delayed break down of the gel polymer chains, and the clay stabilisers are used to prevent the swelling of clay particles in reaction to water-base hydraulic fracturing fluids. Besides this, ammonium is also

likely to be leached from ammonium-containing clays, evaporites, and the thermal degradation of organic matter in the shale oil and gas reservoirs (Liu *et al.* 2012).

Table 2.1 Selected geochemical parameters of Smoky River freshwater and flowback and produced water samples (PW_1 and PW_2); the rest of the chemistry is presented in Appendix 1 Table S2.

Samples	pH	TDS (mg/L)	DOC (mg/L)	TN (mg/L)	Cl (mg/L)	Na (mg/L)
PW_1	4.1	219037	85.3	427.6	104373	62831
PW_2	4.7	216637	200.8	471.0	139820	68176
Smoky River freshwater	7.2	168	13.9	0.1	1.4	1.8

TDS: total dissolved solids, DOC: dissolved organic carbon, TN: total nitrogen

2.4.2 Grouping Samples Based on Beta-Diversity Analysis

The PW_1 groupings of different PW mixing ratios were determined using beta-diversity analysis. NMDS ordination revealed two key clusters (2.5%-5% and >5%) with increasing PW_1 proportions, which were significantly separated from the cluster with PW_1 mixing ratios between 0-0.5% (Figure 2.1). Of note, the high PW_1 proportion group (> 5% PW_1) changed the least from the starting points (all tested mixing ratios at day 0) according to the NMDS ordination. Compared to clusters with higher PW_1 mixing ratios, the data points with <2.5% PW_1 were more heterogenous. PERMANOVA analyses showed that the PW_1 proportion significantly influenced microbial community structure over the 7 days of incubation ($P < 0.05$). We defined three PW_1 groupings, reflecting the degrees of effects of PW: low PW 1 (<2.5%), intermediate PW 1 (2.5–5%) and high PW 1 proportions (>5%). It is important to note that the community composition of the 0% PW_1 sample changed extensively during the incubation, while those at

higher concentrations did not (Figure 2.1). By comparison, the time effect on microbial community dynamics was overshadowed by the PW concentration effect. We subsequently used these groups for statistical analyses of DOC changes, microbial community diversity and composition shifts, and random forest modelling.

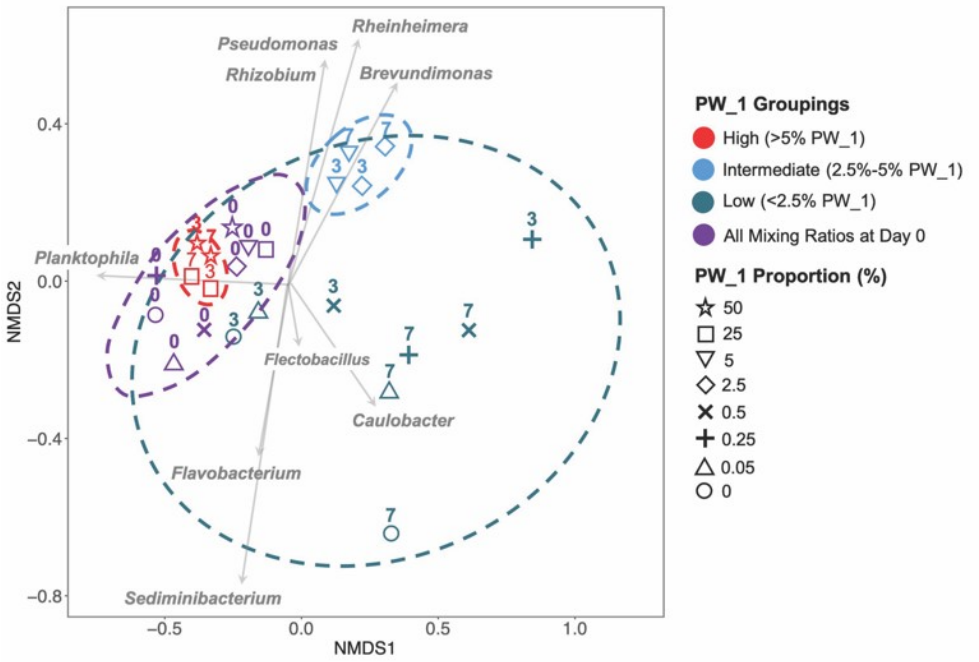

Figure 2.1 Non-metric multidimensional scaling (NMDS) plot (stress: 0.096) showing differences in microbial community composition for freshwater river samples exposed to 0-50% flowback and produced water (PW_1) at day 0, day 3, and day 7. The ten most abundant genera of the entire community were correlated to the dissimilarity of the data points. Time factors (in days) appear as numbers above the data points. Dashed lines represent the 95% confidence intervals of each group.

2.4.3 Changes in DOC Concentration

Geochemical analyses of the fluids suggest that the concentration of organics in PWs were higher than in freshwater. To examine biodegradation potential, we tracked changes in DOC concentrations over 7 days for each mixing ratio. However, ANOVA analyses showed that the relative reductions of DOC in all PW_1 groupings were not significant over 7 days. The largest reduction in DOC over 7 days was an average of 17.5% for the 2.5%-5% group (a detailed description of the DOC changes is presented in Appendix 1 and Appendix 1 Figure S2). Our results suggest no noteworthy or relevant biodegradation of the PW-related organics during the 7 days of incubation. In contrast to our results using the field-collected PW, the DOC reduction for synthetic hydraulic fracturing fluids can be up to 90% within 7 days of incubation (Kekacs *et al.* 2015), possibly indicating that the compounds in the field-collected PW can be more difficult to biodegrade than those in synthetic hydraulic fracturing fluids. The reason for the poor biodegradability of field-collected PW is not yet clear. Many factors such as the presence of recalcitrant organics and more complex mixtures of compounds may cause less reduction of DOC in field-collected PW (Kekacs *et al.* 2015; McLaughlin, Borch and Blotevogel 2016).

2.4.4 Microbial Community Shifts

The trends of changes in microbial community diversity and compositions within each PW_1 proportion category were consistent; namely, the mixtures with higher volumes of PW_1 tend to have higher microbial richness and diversity after 7 days of incubation (Figure 2.2). ANOVA analysis showed that alpha-diversity indices were significantly ($P < 0.05$) different between the three PW_1 groupings over 7 days. Following 7 days of incubation, the number of observed ASVs and Chao1 Index in the high PW proportion group (>5% PW 1) was significantly

higher ($P < 0.05$) than those with a lower PW 1 proportion. The Shannon Diversity Index in the >5% PW_1 group was significantly higher ($P < 0.05$) than for the group with <2.5% PW_1. The Inverse Simpson's Index values were significantly higher ($P < 0.05$) in the 2.5–5% PW_1 mixtures than in the other two PW_1 groupings. The full results of Tukey's tests, which compared the difference between two groups at day 0, day 3 and day 7, are presented in Appendix 1 Table S3. The increased diversity with higher proportions of PW shows uneven diversity, with dominance by a relatively small number of genera. Upon exposure to the disturbance of high levels of PW, these dominant freshwater OTUs are unable to survive, unmasking the hidden diversity present in the rarer biosphere in these samples. Further, some of these surviving bacteria may grow in response to the increased organics in the PW. The combination of removal of dominant species and increased growth leads to higher overall diversity. This follows the concept of the intermediate disturbance hypothesis, which states that the highest biodiversity will be found at intermediate levels of disturbance (Bendix, Wiley and Commons 2017).

The genus *Flavobacterium* within the phylum Bacteroidetes is known to be prevalent in freshwater environments (Bernardet and Bowman 2006) and was the most abundant bacterium across all the samples at day 0. They consistently constituted the largest proportion of the microbial community in mixtures containing < 2.5% PW_1 throughout 7 days of incubation (Figure 2.3). The genera *Methylotenera* and *Caulobacter* were also a higher fraction of sequences than in other genera in mixtures containing < 2.5% PW_1. Compared to the pure freshwater sample, the trends of microbial community dynamics were similar in samples with low PW_1 mixing ratios over 7 days (Figure 2.3). Previously characterized members of these genera have been reported to use glucose and methylamine in natural aquatic environments (Wright and Cain 1969; Entcheva-Dimitrov and Spormann 2004; Bernardet and Bowman 2006; Kalyuzhnaya *et al.* 2010). Our results

suggest that relatively low PW concentrations may not dramatically influence the freshwater community. This is likely because the concentrations of PW_1 were too low for there to be a toxicity effect on the community (e.g., pH > ~7.1, salinity < ~5640 mg/L, DOC < ~15.7 mg/L; values based on calculation of the PW proportion), so the indigenous microorganisms were still present.

Compared to the low PW_1 proportion group (< 2.5% PW_1), a pronounced influence on taxonomy compositions began at 2.5% PW_1. The relative abundance of *Flavobacterium* decreased in experiments having PW_1 proportions between 2.5%-5% by day 3 (the average of *Flavobacterium* decrease from 75% to 20% and 9% in 2.5% PW_1 and 5% PW_1 mixtures, respectively) and remained at lower abundance levels at day 7 (29% and 4% in 2.5% PW_1 and 5% PW_1 mixtures, respectively). The genera *Pseudomonas*, *Rheinheimera*, *Rhizobium*, and *Brevundimonas* were significantly ($P < 0.05$) correlated to the 2.5%-5% PW_1 proportion group (Figure 2.1 and Figure 2.3). The enrichments of these genera may be related to the organic constituents introduced by PW. Some previously characterized members of these genera are capable of degradation of a wide variety of hydrocarbons and organics found in hydraulic fracturing fluids such as isopropanol and PEGs (Williams and Sayers 1994; Ahmad, Mehmannavaz and Damaj 1997; Chaîneau *et al.* 1999; Táncsics *et al.* 2010; Kekacs *et al.* 2015; Nuria Obradors 2015). The increase in the relative abundance of these genera is consistent with the higher DOC reduction observed for the 2.5%-5% PW_1 proportion group. The shift in microbial community composition is likely to benefit biodegradation processes. However, significant reductions in PW organic concentrations may take a greater length of time. Functional analysis through metagenomics in the future may allow us to better understand the role of particular microorganisms in the biodegradation process. Additional details of the relative

abundance of the 10 most abundant bacteria are presented in Appendix 1 Table S4. The increase of abundant microorganisms (Figure 2.3) plus some less abundant microorganisms (comprising less than 1% of the total sequences) such as genera *Aquicella*, *Geobacter*, *Massilia*, *Pedobacter*, *Planctomyces*, and *Sphingomonas* was consistent with the increasing diversity in the mixtures with >0.5% PW_1. This shift suggests that the medium to higher concentrations of PW may inhibit the indigenous freshwater species at natural conditions, while providing substrates for more types of microorganisms to grow. Alternatively, higher concentrations of PW may remove abundant members of the community, allowing for detection of rare community members. *Halanaerobium* was observed in all mixtures following the addition of PW_1, although at low abundance (less than 1% of the total sequences). They are the most prevalent bacteria in the PW and are capable of using PW-related organics (Daly *et al.* 2016; Zhong *et al.* 2019).

For mixtures at the highest concentrations of PW_1 (25% and 50%) (e.g., pH < ~6.4, salinity > ~54,900 mg/L, DOC > ~32 mg/L; values based on calculation of the PW proportion), the microbial response is heavily restricted. No significant shift was observed in microbial community compositions for both 25% and 50% mixtures after 7 days of incubation (Figure 2.3). It has been demonstrated that elevated salinity (>40,000 mg/L) can inhibit the aerobic degradation of hydraulic fracturing fluid chemicals by microbial communities derived from surface aquatic environments over the course of 7 days (Kekacs *et al.* 2015). The input salinity in the 25% and 50% mixtures is above this threshold. Thus, the community likely did not shift significantly because the growth of all of its members was inhibited by high salinity.

The overall abundance of Archaea is less than 1%, but their presence also appears to be associated with the presence of PW_1 chemical constituents in mixtures. Within the archaea, ASVs related to the genus *Nitrosopumilus* increased from 0.03% to 0.23% in both the 25% and 50%

mixtures after 7 days of incubation. The previous study showed that certain strains of the genus *Nitrosopumilus*, such as *Nitrosopumilus maritimus*, are able to oxidize ammonia (Walker *et al.* 2010). Thus, this increase may be associated with the high ammonium concentration found in PW_1. The rest of the Archaea genera changed in their relative abundance by less than 0.1% of all reads. Additional results for the archaea are presented in the Appendix 1 Table S5.

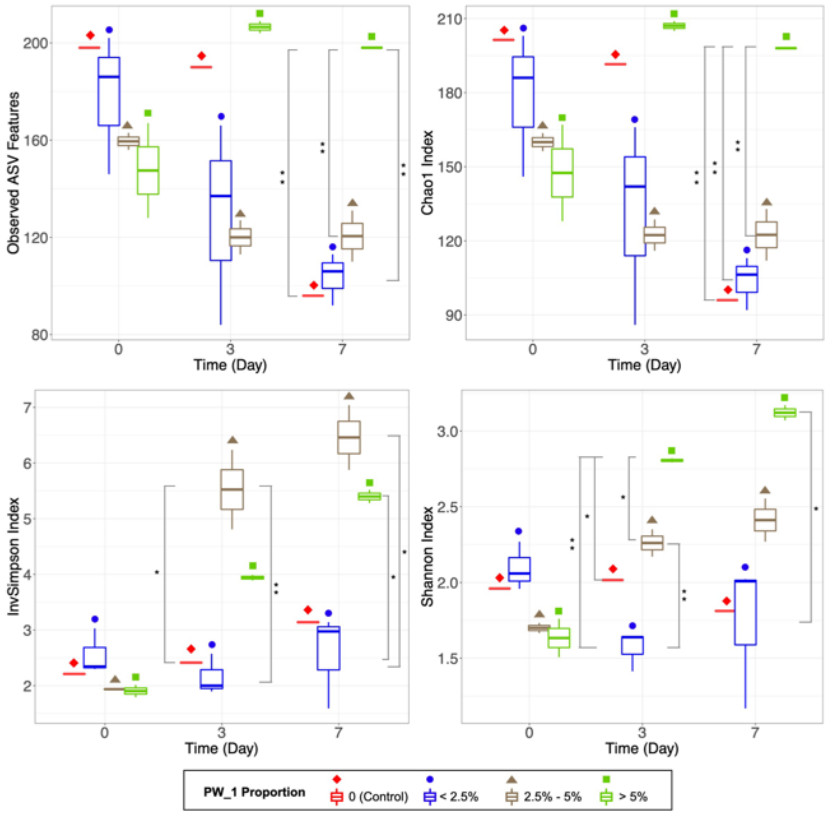

Figure 2.2 Temporal changes in microbial diversity between flowback and produced water (PW_1) proportion ranges from <2.5%, 2.5%-5%, >5% and a control group at day 0, day 3, and day 7, which suggest that increasing PW concentration may increase the overall taxonomic richness and diversity. Taxonomic richness was represented by (a) Observed ASVs Index and (b) Chao1 Richness Index, while taxonomic diversity was represented by (c) Inverse Simpson's Index

and (d) Shannon Diversity Index. The significant thresholds are $P < 0.05$ (*), $P < 0.01$ (**), and $P < 0.001$ (***).

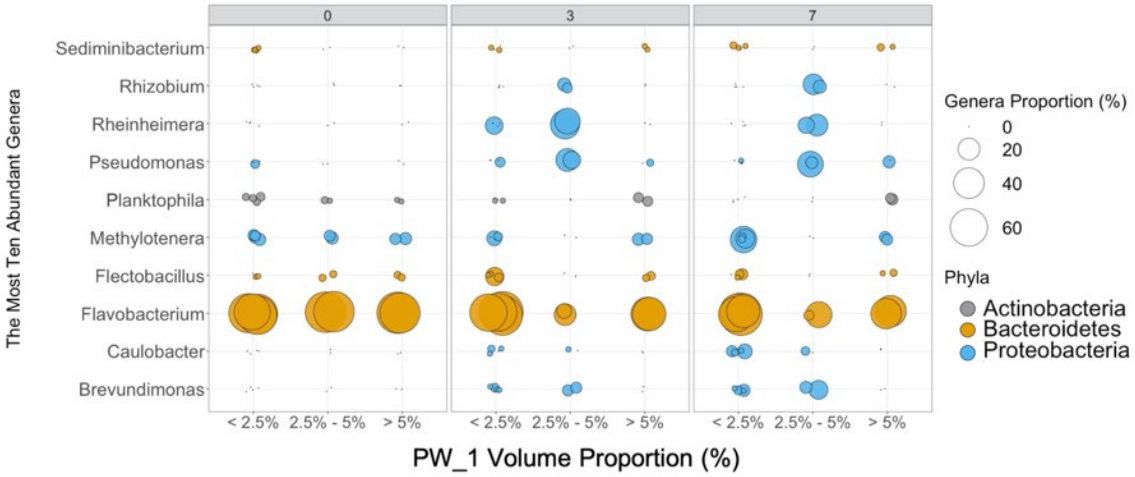

Figure 2.3 Three-factor plot showing temporal changes of the relative abundance in the 10 most abundant genera (y-axis) as a function of time and flowback and produced water (PW_1) mixing ratio. The PW_1 beta-diversity groups (<2.5%, 2.5%-5%, >5%) were labelled on the lower x-axis. Day 0, 3, and 7 exposures are labelled on the higher x-axis and are split into three sub-plots, one for each time point. The relative abundance (%) of a genus is represented by the bubble size, and colors on the y-axis represent the phyla of the 10 genera. Each bubble is represented a data point (the detailed description of each data point and their corresponding values of relative abundance are presented in Appendix 1 Figure S3 and Appendix 1 Table S4).

2.4.5 Cell Viability Kinetics

We mixed 10% PW_2 into hydraulic fracturing source water and North Saskatchewan River water and observed changes in live/dead cell numbers and ratios over 1 month. Here, these results are incorporated into this study to further discuss the potential effect of medium-high

mixing ratios of PW on the microbial communities. Compared to the live cell status in the natural conditions, ANOVA tests showed that the treatment of adding 10% PW_2 had significant effects on live cell numbers ($P < 0.01$) and survived cell proportions ($P < 0.05$) (Figure 2.4). Student's t-tests showed that the trends of live cell numbers and live cell proportions in the treatments remained at statistically lower levels within a day compared to the controls ($P < 0.05$), then fully recovered in the following days. Specifically, for the treatments of hydraulic fracturing source water and North Saskatchewan River water, the live cell ratios were significantly lower than their control groups immediately after adding 10% PW_2 (Figure 2.4A). Live cell ratios in hydraulic fracturing source water increased from 8% ± 13% at the very beginning to the peaks of 52% ± 18% at day 7. Similarly, live cell ratios in North Saskatchewan river water increased from 5% ± 4% to the peaks of 43% ± 11% at day 3. By contrast, live cell ratios in the two freshwater controls were either remained consistent or decreased over the observation period (Figure 2.4A). Live cell numbers in the treatment groups reduced to ~10^4 cells/mL at the very beginning, which were one magnitude and two magnitudes lower than their controls, respectively. Ultimately live cell numbers increased to above 10^5 cell/mL at day 7 in the hydraulic fracturing source water treatment group, as well as day 3 in the North Saskatchewan river water treatment group. Live cell numbers in both the freshwater controls were either remained consistent or decreased over the observation period (Figure 2.4B).

The results suggest that PW_2 may immediately kill many of the original freshwater live cells, resulting in dead or metabolically inactive cells shortly after exposure. At relatively low concentration, PW_2 could also provide nutrients for certain microorganisms to be enriched in the mixed conditions. As shown previously for the Duvernay Formation, increasing salinity is highly correlated to decreased cell viability in the fluids produced during the first few days of well

flowback (Zhong *et al.* 2019). The viability tests imply that relatively rapid loss of cell viability due to PW effects could lead to less observed changes in microbial community composition and DOC reduction in the high PW_1 proportion group. Of note, we did not aim to correlate changes in cell viability with the changes in community compositions between the two independent experiments (PW_1 and PW_2). How chemical differences in PW_1 and PW_2 influence dynamic changes in microbial communities, such as the viability of cells in differing DOC concentrations, needs further investigation.

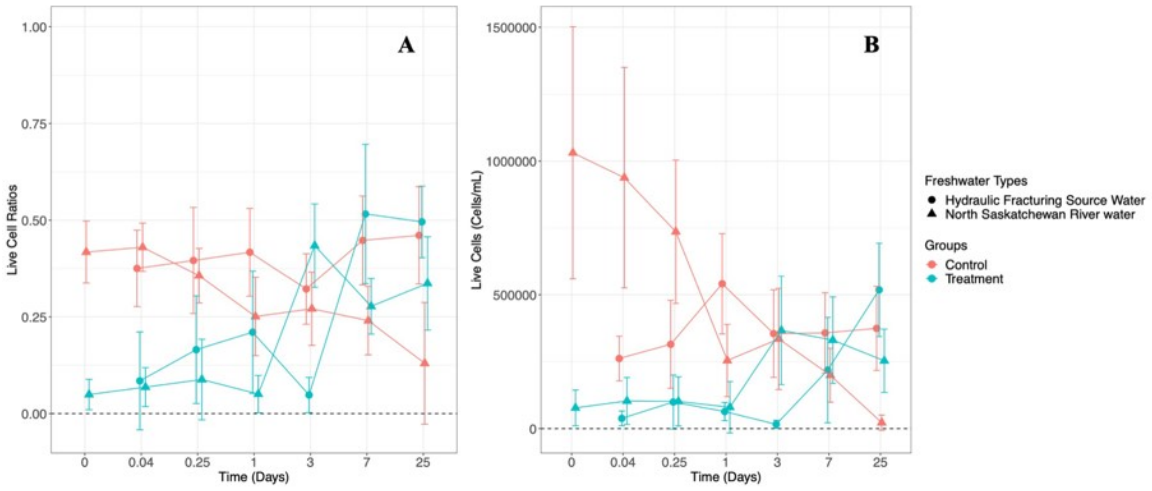

Figure 2.4 Temporal changes in (a) live cell ratios (n=15), and (b) live cell numbers (n=15) over 30 days in experiments that added 10% flowback and produced water (PW_2) to the hydraulic fracturing source water sample (circle) and the North Saskatchewan River water sample (triangle). The treatment group and the control group are represented by green and red symbols, respectively.

2.4.6 Microbial Community Indicators

Analytical methods to fully identify chemical constituents of PW are limited, which may impede evaluation of the impacts of PW releases to freshwater (He *et al.* 2017a). Sequencing

technologies coupled with random forests modelling may allow for microorganisms to be additional indicators for the assessment of PW contaminated water and/or soil (Ulrich *et al.* 2018). In our random forests models, 429 bacterial genera and 17 archaeal genera were used to predict the three PW_1 groupings (< 2.5% PW_1, 2.5%-5.0% PW_1, and > 5% PW_1) observed by beta-diversity analyses (Figure 2.1). The top 20 important predictors based on the Gini Index score, which were generated by random forest modelling, are shown in Figure 2.5. The results show that *Pseudomonas*, *Rhizobium*, *Sediminibacterium*, and *Brevundimonas* are important predictors generated by random forest modelling. Genera such as *Brevundimonas*, *Rhizobium*, and *Pseudomonas* were significantly correlated to the 2.5%-5% PW_1 group (Figure 2.1, Appendix 1 Table S6), indicating they could be effective indicators in identifying effects of spills at the intermediate PW_1 proportion group (2.5%-5%). *Flavobacterium* was the most shifted bacterial genus from the low PW_1 proportion group (<2.5% PW_1) to the intermediate 2.5%-5% PW_1 beta-diversity sample group (Figure 2.3). The high Gini Index value suggests that a drop in abundance of *Flavobacterium* may be an important negative indictor of a spill. Compared with the traditional technique that uses a single or few microbes to be the indictors for pollution, random forest modelling allows for numerous genera to be predictors as an ensemble, including those in relatively less abundance. Here, we found that minor genera such as *Cellvibrio* (1.4% and 2.6% of the total sequences in the 2.5% PW_1 at day 3 and day 7, respectively) and *Shewanella* (1.8% and 3% of the total sequences in the 5% PW_1 at day 3 and day 7, respectively) could act as effective predictors of PW exposure. Consistently, indicators (e.g., *Rhizobium*, *Pedobacter*, and *Cystobacter*) selected in our random forests model were similar at the family level (e.g., Rhizobiaceae, Sphingobacteriaceae, and Cystobacteraceae) to indicator organisms found in streams impacted by hydraulic fracturing in Pennsylvania (Ulrich *et al.* 2018). Additionally, *Rhizobium*, *Cystobacter*,

Pedobacter, *Herminiimonas*, *Sediminibacterium*, and *Desulfosporosinus* were consistently within the top 20 indicators in our model, and were also similar at the order level (e.g., Clostridiales, Rhizobiales, Myxococcales, and Sphingobacteriales), organisms which were enriched in an impacted stream near a shale gas disposal facility in central West Virginia (Akob *et al.* 2016). As machine learning advances, training additional data in future studies shows promise in improving the spill classification accuracy and precision.

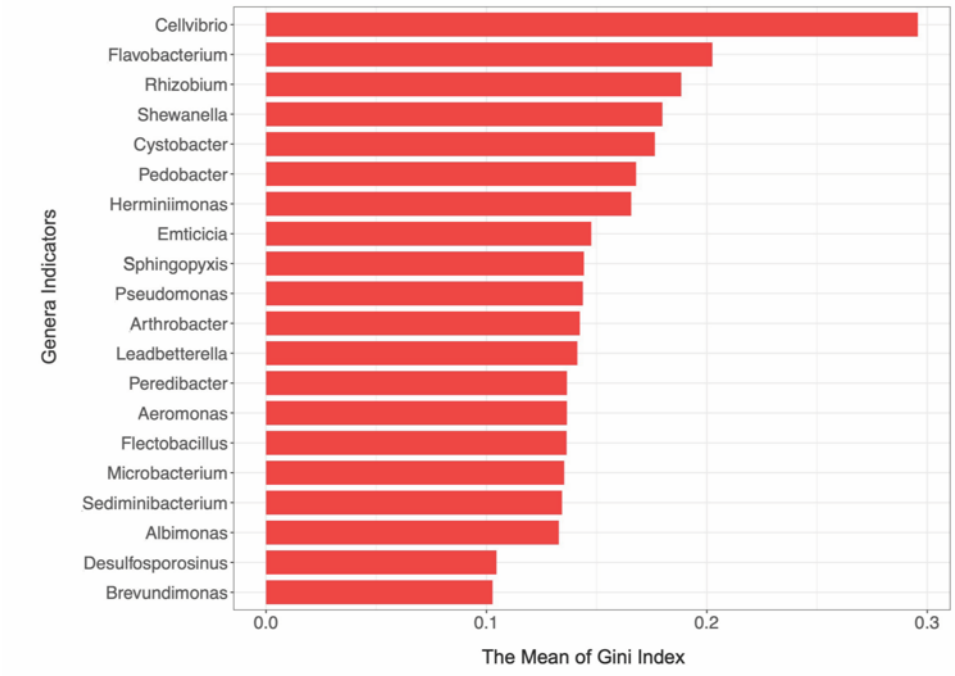

Figure 2.5 The Gini Index generated by random forests showing the 20 most important genera in predicting PW spills. The random forests technique examines a large ensemble of decision trees, which has considered all the genera represented in the 16S rRNA gene-based sequences. The Gini Index is a measure of how each variable contributes to homogeneity of the nodes and leaves in the resulting random forests, from 0 (homogeneous) to 1 (heterogeneous). A genus having a larger Gini Index is more likely to be a variable that separates the targeted groups.

2.4.7 Environmental Implications and Future Steps

Our study demonstrates that microbial community compositions in aquatic environments are sensitive to the scale of a PW spill. Our results suggest that large volume spills (leading to 25+% PW_1 concentrations), while rare, could have considerable impact. The cell viability test, as a separate experiment, implies that the community may lose its ability to adapt and may not be viable after exposure. We found that the degradation rate for organic constituents of field-collected PW is considerably less than those measured in laboratory-synthesised brines, which calls for more studies using real PW and for long-term monitoring of recalcitrant organic pollutants at contaminated sites. Microorganisms such as *Pseudomonas*, *Rheinheimera*, *Rhizobium*, and *Brevundimonas* may serve as key players in remediation processes, which may also be used as biosensors to assess water bodies that have experienced a PW spill. Moreover, through building a low-error random forests model (Appendix 1 Table S7), the sets of genera uncovered in 16S rRNA gene analyses show promise as bioindicators to represent changes in aquatic ecosystems due to PW releases into freshwater. In the future, these bioindicators may compliment traditional analytical methods such as chemical analyses in assessing the magnitude or severity of a release. Additional work is necessary to identify the site-specific biomarkers, since the components of PW such as salinity and organic compound identities vary by extraction site and by time of flowback. Moreover, more research is needed to determine whether these bioindicators are universal and how long-lasting they are.

2.5 Statement of Contribution

In this study, sample preparation, cell viability, bioinformatics, random forests modeling, data analyses, and paper writing were conducted by me. Samples were sent to the Natural Resources Analytical Laboratory at University of Alberta for chemical characterization.

3. Chapter Three: Impacts of Hydraulic Fracturing Flowback and Produced Water on Soil Microbial Communities [3]

3.1. Summary

Improper handling of flowback and produced water (FPW) generated by hydraulic fracturing can result in spills that pose risks to soil environments. However, little is known about the response of the soil microbiota to FPW. Here, we investigated the effects of one month of aerobic FPW exposure on soil microbial communities using 16S amplicon, metagenomic, and respiration analyses in luvisol and chernozem soils that are common in the region of hydraulic fracturing in Alberta, Canada. We found that the luvisol was more susceptible than the chernozem of more nutrients regarding decreases of biodiversity and respiration activity following FPW exposure. In >50% FPW amended luvisol, *Marinobacter* spp. were significantly enriched in the microbial communities, biodegradation genes (e.g., alcohol dehydrogenases and alkane 1-monooxygenase) increased in the communities, and metagenome-assembled genomes containing genes to resist salt effects formed unique clade with their phylogenetic relatives identified in other FPW related environments. However, these responses were not evident in chernozem, but we found antibiotic genes increased of the communities in the 50% FPW amended chernozem. Our

[3] This Chapter is written based on a complete manuscript: **Cheng Zhong**, Konstantin von Gunten, Camilla L. Nesbø, Yifeng Zhang, Xiaoqing Shao, Rong Jin, Kurt O. Konhauser, Greg G. Goss, Brian D. Lanoil, Daniel S. Alessi

study suggests that soil microbial communities have strong capacity to respond FPW-induced stress. While high nutrient soil may buffer these responses, resulting uncertainty in natural attenuation of FPW substance.

3.2. Introduction

Hydraulic fracturing flowback and produced water (HF-FPW) is generated during unconventional oil and gas development (Vengosh *et al.* 2014). FPW spills during handling, transport and treatment at the surface can cause the contamination of nearby environments. In the United States (US) alone, 6,622 spills were reported from 21,300 unconventional wells across Pennsylvania, Colorado, North Dakota, and New Mexico from 2005 to 2014 (Maloney *et al.* 2017). Another estimate for the same studied regions and time period reported 2-16% of 31,481 unconventional wells had a spill each year (Patterson *et al.* 2017). FPW spills directly impact soil, surface water bodies and shallow aquifers (U.S. Environmental Protection Agency 2015). The complex chemistry of HF-FPW, including a wide range of inorganic elements, hydrocarbons, chemical additives, and radioactive compounds, poses challenges to its characterization and understanding the transport and toxicity of FPW in natural environments (Warner *et al.* 2013; Haluszczak, Rose and Kump 2013; Engle, Cozzarelli and Smith 2014; Ferrer and Thurman 2015b; Lauer, Harkness and Vengosh 2016; Cozzarelli *et al.* 2017; Orem *et al.* 2017; Shrestha *et al.* 2017; Tasker *et al.* 2018; Flynn *et al.* 2019; Preston *et al.* 2019), particularly since many organic compounds in FPW are not well characterized (He *et al.* 2018a; Sun *et al.* 2019a). FPW spills may cause adverse effects to aquatic ecosystems; for example, exposure to FPW may lead to reductions in the growth, survival, and abundance of invertebrate and vertebrate aquatic species (Hossack *et al.* 2018; Folkerts *et al.* 2019; Wang *et al.* 2019; Folkerts, Blewett and Goss 2020). Despite the fact that soil most frequently receives FPW contamination (64%) in the environment (U.S.

Environmental Protection Agency 2015), the understanding of soil contamination by FPW is limited as compared to studies of aquatic environments.

An understanding of changes in soil microbial ecology is needed to evaluate of the fate of FPW contaminated ecosystems because microorganisms are fundamental in maintaining a habitable ecosystems via immobilizing toxic metals and degrading organic pollutants (Delgado-Baquerizo *et al.* 2016a). Previous studies have shown that microbes can use chemical additives in fracturing fluid as carbon and nitrogen sources, and exposure to FPW may lead to compositional shifts and growth reduction in soil and aquatic microbial communities (Akob *et al.* 2015; Kekacs *et al.* 2015; McLaughlin, Borch and Blotevogel 2016; Heyob *et al.* 2017; Ulrich *et al.* 2018; Zhong *et al.* 2020); yet many of these studies used synthetic brines that are chemically simple when compared to field-collected FPW. FPW constituents, such as salt and biocides, can restrict the performance of microbial communities during natural attenuation or bioremediation at contaminated sites and may impact their contributions to normal ecosystem functioning (Kekacs *et al.* 2015; McLaughlin, Borch and Blotevogel 2016; Zhong *et al.* 2020). However, little is known about changes in the soil microbiome upon exposure to FPW. Recent advances in metagenomics allow for simultaneous examination of multiple genes and the discovery of genomes of novel uncultivated species, which together reveal the complex interactions between microorganisms in an ecosystem (Quince *et al.* 2017). Metagenomics is a tool that can aid in uncovering mechanisms by which microbial communities are influenced by FPW, the sources of toxicity in FPW, and the genetic potential of microbial communities to resist and remediate FPW-influenced environments such as soils. More importantly, metagenomics can compare the genomic similarity of a system to that characterized previously in other environments. This approach allows for more accurate

predictions of the functions of detected microbes and a better understanding of how an environment function.

Soil, containing often diverse and abundant microbial communities, is often the first natural environment exposed to contaminants when FPW is spilled (Chen *et al.* 2016; Heyob *et al.* 2017). The soil microbiome is likely to play a key role in mitigating the transport of contaminants in FPW by stabilizing metals and inducing the biodegradation of synthetic and natural organic compounds (Chen *et al.* 2016; McLaughlin, Borch and Blotevogel 2016). Further, the soil microbiome is critical to normal functioning of soils and spills may disrupt their functioning (Delgado-Baquerizo *et al.* 2016b). To shed light on the response of soil microbiomes to FPW, we investigated the taxonomic and genomic compositions of microbial communities exposed to FPW from two physiochemically distinct soils near to active HF regions. The genome-based investigation was coupled with soil respiration tests to assess integrated microbial activity at various exposure levels. We demonstrate here that FPW exposure leads to changes in community structure, functional gene abundance, and microbial respiration activity. Our results provide some of the first insights into the impacts of FPW on the soil environment and provide essential knowledge for risk assessment of FPW contamination (e.g., spills) and bioremediation processes.

3.3. Materials and Methods

3.3.1 Field Sampling

The FPW sample used in this study was collected in July 2019 from a horizontally-oriented hydraulic fractured well (well ID: 02-12-81-W6) in the Montney Formation of the Western Canadian Sedimentary Basin near Dawson Creek, British Columbia, Canada. The Montney Formation deposited in offshore to shoreface, consisting of shale, siltstone, sandstone, and

grainstone in Lower Triassic age (Zonneveld *et al.* 2010). The FPW was collected 12 days after the initial flowback (June 2nd, 2019) at the gas and water separators. The shut-in period was about 10 days (HF complete at May 23th, 2019). The FPW sample, with a thin oil layer on top, was stored in a 20L sealed pails without headspace at room temperature for two months until the experiment began.

Soil samples were collected in August 2019 in Alberta, Canada. Luvisol samples was taken from the Ah horizon (0-10 cm organic-rich soil) at the Fox Creek site (FC luvisol, elevation: 870 m, N: 54.3455 W: 116.8596) and chernozem was taken from the Ap horizon (1-29 cm, organic-rich soil) at the Grande Prairie site (GP chernozem, elevation: 682 m, N: 55.21418 W: 118.93912) (Appendix 2 Figure S1). These sites represent typical Albertan soils that could be potentially exposed to spills from unconventional oil and gas wells. Soils were sieved on site (2 mm) to remove coarse grains and plant material. The pH and conductivity of soil slurry (1:2 ratio milli-Q water) were measured on-site. The soil sampling details and descriptions of the surrounding vegetation at the sampling sites are given in Appendix 2.

3.3.2. Soil and FPW Characterization

Particle size, moisture, total carbon and nitrogen The particle sizes of soil samples were analysed using a Laser Diffraction Particle Size Analyzer (Beckman Coulter, LS 13 320, California, United States). Soil moisture was determined gravimetrically after drying at 105 °C overnight. The total organic carbon (TOC) and total nitrogen (TN) in the FPW samples were analyzed using a Shimadzu TOC-L with ASI-L and TNM-L modules (Shimadzu, Kyoto, Japan). The TOC and TN of soil samples were analyzed using a Thermo Scientific, Flash 2000 Organic Elemental Analyzer (Thermo Scientific, Massachusetts, United States).

Inorganic chemistry analyses Total dissolved solids (TDS) in FPW were determined by weighing the residual solids after evaporating 10 mL of fluid at 200°C. Well mixed, oven-dried, and ground (mortar and pestle) soil samples were digested in aqua regia for elemental analysis. For this purpose, 0.1 g of sample was put into 30 mL PTFE digestion tubes and amended with 6 mL 37% HCl and 3 mL 70% HNO_3. The mixture was heated to 130°C on a hotplate until the reaction seized (approximately 1 h). After, the mixtures were heated at 175°C to reduce the volume and the remains were diluted to 50 mL using 2% HNO_3 and 0.5% HCl. The solution was filtered with nylon syringe filters (0.45 µm) and analyzed on an Agilent 8800 Triple Quad ICP-MS (Agilent, California, United States) according to the manufacturer's recommendations (Sugiyama and Nakano 2014; Sakai 2015). Additionally, metals of FPW and digested soil were subsequently analyzed in triplicate by using a Thermo iCAP6300 Duo inductively coupled plasma-optical emission spectrometer (ICP-OES) (Thermo Scientific, Cambridge, United Kingdom). Other common ions in soil and FPW samples, including NH_4^+, Cl^-, SO_4^{2-}, PO_4^{3-}, NO_2^-, NO_3^- were analyzed in triplicate using colorimetric methods and using the Thermo Fisher Gallery Beermaster Plus (ThermoFisher, Massachusetts, United States). The detailed methods are presented in Appendix 2.

Nontarget organic analyses Twenty milliliters of the FPW sample and a source water sample (control) were syringe filtered and solid-phase extracted in preparation for mass spectrometry (Appendix 2). High Performance Liquid Chromatography with Orbitrap Mass Spectrophotometry (HPLC/Orbitrap MS) analysis was utilized for nontarget analysis of the aqueous phase organic compounds in the FPW sample and source water sample. Ten microliters of the organic extract (equivalent to 140 µL of the original FPW) was injected for analysis. The HPLC utilized a C18 analytical column (Poroshell 120 EC-C18, 2.1 × 100 mm, particle size 2.7

μm, Agilent Technologies) with a flow rate of 250 μl/min. The elution gradient started from 99% A (LC-MS grade water with ammonium formate) and 1% B (methanol), which was held for 1 min, ramped to 100% B by 36 min, held until 39 min, and returned to initial conditions by 42 min.

The mass spectrometer was operated in positive electrospray ionization mode, acquiring in full scan mode (m/z 100 to 2000) at 2.3 Hz) with nominal resolving power of 120,000 at m/z 400. The instrumental methods followed a previous study (Sun *et al.* 2019a), which used accurate mass measurement and tandem MS analysis. The various groups of organic compounds are also referred to the results of a previous study (Sun *et al.* 2019a).

3.3.3. Soil FPW Exposure

FPW was diluted with sterile deionized water to 100%, 50%, 25%, 5%, 0.5% FPW (v/v). Treatment groups included luvisol and chernozem soil slurries comprised of 1/2 part soil and 1/2 part FPW (10 g soil + 10 g FPW). Control groups were soil slurry consisting of 10 g soil + 10 g sterile deionized water. The soil slurry samples were incubated in 100 mL pre-sterilized borosilicate glass serum bottles sealed with butyl rubber stoppers. The incubations were prepared in triplicate. Bottles were sacrificed for DNA extracts at days 0, 3, 9, and 27. These samples are labeled according to the following convention: soil site_FPW percentage_incubation days (e.g. FC_50_27).

3.3.4. DNA Extraction, PCR, 16S rRNA Sequencing, and Bioinformatics

DNA was extracted from approximately 500g of soil using the FastDNA Spin Kit for Soil (MP Biomedicals, Solon, USA). DNA extracts were pooled in triplicate before PCR. DNA quality was evaluated visually via gel electrophoresis and quantified using a Qubit 3.0 fluorometer

(Thermo-Fischer, Waltham, MA, USA). Bacterial 16S rRNA genes were PCR-amplified with dual-barcoded primers (515F/806R) targeting the V4 region (5'-GTGCCAGCMGCCGCGGTAA-3' and 5'-GGACTACHVGGGTWTCTAAT-3'), as per the protocol of a previous study (Kozich *et al.* 2013). Amplicons were sequenced with an Illumina MiSeq using the 300-bp paired-end kit (v.3). The raw data was processed using QIIME2 (Bolyen *et al.* 2019). Quality control was conducted using the DADA2 pipeline implemented in QIIME2. Filtered sequences were aligned to Amplicon Sequence Variants (ASVs) features and further assigned to different taxonomic levels using SILVA 132 at a 99% similarity threshold (Quast *et al.* 2013; Yilmaz *et al.* 2014).

3.3.5. Shotgun Metagenomic Sequencing and Bioinformatics

Luvisol and chernozem soil samples exposed to 50% FPW at day 0 and day 27 and their controls (0% FPW) at day 0 and day 27 were selected for shotgun metagenomic sequencing on the Illumina NextSeq 500 Platform. Libraries were prepared using an Illumina Nextera library preparation kit (Illumina, San Diego, CA, USA). Raw reads were trimmed and checked for quality using Trimmomatic v0.39 (Bolger, Lohse and Usadel 2014). Reads passing quality control were taxonomically classified using phyloFlash (Gruber-Vodicka, Seah and Pruesse 2019). Then, high quality reads were separately- and co-assembled into contigs and further genome bins using the MEGAHIT v1.1.2 (Li *et al.* 2015) and MetaSPAdes v3.12.0 assemblers implemented in the Anvio snakemake workflow (Eren *et al.* 2015; Nurk *et al.* 2017). The MEGAHIT assembled contigs were annotated in Integrated Microbial Genomes & Microbiomes (IMG) (Appendix 2 Table S1) (Chen *et al.* 2019). Each assembly was binned using MaxBin v2.2.7 (Wu *et al.* 2014) and MetaBAT v1.7 (Kang *et al.* 2015) with default parameters. Constructed genome bins from all combinations of assembly and binning software were pooled and dereplicated with dRep v2.3.2 (Olm *et al.* 2017). Dereplicated bins were assessed for quality using CheckM v1.1.0 (Parks *et al.* 2015) and then

assigned to taxonomy using GTDB-Tk v0.2.2 (Chaumeil *et al.* 2019). Bins with >90% completeness and <10% contamination scores were selected for further manual refinement (Appendix 2). Phylogenetic trees were reconstructed by using GtoTree based on 25 bacterial and archaeal genes of single-cell genomes implemented in GtoTree (Lee and Ponty 2019). The similarity of the bins to their close neighborhood in the phylogenetic tree was assessed using FastANI (Jain *et al.* 2018). The quality of finished MAGs was verified using Anvi'o and annotated using IMG (Chen *et al.* 2019). In addition to the IMG annotation, we manually blast important genes in National Center for Biotechnology Information to confirm their presence and absence. For bins that were not able to annotate in IMG, we used MetaErg annotation for their functional profiling (Dong and Strous 2019).

3.3.6. Soil Respiration Assay

Luvisol and chernozem soil samples were incubated in sealed serum bottles with 0%, 5%, and 50% FPW at room temperature for 30 days. Abiotic controls were prepared by autoclaving a soil twice and afterwards by adding 2.5 g/L NaN_3. O_2 and CO_2 concentrations were measured at days 0, 2, 3, 5, 9, 18, and 27. O_2 concentration was analyzed using an Oxygen Sensor Spot SP-PSt3-NAU (Regensburg, Germany). The concentration of produced CO_2 was determined by manual injection of a sample of the microcosm headspace into a Thermo Fisher Scientific Trace 1300 Gas Chromatographer equipped with a Thermal Conductivity Detector (TCD) (Massachusetts, United States) and fitted with a capillary column (TG-BOND Q) of 30 m length, 0.32 mm internal diameter and 0.10 μm film thickness, following the standard instrument manual.

3.3.7. Statistical Analyses

Alpha- and Beta-diversity of even-depth 16S rRNA amplicon data were analyzed using Phyloseq (McMurdie and Holmes 2013). Envfit implemented in the Vegan package was used for correlating ASVs to sample dissimilarities on the components analysis (Oksanen *et al.* 2018). Gene counts and gene abundance (estimated gene copy numbers inferred from coverage) were normalized according to samples' genome size (library size) before comparative analyses (Appendix 2 Table S1) (Chen *et al.* 2019). PCoA determined the community beta-diversity and similarity of protein-coding gene distribution based on major functional categories of the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Significant tests analyses used in this study were PERMANOVA and ANOVA combined with Tukey HSD analyses (p=0.05 was used as cut-off value). The detailed methods and additional software used for statistical analyses were provided in Appendix 2.

3.3.8. Data Availability

Raw 16S rRNA gene sequences are available on the National Center for Biotechnology Information (NCBI) GenBank database under BioProject PRJNA640927. Shotgun metagenome assemblies (Taxon Object ID: 3300041026-3300041031, 3300041038, 3300041039) and high-quality (>90% completeness and <5% contamination) metagenome-assembled genomes (MAGs) are available (Taxon ID: 2886275984 and 2886272890) were deposited in Integrated Microbial Genomes & Microbiome (IMG).

3.4. Results and Discussion

3.4.1. FPW and Soil Sample Characteristics

The total salinity (measured by TDS) of the FPW sample was 109 ± 4 g L^{-1}. FPW has abundant total C, N, and S, but is low in P in relatively reduced redox conditions (Table 3.1, Appendix 2 Table S2), while the other bulk and trace elements in FPW were not abundant relative to the inherent concentrations of the FC luvisol and GP chernozem (Appendix 2). Polyethylene glycols (PEGs) and alkyl ethoxylates (AEOs) were the two major groups of additives identified in our FPW sample (Figure 3.1), which are regularly detected as surfactants from other FPW samples (Thurman *et al.* 2014; He *et al.* 2018a; Evans *et al.* 2019a; Sun *et al.* 2019a).

The soil characterization suggests that GP chernozem may have higher nutrients for microbial communities (Table 3.1, Appendix 2 Table S2). In comparison, GP chernozem contained greater fractions of silt and clay and a lower fraction of sand, higher moisture and organic carbon contents and higher total C, N, S, and P, compared to FC luvisol. GP chernozem had higher abundances of Al, K, S, and Fe, and the trace elements Se, Rb, Cd, and Cs than did the FC luvisol. While FC luvisol contained more Ca and Zn, and had a higher pH value of 7.9 compared to pH 6.1 for GP chernozem.

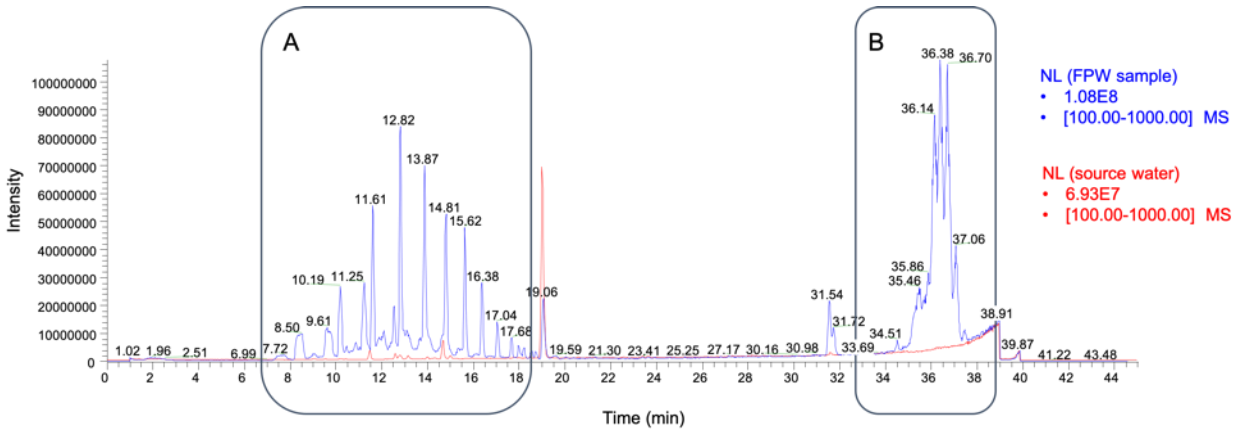

Figure 3.1 HPLC-Orbitrap MS total ion chromatograms of FPW and a corresponding source water. NL is the normalized total ion abundance. Box A covers a group of polyethylene glycol (PEG) and Box B covers a group of alkyl ethoxylates (AEOs).

Table 3.1 Selected physicochemical properties of FPW and two types of soil samples collected from FC luvisol and GP chernozem. BDL: below detection limit; NM: not measured; pH and conductivity were measured on site (n=4). Rest of the analyses were measured in the laboratory conditions (where available data shown as average ± standard deviation, n=3).

Profile	FPW	Soils	
		FC luvisol	GP chernozem
Clay (%)	NM	11.19 ± 0.46	17.77 ± 1.00
Silt (%)	NM	17.47 ± 3.20	33.97 ± 0.71
Sand (%)	NM	71.34 ± 2.82	48.26 ± 1.37
pH	6.14 ± 0.02	7.89 ± 0.14	6.06 ± 0.07
Moisture (%)	NM	14.7 ± 0.4	29.7 ± 0.7
Na	30900 ± 290 mg L^{-1}	135 mg kg^{-1}	209 mg kg^{-1}
Fe	24.97 ± 1.57 mg L^{-1}	18500 mg kg^{-1}	25100 mg kg^{-1}
Pb	12.23 ± 1.90 mg L^{-1}	4.55 mg kg^{-1}	13.6 mg kg^{-1}
Ca	6340.82 ± 64.76 mg L^{-1}	9340 mg kg^{-1}	4390 mg kg^{-1}
As	30.81 ± 2.03 mg L^{-1}	4.34 mg kg^{-1}	6.22 mg kg^{-1}
P	BDL	420 mg kg^{-1}	947 mg kg^{-1}
S	311.64 ± 10.45 mg L^{-1}	130 mg kg^{-1}	650 mg kg^{-1}
Cl^-	63100 ± 261 mg L^{-1}	5.72 ± 0.63 mg kg^{-1}	29.70 ± 15.78 mg kg^{-1}
NH_4^+	412.72 ± 10.43 mg L^{-1}	0.43 ± 0.05 mg kg^{-1}	0.65 ± 0.16 mg kg^{-1}
SO_4^{2-}	58.73 ± 0.92 mg L^{-1}	2.19 ± 0.28 mg kg^{-1}	33.58 ± 0.94 mg kg^{-1}
TOC	344.57 ± 50.93 mg L^{-1}	0.84 ± 0.02 w/w%	5.48 ± 0.05 w/w%
TN	366.07 ± 8.47 mg L^{-1}	0.07 ± 0.00 w/w%	0.50 ± 0.01 w/w%

3.4.2. FPW Impact on Community Structure of Soil Microbiota

The geochemical results suggested that FC luvisol and GP chernozem may have distinct microbial community structures. The 16S rRNA amplicon data showed that the *in situ* FC luvisol had higher biodiversity than GP chernozem (Appendix 2). For example, the Shannon Diversity for FC luvisol and GP chernozem were 4.48 and 3.35, respectively. FC luvisol enriched in Proteobacteria, while, GP chernozem community compositions (e.g., Acidobacteria and Actinobacteria) were profoundly influenced by the lower pH (Appendix 2) (Fierer and Jackson 2006). This compositional difference inherent to soil types present throughout the exposure experiment during incubation (Figure 3.2A, Appendix 2 Figure S2).

Our results suggested that soil microbial communities have strong tolerance toward FPW because both soil communities only showed large shifts in samples exposed to 50-100% FPW. The 16S rRNA amplicon data showed that increasing exposure lead to biodiversity lost (Figure 3.2B) and community compositional shifts (Figure 3.2C), and that this effect is stronger in FC luvisol than GP chernozem. The PCoA analyses showed that 60% of the variance in taxonomic compositions was due to differences between the soils (R^2=0.578, p=0.001), which is unsurprising given that the two soils had distinct native microbial community compositions. FPW exposure levels, especially for both soils exposed to 50-100% FPW, were significantly (R^2=0.108, p=0.007) influence community structure. The significant effect of the incubation time (R^2=0.07-0.09, p=0.002-0.005) responsible for the compositional changes can only be resolved without considering site dissimilarity between two soil types (Appendix 2 Figure S3).

Upon 50%-100% FPW exposure of the FC luvisol, the relative abundance of the bacterial genus *Marinobacter* dramatically increased by 13-23% after 27 days incubation (Envfit, p=0.001,

Appendix 2 Table S3, Appendix 2 Figure S4). For example, *Marinobacter* increased from 2.5% to 25% in FC luvisol exposed to 100% FPW from day 0 to 27. Despite the presence of O_2, the strictly anaerobic bacterium *Halanaerobium*, commonly found in North American FPW (Daly *et al.* 2016; Mouser *et al.* 2016; Zhong *et al.* 2019), was slightly enriched from 0.2% to 1% in their relative abundance in luvisol exposed to >50% FPW.

The taxonomic classification of our metagenome data was consistent with the 16S rRNA amplicon analyses of compositional changes (Figure 3.2A, detailed description is presented in Appendix 2). Importantly, we confirmed the shift of *Marinobacter* in FC luvisol; the *Marinobacter* increased from 0.7% to 3.7% in luvisol exposed to 50% FPW after 27 days, while no *Marinobacter* were detected for luvisol without exposing to FPW. However, metagenome-based analyses showed lower relative abundance of the most abundant taxa in the 16S rRNA amplicon-based analyses and higher abundance of minor taxa such as Firmicutes and Myxococcot compared to 16S (Appendix 2 Figure S5). These differences may result from bias during amplification of 16S rRNA genes (Suzuki and Giovannoni 1996), or the highly abundant taxa having on average more 16S rRNA genes per genome (Větrovský and Baldrian 2013).

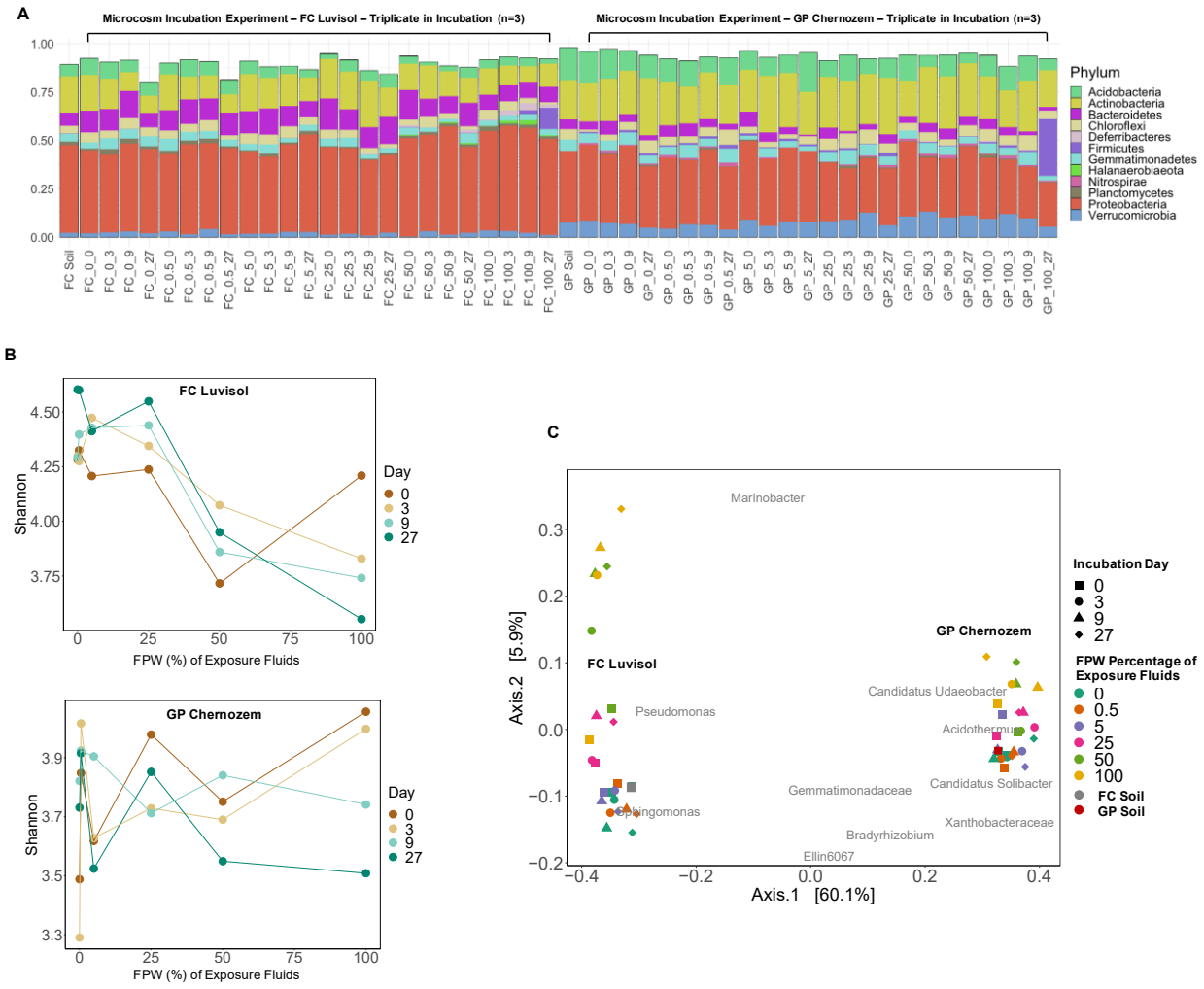

Figure 3.2 16S rRNA amplicon-based analyses showing response of microbial community in taxanomic composition and diversity to exposure to a range from 0-100% FPW from day 0 to 27 (triplicate in incubation n=3). (A) microbial community profile showing the relative abundance of the microbial phyla in FC luvisol and GP chernozem exposed to 0-100% FPW before and after incubation (the changes of major microbial genera >5% of the total community were provided in Appendix 2 Figure S4), (B) changes in FC luvisol and GP chernozem community diversity (as measured by the Shannon diversity index) with increasing FPW exposure levels, (C) PCoA ordination showing changes in FC luvisol and GP chernozem taxonomic compositions at increased FPW exposure levels. Vectors indicate the influence of the

10 most abundant ASVs on the composition. For Figure 3.1A, sample labels: location_FPW fraction (%) of the total volume fluids_incubation day (e.g. FC_50_27); the average relative abundance of ASV of the total sequence dataset below 0.1% were removed to reduce the total phylum types in this Figure.

3.4.3. FPW Impact on Gene Abundance of Soil Microbial Communities

To further investigate changes in functions of communities impacted by FPW exposure, we compared the abundance of functional genes in soils exposed to 50% FPW for 27 days relative to changes in soils not exposed to FPW. PCoA analyses showed that functional gene abundance changes between FPW exposed and unexposed soils were diverged after 27 days incubation (Figure 3.3A). Gene abundance related to xenobiotics biodegradation and metabolism, energy metabolism, carbohydrate metabolism, amino acid metabolisms, and lipid metabolisms were generally higher in GP chernozem than FC luvisol (Figure 3.3B), which may be linked to higher concentrations of nutrients in the GP chernozem. Gene abundance related to these metabolisms profoundly dropped in FC luvisol after 27 days without FPW exposure. For other sample groups, including both soils exposed to FPW, genes related to these above-mentioned metabolisms stayed the same or increased after incubation. This may be consistent with previous studies, inferring that FPW derived chemical additives may be used by soil microbiota and enhanced their activity (Chen *et al.* 2016, 2017). Note that, genes related to cell motility increased in FPW-exposed GP chernozem. Along the same line, we found multidrug efflux pump genes (e.g., *mexY*, *mexF*, *mdtC*, *acrA*) of antibiotic resistant increased in abundance in GP chernozem exposed to 50% FPW after 27 days incubation. This suggests that microbial communities in GP chernozem may actively respond to reduce xenobiotic toxicity from FPW.

The response of counts and estimated copies of essential genes could be related to exposure to FPW-derived organic compounds (e.g., PEGs and AEOs), unmeasured hydrocarbons, and resistance to FPW stress (e.g., salt) (Sleator and Hill 2002; Wadhams and Armitage 2004; Nau-Wagner *et al.* 2012; Ferrer and Thurman 2015a; Daly *et al.* 2016; Heyob *et al.* 2017; Borton *et al.* 2018a; He *et al.* 2018a; Evans *et al.* 2019a; Sun *et al.* 2019a). Thus, we examined aldehyde and alcohol dehydrogenase genes that could be responsible for degrading PEGs and AEOs and their derivatives (Heyob *et al.* 2017; Rogers *et al.* 2019). Our results showed that short-chain alcohol dehydrogenase family, alcohol dehydrogenase, and alcohol dehydrogenase (NADP+) increased in FC luvisol exposed to 50% FPW while the same genes showed a reduction in abundance or were maintained at low levels in the controls after 27 days of incubation (Figure 3.3C). The increase of alcohol dehydrogenase genes have been previously found to be related to degradation of PEGs associated with simulated spills of FPW in groundwater (Rogers *et al.* 2019). However, no clear trend of alcohol dehydrogenase changes was observed for GP chernozem. For GP chernozem, abundant genes were annotated as short-chain alcohol dehydrogenase family [COG 1028], a few was classified to alcohol dehydrogenase [EC: 1.1.1.1] and alcohol dehydrogenase (NADP+) [EC: 1.1.1.2]. The abundance of aldehyde dehydrogenase also fell for FC luvisol samples without FPW exposure after incubation and the changes for the rest of the experimental groups were relatively small. *pduCDE*, the gene sets previously reported to be encoded in Firmicutes that can shorten ethoxylate chains (Heyob *et al.* 2017; Evans *et al.* 2019a), was absent or detected in low abundance in our samples (highest values were 0.34% gene counts and 3 gene copies per million reads in GP_0_27).

Various hydrocarbons were commonly detected in FPW according to previous studies, including polycyclic aromatic hydrocarbons of less abundance and are more recalcitrant (Ferrer

and Thurman 2015a; Orem *et al.* 2017; He *et al.* 2018a; Wang *et al.* 2020a). We found that gene counts and gene copy abundance of alkane 1-monooxygenase (*alkB1_2*) increased in FC luvisol exposed to FPW after 27 days, which was not found in its control and for the GP chernozem (Figure 3.2C). There is no difference in the abundance of genes involved in metabolism of more structural complex hydrocarbons between exposure and control samples. For example, for both soils, cytochrome p450 and various monooxygenases and dioxygenases, were detected in low abundance and did not change in abundance after exposure to FPW (Appendix 2 Figure S6). Microbial communities may not target the recalcitrant compounds in environments where other carbon sources are present, such as in GP chernozem with rich nutrients and chemical surfactants (Kleindienst *et al.* 2015).

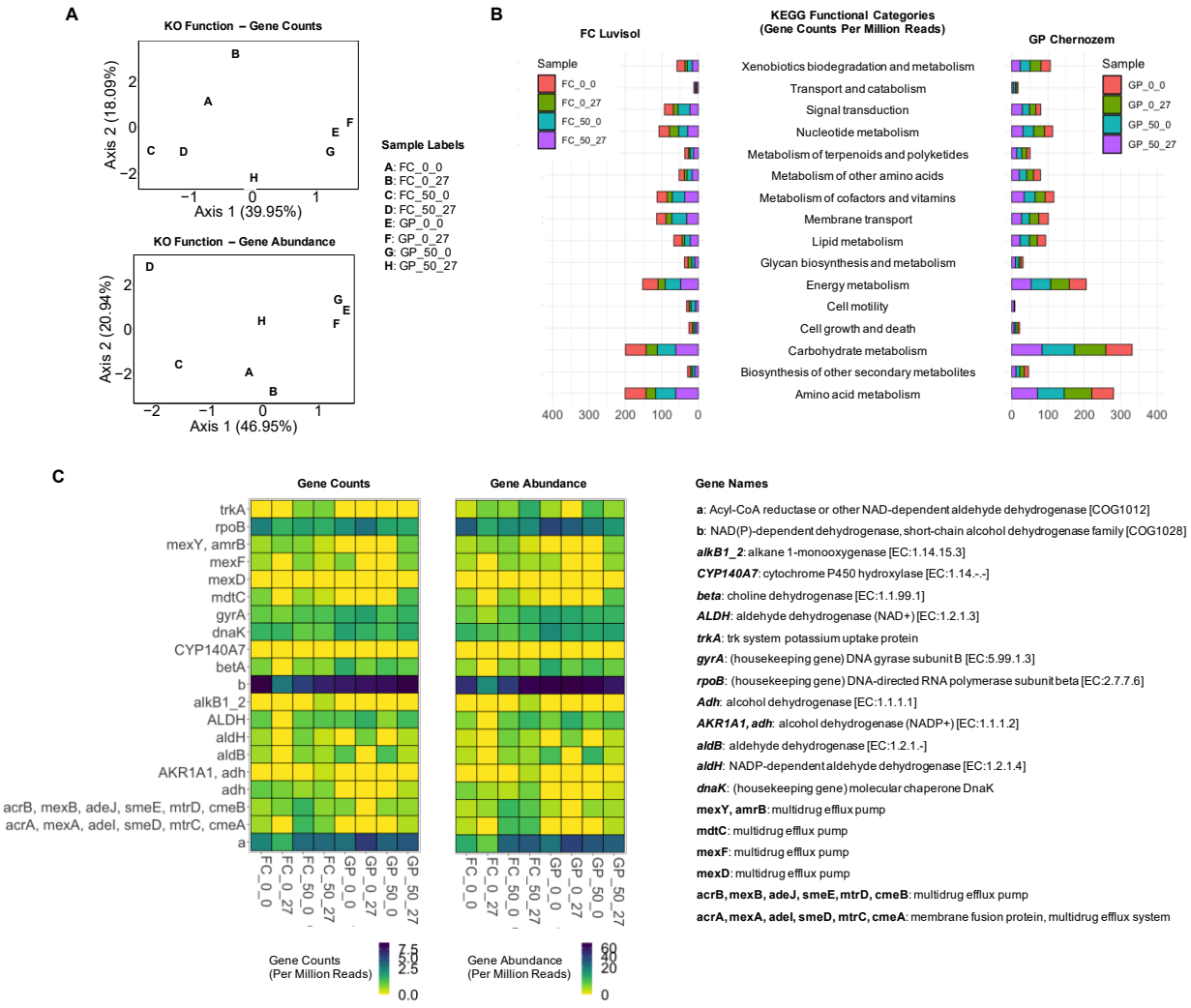

Figure 3.3 Functional analyses for FC luvisol and GP chernozem samples exposed to 50% FPW versus control groups (triplicate in incubation n=3) (A) PCoA ordination of Bray-Curtis distance of normalized gene counts and gene abundance showing function gene structure dissimilarity based on KO functions, (B) effects of FPW exposure on the abundance of gene counts of KEGG level 1 functional categories related to metabolism, environmental information processing, and cellular processes, (C) functions of selected genes that may involve in response to FPW constituents (full lists of examined genes were present in Appendix 2 Figure S6). Sample labels: location_FPW fraction (%) of the total volume fluids_incubation day (e.g. FC_50_27).

3.4.4. Unique Clades of Genomes under (HF) FPW Effects

Metagenome assembled genomes (MAGs) allow for the linking of specific functions to genomes and their taxonomic lineages. Three MAGs, bin 6, bin 5, and bin 2, were reconstructed from the FC_50_27 sample (Appendix 2 Table S4). Bin 6, classified as *Marinobacter persicus*, was observed at high abundance, consistent with the 16S rRNA based analyses. The genome related to bin 6 is most closely related (ANI=95%) to the MAG detected in the FPW obtained from fractured shale formations (Figure 3.4A), which are predicted to degrade carbohydrates, such as glucose, fructose, lactate, and acetate, and hydrocarbons, such as alkanes, benzene, toluene, and xylenes (Daly *et al.* 2016). Other phylogenetically close genomes such as *Marinobacter hydrocarbonoclasticus* are widely detected in hydrocarbon impacted environments, e.g., from Mediterranean seawater near a petroleum refinery with high hydrocarbonoclastic potentials (Christen, Fernandez and Acquaviva 1992). Additionally, increasing abundance of *Marinobacter* was observed for simulated deep sea plume to response to oil effects (Kleindienst *et al.* 2015). Genomes related to bin 5 and bin 2 were classified as *Salegentibacter* (Figure 3.4B) and *Erythrobacter* (Appendix 2 Figure S7), respectively. Both bacterial genera were previously detected in FPW from shale formations (Lipus 2017), and strains of these genera were shown to be halotolerant to halophilic in saline environments (McCammon and Bowman 2000; Xu *et al.* 2018). The closest (ANI=98%) known relative to bin 5 is *Salegentibacter sp. 24*, detected in HF-impacted freshwater and sediment (IMG Submission ID, 182861). The unique clade of these two bins in the phylogenetic tree suggest some of the microbes could be effective indicators for heavy FPW impacts.

Annotation of the *Marinobacter*, *Salegentibacter*, and *Erythrobacter* MAGs, showed that these genomes contained genes that may involve in the degradation of organic compounds (Figure 3.4C). Dehydrogenase genes such as aldehyde and alcohol dehydrogenase genes were present, which has been also detected in the closest MAGs reported from fractured shale FPW (Daly *et al.* 2016). The presence of these dehydrogenase genes detected suggested these organisms may degrade the PEGs and AEOs and to CO_2 via Citric acid (TCA) cycle (Acetyl-CoA and Succinyl-CoA were detected) under aerobic conditions. However, further study is needed to confirm whether they are specifically targeted to PEGs and AEOs, or additional unreported organic compounds due to the potential broad use of aldehyde and alcohol dehydrogenase. While, pduCDE and pegA, genes more specifically associated with degradation of chemicals such as PEGs and AEOs (Sugimoto *et al.* 2001; Heyob *et al.* 2017) were also not detected in the three bins.

Notably, genes that may be involved in the first few steps of degradation of benzene, toluene, and xylenes detected from the most closely related MAGs of *Marinobacter* from fractured shale FPW (Daly *et al.* 2016), were not detected in bin 6 related to *Marinobacte* strains, as well as for the other two MAGs. Several key genes for cyclohexane and alkane degradation, including Alkane 1-monooxygenase 1 (*alkB1_2)*, Probable FAD-binding monooxygenase (*Alma*), and Cytochrome P450 52A1 (*CYP52A1)* were not deteted in the three bins (Liu *et al.* 2019). For example, *alkB1_2* contained by *Marinobacter hydrocarbonoclasticus* was not found in bin 6.

Glycine betaine and K^+ based metabolisms are both important strategies for microbes such as *Marinobacter* that survive in saline FPW in the hydraulically fractured subsurface and other similar saline environments (Daly *et al.* 2016; Lipus *et al.* 2017; Borton *et al.* 2018a, 2018b; Nixon *et al.* 2019). Our reconstructed high-quality bin 6 and bin 5 contained functional genes related to

K^+ uptake and glycine betaine pathways (Figure 3.4C, Appendix 2 Table S5). The annotation of the bin 6 for these osmoprotectent genes were consistent with previous studies (Daly *et al.* 2016). Choline dehydrogenase and betaine aldehyde dehydrogenase were detected in bin 6. Choline is converted to betaine aldehyde by choline dehydrogenase and then ultimately converted to glycine betaine by betaine aldehyde dehydrogenase (Daly *et al.* 2016). These osmoprotectant genes highlights that soil bacteria may persist in maintaining their functions after FPW exposure. However, no evidence suggests that FPW exposure induced increase abundance of the genes related to glycine betaine and K^+ based metabolisms over the community (Figure 3.3C). Future studies that conducting gene expression analyses and isolating these persistent bacteria (e.g., *Marinobacter*) for functional analysis may validate this conjecture.

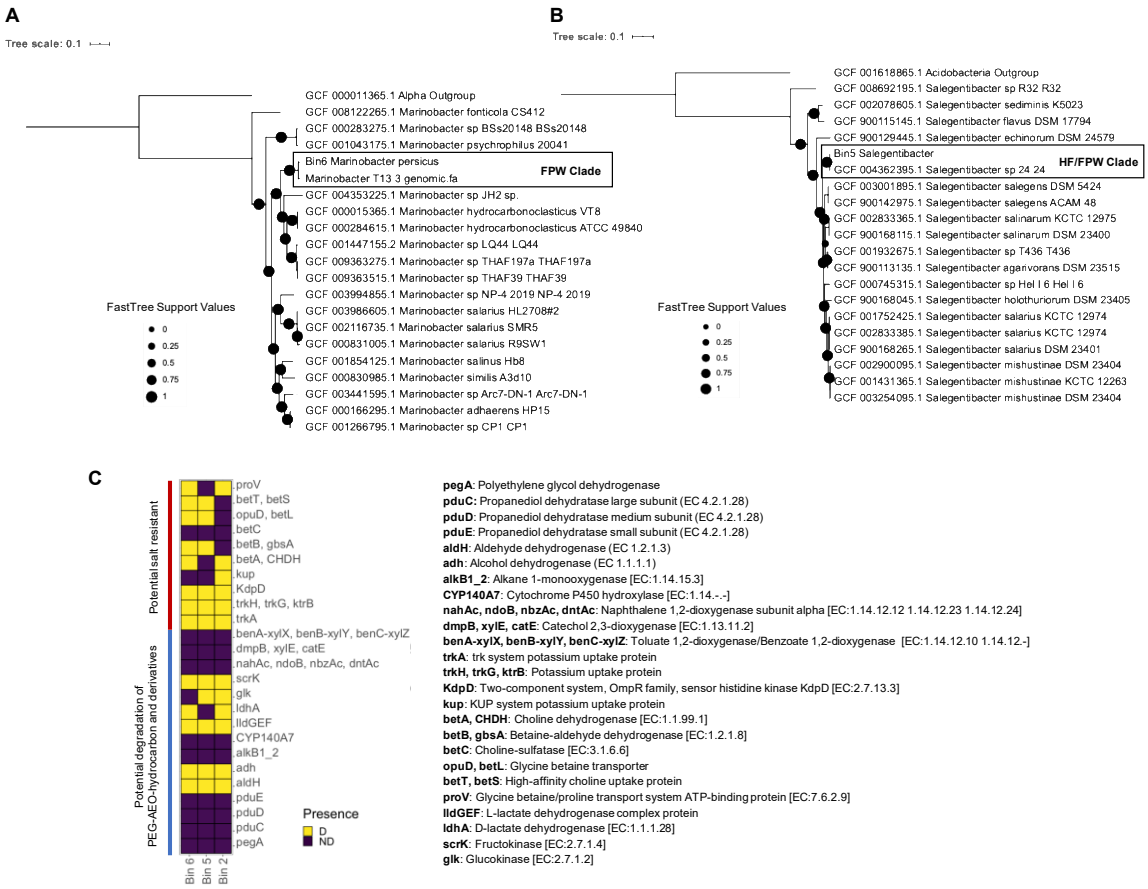

Figure 3.4 Assembled genomes clades within bacterial genera *Marinobacteria*, *Erythrobacter*, and *Salegentibacter* and functional annotation for high-quality bin 6, bin 5, and bin 2. (A) FastTree phylogenomic tree computed using a concatenated alignment using 172 Gammaproteobacteria.hmm genes for bin6 and (B) FastTree phylogenomic tree computed using a concatenated alignment using 90 Bacteroidetes.hmm genes for bin5 (C) Binary heatmap (D: detected, ND: not detected) of the essential genes may be involved in PEGs, AEOs, hydrocarbon degradation, and their derivative compounds and essential genes that may be involved in salt resistant of the FPW effects. Full lists of the examined genes for bin 2, 5, and 6 related to glycine betaine and K^+ based metabolisms were provided in Appendix 2 Table S5. The annotation of the bin 2 is used MetaErg due to insufficient bin quality required for IMG annotation.

3.4.5. Temporal Changes in Soil Respiration Activity

Increased levels of FPW inhibited CO_2 production and O_2 consumption for both soil types (Figure 3.5). The respiration decrease is likely to be caused by loss of viable cells, an effect that was previously reported for freshwater communities in our former study (Zhong *et al.* 2020). This indicates that the increase in relative abundance of bacteria such as *Marinobacter* may be due to their persistent and the loss of other organisms. The changes, caused by increasing levels of FPW, were significant (ANOVA, $p<0.001$) for both soils. The respiration capacity of unexposed GP chernozem is higher than for FC luvisol; however, this difference was reduced at increased FPW concentrations. This is consistent with higher nutrient in GP chernozem, less loss of biodiversity, and initially higher functional genes that may be related to organic degradation, suggesting that GP chernozem may have a greater capacity to resist FPW-induced stress.

The aerobic respiration rates with either 5% or 50% FPW treatment did not recover to the respiration levels observed in non-FPW groups during the incubation period, indicating that the ecological impact caused by FPW may not be mitigated by microorganisms over a short period of time. The degree of FPW exposure significantly impacted CO_2 generation over the month of incubation (ANOVA, $p<0.001$). Total CO_2 release from FC luvisol and GP chernozem exposed to 5% FPW was about 1.5-fold less than in the natural soils. The release was 3-fold less in FC luvisol and 2-fold less in GP chernozem with exposure to 50% FPW (for details and Tukey HSD analyses see Appendix 2 Table S6). Our results imply declines in biogeochemical cycling efficiency in FPW-amended soils.

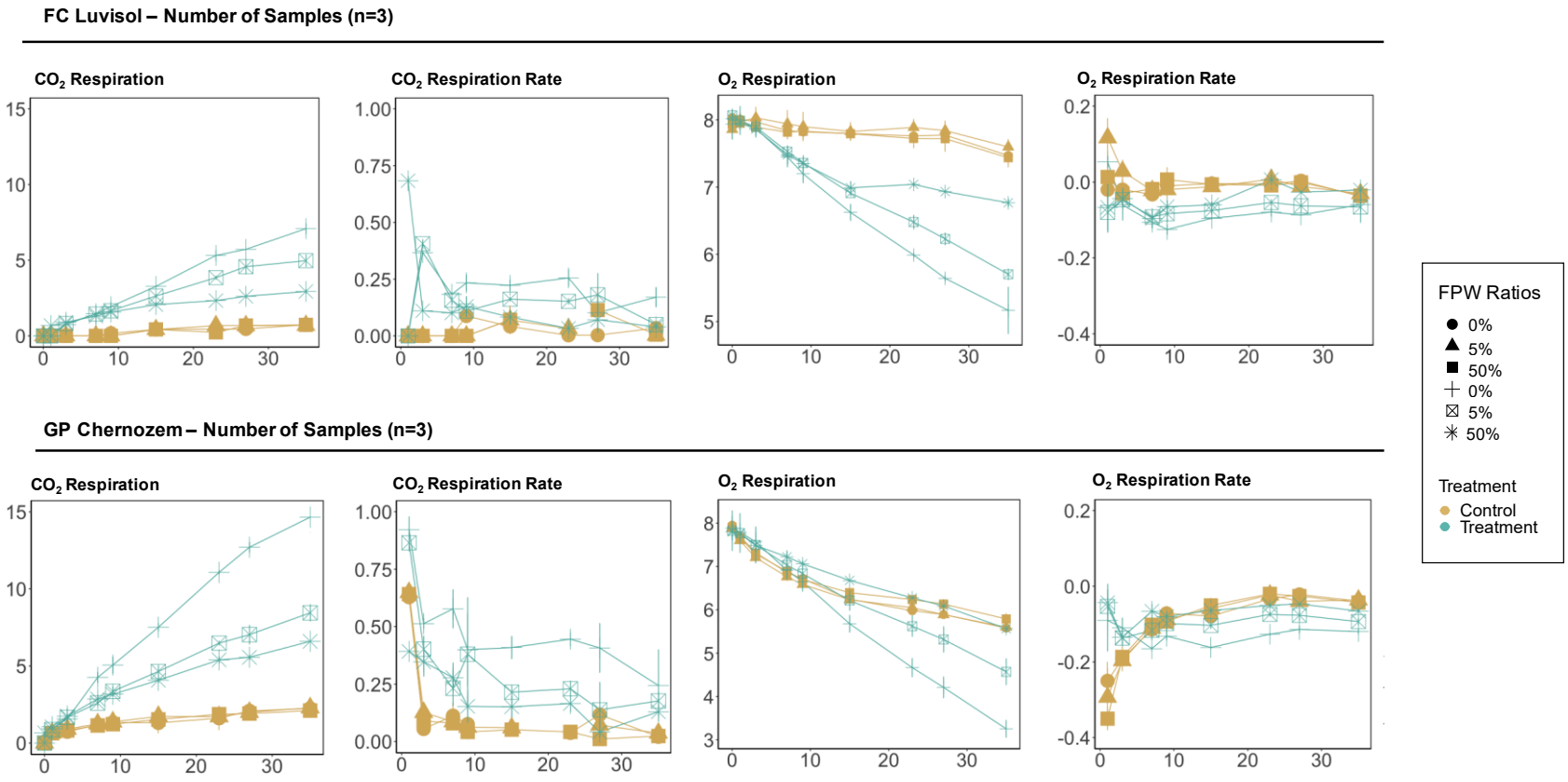

Figure 3.5 Soil respiration (measured as CO_2 production and O_2 consumption) in Fox Creek luvisol and Grand Prairie chernozem soils exposed to 50% FPW over a month of incubation versus control groups. For each soil and each gas, the increase/decrease of total concentration is shown as well as the daily change rate.

3.5. Environmental Implications

Soil microbiota are expected to mitigate FPW spills and to prevent further transport to groundwater or other nearby environments. Our study suggests that soil microbial community have relatively strong resistant to FPW-induced stress. Heavy FPW exposure may impede geochemical cycling and biodiversity. In this case, shifting community structure and consequent functional capacity made microbial communities are key player for soil to response the high FPW-induced stress. Soil types were vitally important to influence above mentioned responses. Future studies need to link these observed responses to specific FPW compounds, including longer time observation of the microbial response to recalcitrant compounds of FPW. This knowledge promises to enhance our understanding of the impacts of FPW spills on different soil systems, providing essential information to develop effective regulations to mitigate FPW risks in regions with various types of soils, to develop risk assessment strategies for FPW pollutants, and for optimizing practises for bioremediation.

3.6. Statement of Contribution

In this study, sample preparation, respiration tests, 16S rRNA gene bioinformatics, data analyses, and paper writing were conducted by me. Samples were sent to Natural Resources Analytical Laboratory at University of Alberta for chemical characterization. Dr. Camilla L. Nesbø assisted me with metagenome bioinformatics. Non-target organic analyses were conducted by Dr. Yifeng Zhang in the Laboratory Medicine & Pathology of the University of Alberta.

4. Chapter Four: Comparative Metagenomics Uncover Distinct Shale Microbiome in Deep Fractured Subsurface [4]

4.1. Summary

Multiple-stage hydraulic fracturing has been applied widely in North America and recently extends to China for unconventional resource recovery. The activities of microbes in the deep fractured subsurface have economic and environmental implications. Here, we investigated microbial and geochemical consistency in flowback and produced water derived from major shale basins in China, the United States, and Canada. Our results revealed two distinct shale ecosystems: a low diversity microbial community in higher salinity samples from North America dominated by known halophiles, and a high diversity community in lower salinity samples from China. While the capacity for sulfidogenesis and methanogenesis were common to the two regions, the organisms responsible for these processes and the mechanisms and environmental constraints in samples from China differed from North American samples. Our study suggests that the effective management for fractured subsurface microbial communities may differ between China and North America during hydrocarbon production.

---

[4] This Chapter is written based on a complete manuscript: **Cheng Zhong**, Mikayla A. Borton, Camilla L. Nesbø, Fu Chen, Malcolm D. Forster, Liaozi Han, Greg G. Goss, Kelly C. Wrighton, Brian D. Lanoil, Daniel S. Alessi

4.2. Introduction

The terrestrial subsurface is a vital biomass reservoir on Earth, estimated to host 2 to 6 × 10^{29} cells (Magnabosco *et al.* 2018). This large biomass plays a critical role in driving elemental cycling, producing by-products such as methane that can be harvested for energy generation (Head, Jones and Larter 2003). Horizontal drilling combined with multiple-stage hydraulic fracturing (HF) has enabled hydrocarbon extraction from typically low permeability shale formations and results in the linking of subsurface and surface microbial ecosystems (United States Environmental Protection Agency 2016; Alessi *et al.* 2017). On average, 15,000 m^3 of freshwater is injected per well into geologic formations that lie kilometers below the surface to create fractures, and HF fluid combined with formation brine return to the surface as flowback and produced water (FPW) after the fracturing operation is completed (Kondash and Vengosh 2015). Understanding the activities of microbes in these deep terrestrial ecosystems, and how the process of HF impacts them, promises to provide insights that have economic and environmental importance.

Concerns about undesirable microbial activities during shale resource recovery include biofouling and generation of biogenic sulfide that may cause reservoir souring and infrastructure corrosion (Booker *et al.* 2017; Nixon *et al.* 2017). Early studies that investigated FPW from unconventional resource basins in North America showed that HF introduces microorganisms from surface environments that may inhabit fractured shales, and that chemical additives in the injected HF fluids may support microbial growth and colonization (Daly *et al.* 2016; Evans *et al.* 2019a). Studies to date from the Marcellus and Utica (US) and Duvernay (Canada) shale formations revealed a persistent, stable microbial community dominated by members of the sulfidogenic bacterial genus *Halanaerobium*, with methanogens of the archaeal genus *Methanohalophilus* also detected, albeit at lower abundance (Murali Mohan *et al.* 2013a; Cluff *et*

al. 2014; Daly *et al.* 2016; Zhong *et al.* 2019). The prevalent microbes may use chemical additives as carbon sources (Mohan *et al.* 2014; Booker *et al.* 2017; Lipus *et al.* 2017; Borton *et al.* 2018a; Booker *et al.* 2019). For example, glycolysis and pyruvate metabolisms are often cited as primary fermentation pathways for *Halanaerobium congolense* to degrade FPW dissolved organic matter (e.g. ethylene glycol, guar gum, to sugars) to corrosive short-chain acids (Booker *et al.* 2017; Lipus *et al.* 2017; Booker *et al.* 2019). Glycine betaine pathways are important for microorganisms such as *Halanaerobium* and *Methanohalophilus* to to obtain energy or form an osmoprotectant to counter high salinity in the Marcellus and Utica (Daly *et al.* 2016; Borton *et al.* 2018b, 2018a). Beyond North America, China has become the third-largest producer of shale gas. In China, 95% of the country's total shale gas production is from the Sichuan Basin (Chinese National Bureau of Statistics 2019). The recent study showed that methanogens co-existing with sulfate-reducing microbes present in the Sichuan Basin fractured shale (Zhang *et al.* 2017, 2020). However, knowledge of the deep subsurface ecosystem in the Sichuan Basin shales, as well as its genotype consistency to that for North America is poorly understood.

Here we present the first metanalyses of metagenomes in FPW derived from shale formations at basin scales. The studied locations included the Sichuan Basin (China), the Marcellus and Utica (US), and the Duvernay (Canada) shale development regions. The Duvernay Formation is of Devonian age and is laterally equivalent to the Leduc Formation, a carbonate platform and reef complex (Flynn *et al.* 2019). The sampling locations of the Sichuan Basin FPW were in Weiyuan and Zhaotong shale gas fields, which have the production shale layer of the Upper Ordovician Wufeng Formation and Lower Silurian Longmaxi Formation, deposited in the continental shelf (Feng *et al.* 2018). The Marcellus and Utica Formations are located in the Appalachian Basin. The organic-rich black shales of the Devonian-aged Marcellus Formation and

Upper Ordovician-aged Utica Formation were deposited in foreland basin (Zagorski, Wrightstone and Bowman 2012; U.S. Energy Information Administration 2017).

The target formations have vertical depths ranging from 2-4 km. We analyzed geochemical, 16S rRNA gene amplicon, and metagenomic data from eight samples collected from FPW derived from the Sichuan Basin (China). We compared these results to reanalyzed data collected from the North American shale development regions. We used 16S rRNA gene sequences to compare microbial community structure in FPW from China and North America, and gene-centric and genome-resolved metagenomic approaches to determine inferred functions encoded by genomes of the shale microbial communities. These comparative analyses aim to enhance the understanding of global relevance of phylogeny and functionalities of the microbial communities of the fractured subsurface and their potential impacts on resource recovery and to the environment.

4.3. Methods

4.3.1. Samples Collection for Comparative Analysis

FPW and input water, used to make HF fluids, were collected from four shale development regions in China, US, and Canada, including the China's Sichuan Basin shale gas development region, Utica and Marcellus shale gas development regions, US, and the Duvernay shale (tight sandstone) oil and gas development region, Canada. The 16S rRNA gene amplicon dataset included 8 FPW and 1 input water samples from the Sichuan Basin, 8 FPW and 2 input water samples from the Duvernay, 34 FPW and 1 input water samples from the Marcellus, and 20 FPW and 5 input water samples from the Utica. The shotgun metagenome dataset includes 2 FPW samples from the Sichuan Basin, 7 FPW samples from the Marcellus, 20 FPW samples from the

Utica, and 5 input water samples from Utica. DNA concentrations in FPW collected from the Duvernay Formation were insufficient for the metagenomic sequencing.

The Duvernay formation FPW samples were collected from gas–fluid separators from two shale wells near the town of Fox Creek, Alberta, Canada. The subsequent 16S rRNA sequencing, salinity, and DOC concentrations were presented in a previous study (Zhong *et al.* 2019). Input water and FPW of from shale wells in the US were collected from wellheads and gas–fluid separators. These fluids were collected from five wells in the Utica and Marcellus shales in Ohio (n = 2), West Virginia (n=2), and Pennsylvania (n=1). DNA was extracted and sequenced from all produced and input fluids, as previously described in a previous study (Daly *et al.* 2016).

This study is the first reported metagenomics and geochemistry of the FPW samples from the Sichuan Basin shales. In China, eight FPW samples collected from between 75-156 days after the initial flowback were collected in 2018 from gas–fluid separators of five shale wells from the Weiyuan shale gas play (n=3) and Zhaotong shale gas play (n=2) in/near the Sichuan Basin. There was no shut-in time after HF operations were completed for all sampled wells in the Sichuan Basin. Additionally, an input water sample from the freshwater storage tank was collected from the Weiyuan shale gas play. Fluids were collected in 500 mL sterile polypropylene containers without headspace and transported on dry ice from the production sites to the Southwest Petroleum University, Chengdu, China, within a week. The fluid samples were filtered through 0.22 μm pore size hydrophilic polypropylene membranes and stored at -20ºC until DNA extraction.

4.3.2. Geochemical Analyses

Chemical analyses of Sichuan Basin shale FPW were conducted at the Southwest Petroleum University in Chengdu, China. Briefly, pH was measured by standard methods. The

chemical oxygen demand (COD) of water samples was measured using a spectrophotometer (COD-571-1, INESA, Shanghai, China) and converted into TOC using the method described in Dubber and Gray (2010) (Dubber and Gray 2010). The concentrations of cations including Ca^{2+}, K^{+}, Na^{+}, Mg^{2+}, and Sr^{2+} were analysed using an atomic absorption spectrophotometer (AA-7020, EWAI, Beijing, China), and concentrations of the anions F^{-}, SO_4^{2-}, NO^{3-} and Cl^{-} were measured using ion chromatography (883, Metrohm, Herisau, Switzerland) following the standard protocols. Corresponding chemical analyses for Utica and Marcellus fluids are outlined in previous studies (Daly *et al.* 2016; Borton *et al.* 2018a). Additional DOC data were collected from a previous study (Cluff *et al.* 2014). Additional descriptions of the methodology are presented in Appendix 3.

4.3.3. DNA Extraction, PCR, and 16S rRNA Gene Sequencing from Sichuan Basin Fluids

Total genome DNA from samples was extracted using the combined cetyl tri-methyl ammonium bromide (CTAB) and sodium dodecyl sulfate (SDS) method. 16S rRNA genes were amplified using 341F-806R primers (5′-CCTAYGGGRBGCASCAG-3′, 5′-GGACTACNNGGGTATCTAAT-3′) targeting V4 regions (Yu *et al.* 2005). Sequencing libraries were generated using TruSeq DNA PCR-Free Sample Preparation Kit (Illumina, California, USA) following the manufacturer's recommendations. The library quality was assessed using a Qubit 2.0 Fluorometer (Life Technologies, California, USA) and an Agilent 2100 Bioanalyzer system (Agilent Technologies, California, USA). Finally, the library was sequenced on an Illumina HiSeq platform (Illumina, California, USA) and 250 bp paired-end reads were generated as recommended by the manufacturer.

4.3.4. Bioinformatics for 16S rRNA Datasets

For samples from Marcellus and Utica, 16S rRNA gene sequences were reconstructed from trimmed unassembled Illumina reads using EMIRGE v 0.61.0 with 50 iterations. EMIRGE sequences were chimera checked before phylogenetic gene analyses. To make EMIRGE 16S rRNA gene data comparable to 16S rRNA gene amplicon data, full-length EMIRGE 16S rRNA gene sequences were trimmed to the V4 amplified region and ran through an in-house pipeline to create QIIME2 input files from EMIRGE data. This pipeline uses the NormPrior value calculated by EMIRGE to generate seqs.fna files for QIIME2 with the number of EMIRGE sequences reflected by NormPrior abundance with each sample having 1,000,000 sequences. For instance, if the NormPrior was 90%, it would have 900,000 sequences in the seqs.fna file, and 90% relative abundance in the QIIME2 ASV table. The seqs.fna files are directly imported into QIIME2. The methods for 16S rRNA sequencing of samples from Duvernay Formation is presented in Zhong et al. (2019) (Zhong *et al.* 2019). For 16S rRNA gene analyses from amplicon (Sichuan Basin and Duvernay) and metagenomic data (Marcellus and Utica), reads and seq.fna files were imported into QIIME2 v2018.11. Amplicon reads were demultiplexed and analyzed using DADA2 to produce amplicon sequence variants (ASV) by sample. EMIRGE-derived seq.fna files were converted to a biom file. Subsequently, both the EMIRGE-derived biom file and seqs.fna were imported into QIIME2. Both 16S rRNA gene analyses from amplicon and metagenomic data were classified using SILVA132, and then merged. The resulting ASV table was collapsed to the genus level and analyzed for biogeographical features.

4.3.5. Genome Assembly, Binning, Taxonomic Classification, and Functional Annotation

For metagenomic sequencing of FPW from Sichuan Basin shale, sequencing libraries were generated using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina (New England Biolabs, Massachusetts, USA) following the manufacturer's recommendations, and paired-end sequenced on an Illumina HiSeq platform (Illumina, California, USA). The raw data was cleaned, trimmed, and quality checked before assembly (Appendix 3). Reads passing quality control were assembled into contigs using the MEGAHIT assembler v1.1.2 with default settings (Li *et al.* 2015). The assembled reads were uploaded to Integrated Microbial Genomes and Microbiomes (IMG) in order to conduct gene-centric based approaches. The metagenome sequences assigned genes were classified using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways via KEGG Ortholog (KO). This is used to evaluate a broad-scale functional potential of the microbial communities detected from Sichuan Basin to that detected in Marcellus and Utica. Concurrently, raw reads from both wells were also processed using the Anvi'o (v6.2) pipeline's Metagenomic workflow (Eren *et al.* 2015), which relies on read quality control methods proposed by Minoche et al. (2011) (Minoche, Dohm and Himmelbauer 2011) as well as the MetaSPAdes v3.12.0 assembler (Nurk *et al.* 2017). MetaSPAdes was used to assemble reads from each of the two wells individually, as well as to generate a co-assembly of reads from both wells. Read mapping to assembled contigs was performed using BBMap v37.24 (https://sourceforge.net/projects/bbmap/) for MEGAHIT-assembled contigs and Bowtie v1.1.2 (Langmead and Salzberg 2012) for MetaSPAdes-assembled contigs. Contigs >2500 bp from each assembly were binned using MaxBin v2.2.7 (Wu *et al.* 2014) and MetaBAT v1.7 (Kang *et al.* 2015) with default parameters. Constructed genome bins from all combinations of assembly and binning software were pooled and dereplicated with dRep v2.3.2 (Olm *et al.* 2017) using default settings. Dereplicated bins were

assessed for quality using CheckM v1.1.0 (Parks *et al.* 2015) then assigned to taxonomy using GTDB-Tk v0.2.2 (Chaumeil *et al.* 2019). Bins with >90% completeness and <5% contamination scores with relatively less contigs were selected for further manual refinement (Appendix 3). The phylogenetic tree was reconstructed by using GtoTree based on Universal Hug et al. single copy gene set (Hug *et al.* 2016; Lee and Ponty 2019). The refined bins were annotated using the DRAM (Distilled and Refined Annotation of Metabolism) pipeline with default parameters (Shaffer *et al.* 2020). The annotation is used to predict the functional capacity of the reconstructed MAGs.

Fluid samples from Marcellus and Utica shales that were used for metagenomics analyses were sequenced at the Joint Genome Institute. Raw reads from FPW and input water were trimmed in IMG/M by the Joint Genome Institute. Assemblies were subsequently annotated using the IMG Annotation Pipeline. The genome bins of high abundance from the Utica and Marcellus which the binning and annotation methods were previously described (Daly *et al.* 2016; Borton *et al.* 2018b), were used as references for phylogenetic and functional analyses.

4.3.6. Statistical Analyses

Microbiology data and geochemistry data of FPW samples were grouped by Sichuan Basin, Marcellus, Utica, and Duvernay to support the analyses at regional scales. Salinity of these four studied shales was measured by chloride and TDS. The chloride levels of different shales were compared using ANOVA analysis combined with and TukeyHSD analysis. The differences in microbial diversity and richness among the four studied shales were measured by Shannon diversity, Inverse Simpson diversity, and Observed ASV indexes, using diversity function implemented in R. ANOVA analysis combined with and TukeyHSD analysis were used to compare these diversity and richness indexes. The diversity indexes were correlated to the salinity.

The methods of the correlation analyses are presented in Appendix 3. Principal Coordinates Analysis (PCoA) was used to measure the similarity of shale taxonomy of Sichuan Basin, Marcellus, Utica, Duvernay FPW samples, and their corresponding input samples. Besides, PCoA analyse was used to measure the similarity of functional gene composition (KO-based KEGG pathways) between FPW samples of Sichuan Basin, Marcellus, and Utica, and Utica input water samples. Permutational Multivariate Analysis of Variance (PERMANOVA) was used to test significance of two factors – the location of a well and the time a sample was collected following initial flowback – on the taxa distributions. Spearman correlation implemented in the R package, Vegan, was used were used to correlate specific microorganisms to overall taxonomic differences (Oksanen *et al.* 2018). The relative abundance of the estimated gene copy numbers classified to KEGG level 1 and level 2 functional pathways for FPW samples of Sichuan Basin, Utica, and Marcellus and Utica input samples were compared using ANOVA analysis and TukeyHSD analysis. The Kruskal-Wallis Rank Sum Test implemented in IMG was used to compare the differences in KEGG functional modules (a series of genes) and KO based functional genes among. For ANOVA, TukeyHSD, and Kruskal-Wallis Rank Sum Tests, $p<0.05$ was used to infer a statistical significance.

4.3.7. Data Availability

For newly generated sequences, 16S rRNA gene sequences of Sichuan Basin FPW and input water samples were deposited in NCBI under BioProject PRJNA628814 (SAMN14751469-SAMN14751477). The shotgun metagenomes of two Sichuan Basin FPW samples were deposited with IMG accession numbers 3300031260 and 3300031485. High quality MAGs derived from Sichuan Basin assembled sequences were deposited in IMG under GOLD Studies Gs0135899.

For reanalyzed sequences, 16S rRNA gene sequences of Duvernay samples were deposited in NCBI under BioProject PRJNA407226 (SRX3204453, SRX3204457, SRX3204459-SRX3204466), 16S rRNA gene sequences of Marcellus samples were deposited in NCBI under BioProject PRJNA308326 (SAMN04417539, SAMN04417540, SAMN04417544-SAMN04417546). The assembled sequences of Utica and Marcellus FPW samples were deposited in IMG (3300006628-3300006633, 3300006639, 3300006780, 3300006782, 3300006798, 3300006807, 3300006866, 3300007157, 3300007158, 3300007162, 3300007165, 3300009194, 3300009419, 3300009574- 3300009576, 3300009625, 3300009744, 3300010372, 3300010374, 3300013015-3300013021). The deposit access numbers of the MAGs derived from Utica and Marcellus assembled sequences were presented in previous studies (Daly *et al.* 2016; Borton *et al.* 2018a).

4.4. Results and Discussion

4.4.1. Reconstruction of Datasets for Metanalyses of Persisting Shale Genomes from Sichuan Basin, Marcellus, Utica, and Duvernay Shales

The 16S rRNA gene-based datasets comprises of eight Duvernay (0.04-18 days following HF) (Zhong *et al.* 2019), 34 Marcellus (7-488 days following HF), and 20 Utica (9-302 days following HF) FPW samples on which previously reported (Cluff *et al.* 2014), and eight newly-collected FPW samples from the Sichuan Basin (75-156 days following HF) (Figure 4.1A). At least one input fluid sample of each shale development region was included as a reference sample for measuring the compositional distance between two energy extraction phases (before and after HF). 16S rRNA gene sequences from Duvernay and Sichuan Basin were used 16S rRNA amplicon

sequencing while 16 rRNA gene sequences from Utica and Marcellus were extracted from metagenomes.

An earlier survey of seven Marcellus, 20 Utica FPW samples, five Utica input samples, and two newly collected day 82 and 156 Sichuan Basin FPW samples were chosen for paired shotgun metagenomics analyses. Duvernay samples are not included as they failed to generate DNA of sufficient quality for metagenomics sequencing.

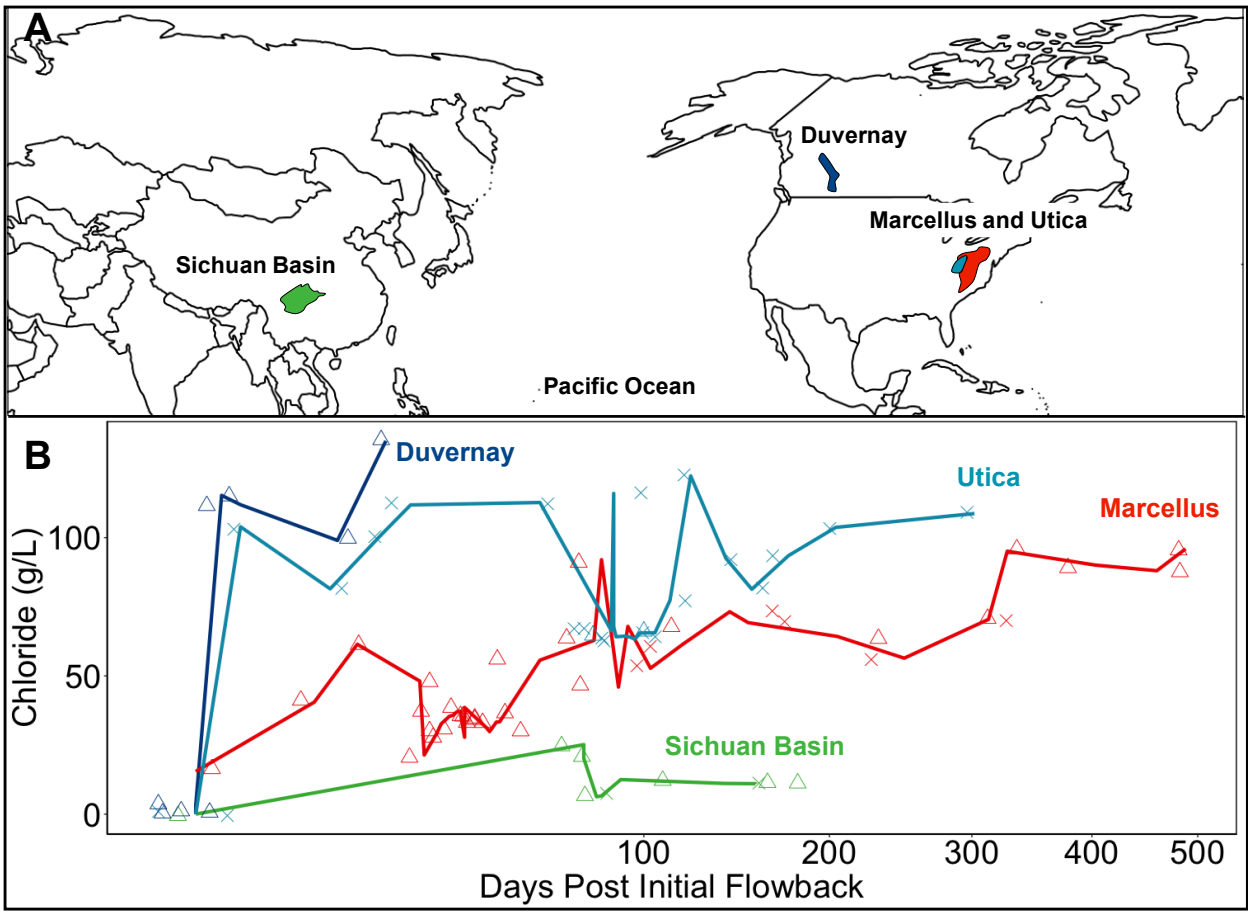

Figure 4.1 Reconstruction of persisting shale genome and salinity of Sichuan Basin, Utica, Marcellus, and Duvernay shale formations. (A) Locations of studied shale formations; Marcellus and Utica shales, Appalachian Basins (United States), Duvernay shale, Western Canadian

Sedimentary Basin (Canada), and Sichuan Basin shale (China) are the major active shale oil and gas plays in each country. (B) Genomic sampling over time from the Sichuan Basin shales of lower salinity to the North American Basin shales of higher salinity. Trends of chloride concentrations corresponded biosamples (jittered) showing they well-covered the elevated salinity in time series, representing the stable shale microbiome in the late stage of the extraction process. "Δ" represents samples used for 16S rRNA gene based on analyses; "×" represents samples used for metagenomic analyses and have paired 16S rRNA gene data. Each biosample has corresponding salinity data with exception of three (Utica) input samples without measurement.

4.4.2. Distinct Geochemical Characteristics of FPW Between the Sichuan Basin, Marcellus, Utica, and Duvernay

To support the metanalyses of persistent shale genomes, the pH, inorganic elements, and total organic contents of the eight Sichuan Basin FPW samples were measured, and were compared to the FPW samples previously reported from Utica, Marcellus, and Duvernay Formations (Flynn *et al.* 2019; Folkerts *et al.* 2019; Zhong *et al.* 2019; Mehler *et al.* 2020). The results showed that the salinities (measured as chloride) are significantly different (p<0.001) among the four formations. The chloride concentrations Sichuan Basin FPW samples were 13,000±6,500 mg L^{-1}, significantly lower (p<0.001) than the 53,800±21,600 mg L^{-1} of Marcellus samples, 88,800±21,900 mg L^{-1} of Utica samples, and 102,700±31,800 mg L^{-1} of Duvernay samples. The time following initial flowback contributed to most of the variance (p<0.001) in salinity observed among the samples (Figure 4.1B). This basinal salinity difference is also evident when comparing total dissolved solids (TDS) of Sichuan Basin FPW to those previously reported in other studies as an estimate of salinity (Appendix 3).

Correlated to the chloride levels, the concentrations of most cations in Sichuan Basin FPW samples (Appendix 3 Table S1) were lower than the ranges measured in FPW samples collected from North America (Haluszczak, Rose and Kump 2013; Oetjen *et al.* 2018; Flynn *et al.* 2019). The sulfate concentrations in Sichuan Basin FPW samples were between 47.4-127.4 mg L^{-1}, within the ranges observed in the Marcellus (Haluszczak, Rose and Kump 2013) and Duvernay shales (Flynn *et al.* 2019). The dissolved organic carbon (DOC) concentrations of Sichuan Basin FPW were 686 ± 396 mg L^{-1}, while DOC concentrations were 266 ± 126 mg/L for FPW from the Duvernay shale (Zhong *et al.* 2019), and 128 ± 66 mg/L for FPW from the Marcellus shale (Cluff *et al.* 2014).

The pH of Sichuan Basin FPW ranged between 6.9 and 8.3 (Appendix 3 Table S1), and was within the range of the Marcellus samples (5.1-8.4) (Barbot *et al.* 2013), and relatively higher compared to that for Duvernay samples (5.2-6.1) (Zhong *et al.* 2019). Subsurface temperature is known to strongly influence subsurface microbial taxonomic compositions (Gaspar *et al.* 2016). Previous studies indicate that the reservoir temperature of these target formations are within broad ranges of 20-200ºC, depending on the well locations (Gaspar *et al.* 2014; Taylor *et al.* 2014; Daly *et al.* 2016; Chen *et al.* 2018; Zhong *et al.* 2019). While, the reservoir temperatures of these wells studied in this study were not tracked. Distinct geochemical differences, especially the dramatically lower salinity in FPW from the Sichuan Basin as compared to the shales in North America is indicative of differences in the subsurface environments. Therefore, we postulated these differences in the physiochemical environment provided alternative environmental filtering of the subsurface microbial communities, resulting in different microbial community membership between Sichuan Basin and North American Basins. Whether functions change despite the community changes is of importance to understand.

4.4.3. Regionally-Distinct Shale Microbial Communities

Consistent with the geochemistry, our broad-scale 16S rRNA gene based genome survey revealed that diversity metrics and membership of the microbial communities in the Sichuan Basin were distinct from those detected in the North American Basins (Figure 4.2A). The microbial richness and diversity of the Sichuan Basin FPW samples were significantly higher (p<0.001) than that for the Marcellus, Utica, and Duvernay (Appendix 3). For instance, the Shannon diversity was 3.84 ± 0.28 for Sichuan Basin FPW samples and were 1.59 ± 1.40, 1.59 ± 1.05, 0.64 ± 0.75 for the Marcellus, Duvernay, and Utica FPW samples, respectively. The reduced biodiversity was correlated (polynomial regression, p<0.001, R^2=0.5-0.6) to increased salinity in North American versus Sichuan Basin FPW (Appendix 3 Figure S1, Appendix 3). We note that these differences in microbial diversity were not driven by different input biomass, as there were only small and not significant differences in diversity of input water among the studied shales. Additionally, diversity matrices remained significantly different between China and North American samples when we limited our analyses to persistent microbial communities, using samples produced > 50 days after initial flowback (Appendix 3).

The compositional differences (p=0.001, R^2=32%) were shown by clusters in the PCoA ordination, which were significantly separated by studied formations (Figure 4.2B). High abundances of members of the halophilic genus *Halanaerobium* in the Marcellus, Utica, and Duvernay shales on the right side of the ordination discriminated them (p=0.002, Appendix 3 Table S2) from the Sichuan Basin samples. This is also supported by a distinguished cluster observed in hierarchical clustering analysis (Appendix 3 Figure S2) where *Halanaerobium* consisted of >75% of the total community in the samples of this cluster. In this cluster of North American FPW samples, the lowest salinity at which *Halanaerobium* became predominant is >63,000 mg L^{-1}

chloride, found in Utica FPW at 96 days following HF. The relative abundance of *Halanaerobium* increased in Marcellus, Utica, and Duvernay samples with increasing FPW production time, which correlates to increasing salinity (Murali Mohan *et al.* 2013a; Cluff *et al.* 2014; Daly *et al.* 2016; Zhong *et al.* 2019). Eventually, the salinity levels stabilized, at which point the *Halanaerobium* were highly abundant in FPW from all of these plays (Murali Mohan *et al.* 2013a; Cluff *et al.* 2014; Daly *et al.* 2016; Zhong *et al.* 2019).

In the cluster of a group on the upper-left side of the PCoA ordination consisted of mostly Marcellus FPW samples and a few Utica FPW samples. *Arcobacter* (average 16.4%) and *Marinobacter* (average 4.15%) were abundant in the Marcellus FPW samples in this cluster. This could be associated with the abundance of these species in the original input samples. Average of the *Arobacter* and *Marinobacter* in Marcellus input samples were 5.70% and 6.21%, respectively. Thus, these clusters of samples collected from Utica and Marcellus shales may generally represent the transitional period of FPW (early to middle flowback times) before reaching the end of the trajectory, at which time *Halanaerobium* became predominant.

In the ordination, microbial community compositions in Sichuan Basin FPW samples are similar to input samples with less salinity. These results further suggest that salinity has a potent effect in determining microbial community structure in produced fluids. Correlated to the high diversity index, a wide range of microbes were detected across ten Sichuan Basin FPW samples (Appendix 3 Figure S3). This included relatively more abundant bacterial genera such as *Thermovirga*, *Sulfurospirillum*, *Sphaerochaeta*, *Desulfomicrobium*, *Bacillus*, *Arcobacter*, and *Marinobacterium*. The less abundant bacterial genera *Marinobacter* and *Shewanella* were detected and previously reported to be in surface HF water systems of the Sichuan Basin shale gas plays (Zhang *et al.* 2017). Methanogens including *Methanothermobacter* and *Methanolobus* were also

detected. The microbial community structure in the Sichuan Basin samples were similar to those found in conventional oil fields (e.g., *Thermovirga* and *Methanothermobacter*) in contrast to the North American FPW samples (Dahle and Birkeland 2006; Cheng *et al.* 2011). The community structure of corresponding samples for North American shale basins previously described in previous studies (Cluff *et al.* 2014; Daly *et al.* 2016). Prevalent halophilic members of *Halanaerobium* and *Methanohalophilus* found in Marcellus, Utica, and Duvernay, were not detected in Sichuan Basin samples, including the sample collected >150 days after initial flowback when salinity had stabilized at >10,000 chloride mg L^{-1}. *Halanaerobium* strains produce sulfide from thiosulfate reduction (Liang *et al.* 2016) and are the predominant bacteria in North American shales. In fact, our results show that more types of sulfide-producing bacteria are present in the low-salinity fractured shales (Appendix 3 Figure S3). The optimal grow salinity for strains of *Halanaerobium* isolated from oil and gas field sites is between 1.7-2.5M NaCl (approximately 60,000-90,000 mg/L Cl) (Oren 2015; An, Shen and Voordouw 2017; Booker *et al.* 2017), suggesting that the relatively low salinity of the Sichuan Basin shale may not provide an optimal salinity for the growth of *Halanaerobium*.

PCoA analyses reveal the effects of temperature on the shale microbial communities, in which distinct clusters may correlate to the detection of *Thermotoga* (average 0.53%) and *Thermococcus* (average 3.31%) in Utica FPW samples, and *Thermoanaerobacter* (p=0.001) in both Utica (average 0.49%) and Marcellus (average 1.25%) FPW samples on the right-down side of the PCoA ordination (Figure 4.2B, Appendix 3 Table S2). Previously characterized species of these genera are thermophilic and hyperthermophilic, and can grow at temperatures of up to 90°C (Klingeberg *et al.* 1995; Xue *et al.* 2001; Frock, Notey and Kelly 2010). Many enriched genera in

the Sichuan Basin samples, such as *Thermovirga* and *Methanothermobacter*, are also thermophilic microorganisms (Dahle and Birkeland 2006; Cheng *et al.* 2011).

Although major differences of the community structure from two regions is illustrated, it should be noted that the analyses of the microbial community structure are based on both 16S rRNA genes amplicon and 16S rRNA genes extracted from metagenome, which may cause estimation variance at finer scales. This variance has been demonstrated from the overall community structure distance (Figure 4.2B) and the major taxonomic compositions profiles (Appendix 3 Figure S3) of two paired Sichuan Basin FPW samples (Weiyuan 2_80 and Wiyuan 2_156). For example, *Thermovirga*, *Desulfomicrobium*, and *Bacillus* that detected in the 16S rRNA amplicon sequencing were not detected in their paired metagenome samples. This estimation variance is likely to be attributed to amplification bias of microbes of different abundance.

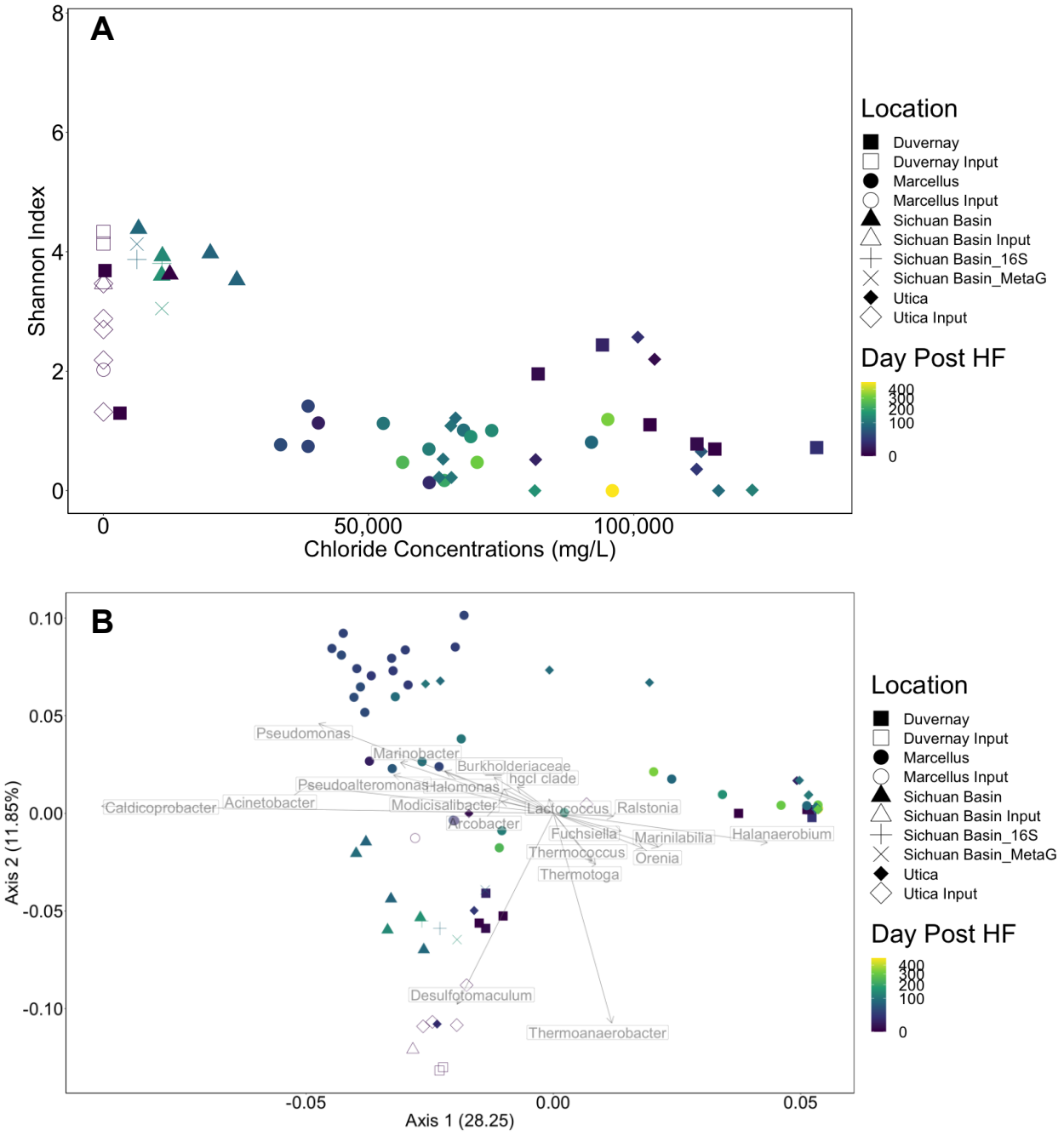

Figure 4.2 Metanalyses of diversity and taxonomy of the microbial community matrices between China and North American shale basins. (A) Microbial community diversity (measured as Shannon Diversity) between four studied shales versus and the diversity changes in increased salinity, (B) PCoA analysis (Bray-Curtis distance) showing taxonomic similarity between Sichuan,

Utica, Marcellus, and Duvernay samples. Samples collected from Sichuan Basin and Duvernay were obtained by 16S rRNA gene amplicon sequencing with exception labelled; samples from collected from Utica and Marcellus were obtained by extracts 16S rRNA gene from metagenome. Time post initial flowback were represented by color from dark blue to light yellow.

4.4.4. Function Capacity Conserved in Shale Microbiota Between FPW Samples from Sichuan Basin, and Marcellus and Utica (Appalachian Basin)

We examined the relative abundance of protein coding sequences of samples from the two regions of the Sichuan Basin and Appalachian Basin (Marcellus and Utica) to predict and compare the metabolic capacities of the shale microbiota between regions, under the context of distinct taxonomy revealed by 16S rRNA gene-based analyses. The PCoA analysis of KEGG Orthology based KEGG pathways showed FPW samples of Appalachian Basins have functional capacities generally different from the input samples with a few exceptions, and functional similarity of Sichuan Basin samples were in between Marcellus and Utica FPW and input samples (Figure 4.3, annotation of KEGG Level 1 and 2 pathways were presented in Appendix 3 Figure S4).

We compared relative abundance of estimated gene copy numbers associated with methanogenesis and sulfidogenesis that may be economically important to energy development. In terms of carbon cycling, unsurprisingly gene abundance of the fermentation pathways between two regions were in similar range. As an final step for converting carbon compounds to methane, methanogenesis has been noted as a vitally important biogeochemical process in organic-rich shales (Cokar *et al.* 2013; Mouser *et al.* 2016). The abundance of various genes that can convert methanol, CO_2, acetate, trimethylamine, dimethylamine, and methylamine to methane, and key methanogenesis genes Methyl–coenzyme M reductase complexes (*mcrABG*) were dramatically

higher in Sichuan Basin shale microbial communities than those found in the Appalachian Basin samples (Figure 4.3).

Along these lines, Sichuan Basin FPW metagenomes also showed a capacity to produce corrosive sulfidogenic products like their Appalachian Basin counterparts, consistent with increased sulfate (47.4-127.4 mg/L SO_4^{2-} in FPW samples) as an electron acceptor (Mouser *et al.* 2016). Assimilatory and dissimilatory sulfate reduction pathway gene modules, sulfate transporters (e.g. *cysA*, *cysW*, and *cysU*), as well as thiosulfate gene (e.g., *rdl*) detected in Sichuan Basin FPW samples were within the range or lower that of Appalachian Basin FPW samples (Figure 4.3). However, Sichuan Basin FPW contained significantly (adjusted p=0.004-0.010) higher abundance of *dsrAB* than Appalachian Basin FPW, which catalyzes the main energy-conserving step in the dissimilatory sulfate reduction pathway, resulting in metabolically produced sulfide. These broad scale pathway analyses suggest that sulfidogenesis and methanogenesis are core metabolic capacities conserved across shale microbial communities of these two regions.

Harsh subsurface environments plus biocides are not efficient in restricting the growth of potentially detrimental microorganisms in downhole environments in many cases (Kahrilas *et al.* 2015, 2016). Furthermore, we examined the biofilm and sporulation genes (Appendix 3 Figure S5), which have been reported in site-specific studies from Appalachian Basin that may cause negative impacts on the energy development (Mohan *et al.* 2014; Booker *et al.* 2017, 2019; Lipus *et al.* 2017). Genes involved in biofilm synthesis (*pgaAD*), quorum sensing (*qseBC*), and flagellin (*fliC* and *flrBC*), and chemotaxis motility (*motAB*), had little difference in abundance between samples of the two studied regions, with a few exceptions of genes involved in biofilm synthesis (*pgaC*), quorum sensing (*luxQN*/*cqsS*-*luxU*-*luxO*) and cellular attachment (*adrA*) where Sichuan Basin samples had higher abundance. Genes involved in sporulation (*kinABCDE-spo0FA*

complexes) were also within similar ranges of abundance. While these genes open the possibility for microbial communities to form biofilms or sporulate, may support the resilience of microbial communities to high temperature, salinity, pressure, and added biocides (Leggett *et al.* 2012), the examined biofilm-related genes were commonly absent in FPW samples; rather, sporulation related genes (spo0A) were shown prevalently enriched (adjusted p=0.005-0.025) in FPW samples as compared to those found in input samples.

4.4.5. No Enrichment of Genes Related to Salt-Tolerance in Sichuan Basin FPW

Glycine betaine and potassium uptake metabolisms are important for microbes to survive in the high salinity FPW of the Marcellus and Utica shales (Daly *et al.* 2016; Lipus *et al.* 2017). In general, most of the examined genes related to the two metabolisms in Sichuan Basin and Appalachian Basin shales were shown higher relative abundance than input samples (Appendix 3 Figure S5). However, lower salinity of Sichuan Basin FPW suggests that these pathways may not be used as equally important as found from Appalachian Basin. Consistent with the geochemical evidence, most of genes related to glycine betaine pathways such as glycine reductase (*grdA*), glycine betaine transporter (*opuD*, *betL*), glycine hydroxymethyltransferase (*glyA*), glycine dehydrogenase (*GLDC*), threonine 3-dehydrogenase (*tdh*), betaine reductase (*grdB*), and glycine betaine/proline betaine transport system (*proV*, *proW*, *proX*) examined in Sichuan Basin FPW were in a similar range or lower that those found in the Marcellus and Utica shales. Additionally, a series of *trk* system potassium uptake genes (*trkH*, *trkG*, *ktrB*, *trkA*) were also in a similar range between the two regions. The results indicate use potential of the glycine betaine and potassium uptake metabolisms may correlate to the salinity of fractured formation.

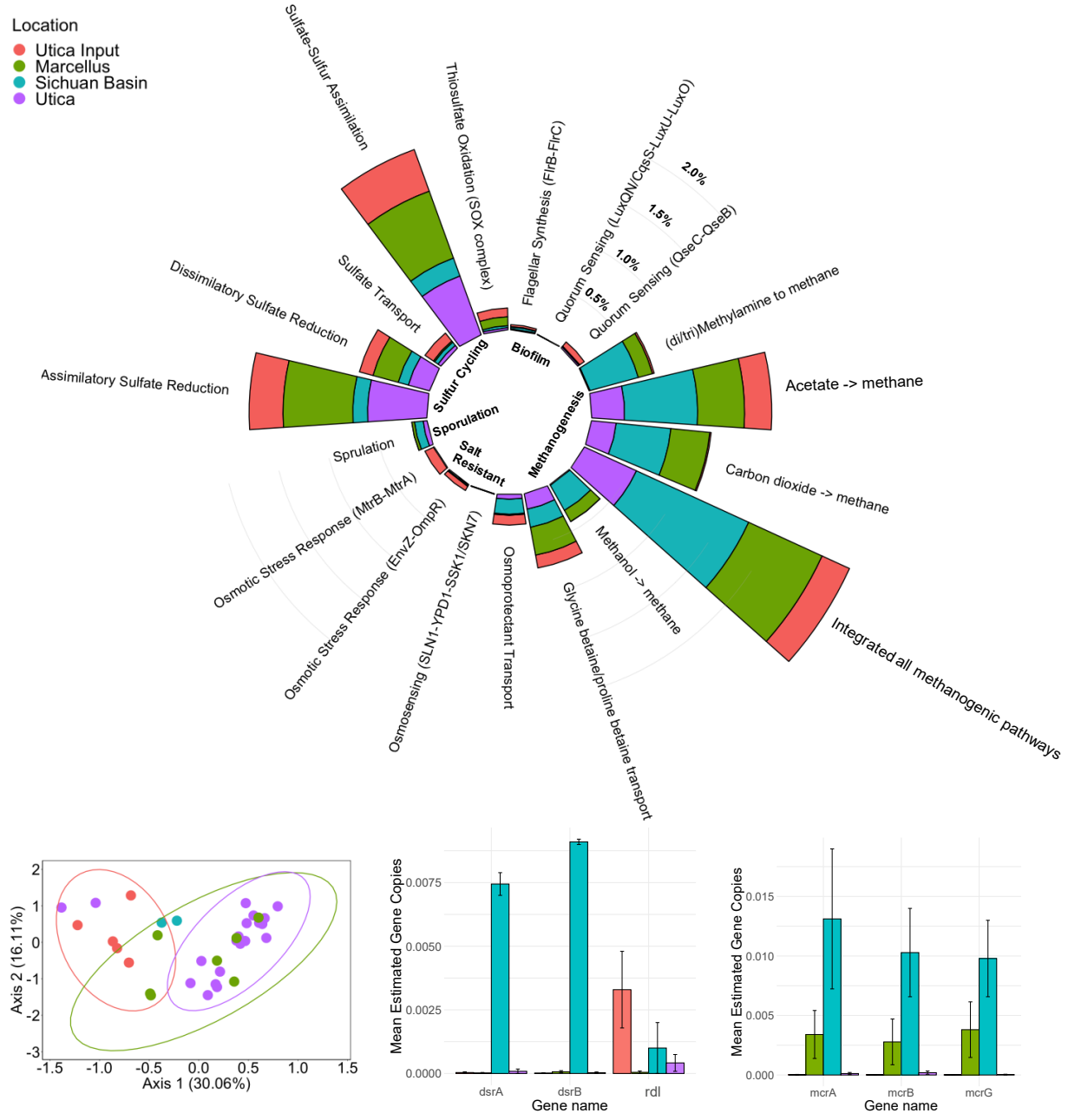

Figure 4.3 A broad-scale overview of the predicted functional capacities between Sichuan and Appalachian Basins. The relative abundance of the studied KEGG modules related to sulfur cycling, salt tolerance, biofilm formation, sporulation, and methane production. The PCoA ordination of KEGG pathway based on KO between the Marcellus, Utica, and Sichuan Basin shales, circles are 95% intervals. Abundance of key genes involved in sulfate-reducing (*dsrAB*),

thiosulfate-reducing (*rdl*), and methanogenesis (*mcr*) were shown to support the major finding of the broad-scale overview.

4.4.6. Phylogeny and Predicted Functional Capacities of Shale Metagenome-Assembled Genomes (MAGs)

The more diverse community in the Sichuan Basin samples may have more functional pathways to survive in the fractured subsurface, compared to the relatively low diversity community of the North America. To better anchor the predicted microbial functional capacities observed in our broad-scale metagenomic pathway analysis to specific microbial genomes (Hu *et al.* 2016; Probst *et al.* 2018; Boyd *et al.* 2019; Liang *et al.* 2020; Yan *et al.* 2020), we reconstructed metagenome assembled genomes (MAGs) from the Sichuan Basin shales and compared them to previously reported MAGs from the Appalachian Basin (Figure 4.4). 36 genome bins with medium to high quality (Bowers *et al.* 2017) were reconstructed (manually refined bins with low numbers of contigs are presented in Appendix 3 Table S3). MAGs derived from Sichuan Basin with relatively higher abundance have assigned to established genera closely to lineages detected from the conventional oil reservoirs (e.g., *Thermoanaerobacter*, *Methanothermobacter*, and *Mesotoga*) (Hu *et al.* 2016). In the Sichuan Basin samples, the genomes of high abundance include those classified to *Thermovirga lienii* (0.74-5.35%) and *Methanothermobacter thermautotrophicus* (0.16-5.38%), in contrast to the dominance of *Halanaerobium* and *Methanohalophilus euhalobius* in the Marcellus, Utica, and Duvernay samples. We note that bin 27 and bin 30 of the Sichuan Basin derived MAGs shared genus-level taxonomic assignment, *Methanolobu,* to a MAG from Appalachian Basin samples. Comparison of these genomes revealed 77-78% Average Nucleotide Identity (ANI) similarity between the Sichuan and Marcellus genomes (Daly *et al.* 2016).

Functions consistently present in the shale microbial communities through gene-centric analyses, such as fermentation, sulfidogenesis, and methanogenesis, were validated using these recovered MAGs (Figure 4.4). These metabolisms were major conserved processes in both regions, forming similar symbiosis in the fractured subsurface between regions despite the community structure are distinct (Appendix 3) (Langendijk *et al.* 2001; Copeland *et al.* 2009; Kaster *et al.* 2011; Miyazaki *et al.* 2014). While methanogenesis was conserved across China and North American FPW, binning revealed possible differences in substrate use preferences. Methanogen related bins from Sichuan Basin included both the capacity to use hydrogen (e.g. *Methanothermobacter thermautotrophicus*, *Methanobacterium fomicicum*) (Örlygsson *et al.* 1996; Wasserfallen *et al.* 2000), as well as methyl-C1 compounds (e.g. *Methanolobus vulcani*) (Kadam and Boone 1995). This is supported by the functional annotation shown in Figure 4.4. Bin 27 *Methanolobus vulcani* contained methanogenesis genes only via acetate to methane, but key methanogenesis gene *mcrA* was not detected.

In comparison, methylotropic and hydrogenotrophic pathways were detected in this study (Figure 4.4), supporting by previous studies showing Appalachian Basin MAGs were restricted to methyl compounds (e.g. *Methanolobus sp001577565 and Methanohalophilus euhalobius*) (Daly *et al.* 2016; Borton *et al.* 2018b). The possible broader methanogenic pathways were supported by a recent site-specific study showing that methylotropic, hydrogenotrophic and acetoclastic methanogenic pathways can be used to produce methane for microbial communities from the FPW of the Sichuan Basin shales (Zhang *et al.* 2020). In addition, our annotation is consistent with a previous study showed that methylotropic gene complexes (e.g., *mtmBC*, *mtbABC*, *mtaBC*, *mttBC*) were detected in *Methanohalophilus euhalobius*, but were absent in *Methanothermobacter thermautotrophicus* and *Methanobacterium fomicicum* (Evans *et al.* 2019b), further suggesting

differing preferences of substrate for methanogenesis in microbial communities of different regions.

Similarly, we recovered two MAGs (bin 28 *Desulfomicrobium orale* and bin 7 *Desulfomicrobium escambiense*) with the capacity for dissimilatory sulfate reduction only from the Sichuan Basin samples (Figure 4.4). These two bins harbored genes encoding sulfate reduction (e.g., *dsrAB*) and thiosulfate reduction functions (e.g., *rdl*). In contrast, sulfide production in the Appalachian Basin was derived from thiosulfate reduction from dominant, fermenting strains of *Halanaerobium* and others via contained thiosulfate-reducing genes (e.g., *rdl*, *phsA*, *psrA*) (Booker *et al.* 2017). Our study is consistent with previous studies showing that markers for sulfate-reducing potential such as the key sulfate-reducing gene *dsrAB* were not detected in the predominant sulfidogenic bacteria *Halanaerobium* in the Utica and Marcellus samples (Booker *et al.* 2017; Lipus *et al.* 2017). Consistently, a recent study showed no sulfate-reducing bacteria were detected through culturing, future suggesting that sulfate-reducing pathway may not a major metabolism used by microbes from Marcellus (Cliffe *et al.* 2020). This restricted substrate preference of methanogens and an absence of sulfate reducing capability observed in the Appalachian Basin relative to Sichuan Basin FPW metagenome and MAGs may be attributed to the increased salinity of the Appalachian FPW, as these metabolisms are reported to exclusively occur (methylotrophic methanogenesis) or be inhibited (hydrogenotrophic, acetoclastic, sulfate reduction) at elevated salinities, where increased energy generated may be required to offset the energy expenditure of salinity adaption (Oren 2011).

To further investigate the importance of glycine betaine pathways for Sichuan Basin FPW microbial communities, we examined relevant genes from the manually refined MAGs derived from Sichuan Basin FPW. Among the bins, only the bin classified to *Thermovirga lienii* encoded

relevant genes for glycine betaine pathways such as *grd* complexes (ABCDEHI), *tdh*, *gcvP AB*, and *opu ABC*, while, key enzymes specifically related to glycine betaine synthesis such as *betAB*, *gsdmt*, and *sdmt* were not detected.

The increased abundance of sporulation gene spo0A and starvation of certain microorganisms such as *Bacillus subtilis* in late-stage of FPW may trigger sporulation due to dilution of potential carbon sources (González-Pastor 2011; Zhong *et al.* 2019). We also examined spo0A genes in the manually refined bins and found the bin classified to *Bacillus subterraneus* harbored this key gene for sporulation. However, previous studies suggested that *Bacillus subterraneus*, the closet relatives of our MAG, isolated from a deep subsurface thermal aquifer was non-spore-forming (Kanso, Greene and Patel 2002). High salinity of the FPW may also inhibit the sporulation efficiency (Widderich *et al.* 2016). Thus, future laboratory-based experiments may be required to prove the hypothesis that microbes form spores in the fracture subsurface.

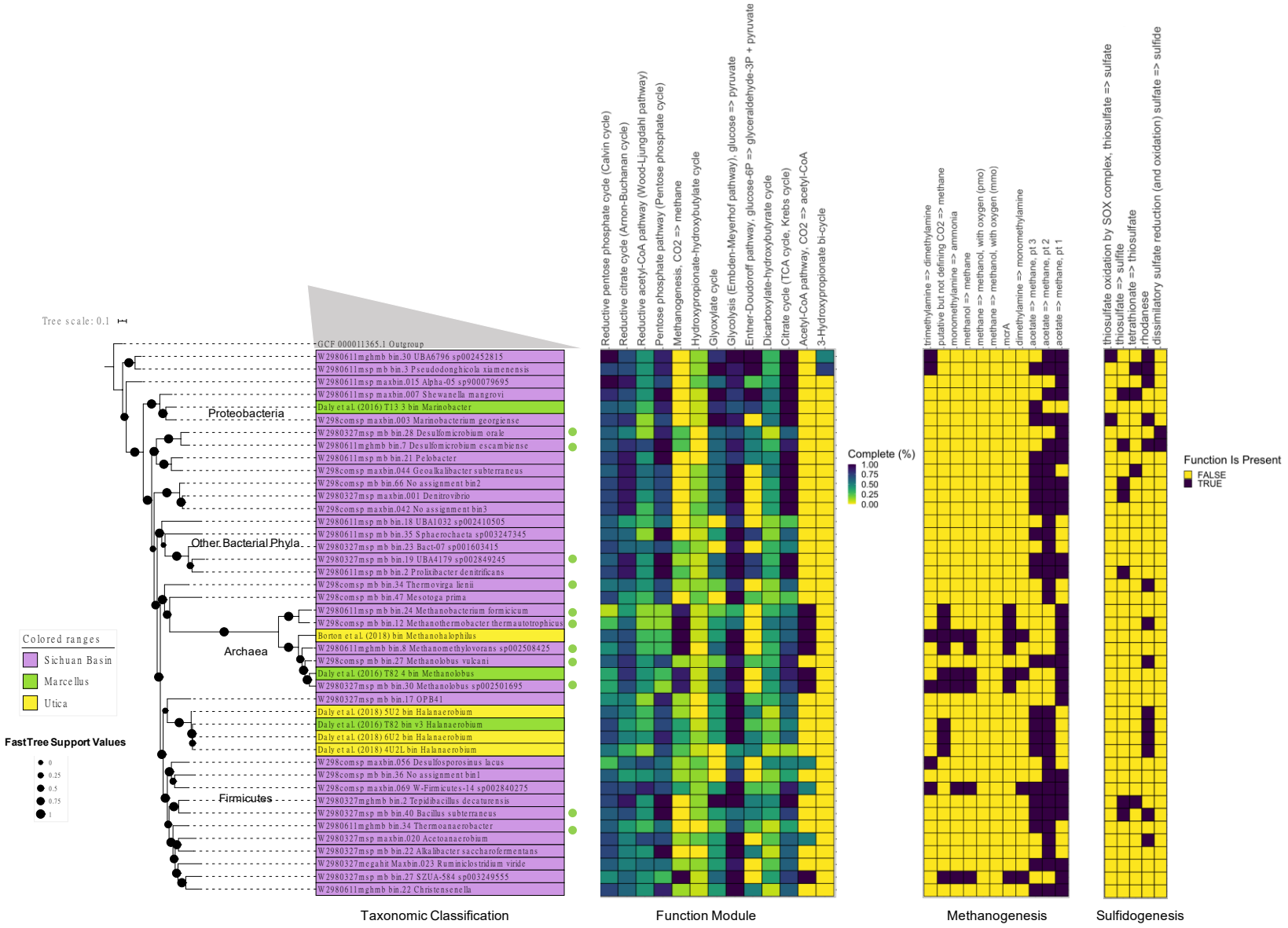
Tree scale: 0.1
GCF 000011365.1 Outgroup
W2980611mghmb bin.30 UBA6796 sp002452815
W2980611msp mb bin.3 Pseudodonghicola xiamenensis
W2980611msp maxbin.015 Alpha-05 sp900079695
W2980611msp maxbin.007 Shewanella mangrovi
Daly et al. (2016) T13 3 bin Marinobacter
W298comsp maxbin.003 Marinobacterium georgiense
W2980327msp mb bin.28 Desulfomicrobium orale
W2980611mghmb bin.7 Desulfomicrobium escambiense
W2980611msp mb bin.21 Pelobacter
W298comsp maxbin.044 Geoalkalibacter subterraneus
W298comsp mb bin.66 No assignment bin2
W2980327msp maxbin.001 Denitrovibrio
W298comsp maxbin.042 No assignment bin3
W2980611msp mb bin.18 UBA1032 sp002410505
W2980611msp mb bin.35 Sphaerochaeta sp003247345
W2980327msp mb bin.23 Bact-07 sp001603415
W2980327msp mb bin.19 UBA4179 sp002849245
W2980611msp mb bin.2 Prolixibacter denitrificans
W298comsp mb bin.34 Thermovirga lienii
W298comsp mb bin.47 Mesotoga prima
W2980611msp mb bin.24 Methanobacterium formicicum
W298comsp mb bin.12 Methanothermobacter thermautotrophicus
Borton et al. (2018) bin Methanohalophilus
W2980611mghmb bin.8 Methanomethylovorans sp002508425
W298comsp mb bin.27 Methanolobus vulcani
Daly et al. (2016) T82 4 bin Methanolobus
W2980327msp mb bin.30 Methanolobus sp002501695
W2980327msp mb bin.17 OPB41
Daly et al. (2018) 5U2 bin Halanaerobium
Daly et al. (2016) T82 bin v3 Halanaerobium
Daly et al. (2018) 6U2 bin Halanaerobium
Daly et al. (2018) 4U2L bin Halanaerobium
W298comsp maxbin.056 Desulfosporosinus lacus
W298comsp mb bin.36 No assignment bin1
W298comsp maxbin.069 W-Firmicutes-14 sp002840275
W2980327mghmb bin.2 Tepidibacillus decaturensis
W2980327msp mb bin.40 Bacillus subterraneus
W2980611mghmb bin.34 Thermoanaerobacter
W2980327msp maxbin.020 Acetoanaerobium
W2980327msp mb bin.22 Alkalibacter saccharofermentans
W2980327megahit Maxbin.023 Ruminiclostridium viride
W2980327msp mb bin.27 SZUA-584 sp003249555
W2980611mghmb bin.22 Christensenella
Proteobacteria
Other Bacterial Phyla
Archaea
Firmicutes
Colored ranges
Sichuan Basin
Marcellus
Utica
FastTree Support Values
0
0.25
0.5
0.75
1
Taxonomic Classification
Reductive pentose phosphate cycle (Calvin cycle)
Reductive citrate cycle (Arnon-Buchanan cycle)
Reductive acetyl-CoA pathway (Wood-Ljungdahl pathway)
Pentose phosphate pathway (Pentose phosphate cycle)
Methanogenesis, CO2 => methane
Hydroxypropionate-hydroxybutylate cycle
Glyoxylate cycle
Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate
Entner-Doudoroff pathway, glucose-6P => glyceraldehyde-3P + pyruvate
Dicarboxylate-hydroxybutyrate cycle
Citrate cycle (TCA cycle, Krebs cycle)
Acetyl-CoA pathway, CO2 => acetyl-CoA
3-Hydroxypropionate bi-cycle
Complete (%)
1.00
0.75
0.50
0.25
0.00
Function Module
trimethylamine => dimethylamine
putative but not defining CO2 => methane
monomethylamine => ammonia
methanol => methane
methane => methanol, with oxygen (pmo)
methane => methanol, with oxygen (mmo)
mcrA
dimethylamine => monomethylamine
acetate => methane, pt 3
acetate => methane, pt 2
acetate => methane, pt 1
Methanogenesis
thiosulfate oxidation by SOX complex, thiosulfate => sulfate
thiosulfate => sulfite
tetrathionate => thiosulfate
rhodanese
dissimilatory sulfate reduction (and oxidation) sulfate => sulfide
Sulfidogenesis
Function Is Present
FALSE
TRUE

Figure 4.4 Phylogeny and predicted functional capacities of shale metagenome-assembled genomes. Maximum likelihood phylogenetic tree calculated using FastTre v.2.1.11 with FastTree support values showing the relationships between medium-high quality MAGs from produced fluids, with purple, green, and yellow colors denoting shale source. The green circles are manually refined MAG with <100 contigs. The heatmap denotes the complete (%) of major functional modules and the presence and absence of the genes related to sulfidogenesis and methanogenesi

4.5. Environmental Implications

Our study illustrates that the genotype of persistent microbial communities inhabiting hydraulically fractured shales are distinct between regions, indicating that at least two unique ecosystems are present in the fractured deep subsurface. Microbial communities in the Sichuan Basin shale are more diverse compared to those found in the shales of the North American Basins. Relatives of more extreme halophiles and thermophiles were detected in the FPW samples collected from North American fractured shales. These patterns were substantially influenced by salinity and the temperature inherent to the formation geology (Zhong *et al.* 2019). Notably, many microbial genera detected in Sichuan Basin FPW have been found in conventional oil reservoirs, suggesting previous knowledge of microbiology of oil reservoirs can be inferred in the studies of these highly diverse microbial communities in FPW derived from the Sichuan Basin shale.

Our study provides a perspective from a large-scale view on the roles that microbial communities have in resource turnover, chemical degradation, and downhole infrastructure damage in the fractured subsurface. These microbial communities are effective indicators to reflect the subsurface geochemistry characteristics. Conserved functions of methanogenesis and sulfidogenesis between regions demonstrate a general similarity that microbial communities are key player of geochemical cycling in fractured subsurface. The biogenic methane as an additional supply to improve the energy yield. However, biogenic sulfide is of significant concern because it may lead to infrastructure damage, gas quality degradation, and health risks to on-site personnel (Lipus *et al.* 2017). Prevalent enriched genes such as sporulation gene indicate potential challenges in subsequent water treatment, recycling, transportation, and disposal of FPW. Despite this, the different pathways leading to sulfide and methane generation suggest that designing proper FPW

management that can adapt to formation geochemistry is essential to improve the efficiency of life cycle of HF water cycles as HF activities continue to grow.

4.6. Statement of Contribution

In this study, experimental design, sample preparation and 16S rRNA amplicon for samples from the Duvernay Formation, and meta-analyses were conducted by me. Sample preparation, 16S rRNA amplicon and metagenomics sequencing for samples from Sichuan Basin were conducted at the Southwest Petroleum University. Sample preparation, 16S rRNA amplicon and metagenomics sequencing for samples from Marcellus and Utica were conducted by Mikayla A. Borton of the Wrighton laboratory microbiome research group at Colorado State University. Metagenome 16S rRNA gene extraction, taxonomic classification, and MAGs annotation were conducted by Mikayla A. Borton.

5. Conclusions and Future Steps

Every ecosystem, from water, soils to the human body, is home to a diverse community of microorganisms that are not visible to the human eye. Understanding how microbes influence and mold their ecosystems is of significant interest in understanding elemental cycling, which can aid in addressing environmental concerns in activities such as energy recovery and pollution remediation. Hydraulic fracturing is the key technology in extracting hydrocarbon resources from low permeability formations. This technology forms a water cycle that connects the surface environment to the subsurface. One of the key environmental issues is management flowback and produced water (FPW) produced after hydraulic fracturing operations. FPW has a complex composition, including various hydrocarbons, chemical additives, and lithospheric elements. Understanding the microbial ecology in FPW is vitally important for assessing biofouling and biocorrosion issues in downhole environments. Microbes also have the potential to remediate FPW pollutants released to soil and water in the event of a spill, and spills themselves may impact the contributions of microbes to normal ecosystem functioning. Thus, this dissertation focused on microbial ecology and geochemistry in aquatic and soil environments under FPW-induced stress (Chapters 2 and 3), and regional-scale subsurface environments impacted by hydraulic fracturing (Chapter 4).

In Chapters 2 and 3, my results show that increasing FPW exposure levels caused reductions in biomass and respiration activity and shifts of microbial community compositions and diversity. Bacteria are sensitive to organic compounds derived from FPW and thus are effective indicators for >2.5% FPW contamination. For example, in experiments in aquatic environments, *Pseudomonas*, *Rheinheimera,* and *Brevundimonas* become dominant under FPW-induced stress (2.5%-5% by volume); *Marinobacter* in the soil environment experiments became dominant under

FPW-induced stress (FPW consists of >50% of the total exposure fluids). The results in Chapter 2 illustrate that machine learning techniques such as random forests modeling combined with high throughput sequencing are promising tools for selecting bioindicators at the FPW contaminated sites. The results in Chapter 3 show that luvisol soil of higher pH, larger particle size, and lower carbon, nitrogen, and moisture are more vulnerable to FPW-induced stress compared to chernozem soil. This effect is seen in greater losses in biodiversity and respiration activity in the luvisol soil. Persistent bacteria in FPW, such as *Marinobacter*, contained genes essential to the degradation of FPW-derived chemicals and that have the potential to resist salinity stress, essential to maintaining function in the soils. Thus, Chapter 3 suggests that chernozem has a considerable capacity to resist FPW exposure, and the soil microbiota can retain certain functions in susceptible soils exposed to even fairly high concentrations of FPW. Together, these two Chapters 2 and 3 suggest that both aquatic and soil microbial communities are likely to be influenced (e.g., changes in microbial community compositions and diversity), while soil microbiota may have greater resistance capacity than aquatic microbiota in general. Upon exposure to FPW, both soil and aquatic microbiota shifted to microbes that have potential to degrade FPW-chemicals within certain ranges of FPW exposure, thus these results indicated that bioremediation (e.g., natural attenuation) may be effective methods to remediate certain organic pollutants derived from FPW.

Chapter 4 compares shale microbial communities between the Marcellus, Utica, Duvernay (North America), and Sichuan Basin (China) and illustrates two distinct genotypes between North American and Chinese (Sichuan Basin) plays. Microbial communities are of higher diversity in FPW generated from the Sichuan Basin than the North America Basins. In addition, microbial community compositions are dramatically different between the two regions, wherein community compositions in FPW derived from North America Basins have more halophiles such

as *Halanaerobium*, while communities derived from in FPW of the Sichuan Basin shale gas formations are similar to those found in conventional oil reservoirs. The lower biodiversity and presence of halophiles in the North America Basins is correlated to significantly higher salinity. Core functional genes related to fermentation, sulfidogenesis, and methanogenesis are conserved between the two regions. The more biodiverse community in the Sichuan Basin contained more functional pathways that can lead to sulfidogenesis and methanogenesis as compared to communities in FPW of the North America Basins. Chapter 4 provides the first evidence that microbial functions related to infrastructure degradation and biogenic methane production are conserved across major North America and Chinese basins. The results imply that the core functions of the shale microbiome and their impacts to energy recovery might be generally similar. However, the treatment strategies in different regions of production should be cautiously considered to target specific communities that can adapt to local downhole conditions.

Together, the series of studies in my dissertation enhances our overall understanding of the microbial ecology in environments closely related to the HF water cycle, which has implications for energy extraction efficiency, predicting the impacts of FPW spills, and developing strategies to increase energy yield and risk assessment and management of FPW. Based on my research, future work should be done in the areas of: (1) using bioindicators coupled with the recent advances in computational techniques for FPW contaminants to support traditional chemical analyses (Chapter 2); (2) investigating the pathways of biodegradation of organic pollutants and the fate of metals derived from FPW, and identifying the key microbes responsible for mitigating these FPW pollutants; (3) developing baseline environmental assessments to identify vulnerable soil and water environments to detect and remediate the adverse effects of FPW contamination (Chapter 3); and (4) tracking the biogeochemical processes (e.g., degradation pathways of

chemical additives in the downhole environments from Chinese shale hydrocarbon development), and promoting the selection of proper FPW treatment strategies based on the distinct microbial communities in different regions of unconventional hydrocarbon development (Chapter 4).

This dissertation is mainly comprised of genome-based studies. Thus, a significant next step is to link speculations made on genotype to the phenotype by isolating key bacteria observed in these studies or conducting studies based on the RNA-based gene expression approaches such as metatranscriptomics and proteomics. Despite the genome-based studies in Chapters 2 and 3 that give permissive evidence that microbial communities can degrade organic contaminants derived from HF chemicals (Chapter 2 and 3), the degradation pathways are not well understood. Developing such an understanding requires in-depth chemical analyses on the degradation intermediates and higher sequencing resolution to link specific organic compounds to the key genes responsible for biodegradation. Chapters 2 and 3 used simulated aquatic and soil environments in the laboratory, and thus, investigation of the *in situ* microbial ecology in the field is needed to validate these simulations. In Chapter 2, I found that adding FPW can quickly kill most of the healthy cells in water, although the number of viable cells recovered after several days of incubation. These results suggest that FPW may temporarily inhibit the growth of the majority of the indigenous microbes in freshwater environments. This knowledge is useful in designing more sustainable wastewater treatment strategies.

Finally, monitoring of changes in microbial communities over a more extended period and monitoring of the microbial changes under different conditions such as pH and temperature is needed to more accurately understand the response of microbial communities to FPW exposure to improve bioremediation and natural attenuation. Future research should study the chronic impacts

of low dose exposure on the microbial communities, because these chronic impacts may take longer time to be effectively detected.

Chapter 4 demonstrates that the microbial community structures in FPW from China are significantly different from those in North American Basins. However, my study was focused on a large-scale comparison detailing the key differences between North America and Chinese shale microbiomes in FPW. Future studies may reconstruct detailed metabolic pathways of the metagenome-assemble genomes identified in my study in order to confirm the conjectures in my study; for example, an analysis of the genome bins related to methanogens found in Chapter 4. Further, isolating the abundant microbes identified from my genome-based studies may be needed in future investigations of the functions of the microbial communities. Despite the fact that my study showed the presence of thermotolerant and thermophilic organisms, the temperature was not measured in this study. Thus, future studies should investigate the impacts of temperature differences between formations in China and North America on the distribution and functions of microbial communities. Currently, microbiological data on FPW from China is limited, so more studies are needed on the major bacteria observed in this thesis to assess their impacts to energy extraction and to develop effective mitigation practices where necessary. Compared to the studies from the United States, the biogeochemical processes in both China and Canada are less understood. For example, the impacts of chemical additives and their transformation products at downhole conditions are poorly understood. Therefore, future studies should track temporal changes in the input chemicals. Cumulatively, this knowledge will support observations from the genome-based studies and enhance our understanding of biogeochemical processes in environments impacted by hydraulic fracturing.

6. References

Ahmad D, Mehmannavaz R, Damaj M. Isolation and Characterization of Symbiotic N2-Fixing Rhizobium Meliloti From Soils Contaminated with Aromatic and Chloroaromatic Hydrocarbons: PAHs and PCBs. *Int Biodeterior Biodegrad* 1997;**39**:33–43.

Akob DM, Cozzarelli IM, Dunlap DS *et al.* Organic and Inorganic Composition and Microbiology of Produced Waters from Pennsylvania Shale Gas Wells. *Appl Geochemistry* 2015;**60**:116–25.

Akob DM, Mumford AC, Orem W *et al.* Wastewater Disposal from Unconventional Oil and Gas Development Degrades Stream Quality at a West Virginia Injection Facility. *Environ Sci Technol* 2016;**50**:5517–25.

Alberta Energy Regulator. *Alberta Energy Industry Water Use Report*., 2019.

Alessi DS, Zolfaghari A, Kletke S *et al.* Comparative Analysis of Hydraulic Fracturing Wastewater Practices in Unconventional Shale Development: Water Sourcing, Treatment and Disposal Practices. *Can Water Resour J* 2017;**42**:105–21.

Allshouse WB, McKenzie LM, Barton K *et al.* Community Noise and Air Pollution Exposure during the Development of a Multi-Well Oil and Gas Pad. *Environ Sci Technol* 2019;**53**:7126–35.

An BA, Shen Y, Voordouw G. Control of Sulfide Production in High Salinity Bakken Shale Oil Reservoirs by Halophilic Bacteria Reducing Nitrate to Nitrite. *Front Microbiol* 2017;**8**, DOI: 10.3389/fmicb.2017.01164.

Arnold G, Long LAN. Policy Expansion in Local Government Environmental Policy Making. *Public Adm Rev* 2019;**79**:465–76.

Atkinson GM, Eaton DW. Developments in Understanding Seismicity Triggered by Hydraulic Fracturing. *Nat Rev Earth Environ* 2020;**1**, DOI: 10.1038/s43017-020-0049-7.

Avidan M, Etzion D, Gehman J. Opaque Transparency: How Material Affordances Shape Intermediary Work. *Regul Gov* 2019;**13**:197–219.

Barabote RD, Xie G, Leu DH *et al.* Complete Genome of the Cellulolytic Thermophile Acidothermus cellulolyticus IIB Provides Insights into its Ecophysiological and Evolutionary Adaptations. *Genome Res* 2009;**19**:1033–42.

Barbot E, Vidic NS, Gregory KB *et al.* Spatial and Temporal Correlation of Water Quality Parameters of Produced Waters from Devonian-Age Shale Following Hydraulic Fracturing. *Environ Sci Technol* 2013;**47**:2562–9.

Barth-Naftilan E, Sohng J, Saiers JE. Methane in Groundwater Before, During, and After Hydraulic Fracturing of the Marcellus Shale. *Proc Natl Acad Sci U S A* 2018;**115**:6970–5.

Bendix J, Wiley JJ, Commons MG. Intermediate Disturbance and Patterns of Species Richness. *Phys Geogr* 2017;**38**:393–403.

Bernardet J, Bowman JP. The Genus Flavobacterium. *Prokaryotes* 2006;**7**:481–531.

Blewett TA, Delompré PLM, He Y *et al.* Sublethal and Reproductive Effects of Acute and Chronic Exposure to Flowback and Produced Water from Hydraulic Fracturing on the Water Flea Daphnia magna. *Environ Sci Technol* 2017a;**51**:3032–9.

Blewett TA, Weinrauch AM, Delompré PLM *et al.* The Effect of Hydraulic Flowback and Produced Water on Gill Morphology, Oxidative Stress and Antioxidant Response in Rainbow Trout (Oncorhynchus Mykiss). *Sci Rep* 2017b;**7**:46582.

Blondes MS, Shelton JL, Trembly J *et al.* The Geochemical Composition of Produced Water from the Utica Shale Play in the Appalachian Basin: Implications for Treatment and Reuse. *Am Geophys Union, Fall Meet Abstr* 2018.

Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 2014;**30**:2114–20.

Bolyen E, Rideout JR, Dillon MR *et al.* Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using QIIME 2. *Nat Biotechnol* 2019;**37**:852–7.

Booker AE, Borton MA, Daly RA *et al.* Sulfide Generation by Dominant Halanaerobium Microorganisms in Hydraulically Fractured Shales. *mSphere* 2017;**2**:e00257-17.

Booker AE, Hoyt DW, Meulia T *et al.* Deepsubsurface Pressure Stimulates Metabolic Plasticity in Shale-Colonizing Halanaerobium spp. *Appl Environ Microbiol* 2019;**85**:1–16.

Borton MA, Daly RA, Banion BO *et al.* Comparative Genomics and Physiology of the Genus Methanohalophilus, a Prevalent Methanogen in Hydraulically Fractured Shale. *Environ Microbiol* 2018a;**20**:4596–611.

Borton MA, Hoyt DW, Roux S *et al.* Coupled Laboratory and Field Investigations Resolve Microbial Interactions that Underpin Persistence in Hydraulically Fractured Shales. *Proc Natl Acad Sci* 2018b;**115**:E6585–E6594.

Bower CE, Holm-Hansen T. A Salicylate–Hypochlorite Method for Determining Ammonia in Seawater. *Can J Fish Aquat Sci* 1980;**37**:794–8.

Bowers RM, Kyrpides NC, Stepanauskas R *et al.* Minimum Information About A Single Amplified Genome (MISAG) and A Metagenome-Assembled Genome (MIMAG) of Bacteria and Archaea. *Nat Biotechnol* 2017;**35**:725–31.

Boyd JA, Jungbluth SP, Leu AO *et al.* Divergent Methyl-Coenzyme M Reductase Genes in A Deep-Subseafloor Archaeoglobi. *ISME J* 2019;**13**:1269–79.

Brantley SL, Yoxtheimer D, Arjmand S *et al.* Water Resource Impacts During Unconventional Shale Gas Development: The Pennsylvania Experience. *Int J Coal Geol* 2014;**126**:140–56.

Breiman L, Cutler A, Liaw A *et al.* Breiman and Cutler's Random Forests for Classification and Regression. 2018.

Buono RM, Gunn EL, McKay J *et al. Regulating Water Security in Unconventional Oil and Gas*., 2019.

Butkovskyi A, Bruning H, Kools SAE *et al.* Organic Pollutants in Shale Gas Flowback and Produced Waters: Identification, Potential Ecological Impact, and Implications for Treatment Strategies. *Environ Sci Technol* 2017;**51**:4740–54.

Chaîneau CH, Morel J, Dupont J *et al.* Comparison of The Fuel Oil Biodegradation Potential of Hydrocarbon-Assimilating Microorganisms Isolated from A Temperate Agricultural Soil. *Sci Total Environ* 1999;**227**:237–47.

Chaumeil P-A, Mussig AJ, Hugenholtz P *et al.* GTDB-Tk: A Toolkit to Classify Genomes with

The Genome Taxonomy Database. *Bioinformatics* 2019;**36**:1925–7.

Chen F, Lu S, Ding X *et al.* Shale Gas Reservoir Characterization: A Typical Case in The Southeast Chongqing of Sichuan Basin, China. *PLoS One* 2018;**13**:1–16.

Chen IMA, Chu K, Palaniappan K *et al.* IMG/M v.5.0: An Integrated Data Management and Comparative Analysis System for Microbial Genomes and Microbiomes. *Nucleic Acids Res* 2019;**47**:D666–77.

Chen S, Zhu Y, Wang H *et al.* Shale Gas Reservoir Characterisation: A Typical Case in The Southern Sichuan Basin of China. *Energy* 2011;**36**:6609–16.

Chen SS, Sun Y, Tsang DCW *et al.* Potential Impact of Flowback Water from Hydraulic Fracturing on Agricultural Soil Quality: Metal/Metalloid Bioaccessibility, Microtox Bioassay, and Enzyme Activities. *Sci Total Environ* 2016;**579**:1419–26.

Chen SS, Sun Y, Tsang DCW *et al.* Insights into the Subsurface Transport of As(V) and Se(VI) in Produced Water from Hydraulic Fracturing Using Soil Samples from Qingshankou Formation, Songliao Basin, China. *Environ Pollut* 2017;**223**:449–56.

Cheng L, Dai L, Li X *et al.* Isolation and Characterization of Methanothermobacter crinale sp. nov., A Novel Hydrogenotrophic Methanogen from the Shengli Oil Field. *Appl Environ Microbiol* 2011;**77**:5212–9.

Chermak JA, Schreiber ME. Mineralogy and Trace Element Geochemistry of Gas Shales in The United States: Environmental implications. *Int J Coal Geol* 2014;**126**:32–44.

Chinese National Bureau of Statistics. The Natural Gas Production by Years (In Chinese). 2019.

Christen R, Fernandez L, Acquaviva M. Marinobacter hydrocarbonoclasticus gen. nov., sp. nov., a New, Extremely Halotolerant, Hydrocarbon-Degrading Marine Bacterium. *Growth Lakel* 1992;**42**:568–76.

Clark MP, Nijssen B, Lundquist JD *et al.* Hydraulic Fracturing Fluid Migration in the Subsurface: A Review and Expanded Modeling Results. *Water Rerouce Res* 2015;**51**:7159–88.

Cliffe L, Nixon SL, Daly RA *et al.* Identification of Persistent Sulfidogenic Bacteria in Shale Gas Produced Waters. *Front Microbiol* 2020;**11**:1–13.

Cluff MA, Hartsock A, Macrae JD *et al.* Temporal Changes in Microbial Ecology and Geochemistry in Produced Water from Hydraulically Fractured Marcellus Shale Gas Wells. *Environ Sci Technol* 2014;**48**:6508–17.

Cokar M, Ford B, Kallos MS *et al.* New Gas Material Balance to Quantify Biogenic Gas Generation Rates from Shallow Organic-Matter-Rich Shales. *Fuel* 2013;**104**:443–51.

Cook M, Webber M. Food, Fracking, and Freshwater: The Potential for Markets and Cross-Sectoral Investments to Enable Water Conservation. *Water* 2016;**8**:45.

Copeland A, Spring S, Göker M *et al.* Complete Genome Sequence of Desulfomicrobium Baculatum Type Strain (XT). *Stand Genomic Sci* 2009;**1**:29–37.

Cozzarelli IM, Skalak KJ, Kent DB *et al.* Signatures and Effects of An Oil and Gas Wastewater Spill in The Williston Basin, North Dakota. *Sci Total Environ* 2017;**579**:1781–93.

Dahle H, Birkeland NK. Thermovirga lienii gen. nov., sp. nov., a Novel Moderately Thermophilic, Anaerobic, Amino-Acid-Degrading Bacterium Isolated from a North Sea Oil Well. *Int J Syst*

Evol Microbiol 2006;**56**:1539–45.

Dai C, Wang K, Liu Y *et al.* Reutilization of Fracturing flowback Fluids in Surfactant Flooding for Enhanced Oil Recovery. *Energy and Fuels* 2015;**29**:2304–11.

Daly RA, Borton MA, Wilkins MJ *et al.* Microbial Metabolisms in a 2.5-km-Deep Ecosystem Created by Hydraulic Fracturing in Shales. *Nat Microbiol* 2016;**1**:16146.

Daly RA, Roux S, Borton MA *et al.* Viruses Control Dominant Bacteria Colonizing the Terrestrial Deep Biosphere after Hydraulic Fracturing. *Nat Microbiol* 2019;**4**:352–61.

Davis JP, Struchtemeyer CG, Elshahed MS. Bacterial Communities Associated with Production Facilities of Two Newly Drilled Thermogenic Natural Gas Wells in the Barnett Shale (Texas, USA). *Microb Ecol* 2012;**64**:942–54.

Delgado-Baquerizo M, Maestre FT, Reich PB *et al.* Microbial Diversity Drives Multifunctionality in Terrestrial Ecosystems. *Nat Commun* 2016a;**7**:1–8.

Delgado-Baquerizo M, Maestre FT, Reich PB *et al.* Microbial Diversity Drives Multifunctionality in Terrestrial Ecosystems. *Nat Commun* 2016b;**7**:1–8.

Delompré PLM, Blewett TA, Snihur KN *et al.* The Osmotic Effect of Hyper-Saline Hydraulic Fracturing Fluid on Rainbow Trout, Oncorhynchus Mykiss. *Aquat Toxicol* 2019;**211**:1–10.

Dokshin FA. Whose Backyard and What's at Issue? Spatial and Ideological Dynamics of Local Opposition to Fracking in New York State, 2010 to 2013. *Am Sociol Rev* 2016;**81**:921–48.

Dong X, Strous M. An Integrated Pipeline for Annotation and Visualization of Metagenomic Contigs. *Front Genet* 2019;**10**:1–10.

Dubber D, Gray NF. Replacement of Chemical Oxygen Demand (COD) with Total Organic Carbon (TOC) for Monitoring Wastewater Treatment Performance to Minimize Disposal of Toxic Analytical Waste. *J Environ Sci Heal - Part A Toxic/Hazardous Subst Environ Eng* 2010;**45**:1595–600.

Engle MA, Cozzarelli IM, Smith BD. *USGS Investigations of Water Produced During Hydrocarbon Reservoir Development*., 2014.

Entcheva-Dimitrov P, Spormann AM. Dynamics and Control of Biofilms of The Oligotrophic Bacterium Caulobacter Crescentus. *J Bacteriol* 2004;**186**:8254–66.

Eren AM, Esen OC, Quince C *et al.* Anvi'o: An Advanced Analysis and Visualization Platformfor 'Omics Data. *PeerJ* 2015:PeerJ 3:e1319.

Evans M V., Getzinger G, Luek JL *et al.* In Situ Transformation of Ethoxylate and Glycol Surfactants by Shale-Colonizing Microorganisms During Hydraulic Fracturing. *ISME J* 2019a;**13**:2690–700.

Evans PN, Boyd JA, Leu AO *et al.* An Evolving View of Methane Metabolism in The Archaea. *Nat Rev Microbiol* 2019b;**17**:219–32.

Feng Z, Dong D, Tian J *et al.* Geochemical Characteristics of Longmaxi Formation Shale Gas in the Weiyuan Area, Sichuan Basin, China. *J Pet Sci Eng* 2018;**167**:538–48.

Ferrer I, Thurman EM. Chemical Constituents and Analytical Approaches for Hydraulic Fracturing Waters. *Trends Environ Anal Chem* 2015a;**5**:18–25.

Ferrer I, Thurman EM. Analysis of Hydraulic Fracturing Additives by LC/Q-TOF-MS. *Anal*

Bioanal Chem 2015b;**407**:6417–28.

Fierer N, Jackson RB. The Diversity and Biogeography of Soil Bacterial Communities. *Proc Natl Acad Sci U S A* 2006;**103**:626–31.

Flynn SL, Gunten K Von, Warchola T *et al.* Characterization and Implications of Solids Associated with Hydraulic Fracturing Flowback and Produced Water from the Duvernay Formation. *Environ Sci Process Impacts* 2019;**21**:242–55.

Folkerts EJ, Blewett TA, Delompré P *et al.* Toxicity in Aquatic Model Species Exposed to A Temporal Series of Three Different Flowback and Produced Water Samples Collected from A Horizontal Hydraulically Fractured Well. *Ecotoxicol Environ Saf* 2019;**180**:600–9.

Folkerts EJ, Blewett TA, Goss GG. *Investigating the Potential Toxicity of Hydraulic Fracturing Flowback and Produced Water Spills to Aquatic Animals in Freshwater Environments: A North American Perspective*. Springer, Cham, 2020.

Folkerts EJ, Blewett TA, He Y *et al.* Alterations to Juvenile Zebrafish (Danio rerio) Swim Performance after Acute Embryonic Exposure to Sub-lethal Exposures of Hydraulic Fracturing Flowback and Produced Water. *Aquat Toxicol* 2017a;**193**:50–9.

Folkerts EJ, Blewett TA, He Y *et al.* Cardio-Respirometry Disruption in Zebrafish (Danio rerio) Embryos Exposed to Hydraulic Fracturing Flowback and Produced Water. *Environ Pollut* 2017b;**231**:1477–87.

Frock AD, Notey JS, Kelly RM. The genus Thermotoga: Recent developments. *Environ Technol* 2010;**31**:1169–81.

Funk SP, Du L, He Y *et al.* Assessment of Impacts of Diphenyl Phosphate on Groundwater and Near- Surface Environments: Sorption and Toxicity. *J Contam Hydrol* 2019;**221**:50–7.

Gagnon GA, Krkosek W, Anderson L *et al.* Impacts of Hydraulic Fracturing on Water Quality: A Review of Literature, Regulatory Frameworks and An Analysis of Information Gaps. *Environ Rev* 2016;**24**:122–31.

Gaspar J, Davis D, Camacho C *et al.* Biogenic versus Thermogenic H2S Source Determination in Bakken Wells: Considerations for Biocide Application. *Environ Sci Technol Lett* 2016;**3**:127–32.

Gaspar J, Mathieu J, Yang Y *et al.* Microbial Dynamics and Control in Shale Gas Production. *Environ Sci Technol Lett* 2014;**1**:465–73.

Gehman J, Lefsrud LM, Fast S. Social License to Operate: Legitimacy by Another Name? *Can Public Adm* 2017;**60**:293–317.

Gehman J, Thompson DY, Alessi DS *et al.* Comparative Analysis of Hydraulic Fracturing Wastewater Practices in Unconventional Shale Development: Newspaper Coverage of Stakeholder Concerns and Social License to Operate. *Sustain* 2016;**8**, DOI: 10.3390/su8090912.

González-Pastor JE. Cannibalism: A Social Behavior in Sporulating Bacillus Subtilis. *FEMS Microbiol Rev* 2011;**35**:415–24.

Goss G, Alessi D, Allen D *et al.* Unconventional Wastewater Management: A Comparative Review and Analysis of Hydraulic Fracturing Wastewater Management Practices across Four

North American Basins. *Can Water Netw* 2015:187 pp.

Graham J, Irving J, Tang X *et al.* Increased Traffic Accident Rates Associated with Shale Gas Drilling in Pennsylvania. *Accid Anal Prev* 2015;**74**:203–9.

Gregory KB, Vidic RD, Dzombak DA. Water Management Challenges Associated with the Production of Shale Gas by Hydraulic Fracturing. *Elements* 2011;**7**:181–6.

Gruber-Vodicka H, Seah BK, Pruesse E. PhyloFlash — Rapid SSU rRNA Profiling and Targeted Assembly from mMetagenomes. *bioRxiv* 2019:521922.

Guo C, Chang H, Liu B *et al.* A Combined Ultrafiltration-Reverse Osmosis Process for External Reuse of Weiyuan Shale Gas Flowback and Produced Water. *Environ Sci Water Res Technol* 2018;**4**:942–55.

Guo M, Lu X, Nielsen CP *et al.* Prospects for Shale Gas Production in China: Implications for Water Demand. *Renew Sustain Energy Rev* 2016;**66**:742–50.

Guo Y, Zhang X, Wang Q *et al.* Temporal Changes in Vegetation Around A Shale Gas Development Area in A Subtropical Karst Region in Southwestern China. *Sci Total Environ* 2020;**701**:134769.

Haluszczak LO, Rose AW, Kump LR. Geochemical Evaluation of Flowback Brine from Marcellus Gas Wells in Pennsylvania, USA. *Appl Geochemistry* 2013;**28**:55–61.

Harkness JS, Dwyer GS, Warner NR *et al.* Iodide, Bromide, and Ammonium in Hydraulic Fracturing and Oil and Gas Wastewaters: Environmental Implications. *Environ Sci Technol* 2015;**49**:1955–63.

He Y, Flynn SL, Folkerts EJ *et al.* Chemical and Toxicological Characterizations of Hydraulic Fracturing Flowback and Produced Water. *Water Res* 2017a;**114**:78–87.

He Y, Folkerts EJ, Zhang Y *et al.* Effects on Biotransformation, Oxidative Stress, and Endocrine Disruption in Rainbow Trout (Oncorhynchus Mykiss) Exposed to Hydraulic Fracturing Flowback and Produced Water. *Environ Sci Technol* 2017b;**51**:940–7.

He Y, Sun C, Zhang Y *et al.* Developmental Toxicity of the Organic Fraction from Hydraulic Fracturing Flowback and Produced Waters to Early Life Stages of Zebrafish (Danio Rerio). *Environ Sci Technol* 2018a;**52**:3820–30.

He Y, Zhang Y, Martin JW *et al.* In Vitro Assessment of Endocrine Disrupting Potential of Organic Fractions Extracted from Hydraulic Fracturing Flowback and Produced Water (HF-FPW). *Environ Int* 2018b;**121**:824–31.

Head IM, Jones DM, Larter SR. Biological Activity in the Deep Subsurface and the Origin of Heavy Oil. 2003;**426**.

Heyob KM, Blotevogel J, Brooker M *et al.* Natural Attenuation of Nonionic Surfactants Used in Hydraulic Fracturing Fluids: Degradation Rates, Pathways, and Mechanisms. *Environ Sci Technol* 2017;**51**:13985–94.

Hill LL, Czolowski ED, DiGiulio D *et al.* Temporal and Spatial Trends of Conventional and Unconventional Oil and Gas Waste Management in Pennsylvania, 1991–2017. *Sci Total Environ* 2019;**674**:623–36.

Hossack BR, Smalling KL, Anderson CW *et al.* Effects of Persistent Energy-Related Brine

Contamination on Amphibian Abundance in National Wildlife Refuge Wetlands. *Biol Conserv* 2018;**228**:36–43.

Hou D, Luo J, Al-Tabbaa A. Shale Gas Can Be a Double-Edged Sword for Climate Change. *Nat Clim Chang* 2012;**2**:385–7.

Hu P, Tom L, Singh A *et al.* Genome-Resolved Metagenomic Analysis Reveals Roles for Candidate Phyla and Other Microbial Community Members in Biogeochemical Transformations in Oil Reservoirs. *MBio* 2016;**7**:1–12.

Huang T, Pang Z, Li Z *et al.* A Framework to Determine Sensitive Inorganic Monitoring Indicators for Tracing Groundwater Contamination by Produced Formation Water from Shale Gas Development in the Fuling Gasfield, SW China. *J Hydrol* 2020;**581**:124403.

Hug LA, Baker BJ, Anantharaman K *et al.* A New View of The Tree of Life. *Nat Microbiol* 2016;**1**:1–6.

Jackson RB, Vengosh A, Darrah TH *et al.* Increased Stray Gas Abundance in A Subset of Drinking Water Wells. *Proc Natl Acad Sci U S A* 2013;**110**:11250–5.

Jain C, Rodriguez-R LM, Phillippy AM *et al.* High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries. *Nat Commun* 2018;**9**:1–8.

Jasechko S, Perrone D. Hydraulic Fracturing near Domestic Groundwater Wells. *Proc Natl Acad Sci U S A* 2017;**114**:13138–43.

Kadam PC, Boone DR. Physiological Characterization and Emended Description of Methanolobus vulcani. *Int J Syst Bacteriol* 1995;**45**:400–2.

Kahrilas GA, Blotevogel J, Corrin ER *et al.* Downhole Transformation of the Hydraulic Fracturing Fluid Biocide Glutaraldehyde: Implications for Flowback and Produced Water Quality. *Environ Sci Technol* 2016;**50**:11414–23.

Kahrilas GA, Blotevogel J, Stewart PS *et al.* Biocides in Hydraulic Fracturing Fluids: A Critical Review of Their Usage, Mobility, Degradation, and Toxicity. *Environ Sci Technol* 2015;**49**:16–32.

Kalyuzhnaya MG, Beck DAC, Suciu D *et al.* Functioning In Situ: Gene Expression in Methylotenera Mobilis in Its Native Environment as Assessed Through Transcriptomics. *ISME J* 2010;**4**:388–98.

Kang DD, Froula J, Egan R *et al.* MetaBAT, An Efficient Tool for Accurately Reconstructing Single Genomes from Complex Microbial Communities. *PeerJ* 2015;**2015**:1–15.

Kanso S, Greene AC, Patel BKC. Bacillus subterraneus sp. nov., An Iron- and Manganese-Reducing Bacterium from A Deep Subsurface Australian Thermal Aquifer. *Int J Syst Evol Microbiol* 2002;**52**:869–74.

Karlsson FH, Fåk F, Nookaew I *et al.* Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nat Commun* 2012, DOI: 10.1038/ncomms2266.

Karlsson FH, Tremaroli V, Nookaew I *et al.* Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* 2013, DOI: 10.1038/nature12198.

Kaster AK, Goenrich M, Seedorf H *et al.* More Than 200 Genes Required for Methane Formation from H2 and CO2 and Energy Conservation Are Present in Methanothermobacter

marburgensis and Methanothermobacter Thermautotrophicus. *Archaea* 2011;**2011**, DOI: 10.1155/2011/973848.

Kearse M, Moir R, Wilson A *et al.* Geneious Basic: An Integrated and Extendable Desktop Software Platform for The Organization and Analysis of Sequence Data. *Bioinformatics* 2012;**28**:1647–9.

Kekacs D, Drollette BD, Brooker M *et al.* Aerobic Biodegradation of Organic Compounds in Hydraulic Fracturing Fluids. *Biodegradation* 2015;**26**:271–87.

Kleindienst S, Seidel M, Ziervogel K *et al.* Chemical Dispersants Can Suppress the Activity of Natural Oil-Degrading Microorganisms. *Proc Natl Acad Sci U S A* 2015;**112**:14900–5.

Klingeberg M, Galunsky B, Sjoholm C *et al.* Purification and Properties of A Highly Thermostable, Sodium Dodecyl Sulfate-Resistant and Stereospecific Proteinase from The Extremely Thermophilic Archaeon Thermococcus Stetteri. *Appl Environ Microbiol* 1995;**61**:3098–104.

Kondash A, Vengosh A. Water Footprint of Hydraulic Fracturing. *Environ Sci Technol Lett* 2015;**2**:276–80.

Kondash AJ, Albright E, Vengosh A. Quantity of Flowback and Produced Waters from Unconventional Oil and Gas Exploration. *Sci Total Environ* 2017;**574**:314–21.

Konschnik K, Dayalu A. Hydraulic Fracturing Chemicals Reporting: Analysis of Available Data and Recommendations for Policymakers. *Energy Policy* 2016;**88**:504–14.

Kozich JJ, Westcott SL, Baxter NT *et al.* Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the Miseq Illumina Sequencing

Platform. *Appl Environ Microbiol* 2013;**79**:5112–20.

Krupnick A, Wang Z, Wang Y. Environmental Risks of Shale Gas Development in China. *Energy Policy* 2014;**75**:117–25.

Langendijk PS, Kulik EM, Sandmeier H *et al.* Isolation of Desulfomicrobium orale sp. nov. and Desulfovibrio Strain NY682, Oral Sulfate-Reducing Bacteria involved in Human Periodontal Disease. *Int J Syst Evol Microbiol* 2001;**51**:1035–44.

Langmead B, Salzberg S. Fast Gapped-read Alignment with Bowtie 2. *Nat Methods* 2013;**9**:357–9.

Lauer NE, Harkness JS, Vengosh A. Brine Spills Associated with Unconventional Oil Development in North Dakota. *Environ Sci Technol* 2016;**50**:5389–97.

Lee MD, Ponty Y. GToTree: A User-Friendly Workflow for Phylogenomics. *Bioinformatics* 2019;**35**:4162–4.

Leggett MJ, Mcdonnell G, Denyer SP *et al.* Bacterial Spore Structures and Their Protective Role in Biocide Resistance. *J Appl Microbiol* 2012;**113**:485–98.

Lester Y, Ferrer I, Thurman EM *et al.* Characterization of Hydraulic Fracturing Flowback Water in Colorado: Implications for Water Treatment. *Sci Total Environ* 2015;**512**–**513**:637–44.

Li D, Liu CM, Luo R *et al.* MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015, DOI: 10.1093/bioinformatics/btv033.

Liang JL, Liu J, Jia P *et al.* Novel Phosphate-Solubilizing Bacteria Enhance Soil Phosphorus

Cycling Following Ecological Restoration of Land Degraded by Mining. *ISME J* 2020, DOI: 10.1038/s41396-020-0632-4.

Liang R, Davidova IA, Marks CR *et al.* Metabolic Capability of a Predominant Halanaerobium sp. in Hydraulically Fractured Gas Wells and its Implication in Pipeline Corrosion. *Front Microbiol* 2016;**7**:988.

Lipus D. Microbiology of Hydraulic Fracturing Wastewater. 2017, DOI: 10.1360/zd-2013-43-6-1064.

Lipus D, Vikram A, Ross D *et al.* Predominance and Metabolic Potential of Halanaerobium spp. in Produced Water from Hydraulically Fractured Marcellus Shale wells. *Appl Environ Microbiol* 2017;**83**, DOI: 10.1128/AEM.02659-16.

Liu H, Yang X, Liu H *et al.* Processing and Recycling of Waste Flowback Fracturing Fluids in Tarim Basin, China. *Soc Pet Eng - SPE Heal Safety, Secur Environ Soc Responsib Conf - North Am 2017* 2017:271–6.

Liu J, Zheng Y, Lin H *et al.* Proliferation of Hydrocarbon-Degrading Microbes at the Bottom of the Mariana Trench. *Microbiome* 2019;**7**:1–13.

Liu Q, Zhijun J, Jianfa C *et al.* Origin of Nitrogen Molecules in Natural Gas and Implications for the High Risk of N2 Exploration in Tarim Basin, NW China. *J Pet Sci Eng* 2012;**81**:112–21.

Llewellyn GT, Dorman F, Westland JL *et al.* Evaluating A Groundwater Supply Contamination Incident Attributed to Marcellus Shale Gas Development. *Proc Natl Acad Sci* 2015;**112**:6325–30.

Luek JL, Harir M, Schmitt-Kopplin P *et al.* Temporal dynamics of halogenated organic compounds in Marcellus Shale flowback. *Water Res* 2018;**136**:200–6.

Magnabosco C, Lin L, Dong H *et al.* The biomass and biodiversity of the continental subsurface. *Nat Geosci* 2018;**11**, DOI: 10.1038/s41561-018-0221-6.

Maloney KO, Baruch-mordo S, Patterson LA *et al.* Unconventional Oil and Gas Spills: Materials, Volumes, and Risks to Surface Waters in Four States of The U.S. *Sci Total Environ* 2017;**581**–**582**:369–77.

Mauter MS, Alvarez PJJ, Burton A *et al.* Regional Variation in Water-Related Impacts of Shale Gas Development and Implications for Emerging International Plays. *Environ Sci Technol* 2014;**48**:8298–306.

Mazur A. How Did the Fracking Controversy Emerge in The Period 2010-2012? *Public Underst Sci* 2016;**25**:207–22.

McCammon SA, Bowman JP. Taxonomy of Antarctic Flavobacterium Species: Description of Flavobacterium gillisiae sp. nov., Flavobacterium tegetincola sp. nov. and Flavobacterium xanthum sp. nov., nom. rev. and Reclassification of [Flavobacterium] salegens as Salegentibacter salegen. *Int J Syst Evol Microbiol* 2000;**50**:1055–63.

McLaughlin MC, Borch T, Blotevogel J. Spills of Hydraulic Fracturing Chemicals on Agricultural Topsoil: Biodegradation, Sorption, and Co-contaminant Interactions. *Environ Sci Technol* 2016;**50**:6071–8.

McMurdie PJ, Holmes S. Phyloseq : An R Package for Reproducible Interactive Analysis and

Graphics of Microbiome Census Data. *PLoS One* 2013;**8**:e61217.

Mehler WT, Nagel A, Flynn S *et al.* Understanding the Effects of Hydraulic Fracturing Flowback and Produced Water (FPW) to The Aquatic Invertebrate, Lumbriculus Variegatus Under various Exposure Regimes. *Environ Pollut* 2020;**259**:113889.

Ministry of Ecology and Environment of the People's Republic of China. *HJ/T 51-1999 Water Quality: Determination of Total Salt-Gravimetric Method*., 1999.

Ministry of Ecology and Environment of the People's Republic of China. *HJ 828-2017 Water Quality: Determination of the Chemical Oxygen Demand - Dichromate Method*., 2017.

Minoche AE, Dohm JC, Himmelbauer H. Evaluation of Genomic High-Throughput Sequencing Data Generated on Illumina HiSeq and Genome Analyzer Systems. *Genome Biol* 2011;**12**, DOI: 10.1186/gb-2011-12-11-r112.

Miyazaki M, Sakai S, Ritalahti KM *et al.* Sphaerochaeta multiformis sp. nov., An Anaerobic, Psychrophilic Bacterium Isolated from Subseafloor Sediment, and Emended Description of the Genus Sphaerochaeta. *Int J Syst Evol Microbiol* 2014;**64**:4147–54.

Mohagheghi A, Grohmann K, Himmel M. Isolation and Characterization of Acidothermus cellulolyticus gen. nov., sp. nov., A New Genus of Thermophilic, Acidophilic, Cellulolytic Bacteria. *Int J Syst Bacteriol* 1986;**36**:435–43.

Mohan AM, Bibby KJ, Lipus D *et al.* The Functional Potential of Microbial Communities in Hydraulic Fracturing Source Water and Produced Water from Natural Gas Extraction Characterized by Metagenomic Sequencing. *PLoS One* 2014;**9**, DOI:

10.1371/journal.pone.0107682.

Moritz A, Hélie JF, Pinti DL *et al.* Methane Baseline Concentrations and Sources in Shallow Aquifers from the Shale Gas-Prone Region of The St. Lawrence Lowlands (Quebec, Canada). *Environ Sci Technol* 2015;**49**:4765–71.

Mouser PJ, Borton M, Darrah TH *et al.* Hydraulic Fracturing Offers View of Microbial Life in the Deep Terrestrial Subsurface. *FEMS Microbiol Ecol* 2016;**92**:fiw166.

Mumford AC, Maloney KO, Akob DM *et al.* Shale Gas Development Has Limited Effects on Stream Biology and Geochemistry in A Gradient-Based, Multiparameter Study in Pennsylvania. *Proc Natl Acad Sci U S A* 2020;**117**:3670–7.

Murali Mohan A, Hartsock A, Bibby KJ *et al.* Microbial Community Changes in Hydraulic Fracturing Fluids and Produced Water from Shale Gas Extraction. *Environ Sci Technol* 2013a;**47**:13141–50.

Murali Mohan A, Hartsock A, Hammack RW *et al.* Microbial Communities in Flowback Water Impoundments from Hydraulic Fracturing for Recovery of Shale Gas. *FEMS Microbiol Ecol* 2013b;**86**:567–80.

National Environmental Methods Index. *Nitrogen, Nitrate-Nitrite; Method 353.1 (Colorimetric, Automated, Hydrazine Reduction)*., 1978a.

National Environmental Methods Index. *Chloride (Colorimetric, Automated Ferricyanide AII)*., 1978b.

National Environmental Methods Index. *Phosphorus, All Forms (Colorimetric, Automated,*

Ascorbic Acid)., 1993.

Nau-Wagner G, Opper D, Rolbetzki A *et al.* Genetic Control of Osmoadaptive Glycine Betaine Synthesis in Bacillus subtilis through the Choline-Sensing and Glycine Betaine-Responsive GbsR Repressor. *J Bacteriol* 2012;**194**:2703–14.

Newell RG, Raimi D. Implications of Shale Gas Development for Climate Change. *Environ Sci Technol* 2014;**48**:8360–8.

Nicot JP, Scanlon BR. Water Use for Shale-Gas Production in Texas, U.S. *Environ Sci Technol* 2012;**46**:3580–6.

Nixon SL, Daly RA, Borton MA *et al.* Genome-Resolved Metagenomics Extends the Environmental Distribution of the Verrucomicrobia Phylum to the Deep Terrestrial Subsurface. *mSphere* 2019;**4**:1–18.

Nixon SL, Walker L, Streets MDT *et al.* Guar Gum Stimulates Biogenic Sulfide Production at Elevated Pressures: Implications for Shale Gas Extraction. *Front Microbiol* 2017;**8**:679.

Notte C, Allen DM, Gehman J *et al.* Comparative Analysis of Hydraulic Fracturing Wastewater Practices in Unconventional Shale Developments: Regulatory Regimes. *Can Water Resour J* 2017;**42**:122–37.

Nuria Obradors JA. Efficient Biodegradation of High-Molecular-Weight Polyethylene Glycols by Pure Cultures of Pseudomonas Stutzeri. *Appl Environ Microbiol* 2015;**57**:2–8.

Nurk S, Meleshko D, Korobeynikov A *et al.* MetaSPAdes: A New Versatile Metagenomic Assembler. *Genome Res* 2017;**27**:824–34.

OECD C. *Organisation for Economic Co-Operation and Development (OECD) 301 Methods*., 1992.

Oetjen K, Chan KE, Gulmark K *et al.* Temporal Characterization and Statistical Analysis of Flowback and Produced Waters and Their Potential for Reuse. *Sci Total Environ* 2018;**619–620**:654–64.

Oksanen AJ, Blanchet FG, Friendly M *et al.* Community Ecology Package 'Vegan.' 2018.

Olm MR, Brown CT, Brooks B *et al.* DRep: A Tool for Fast and Accurate Genomic Comparisons That Enables Improved Genome Recovery from Metagenomes Through De-Replication. *ISME J* 2017;**11**:2864–8.

Orem W, Tatu C, Varonka M *et al.* Organic Substances in Produced and Formation Water from Unconventional Natural Gas Extraction in Coal and Shale. *Int J Coal Geol* 2014;**126**:20–31.

Orem W, Varonka M, Crosby L *et al.* Organic Geochemistry and Toxicology of A Stream Impacted by Unconventional Oil and Gas Wastewater Disposal Operations. *Appl Geochemistry* 2017;**80**:155–67.

Oren A. Thermodynamic Limits to Microbial Life at High Salt Concentrations. *Environ Microbiol* 2011;**13**:1908–23.

Oren A. *Halanaerobium*., 2015.

Örlygsson J, Krooneman J, Collins MD *et al.* Clostridium Acetireducens sp. nov., A Novel Amino Acid-Oxidizing, Acetate-Reducing Anaerobic Bacterium. *Int J Syst Bacteriol* 1996;**46**:454–9.

Parker KM, Zeng T, Harkness J *et al.* Enhanced Formation of Disinfection Byproducts in Shale Gas Wastewater-Impacted Drinking Water Supplies. *Environ Sci Technol* 2014;**48**:11161–9.

Parks DH, Imelfort M, Skennerton CT *et al.* CheckM: Assessing The Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes. *Genome Res* 2015;**25**:1043–55.

Parks DH, Rinke C, Chuvochina M *et al.* Recovery of Nearly 8,000 Metagenome-Assembled Genomes Substantially Expands The Tree of Life. *Nat Microbiol* 2017;**2**:1533–42.

Patterson LA, Konschnik KE, Wiseman H *et al.* Unconventional Oil and Gas Spills: Risks, Mitigation Priorities, and State Reporting Requirements. *Environ Sci Technol* 2017;**51**:2563–73.

Piao S, Ciais P, Huang Y *et al.* The Impacts of Climate Change on Water Resources and Agriculture in China. *Nature* 2010;**467**:43–51.

Preston TM, Anderson CW, Thamke JN *et al.* Predicting Attenuation of Salinized Surface- and Groundwater-Resources from Legacy Energy Development in the Prairie Pothole Region. *Sci Total Environ* 2019;**690**:522–33.

Probst AJ, Ladd B, Jarett JK *et al.* Differential Depth Distribution of Microbial Function and Putative Symbionts Through Sediment-Hosted Aquifers in the Deep Terrestrial Subsurface. *Nat Microbiol* 2018;**3**:328–36.

Prosser JI, Bohannan BJM, Curtis TP *et al.* The Role of Ecological Theory in Microbial Ecology. *Nat Rev Microbiol* 2007;**5**:384–92.

Quast C, Pruesse E, Yilmaz P *et al.* The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools. *Nucleic Acids Res* 2013;**41**:590–6.

Quince C, Walker AW, Simpson JT *et al.* Shotgun Metagenomics, from Sampling to Analysis. *Nat Biotechnol* 2017;**35**:833–44.

R Core Team. R: A Language and Environment for Statistical Computing. 2018.

Rivard C, Lavoie D, Lefebvre R *et al.* An Overview of Canadian Shale Gas Production and Environmental Concerns. *Int J Coal Geol* 2014;**126**:64–76.

Rogers JD, Thurman EM, Ferrer I *et al.* Degradation of Polyethylene Glycols and Polypropylene Glycols in Microcosms Simulating A Spill of Produced Water in Shallow Groundwater. *Environ Sci Process Impacts* 2019;**21**:256–68.

Rokosh CD, Lyster S, Anderson SD a *et al. Summary of Alberta's Shale- and Siltstone-Hosted Hydrocarbon Resource Potential*., 2012.

Rosenblum J, Thurman EM, Ferrer I *et al.* Organic Chemical Characterization and Mass Balance of a Hydraulically Fractured Well: From Fracturing Fluid to Produced Water over 405 Days. *Environ Sci Technol* 2017;**51**:14006–15.

Sakai K. *Routine Soil Analysis Using An Agilent 8800 ICP-QQQ. Application Note.*, 2015.

Scanlon BR, Reedy RC, Xu P *et al.* Datasets Associated with Investigating the Potential for Beneficial Reuse of Produced Water from Oil and Gas Extraction Outside of the Energy Sector. *Data Br* 2020;**30**:105406.

Shaffer M, Borton MA, Mcgivern BB *et al.* DRAM for Distilling Microbial Metabolism to

Automate The Curation of Microbiome Function. *Nucleic Acids Res* 2020:1–18.

Sherwood OA, Rogers JD, Lackey G *et al.* Groundwater Methane in Relation to Oil and Gas Development and Shallow Coal Seams in the Denver-Julesburg Basin of Colorado. *Proc Natl Acad Sci U S A* 2016;**113**:8391–6.

Shi W, Wang X, Guo M *et al.* Water Use for Shale Gas Development in China's Fuling Shale Gas Field. *J Clean Prod* 2020:120680.

Shimadzu Corporation. *Total Organic Carbon Analyzer TOC-V User Manual*., 2001.

Shrestha N, Chilkoor G, Wilder J *et al.* Potential Water Resource impacts of Hydraulic Fracturing from Unconventional Oil Production in the Bakken Shale. *Water Res* 2017;**108**:1–24.

Siddique T, Penner T, Klassen J *et al.* Microbial Communities Involved in Methane Production from Hydrocarbons in Oil Sands Tailings. *Environ Sci Technol* 2012;**46**:9802–10.

Sleator RD, Hill C. Bacterial Osmoadaptation: The role of Osmolytes in Bacterial Stress and Virulence. *FEMS Microbiol Rev* 2002;**26**:49–71.

Smalling KL, Anderson CW, Honeycutt RK *et al.* Associations between Environmental Pollutants and Larval Amphibians in Wetlands Contaminated by Energy-Related Brines are Potentially Mediated by Feeding Traits. *Environ Pollut* 2019;**248**:260–8.

Song M, Liu H, Wang Y *et al.* Enrichment rules and exploration practices of Paleogene shale oil in Jiyang Depression, Bohai Bay Basin, China. *Pet Explor Dev* 2020;**47**:242–53.

Song W, Chang Y, Liu X *et al.* A Multiyear Assessment of Air Quality Benefits from China's Emerging Shale Gas Revolution: Urumqi as A Case Study. *Environ Sci Technol*

2015;**49**:2066–72.

Stolz JF, Ellis DJ, Switzer J *et al.* Sulfurospirillum barnesii sp. nov. and Sulfurospirillum arsenophilum sp. nov., New Members of the Sulfurospirillum Clade of the Proteobacteria. *Int J Syst Bacteriol* 1999;**49**:1177–80.

Strong LC, Gould T, Kasinkas L *et al.* Biodegradation in Waters from Hydraulic Fracturing: Chemistry, Microbiology, and Engineering. *J Environ Eng* 2013;**140**:B4013001.

Struchtemeyer CG, Elshahed MS. Bacterial Communities Associated with Hydraulic Fracturing Fluids in Thermogenic Natural Gas Wells in North Central Texas, USA. *FEMS Microbiol Ecol* 2012;**81**:13–25.

Sugimoto M, Tanabe M, Hataya M *et al.* The First Step in Polyethylene Glycol Degradation by Sphingomonads Proceeds via A Flavoprotein Alcohol Dehydrogenase Containing Flavin Adenine Dinucleotide. *J Bacteriol* 2001;**183**:6694–8.

Sugiyama N, Nakano K. *Reaction Data for 70 Elements Using O2, NH3 and H2 Gases with the Agilent 8800 Triple Quadrupole ICP-MS.*, 2014.

Sun C, Zhang Y, Alessi DS *et al.* Nontarget Profiling of Organic Compounds in A Temporal Series of Hydraulic Fracturing Flowback and Produced Waters. *Environ Int* 2019a;**131**:104944.

Sun Y, Wang D, Tsang DCW *et al.* A Critical Review of Risks, Characteristics, and Treatment Strategies for Potentially Toxic Elements in Wastewater from Shale Gas Extraction. *Environ Int* 2019b;**125**:452–69.

Suzuki MT, Giovannoni SJ. Bias Caused by Template Annealing in The Amplification of Mixtures

of 16S rRNA Genes by PCR. *Appl Environ Microbiol* 1996;**62**:625–30.

Tabatabai MA, Frankenberger WT. *Ion Chromatography: Methods of Soil Analysis, Part 3 - Chemical Methods*. Sparks DL (ed.)., 1996.

Táncsics A, Szabó I, Baka E *et al.* Investigation of Catechol 2,3-Dioxygenase and 16S rRNA Gene Diversity in Hypoxic, Petroleum Hydrocarbon Contaminated Groundwater. *Syst Appl Microbiol* 2010;**33**:398–406.

Tasker TL, Burgos WD, Piotrowski P *et al.* Environmental and Human Health Impacts of Spreading Oil and Gas Wastewater on Roads. *Environ Sci Technol* 2018;**52**:7081–91.

Taylor RS, Stobo B, Niebergall G *et al.* Optimization of Duvernay Fracturing Treatment Design Using Fully Compositional Dual Permeability Numeric Reservoir Simulation. *Soc Pet Eng* 2014:13pp.

Thurman EM, Ferrer I, Blotevogel J *et al.* Analysis of Hydraulic Fracturing Flowback and Produced Waters Using Accurate Mass: Identification of Ethoxylated Surfactants. *Anal Chem* 2014;**86**:9653–61.

Tonglou GUO, Hanrong Z. Formation and Enrichment Mode of Jiaoshiba Shale Gas Field, Sichuan Basin. *Pet Explor Dev* 2014;**41**:31–40.

Townsend-Small A, Marrero JE, Lyon DR *et al.* Integrating Source Apportionment Tracers into a Bottom-up Inventory of Methane Emissions in the Barnett Shale Hydraulic Fracturing Region. *Environ Sci Technol* 2015;**49**:8175–82.

U.S. Energy Information Administration. *World Shale Resource Assessments*., 2013.

U.S. Energy Information Administration. *Utica Shale Play Geology Review*., 2017.

U.S. Environmental Protection Agency. *Review of State and Industry Spill Data: Characterization of Hydaulic Fracturing-Related Spills*., 2015.

U.S. Environmental Protection Agency. *The Hydraulic Fracturing Water Cycle*., 2016.

Ulrich AC, Guigard SE, Foght JM *et al.* Effect of Salt on Aerobic Biodegradation of Petroleum Hydrocarbons in Contaminated Groundwater. *Biodegradation* 2009;**20**:27–38.

Ulrich N, Kirchner V, Drucker R *et al.* Response of Aquatic Bacterial Communities to Hydraulic Fracturing in Northwestern Pennsylvania: A Five-Year Study. *Sci Rep* 2018;**8**:1–12.

United States Environmental Protection Agency. *The Hydraulic Fracturing Water Cycle*., 2016.

Vasi IB, Walker ET, Johnson JS *et al.* "No Fracking Way!" Documentary Film, Discursive Opportunity, and Local Opposition against Hydraulic Fracturing in the United States, 2010 to 2013. *Am Sociol Rev* 2015;**80**:934–59.

Vengosh A, Jackson RB, Warner N *et al.* A Critical Review of the Risks to Water Resources from Shale Gas Development and Hydraulic Fracturing in the United States. *Environ Sci Technol* 2014;**48**:8334–8348.

Větrovský T, Baldrian P. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLoS One* 2013;**8**:1–10.

Vidic RD, Brantley SL, Vandenbossche JM *et al.* Impact of Shale Gas Development on Regional Water Quality. *Science (80-)* 2013;**340**:1235009.

Vikram A, Lipus D, Bibby K. Produced Water Exposure Alters Bacterial Response to Biocides. *Environ Sci Technol* 2014;**48**:13001–9.

Vikram A, Lipus D, Bibby K. Metatranscriptome Analysis of Active Microbial Communities in Produced Water Samples from the Marcellus Shale. *Microb Ecol* 2016;**72**:571–81.

Wadhams GH, Armitage JP. Making Sense of it All: Bacterial Chemotaxis. *Nat Rev Mol Cell Biol* 2004;**5**:1024–37.

Walker CB, Torre JR De, Klotz MG *et al.* Nitrosopumilus Maritimus Genome Reveals Unique Mechanisms for Nitrification and Autotrophy in Globally Distributed Marine Crenarchaea. *Proc Natl Acad Sci U S A* 2010;**107**:8818–23.

Wang B, Xiong M, Wang P *et al.* Chemical Characterization in Hydraulic Fracturing Flowback and Produced Water (HF-FPW) of Shale Gas in Sichuan of China. *Environ Sci Pollut Res* 2020a.

Wang G, Jin Z, Liu G *et al.* Geological Implications of Gamma Ray (GR) Anomalies in Marine Shales: A Case Study of The Ordovician-Silurian Wufeng-Longmaxi Succession in The Sichuan Basin and Its Periphery, Southwest China. *J Asian Earth Sci* 2020b;**199**:104359.

Wang N, Kunz JL, Cleveland D *et al.* Biological Effects of Elevated Major Ions in Surface Water Contaminated by a Produced Water from Oil Production. *Arch Environ Contam Toxicol* 2019;**76**:670–7.

Ward NL, Challacombe JF, Janssen PH *et al.* Three Genomes from the Phylum Acidobacteria Provide Insight into the Lifestyles of These Microorganisms in Soils. *Appl Environ Microbiol*

2009;**75**:2046–56.

Warner NR, Christie CA, Jackson RB *et al.* Impacts of Shale Gas Wastewater Disposal on Water Quality in Western Pennsylvania. *Environ Sci Technol* 2013;**47**:11849–57.

Wasserfallen A, Nölling J, Pfister P *et al.* Phylogenetic Analysis of 18 Thermophilic Methanobacterium Isolates Supports the Proposals to Create A New Genus, Methanothermobacter gen. nov., and to Reclassify Several Isolates in Three Species, Methanothermobacter Thermautotrophicus comb. nov., Methano. *Int J Syst Evol Microbiol* 2000;**50**:43–53.

Westco Scientific. *SmartChem 200 Method 410-200B*., 2007.

Wickham H. *Ggplot2 Elegant Graphics for Data Analysis*., 2009.

Widderich N, Rodrigues CDA, Commichau FM *et al.* Salt-Sensitivity of σH and Spo0A Prevents Sporulation of Bacillus subtilis at High Osmolarity Avoiding Death During Cellular Differentiation. *Mol Microbiol* 2016;**100**:108–24.

Williams PA, Sayers JR. The Evolution of Pathways for Aromatic Hydrocarbon Oxidation in Pseudomonas. *Biodegradation* 1994;**5**:195–217.

Wright KA, Cain RB. Microbial Formation of Methylamine from 4-Carboxy-1-Methylpyridinium Chloride, A Photolytic Product of Paraquat. *Soil Biol Biochem* 1969;**1**:5–14.

Wu Q, Tang L, Wang L *et al. Key Technologies for Wastewater Sewage Treatment in Oil and Gas Fields (in Chinese)*., 2017.

Wu Y-W, Tang Y-H, Tringe SG *et al.* MaxBin: An Automated Binning Method to Recover

Individual Genomes from Metagenomes Using An Expectation Maximization Algorithm. *Microbiome* 2014;**2**:4904–9.

Wuchter C, Banning E, Mincer TJ *et al.* Microbial diversity and methanogenic activity of antrim shale formation waters from recently fractured wells. *Front Microbiol* 2013;**4**:1–14.

Xiong B, Miller Z, Roman-White S *et al.* Chemical Degradation of Polyacrylamide during Hydraulic Fracturing. *Environ Sci Technol* 2018;**52**:327–36.

Xiong B, Purswani P, Pawlik T *et al.* Mechanical degradation of polyacrylamide at ultra high deformation rates during hydraulic fracturing. *Environ Sci Water Res Technol* 2020;**6**:166–72.

Xu L, Wu YH, Cheng H *et al.* Complete Genome Sequence of Erythrobacter seohaensis SW-135T Sheds Light on The Ecological Role of The Genus Erythrobacter for Phosphorus Cycle in The Marine Environment. *Mar Genomics* 2018;**40**:21–4.

Xue Y, Xu Y, Liu Y *et al.* Thermoanaerobacter tengcongensis sp. nov., A Novel Anaerobic, Saccharolytic, Thermophilic Bacterium Isolated from A Hot Spring in Tengcong, China. *Int J Syst Evol Microbiol* 2001;**51**:1335–41.

Yan L, Herrmann M, Kampe B *et al.* Environmental Selection Shapes The Formation of Near-Surface Groundwater Microbiomes. *Water Res* 2020;**170**:115341.

Yilmaz P, Parfrey LW, Yarza P *et al.* The SILVA and "All-Species Living Tree Project (LTP)" Taxonomic Frameworks. *Nucleic Acids Res* 2014;**42**:643–8.

Yu M, Weinthal E, Patiño-Echeverri D *et al.* Water Availability for Shale Gas Development in

Sichuan Basin, China. *Environ Sci Technol* 2016;**50**:2837–45.

Yu Y, Lee C, Kim J *et al.* Group-Specific Primer and Probe Sets to Detect Methanogenic Communities Using Quantitative Real-Time Polymerase Chain Reaction. *Biotechnol Bioeng* 2005;**89**:670–9.

Zagorski WA, Wrightstone GR, Bowman DC. The Appalachian Basin Marcellus Gas Play: Its History of Development, Geologic Controls on Production, and Future Potential As A World-Class Reservoir. *AAPG Mem* 2012;**97**:172–200.

Zhang Y, Yu Z, Zhang H *et al.* Microbial Distribution and Variation in Produced Water from Separators to Storage Tanks of Shale Gas Wells in Sichuan Basin, China. *Environ Sci Water Res Technol* 2017;**3**:340–51.

Zhang Y, Yu Z, Zhang Y *et al.* Regeneration of Unconventional Natural Gas by Methanogens Co-Existing with Sulfate-Reducing Prokaryotes in Deep Shale Wells in China. *Sci Rep* 2020;**10**:1–13.

Zhong C, Li J, Flynn SL *et al.* Temporal Changes in Microbial Community Composition and Geochemistry in Flowback and Produced Water from the Duvernay Formation. *ACS Earth Sp Chem* 2019;**3**:1047–57.

Zhong C, Nesbø CL, Goss GG *et al.* Response of Aquatic Microbial Communities and Bioindicator Modelling of Hydraulic Fracturing Flowback and Produced Water. *FEMS Microbiol Ecol* 2020;**96**:fiaa068.

Zonneveld JP, Macnaughton RB, Utting J *et al.* Sedimentology and Ichnology of The Lower

Triassic Montney Formation in The Pedigree-Ring/Border-Kahntah River Area, Northwestern Alberta and Northeastern British Columbia. *Bull Can Pet Geol* 2010;**58**:115–40.

Zou C, Ni Y, Li J *et al.* The Water Footprint of Hydraulic Fracturing in Sichuan Basin, China. *Sci Total Environ* 2018;**630**:349–56.

7. Appendix

Appendix 1. Supplementary Information for Chapter 2

Result and Discussion

Changes in DOC over 7 Days Incubation The results showed DOC concentrations were reduced from 54.7 mg/L ± 27.5 mg/L to 52.3 mg/L ± 26.8 mg/L (decrease of 4.9% ± 2.7%) for the high PW_1 proportion group (> 5% PW_1) after 7 days. DOC concentrations were reduced from 14.2 mg/L ± 2.0 mg/L to 11.7 mg/L ± 0.7 mg/L (decreases of 17.5% ± 13.6%) for the intermediate PW_1 group (2.5%-5% PW_1) after 7 days. No reduction in DOC concentrations was observed for the low PW_1 proportion group (< 2.5% PW_1), which was initially 10.6 mg/L ± 1.1 mg/L, after 7 days incubation (Appendix 1 Figure S2). DOC concentrations increased (up to 10% at day 7) in the abiotic samples as compared to the initial time point, likely due to oxidation of hydrocarbons that were originally less soluble and thus not detected during DOC measurements.

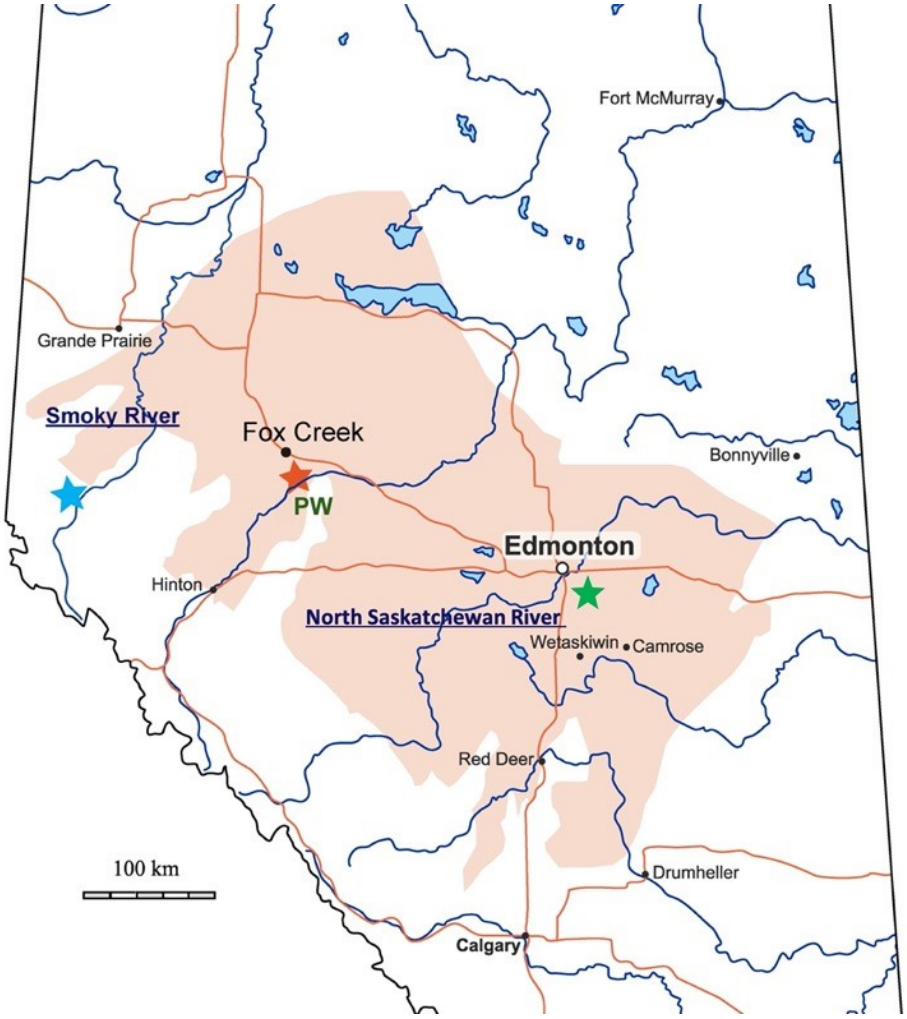

Figure S1 Regional map showing the sampling locations of freshwater samples and flowback and produced water (PW) samples. The PW_1, PW_2 samples, and the hydraulic fracturing source water sample were collected from a shale oil and gas extraction site near the town of Fox Creek, Alberta, Canada (red star). The Smoky River water sample was collected from the blue star location. The North Saskatchewan River sample was collected in the City of Edmonton near the University of Alberta (green star). The red shaded area represents the distribution of the Duvernay Formation. The map is modified from (Rokosh *et al.* 2012).

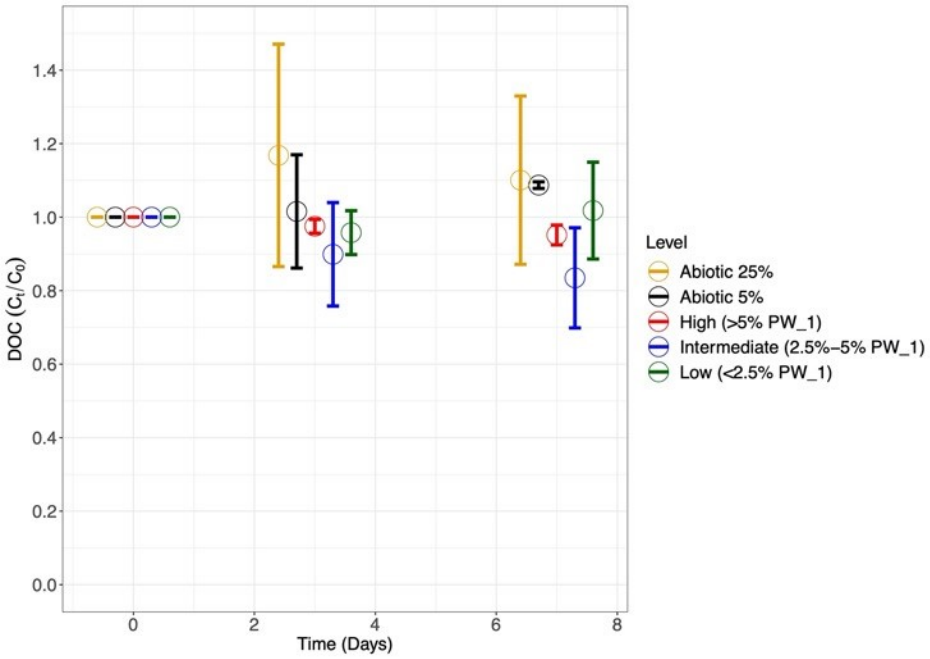

Figure S2 Temporary changes in DOC concentrations over the 7 days of incubation, including five groups: 5% PW_1 and 25% PW_1 abiotic controls (n=2), low PW_1 proportion group (<2.5% PW_1) (n=8), intermediate group (2.5%-5% PW_1) (n=4), and high proportion group (>5% PW_1) (n=4).

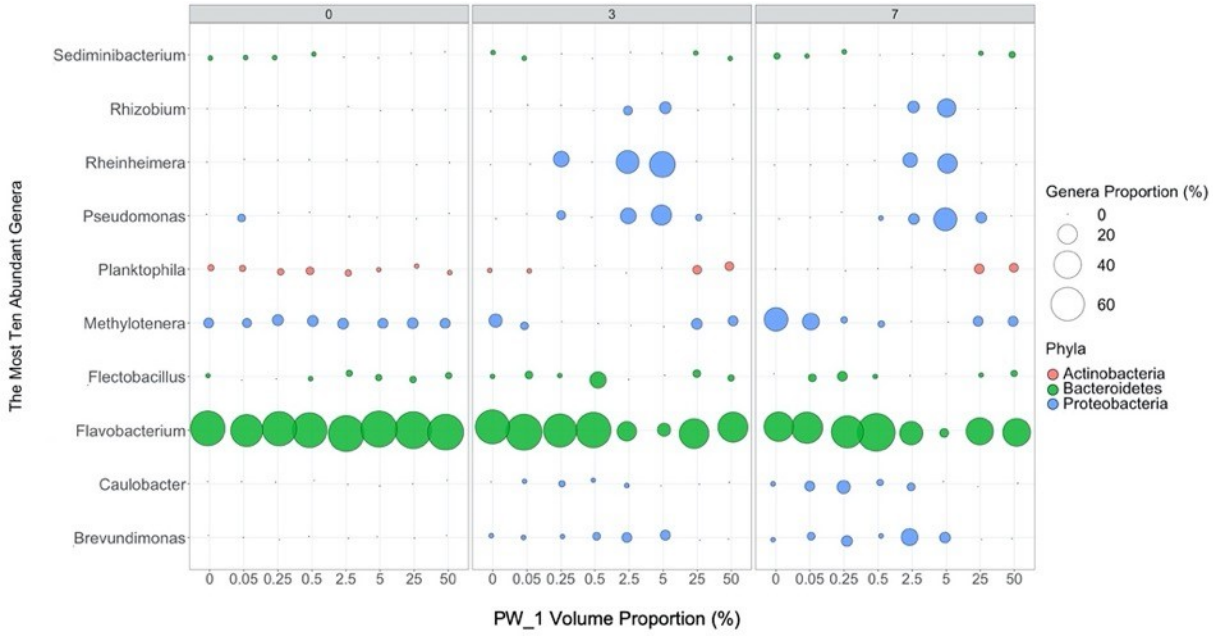

Figure S3 Three-factor plot showing temporal changes of the relative abundance in the ten most abundant genera (y-axis) as a function of time and flowback and produced water (PW_1) mixing ratio. The PW_1 mixing ratios (0, 0.05%, 0.25%, 0.5%, 2.5%, 5%, 25%, and 50%) were labelled on the lower x-axis. Day 0, 3, and 7 exposures are labelled on the higher x-axis and are split into three sub-plots, one for each time point. The relative abundance (%) of a genus is represented by the bubble size, and colors on the y-axis represent the phyla of the ten genera.

Table S1 Total numbers of non-chimeric sequences for each mixture sample used for taxonomic

Time (Day)	PW_1 Proportion (%)	Non-chimeric Sequences
0	0	381597
0	0.05	309857
0	0.25	311813
0	0.5	305853
0	2.5	305644
0	5	304817
0	25	265476
0	50	303489
3	0	352290
3	0.05	323176
3	0.25	421973
3	0.5	403394
3	2.5	350604
3	5	326140
3	25	303656
3	50	336612
7	0	207035
7	0.05	233575
7	0.25	305377

7	0.5	345418
7	2.5	345408
7	5	309564
7	25	281190
7	50	234972

Table S2 Inorganic elements in flowback and produced water (PW_1 and PW_2) and Smoky River freshwater

Sample Types	Mg (mg/L)	K (mg/L)	Ca (mg/L)	Mn (mg/L)	Fe (mg/L)	Zn (mg/L)	S (mg/L)
PW_1	924	2325	12564	6.23	12.26	1.44	236
PW_2	942	2115	12295	16	7	0.41	49
Smoky River freshwater	9.26	1.07	28.8	BDL	BDL	BDL	14.27

BDL: below the detection limit, Table S2 Detailed inorganic chemistry profiles for both PW samples and freshwater. PW_1 and PW_2 are highly similar in composition, while PWs were significantly different from the freshwater.

Table S3 Full list of Tukey's Test results for comparing microbial diversity between the three representative groups (< 2.5% PW_1, 2.5%-5% PW_1, and > 5% PW_1). Pure freshwater (0% PW_1) was used for control groups. Tukey's Test was a post-hoc analysis for ANOVA analyses. Tukey's test compares the means of all treatments to the mean of every other treatment. The significant thresholds are $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

Groups	Diversity Indexes	Time (day)	diff	lwr	upr	P Value
0.05-0.5 vs 0	Observed ASVs	3	-61.0000	-201.4002	79.4002	0.4013
2.5-5 vs 0	Observed ASVs	3	-70.0000	-218.9170	78.9170	0.3498
25-50 vs 0	Observed ASVs	3	16.5000	-132.4170	165.4170	0.9660
2.5-5 vs 0.05-0.5	Observed ASVs	3	-9.0000	-119.9961	101.9961	0.9859
25-50 vs 0.05-0.5	Observed ASVs	3	77.5000	-33.4961	188.4961	0.1439
25-50 vs 2.5-5	Observed ASVs	3	86.5000	-35.0902	208.0902	0.1369
0.05-0.5 vs 0	Observed ASVs	7	7.6667	-42.1447	57.4781	0.9184
2.5-5 vs 0	Observed ASVs	7	24.5000	-28.3330	77.3330	0.3585
25-50 vs 0	Observed ASVs	7	102.0000	49.1670	154.8330	0.0049
2.5-5 vs 0.05-0.5	Observed ASVs	7	16.8333	-22.5460	56.2127	0.4122
25-50 vs 0.05-0.5	Observed ASVs	7	94.3333	54.9540	133.7127	0.0022
25-50 vs 2.5-5	Observed ASVs	7	77.5000	34.3621	120.6379	0.0064
0.05-0.5 vs 0	Shannon	3	-0.4501	-0.9840	0.0837	0.0847
2.5-5 vs 0	Shannon	3	0.2446	-0.3216	0.8108	0.4052
25-50 vs 0	Shannon	3	0.7898	0.2236	1.3560	0.0162
2.5-5 vs 0.05-0.5	Shannon	3	0.6947	0.2727	1.1168	0.0089

25-50 vs 0.05-0.5	Shannon	3	1.2399	0.8179	1.6620	0.0010
25-50 vs 2.5-5	Shannon	3	0.5452	0.0829	1.0075	0.0290
0.05-0.5 vs 0	Shannon	7	-0.0790	-1.7792	1.6212	0.9972
2.5-5 vs 0	Shannon	7	0.6002	-1.2031	2.4036	0.5813
25-50 vs 0	Shannon	7	1.3092	-0.4942	3.1125	0.1296
2.5-5 vs 0.05-0.5	Shannon	7	0.6792	-0.6649	2.0234	0.3047
25-50 vs 0.05-0.5	Shannon	7	1.3881	0.0440	2.7323	0.0451
25-50 vs 2.5-5	Shannon	7	0.7089	-0.7635	2.1813	0.3346
0.05-0.5 vs 0	Chao1	3	-60.1667	-198.3938	78.0605	0.4001
2.5-5 vs 0	Chao1	3	-69.1667	-215.7787	77.4454	0.3474
25-50 vs 0	Chao1	3	15.5000	-131.1120	162.1120	0.9701
2.5-5 vs 0.05-0.5	Chao1	3	-9.0000	-118.2782	100.2782	0.9852
25-50 vs 0.05-0.5	Chao1	3	75.6667	-33.6115	184.9448	0.1471
25-50 vs 2.5-5	Chao1	3	84.6667	-35.0416	204.3749	0.1391
0.05-0.5 vs 0	Chao1	7	7.7778	-41.9788	57.5344	0.9151
2.5-5 vs 0	Chao1	7	26.4375	-26.3374	79.2124	0.3100
25-50 vs 0	Chao1	7	102.0000	49.2251	154.7749	0.0049
2.5-5 vs 0.05-0.5	Chao1	7	18.6597	-20.6763	57.9958	0.3439
25-50 vs 0.05-0.5	Chao1	7	94.2222	54.8862	133.5583	0.0022
25-50 vs 2.5-5	Chao1	7	75.5625	32.4720	118.6530	0.0070
0.05-0.5 vs 0	Inverse Simpson	3	-0.2571	-2.9225	2.4083	0.9768
2.5-5 vs 0	Inverse Simpson	3	3.1086	0.2815	5.9357	0.0367
25-50 vs 0	Inverse Simpson	3	1.5254	-1.3017	4.3525	0.2663
2.5-5 vs 0.05-0.5	Inverse Simpson	3	3.3657	1.2585	5.4729	0.0099
25-50 vs 0.05-0.5	Inverse Simpson	3	1.7825	-0.3247	3.8897	0.0839

25-50 vs 2.5-5	Inverse Simpson	3	-1.5832	-3.8915	0.7251	0.1508
0.05-0.5 vs 0	Inverse Simpson	7	-0.5711	-4.0173	2.8751	0.9017
2.5-5 vs 0	Inverse Simpson	7	3.3183	-0.3369	6.9736	0.0678
25-50 vs 0	Inverse Simpson	7	2.2584	-1.3968	5.9137	0.1959
2.5-5 vs 0.05-0.5	Inverse Simpson	7	3.8894	1.1650	6.6139	0.0149
25-50 vs 0.05-0.5	Inverse Simpson	7	2.8295	0.1051	5.5540	0.0442
25-50 vs 2.5-5	Inverse Simpson	7	-1.0599	-4.0444	1.9246	0.5381

diff: difference between means of the two groups; lwr, upr: the lower and the upper end points of the confidence intervals at 95%

Table S4 Relative abundance of the ten most abundant bacteria of the entire community. Multiple amplicon sequence variants (ASVs) could be classified to the same Genus.

ASV	Relative Abundance (%)	Time	Ratio	Kingdom	Genus
ASV16598	74	0	25	Bacteria	*Flavobacterium*
ASV16598	71	0	2.5	Bacteria	*Flavobacterium*
ASV16598	71	0	5	Bacteria	*Flavobacterium*
ASV16598	70	0	50	Bacteria	*Flavobacterium*
ASV16598	66	0	0	Bacteria	*Flavobacterium*
ASV16598	66	0	0.5	Bacteria	*Flavobacterium*
ASV16598	64	0	0.25	Bacteria	*Flavobacterium*
ASV16598	56	0	0.05	Bacteria	*Flavobacterium*
ASV18585	6	0	0.5	Bacteria	*Methylotenera*
ASV18585	6	0	0.25	Bacteria	*Methylotenera*
ASV18585	6	0	25	Bacteria	*Methylotenera*
ASV18585	6	0	2.5	Bacteria	*Methylotenera*
ASV18585	5	0	50	Bacteria	*Methylotenera*
ASV18585	5	0	0	Bacteria	*Methylotenera*
ASV18585	5	0	5	Bacteria	*Methylotenera*
ASV18585	4	0	0.05	Bacteria	*Methylotenera*
ASV18634	3	0	0.05	Bacteria	*Pseudomonas*
ASV13804	3	0	0.5	Bacteria	*Planktophila*
ASV13804	2	0	0	Bacteria	*Planktophila*
ASV16930	2	0	50	Bacteria	*Flectobacillus*
ASV13804	2	0	0.25	Bacteria	*Planktophila*
ASV16930	2	0	25	Bacteria	*Flectobacillus*
ASV16930	2	0	5	Bacteria	*Flectobacillus*
ASV16930	2	0	2.5	Bacteria	*Flectobacillus*
ASV13804	2	0	2.5	Bacteria	*Planktophila*
ASV13804	2	0	0.05	Bacteria	*Planktophila*
ASV16598	71	3	0.05	Bacteria	*Flavobacterium*
ASV16598	68	3	0.5	Bacteria	*Flavobacterium*
ASV16598	63	3	0	Bacteria	*Flavobacterium*
ASV16598	60	3	0.25	Bacteria	*Flavobacterium*
ASV16598	49	3	50	Bacteria	*Flavobacterium*
ASV16598	48	3	25	Bacteria	*Flavobacterium*
ASV19210	36	3	5	Bacteria	*Rheinheimera*
ASV19210	28	3	2.5	Bacteria	*Rheinheimera*
ASV18634	22	3	5	Bacteria	*Pseudomonas*
ASV16598	20	3	2.5	Bacteria	*Flavobacterium*

ASV16930	14	3	0.5	Bacteria	*Flectobacillus*
ASV18634	13	3	2.5	Bacteria	*Pseudomonas*
ASV19210	13	3	0.25	Bacteria	*Rheinheimera*
ASV16598	9	3	5	Bacteria	*Flavobacterium*
ASV18585	9	3	0	Bacteria	*Methylotenera*
ASV15854	7	3	5	Bacteria	*Rhizobium*
ASV18585	6	3	25	Bacteria	*Methylotenera*
ASV15173	5	3	5	Bacteria	*Brevundimonas*
ASV15173	5	3	2.5	Bacteria	*Brevundimonas*
ASV18585	5	3	50	Bacteria	*Methylotenera*
ASV15854	4	3	2.5	Bacteria	*Rhizobium*
ASV18634	4	3	0.25	Bacteria	*Pseudomonas*
ASV13804	4	3	50	Bacteria	*Planktophila*
ASV13804	4	3	25	Bacteria	*Planktophila*
ASV16930	3	3	0.05	Bacteria	*Flectobacillus*
ASV15173	3	3	0.5	Bacteria	*Brevundimonas*
ASV18585	3	3	0.05	Bacteria	*Methylotenera*
ASV16930	3	3	25	Bacteria	*Flectobacillus*
ASV16930	2	3	50	Bacteria	*Flectobacillus*
ASV18634	2	3	25	Bacteria	*Pseudomonas*
ASV15177	2	3	0.25	Bacteria	*Caulobacter*
ASV16598	78	7	0.5	Bacteria	*Flavobacterium*
ASV16598	56	7	0.25	Bacteria	*Flavobacterium*
ASV16598	53	7	0.05	Bacteria	*Flavobacterium*
ASV16598	47	7	0	Bacteria	*Flavobacterium*
ASV16598	42	7	50	Bacteria	*Flavobacterium*
ASV16598	40	7	25	Bacteria	*Flavobacterium*
ASV18585	30	7	0	Bacteria	*Methylotenera*
ASV16598	29	7	2.5	Bacteria	*Flavobacterium*
ASV18634	28	7	5	Bacteria	*Pseudomonas*
ASV19210	20	7	5	Bacteria	*Rheinheimera*
ASV15854	18	7	5	Bacteria	*Rhizobium*
ASV18585	15	7	0.05	Bacteria	*Methylotenera*
ASV15173	15	7	2.5	Bacteria	*Brevundimonas*
ASV19210	11	7	2.5	Bacteria	*Rheinheimera*
ASV15177	9	7	0.25	Bacteria	*Caulobacter*
ASV15854	7	7	2.5	Bacteria	*Rhizobium*
ASV15173	6	7	0.25	Bacteria	*Brevundimonas*
ASV18634	6	7	25	Bacteria	*Pseudomonas*
ASV18634	6	7	2.5	Bacteria	*Pseudomonas*
ASV15173	6	7	5	Bacteria	*Brevundimonas*

ASV18585	5	7	50	Bacteria	*Methylotenera*
ASV18585	5	7	25	Bacteria	*Methylotenera*
ASV15177	5	7	0.05	Bacteria	*Caulobacter*
ASV16930	5	7	0.25	Bacteria	*Flectobacillus*
ASV13804	5	7	25	Bacteria	*Planktophila*
ASV16598	4	7	5	Bacteria	*Flavobacterium*
ASV13804	4	7	50	Bacteria	*Planktophila*
ASV15177	3	7	2.5	Bacteria	*Caulobacter*
ASV15173	3	7	0.05	Bacteria	*Brevundimonas*
ASV16930	3	7	0.05	Bacteria	*Flectobacillus*
ASV16557	2	7	0	Bacteria	*Sediminibacterium*
ASV16930	2	7	50	Bacteria	*Flectobacillus*
ASV16557	2	7	50	Bacteria	*Sediminibacterium*
ASV15177	2	7	0.5	Bacteria	*Caulobacter*
ASV18585	2	7	0.25	Bacteria	*Methylotenera*
ASV18585	2	7	0.5	Bacteria	*Methylotenera*

Table S5 Relative abundance of archaea. Multiple amplicon sequence variants (ASVs) could be classified to the same Genus.

ASV	Relative Abundance (%)	Time	Ratio	Kingdom	Genus
ASV5527	0.29	0	0	Archaea	*Nitrosopumilus*
ASV5527	0.26	0	0.5	Archaea	*Nitrosopumilus*
ASV5527	0.24	0	0.25	Archaea	*Nitrosopumilus*
ASV5527	0.09	0	0.05	Archaea	*Nitrosopumilus*
ASV5527	0.05	0	2.5	Archaea	*Nitrosopumilus*
ASV5527	0.04	0	5	Archaea	*Nitrosopumilus*
ASV5452	0.04	0	0.5	Archaea	*Methanosaeta*
ASV5546	0.04	0	0	Archaea	*Candidatus Nitrososphaera*
ASV5434	0.03	0	0.5	Archaea	*Candidatus Methanoregula*
ASV5527	0.03	0	25	Archaea	*Nitrosopumilus*
ASV5527	0.03	0	50	Archaea	*Nitrosopumilus*
ASV5468	0.02	0	0	Archaea	*Coxiella*
ASV5653	0.02	0	0.5	Archaea	*Halobacteriaceae*
ASV5446	0.02	0	0.25	Archaea	*Methanohalophilus*
ASV5546	0.02	0	0.05	Archaea	*Candidatus Nitrososphaera*
ASV5434	0.02	0	0	Archaea	*Candidatus Methanoregula*
ASV5546	0.02	0	2.5	Archaea	*Candidatus Nitrososphaera*
ASV5434	0.02	0	0.25	Archaea	*Candidatus Methanoregula*
ASV5449	0.02	0	5	Archaea	*Methanosarcina*
ASV5546	0.02	0	50	Archaea	*Candidatus Nitrososphaera*
ASV5452	0.02	0	0.25	Archaea	*Methanosaeta*
ASV5490	0.01	0	0.25	Archaea	*Methanobacterium*
ASV5935	0.01	0	0.5	Archaea	*Methermicoccus*
ASV5455	0.01	0	0.5	Archaea	*Methanofollis*
ASV5427	0.01	0	0	Archaea	*Methanospirillum*
ASV5517	0.01	0	0.25	Archaea	*Nitrosotalea*
ASV5434	0.01	0	50	Archaea	*Candidatus Methanoregula*
ASV5449	0.01	0	0.25	Archaea	*Methanosarcina*
ASV5490	0.01	0	0	Archaea	*Methanobacterium*
ASV5434	0.01	0	0.05	Archaea	*Candidatus Methanoregula*
ASV5527	0.16	3	25	Archaea	*Nitrosopumilus*
ASV5527	0.15	3	50	Archaea	*Nitrosopumilus*
ASV5527	0.15	3	0	Archaea	*Nitrosopumilus*
ASV5527	0.14	3	0.05	Archaea	*Nitrosopumilus*
ASV5546	0.06	3	50	Archaea	*Candidatus Nitrososphaera*
ASV5546	0.05	3	25	Archaea	*Candidatus Nitrososphaera*

ASV5546	0.04	3	0	Archaea	*Candidatus Nitrososphaera*
ASV5434	0.04	3	50	Archaea	*Candidatus Methanoregula*
ASV5452	0.03	3	25	Archaea	*Methanosaeta*
ASV5527	0.03	3	2.5	Archaea	*Nitrosopumilus*
ASV5434	0.03	3	25	Archaea	*Candidatus Methanoregula*
ASV5527	0.02	3	5	Archaea	*Nitrosopumilus*
ASV5517	0.02	3	0	Archaea	*Nitrosotalea*
ASV5527	0.02	3	0.5	Archaea	*Nitrosopumilus*
ASV5452	0.02	3	50	Archaea	*Methanosaeta*
ASV5449	0.01	3	0	Archaea	*Methanosarcina*
ASV5490	0.01	3	50	Archaea	*Methanobacterium*
ASV5446	0.01	3	2.5	Archaea	*Methanohalophilus*
ASV5449	0.01	3	0.05	Archaea	*Methanosarcina*
ASV5490	0.01	3	0.05	Archaea	*Methanobacterium*
ASV5935	0.01	3	0	Archaea	*Methermicoccus*
ASV5449	0.01	3	0.5	Archaea	*Methanosarcina*
ASV5435	0.01	3	50	Archaea	*Methanocella*
ASV5527	0.23	7	25	Archaea	*Nitrosopumilus*
ASV5527	0.23	7	50	Archaea	*Nitrosopumilus*
ASV5527	0.06	7	0	Archaea	*Nitrosopumilus*
ASV5546	0.05	7	50	Archaea	*Candidatus Nitrososphaera*
ASV5434	0.04	7	50	Archaea	*Candidatus Methanoregula*
ASV5546	0.04	7	25	Archaea	*Candidatus Nitrososphaera*
ASV5527	0.04	7	0.05	Archaea	*Nitrosopumilus*
ASV5527	0.04	7	5	Archaea	*Nitrosopumilus*
ASV5452	0.03	7	25	Archaea	*Methanosaeta*
ASV5517	0.03	7	25	Archaea	*Nitrosotalea*
ASV5434	0.02	7	25	Archaea	*Candidatus Methanoregula*
ASV5527	0.02	7	2.5	Archaea	*Nitrosopumilus*
ASV5527	0.02	7	0.5	Archaea	*Nitrosopumilus*
ASV5452	0.02	7	50	Archaea	*Methanosaeta*
ASV5446	0.02	7	25	Archaea	*Methanohalophilus*
ASV5487	0.02	7	50	Archaea	*Methanosphaera*
ASV5427	0.01	7	50	Archaea	*Methanospirillum*
ASV5490	0.01	7	25	Archaea	*Methanobacterium*
ASV5449	0.01	7	25	Archaea	*Methanosarcina*

Table S6 Envfit correlation of the top 10 genera to the sample dissimilarities in the NMDS ordination

	NMDS1	NMDS2	r2	Pr(>r)			
Planktophila	-0.99978	-0.02114	0.6561	0.001	***		
Brevundimonas	0.55169	0.83405	0.4252	0.011	*		
Caulobacter	0.81823	-0.57489	0.3984	0.012	*		
Rhizobium	0.23506	0.97198	0.4013	0.017	*		
Sediminibacterium	-0.2141	-0.97681	0.7354	0.001	***		
Flavobacterium	-0.24078	-0.97058	0.2357	0.064	.		
Flectobacillus	0.19912	-0.97998	0.0295	0.728			
Methylotenera	-0.02844	-0.9996	0.6114	0.001	***		
Pseudomonas	0.22731	0.97382	0.401	0.011	*		
Rheinheimera	0.34883	0.93718	0.5116	0.005	**		
Significance	codes:	0	'***'	0.001	'**'	0.01	'*'
Permutation:	free						
Number	of	permutations:	999				

Table S7 Confusion matrix of the average errors for each spill status corresponding to the three represented groups. The table shows the average error of 1000 runs of prediction for each group.

Groups (PW_1 Treatment Groups)	< 2.5%	2.5%-5%	> 5%
Average Error (1000 Runs)	0.012%	0	0.325%

Appendix 2. Supplementary Information for Chapter 3

Methods

Sampling To minimize contamination, all equipment used for sampling soils was thoroughly sterilized using 70% ethanol in advance. Soil sampling was done on 12 August 2019. Surface vegetation was removed before collecting topsoil. Fox Creek (FC) soil was collected in the morning (about 11:00). Sample was taken from the Ah horizon (Appendix 2 Figure S1). The soil was identified to be an orthic grey luvisol. Grand Prairie (GP) soil was collected was collected in the afternoon (about 17:00). The sample was taken from the Ap horizon (Appendix 2 Figure S1). The soil is likely a gleyed solonetzic black chernozem (https://soil.agric.gov.ab.ca/agrasidviewer/). Before collecting GP soil, all the collection tools were sterilized again using ethanol. For both soils, large particles such as rocks and roots were sieved out on site by using a 2mm stainless steel sieve. After collecting a sufficient amount of soil, the sampling hole was filled back again.

Inorganic analyses For soil samples, water soluble ions were extracted using a 1:5 soil to milli-Q water. SO_4^{2-} concentration was analyzed using the EPA Method 375.4 using barium chloride extraction (Guo *et al.* 2018). NH_4^+ concentration was analyzed using the Salicylate-Hypochlorite Method (Bower and Holm-Hansen 1980). NO_2^- and NO_3^- concentrations were analyzed using the EPA method 353.1 (National Environmental Methods Index 1978a). PO_4^{3-} concentration was analyzed using the EPA method 365.1 with Molybdenum Blue (National Environmental Methods Index 1993). Cl^- concentration was analyzed using EPA method 325.2 with ferrithiocyanate (National Environmental Methods Index 1978b).

Pre-processing samples for nontarget organic analyses Twenty milliliters of the FPW sample and a source water sample were filtered using a glass fiber membrane (Glass Fiber Store, 90 mm diameter, pore size: 0.4 μm). The aqueous filtrate was processed using solid phase extraction. Briefly, Oasis-HLB cartridges (Waters, 150 mg/6 mL) were conditioned with 2 mL of methanol (HPLC grade) followed by 2 mL of 1% ammonium hydroxide, 2 mL of 0.1M formic acid, and then 2 mL of pure water (LC-MS grade). Afterwards, 7 mL of aqueous filtrate was loaded to the cartridge and subsequently washed with 7 mL of pure water and 2 mL of 0.1 M formic acid. The cartridge was vacuum dried and organic compounds eluted using 2 mL of methanol followed by 2 mL of 0.2% ammonium hydroxide in methanol. The eluate was evaporated to near-dryness under a gentle stream of high-purity nitrogen at 40 °C. The sample was then reconstituted with 300 μL of methanol and 200 μL of pure water for analysis using an HPLC-Orbitrap-MS.

Manual refinement for metagenome-assembled genomes For bins used for functional annotation, bins were visually inspected for contamination using the anvi-refine interface in Anvi'o (Eren *et al.* 2015). To remove contaminating sequences not detected in Anvi'o, bins were processed using refineM v0.1.1 (Parks *et al.* 2017). Bins were then imported into Geneious v7.0.6 for further rounds of assembly and refinement (Kearse *et al.* 2012).

Statistical analyses Data processing was conducted using R (v.4.0.1) (R Core Team 2018) using ggplot2 for data visualization.(Wickham 2009) For 16S rRNA gene sequencing analyses, alpha-diversity analyses were conducted in Phyloseq,(McMurdie and Holmes 2013) including assessments of observed ASVs, Chao1 richness, Shannon diversity, and Inverse Simpson diversity. Samples were rarefied to even depth for beta-diversity analyses using Bray-Curtis distance in Principal Coordinates Analysis (PCoA) in Phyloseq.(McMurdie and Holmes 2013) The FANTAXTIC package was used to construct taxonomic bar charts

(https://rdrr.io/github/gmteunisse/Fantaxtic/). The significant factors that influence the PCoA ordinations were tested by PERMANOVA analysis. ANOVA combined with Tukey HSD analyses was used to test the significance of results (p=0.05 was used as cut-off value) related to physicochemical, 16S rRNA gene sequencing, shotgun metagenomics, and soil respiration analyses.

Results and Discussion

Vegetation at the sampling sites At the FC site, major trees were *Populus tremuloides*, *Picea glauca*, and *Pinus contorta*. Identified shrubs were *Alnus viridis* and *Salix* spp. Floor vegetation consisted of *Taraxacum officinale*, *Trifolium repens*, *Trifolium pratense*, *Chamerion angustifolium*, *Eurybia conspicua*. At the GP site, the major vegetation was made of *Populus tremuloides* trees, shrubs like *Cornus sericeae*, *Salix* sp., and *Rosa* sp., and *Eriophorum chamissonis*, *Agrostis scabra*, *Phleum pratense*, *Poa pratensis* on the floor.

FPW characteristics For FPW, NH_4^+ made up the largest fraction (412.72 ± 10.43 mg L^{-1}) of total N, with lower concentrations of NO_2^- (below the detection limit) and NO_3^- (0.03 ± 0.00 mg L^{-1}) (Table 3.1, Appendix 2 Table S1). Similarly, S present as SO_4^{2-} was only a small fraction of the total S. Notably, FPW samples were collected from anoxic subsurface environments and an oil layer on top of the fluid samples inhibited oxidation, which may contribute to the predominant presence of these detected ligands in their reduced forms.

FC luvisol and GP chernozem soil taxonomic diversity and compositions The observed ASV and Shannon diversity for FC luvisol were 133 and 4.48, respectively. While, 40 and 3.35, respectively, for the GP chernozem. FC luvisol soil contained more diverse microbial species than that for GP chernozem soil in terms of the top 100 ASV (Appendix 2 Figure S2).

The predominant bacterial phyla in FC luvisol were Actinobacteria and Proteobacteria, together comprising 39.4% of the total community, while GP chernozem community primarily consisted of the bacterial phyla Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia, together comprising 77.66% of the total sequences. The higher biodiversity in FC luvisol (Figure 3.2B, Appendix 2 Figure S2) may be attributed to diverse species of low abundance (relative abundance < 1% of the total sequences). More abundant phyla Acidobacteria and Actinobacteria, including predominant acidophilic bacterial genera *Acidothermus* (10.87% of the total sequences) and *Candidatus Solibacter* (6.37% of the total sequences), may reflect the effect of the lower pH of GP chernozem.(Fierer and Jackson 2006)

pH is one of the major factors that driving the taxonomic compositions and diversity in soil (Fierer and Jackson 2006). GP chernozem soil has higher relative abundance of Acidobacteria and Actinobacteria and their species are widely found in acidic to extreme acidic environment. In the genus level, bacterial genera *Acidothermus* (10.87% of the total sequences), *Candidatus Udaeobacter* (7.78% of the total sequences), *Candidatus Solibacter* (6.37% of the total sequences), Xanthobacteraceae (5.90% of the total sequences), Gemmatimonadaceae (5.44% of the total sequences), Ellin6067 (5.34 of the total sequences). Within the Acidobacteria and Actinobacteria, the phyla shown higher abundance in chernozem soil than FC luvisol, *Acidothermus cellulolyticus* as the sole species of the genus *Acidothermus* is acidophilic (optimal growth: pH 5-6).(Mohagheghi, Grohmann and Himmel 1986; Barabote *et al.* 2009) *Candidatus Solibacter usitatus* was previously reported to grow at pH 3.5 to 6.5.(Ward *et al.* 2009) The growth pH ranges of these species potentially present or phylogenetic close to the ones within in the abundant bacterial genera observed in this study are within the pH of GP chernozem soil, indicating that pH is one of the major driving factors in shaping taxonomic compositions in the GP chernozem.

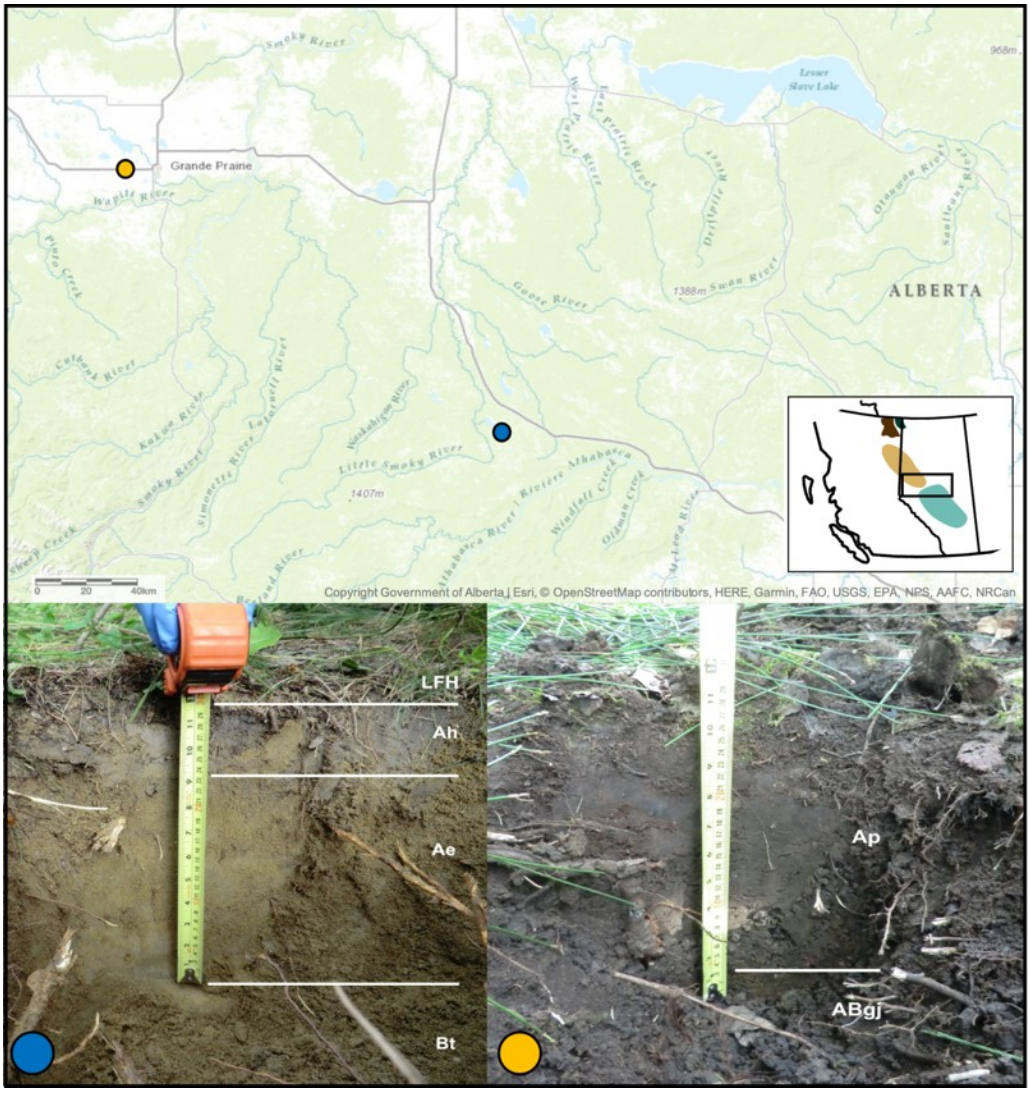

Figure S1 Top: Location of soil sampling sites in Fox Creek and Grande Prairie in the Western Sedimentary Canadian Basin, Canada. Sample locations are marked with a blue dot for FC luvisol and a yellow dot for chernozem soil. The boundaries of the Duvernay and Montney unconventional plays are shaded in light green and light brown, respectively (the Horn River and Cordova unconventional plays are shaded in dark brown and green, respectively). Bottom: Soil profiles for FC luvisol (left) and GP chernozem (right).

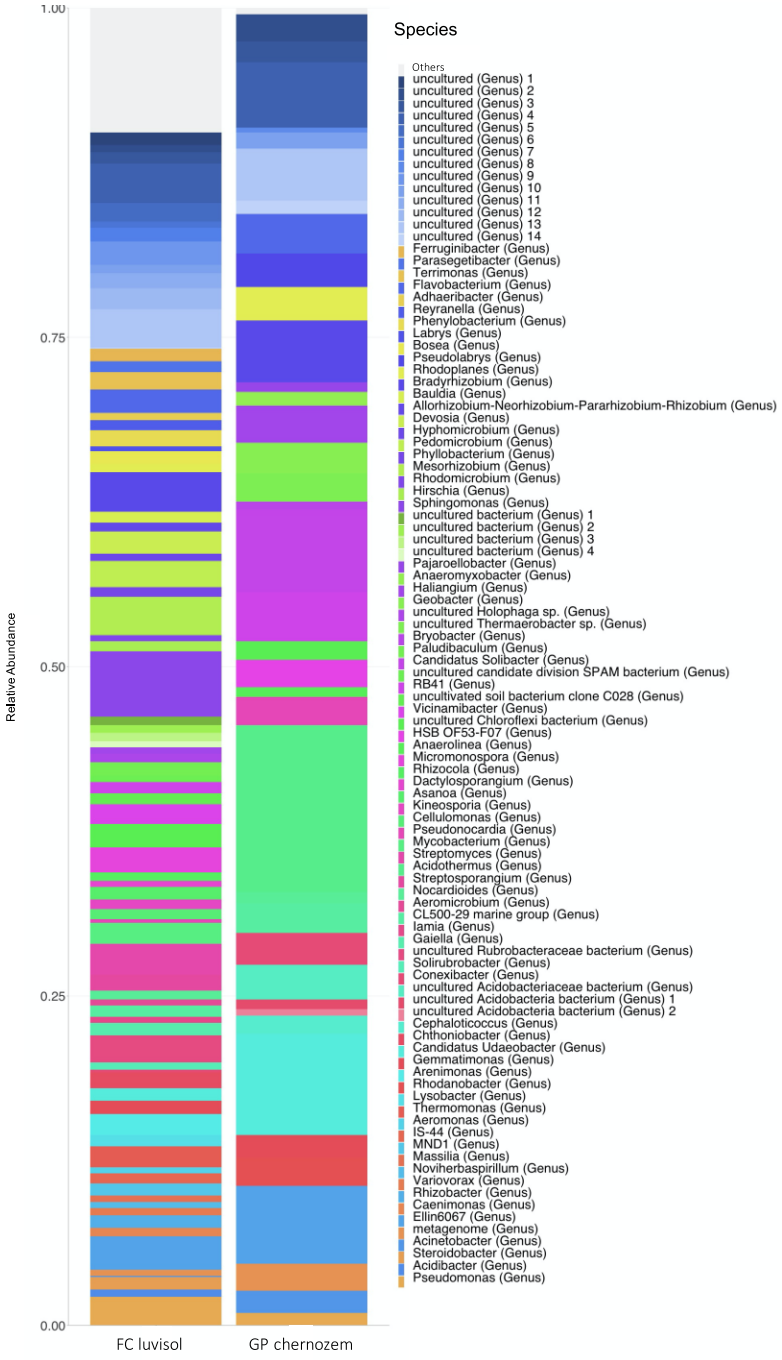

Figure S2 Microbial community profile revealed by 16S rRNA amplicon-based analysis showing the relative abundance of the top 100 microbial species in FC luvisol and GP chernozem. Genera are grouped by similar colour.

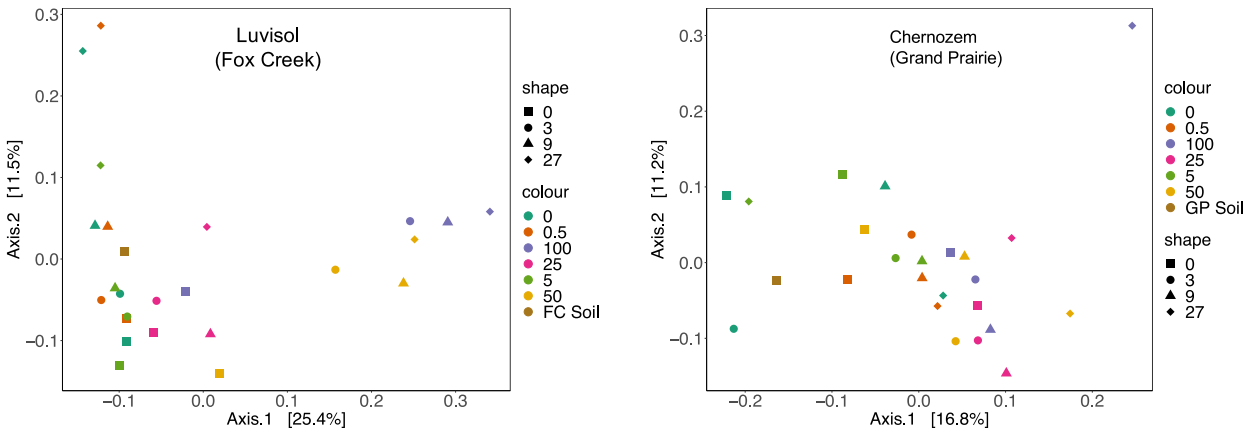

Figure S3 PCoA ordinations showing changes of microbial communities of FC luvisol and GP chernozem exposing to different concentrations of FPW; the changes of microbial community compositions in FC luvisol exposing to high concentrations of FPW were more potent than GP chernozem.

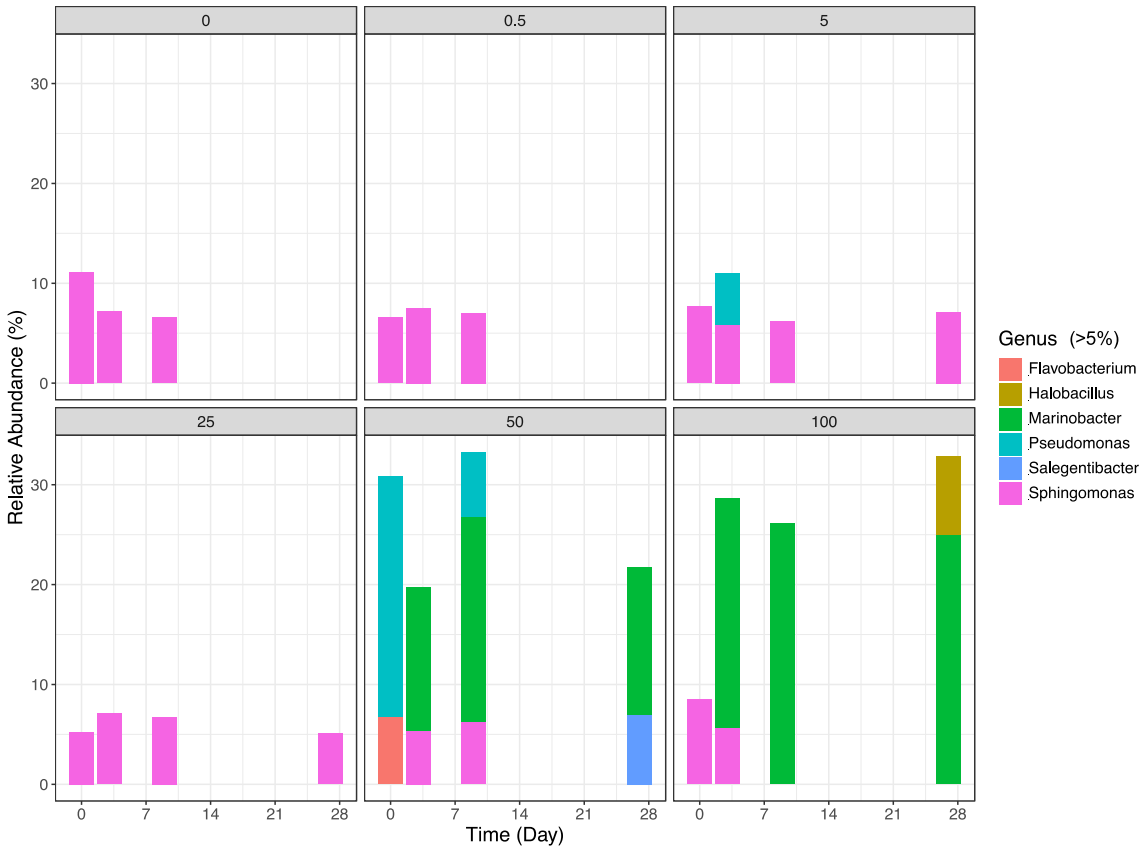

Figure S4 Change in relative abundance of the *Marinobacter* of the microbial communities in the FC luvisol using 16S rRNA amplicon-based sequences. Genus of relative abundance > 5% of the total sequences were shown. The grids titles represent different FPW fraction (%) of the total volume fluids.

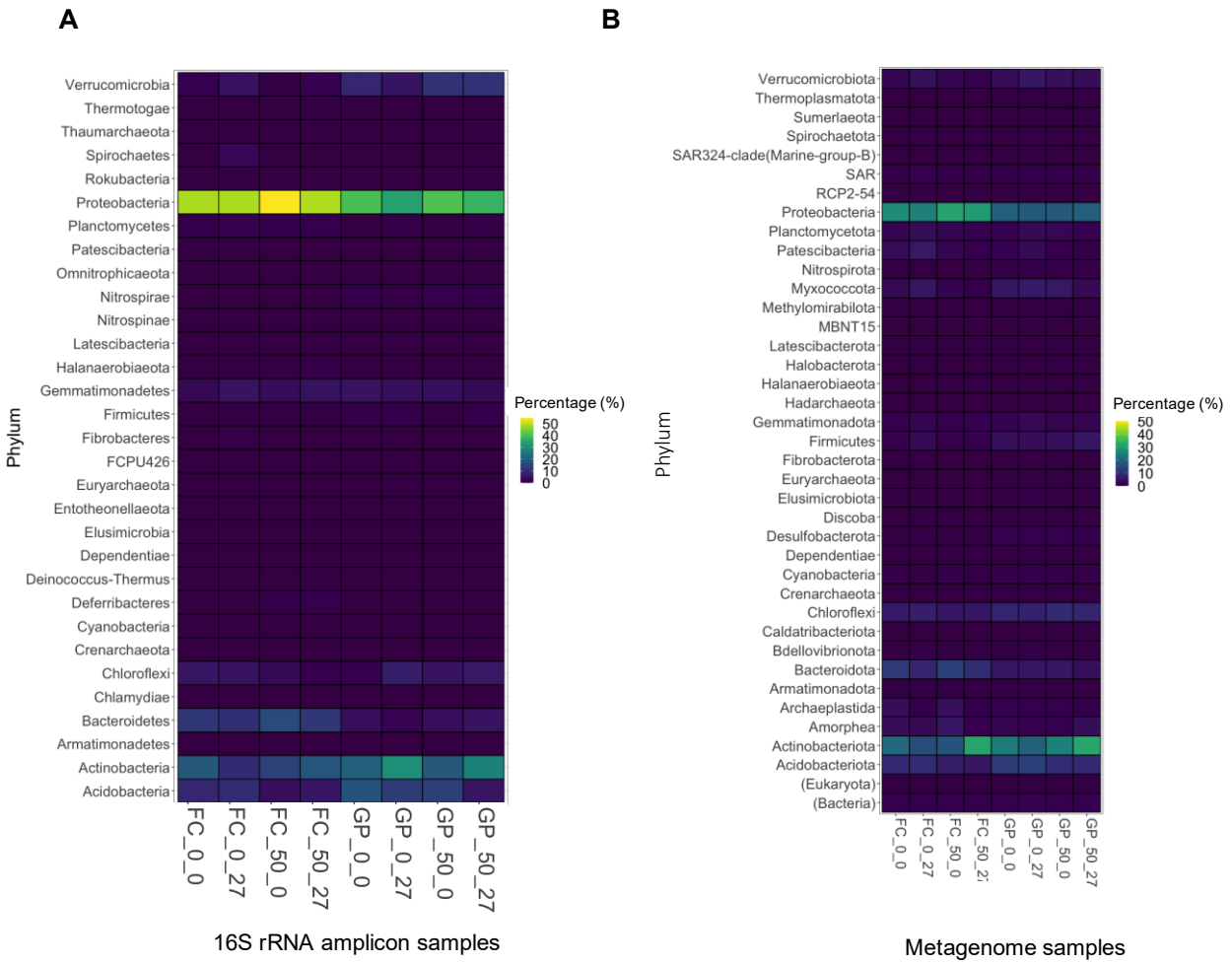

Figure S5 Overview comparison of changes in microbial taxa in FC luvisol and GP chernozem exposed to 50% FPW at day 0 and 27 using 16S rRNA amplicon-based analyses and metagenome-based analyses. Taxa were shown at the Phyla level.)

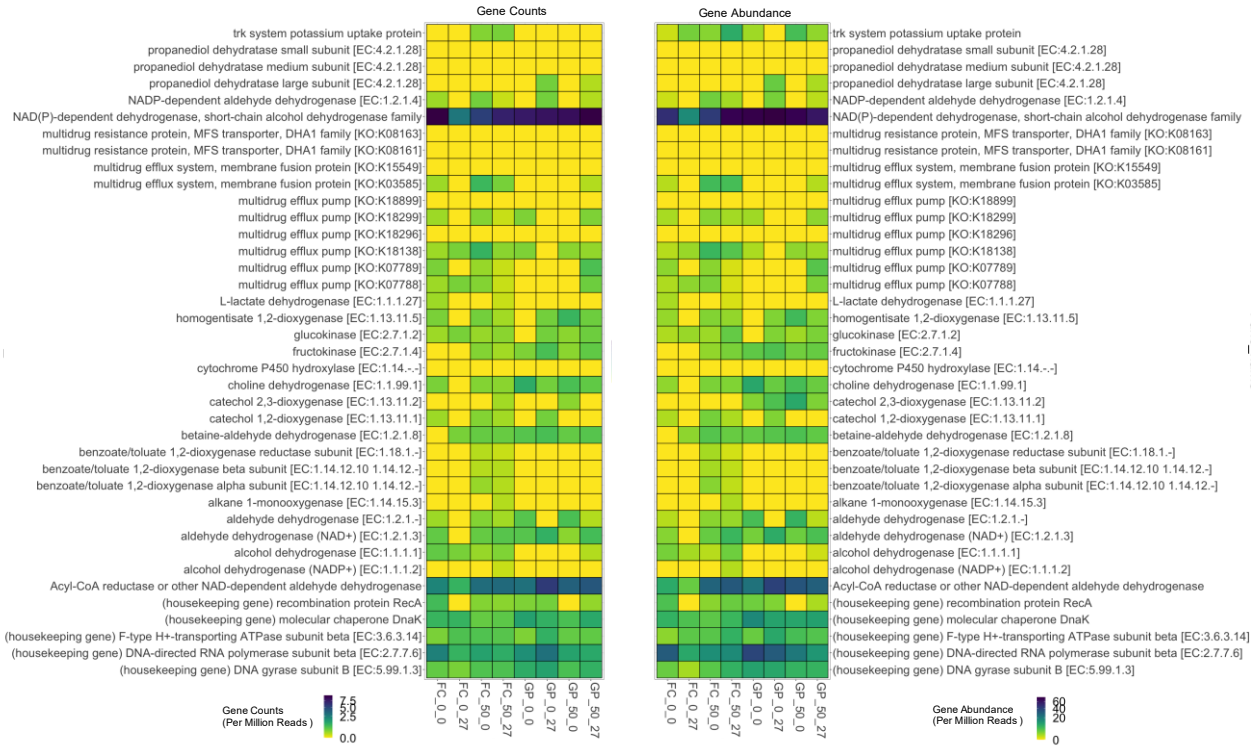

Figure S6 Effects of FPW exposure on the gene counts and gene abundance (estimated gene copy numbers inferred from coverage in IMG) of key genes that may be involved in degradation of organic compounds detected in FPW, including a wide range of organic acids, alcohols, and hydrocarbons, salt-tolerance, antibiotic resistance, and various common housekeeping genes as controls. Sample labels: location_FPW fraction (%) of the total volume fluids_incubation day (e.g., FC_50_27).

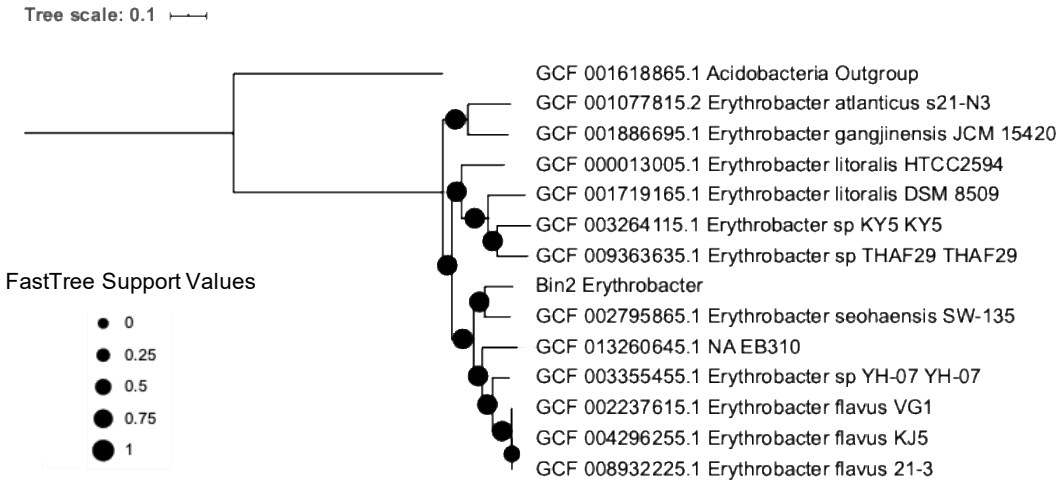

Figure S7 FastTree phylogenomic tree computed using a concatenated alignment using 117 Alphaproteobacteria.hmm genes for bin2

Table S1 Basic statistics of the metagenome used for analyses in this study. Sample labels: location_FPW fraction (%) of the total volume fluids_incubation day (e.g., FC_50_27). KEGG: Kyoto Encyclopedia of Genes and Genomes, KO: KEGG Orthology, COG: Clusters of Orthologous Groups of proteins

Genome Name	Total gene count	Number of RNA genes	Number of 16S rRNA genes	Number of Genes with Predicted Protein Product	Genes with Predicted Protein Product (%)	Number of genes in KEGG	Number of genes in KO	Number of genes in COG	Total number of bases
FC_50_0	15126	355	22	10038	66.36	3653	6403	9659	11820294
FC_50_27	28572	509	36	19230	67.3	7067	11864	18635	22257146
GP_50_0	6548	88	16	4362	66.62	1483	2398	4187	4539729
GP_50_27	15413	198	16	10563	68.53	3737	6342	10102	10842168
FC_0_0	10223	182	21	5994	58.63	1942	3235	5738	6652796
FC_0_27	5004	123	19	2016	40.29	583	960	1963	3220802
GP_0_0	5252	85	15	3443	65.56	1053	1814	3290	3635267
GP_0_27	4142	74	11	2737	66.08	933	1511	2620	2955936

Table S2 Additional physicochemical results for FPW and soil digestion samples collected from FC luvisol and GP chernozem. BDL: below detection limit, NM: not measured, where available data shown as average ± standard deviation, n=3.

Profile	FPW	Soils	
		Fox Creek	Grande Prairie
Conductivity (µS/cm)	NM	56 ± 11	44 ± 11
TDS	109 ± 4 g L^{-1}	NM	NM
Al	BDL	9830 mg kg^{-1}	21800 mg kg^{-1}
Zn	BDL	63.6 mg kg^{-1}	45.1 mg kg^{-1}
Cd	BDL	0.0826 mg kg^{-1}	0.278 mg kg^{-1}
V	BDL	39.2 mg kg^{-1}	72.2 mg kg^{-1}
Cs	NM	0.83 mg kg^{-1}	2.11 mg kg^{-1}
Se	NM	0.08 mg kg^{-1}	0.65 mg kg^{-1}
Rb	NM	7.95 mg kg^{-1}	30.5 mg kg^{-1}
Mg	843.51 ± 19.54 mg L^{-1}	5260 mg kg^{-1}	4610 mg kg^{-1}
Si	NM	2150 mg kg^{-1}	3530 mg kg^{-1}
K	1208.28 ± 8.47 mg L^{-1}	1190 mg kg^{-1}	3780 mg kg^{-1}
Mn	4.56 ± 0.10 mg L^{-1}	201 mg kg^{-1}	257 mg kg^{-1}
B	30.44 ± 2.59 mg L^{-1}	<209 mg kg^{-1}	<212 mg kg^{-1}
Ti	NM	260 mg kg^{-1}	271 mg kg^{-1}
Co	BDL	23.5 mg kg^{-1}	27.7 mg kg^{-1}
Ni	BDL	11.8 mg kg^{-1}	14.9 mg kg^{-1}
Br	NM	<27.1 mg kg^{-1}	<27.4 mg kg^{-1}
Sr	NM	15.9 mg kg^{-1}	33.9 mg kg^{-1}
Th	NM	4.09 mg kg^{-1}	3.84 mg kg^{-1}
Li	NM	<12.7 mg kg^{-1}	25.1 mg kg^{-1}
U	NM	0.496 mg kg^{-1}	0.697 mg kg^{-1}
Cr	BDL	15.7 mg kg^{-1}	24.3 mg kg^{-1}
Ce	NM	20.7 mg kg^{-1}	29.3 mg kg_{-1}
Ba	NM	128 mg kg^{-1}	265 mg kg^{-1}
Cu	BDL	8.83 mg kg^{-1}	17.7 mg kg^{-1}
Mo	46.24 ± 5.89 mg L^{-1}	0.53 mg kg^{-1}	0.877 mg kg^{-1}
PO_4^{3-}	0.03 ± 0.00 mg L^{-1}	0.18 ± 0.01 mg kg^{-1}	0.52 ± 0.02 mg kg^{-1}
NO_2^-	BDL	0.01 ± 0.01 mg kg_{-1}	0.01 ± 0.00 mg kg^{-1}
NO_3^-	0.03 ± 0.00 mg L^{-1}	0.36 ± 0.15 mg kg^{-1}	5.42 ± 0.02 mg kg^{-1}

Table S3 Envfit analyses showing correlation between top 10 ASVs to the PCoA ordination using 16S rRNA amplicon-based analyses

Top 10 ASVs	Axis 1	Axis 2	R^2	p Value	Significance
Xanthobacteraceae	0.91915	-0.39391	0.6465	0.001	***
Bradyrhizobium	0.79858	-0.60189	0.3637	0.001	***
Sphingomonas	-0.94336	-0.33178	0.88	0.001	***
Candidatus Solibacter	0.95106	-0.309	0.621	0.001	***
Acidothermus	1	0.00217	0.9076	0.001	***
Candidatus Udaeobacter	0.98676	0.16217	0.8326	0.001	***
Gemmatimonadaceae	0.41156	-0.91138	0.0882	0.109	
Ellin6067	0.44349	-0.89628	0.2525	0.002	**
Pseudomonas	-0.98087	0.19468	0.3047	0.002	**
Marinobacter	-0.14382	0.9896	0.8681	0.001	***

Table S4 Summary statistics of the quality of MAGs reconstructed from FC_50_27 (FC luvisol exposed to 50% FPW after 27 days incubation). bin 6 (IMG Taxon ID: 2886272890) was classified as *Marinobacter persicus*, bin 5 (IMG Taxon ID: 2886275984) was classified as *Salegentibacter*, bin 2 was classified as *Erythrobacter*.

Bin	Completeness (%)	Contamination (%)	N50	No. of Contigs	GC content (%)	Abundance (%)
bin 6	97.18	1.38	63,086	56	58.41	1.63
bin 5	97.18	0.08	86,172	61	37.39	1.14
bin 2	87.32	1.57	149,982	23	60.43	1.93

Table S5 Full list of gene detection related to glycine betaine and K^+ based metabolisms, the candidate genes predicted to be involved in glycine betaine and K^+ based metabolisms refers to Daly et al. (2016).(Daly *et al.* 2016) KEGG: Kyoto Encyclopedia of Genes and Genomes

Pathway	Function Names	KEGG Number	Gene	Bin 6	Bin 5
K^+ based metabolisms	2-component system OmpR family sensor histidine kinase	K07646	kdpD	1	1
		K07667	kdpE	1	1
	K+transporting ATPase ATPase F chain	K01545	kdpF	0	0
	K+transporting ATPase ATPase A chain	K01546	kdpA	1	0
	K+transporting ATPase ATPase B chain	K01547	kdpB	1	0
	K+transporting ATPase ATPase C chain	K01548	kdpC	1	0
	Na+:H+ antiporter	K03315, K07084	nhaC	0	1
	multicomponent Na+:H+ antiporter	K05565	mnhA, mrp	0	1
		K05566	mnhB	0	1
		K05567	mnhC	0	1
		K05568	mnhD	0	1
		K05569	mnhE	0	1
		K05570	mnhF	0	1
		K05571	mnhG	0	1
	Polyhydroxyalkanoate (PHA) K transporters (multi-subunit)	K05559	phaAB	1	0
		K05560	phaC	1	0
		K05561	phaD	1	0
		K05562	phaE	1	0
		K05563	phaF	1	0
		K05564	phaG	1	0

	trkA type sodium hydrogen exchanger	K03499	trkA, ktrA	2	2
	trkH type K uptake	K03498	trkH, ktrB	3	1
Glycine betaine	choline/glycine/proline betaine transport protein	K02168	betP,T,S	0	0
	choline/carnitine/betaine/glycine/BCCT transporter	K03451	betT, betS	2	1
	L-carnitine/gamma-butyrobetaine antiporter	K05245		0	0
	choline dehydrogenase	K00108, K11440	betA	1	1
	betaine-aldehyde dehydrogenase	K00130	betB	1	0
	choline-glycine betaine transporter		betL	1	1
	Proline/betaine (ProP) MFS transporter, PPII, proline permease II	K03762	proP	0	0
	proline/betaine (ProP effector)	K03607	proQ	0	0
	glycine-betaine/proline transport system ATP binding protein	K02000	proV	1	0
	glycine-betaine/proline transport system permease protein	K02001	proW	1	0
	glycine-betaine/proline transport system substrate binding protein	K02002	proX	0	0
	glycine-betaine transporter	K05020	opuD	1	1
	glutamine transport system substrate-binding protein	K10036	glnH	1	0
	glutamine transport system permease protein	K10037	glnP	1	0
	glutamine transport system ATP-binding protein	K10038	glnQ	1	0
	Osmoprotectant transport system substrate binding protein	K05845	opuC	1	0
	Osmoprotectant transport system permease	K05846	opuBD	1	0
	Osmoprotectant transport system ATP-binding protein	K05847	opuA	1	0
	glycine/sarcosine N-methyltransferase	K18896		0	0
	sarcosine/dimethylglycine N-methyltransferase	K18897		0	0

Table S6 CO_2 generation in FC luvisol and GP chernozem exposed to diluted fluids with 0%, 5%, and 50% FPW (triplicate in incubation n=3). Tukey HSD analyses; "a" indicates significant differences between T-FC-0% and T-FC-5%, "b" indicates significant differences between T-FC-0% and T-FC-50%, "c" indicates significant differences between T-FC-5% and T-FC-50%, "d" indicates significant differences between T-GP-0% and T-GP-5%, "e" indicates significant differences between T-GP-0% and T-GP-50%, "f" indicates significant differences between T-GP-5% and T-GP-50%. Sample labels: C-FC-0% (control-Fox Creek-0%) and T-FC-0% (treatment-Fox Creek-0%)

Sample ID (Treatment-Location-FPW%)	Time	Treatment	Location	CO_2 (mg L^{-1})
C-FC-0%	35	Control	Fox Creek	0.73±0.05
C-FC-5%	35	Control	Fox Creek	0.70±0.02
C-FC-50%	35	Control	Fox Creek	0.73±0.02
T-FC-0%	35	Treatment	Fox Creek	7.08±0.35 a,b
T-FC-5%	35	Treatment	Fox Creek	4.97±0.40 a,c
T-FC-50%	35	Treatment	Fox Creek	2.93±0.13 b,c
C-GP-0%	35	Control	Grande Prairie	2.25±0.08
C-GP-5%	35	Control	Grande Prairie	2.29±0.05
C-GP-50%	35	Control	Grande Prairie	2.11±0.06
T-GP-0%	35	Treatment	Grande Prairie	14.66±0.99 d,e
T-GP-5%	35	Treatment	Grande Prairie	8.45±0.55 d,f
T-GP-50%	35	Treatment	Grande Prairie	6.60±0.02 e,f

Appendix 3. Supplementary Information for Chapter 4

Methods

pH The electrode was first calibrated using a standard buffer solution pH=6.86. The electrode was then cleaned with filter paper, and used to measure the pH of the water samples. The pH values werre recorded when the values were stable on the instrument.

COD measurements (dichromate method)(Ministry of Ecology and Environment of the People's Republic of China 2017) A 10.0 mL aliquot of a water sample was mixed thoroughly with 5 mL of mercuric sulfate solution (100 g/L), potassium dichromate standard solution (0.250 mol/L), and several explosion-proof boiling glass beads. Conical bottles were connected to the lower end of the condensation pipe of the reflux device. 15 mL silver sulfate sulfuric acid reagent (10 g/L) was slowly added from the upper end of the condenser tube to prevent the escape of organic matter with a low boiling point. Solutions were thoroughly mixed by rotating conical flasks. After reflux cooling, 45 mL water was added from the upper end of the condenser tube to wash it, after which the conical flasks were removed. Three drops of ferrous indicator solution, titrated with ammonium ferrous sulfate standard titration solution (0.05 mol/L), were added after the solution cooled to room temperature. The end point was determined when the color of the solution changed from yellow to reddish brown through blue-green. The milliliter consumption (V_1) of ferrous ammonium sulfate standard titration solution was recorded. Following the same steps listed above, water samples were replaced with 10.0 mL of experimental water for blank tests, and the volume of blank ferrous ammonium standard solution (V_0) was recorded.

According to the formula, ρ (mg/L):

$$\rho = \frac{C \times (V_0 - V_1) \times 8000}{V_2} \times f$$

C—Standard solution concentration of ammonium ferrous sulfate, mol/L;

V_0—Volume of standard solution of ammonium ferrous sulfate consumed in blank test, mL;

V_1—Volume of standard solution of ferrous ammonium sulfate consumed in water sample determination, mL;

V_2—Volume of water sample taken during heating and reflux, mL;

f—Sample dilution ratio;

8000—Molar mass of $\frac{1}{4}O_2$ in mg/L.

To unify the indicator used to evaluate the organic contents in each shale formation, we converted the chemical oxygen demand (COD) into total organic carbon (TOC), based on a model of influent wastewater (COD = 49.2 + 3.00×TOC) (Dubber and Gray 2010). Since all the samples prefiltered through 0.3 um filter, all the TOC are all assumed as DOC for comparative analysis.

TDS (gravimetric method)(Ministry of Ecology and Environment of the People's Republic of China 1999) Evaporating dishes were dried at 105℃ ± 2℃ for 2 hours and the weights were measured to make sure the weights were constant (the weight difference is no more than 0.5g). Water samples were filtered through 0.45 μm and filtrates were collected in dry and clean glassware. Filtered water samples were transferred to porcelain evaporating dishes and were

evaporated. After the evaporating dish was cooled, a few drops of 1+1(v/v) hydrogen peroxide solution were added. The evaporating dishes were slowly rotated until the bubbles disappeared, and the dishes were then put on the steam bath to be evaporated. This step was repeated for several times until the residue turned white or the color became stable. The evaporated dishes were dried at 105℃±2℃ for 2 hours and the weights measured.

TDS:

$$C = \frac{W - W_0}{V} \times 10^6$$

C—Total dissolved solids content, mg/L;

W—Total weight of evaporating dish and residue, g;

W_0—Weight of evaporating dish, g;

V—Volume of Water sample, mL.

Cation analyses The major cations were measured using an Atomic Absorption Spectrophotometer. 10 mL of water samples were filtered using 0.45 μm pore organic microporous membranes and the filtered samples were injected into the instrument test tank. The absorption wavelengths of each ion (K^+:766.5nm; Mg^{2+}:285.2nm; Na^+:589nm; Ca^{2+}:422.7nm; Sr^{2+}:430.2nm) were selected using the SP-3500AA (4AT) test software. The energy was automatically adjusted in the instrument (the energy balance is 50%; if the energy is too low or too high, the lamp current is adjusted). Then, the air compressor, acetylene and water-sealed instrument were opened to start

the test (including blank and standard sample determination). Data and standard curve data were saved.

Anion analyses Preparation of 1000ppm standard solutions: F^- (0.2210g NaF added to 100mL deionized water); SO_4^{2-} (0.1480g Na_2SO_4 added to 100mL deionized water); NO_3^- (0.1371g $NaNO_3$ added to 100mL deionized water); Cl^- (0.1651g NaCl added to 100mL deionized water). All standard solutions were stored in polyethylene bottles at 4 ºC. Preparation of mixed standard solutions: take 0.5mL F- solution, 0.75mL Cl^- solution, 2.5mL $NO3^-$ solution, and 2.5mL SO_4^- solution, and add them into a 250mL volumetric flask; bring the total volume to 250 mL with ultrapure water. 5mL aliquots (the standard sample was injected directly, and the water samples were filtered by 0.45 μm pore size filter membrane) were used for Ion Chromatography analyses. Balance mode was selected to balance the baseline (Baseline noise:10Hz; Resolution ratio:0.0047nS/cm). Samples were automatedly loaded into the instruments and analyzed after the instrument signal was baselined for 0.5-1h. Samples and standard solutions were processed and results recorded using the Magic Net software implemented in Ion Chromatography.

Cleaning, trimming, and checking quality of raw shotgun metagenomes derived from Sichuan Basin PW samples Preprocessing of the raw data obtained from the Illumina HiSeq sequencing platform was done using Readfq (https://github.com/cjfields/readfq) to acquire clean data for subsequent analysis. Clean data was blasted to the host database using Bowtie 2 software v2.2.4 to filter the reads that are of host origin (Karlsson *et al.* 2012, 2013; Langmead and Salzberg 2013). Then, reads were trimmed and checked for quality using Trimmomatic v0.39 (Bolger, Lohse and Usadel 2014).

Manual refinement for MAG derived from Sichuan Basin PW samples For bins used for functional annotation, bins were visually inspected for contamination using the anvi-refine interface in Anvi'o (Eren *et al.* 2015). To remove contaminating sequences not detected in Anvi'o, bins were processed using refineM v0.1.1 (Parks *et al.* 2017), which identifies incorrectly binned contigs based on sequence composition and gene taxonomy. Bins were then imported into Geneious v7.0.6 for further rounds of assembly and refinement (Kearse *et al.* 2012). Quality of finished MAGs was verified using CheckM with lineage-specific marker gene sets.

Results and Discussion

Total dissolved solid concentrations of each formation studied To further investigate the salinity difference, alternatively, we examined the total dissolved solid contents of each formation studied in this study. The PW TDS of Sichuan Basin sample was 24,796 mg/L ± 13,113 mg/L (n=8), consisting of 23% of the measured TDS from the Marcellus samples (Barbot *et al.* 2013), 15% of the measured TDS from the Duvernay samples (n=65) (Flynn *et al.* 2019; Zhong *et al.* 2019), and 9-10% of the measured TDS from the Utica samples (Blondes *et al.* 2018).

Diversity differences between Sichuan Basin to Appalachian Basin Low microbial community diversities were detected in FPW samples from the Marcellus, Utica, and Duvernay shales; the Shannon diversity was 3.84 ± 0.28 for Sichuan Basin samples and were 1.09 ± 0.65 and 1.59 ± 1.05 for the Marcellus and Duvernay samples, respectively. The lowest Shannon diversity was 0.64 ± 0.75 for Utica samples. The Inverse Simpson diversity was generally consistent with the Shannon diversity; 30.6 ± 17.3 for Sichuan Basin samples, and 2.76 ± 1.70, 3.56 ± 4.60 and 2.14 ± 2.05 for the Marcellus, Duvernay and Utica samples, respectively.

To support the overall diversity differences for the persisting and stable microbial communities inhabited in the shale formations, we performed diversity analyses using only FPW samples collected > 50 days following the commencement of well flowback. These FPW samples have relatively stable salinity, which was far higher than for the injected fluids or initial flowback samples. The results showed that Shannon diversity for Sichuan Basin FPW samples were 3.87±0.282, and for Marcellus and Utica is 0.958±0.623 and 0.453±0.524. The diversity in Sichuan Basin samples is significantly higher (p<0.001) than for the Marcellus and Utica. This is also consistent with similar analyses of using Inverse Simpson and Observed ASV as diversity and richness indicators. Notably, diversity in samples from the Marcellus was also significantly higher (p=0.02) than in samples from the Utica using the Shannon diversity index; however, it was not significantly different using the Inverse Simpson (p=0.997) or Observed ASV (p=0.16) as indicators. Duvernay FPW samples from the Western Canadian Sedimentary Basin are excluded in this analysis since DNA in FPW samples >50 days was insufficient for sequencing; thus no FPW samples >18 days following well flowback were obtained for this study. The sampling and DNA extraction details have been documented in Zhong et al. (2019) (Zhong *et al.* 2019). However, given the rapid climb of salinity (Figure 4.1), long shut-in time (Zhong *et al.* 2019), and dominance of the typical end-point bacteria *Halanaerobium* in early FPW samples from the Duvernay Formation, the relatively early Duvernay FPW samples may be representative of the stable microbial genomes.

We attempt to quantitively link microbial biodiversity to environmental constraints in order to further explore the principles behinds the distinct diversity metrics. The higher diversity of the Sichuan Basin shale microbial communities was correlated to lower formation water salinity. The results showed that salinity effects were significant on 16S rRNA gene diversity (p<0.05) and

richness (p=0.015). This is evident in the modest correlations between Shannon (p<0.001, R^2=0.58) and Inverse Simpson diversity (p<0.001, R^2=0.50) values and chloride concentrations (Appendix 3 Figure S1). Linear regression may not explain the relationship between Observed ASV and chloride concentrations (R^2=0.37).

Potential functions of microbes detected from Sichuan Basin samples Microbial functions are predicted by comparing 16S rRNA gene amplicons to previously reported microbes as well as functional annotations of key genes for MAGs detected in this study. Within the predominant microorganisms (measured by 16S rRNA gene sequencing), strains of the abundant genus *Desulfomicrobium* are sulfate-reducing bacteria and can produce acetate and CO_2 (Langendijk *et al.* 2001; Copeland *et al.* 2009). Strains of the abundant genus *Sphaerochaeta* are able to produce acetate, ethanol, hydrogen, and CO_2 through glucose fermentation (Miyazaki *et al.* 2014). Acetate and hydrogen can be utilized by strains of the sulfur and thiosulfate bacteria *Sulfurospirillum* to grow (Stolz *et al.* 1999). Strains of the abundant *Thermovirga* can reduce sulfur to sulfide, mainly using amino acids (Dahle and Birkeland 2006). Within the Archaea, *Methanothermobacter* are known to grow on hydrogen and CO_2 and produce methane (Kaster *et al.* 2011).

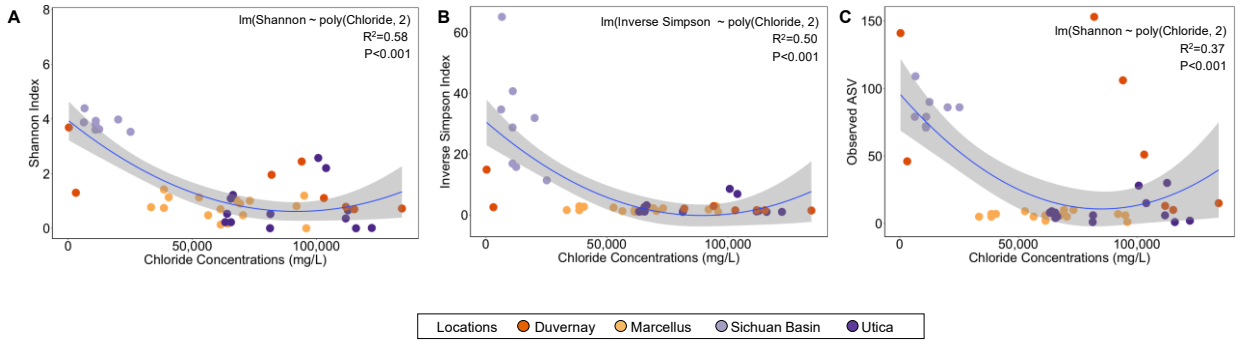

Figure S1 Microbial community diversity across the four studied formations and links between diversity, and FPW salinity and organic contents revealed by correlations between (A) Shannon diversity and chloride concentrations, (B) Inverse Simpson diversity and chloride concentrations, (C) Observed ASV and chloride concentrations. P values represent the significant effects of chloride and DOC on microbial community diversity and R^2 is the goodness-of-fit measure for regression model. Grey regions represent 95% confidence intervals.

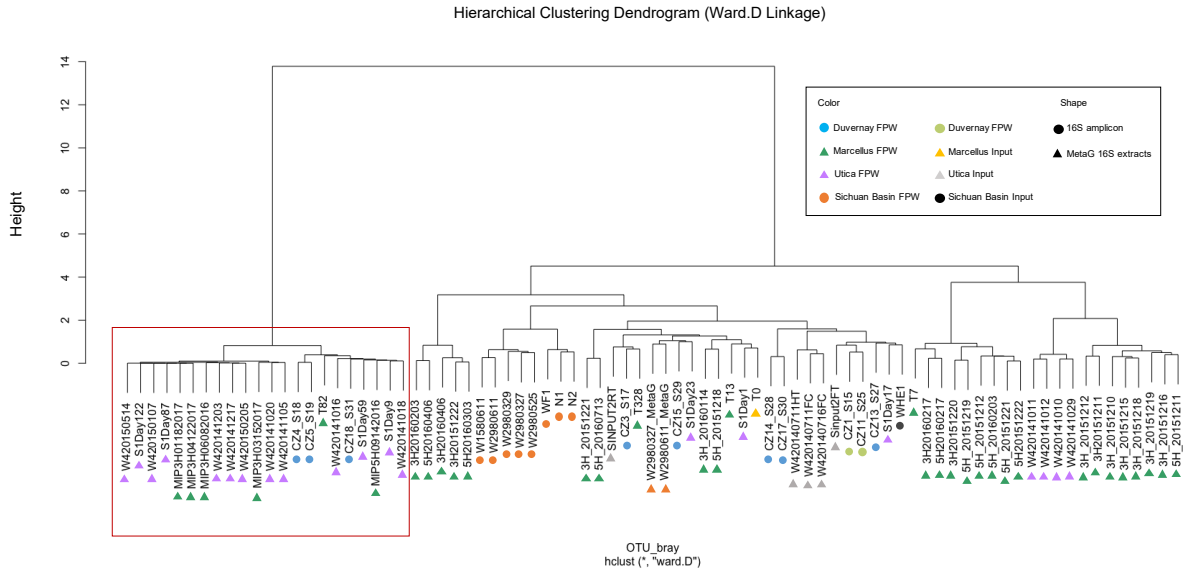

Figure S2 Ward.D linkage type hierarchical clustering based on Bray-Curtis distance of ASVs. The red box is samples enriched in bacterial genus *Halanerobium* (minimal >75% of the total community). Colors represent sample collected locations; shape represents methods for obtaining 16S rRNA gene for community survey (circles represent sequencing from 16S rRNA amplicon while triangles represent metagenome 16S rRNA gene extracts). We note that Sichuan Basin FPW samples with both 16S rRNA amplicon and metagenome 16S rRNA gene extracts were not clustered together, suggesting potential estimate differences between the two types of methods.

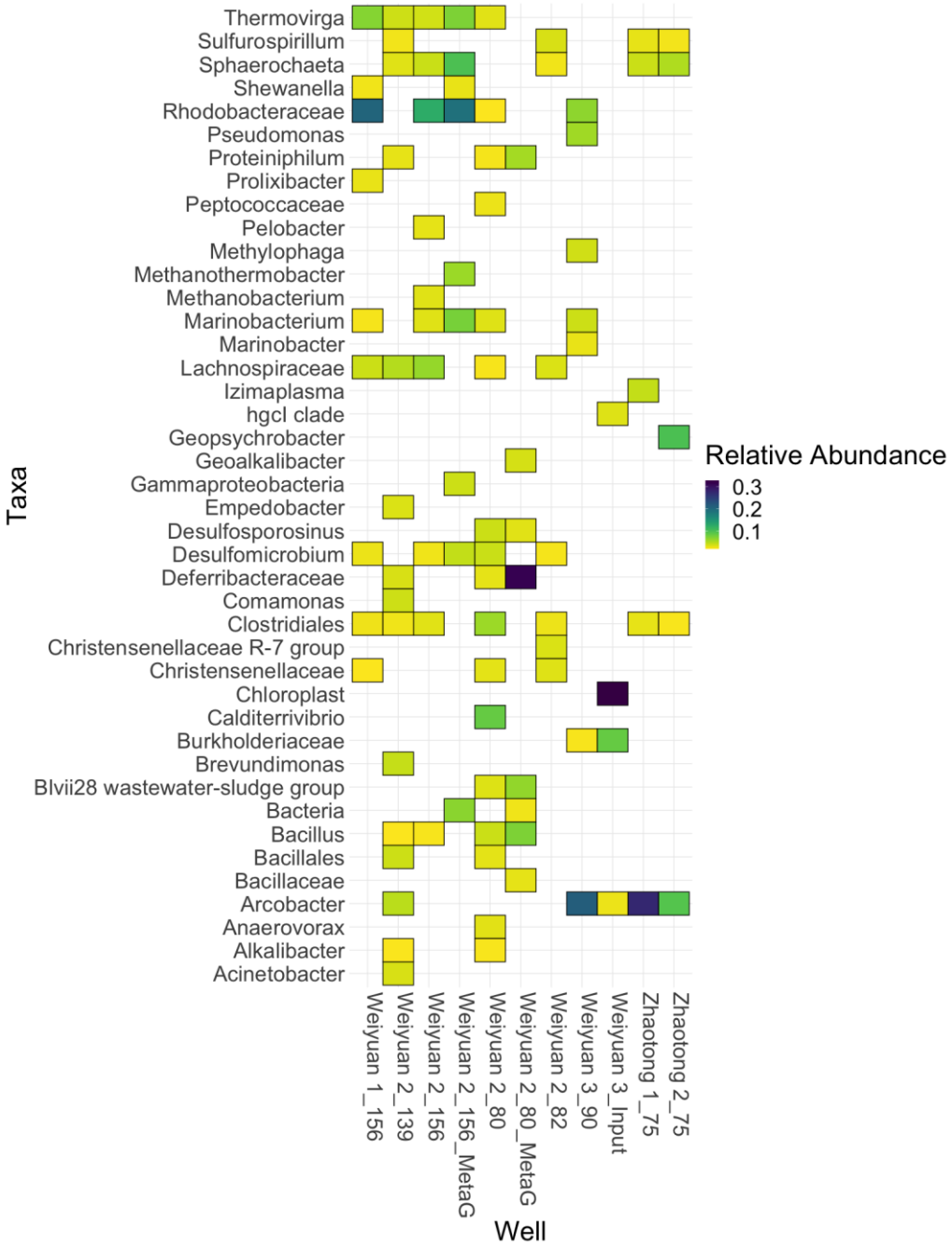

Figure S3 Taxonomic compositions of FPW samples newly collected from Sichuan Basin shales showing abundant microbial taxa (>2% of the total community). The paired samples with two methods (metagenome 16S rRNA gene extracts are represented as MetaG) are presented in this Figure to infer potential variance for results caused by different methods.

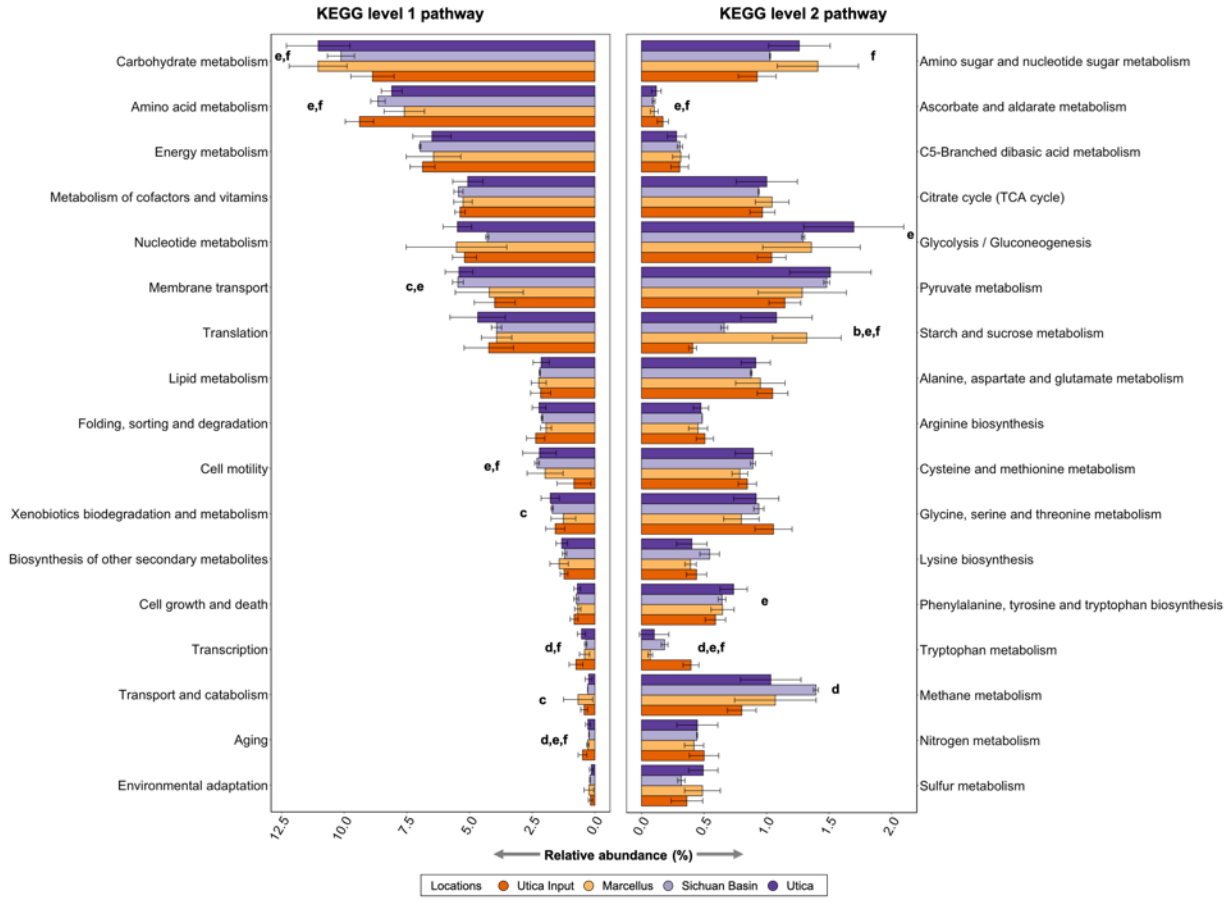

Figure S4 The relative abundance of estimated gene copies of selected (environmental-related) KEGG level 1 pathways, and relative abundance of estimated gene copies of predominant carbohydrate, amino acid, and energy sublevel-metabolic pathways (KEGG level 2 pathways) for the Marcellus, Utica, and Sichuan Basins, and (Utica) input samples. a: significance between Sichuan Basin and Utica samples, b: significance between Sichuan Basin and Marcellus samples, c: significance between Utica and Marcellus samples. d: significance between Sichuan Basin and (Utica) input samples, e: significance between Utica and (Utica) input samples, f: significance between Marcellus and (Utica) input samples. $p<0.05$ is used for these letters of significance.

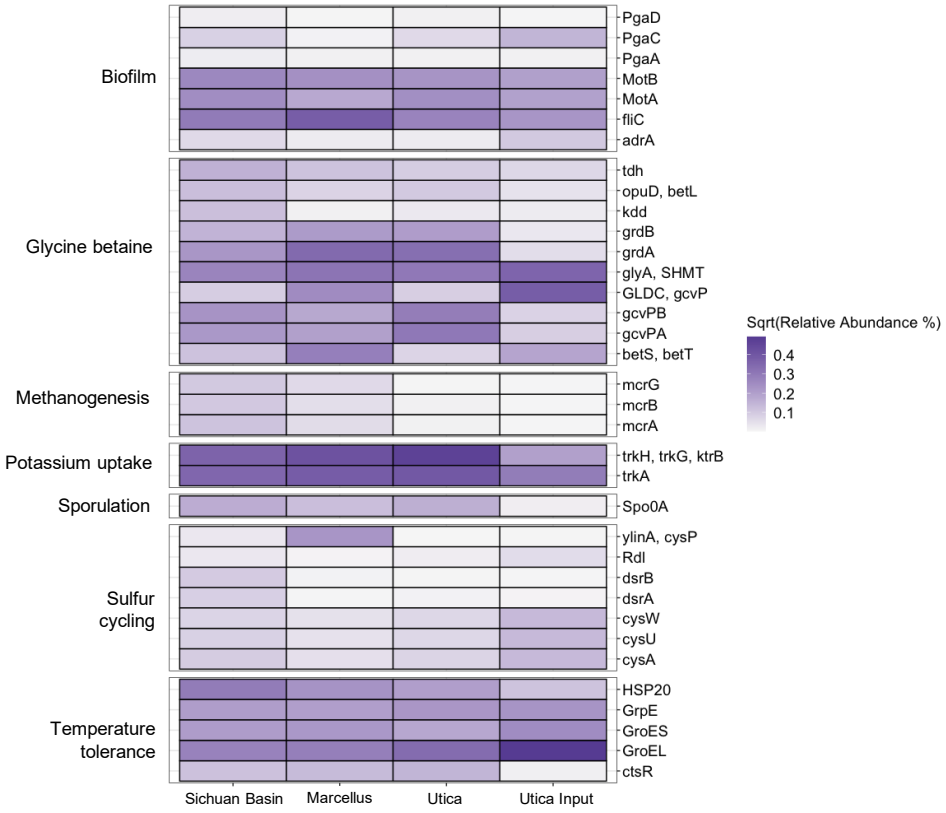

Figure S5 Heatmap for the key genes examined int this study that may be involved in biofilm, glycine betaine pathways, methanogenesis, potassium uptake, sporulation, and sulfate reduction, and temperature tolerance.

Table S1 Summary of the selected chemistry of Sichuan Basin FPW

Basin	Sichuan Basin								
Regions	Weiyuan							Zhaotong	
Samples	Weiyuan 3_Input	Weiyuan 1_156	Weiyuan 2_80	Weiyuan 2_82	Weiyuan 2_139	Weiyuan 2_156	Weiyuan 3_90	Zhaotong 1_75	Zhaotong 2_75
pH	7.59	8.33	8.14	8.22	7.8	7.87	8.08	6.93	7.75
COD (mg L^{-1})	64	1696	1490	1876	1783	1946	5109	1765	1182
TDS (mg L^{-1})	1905	21740	11420	10880	16270	21157	30675	47002	39225
Cl (mg L^{-1})	25.05	10987	6302	6609	11111	11013	12509	25147	20142
SO_4^{2-} (mg L^{-1})	NA	47.4	54	77	49.7	48.1	127.4	98.6	71.9
Br (mg L^{-1})	NA	166.5	135.8	139.2	165.2	158.6	141.6	147.8	116.2
Na (mg L^{-1})	30	6416	4332	3753	6890	6714	7174	12412	10874
K (mg L^{-1})	2	168	122	130	168	177	275	443	341
Ca (mg L^{-1})	57	265	78	119	278	273	188	769	406
Mg (mg L^{-1})	3	15	11	12	16	16	30	68	60
Sr (mg L^{-1})	1	55	9	13	56	54	43	121	72
NO_3^- (mg L^{-1})	/	75	75	75	75	76	/	/	/

Note: sample name convention is region well_day post initial flowback (e.g., Weiyuan 1_156)

Table S2 Envfit correlation of the most 20 abundant sequences of the Sichuan Basin FPW samples to PCoA ordination

Taxa	PCoA 1	PCoA 2	R^2	P value	
Burkholderiaceae	-0.54587	0.83787	0.0351	0.17	
Acinetobacter	-0.96314	0.26901	0.1997	0.025	*
Arcobacter	-0.87564	0.48297	0.0109	0.622	
Caldicoprobacter	-0.99904	0.04373	0.6213	0.001	***
Desulfotomaculum	-0.18658	-0.98244	0.7376	0.001	***
Fuchsiella	0.82649	-0.56295	0.0205	0.231	
Halanaerobium	0.95131	-0.30823	0.1556	0.003	**
Halomonas	-0.71133	0.70286	0.0712	0.072	
hgcI clade	-0.47566	0.87963	0.0172	0.308	
Lactococcus	-0.09951	0.99504	0.004	0.84	
Marinilabilia	0.7741	-0.63306	0.0565	0.1	.
Marinobacter	-0.76654	0.64219	0.121	0.046	*
Modicisalibacter	-0.62926	0.77719	0.0187	0.2	
Orenia	0.70808	-0.70613	0.0433	0.159	
Pseudoalteromonas	-0.8552	0.5183	0.1062	0.046	*
Pseudomonas	-0.71883	0.69519	0.328	0.002	**
Ralstonia	0.98762	-0.15686	0.0109	0.509	
Thermoanaerobacter	0.11843	-0.99296	0.8688	0.001	***
Thermococcus	0.30864	-0.95118	0.0498	0.116	
Thermotoga	0.31922	-0.94768	0.0565	0.116	

Table S3. Summary statistics of metagenome-assembled genomes (MAGs) reconstructed from the Sichuan Basin sample datasets after manual refinements

MAGs	Taxa	No. of Contigs	N50, bp	GC content (%)	Completeness (%)	Abundance (%)		Contamination (%)
						W2_80	W2_156	
W2980611mghmb_bin.8	Methanomethylovorans sp002508425	26	93587	37.71	81.58		0.16	0
W2980611mghmb_bin.7	Desulfomicrobium escambiense	72	72827	63.55	81.69		0.13	0
W2980327msp_mb_bin.28	Desulfomicrobium orale	18	118044	62.85	81.69		0.64	0
W2980327msp_mb_bin.30	Methanolobus sp002501695	16	175801	49.31	93.42		1.07	0
W2980327msp_mb_bin.40	Bacillus subterraneus	32	123696	42.43	87.32		1.12	0
W2980611msp_mb_bin.24	Methanobacterium formicicum	35	33579	42.83	60.53	0.26		1.35
W2980611mghmb_bin.34	Thermoanaerobacter	50	26223	34.76	80.28	2.27		0
W2980327msp_mb_bin_19	UBA4179 sp.	53	62058	34.17	91.80	25.55	0.02	0.55
W298comsp_mb_bin_27	Methanolobus vulcani	151	21560	40.64	92.81	0.01	0.52	0
W2980327msp_mb_bin.34	Thermovirga lienii	40	75813	47.04	98.31	0.74	5.35	0
W298comsp_mb_bin.12	Methanothermobacter thermautotrophicus	11	369779	49.60	100.00	0.16	5.38	0.25

W2_80, Weiyuan 2_80; W2_156, Weiyuan 2_156, Samples have abundance for both wells are co-assembled genomes. The definition of the quality standard based on the value suggested by a previous study (Bowers *et al.* 2017)