Methylo trophic Methanogenesis in Hydraulically Fractured Shales

Master’s Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of The Ohio State University

By:
Daniel Nimmer Marcus, B.S.
Graduate Program in Microbiology

The Ohio State University
2016

Thesis Committee:
Kelly Wrighton PhD, Advisor
Joseph Krzycki PhD
Charles Daniels PhD
Matthew Sullivan PhD
ABSTRACT

Over the last decade shale gas obtained from hydraulic fracturing of deep shale formations has become a sizeable component of the US energy portfolio. There is a growing body of evidence indicating that methanogenic archaea are both present and active in hydraulically fractured shales. However, little is known about the genomic architecture of shale derived methanogens. Here we leveraged natural gas extraction activities in the Appalachian region to gain access to fluid samples from two geographically and geologically distinct shale formations. Samples were collected over a time series from both shales for a period of greater than eleven months. Using assembly based metagenomics, two methanogen genomes from the genus *Methanohalophilus* were recovered and estimated to be near complete (97.1 and 100%) by 104 archaeal single copy genes. Additionally, a *Methanohalophilus* isolate was obtained which yielded a genome estimated to be 100% complete by the same metric. Based on metabolic reconstruction, it is inferred that these organisms utilize C-1 methyl substrates for methanogenesis. The ability to utilize monomethylamine, dimethylamine and methanol was experimentally confirmed with the *Methanohalophilus* isolate. In situ concentrations of C-1 methyl substrates, osmoprotectants, and Cl⁻ were measured in parallel with estimates of community membership. The appearance of the methanogenic genus *Methanohalophilus* followed different geochemical trends in the two shale environments. To adapt to the high osmolarities measured in the two shale systems, all recovered
genomes encoded three distinct strategies for osmoprotection. This includes the synthesis of the compatible organic solute glycine betaine from glycine, which has been shown to be a capability in some methanogenic archaea. The overall scheme of carbon cycling for the three analyzed *Methanohalophilus* genomes is described. Collectively, this investigation constitutes the first in depth genomic and complementary physiological analysis of methanogens obtained from the economically important shale environment.
Dedication

I would like to dedicate this document to my family & close friends for their support, and for constantly reminding me of the person I strive to be.

I would also like to recognize the positive influence 李琪 (Amelia) has had on my life; I look forward to what will come next…
I would like to acknowledge individuals who assisted with my work:

- Dr. Wrighton for serving as my advisor
- Dr. Krzycki, Dr. Daniels, and Dr. Sullivan for serving as members of my thesis committee
- Rebecca Daly for training and handling logistics in the laboratory
- David Hoyt from Pacific Northwest National Lab for help in performing and analyzing Nuclear Magnetic Resonance (NMR) measurements
- Pearlly Yan from the OSU Comprehensive Cancer Center for assistance in sequencing my *Methanohalophilus* isolate

I would also like to acknowledge my funding sources:

- National Science Foundation DEB Grant #1342701
- Department of Energy BER EMSL/JGI Grant #48483
- The Marcellus Illumina sequencing was made possible by the Deep Carbon Observatory’s Census of Deep Life supported by the Alfred P. Sloan Foundation
- The Ohio State Comprehensive Cancer Center core support grant #P30CA016058 provided sequencing of *Methanohalophilus fracturphilus*
Vita

The Ohio State University – Columbus, OH 2014–present
M.S. with Thesis - Department of Microbiology, GPA = 3.9

University of Michigan – Ann Arbor, MI 2006–2010
B.S. - Program in the Environment, School of Natural Resources, GPA = 3.5

Research Laboratory Specialist 2013–2014
Raskin Laboratory, Department of Environmental and Water Resources Engineering, University of Michigan

Research Laboratory Specialist 2013–2014
Dick Geomicrobiology Laboratory, Department of Earth and Environmental Sciences, University of Michigan

Lab Manager/Research Technician 2010–2014
Dick Geomicrobiology Laboratory, Department of Earth and Environmental Sciences, University of Michigan

Network Administrator/Computer Support Staff 2006–2008
Department of Earth and Environmental Sciences, University of Michigan

PUBLICATIONS


Fields of Study

Major Field: Microbiology
# Table of Contents

Abstract………………………………………………………………………….….ii
Dedication………………………………………………………………………..iv
Acknowledgements…………………………………………………………..v
Vita………………………………………………………………………………vi
Table of Contents……………………………………………………………..vii
List of Tables…………………………………………………………………viii
List of Figures………………………………………………………………ix

Chapter 1: Introduction……………………………………………………..1
Chapter 2: Genome Announcement……………………………………….10
Chapter 3: Main Thesis Text………………………………………………..13
  Introduction………………………………………………………………13
  Results and Discussion…………………………………………………..13
    Geochemistry…………………………………………………………….13
    16S rRNA Gene Analysis……………………………………………21
    Methanohalophilus Isolate Recovered……………………………24
    Methanohalophilus Genomes Recovered…………………………30
  Genomic Analysis……………………………………………………….31
    Substrate Catabolism………………………………………………….32
    Oxidative Branch Methanogenesis…………………………………36
    Reductive Branch Methanogenesis………………………………36
    Chemiosmotic Energy Conservation……………………………37
    Inputs Carbon Assimilation…………………………………………38
    Citric Acid Cycle……………………………………………………….39
    Gluconeogenesis……………………………………………………….42
    Pentose Sugar Synthesis…………………………………………….43
    Acetyl-CoA Synthesis……………………………………………….46
    Mechanisms of Osmo-Tolerance…………………………………….48
    Gene Synteny for Glycine Betaine Synthesis……………………49
    Global Genome Comparison………………………………………..53
  Conclusions………………………………………………………………55
  Materials and Methods…………………………………………………..56

Chapter 4: Future Directions………………………………………………64
References……………………………………………………………………67
Appendix A……………………………………………………………………75
List of Tables

Table 1.1 Energetics of Methanogenesis..................................................4

Table 3.1 Media Revision 1 Recipe.........................................................25
Table 3.2 Media Revision 2 Recipe.........................................................26
Table 3.3 Primers used for PCR.............................................................27
Table 3.4 Comparison of 16S rRNA Gene Sequences..............................27
Table 3.5 Genome Statistics..................................................................31
Table 3.6 Description of Chemical Reactions Carried out by Bifunctional Methyltransferases Involved in Glycine Betaine Synthesis.........................49
Table 3.7 Alignment Statistics of Bifunctional Methyltransferases Involved in Glycine Betaine Synthesis.................................................................49
Table 3.8 Average Nucleotide Identity (ANI) Comparison of
Methanohalophilus genomes.................................................................54
List of Figures

Figure 1.1 The Distribution of and Estimated Reserves in U.S. Shales...........2

Figure 1.2 Prior Microbiological Investigations of U.S. Shales.................5

Figure 1.3 16S rRNA Gene Tree of Methanogens and Optimum NaCl Concentrations for Growth.................................................................8

Figure 3.1 Time Series Profile of Marcellus Production Fluids................14

Figure 3.2 Time Series Profile of Utica Production Fluids......................19

Figure 3.3 Utica Sample Read Mappings to Methanolophilus Genomes......22

Figure 3.4 Plot of Halanaerobium Dominance in Utica Production Water...24

Figure 3.5 Brightfield Micrograph of Methanolophilus fracturphilus........28

Figure 3.6 Growth Curve of Methanolophilus fracturphilus ..................28

Figure 3.7 16S rRNA Gene Tree of Methanolophilus with Substrate Range..29

Figure 3.8 Amino Acid Alignments of Methylamine Methyltransferases.....32

Figure 3.9 Genome Cartoon (Energy Metabolism) of Methanolophilus Genome.................................................................34

Figure 3.10 Central Carbon Metabolism from Methanolophilus Genomes...40

Figure 3.11 Citric Acid Cycle from Methanolophilus Genomes..............41

Figure 3.12 Gluconeogenesis from Methanolophilus Genomes.............43

Figure 3.13 Partial Pentose Phosphate Pathway from Methanolophilus Genomes.................................................................45

Figure 3.14 Proposed Alternate Entrance into the Shikimic Acid Pathway....45
Figure 3.15 Acetyl-CoA Biosynthetic Pathway from *Methanothalophilus* Genomes…………………………………………………………………47

Figure 3.16 Phylogenetic Tree of Glycine Sarcosine Methyltransferases……51

Figure 3.17 Proposed Supportive Metabolic Framework for Glycine Betaine Synthesis………………………………………………………53
Chapter 1 - Introduction

Methylo trophic Methanogenesis in Hydraulically Fractured Shales

An introduction to the environment, metabolism and the genus *Methanohalophilus*

Natural gas plays an enormous role in supporting the contemporary lifestyle of US citizens both as individuals and collectively as a society. Most notably, it is used to generate approximately 33% of the US’ annual 4 trillion kilowatt hours of electricity, it fuels commercial transportation, and is used for domestic heating (EIA 2014). The importance of natural gas in modern industrialized life cannot be overstated. Over the last decade, advances in hydraulic fracturing technology have enabled the U.S. energy industry to recover enormous quantities of oil and natural gas from previously uneconomic unconventional shale deposits (EIA 2014). The United States is estimated to contain the second highest global quantities of technically recoverable shale oil (50 billion barrels) and the fourth highest quantities of technically recoverable shale gas (665 trillion cubic feet) (EIA 2013). Shale formations containing economic quantities of hydrocarbons (hereafter referred to as shale plays) from different geologic origins (Osborn et al. 2011) occupy geographically large portions of the continental U.S. (Figure 1.1A) and constitute a sizeable fraction of domestic energy resources (Figure 1.1B) (EIA 2013).
Recent molecular evidence has demonstrated that methanogens are present in these systems (Waldron et al. 2007, Wuchter et al. 2013, Cluff et al. 2014, Akob et al. 2015). This raises the question of what fraction of recoverable shale gas may be the product of microbial metabolism as a response to the anthropogenic fracturing process.

Hydraulically fractured shales are in many ways extreme environments. These anthropogenically modified subsurface environments are thought to have down well temperatures in excess of 50° C, pressures greater than 25 MPa (~3.6k PSI), and brine level salinities (Cluff et al. 2014, Daly et al. submitted). Recorded down hole temperatures are believed to be below the threshold required for subsurface sterilization (Rothschild and Mancinelli 2001, Takai et al. 2008), and laboratory investigations have revealed surprisingly high pressure tolerances for common model organisms (Sharma et al. 2002). Salinity on the other hand, is thought to play a large role in structuring terrestrial subsurface communities (Waldron et al. 2007, Cluff et al. 2014). Canonically,
cells in saline environments follow one of two strategies to persist within osmotically challenging systems. Organisms using the “salt-in” strategy accumulate high levels of inorganic solutes in order to reach osmotic equilibrium with the surrounding environment. While this strategy is energetically inexpensive for an individual cell, the enzymatic cellular machinery must be adapted to high salt concentrations. Therefore, the energetic expense for the salt-in strategy is likely invested over evolutionary time as opposed to being expended over the course of an individual cell’s lifespan. A hallmark of salt adapted proteins is a high ratio of acidic to hydrophobic amino acids. Alternatively, the de-novo synthesis or accumulation of small organic osmolytes (aka compatible organic solutes) from the environment is the other main strategy of halotolerant adaptation. While this approach does not require large scale enzymatic adaptation to high inorganic solute concentrations, the synthesis and transport of compatible organic solutes does come with an appreciable energetic cost (Oren 1999, Head et al. 2014). Halophilic methylotrophic methanogens are known to utilize a combination of the two approaches by accumulating both compatible organic solutes and potassium ions (Lai et al. 1991).

The conspicuous absence of hydrogenotrophic (using hydrogen and CO₂) and aceticlastic (using acetate) methanogens in high salt environments has been documented with few exceptions. It has been suggested that only methylotrophic methanogens (capable of both oxidizing and reducing methylated substrates) are able to acquire enough energy through substrate catabolism to offset the energetic expense associated with de-novo synthesis and or transport of compatible organic solutes in highly saline systems (Oren 1999). Table 1.1 illustrates that methylotrophic methanogens produce more
methane per mole of (oxidizable) substrate consumed, and acquire a higher net yield of energy from a comparable molar amount of substrate consumed. It should be noted that homoacetogens have been shown to grow in up to 250g/L NaCl, which indicates that growth on H₂ and CO₂ can support life in saline systems. One posited explanation for why homoacetogens are capable of growth on H₂ and CO₂ at high osmolarities is that characterized strains accumulate inorganic ions as their primary means of osmotolerance (Oren 1999). One additional caveat for the values provided in Table 1.1 is that the partial pressures of H₂, CO₂ and CH₄ would affect the energetics of hydrogenotrophic growth. To the knowledge of the author, in situ hydrogen concentrations have not been measured in shale, and for this reason calculations were made at standard conditions.

Table 1.1 Energetics of methanogenesis from substrates at standard conditions.
* Values calculated from ΔG°' sourced from (Thauer et al. 1977).
** For hydrogenotrophic growth only moles of oxidizable H₂ are considered in these calculations.

<table>
<thead>
<tr>
<th>Substrate</th>
<th><strong>H₂ : CO₂</strong></th>
<th>Acetate</th>
<th>CH₃OH</th>
<th>MMA</th>
<th>DMA</th>
<th>TMA</th>
</tr>
</thead>
<tbody>
<tr>
<td># moles substrate consumed</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td># moles CH₄ formed</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Molar ratio substrate oxidized : CH₄ formed</td>
<td>4:1</td>
<td>1:1</td>
<td>1:3</td>
<td>1:3</td>
<td>1:3</td>
<td>1:3</td>
</tr>
<tr>
<td>ΔG°' (kJ)/mole CH₄ formed</td>
<td>-131*</td>
<td>-36*</td>
<td>-107*</td>
<td>-76*</td>
<td>-75*</td>
<td>-76*</td>
</tr>
<tr>
<td>ΔG°' (kJ) from consumption of 4 moles substrate</td>
<td>-131*</td>
<td>-36*</td>
<td>-319*</td>
<td>-230*</td>
<td>-449*</td>
<td>-683*</td>
</tr>
</tbody>
</table>
There have been a limited number of investigations into the microbiological composition of hydraulically fractured shales. 16S rRNA gene sequencing has shown two obligate methylotrophic methanogenic genera (\textit{Methanohalophilus} and \textit{Methanolobus}) within the order \textit{Methanosarcinales} to be present in both raw and amended fluid samples originating from numerous geographically and geologically distinct shale formations (Waldron \textit{et al.} 2007, Fichter \textit{et al.} 2012, Wuchter \textit{et al.} 2013, Cluff \textit{et al.} 2014, Akob \textit{et al.} 2015) (Figure 1.2). The \textit{Methanosarcinales} have been shown to be the only metabolically versatile group of cultured methanogens capable of utilizing methylated substrates without requiring an additional electron donor (Thauer \textit{et al.} 2008, Borrel \textit{et al.} 2013).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.2}
\caption{Previous microbiological assessments of hydraulically fractured shales. \textbf{A} Geographic map illustrating the locations of existing studies of shale microbial communities. \textbf{B} Detection of select microbial community members across shale sites. The Green rectangle points out the occurrence of the methylotrophic methanogenic genus \textit{Methanohalophilus}. Modified from Mikayla Borton}
\end{figure}
Methanohalophilus has been detected in shale environments more than twice as many times as any other methanogenic genera. Members of the Methanohalophilus have been isolated from various saline environments including saline lakes (Paterek and Smith 1988), salterns (Mathrani and Boone 1985), natural gas fields (Katayama et al. 2014) and cyanobacterial mats (Boone et al. 1993). One striking feature of Methanohalophilus is that it appears to be the only cultured methanogenic genus that consistently (with one notable exception) grows optimally at high concentrations of salt (studies specified either NaCl or Na\(^+\)) (Figure 1.3). Since most cultured Methanohalophilus isolates grow optimally with a salt concentration in the range of 0.5 – 2.5M, this group is classified as being moderately halophilic (Oren, 2008). It should be noted that the 16S rRNA gene tree in Figure 1.3 was not an exhaustive sampling of cultured methanogens; however, the tree does illustrate that most analyzed hydrogenotrophic methanogens grow optimally at a low NaCl concentration.

The absence of detectable Methanohalophilus in input fluids and early periods of sampling in shale systems where it is later found suggests that it is actively reproducing in these environments. This begs the question of where its substrates originate. Operators of wells undergoing energy extraction have been known to put a cocktail of chemical additives in the input fracturing liquid. Methanol is a common additive included as a corrosion/scale inhibitor at final concentrations under 0.001% (Saba et al. 2012). Di- (DMA) and monomethylamine (MMA) have been found by our research group to be put down wells directly (Daly et al. submitted). These three compounds have been shown to be utilized directly for growth by nearly all members of the Methanohalophilus (Paterek and Smith 1985, Boone et al. 1993, Katayama et al. 2014)
Quaternary amines such as choline are frequently added to fracking chemicals at final concentrations of 0.05 – 0.2% for various abiotic operational purposes (McCurdy 2011). While choline can be directly used as a substrate for some methanogens (Watkins et al. 2012), this has never been shown for *Methanohalophilus*. However, a recent metagenomic investigation into the metabolic interactions in the Marcellus shale microbial community conducted by the Wrighton research group posited that the consortia present in the system could potentially convert choline to glycine betaine. The genomically inferred microbial conversion of choline to glycine betaine and subsequent fermentation to TMA supporting methanogenic growth was verified experimentally with a Utica enrichment sample (Daly et al. submitted).

Collectively, these studies all point to the possibility that biogenic methane is produced in situ by *Methanohalophilus* in fractured shales. This investigation aims to add to the scarce body of knowledge concerning *Methanohalophilus* and its recurring presence in shale. Geochemistry and genomics have been used to further characterize this unique halophilic methylotrophic methanogenic group.
Figure 1.3 16S rRNA phylogenetic tree of cultured Methanogens with substrate utilization and optimum NaCl concentration for growth. \( \text{H}_2\text{CO}_2 \) = hydrogenotrophic growth, Formate = growth on formate, and Alcohols* indicates that alcohols are utilized by these strains as an electron donor for hydrogenotrophic growth. \( \text{C}1 \) = C-1 methylated substrates (eg. \( \text{CH}_3\text{OH}, \text{TMA}, \text{DMA}, \) and \( \text{MMA} \)) and Acetate refers to growth on acetate. The two MX organisms listed both require methylated compounds and hydrogen together for methanogenesis; salt physiology has not been determined for either MX organism listed which is why this data is not presented. For brevity, references for this data have not been included, but all values were obtained from peer reviewed papers characterizing the methanogens above.
Chapter 2 – Genome Announcement

Draft genome sequence of *Methanohalophilus fracturphilus*, an obligate methylotrophic methanogen obtained from fractured shale.

Daniel N. Marcus¹ and Kelly C. Wrighton¹

¹Department of Microbiology, The Ohio State University, Columbus, OH, 43214, USA

Abstract:
*Methanohalophilus fracturphilus* is a methylotrophic methanogenic archaeon and a representative of the family *Methanosarcinales*. Here, we report a 1.9Mb draft genome with a G+C content of 42.2%.

Main Text:

*Methanohalophilus fracturphilus* is an obligate methylotrophic methanogen, consistent with all other *Methanohalophilus* isolates characterized to date (Paterek and Smith 1985, Boone et al. 1993, Katayama et al. 2014). From the recovered 16S rRNA gene sequence, *Methanohalophilus fracturphilus* is 99% identical to both existing genome sequenced *Methanohalophilus* (*mahii* and *halophilus* sp. Z-7982). By average nucleotide identity (ANI) (Konstantinidis and Tiedje 2005), the genome of *Methanohalophilus fracturphilus* is most closely related to *Methanohalophilus halophilus* sp. Z-7982 with a two ANI of 92%. Although *Methanohalophilus fracturphilus*’ 16S rRNA gene similarity is within the commonly accepted 97% threshold for species delineation, the assessment of total genomic similarity indicates that *Methanohalophilus*
fracturphilus is sufficiently distinct (less than 95% ANI) to the most closely related Methanohalophilus genome and thus warrants being classified as its own species (Goris et al., 2007). To date, the Methanohalophilus genus is comprised of 5 named species. Although these strains were taxonomically identified, the phylogeny and genomic features shared within the genus are largely unclear. Here we report a draft genome sequence of Methanohalophilus fracturphilis, a species isolated from production water obtained from a hydraulic fracturing shale well during the course of energy extraction in Ohio.

Whole-genome shotgun sequencing was performed using the Illumina HiSeq platform (Illumina, California) at the Ohio State Comprehensive Cancer Center’s genomics facility. We prepared and sequenced a 100 bp paired-end KAPA Hyper (Kapa, Massachusetts) library with a mean insert size of 225 bp. After quality trimming, the sequencing yielded ~15.6 kb read pairs. This data was assembled using IDBA-UD (Peng et al. 2012) and produced an assembly with 2808x average coverage of the draft genome.

The draft genome is 1.9 Mb, has a G+C content of 42.2% and is spread over 113 scaffolds. The longest scaffold is 367.6 kb and the assembly has an N50 of 109.4 kb. MetaProdigal (Hyatt et al. 2012) predicted 1986 protein-coding genes. These open reading frames were compared to KEGG and UniRef90 using USEARCH (Edgar 2010). 1440 of the predicted protein sequences contained annotations with predicted function and 546 were identified only as hypotheticals. Using t-RNA Scan-SE (Schattner et al. 2005) we identified 47 tRNAs. The genome was inferred to be complete as we recovered only one copy of the 104 single copy housekeeping genes previously described in Archaea (Wu et al., 2012).
The genome encoded a full methanogenesis pathway including all genes canonically involved in the oxidative and reductive branch of methanogenesis. Methyltransferases and cognate corrinoid proteins for the utilization of methanol, MMA and DMA were confirmed by genomics and physiological verification. Notably, the capacity to utilize trimethylamine was not encoded in the genome and laboratory growth experiments using the substrate corroborated this finding. The genome also contains pyrrolysine biosynthetic genes (PylB,C,D) and tRNA synthetase which are required for the catalytic functioning of the organism’s catabolic methylamine methyltransferases.

Genomic information from *Methanohalophilus fracturphilus* will complement other genomes in this genus and may provide insight into the genomic features enabling adaptation to the deep subsurface. This draft-genome sequence has been deposited in DDBJ/GenBank/EMBL under accession number [PENDING].
Chapter 3 – Main Thesis Text

Introduction
In this investigation, produced water (also known as the fluid that that is recovered from a hydraulically fractured shale during the process of energy extraction) was sampled over time from geographically and geologically distinct shales. The two shales discussed in this text are the Marcellus and Utica; both geologic deposits span multiple states in the Appalachian region, and cover substantial portions of Pennsylvania, West Virginia, and Ohio. Samples were analyzed geochemically and sequenced using shotgun metagenomics for reconstruction of microbial community structure and function. Efforts to isolate a methanogen from shale were successful and yielded a *Methanohalophilus* culture and accompanying genome. Metabolic reconstruction was attempted on the three Methanohalophilus methanogen genomes recovered in this study and the results are discussed in the following sections.

Results and Discussion

Geochemistry of Shale Produced Fluids can Support Methylo trophic Methanogens
In production fluids collected from the Marcellus shale a dramatic increases in Cl⁻ concentrations over that measured in the input fluid was observed in the first four produced water samples (Figure 3.1).
Figure 3.1 Time series produced water samples were profiled from the Marcellus shale. Day 0 values are input fluids; they were included to illustrate the initial chemical and methanogen composition of fluids injected into the subsurface. The well was shut in for less than 24 hours before produced water was allowed to return to the surface. Chloride and community abundance values were reproduced from (Cluff et al., 2014). Methanohalophilus (pink) and Methanolobus (brown) 16S rRNA gene relative abundances are shown in the bar chart.

After the fourth sample was collected on day 82, the next sample was not collected until day 328 when the Cl⁻ concentration had stabilized near values obtained on day 82. This increase in salinity over time is consistent with prior reports of injected fluid from the surface equilibrating with subsurface shale formations waters and becoming increasingly saline (McIntosh et al. 2002, Rowan et al. 2015). The source of
inorganic osmolytes in the subsurface is believed to be paleoseawater evaporite deposits embedded in the shale rock (Rowan et al. 2015).

Corresponding with the increase in salinity, glycine-betaine (GB) was not detected in the input fluid, but did accumulate in the environment reaching a maximum concentration at day 82 of 5.7uM (measured with $^1$H NMR). GB is an important zwitterionic quaternary amine that can function as an osmoprotectant for many phylogenetically distinct organisms (including some halophilic methanogens) (Lai and Lai 2011). The lack of GB detected in the input fluid is consistent with the hypothesis that this compound is produced in situ by active community members in order to offset osmotic stress.

Based on genomic reconstruction of methanogens in the system (results below) methanogenic methyl C-1 substrates were profiled using $^1$H NMR. Di- (DMA) and monomethylamine (MMA) in addition to methanol were present in the input fluids and were depleted over time. As shown in Figure 3.1, MMA was measured at 1.3mM in the input fluid, approximately a four fold higher concentration than methanol (0.33mM). DMA was not shown as when it was most concentrated in the input sample, it was only measured at 4.1uM. According to recent analysis of the genomic makeup the Marcellus microbial community, it is currently believed that some of the aforementioned methylotrophic substrates could possibly be utilized by non-methanogenic taxa. It is thought that some of these compounds could be assimilated as a nitrogen source (MMA) or oxidized by early community members (Daly et al. submitted). However, in the Marcellus shale the relationship between methyl C-1 substrate depletion and methanogen
abundance is consistent with the idea that methanogens could be converting an appreciable fraction of methyl C-1 compounds into biogenic methane.

Produced water samples collected from the Utica shale had considerably higher concentrations of Cl\(^-\) relative to the freshwater source used as a base for input fluids (up to 13,245x) (Figure 3.2). Unfortunately, the blended input fluid, which was injected into the subsurface was not suitable for ion chromatography analysis and thus Cl\(^-\) was not quantified. However, the input fluid source was a freshwater lake (Goldman et al. 1983) that contained low levels of Cl\(^-\) (8.2mg/L). The trajectory of Cl\(^-\) in Utica produced water samples through time is consistent with what was seen previously in the Marcellus shale (Figure 3.2).

Of the four methylotrophic substrates detected in the Utica samples, only methanol was seen in the input fluid; however, the concentrations of methanol increased dramatically (~41x) in the first produced fluid sample recovered. While trimethylamine (TMA), (DMA) and (MMA) were not present in the input fluids, they were measured in subsequent produced water samples. MMA concentrations reached a maximum of 2.8uM 105 days after hydraulic fracturing, and were found to be depleted in the terminal produced water sample (day 302) at 1uM; the final produced water sample had the lowest concentration of MMA in any of the produced water samples. DMA concentrations were highest at day 96 at 1.3uM. However, in samples closely preceding and following day 96 DMA was observed at 0.4uM, its lowest concentration in all produced water samples. TMA was found in its highest concentration in the first produced water sample (day 86) at 1.7uM and became increasingly depleted in the system over time, with its lowest concentration measured in the terminal produced water sample at 0.4uM (Figure 3.2). As
is discussed in the methods section, glycine betaine could not be confidently quantified from Utica production fluids as a result of the fluid composition. However, despite the inability to determine glycine betaine concentrations, the NMR data does indicate its presence in produced water samples.

One difference worth mentioning between the two shale sites is the length of time injected fluids were trapped in the subsurface. For the Marcellus shale, input fluids were sealed in the rock matrix briefly (less than one day) at a depth of approximately 2.5km before being allowed to return to the surface as production fluids. This is in contrast to the 86 days fluids were kept in contact with the Utica shale prior to being allowed to return to the surface. Prior to the time when production fluids were liberated, no sampling was possible; thus, for the Utica shale there was a more extended initial period devoid of measurements.

Unlike the clear trends observed from the Marcellus shale that appear largely consistent with methylotrophic methanogenic growth arising from input fluid material, the measured increases in Utica methyl C-1 substrate concentrations through time are difficult to explain. For all four of these methylotrophic substrates, the increase in concentrations over either their absence (TMA, DMA, MMA) or low concentration in the input fluid (methanol) suggests that they are potentially being leached from the system or are generated in situ by other community members. A recent laboratory investigation by colleagues has confirmed the hypothesis that input fluids amended with glycine betaine give rise to methylated methanogenic substrates that are subsequently consumed, resulting in the production of biogenic methane (Daly et al. submitted). In the future, radiolabelled substrates could be used to track the in situ fate of carbon and nitrogen in
order to more clearly understand the ecological network that may support methanogenic growth.
Figure 3.2 Time series produced water samples were profiled from the Utica shale. Day 0 Cl− (determined using ion chromatography) and microbial community values are from the freshwater input fluid source, all other day 0 measurements are from the mixed input fluid (this includes the freshwater base fluid and fracturing additives). All other time points are produced water samples. The well was shut in for 86 days before produced water was collected at the wellhead. Trimethylamine (TMA), dimethylamine (DMA), monomethylamine (MMA), and methanol (CH$_3$OH) concentrations were determined with $^1$H NMR. *Methanohalophilus* relative abundances were calculated from read mappings to EMIRGE reconstructed 16S rRNA gene sequences and are shown in the bottom panel. Relative abundance values are provided to the right of the bottom panel. Black rectangles enclose samples with detected *Methanohalophilus*. 
Utica Well Trajectory

- **Days Following the Onset of Produced Water**
- **Community Abundance**
- **Methanohalophilus %**
- **g/L Cl−**
- **μM CH₃OH**
- **μM TMA**
- **μM DMA**
- **μM MMA**

- **Day 94:** 0.61%
- **Day 96:** 0.14%
- **Day 98:** 0.08%
16S rRNA Gene Analyses Reveals Methylotrophic Methanogens are Present in Shale Derived Fluids

From 454 pyrosequencing of raw Marcellus fluid samples (described previously in Cluff et al., 2014), two methanogens were detected, *Methanohalophilus* and *Methanolobus*; both genera have been shown to be obligate methylotrophs (Konig and Stetter 1982, Boone et al. 1993) belonging to the metabolically versatile order *Methanosarcinales* (Thauer et al. 2008). At day 82, *Methanolobus* was the dominant methanogen at 1.9% total community abundance and *Methanohalophilus* was near the limit of detection at 0.19% of sequences obtained. This trend inverted for the terminal sample collected at day 328, where *Methanohalophilus* was present at 1.7% of the community and *Methanolobus* was just above the limit of detection at 0.05% (Figure 3.1).

The trends observed from samples collected from the Marcellus shale (Figure 3.1) are consistent with what would be expected if active methanogenic growth were occurring in the subsurface. This includes proliferation of *Methanohalophilus* only above a salinity threshold greater than that present in the input water (the lowest optimum [NaCl] reported for a *Methanohalophilus* isolate is approximately 20g/L (Katayama et al. 2014)), a depletion of methyl C-1 substrates and an increase in GB over time. However, it should be noted that that both *Marinobacter* and a novel *Halomonadaceae* were recently genomically investigated from these very same samples by colleagues and both of these genomes were found to contain the necessary genes to de-novo synthesize GB (Daly et al. submitted). Given that other community members potentially capable of producing GB are present in much higher community abundances, the trend of increasing GB over time cannot be directly linked to the methanogens present in the system.
One notable distinguishing feature between the two shale plays is that methanogens were not detected in the final Utica samples collected. Metagenomic sequence reads from the Utica shale were trimmed for quality and assembled into near full-length 16S rRNA genes for the resident microbial community using EMIRGE (Miller et al., 2011). From reconstructed 16S rRNA gene sequences, *Methanohalophilus* was seen for a period of five days (concluding on day 98). Following this span of time, *Methanohalophilus* was not detected in any of the subsequently collected samples, which were taken over a period of greater than 6 months after day 98. Read mappings independently confirmed the absence of *Methanohalophilus* in all other Utica samples collected (Figure 3.3). Days 154 and 175 contained the greatest number of reads mapping to *Methanohalophilus* genomes amongst the post day 98 samples; ESOM binning of scaffolds greater than 5kb from these datasets corroborated the absence of *Methanohalophilus*.

![Figure 3.3](image)

**Figure 3.3** The genomes from *Methanohalophilus mahii*, *Methanohalophilus halophilus* and the Marcellus day 328 *Methanohalophilus* genome bin were concatenated together and collectively used as a reference for read mapping for each one of the recovered Utica metagenome samples. The total number of reads satisfying default bwa read mapping criteria are displayed above.
The disappearance of methanogens in the Utica produced water samples prior to the end of sampling was a puzzling finding; particularly in light of C-1 substrate concentrations remaining close to the values present in samples with detected methanogens. One possibility for the lack of *Methanohalophilus* detected after day 98 is insufficient sequencing depth (Rodriguez and Konstantinidis 2014). The dramatic increase in the dominance of the taxa *Halanaerobium* over time is apparent when it is assessed against other community members (Fig 3.4). In the first sample containing *Methanohalophilus* (day 94), *Halanaerobium* accounted for only 89% of the total recovered microbial community. This dominance increased in day 96 (96%) and again in day 98 (97%). Following day 98 (the last sample with *Methanohalophilus*) *Halanaerobium* increased in community abundance (excluding one outlier sample) and for approximately the final 6 months, *Halanaerobium* made up greater than 98% of all recovered 16S rRNA sequences. *Halanaerobium*’s increasingly high level of dominance could potentially have saturated what sequencing was obtained from later Utica produced water samples preventing the detection of *Methanohalophilus*. An additional possible explanation for the disappearance of Methanohalophilus could be insufficient quantities of trace metals. Methanogens are known to utilize a variety of metals as cofactors, such as Cobalt in corrinoid proteins, Nickel in the methyl coenzyme M reductase complex and Zinc at least one of the two heterodisulfide complexes (Glass and Orphan, 2012). Data for all relevant metals was not available and it cannot be ruled out that trace nutrient availability could affect the ability of *Methanohalophilus* to grow in situ.
Methanohalophilus Isolate Recovered from Utica Shale Fluids

A raw production fluid sample collected on day 98 from the Utica shale was used as an inoculum for establishing enrichment cultures on TMA and MMA. Given that only MMA amended fluids yielded methane, these samples were serially diluted to isolate methanogens in shale fluids. However in our original complex media recipe (revision 1 – see Table 3.1), we were unable to enrich methanogens simply on the basis of dilution (from undiluted to $10^{-6}$, the genus *Methanohalophilus* was below 1% microbial community abundance in MiSeq 16S rRNA amplicon data – for primers see Table 3.3B). After evaluating the sequencing results from the dilution series described above, the media recipe was altered to remove trypticase peptone, fatty acids, casamino acids and the amount of yeast extract was lowered from 2.0g/L to 0.1g/L (revision 2 media recipe – see Table 3.2). Multiple antibiotics were also added to new transfers (50ug/mL
kanamycin, 100μg/mL ampicillin, and 200μg/mL streptomycin) as previously described for methanogen isolation (Ticak et al. 2015)). The repeated transfer of the MMA enrichment using media recipe 2 with added selective agents nearly eliminated the previously dominant (67-87%) community member *Halanaerobium* to the level of detection. In these same cultures, the genus *Methanohalophilus* accounted for 99%+ of the microbial community in dilution samples sequenced.

**Table 3.1**

| Media recipe revision 1 (based on DSMZ #479) used for original enrichment of *Methanohalophilus* from Utica shale production fluids. |
|---|---|
| **Compound** | g/L |
| NaCl | 87.00 |
| KCl | 1.50 |
| MgCl₂-6 H₂O | 6.00 |
| CaCl₂-2 H₂O | 0.10 |
| NH₄Cl | 1.00 |
| K₂HPO₄ | 0.305 |
| CoM sodium salt | 0.20 |
| NaHCO₃ | 4.00 |
| DSMZ #141 Trace Elements Mix | 10.00mL |
| Yeast Extract | 2.00 |
| Trypticase Peptone | 2.00 |

<table>
<thead>
<tr>
<th><strong>Post Autoclaving ingredients added</strong></th>
<th>Final Concentration [mM]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na₂S-9 H₂O</td>
<td>104</td>
</tr>
<tr>
<td>L-cysteine-HCl-H₂O</td>
<td>1.42</td>
</tr>
<tr>
<td>Valeric acid</td>
<td>2.45</td>
</tr>
<tr>
<td>Isovaleric acid</td>
<td>2.45</td>
</tr>
<tr>
<td>Isobutyric acid</td>
<td>2.84</td>
</tr>
<tr>
<td>2-Methylbutyric acid</td>
<td>2.45</td>
</tr>
<tr>
<td>monomethylamine-HCl</td>
<td>10.00</td>
</tr>
</tbody>
</table>
Table 3.2
Media recipe revision 2 used for isolation, substrate testing and growth curve (Figure 3.6) of *Methanohalophilus fracturphilus*.

<table>
<thead>
<tr>
<th>Compound</th>
<th>g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>87.00</td>
</tr>
<tr>
<td>KCl</td>
<td>1.50</td>
</tr>
<tr>
<td>MgCl₂·6 H₂O</td>
<td>6.00</td>
</tr>
<tr>
<td>CaCl₂·2 H₂O</td>
<td>0.10</td>
</tr>
<tr>
<td>NH₄Cl</td>
<td>1.00</td>
</tr>
<tr>
<td>K₂HPO₄</td>
<td>0.305</td>
</tr>
<tr>
<td>CoM sodium salt</td>
<td>0.20</td>
</tr>
<tr>
<td>NaHCO₃</td>
<td>4.00</td>
</tr>
<tr>
<td>DSMZ #141 Trace Elements Mix</td>
<td>10.00mL</td>
</tr>
<tr>
<td>Yeast Extract</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Post Autoclaving ingredients added

<table>
<thead>
<tr>
<th>Compound</th>
<th>Final Concentration [mM]</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-cysteine-HCl-H₂O</td>
<td>1.42</td>
</tr>
<tr>
<td>monomethylamine-HCl</td>
<td>10.00</td>
</tr>
</tbody>
</table>

To assess if the highly enriched *Methanohalophilus* enrichment had yielded a pure culture, grown up cells were visually inspected using epifluorescence microscopy, which revealed a uniform cocci morphology (Figure 3.5). DNA was extracted from the enrichment, and PCR was performed using archaeal primers (forward designed by Wrighton research group) and universal bacterial primers (Table 3.3A). No bacterial amplification was observed from the MMA enrichment, and archaeal amplicons were purified and Sanger sequenced. A near full length (1326nt) 16S rRNA gene sequence was obtained and found to be within 1% identity to *Methanohalophilus* 16S rRNA gene sequences recovered from both the Marcellus and Utica shales (Table 3.4).
Table 3.3
Primers utilized in this study.

A Primers used for PCR amplification of the 16S rRNA gene to assess purity of *Methanohalophilus fracturphilus*.

<table>
<thead>
<tr>
<th>Archaeal F Primer</th>
<th>Archaea and Bacterial R Primer</th>
<th>Bacterial F Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>4Fa 5’ – TCCGGTTGATCCTGCCR G – 3’</td>
<td>1492R 5’ – GGTACCTTGTTACGA CTT – 3’</td>
<td>27F 5’ – AGAGTTTGATCCTGG CTCAG – 3’</td>
</tr>
</tbody>
</table>

B Primers used for PCR amplification of the 16S rRNA gene to investigate microbial community structures in enrichment samples.

<table>
<thead>
<tr>
<th>Forward Universal Primer</th>
<th>Reverse Universal Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>515F 5’ – GTGCCAGCMGCCGCGGTAA – 3’</td>
<td>805R 5’ – GGACTACHVGGGTWTCTAAT – 3’</td>
</tr>
</tbody>
</table>

Table 3.4
Comparison of nucleotide identity for the 16S rRNA gene sequence obtained from *Methanohalophilus fracturphilus* (BLASTn aligned) against select *Methanohalophilus* pure cultures and metagenomic bins.

<table>
<thead>
<tr>
<th></th>
<th>Utica day 94</th>
<th>Utica day 96</th>
<th>Utica day 98</th>
<th>Marcellus day 328</th>
<th>Methano-halophilus mahii</th>
<th>Methano-halophilus halophilus sp. Z-7982</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleotide Alignment length</td>
<td>1326</td>
<td>1314</td>
<td>1317</td>
<td>1326</td>
<td>1326</td>
<td>1326</td>
</tr>
<tr>
<td>% Identity</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
</tr>
</tbody>
</table>

After obtaining preliminary evidence of isolate purity, efforts were undertaken to assess viable growth substrates for the obtained culture. DMA, MMA, and methanol all supported growth (See Figure 3.6 for growth on MMA and Methanol), while TMA, acetate, choline, tetramethylammonium, H₂:CO₂ (80:20 v/v), and dimethylglycine all failed to elicit increases in optical densities from time zero. The finding that this organism could not utilize TMA was unexpected since all other cultured
Methanohalophilus isolates have been shown to be capable of growth on methanol, TMA, DMA, and MMA (when assessed) (Figure 3.7).

Figure 3.5 Brightfield micrograph of Methanohalophilus fracturphilus wet mount at 1000x total magnification; a red 10 um scale bar is included towards the bottom of the figure.

Figure 3.6 Growth curve of Methanohalophilus fracturphilus demonstrating growth on MMA and methanol.
A member of the genus, *Methanohalophilus levihalophilus* (*Mhlh*), was previously isolated at less than 1 km below ground from the highly productive Japanese Minami-Kanto natural gas field (Katayama et al. 2014). It should be noted that the conditions present in the isolation location of *Mhlh* were notably distinct from those present in the shales described in this manuscript. Most notably, *Mhlh* was isolated from a brackish (15g/L Cl⁻) production fluid collected from a shallow depth of 180-800 meters. This is in stark contrast to the brine level salinity (66g/L Cl⁻) fluid (likely collected from considerably greater depth) used to obtain the newly described (see chapter 2 of this
thesis) *Methanothalophilus* isolate. The *Methanothalophilus* culture described in this study represents the first methanogen isolated from hydraulically fractured shale, and as such, the name *Methanothalophilus fracturphilus* is proposed.

*Methanothalophilus* Genomes from Pure Culture and Produced Fluids

A 1L batch culture of *Methanothalophilus fracturphilus* (grown on methanol in media revision 2 (Table 3.2)) was spun down and cell mass was harvested by centrifugation. Obtained DNA was sequenced yielding a genome estimated to be 100% complete (using 104 archaeal single copy genes) with 1.94Mb of sequence from 113 scaffolds with an average coverage of 2808x (Table 3.5). Out of the 104 archaeal single copy genes recovered, zero genes were present in multiple copies suggesting the genome was free of contamination from other taxa.

To obtain genomes from shale produced fluids we used metagenomic sequencing of DNA originating from Marcellus and Utica fluids. From Marcellus fluids we obtained a 2.09Mb genomic bin spread across 305 scaffolds that is estimated to be 97% complete (using same metric as above) with an average coverage of 125x (Table 3.5). Out of the 101 archaeal single copy genes found, only one was present in two copies, indicating that the genome was largely free of contaminating sequence from other community members.

From three Utica samples, *Methanothalophilus* bins were recovered. The Utica day 98 produced water sample yielded a *Methanothalophilus* bin 1.73Mb in length that was estimated to be 100% complete (same metric as above) and was spread across 418 scaffolds with an average coverage of 17x (Table 3.5). This genome had three instances of single copy genes present in two copies, indicating a possible low level of contamination from other taxa. Another near complete *Methanothalophilus* genome was
recovered from day 94 (1.9MB - 98.1% complete). Finally, a smaller and less complete (1.4MB – 83.7% complete) *Methanohalophilus* bin was recovered from day 96. From the Utica samples, only the day 98 *Methanohalophilus* bin was pursued in the following analysis.

**Table 3.5**

<table>
<thead>
<tr>
<th>Organism</th>
<th>Genome Size</th>
<th>GC %</th>
<th># of Scaffolds</th>
<th>Estimated Completion</th>
<th>Average coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Methanohalophilus mahii</em>**</td>
<td>1.82 MB</td>
<td>42.6%</td>
<td>1</td>
<td>100%</td>
<td>NA</td>
</tr>
<tr>
<td><em>Methanohalophilus halophilus</em> sp. Z-7982***</td>
<td>2.03 MB</td>
<td>42.4%</td>
<td>10</td>
<td>100%</td>
<td>NA</td>
</tr>
<tr>
<td><em>Methanohalophilus fracturphilus</em>**</td>
<td>1.94 MB</td>
<td>42.2%</td>
<td>113</td>
<td>100%</td>
<td>2808x</td>
</tr>
<tr>
<td><em>Methanohalophilus bin Marcellus day 328</em></td>
<td>2.09 MB</td>
<td>42.0%</td>
<td>305</td>
<td>97%</td>
<td>125x</td>
</tr>
<tr>
<td><em>Methanohalophilus bin Utica day 98</em></td>
<td>1.73 MB</td>
<td>43.2%</td>
<td>418</td>
<td>100%</td>
<td>17x</td>
</tr>
</tbody>
</table>

**Carbon and Energy Metabolism and Strategies of Osmoprotection**

The genomes of the two shale derived (Marcellus and Utica) *Methanohalophilus* genomes were analyzed alongside the *Methanohalophilus fracturphilus* isolate genome. The results below are applicable to all three genomes unless otherwise specified. All scaffolds and genes for the genes mentioned are listed in Appendix A. Given that none of these genomes are closed, it can not be ruled out that absent genes could simply be present on portions of the genomes that we do not have appreciable sequencing information for; ergo, gene absences should be taken as preliminary data.
Substrate Catabolism and Preparation for Entry into the two Branches of Methanogenesis

The energy metabolism of the recovered *Methanohalophilus* genomes is consistent with canonical methylotrophic methanogens and other *Methanohalophilus* isolates (Deppenmeier 2002, Spring et al. 2010). Genes for methylamine (Tri-, Di-, and Mono-) and methanol methyltransferases and cognate corrinoid proteins were detected. Protein alignments indicating the position of the in frame pyrrolysine amber codon for the MMA, DMA, and TMA methyltransferases (Gaston et al. 2011) are shown in Figure 3.8. It should be noted that in the Marcellus day 328 *Methanohalophilus* bin irregularities were observed in the genes encoding for both the DMA corrinoid methyltransferase (MtbB) and its cognate corrinoid protein (MtbC). The Marcellus day 328 MtbB gene sequence appears to contain a 4nt deletion when aligned against close reference sequences, which results in a frameshift and premature encoding of stop codons. Interestingly, the gene encoding for the DMA methyltransferase’s cognate corrinoid protein, MtbC contains a starting codon encoding for leucine instead of a canonical methionine. These unanticipated results raise doubt regarding the ability of the Marcellus day 328 *Methanohalophilus* to catabolize DMA in vivo.

Figure 3.8 Amino acid alignments of methylamine corrinoid methyltransferases. The location of the in frame amber codon encoding for pyrrolysine is indicated. Positions in white indicate that less than 60% of the sequences contain the same residue. Positions in grey indicate 60-99% of sequences share the same residue. Positions in black indicate that all sequences share the same residue. The pyrrolysine residue is consistent in all sequences and is shown in white because it is not recognized as one of the 20 canonical amino acids.
Consistent with physiological characterization where TMA did not support growth, the isolate genome lacks genes for the trimethylamine methyltransferase (MttB) and its cognate corrinoid protein (MttC). For all three genomes, the inferred metabolic framework suggests it is possible that methylated corrinoid proteins could transfer a methyl group to coenzyme M (CoM) using either a methylamine (MtbA) or methanol (MtaA) specific methylcobalamin:CoM methyltransferase. Based on prior reports (Deppenmeier 2002), three of four methylated CoM pass through the reductive branch of methanogenesis, ultimately resulting in the liberation of 3 molecules of methane, whereas the remaining single methyl-CoM donates a methyl group that is oxidized to CO₂ (Figure 3.9).
Figure 3.9 Inferred energy metabolism of *Methanohalophilus fracturphilus*, Utica and Marcellus *Methanohalophilus* metagenome bins. The gene sequences for the trimethylamine corrinoid methyltransferase (MttB) and cognate corrinoid protein (MttC) are absent from the genome of *Methanohalophilus fracturphilus* as indicated by the pink star in the diagram. The area shaded in green represents the oxidative branch of methanogenesis. The area shaded in blue is the reductive branch of methanogenesis. The area shaded in orange contains select features of chemiosmotic energy conservation found in the *Methanohalophilus* genomes. Abbreviations used: H$_4$MPT=tetrahydromethanopterin, MF=methanofuran, Fd=ferredoxin, MP and MPH$_2$=methanophenazine, CoM=coenzyme M. Enzyme #1: methyl-tetrahydromethanopterin:coenzyme M methyltransferase, #2: methylene-tetrahydromethanopterin reductase, #3: methylene-tetrahydromethanopterin dehydrogenase, #4: methenyl-tetrahydromethanopterin cyclohydrolase, #5: formyl-methanofuran:tetrahydromethanopterin formyltransferase, #6: formyl-methanofuran dehydrogenase. MttB is the trimethylamine corrinoid methyltransferase. MtbB is the dimethylamine corrinoid methyltransferase. MtmB is the monomethylamine corrinoid methyltransferase. MtaB is the methanol specific corrinoid methyltransferase. MtbA is the methylamine specific methylcobalamin:CoM methyltransferase and MtaA is the methanol specific version. Mcr is the methyl-coenzyme M reductase complex. Hdr is the heterodisulfide reductase complex. Scaffold hits for all illustrated proteins are listed in Appendix A.
Oxidative Branch Methanogenesis

We infer that methyl-CoM’s methyl group would be transferred to the carbon carrier tetrahydromethanopterin (H₄MPT) or tetrahydrosarcinopterin (H₄SPT) through the membrane bound enzyme methyl-H₄MPT:CoM methyltransferase. (To the knowledge of the author, biochemical characterization is required to distinguish between the methanogen carbon carriers H₄MPT and H₄SPT. Thus, the distinction between these two pterin based molecules can not be made in the genomes analyzed and the species in question will hereafter be referred to as H₄MPT for brevity). The methyl-H₄MPT would then be sequentially oxidized to methylene, methenyl and then converted formyl-H₄MPT using the enzymes methylene-H₄MPT reductase, methylene-H₄MPT dehydrogenase, and methenyl-H₄MPT cyclohydrolase respectively. Through the 2 oxidations and non-redox carbon group conversion listed above, two molecules of the reducing equivalent coenzyme-F₄₂₀ (F₄₂₀) would become reduced and available for cellular use. The formyl group of H₄MPT would then be transferred to the unusual cofactor methanofuran (MF) (Dimarco et al. 1990) through the formyl-MF:H₄MPT formyltransferase. The terminal step of the oxidative branch of methanogenesis is to oxidize the formyl group attached to MF liberating it as CO₂ and in doing so it would reduce one molecule of ferredoxin through the formyl-MF dehydrogenase (Figure 3.9) (Deppenmeier 2002).

Reductive Branch Methanogenesis

From methyl-coenzyme M (CoM) the genome encodes the methyl-CoM reductase complex (Mcr), a diagnostic feature of genomes of methanogens and some archaeal methanotrophs (Hallam et al. 2003). The Mcr complex reduces methylated CoM using coenzyme B (CoB) as an electron donor, which results in the formation of methane and a
heterodisulfide composed of CoM attached to CoB through a disulfide linkage. In order for methanogenesis to proceed, the CoM-S-S-CoB heterodisulfide needs to be reduced into its constituent coenzyme thiols (Deppenmeier 2002, Buan and Metcalf 2010). We found genes for two molecular systems that appear to serve this purpose: the cytoplasmic heterodisulfide reductase (Hdr) ABC system and the membrane attached Hdr DE complex. HdrABC uses reduced ferredoxin to catalyze the reaction. It has been hypothesized that the HdrABC complex also reduces F_{420} while reducing the heterodisulfide (shown in Figure 3.9), but this has yet to be experimentally demonstrated (Buan and Metcalf 2010). The membrane attached HdrED complex uses reduced methanophenazine (the source of which will be described below) and proton translocation to accomplish the splitting of the heterodisulfide. The co-occurrence of both Hdr complexes in a single organism has been demonstrated before in *Methanosarcina acetivorans* (Buan and Metcalf 2010).

**Chemiosmotic Energy Conservation**

All three analyzed genomes contain Rnf genes encoding for subunits A,B,C,E and G. We infer that these proteins would oxidize reduced ferredoxin coupled to sodium ion translocation in order to reduce the methanogen electron carrier methanophenazine (Suharti et al. 2014). Subunits A, B, C, D, F, H, I K, L, M, N and O of the F_{420} dehydrogenase (Fpo) were found. This complex is believed to be membrane attached and has been shown to oxidize reduced F_{420} while reducing methanophenazine coupled to proton translocation. It has previously been shown that in addition to its participation in the assembled Fpo complex the FpoF subunit is individually capable of oxidizing reduced
ferredoxin coupled to the reduction of F_{420} (Welte and Deppenmeier 2011). Subunits A, B, C, D, E, F, H and I were found for a V-type ATPase (Figure 3.9).

**Oxidative Branch of Methanogenesis Produces Inputs for Acetyl-CoA Synthesis**

As discussed above, one complete cycle of the oxidative branch of methylotrophic methanogenesis would reduce one molecule of ferredoxin in addition to producing one molecule of CO_{2} (Figure 3.9). Additionally, it is possible methylated tetrahydromethanopterin could be drawn off for cellular purposes distinct from being oxidized by methylene-tetrahydromethanopterin reductase. Thus, it appears that from the oxidative branch of methylotrophic methanogenesis, these organisms could potentially obtain the major necessary constituents (CO_{2}, reduced ferredoxin, and methyl-H_{4}MPT) for the generation of acetyl-CoA (discussed below).

All three genomes possess subunits A, B, D, E, and G of acetyl-CoA decarboxylase/synthase which has been shown to catalyze the synthesis of acetyl-CoA using CO_{2}, reduced ferredoxin and a methylated-carbon carrier (H_{4}SPT has been shown in the literature) (Gencic and Grahame 2003, Dai et al., 1998). From acetyl-CoA we hypothesize that pyruvate could be formed using reduced ferredoxin and CO_{2} through the action of the pyruvate ferredoxin oxidoreductase (subunits A, B, D, and G were recovered) (Furdui and Ragsdale 2000). We believe phosphoenolpyruvate (PEP) could be formed with pyruvate phosphate dikinase, or phosphoenolpyruvate synthetase (Figure 3.10) (Tjaden et al. 2006, Eyzaguirre et al. 1982).

The produced Acetyl-CoA appears to be the starting point for assimilatory carbon processes used to produce the biosynthetic precursor compounds: pyruvate, phosphoenolpyruvate, ribose-5-P, oxaloacetate, and alpha-ketoglutarate (Figure 3.10).
These species are conventionally believed to be critical for the production of key biomolecules needed by the cell (Lengeler et al., 1999).

**Citric Acid Cycle**

We hypothesize that pyruvate could be converted into oxaloacetate (using pyruvate carboxylase), which could then be acted on by enzymes from the TCA cycle (Jitrapakdee et al. 2008). A high confidence citrate synthase could not be recovered through BLAST searches against the annotated amino acid reference sequence from the *Methanohalophilus mahii* genome. However, all three genomes did have 75% amino acid identity BLAST matches to an amino acid sequence from *Methanococcoides burtonii* proposed to be orthologous to an alternately stereospecific citrate synthase previously characterized in *Clostridium* (Li et al. 2007, Allen et al. 2009). In a prior investigation of *Methanococcoides burtonii*, proteomics detected peptides for aconitase and isocitrate dehydrogenase (Goodchild et al. 2004), consistent with the proposal that the organism has a means of producing citrate (Allen et al. 2009).

It appears that many of the necessary enzymes requisite for the remainder of the oxidative branch of the TCA cycle are present. It is likely that citrate could be converted to cis-aconitate and finally to isocitrate with aconitase. Using isocitrate dehydrogenase, isocitrate could be converted to oxalosuccinate, which could also be converted by the same enzyme to alpha-ketoglutarate (Alberts et al. 2002). Alpha-ketoglutarate could then be converted to succinyl-CoA with alpha-ketoglutarate ferredoxin oxidoreductase (Figure 3.11) (Baughn et al. 2009). The absence of malate dehydrogenase from the annotations of all three genomes suggests that a glyoxylate bypass system is not present in the genomes analyzed.
Figure 3.10 Inferred assimilatory carbon scheme in analyzed *Methanohalophilus* genomes. According to this framework, energy metabolism generates CO$_2$, methylated tetrahydromethanopterin and reduced ferredoxin that can be converted into acetyl-CoA. In this scheme, pyruvate would be formed from acetyl-CoA and subsequently used to generate carbon compounds that are needed for Gluconeogenesis. Some chemical conversions performed by enzymes from the citric acid cycle (TCA) and a truncated (indicated by **) portion of the Pentose Phosphate cycle are shown. The groups of four arrows between two chemical species indicates multiple reactions and intermediates, however the number of arrows does not reflect the number of genes in the pathway. Red x’s over a single line in a group of four arrows indicates that the pathway reconstruction was partial, with select genes missing or unknown in methanogens. The large single red x covering the lines between ribose-5-p and erythrose-4-p represent the conspicuous apparent absence of transketolase and transaldolase.
**Figure 3.11** Reconstruction of enzymes from the citric acid cycle (TCA) in the three analyzed *Methanohalophilus* genomes. Red x’s indicate missing genes. The red question mark on Citrate Synthase indicates that there is a possible hit for this gene (discussed in text), but it cannot be annotated with a high degree of confidence. Enzyme Numbers:
1 = pyruvate carboxylase, 2 = citrate synthase, 3 = aconitase, 4 = isocitrate dehydrogenase, 5 = alpha ketoglutarate ferredoxin oxidoreductase, 6 = succinate thiokinase, 7 = succinate dehydrogenase, 8 = fumarase, 9 = malate dehydrogenase.
As mentioned above, both alpha-ketoglutarate and oxaloacetate are important intermediates centrally important to the synthesis of other key biomolecules (Lengeler et al., 1999).

For the reductive branch of the TCA cycle malate dehydrogenase is present which can convert oxaloacetate to malate. From malate, fumarate could be formed with fumarase (Figure 3.11) (Alberts et al. 2002).

As can be gathered from the genes reported above, the three *Methanohalophilus* genomes analyzed do not contain complete TCA cycles. This finding is consistent with previous reports of methanogens containing fragmented TCA cycles (Spring et al. 2010) (Huynen et al., 1999).

**Gluconeogenesis**

All genomes encode an Embden-Meyerhof-Parnas pathway (Lengeler et al., 1999). As such, it is possible that phosphoenolpyruvate could enter gluconeogenesis by being converted into 2-phospho-glycerate through the activity of enolase. From 2-phospho-glycerate 3-phospho-glycerate could be generated with phosphoglycerate mutase. From 3-phospho-glycerate 1,3-bisphosphoglycerate could be formed with phosphoglycerate kinase. Coupled to the oxidation of NADH, 1,3-bisphosphoglycerate could be converted into glyceraldehyde-3-phosphate through the activity of glyceraldehyde-3-phosphate dehydrogenase. Some fraction of glyceraldehyde-3-phosphate would then potentially be converted into dihydroxyacetone phosphate with triose phosphate isomerase. Together, dihydroxyacetone phosphate and glyceraldehyde-3-P could be converted into a single unit of fructose-1,6-bisphosphate with fructose-1,6-bisphosphate aldolase. Finally, fructose-1,6-bisphosphate could be converted to fructose-6-phosphate with either fructose-1,6-
bisphosphatase or an ADP specific phosphofructokinase (Figure 3.12). This produced fructose-6-phosphate would then potentially be able to enter the partially recovered Pentose Phosphate pathway (Figure 3.13) (Lengeler et al., 1999).

![Reconstructed Embden-Meyerhof-Parnas (EMP) pathway of Gluconeogenesis. Enzyme Numbers: 1 = enolase, 2 = phosphoglycerate mutase, 3 = phosphoglycerate kinase, 4 = glyceraldehyde-3-phosphate dehydrogenase, 5 = triose phosphate isomerase, 6 = fructose-bisphosphate aldolase, 7 = fructose-1,6-bisphosphatase or ADP-specific phosphofructokinase, 8 = glucose-6-phosphate isomerase.](image)

**Figure 3.12** Reconstructed Embden-Meyerhof-Parnas (EMP) pathway of Gluconeogenesis. Enzyme Numbers: 1 = enolase, 2 = phosphoglycerate mutase, 3 = phosphoglycerate kinase, 4 = glyceraldehyde-3-phosphate dehydrogenase, 5 = triose phosphate isomerase, 6 = fructose-bisphosphate aldolase, 7 = fructose-1,6-bisphosphatase or ADP-specific phosphofructokinase, 8 = glucose-6-phosphate isomerase.

**Pentose Sugar Synthesis**

Only the beginning of a classical Pentose Phosphate pathway was recovered from the three *Methanothalophilus* genomes analyzed. From the genomes, it is inferred fructose-6-phosphate (generated from gluconeogenesis) could be converted into arabino-3-hexulose-6-
phosphate through the action of 6-phospho-3-hexulose isomerase (Taylor et al., 2001). Arabino-3-hexulose-6-phosphate could then be converted to ribulose-5-P by 3-hexulose-6-phosphate synthase (Kato et al., 2006). Finally, ribulose-5-phosphate could be converted with ribose-5-phosphate isomerase into ribose-5-phosphate (Figure 3.13), which is believed to be used for nucleotide synthesis (Lengeler et al., 1999). Interestingly, the genes required for the synthesis of erythrose-4-phosphate (transketolase and transaldolase) appear absent from the annotations of the three genomes investigated. The absence of transketolase and transaldolase was independently evaluated with BLAST searches against annotated proteins from *Methanocaldococcus jannaschii*; the searches yielded low maximum bit scores (transketolase = 28 and transaldolase = 26). However, despite the apparent lack of genes annotated as aiding in the production of erythrose-4-phosphate, there appears to be an alternative entrance into the Shikimate pathway through L-aspartate-4-semialdehyde, which has been previously documented in *Methanocaldococcus jannaschii* (White 2004). This proposed alternative entrance into the Shikimate pathway is shown in Figure 3.14.
Figure 3.13 Pentose sugar synthesis in the three *Methanohalophilus* genomes analyzed.

Figure 3.14 Proposed alternate entry into the Shikimate pathway that bypasses the need to produce erythrose-4-phosphate. Enzymes Numbers: 2 = aspartate kinase, 3 = aspartate semialdehyde dehydrogenase, 4 = 2-amino-3,7-dideoxy-D-threo-hept-6-ulosonate synthase, 5 = 3-dehydroquinate synthase.
**Acetyl-CoA Synthesis**

Given the central role of coenzyme A (CoA) in the conversion of CO₂ into biologically available acetyl-CoA, the gene sequences putatively involved in the CoA biosynthetic process were investigated in the three studied *Methanohalophilus* genomes. In both archaea and bacteria, it is thought that CoA synthesis begins with the conversion of 2-ketovalerate into ketopantoate, which in turns becomes pantoate. In bacteria, the two enzymes responsible for these conversions (ketopantoate hydroxymethyltransferase and ketopantoate reductase) and their substrates have been well characterized (Begley et al. 2001). To date these enzymes and their substrates have yet to be robustly characterized in methanogens (Genschel 2004, Wang et al. 2014). For the enzyme ketopantoate hydroxymethyltransferase the methyl donor remains elusive in methanogens, but it has been proposed that methylene-\(\text{H}_4\text{MPT}\) could serve this purpose (as is indicated with a question mark adjacent to \(\text{H}_4\text{MPT}\) in Figure 3.15). From pantoate, the studied bacterial and archaeal pathways each create one unique intermediate and then resume with the same intermediates ultimately resulting in biologically available CoA (Wang et al. 2014). In methanogens, the final enzyme in the pathway has also yet to be described (Genschel 2004). Given the lack of sequenced references, this study was limited to searching for these three unknown genes in annotations; none of these three genes were found in any of the three analyzed genomes. All other enzymes believed to be involved in CoA synthesis were found through BLAST searches using pulled *Methanohalophilus mahii* reference sequences. The proposed pathway is shown in Figure 3.15, which was reproduced from Genschel 2004 and modified to reflect known archaeal peculiarities (Wang et al. 2014).
Figure 3.15 The scheme of coenzyme A synthesis was reproduced from Genschel 2004, but was modified to account for subsequent findings presented in Wang et al., 2014.

KPHMT = ketopantoate hydroxymethyltransferase, KPR = ketopantoate reductase, ADC = aspartate decarboxylase, PK = pantoate kinase, PS = pantothenate synthetase, PPCS = phosphopantothenoylcysteine synthetase, PPCDC = phosphopantothenoylcysteine decarboxylase, PPAT = pantetheine-phosphate adenylyltransferase, DPCK = dephosphocoenzyme A kinase.
Mechanisms of Osmo-Tolerance
The analyzed *Methanohalophilus* genomes encode three distinct mechanisms of osmo
tolerance. *Methanohalophilus* have been shown to accumulate potassium ions, N^ε^-acetyl-β-
lysine and glycine betaine in response to elevated environmental salinity (Lai et al. 1991).
The genomes contain the Trk complex (H and I subunits), which has been shown through a
heterologous expression study to be responsible for intracellular K^+ accumulation (Kraegeloh
et al. 2005). The *Methanohalophilus* genomes also contain genes encoding for the
conversion of α-lysine into the compatible organic solute N^ε^-acetyl-β-lysine. The proposed
pathway of N^ε^-acetyl-β-lysine synthesis follows the conversion of α-lysine into β-lysine
through lysine-2,3-aminomutase and then β-lysine would then be converted to N^ε^-acetyl-β-
lysine through the activity of β-lysine acetyltransferase (Pfluger et al. 2003).

Glycine betaine has been shown to be the most abundantly accumulated compatible
organic solute in at least one *Methanohalophilus* species (Lai et al. 1991). Genes Encoding
for glycine sarcosine and sarcosine dimethylglycine methyltransferases were found in all
*Methanohalophilus* genomes analyzed. Glycine sarcosine methyltransferases have been
shown to catalyze the conversion of glycine to sarcosine and sarcosine to dimethylglycine.
The sarcosine dimethylglycine methyltransferase have been shown to facilitate the
conversion of sarcosine to dimethylglycine and dimethylglycine to glycine betaine (Table
3.6). Both enzymes utilize S-adenosyl-L-methionine (SAM) as a methyl-donor (Lai and Lai
2011). The amino acid sequences of the two methyltransferases involved in glycine betaine
synthesis from glycine were recovered from all available *Methanohalophilus* genomes and
compared to the characterized proteins from *Methanohalophilus portucalensis* str. FDF1; the
alignment length and highly similar amino acid identities are presented in Table 3.7.
Table 3.6
EC numbers and function of the two bifunctional methyltransferases found in Methanohalophilus genomes; previously empirical work has verified the ability of these enzymes to convert glycine to glycine betaine.

<table>
<thead>
<tr>
<th>E.C. 2.1.1.156</th>
<th>E.C. 2.1.1.157</th>
</tr>
</thead>
<tbody>
<tr>
<td>glycine -&gt; sarcosine</td>
<td>sarcosine -&gt; dimethylglycine</td>
</tr>
<tr>
<td>sarcosine -&gt; dimethylglycine</td>
<td>dimethylglycine -&gt; glycine betaine</td>
</tr>
</tbody>
</table>

Table 3.7
Amino acid alignment statistics (produced by BLASTp) of recovered Methanohalophilus methyltransferases compared to those previously characterized from Methanohalophilus portucalensis str. FDF1. The abbreviation Mh. is used to denote Methanohalophilus.

<table>
<thead>
<tr>
<th></th>
<th>Methanohalophilus portucalensis str. FDF1 glycine sarcosine methyltransferase (263aa)</th>
<th>Methanohalophilus portucalensis str. FDF1 sarcosine dimethylglycine methyltransferase (278)</th>
</tr>
</thead>
<tbody>
<tr>
<td>%ID:Alignment length(aa)</td>
<td>96% : 263</td>
<td>98% : 278</td>
</tr>
<tr>
<td>Utica day 98 Mh. bin</td>
<td>96% : 263</td>
<td>98% : 278</td>
</tr>
<tr>
<td>Methanohalophilus fracturphilus</td>
<td>96% : 263</td>
<td>98% : 278</td>
</tr>
<tr>
<td>Marcellus day 328 Mh. bin</td>
<td>96% : 263</td>
<td>98% : 278</td>
</tr>
<tr>
<td>Methanohalophilus mahii</td>
<td>95% : 263</td>
<td>91% : 278</td>
</tr>
<tr>
<td>Methanohalophilus halophilus sp. Z-7982</td>
<td>98% : 263</td>
<td>97% : 278</td>
</tr>
</tbody>
</table>

Gene Synteny of Glycine Betaine Synthetic Machinery
It should be noted that aside from the genus Methanohalophilus, the only other methanogen genome found to contain the two enzymes required for the methylation of glycine into the compatible organic solute trimethylglycine or glycine betaine (GB) is Methanohalobium evestigatum. A phylogenetic tree of the glycine sarcosine methyltransferase (Figure 3.16) illustrates the tight clustering of Methanohalophilus sequences and a broader yet distinct clade composed of methanogen sequences. Interestingly, on the 16S rRNA salinity tree presented in the introduction of this thesis it can be seen that Methanohalobium evestigatum is the only other cultured methanogen that has a
high optimum NaCl concentration for growth at 249g/L (Davidova et al., 1997). The synteny for genes involved in glycine betaine synthesis is conserved between all genome sequenced *Methanohalophilus* and *Methanohalobium evestigatum* (Figure 3.17). Beyond containing the two requisite bifunctional methyltransferases for de-novo GB synthesis from glycine (Table 3.6), this syntenic gene cluster contains three additional genes whose synteny is also largely conserved (Figure 3.17). This gene cluster encodes for three proteins whose functions could potentially support the production of glycine betaine (Lai and Lai 2011). The encoded protein products are a single methionine adenosyltransferase, two S-adenosyl-L-homocysteine hydrolases, and one adenosine kinase. The proposed supportive metabolic cycling is shown in Figure 3.17.
Figure 3.16 Phylogenetic tree of glycine sarosine methyltransferase amino acids sequences. This tree has 200 bootstrap replicates; bootstrap support is shown. The red arc at the top of the tree indicates sequences obtained from methanogens.

Figure 3.17 illustrates that each methyl transfer from glycine ultimately resulting in glycine betaine utilizes the methyl carrier SAM and produces the byproduct S-adenosyl-L-homocysteine (SAH). In a study by Lai and Lai, proteins for the two methyltransferases responsible for converting glycine to glycine betaine were heterologously produced in *E. coli*
using genes obtained from *Methanothalophilus portucalensis* str. FDF1. These proteins were assessed in vitro and the results demonstrated that SAH is a potent inhibitor of the methylation activity required to produce glycine betaine (Lai and Lai 2011). The regulation of glycine betaine production by SAH gives credence to the idea that a cell could benefit by co-expressing S-adenosyl-L-homocysteine hydrolase genes along with the methyltransferases necessary for glycine betaine production. The same protein characterization investigation illustrated that potassium and to a lesser extent sodium ions stimulated methylation activity. Most notably, glycine methylation specific activity was shown to be the most dramatically enhanced at 2 M KCl, with a 1353x increase in activity over 0 M (no addition control) KCl. The $K_m$ values for glycine methylation were also shown to be highly effected by KCl concentrations; the $K_m$ at 0M KCl was 2.74 M and the $K_m$ at 2.0 M KCl was 55.0 mM (Lai and Lai 2011). The stimulatory effect of potassium ions on the aforementioned binding and methylation activity is of particular interest as *Methanothalophilus* species have previously been shown to co-accumulate potassium ions alongside compatible organic solutes in proportion to the osmotic strength of the environment (Lai et al. 1991). The potential link between intracellular potassium ion and SAH concentrations and the ability of cells to produce the compatible organic solute glycine betaine is indeed intriguing and warrants further study. While it could be coincidental that all investigated *Methanothalophilus* and *Methanothalobium evestigatum* contain the same gene cluster for GB synthesis, it is also entirely possible this locus could be involved in the preferential growth at high salt concentrations for these organisms.
A: *Methanohalobium evestigatum*
B: *Methanohalophilus mahii*
C: *Methanohalophilus halophilus*
D: Marcellus d328 *Methanohalophilus* Metagenome Bin
E: *Methanohalophilus fracturphilus*
F: Utica day 98 *Methanohalophilus* Metagenome Bin

**Figure 3.17** Syntenous gene cluster recovered from all available *Methanohalophilus* genomes in addition to *Methanohalobium evestigatum* and proposed metabolic scheme for glycine betaine synthesis-numbers in the alignment correspond to genes in the metabolism above. Organisms are identified alphabetically (A-F) using the provided legend in the top left. In the amino acid alignment positions in white indicate that less than 60% of the sequences contain the same residue. Positions in light grey indicate 60-99% of sequences share the same residue and positions in black indicate that all sequences share the same residue.

**Global Genome Comparison**

From the analysis presented above it is clear that the three analyzed *Methanohalophilus* genomes have considerable overlap in their gene contents. However, in this text the shale derived *Methanohalophilus* were not discussed in the context of the two
reference *Methanohalophilus* genomes, which were obtained from a salt lake and a marine bay (Wilharm et al. 1991). While it is beyond the scope of this work to thoroughly investigate the similarities and differences between shale and non-shale derived *Methanohalophilus* genomes, an average nucleotide identity comparison (ANI) cluster analysis (Konstantinidis and Tiedje 2005) was conducted as a preliminary assessment of genome similarity. As is shown in Table 3.8, the shale derived genomes appear considerably more similar to one another when compared to non-shale derived *Methanohalophilus*. This was an interesting result since the shale derived *Methanohalophilus* genomes were sampled from geographically and geologically distinct formations. While this data is far from being conclusive, it is suggestive that shale derived *Methanohalophilus* genomes are distinct from their surface counterparts; this topic warrants further investigation.

### Table 3.8

Average nucleotide identity comparison of all genome sequenced *Methanohalophilus*. The abbreviation *Mh.* is used to denote *Methanohalophilus*. The Marcellus day 328 *Methanohalophilus* genome is the same one presented in (Daly et al. submitted).

<table>
<thead>
<tr>
<th></th>
<th><em>Mh. mahii</em></th>
<th><em>Mh. halophilus</em></th>
<th><em>Mh. fracturphilus</em></th>
<th>Marcellus day 328 <em>Mh. bin</em></th>
<th>Utica day 98 <em>Mh. bin</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Mh. mahii</em></td>
<td>100</td>
<td>91.2</td>
<td>91.0</td>
<td>91.0</td>
<td>91.0</td>
</tr>
<tr>
<td><em>Mh. halophilus</em></td>
<td>100</td>
<td>92.0</td>
<td>92.1</td>
<td>92.2</td>
<td></td>
</tr>
<tr>
<td><em>Mh. fracturphilus</em></td>
<td>100</td>
<td></td>
<td>98.6</td>
<td>98.9</td>
<td></td>
</tr>
<tr>
<td>Marcellus day 328 <em>Mh. bin</em></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>98.7</td>
</tr>
<tr>
<td>Utica day 98 <em>Mh. bin</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>
Conclusions

This investigation has analyzed production fluid samples collected over time series for both geochemical trends and genomic content. The genus *Methanohalophilus* was found in both shale plays studied and near complete genomes were reconstructed from select samples. Additionally, a *Methanohalophilus* isolate was obtained from the Utica shale, which also yielded a near complete genome. In comparing the genes present in the three analyzed *Methanohalophilus* genomes it appears that they share many of the same capabilities in terms of substrate utilization, energy metabolism, strategies of osmoprotection, and overall scheme of carbon cycling. This study has broadened the knowledge of the metabolic potential (as inferred from genomic analysis) for shale derived *Methanohalophilus*. Further study will be required to better understand why this group continues to be found in shale systems more often than any other methanogenic genera.
Materials and Methods

Sample Collection and Processing
Fluid samples were collected at wellheads of actively producing hydraulic fracturing wells with the assistance of industry operators. Fluid samples were stored in autoclaved bottles until returning to the laboratory where 300-1000mL was filtered onto 0.22um pore size PES filter discs (Millipore, Fisher Scientific) and stored at 80C. The Utica day 98 sample which was used to produce the isolate *Methanohalophilus fracturphilus* was transferred to a serum vial, degassed with N2 (99.99% purity gas passed through a Supelco NC9599262 oxygen scrubber) for 20 minutes and then sealed with a butyl rubber stopper and incubated at 37C until it was used for inoculation within 48 hours of sample collection.

DNA Extractions
Marcellus produced water yielded DNA with the MoBio Powersoil DNA Isolation kit (MoBio, California). DNA was extracted from Utica produced water samples using 2 separate DNA extractions (on each individual sample) and resulting DNA extracts were pooled before downstream sequencing preparation. Briefly, filter discs were added to polypropylene tubes along with 1x Phosphate buffered saline solution. Tubes were shaken vigorously by hand, bead beaten at 5.5m/s for 3 30 bursts and then centrifuged at 10k x g for 2 minutes. Residual supernatant was reserved for a syringe driven extraction described below. A modified phenol chloroform extraction (Wrighton et al. 2008) was applied in parallel on the pellet produced by centrifugation. DNA obtained from phenol chloroform
extractions was incubated with 100% ethanol, sodium acetate and linear acrylamide at 4C overnight and allowed to precipitate. The following day precipitated DNA was washed with 70% ethanol, and finally resuspended in 10mM Tris prior to being stored at -80C.

Collected Utica filter disc supernatants (described above) were subjected to the Wizard PCR clean up kit (Promega, Wisconsin) as a form of DNA extraction. DNA was eluted in 10mM Tris and stored at -80C. Finally, the DNA extracts obtained using the Wizard and phenol chloroform extractions were combined together for each Utica time point.

*Methanohalophilus fracturphilus* was grown in a 1L batch culture on methanol and spun down in autoclaved centrifuge bottles that were triple washed. Pelleted cells were loaded into 6 individual MoBio Powersoil DNA extractions; after quantification, DNA was pooled for Illumina HiSeq library preparation.

Five mL of homogenized culture media was removed from enrichments, spun down in microcentrifuge tubes and DNA was obtained using the MoBio Powersoil DNA isolation kit.

**DNA Sequencing**

DNA obtained from Marcellus samples was prepared for sequencing using the Nugen Ovation Ultralow Library system (Nugen, California) with ~8-10 cycles of PCR amplification. The prepared libraries were then sequenced on an Illumina HiSeq 2000 in paired-end mode with 113 cycles. Utica DNA extracts were sent to the Joint Genome Institute and Nextera XT low biomass libraries (Illumina, California) were prepared with 15 cycles of amplification and sequenced on an Illumina HiSeq. A KAPA Hyper library (Kapa, Massachusetts) was prepared using DNA obtained from *Methanohalophilus fracturphilus* using 5 cycles of PCR amplification at the Ohio State University Comprehensive Cancer
The library was sequenced on two separate runs on an Illumina HiSeq in paired ended mode and the data was pooled for downstream analysis. 16S rRNA genes were amplified and barcoded (with primers listed in Table 3.3B) from Utica enrichment cultures. Amplicons were sequenced on an Illumina MiSeq by the Department of Energy.

**Enrichment/Isolation Procedures**

The Utica day 98 sample was inoculated into DSMZ #479 media (media revision #1, see Table 3.1) at 20% of the total volume into anaerobic tubes (chemglass #CLS-4209). Subsequent transfers used 10% of the final culture volume as inoculum. Post autoclaving ingredients and inoculum were added anaerobically on the benchtop by degassing sterile syringes prior to pulling and injecting solutions. Sterilized media without post-autoclaving ingredients was stored in the dark at room temperature until use. Community abundances were profiled by 16S rRNA gene sequencing using the Illumina MiSeq platform described above.

**Isolate 16S rRNA Gene Sequencing**

DNA extracted from *Methanohalophilus fracturphilus* was PCR amplified using bacterial and archaeal primers listed in Table 3.3A. Amplifications were performed with an Eppendorf Nexus Gradient thermocycler (Eppendorf, Germany) with Takara taq (Clontech catalog #R011, Germany), in 50uL reactions. DNA extracted from *Halanaerobium congolense* and *Methanohalophilus mahii* was used as known positive controls for the bacterial and archaeal primer sets respectively. DNA from these two organisms was also used to confirm a lack of amplification with each primer set from the other domain of life.

Both bacterial and archaeal PCR reaction mixes, which were loaded with DNA from *Methanohalophilus fracturphilus* were run on a 1% agarose gel. Bacterial reactions were
blank as expected and the archaeal primers yielded a band of approximately 1488 nt length as expected. The archaeal PCR amplicons were cleaned using the Zymo Research DNA clean and Concentrator kit (Zymo Research, California) and Sanger sequenced with primers 4Fa, 1041R, and 1492R. The resulting chromatograms, were trimmed manually for quality and assembled in Geneious into a single near full length 16S rRNA gene consensus sequence.

**Growth Curve Conditions**

The *Methanohalophilus fracturphilus* growth curve utilized media revision 2 (Table 3.2). Cultures were established in triplicate for the following conditions: MMA with culture, MMA without culture, CH$_3$OH with culture, and CH$_3$OH without culture. Three no substrate controls were established to demonstrate how much of the increase in O.D. was a result of substrate carryover from the inoculum. Two of the three no substrate controls were inoculated from cultures previously grown on MMA, and the third was previously grown on CH$_3$OH. The O.D. of the inoculum cultures was 0.077 for MMA and 0.071 for CH$_3$OH. Optical density measurements were taken at 550nm without requiring culture to be taken out of the growth tube, using a Thermo Spectronic Genesys 20 spectrophotometer (Thermo Fisher Scientific, Massachusetts). The points plotted on the lines in Figure 3.6 represent the average of all measurements taken on replicate treatments. The error bars show the high and low values obtained.

**$^1$H NMR and Associated Sample Preparation**

Vacuum dried Marcellus raw production fluid samples were sent to Pacific Northwest National Laboratory for NMR metabolite analysis. Utica samples were analyzed without drying down. Samples were diluted by 10% (v/v) with 5 mM 2,2-dimethyl-2-silapentane-5-sulfonate-d$_6$ (DSS) as an internal standard. All NMR spectra were collected using a Varian
Direct Drive 600 MHz NMR spectrometer equipped with a 5 mm triple resonance salt-tolerant cold-probe. The 1D $^1$H NMR spectra of all samples were processed, assigned, and analyzed by using Chenomx NMR Suite 8.1 (Edmonton, Canada) with quantification based on spectral intensities relative to the internal standard. Candidate metabolites present in each of the complex mixture were determined by matching the chemical shift, J-coupling and intensity information of experimental NMR signals against the NMR signals of standard metabolites in the Chenomx library. A 1D $^1$H spectra was collected following standard Chenomx data collection guidelines (Weljie et al. 2006), employing a 1D NOESY presaturation experiment with 65536 complex points and at least 512 scans at 298 K. Additionally, 2D spectra (including $^1$H-$^1$H TOCSY) were acquired on most of the fluid samples, aiding in the 1D $^1$H assignments of methanol, and MMA.

Due to its significance in this work as an intermediate linking GB fermentation to methanogenesis, MMA was further confirmed in a series of 1D $^1$H NOESY and 2D $^1$H-$^{13}$C HSQC NMR spectra where “spiking” of several different samples was made using an MMA standard. Two additions of ~25uM MMA were made to fluid samples and only the assigned MMA peak ($^1$H chemical shift ~2.62 ppm and $^{13}$C chemical shift ~ 27.7 ppm) increased in intensity. GB concentrations were too low for confirmation with 2-D NMR experiments in the produced fluids. The GB in the Marcellus produced fluid sample series were resolvable and quantified by only the ~3.30 ppm $^1$H resonance but not at ~3.92 ppm due to spectral overlap with ethanolamine; this was confirmed by spiking using a GB standard. In the Utica produced fluids, both resonances (~3.27 and ~3.92 ppm $^1$H) were overlapped with other resonances and could not be resolved by GB spiking.
**Illumina Sequence Data Processing**

Metagenomic datasets from Marcellus and Utica produced water samples in addition to DNA from the isolate *Methanohalophilus fracturphilus* were subjected to the same analysis pipeline. In short, reads were quality trimmed from both the 5’ and 3’ ends using Sickle. After trimming, reads were assembled using IDBA-UD (Peng et al. 2012) with default parameters. Scaffold coverage was calculated by mapping reads back to the assemblies using Bowtie2 (Langmead and Salzberg 2012). Scaffolds were annotated as described previously (Wrighton et al. 2012, Brown et al. 2015), by predicting open reading frames using MetaProdigal (Hyatt et al. 2012), which were subsequently compared to KEGG and UniRef90 using USEARCH (Edgar 2010). The identities of genes of interest involved in osmoprotection, carbon and central energy metabolism were evaluated by tBLASTn (Altschul et al. 1990) aligning metagenomic scaffolds against pulled reference amino acid sequences from the *Methanohalophilus mahii* genome. A combination of manual binning (based on phylogenetic signal from gene annotations), BLAST matches to *Methanohalophilus mahii* and Emergent Self Organizing Maps (ESOM) (Dick et al. 2009) was employed. Only scaffolds 5kb+ were analyzed using ESOM, lower scaffold lengths were attempted but resulted in poor phylogenetic separation (data not reported). Genome completion was estimated by searching for 104 conserved archaeal single copy genes using Amphora2 (Wu and Scott 2012). Average GC was calculated by multiplying the GC of each scaffold assigned to the *Methanohalophilus* bin in question by the length of the scaffold; this product was then divided by the total length of the bin to yield the average GC.
Methanohalophilus Read Mapping for Utica Samples

The genomes of Methanohalophilus mahii, Methanohalophilus halophilus and the Marcellus day 328 Methanohalophilus genome bin were concatenated together and collectively used as a reference for read mapping. All Utica produced water and input samples were independently mapped against this consolidated Methanohalophilus reference database using bwa (Li and Durbin 2009) version 0.7.12-r1039. The total number of reads satisfying default parameters are displayed in Figure 3.3.

Manual Scaffolding Process

Given that neither of the two Methanohalophilus metagenome bins nor the isolate assemblies yielded a circular closed chromosomal sequence, it is not a surprise that some genes were inevitably split across multiple scaffolds. For many key metabolic genes (see Appendix A), reference protein sequences from Methanohalophilus mahii were used (unless otherwise specified) as tBLASTn queries against the nucleotide scaffold sequences. For many of these genes, the entire amino acid sequence failed to align over its near full length on individual scaffolds. However, different regions of many of these reference sequences did align to multiple scaffolds, collectively yielding near full-length high percentage identity alignments. These multiple candidate scaffolds were brought into Geneious and mapped to the reference nucleotide sequence, only scaffold overlaps with 100% nucleotide identity were considered. From the mapped scaffolds a consensus sequence believed to accurately represent the gene of interest was produced.

Phylogenetic Tree Construction

The 16S rRNA gene sequence tree presented in the introduction was produced by aligning sequences against the mother implementation of the Greengenes reference alignment
(http://www.mothur.org/wiki/Greengenes-formatted_databases). The produced alignment was trimmed using an in-house Python script, which left only the common alignment window. The common alignment window was then used to build a maximum likelihood tree using RAxML (Stamatakis 2014) with a random starting and number seed for parsimony inference with the GAMMA GTR model of evolution. 5000x bootstrap replicates were produced and bootstrap confidence values were shown on the tree with green circles for 60%+ bootstrap support and blue circles for 80%+ bootstrap support.

Protein trees were created by aligning sequences using muscle (Edgar 2004) with default parameters. Again, an in-house Python script was used to trim the alignment down to only the common alignment window. The trimmed alignments were used to create maximum likelihood trees using RAxML, with selection of an evolutionary model using protest 3 (Darriba et al. 2011) through an in-house pipeline (protpipeliner) available @ (https://github.com/rwolfe45/Protpipeliner).

**EMIRGE 16S rRNA Gene Reconstruction**

EMIRGE (Miller et al. 2011) was used on quality trimmed reads that were at least 20 nucleotides in length as inputs and was run for 50 iterations. Resulting sequences were chimera checked using USEARCH (Edgar 2010).
Chapter 4 – Future Directions

While this thesis has aided in further characterizing the genus *Methanohalophilus*, the question of why this group of halophilic methanogens appears more frequently than any other methanogenic genera in hydraulically fractured shales remains largely unanswered. Given that *Methanohalophilus* is the only genus where all sequenced species (more than a single strain) contain the two bifunctional methyltransferases necessary for the de-novo synthesis of glycine betaine from glycine begs the question of whether this molecular system could be in part responsible for the conspicuous occurrence of this phylogenetic group in subsurface saline habitats. The previously documented finding that the activities of these two methyltransferases are effected by K⁺ concentrations makes the hypothesis above even more intriguing since *Methanohalophilus* have been shown to co-accumulate K⁺ along with compatible organic solutes (Lai et al. 1991, Lai and Gunsalus 1992). As an empirical follow up to the largely genomic investigation of shale derived *Methanohalophilus* discussed in this thesis, I propose the following experiments.

1. The intracellular K⁺ and compatible organic solute concentrations in shale derived *Methanohalophilus* are currently unknown. Thus, growing *Methanohalophilus fracturphilus* in triplicate over a gradient of NaCl concentrations representative of those found in the fractured shale environment would provide insight into salinity adaptation mechanisms. The concentration of intracellular K⁺ would be investigated
by inductively coupled argon emission spectroscopy as described previously (Lai et al. 1991). The concentrations of compatible organic solutes would be investigated by both thin layer chromatography (Lai et al. 1999) and separately by nuclear magnetic resonance spectroscopy (Lai et al. 1991). This investigation would, if successful, identify physiologically relevant concentrations of K⁺ ions, and determine both the concentration and relative proportion of compatible organic solutes under numerous salinity regimes.

2. After the physiologically relevant intracellular concentrations of K⁺ have been determined, crude cell extracts would be obtained from mid-log phase *Methanohalophilus fracturphilus* to isolate the glycine sarcosine and sarcosine dimethylglycine methyltransferases. If successful, the isolated proteins would be investigated, if unsuccessful, the follow-up work described below would be performed on crude cell extracts. The methylation activity and binding affinity of the glycine sarcosine and independently, the sarcosine dimethylglycine methyltransferases would be evaluated over a range of K⁺ concentrations. The goal of this work would be to confirm the effect of K⁺ on the activity and binding affinity of these two methyltransferases as previously shown for *Methanohalophilus portucalensis* str. FDF1 (Lai and Lai 2011). From experiment 1, the accumulation of K⁺ along with compatible organic solutes in response to extracellular osmolarity would be determined; experiment #2 would look to identify if K⁺ concentrations affected select protein function in vitro.

3. Finally, using data gathered in experiment 1, I would repeat growth curves over a gradient of select NaCl concentrations for *Methanohalophilus fracturphilus*, but
would preserve biomass in RNA later (Weber et al. 2010). From the RNA later preserved cells, bulk RNA would be extracted, cDNA synthesized (Frias-Lopez et al. 2008), and transcriptomes would be sequenced by Illumina HiSeq. This would investigate correlations in the abundance of transcript levels with changing external salt concentrations. Experiment 2 sought to validate the effects of intracellular K+ concentrations on select enzymatic function; experiment 3 seeks to look at global changes in transcript abundance. As part of this endeavor, genes that appear to have transcript levels modulated by the ionic strength of the media would be assessed to see if they are part of the Methanohalophilus pan genome, or are unique to subsurface Methanohalophilus. It is my hope that a global transcriptomic approach would have the potential to uncover new research targets in Methanohalophilus fracturphilus for further study.
References


20. EIA. (2013)."Technically Recoverable Shale Oil and Shale Gas Resources: An Assessment of 137 Shale Formations in 41 Countries Outside the United States."


Appendix A

See Supplemental File for Alignment Results