An Analysis of Viral Metagenomes in Acetate-fed Anaerobic Reactors

Scott Pease

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2012

Committee:

John Scott Meschke

John Ferguson

Heidi Gough

Program Authorized to Offer Degree:

Environmental Health

University of Washington

Abstract

An Analysis of Viral Metagenomes in Acetate-fed Anaerobic Reactors

Scott Pease

Chair of the Supervisory Committee:
Associate Professor John Scott Meschke
Department of Environmental and Occupational Health Sciences

Anaerobic digesters (ADs) are an important engineered system used to effectively treat wastewater sludge. The ecology of these digesters is dominated by the acetoclastic methanogens, *Methanosaeta spp.* and *Methanosarcina spp*. Though CRISPR sequences have been identified in the genomes of these methanogens, no phages infecting these archaea have been identified. Analysis of the virus size fraction of digester samples by transmission electron microscopy (TEM) revealed the presence of numerous virus-like particles (VLPs), consistent with phages within the family *Siphoviridae*, which contains phages that infect both archaea and bacteria. To further investigate these VLPs, acetate-fed anaerobic reactors were established and samples removed for characterization of the viral size fraction (VSF) by DNA sequencing. Samples from functional ADs were also processed to assess the similarity to the reactor communities. Samples were 0.2 um filtered, concentrated, placed in CsCl density gradient, and DNase treated before DNA extraction. To assess sample purity and diversity, one sample was initially amplified and sequenced. Samples were then sequenced using an Ion Torrent PGM. Sequences were quality trimmed and assembled, and then analyzed using MG-RAST. Taxonomic profiling yielded ~10% of reactor sequences having significant matches, compared to 25-32% of digester sequences, and indicated dominance by the order *Caudovirales*, consistent with TEM images.

Reactor sequence matches to CRISPR spacers of *Methanosaeta concilii* suggest the presence of viruses of acetoclastic methanogens in these systems. Functional profiling indicated an abundance of DNA methylase genes, consistent with a prior study. Taxonomic and functional comparison indicated moderate differences between the hourly fed reactor and other samples. Functional comparison to viral metagenomes from other environments indicated significantly lower metabolic gene frequency in our samples. DNA analysis of the VSFs is consistent with microscopy results that phages are a robust component of the digester communities.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The NSF provided the financial support for this project (Award No. 0930877).

I thank my advisor Scott Meschke and my committee members John Ferguson and Heidi Gough. I thank I-Chieh Chien and David Beck for technical assistance and guidance, and I thank Nicky Beck.

I thank Elizabeth Cooper, Jenna Forsyth, Jesse Billingham, Nicole Van Abel, and Rosebud Beauregard for their friendship and support. I also thank my friends and family for their support.

1. INTRODUCTION

1.1 Background

Investigation of viral communities taken from a wide range of environments has yielded a rapidly expanding understanding of the dominant role that viruses have in ecosystems across the globe. A growing body of evidence indicates that they have a substantial effect on prokaryotic community structure and dynamics. However, there is much that remains to be understood about the function of viruses in ecosystems. Viruses that infect archaea remain particularly poorly characterized, and are considered the least studied component of the biosphere (Rohwer *et al*., 2009). Anaerobic digesters contain a microbial ecosystem in which methanogenic archaea serve critical functions and a minimally characterized, though abundant, viral community. These digester systems occasionally lose functionality for unknown reasons, with characteristics suggesting that the archaeal community was initially disturbed. Considering the major influence that viruses can exert on prokaryotic communities, as well as the minimal characterization of the viruses in anaerobic digesters and of viruses of archaea in general, viral predation of the methanogenic archaea was investigated as a possible explanation for these functionality losses of unknown origin. To investigate this possibility, the viral communities of systems selected to enrich for methanogenic archaea were characterized through metagenomic analysis.

1.2 Viral Metagenomics

Recent advances in molecular biology have enabled a much greater understanding of the microbial world. The development of metagenomics, examining the genetic material taken directly from an environmental sample, independent of any culturing steps, has revealed that culture-based methods of microbial identification had vastly underestimated microbial diversity (e.g. Breitbart *et al*., 2002). Metagenomic analyses can be used to address the questions: Who is there and how much? What are they doing? How does this compare to other areas?

Viral metagenomic analyses indicate that viruses are a driving component in a vast breadth of environments. With an estimated 10^{31} viruses distributed globally, they serve as a major predator for microbes. This predation is thought to exert a significant influence on global biogeochemical cycles and microbial evolution (Rohwer *et al*., 2009). Viruses also appear to harbor many prokaryotic genes. The presence of these genes suggests that they serve as important means of storing and sharing genes among hosts via horizontal gene transfer, thereby influencing global evolution and metabolic processes. The abundance of certain genes may reflect their importance

in the particular environment (Dinsdale *et al*., 2008). Viruses likely tend to use these acquired features to improve their own fitness (Alperovich-Lavy *et al*., 2011). The viral metagenomes of a wide variety of environments have been examined. While marine viruses currently are perhaps the most extensively studied, a significant amount of work has been put into the analysis of communities in other environments, such as solar salterns, rumen, freshwater, and fermented food (Dinsdale *et al*., 2008; Berg Miller *et al*., 2012; Roux *et al*., 2012; Park *et al*., 2010). An analysis has been conducted of an anaerobic digester (Tamaki *et al*., 2011), but it provides little information on specific viruses that may be present or how they may be affecting the system.

The comparison of viromes to identified CRISPRs serves as a potentially valuable means of identifying possible hosts of viruses in a system. While virome analyses have greatly expanded the understanding of the importance of viruses in the world, they are limited by their reliance on previously characterized genomes. It is typical for >70% of a virome to have no significant matches in database searches (Rowher *et al*., 2009). The limitation is perhaps greatest in attempting to successfully identify potential hosts for the viruses. By comparing viromes to CRISPRs, it is possible to gather evidence of the presence of probable hosts for viruses in the system (Garrett *et al*., 2010; Anderson *et al*., 2011).

1.3 Archaeoviruses

As with bacteria and eukaryotes, archaea are subject to predation by a wide array of viruses. Though first reports of viruses infecting archaea emerged several decades ago, they are currently the least studied component of the biosphere (Rohwer *et al*., 2009). There have only been 42 species described, the majority of which have come from geothermal or hypersaline environments, to which they show adaptation (Pina *et al*., 2011). Over a dozen proviruses have also been discovered in archaeal genomes (Krupovic *et al*., 2010). The observation of CRISPR sequences in ~85% of archaeal genomes (Grissa *et al*., 2007) demonstrates that there is much remains to be learned about archaeoviruses.

Much like their hosts, archaeoviruses tend to show unique characteristics. Despite the small number of archaeoviruses described, the morphological diversity of these viruses appears to extend far beyond what has been observed in other viruses. Observed morphotypes include spindle, bottle, bacilliform, droplet, linear, spherical, pleomorphic, and head-tail (Pina *et al*., 2011). However, these unusual morphotypes are largely associated with *Crenarchaeota*; most viruses of *Euryarchaeota* such as methanogens have morphology consistent with *Siphoviridae*

and *Myoviridae* of order *Caudovirales*, which have been shown to infect bacteria (Prangishvili *et al*., 2006). Similarly, the vast majority of archaeovirus sequences show no significant similarity to sequences in databases. Those genes that code for proteins with relatively frequent homologues in archaeoviruses are related to capsids, DNA replication, recombination, and nucleic acid metabolism. Unlike viruses in other domains, the majority of known archaeoviruses establish a state of chronic infection in their hosts, wherein virions are continuously produced while the host remains alive. However, *Caudovirales* viruses have only been observed as lytic or lysogenic (Pina *et al*., 2011).

The archaeal *Caudovirales* viruses that have been observed infect methanogens and halophiles, of the phylum *Euryarchaeota*. Several of these viruses possess unique genomic features, such as methyl groups on all cytosine residues (Vogelsang-Wenke and Oesterhelt, 1988) or a large proportion of plasmid sequence (Meile *et al*., 1989). As with the bacteriophages *Myoviridae* and *Siphoviridae*, there appears to be a high degree of genomic instability in these viruses. In an analysis of the available euryarchaeal virus genomes by PSI-BLAST, ~50-75% of sequences had no significant matches to database sequences, with approximately half of matches being viral (Prangishvili *et al*., 2006). This analysis also indicated that archaeoviruses and bacteriophages share some similar proteins. The similarities between archaeoviruses and bacteriophages are further supported by an analysis of *Caudovirales*-related prophages detected within archaeal genomes. The prophages were found to be highly mosaic, and similarities to database viruses indicate that the proviruses share assembly, packaging, and maturation strategies with bacterial viruses. Moreover, there appears to be some similarity of genome organization with bacterial viruses (Krupovic *et al*., 2010). These studies demonstrate that archaeal *Caudovirales* viruses are relatively similar to bacterial viruses, facilitating their characterization.

1.4 Viruses of Methanogens

Few viruses that infect methanogens have yet been found, none of which infect acetoclasts. There have been four viruses, all *Siphoviridae*, found to infect methanogens, and a fifth, spindle-shaped virus for which infectivity has not been shown. These viruses have all infected *Methanobacteria*, with the spindle-shaped virus associated with *Methanococcus*. Eight prophages have also been found within the genomes of non-acetoclastic methanogens (Krupovic *et al*., 2010), though there are likely more that are yet to be discovered. The presence of CRISPR

sequences in the genomes of over 30 species of methanogen, including several in both *Methanosaeta* and *Methanosarcina* (Grissa *et al*., 2007), points to a history of infection by several viruses.

The observed viruses of methanogens have been described in some detail. The methanogen *Methanococcus voltae* PS produces a gene transfer agent that is phage-like in appearance, but lacks the replication component of a virus. This phage-like particle was found to mediate gene transfer analogous to general transduction, transferring 4400 bp fragments of DNA. The particles were found to resemble *Siphoviridae* by TEM analysis, though with heads only ~40 nm in diameter and tails ~60 nm in length (Eiserling *et al*., 1999). Similarly, psiM1, which infects *Methanobacterium thermoautotrophicum* Marburg, has been shown to mediate generalized transduction (Meile *et al*., 1990). TEM analysis indicated that this phage resembled *Siphoviridae*. This phage was found to have an unusual characteristic in that an estimated roughly 10-15% of phage particles were found to contain not the phage DNA, but rather multimers of the host plasmid pME2001 of a length approximately equal to the phage genome (Meile *et al*., 1989). psiM1 has been shown to mutate into psiM2 under laboratory growth conditions, arising from a 692 bp deletion (Pfister et al. 1998). A defective prophage infecting *Methanothermobacter wolfeii* has been found, called psiM100 (Luo *et al*., 2001). This prophage had ~71% nucleotide sequence identity to the whole genome of psiM2, and was thought to be psiM2-related. Two viruses, phiF1 and phiF3, were found to infect *Methanobacterium* species (Nolling *et al*., 1993). phiF1 was found to have low host specificity, capable of infecting several *Methanobacterium* species, while phiF3 (like psiM1) was only able to infect a single host. Both phages exhibited morphology consistent with *Caudovirales*.

1.5 Sludge

Waste water treatment plants (WWTPs) generate sludge as a by-product of the water purification processes. Recent estimates indicate a generation rate of 60 to 90 g of dry solids per person per day (Appels *et al*., 2008). Sludge undergoes several treatment processes in order to be more suitable for disposal. The primary goals of sludge treatment are to reduce its total volume, thereby minimizing the amount that needs to be processed; to stabilize readily degradable components, which reduces odor issues; and to meet regulatory requirements for safe disposal through the reduction of pathogens. The majority of volume reduction occurs by separating water from the solids through various thickening and dehydrating processes, with additional reduction

occurring in the solids degradation. Stabilization occurs by chemical treatment, as well as via the metabolism of microorganisms, which occurs primarily in anaerobic digesters. Similarly, the majority of pathogen reduction tends to occur in anaerobic digesters. Sludge disposal can represent as much as 50% of the operating costs of WWTPs (Appels *et al*., 2008). As such, maximizing the efficiency and reliability of sludge treatment is of major concern for WWTPs.

1.6 Anaerobic Digesters

Anaerobic digesters act as a critical step in many sludge disposal processes. Successful anaerobic digestion degrades and stabilizes organic materials, thereby reducing potential downstream odor problems while yielding biogas (53-70% v/v methane), and inactivates a significant proportion of the pathogens present (Appels *et al*., 2008). This process requires relatively little energy (Chen *et al*., 2008), and can operate at a high capacity on concentrated, slowly degradable substrates (Ramirez and Steyer 2008). These benefits have resulted in their implementation as an effective and efficient step in sludge disposal under many situations. Despite these benefits, anaerobic digesters do have their limits, such as poor supernatant quality, the presence of undesirable compounds in the biogas, and a poor ability to deal with inorganic materials (Appels *et al*., 2008). Perhaps the largest limitation is the poor operational stability of anaerobic digesters, which may be to be due to the sensitivity of the microbial community (Ramirez and Steyer, 2008). This potential problem has arguably hampered their widespread implementation (Dupla *et al*., 2004). Given the perceived usefulness of anaerobic digesters, improving the understanding of what causes these instabilities is of significant interest. Recently, several upsets have occurred in the state of Washington, including some at a WWTP in King County (Conklin *et al*., 2006) and one at a WWTP in Port Angeles. The upsets are thought to occur due to perturbations in the complex microbial ecology contained within the anaerobic digesters.

1.7 Prokaryotic Community

Anaerobic digesters contain a complex, interdependent prokaryotic community that is able to convert complex macromolecules into simple compounds such as CO_2, CH_4, and H_2. This community is made up of four major functional groups divided over three trophic levels. The major groups present are primary fermenting (acidogenic) bacteria, secondary fermenting (acetigenic) bacteria, acetate-utilizing methanogenic archaea (acetoclastic methanogens), and hydrogen-utilizing methanogenic archaea (hydrogenotrophic methanogens). Primary fermenting

bacteria secrete enzymes to break down complex macromolecules into soluble organic substances (oligomers and monomers) via hydrolysis. They proceed to further ferment these soluble organics into simpler organic compounds, such as fatty acids, succinate, alcohols, and lactate. Acetigenic bacteria consume most of these metabolites, producing acetate, H_2, and CO_2. Acetoclastic methanogens metabolize the acetate into methane, while the hydrogenotrophic methanogens metabolize H_2 and CO_2 into methane. Acetoclastic methanogens, which have been shown to produce ~70% of the methane in anaerobic digesters (Smith and Mah, 1965; Jeris and McCarty, 1965), are comprised of only two genera: *Methanosarcina* and *Methanosaeta*. While *Methanosarcina* are capable of using other energy sources, *Methanosaeta* are only able to use acetate (Liu and Whitman, 2008). The methanogens are arguably major drivers of the community (Briones and Raskin, 2003). The primary source of the substrates for methanogenesis is the acetigenic bacteria, but the primary fermenting bacteria produce some of these compounds as well (Schink, 1997). While these are the dominant prokaryotic groups present in anaerobic reactors, other groups are often present.

Other groups of prokaryotes that can exist in the system rely on different metabolic pathways. One group, homoacetogenic bacteria, is capable of converting hydrogen and single-carbon compounds into acetate (Schink, 1997; Appels *et al*., 2008). Though not typically heavily considered in anaerobic digester functionality, they may be quite influential in some instances (Zhang *et al*., 2009). The homoacetogenic reaction can occur in the other direction as well, though only at high temperatures. In environments containing sufficient sulfate, sulfate reducing bacteria can be significant contributors to the system metabolism, able to utilize all of the products of the primary fermenters (Schink, 1997). Attempts to characterize this diverse prokaryotic community have relied on genetic analyses.

The archaeal community is composed of relatively few species. An analysis of the archaeal community in 44 anaerobic digesters treating various wastes in various set-ups indicated that each digester harbored two to nine distinct archaeal sequences: one dominant acetoclast and several coexisting hydrogenotrophs (Leclerc *et al*., 2004). Four known acetoclasts were detected, *Methanosarcina frisus*, *Methanosaeta concilii*, and two unspecified *Methanosaeta*, with *Methanosaeta concilii* dominated, present in 86% of digesters. The group also found that the type of waste being treated did not affect archaeal community composition, but the process (continuously stirred, fluidized bed, etc.) in the digester did, an observation supported by Zhang

et al., (2009). However, the entire prokaryotic community is much more complex.

The composition of anaerobic digester prokaryotic communities is still being described. A metaanalysis of Sanger sequenced 16S rRNA genes from a range of anaerobic digesters indicated the existence of over 6200 species (5900 bacteria and 300 archaea), with 90% of archaeal and 60% of bacterial diversity described as estimated by comparing numbers of species found to the theoretical asymptotes of rarefaction curves (Nelson *et al*., 2011). Nearly 60% of sequences analyzed were not classifiable into an established genus, showing the high degree of undescribed species diversity present in anaerobic digesters. This metaanalysis indicated that bacteria account for ~85% of sequences and archaea account for ~15%. Predominant bacteria included *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, and *Chloroflexi*, while *Methanosaeta* were the most frequently encountered among archaea, comprising over one third of archaeal sequences. A more thorough analysis of a single anaerobic digester was accomplished by a pyrosequencing metagenomic analysis of 16S rRNA genes. This analysis indicated the presence of over 3200 bacterial phylotypes and ~10 archaeal phylotypes (Zhang *et al*., 2009). This wide range of prokaryotes provides the opportunity for the coexistence of a similarly diverse community of viruses.

1.8 Viral Community

Relatively little is known about viruses in anaerobic digesters. Viral abundance and morphology in anaerobic digesters have been characterized to a minimal extent. Virus-like particles (VLPs) have been found to increase ~10-fold in an upflow anaerobic sludge blanket digester treating brewery waste, with $>10^9$ VLPs L^{-1} in the effluent, as estimated by TEM and fluorescence assay. The most common morphology detected was *Siphoviridae*, with *Myoviridae* and *Cystoviridae* also abundant, though ~70% of VLPs were unclassifiable (Park *et al*., 2007). This study was the first to directly detect and characterize phages coming from an anaerobic digester, and provided the first evidence of their generation within digesters. Similarly, in an analysis of VLP abundance in various compartments in a WWTP, Wu and Liu (2009) found that the concentration of VLPs in anaerobic digestion sludge was ~2.7×10^{10} mL^{-1}, an increase of ~25x the concentration of VLPs in the prior compartment, activated sludge, and the largest of any site sampled. TEM imaging indicated that the morphotypes were consistent between the activated sludge and anaerobic digester. PFGE indicated that the distribution of viral genome sizes in the anaerobic digester was in the range of 35-70 kb. While these studies provide strong

evidence of phage activity in anaerobic digesters, they provided no indication of infectivity or potential hosts.

Recently, the viral metagenomes of samples taken from various points in a WWTP were characterized (Tamaki *et al*., 2011). Sample were taken of the influent, activated sludge, anaerobic digester, and effluent. For the anaerobic digester sample, 523 genotypes were observed, with the most abundant genotype comprising 22.5% of the sample, and 14.8% of sequences significantly matched database sequences, of which 13.7% were hits to viruses and 81.3% were hits to bacteria. This high percentage of hits to bacteria is consistent with studies of other environmental viral metagenomes. Hits to viruses were predominantly in *Caudovirales* (*Siphoviridae* (42.4%), *Myoviridae* (36%), *Podoviridae* (16.2%)). The relative abundances of metabolic functions identified were significantly different from other published environmental metagenomes. While the functional similarity was largely different among samples, a beta diversity analysis indicated >89% similarity between the anaerobic digester and the other sampling sites. Despite the large number of virus species that were identified, there is still no association established between these viruses and the many potential hosts in the system.

1.9 Prokaryotic Community Dynamics

The stability of the anaerobic digester ecosystem is driven by its least stable trophic groups. The primary fermenting bacteria in anaerobic digesters are much more diverse and metabolically flexible than the archaea (Briones and Raskin, 2003; Ariesyady *et al*., 2007; Nelson *et al*., 2011), characteristics driven by the acidogens' consumption of much more energetically favorable food (Fernandez *et al*., 1999; Leclerc *et al*., 2004). These factors likely provide greater acidogenic community stability in the form of functional redundancy, whereas the methanogens have relatively little functional redundancy. These traits are demonstrated in stable laboratory reactors by the constant, seemingly random changes in bacterial species dominance, in constrast to the archaeal communities, which tend to remain relatively constant (Fernandez *et al*., 1999; Zumstein *et al*., 2000; Hori *et al*., 2006). The importance of metabolic diversity has been supported by the observation that a system able to process metabolites in parallel was able to recover faster from a spike in feed than a system relying on metabolism in series (Hashsham *et al*., 2000). Thus, the methanogenic community appears to be a more probable start point for anaerobic digester perturbations.

1.10 Anaerobic Digester Upsets

Losses of functionality in anaerobic digesters, also known as upsets, tend to show a suite of common symptoms and have several potential causes. Typical characteristics of digester upsets include a drop in pH and alkalinity, and increase in volatile fatty acids (VFAs), a drop in methane production, sludge washout, and a shift in the microbial community (Ramirez and Steyer, 2008). A wide variety of inhibitors have been implicated in anaerobic digester upsets. The severity of the upset depends on the type, magnitude, duration, and frequency of the perturbations (Leitao *et al*., 2006). These inhibitors can either be introduced with the sludge or generated within the digester. Perhaps the most common reason for upsets has been organics overloads (Ramirez and Steyer, 2008). Other common inhibitors introduced with the sludge include light metal ions, heavy metals, and organic chemicals. Within the digester, potential inhibitors include ammonia and sulfide, both produced in the degradation process, as well as shifts in pH and temperature (Appels *et al*., 2008; Chen *et al*., 2008). With respect to methanogens, sudden spikes in acetate have corresponded to major shifts in the methanogenic communities of some laboratory-scale anaerobic digesters. Shifts in dominance have been observed from the hydrogenotrophic *Methanobacterium* to the acetoclastic *Methanosaeta* (Delbes *et al*., 2001) and from hydrogenotrophic *Methanoculleus* to acetoclastic *Methanosarcina* and hydrogenotrophic *Methanothermobacter* (Hori *et al*., 2006). While these shifts to the community had probable causes, there are reported instances with no apparent cause.

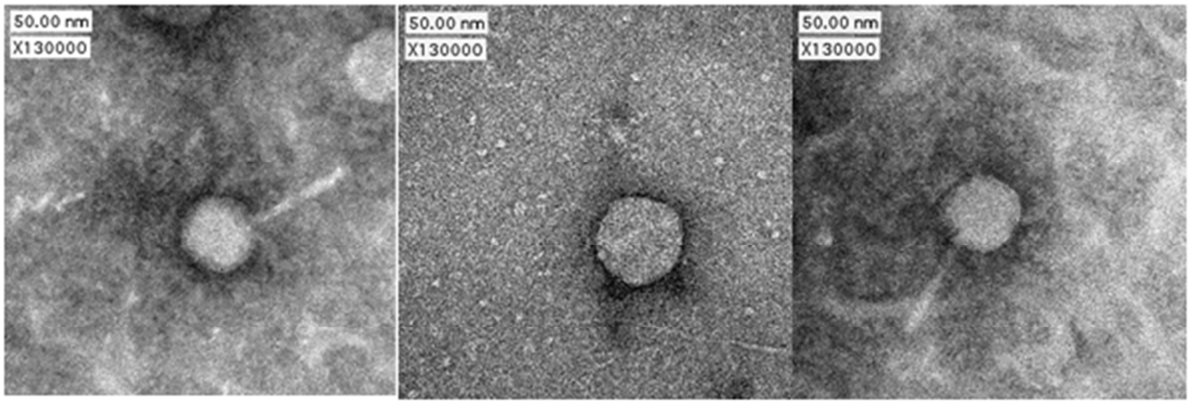
Phage predation may act as a possible mechanism for the shifts in methanogen species dominance in anaerobic digesters have no discernible abiotic cause. In stable laboratory-scale systems, major changes in the archaeal species abundances have been observed, despite consistent abiotic conditions. Shifts in dominance have occurred from one acetoclast to another (Zumstein *et al*., 2000) and from a hydrogenotroph to an acetoclast (Fernandez *et al*., 1999). The consistent abiotic conditions suggest that community interactions brought about the changes in dominance. These observations suggest the possibility of predation by viruses as a potential instigator of shifts in archaeal species dominance in laboratory-scale anaerobic digesters. The possibility that this may occur in full-scale anaerobic digesters is supported by the studies of viruses in anaerobic digesters (Park *et al*., 2008; Wu and Liu, 2009). Furthermore, phages have been observed to regulate microbial abundance and diversity in a membrane bioreactor (Shapiro *et al*., 2010). Given the low diversity of acetoclastic methanogens and the typical characteristics

associated with upset digesters, it seems likely that the imbalance could be initiated by phage predation of this group.

1.11 Study of Acetate-fed Anaerobic Reactors

Anaerobic digesters serve a critical role in many WWTPs. Understanding the mechanisms behind their successful operation will make it easier to address problems that arise in their operation. The observation of unexplained upsets in both laboratory and real-world instances demonstrates the need to investigate possible causes. The emerging realization of the potential for viruses to act as major influences on ecosystems all over the world, combined with the observation of their abundance in anaerobic digesters, suggests that they may play an important role in these systems as well. Acetoclastic methanogens appear to be a likely beginning point for system-wide imbalance, given their low diversity and the observed characteristics of many upsets. As such, we chose to investigate the possibility that viral predation of acetoclastic methanogens could be causing anaerobic digester upsets. To do so, two laboratory-scale anaerobic reactors were established, fed on an acetate-based medium either hourly or daily, and inoculated with samples from WWTP anaerobic digesters. These systems were designed to enrich for acetoclastic methanogens, with the reactor fed daily selecting for *Methanosarcina* and the reactor fed hourly selecting for *Methanosaeta*. However, both systems have become dominated by *Methanosaeta*. TEM image analysis of these systems has demonstrated the presence of numerous VLPs (~5 x 10^7 VLP mL^{-1}), the majority of which appear consistent with the order *Caudovirales* (Figure 1). The purpose of this project was to investigate through metagenomic analyses the viral community that has persisted in these environments.

A.



B.

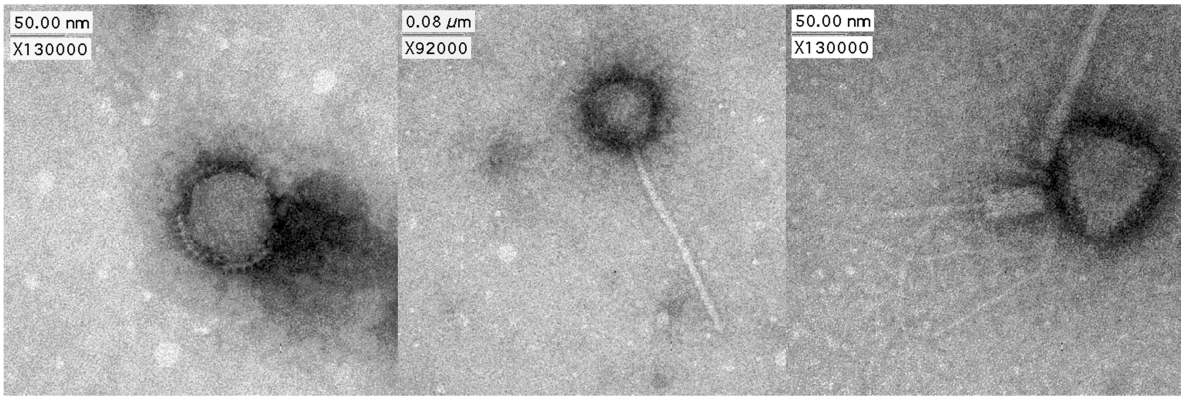


Figure 1. TEM images of VLPs from laboratory anaerobic reactors. Observed VLPs dominated by morphotypes consistent with *Caudovirales*.
A. Daily-fed reactor
B. Hourly-fed reactor

2. MATERIALS AND METHODS

2.1 Sample Description

Anaerobic sludge collected from West Point Treatment Plant in Seattle, WA was used to inoculate two semi-continuous stirred tank reactors in March 2008. The reactors were housed in an environmental chamber at 31-34 °C and stirred continuously. 150 mL of autoclaved, reduced mineral and vitamin medium, containing 234 mM acetate, (Shelton and Tiedje, 1984; Conklin *et al*., 2006) was added to one reactor every 24 h and the other every 1 h (reactor volume: ~2 L). The autoclaved medium was stored at 4 °C and fed into the reactor using Masterflex peristaltic pumps (Cole-Parmer, Chicago, IL) programmed on a timer (ChronTrol Corporation, San Diego, CA). Sterile, anaerobic conditions were maintained during the automated feeding to prevent allochthonous influences on the reactor community. 150 mL samples were taken daily to a total of 900 mL from each reactor in early 2012 for analysis of the viral metagenome. 450 mL samples of anaerobic digester sludge were taken in February 2012 from West Point and South wastewater treatment plants in Seattle, WA.

2.2 Viral Isolation and DNA Extraction

Digester samples were extracted by adding 150 mL of Vertrel XF and 200 mL of PBS (137 NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4), then centrifuging at 6000x g at 4 °C for 20 min. Supernatants were collected and were concentrated by PEG precipitation using 9% w/v PEG 8000 at pH 8.0 and incubating at 4 °C overnight on an orbital shaker at ~250 rpm. Samples were centrifuged at 10000x g at 4 °C for 20 min, and each pellet was resuspended in 5 mL of PBS. Reactor samples were centrifuged at 4000x g at 4 °C for 15 min to pellet prokaryotes and other larger particles. The digester samples and reactor supernatants were sequentially filtered through 0.45-µm and 0.2-µm PVDF and PES syringe membrane filters (Puradisc 25, Whatman) to remove remaining microbial cells (Thurber *et al*., 2009), yielding a virus-sized fraction (VSF). The reactor VSFs were concentrated to ~15 mL using 30 kDa MWCO centrifugal ultrafiltration devices (Amicon Ultra, Millipore). Samples were washed in Amicon 30 kDa devices (Millipore) using 0.02-µm filtered SM buffer (99 mM NaCl, 8 mM $MgSO_4 \cdot 7H_2O$, 0.05 M Tris-Cl (pH 7.5)) and re-concentrating to ~200 µL.

Density gradient centrifugation was performed to further purify the viruses. The samples were re-suspended in 1.15 g/mL 0.02-µm filtered CsCl in SM buffer, loaded on a CsCl step gradient (2 mL each of 0.02-µm filtered CsCl in SM buffer at 1.35, 1.5, and 1.7 g/mL), and centrifuged in

SW-28.1 swinging buckets on an SW-28 rotor at 22000x g at 4 °C for 2 hr (Thurber *et al*., 2009). Fractions were collected in 1 mL aliquots. Viral concentration and purity were assessed by a modification of the protocol developed by Noble and Fuhrman (1998). Subsamples were diluted with 135 mM NaCl, treated with 2% formaldehyde and 10 mM $NaPP_i$, and examined using SYBR Gold staining and epifluorescent microscopy (EFM). The fractions were stored at 4 °C.

The fractions that contained purified VLPs as assessed by EFM were treated with DNase to remove most of the remaining free DNA. Fractions were incubated at 37 °C for 3 hr with 10 U DNase 1 (Sigma-Aldrich, St. Louis, MO) per mL sample (Thurber *et al*., 2009). The sample was washed twice with 0.02-µm filtered SM buffer in an Amicon-30k filter (Millipore), then EDTA was added to a final concentration of 20 mM to inactivate the DNase 1 (Breitbart and Rohwer 2005). DNA was extracted using the High Pure Viral Nucleic Acid Kit (Roche) and stored at -80 °C.

2.3 DNA Amplification and Cloning

DNA concentration was measured by a NanoDrop ND-3300 fluorospectrometer. 1.6 ng of DNA from the reactor fed every 24 h was amplified using the GenomePlex® Complete Whole Genome Amplification (WGA) Kit (Sigma), following the manufacturer's protocol with two exceptions: 17 amplification cycles were run instead of 14, and a 15 min final extension was added. Product yield was measured by fluorospectrometry, and quality was assessed using an Agilent 2100 bioanalyzer.

To assess the practicality of second generation sequencing, sample purity and diversity were assessed by sequencing from a clone library. The sample was cloned using TOPO TA Cloning Kit for subcloning with TOP10 *E. coli* cells and the pCR2.1-TOPO vector (Invitrogen, Carlsbad, CA), following the manufacturer's instructions. 60 colonies were picked and grown up overnight, in 1 mL of LB broth containing 50 ug/mL kanamycin each, in a 96-well deep well plate (Promega, Madison, WI). 10% glycerol stocks were made from the cultures, with 48 submitted for bidirectional sequencing in duplicate, then stored at -80 °C.

2.4 Sequence Analysis

Sequences were processed with Sequencher v. 4.9 (GeneCodes Corporation). Low quality reads were trimmed using default parameters (until the 25 bp on either end contained <4 ambiguous reads and <4 bases of confidence <25), and then remaining cloning vector sequence was trimmed. Consensus sequences of the four reads of each fragment were created using the

Large Gap assembly algorithm, with base call discrepancies adjusted manually, and then were inspected manually to trim the WGA adapter sequences. These consensus sequences were aligned to each other using the Large Gap assembly algorithm, with a minimum match of 80% and a minimum overlap of 20 bp.

Sequences were compared in GenBank to the nucleotide collection (nr/nt) using tBLASTx. Hits with an E-value ≤ 0.001 were considered significant. Sequences were grouped as either phage-derived or non-phage-derived. For sequences with multiple significant hits, the sequence was considered to be phage-derived if phage hits were in the top five (Breitbart *et al*., 2002). Non-phage-derived sequences were further analyzed with a tBLASTx search in the ACLAME database (Leplae *et al*., 2009). Significant hits grouped as non-phage-derived were disregarded if their presence in the system was deemed implausible.

2.5 Ion Torrent Sequencing

Samples were sequenced using an Ion Torrent PGM with a 316 chip and the long reads protocol. Reads were quality filtered using the clip function in Nesoni (v. 58) (Harrison and Seemann, 2011) to trim reads to a length with a minimum quality score of 10 and discard reads <24 bp. Sequences were assembled using Velvet (Zerbino and Birney, 2008), with VelvetOptimiser (v. 2.2.0) (Gladman and Seemann, 2011) used to optimize assembly parameters. Assembled reads were compared to the NCBI GenBank nr/nt and nr databases and to a ribosomal database (Cole *et al*., 2009) to assess sample purity.

2.6 Taxonomic and Functional Analyses

Assembled reads were first analyzed using MG-RAST (Meyer *et al*., 2008). Taxonomic profiles were generated by a BLAT search against the M5NR database (E-value $\leq 10^{-5}$). Taxonomic assignment was used to estimate the alpha diversity in each virome, using the weighted mean of the logarithm of the relative abundances of annotated species. Rarefaction curves were generated for each sample based on matches to the M5NR database (E-value $\leq 10^{-5}$). Functional profiles were generated by a BLAT search against the SEED Subsystems database (E-value $\leq 10^{-5}$). Reactor viromes were further analyzed for the presence of phages or prophages of methanogens by a BLASTX search against the ACLAME database (E-value $\leq 10^{-3}$). Assessment of the most abundant functions annotated was based on BLAT searches against the COG database (E-value $\leq 10^{-5}$). The most abundant COG functions and the level two COG profiles were compared to each other and the anaerobic digester virome from Tamaki *et al*.

(2011). The level two COG profiles were grouped and compared to a collection of publicly available viromes, using normalized data and t-tests with a Bonferroni correction (P-value ≤ 0.002).

2.7 CRISPR Analysis

In order to identify possible hosts of previously undescribed viruses in the acetate-fed reactor viromes, assembled reads were compared to spacers and direct repeats in the CRISPRdb (Grissa *et al*., 2007) using BLASTN (E-value ≤ 0.001). The direct repeat hits were used to ensure that spacer hits were not due to contaminating DNA.

3. RESULTS

3.1 Taxonomic Profile

After quality trimming and assembly, reactor samples contained 1,846-2,481 contigs, each, while anaerobic digester samples contained 36,900-54,031 contigs each, with N50s of 161-319 bp and mean coverage depths of 3.4-13.1 for all samples (Table 1). The percentage of contigs with significant taxonomic matches to the M5NR database (E-value $\leq 10^{-5}$) on MG-RAST was lower for the reactor samples (10.2-10.5%), than for the WWTP samples (25.7-32.1%). In the taxonomic classification of contigs matching the M5NR database, viral matches comprised 4.1-5.8% of samples (0.4-1.9% of all contigs) (Table 2). The taxonomic classification was dominated by matches to bacteria (73.8-82.9%). The reactor samples had a greater percentage of matches to archaea (5.2-7.2%) than the anaerobic digester samples (1-1.5%). Comparison of viromes to 16S rRNA gene databases indicated minimal prokaryotic contamination (0 matches for the reactors, 52 for South WWTP, and 60 for West Point WWTP). Estimates of alpha diversity in each virome were computed using the relative abundances of annotated species in the M5NR database. This estimate indicated 178, 323, 836, and 788 species in the daily, hourly, West Point, and South viromes, respectively. The rarefaction curves indicate that additional sampling may be beneficial in order to better characterize the viral communities of these systems (Figure 2).

Viral sequences were further classified into families using the same search parameters. Both the reactors and the digesters were dominated by the order *Caudovirales* (Figure 3), consistent with TEM image analyses. *Siphoviridae* tended to be the most abundant (36-49%), followed by *Myoviridae* (12-50%) and *Podoviridae* (0-12%). A small number of matches to viruses from other orders was observed for the South WWTP, similar to the anaerobic digester of Tamaki *et al*. (2011). Viral sequences from the reactors were further examined for matches to known methanogen phages or prophages by comparison to the ACLAME database (E-value ≤ 0.001). All matches were to the TerL gene of phage psiM2 of *Methanothermobacterium autotrophicus*. There were five significant matches for the daily fed reactor, three of which were the top hit. The hourly fed reactor had one significant match, which was not the top.

Table 1. Sequence and contig statistics.

	Daily Fed Reactor	Hourly Fed Reactor	West Point WWTP	South WWTP
No. of raw sequences	572,672	2,554,267	2,578,357	2,074,807
Mean raw read length (bp)	215	135	149	233
No. of filtered sequences	453,538	2,121,452	2,146,160	2,008,500
Mean filtered length (SD)	*62 (35)*	84 (43)	91 (48)	110 (54)
No. of contigs	1,846	2,481	54,031	36,900
Mean coverage depth	6.7	13.1	3.4	6.1
Contig N50	316	161	202	319
Total bases in contigs	571,331	458,033	12,945,497	11,185,187
Longest contig	3,425	1,744	3,628	6,943
No of contigs >1000 bp	37	9	141	689
Bases in contigs >1000 bp	61,759	11,414	184,502	1,019,123

Table 2. Functional and taxonomic assignment percentages and distribution of taxonomic assignments

	Daily Fed Reactor	Hourly Fed Reactor	West Point WWTP	South WWTP	Tamaki et al. (2011)
Functional assignment[a]					
Unknown (%)	98.1	96.8	95.8	94.6	95.1
Known (%)	1.9	3.2	4.2	5.4	4.9
Taxonomic assignment[b]					
Unknown (%)	89.5	89.8	74.3	67.9	73.7
Known (%)	10.5	10.2	25.7	32.1	26.3
Taxonomic classification of known[b]					
Viruses (%)	4.1	5.6	5.4	5.8	5.3
Bacteria (%)	80.4	73.8	78.8	82.9	81.8
Archaea (%)	7.2	5.2	1.5	1	2
Eukaryotes (%)	0	2	0.1	0.4	1.1
Unassigned/Unclassified/ Other (%)	8.2	13.5	14.1	9.9	9.7

a. Based on BLAT to the SEED Subsystems database (E < 1e-5).
b. Based on BLAT to the M5NR database (E < 1e-5).

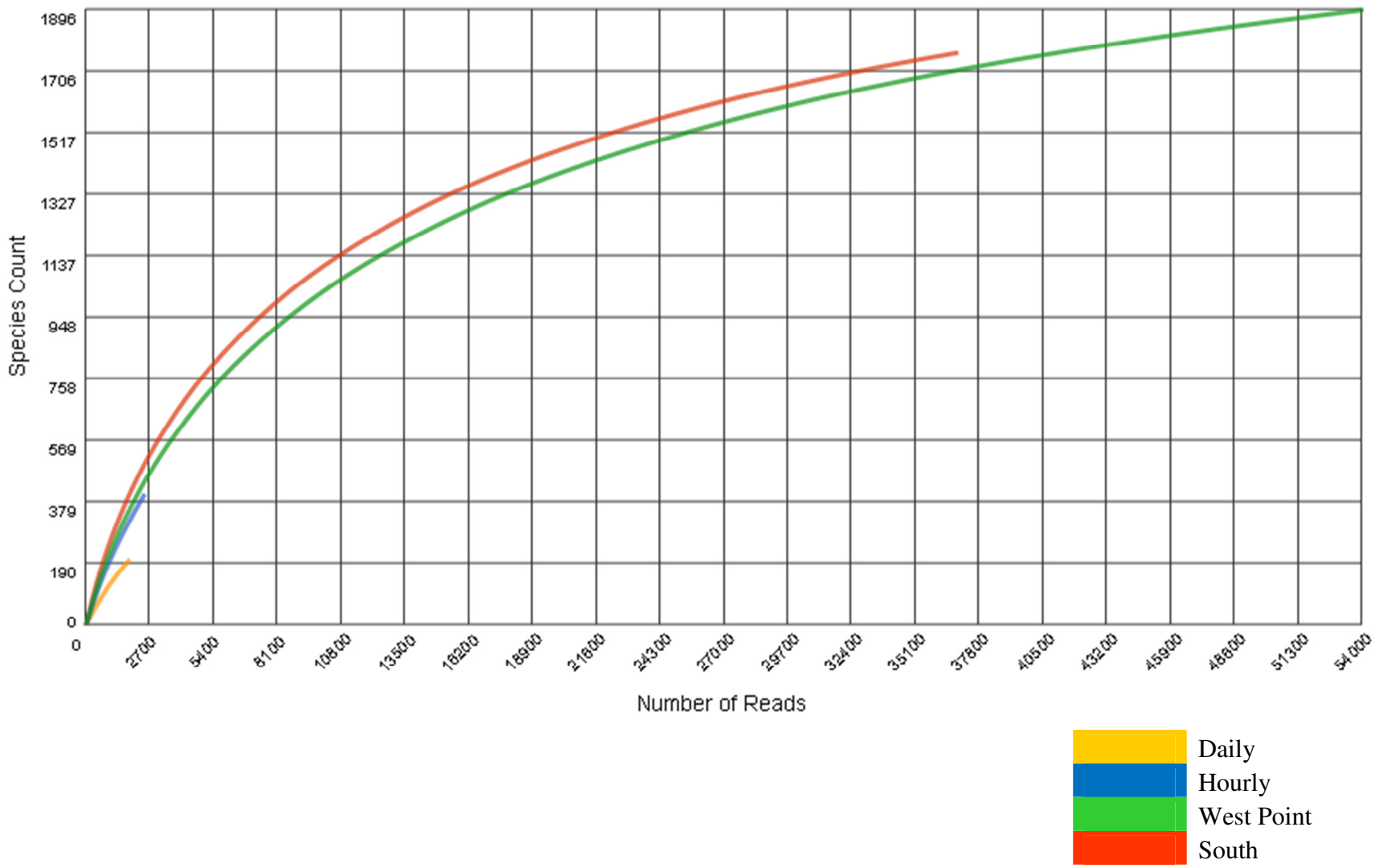


Figure 2. Rarefaction curves of assembled samples based on BLAT to M5NR database (E-value $\leq 10^{-5}$)

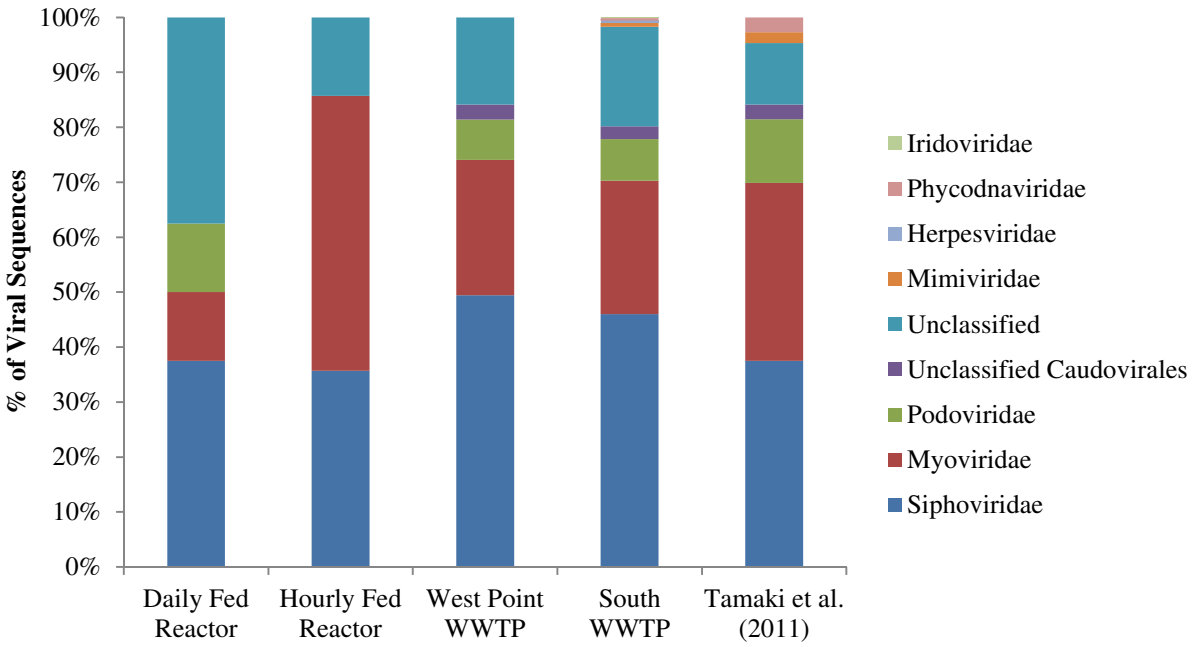


Figure 3. Comparison of virus family abundance based on BLAT to the M5NR database (E < 1e-5)

3.2 Functional Profile

A comparison of the assembled viromes to the SEED subsystems database on MG-RAST (E-value $\leq 10^{-5}$) provided functional annotation for 1.9-5.4% of contigs (Table 1). The COG database was searched through MG-RAST as well (E-value $\leq 10^{-5}$). COG functions related to DNA methylation were among the most abundant, with DNA modification methylase (COG0863) accounting for 6.5-14.7% of reactor COGs and 25.0-25.7% of anaerobic digester COGs. Transposase and inactivated derivatives (COG0675) were the most abundant in the reactor viromes, comprising 17.6-22.0% of assigned COGs. Aside from phage-related COGs, other abundant COGs were related to transcriptional regulation (COG1475) and replication, recombination and repair, such as single-stranded DNA-binding proteins (COG0629) and helicases (COG0305; COG0553).

3.3 Functional Comparisons

A comparison of COG classifications (E-value $\leq 10^{-5}$) among the reactors and anaerobic digesters indicates that they have similar distributions of functions (Figure 4, Figure 5), with the hourly fed reactor sample appearing to be slightly divergent from the other samples. The hourly fed sample appears to have more genes related to nucleotide and carbohydrate transport and

metabolism and fewer related to replication, recombination, and repair. In comparing the collective COG classifications of the reactors and digesters from this study with 42 other viromes, there are several significant differences (Figure 6). The reactors and digesters had significantly lower abundance of genes related to signal transduction, cytoskeleton, post-translational modification, lipid transport and metabolism, inorganic ion transport and metabolism, secondary metabolite synthesis, amino acid transport and metabolism, and energy production and conversion (P-value ≤ 0.002).

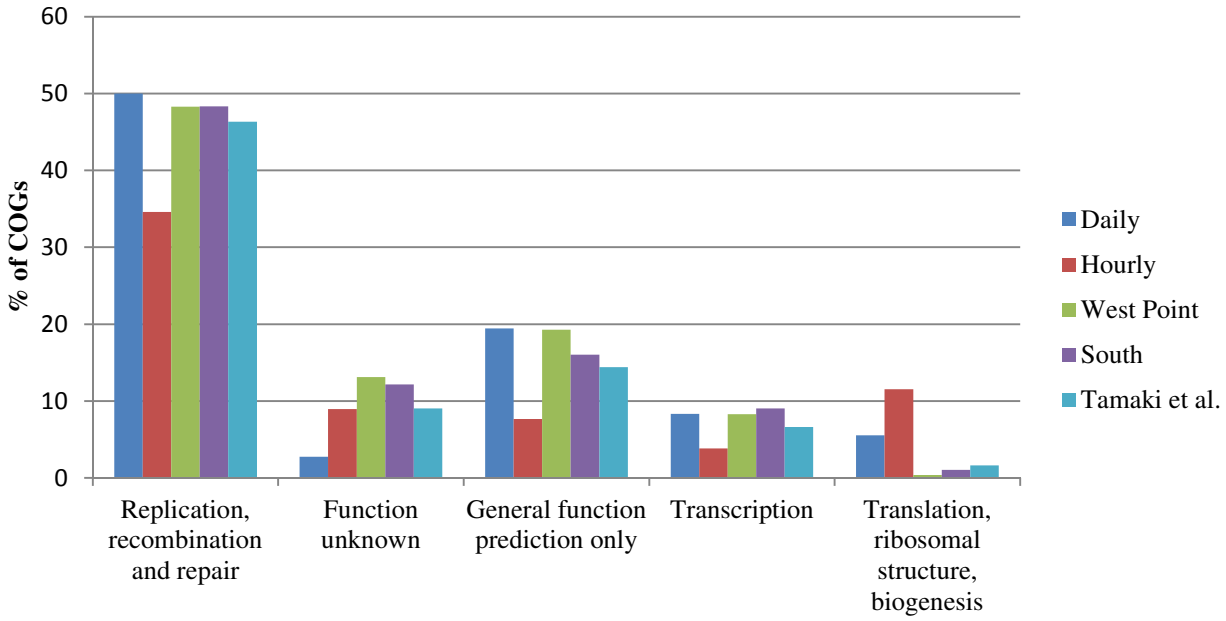


Figure 4. Comparison of relatively abundant categories of functional profile among the samples from this study and the anaerobic digester of Tamaki *et al*. (2011) based on level two COG assignments (E-value $\leq 10^{-5}$).

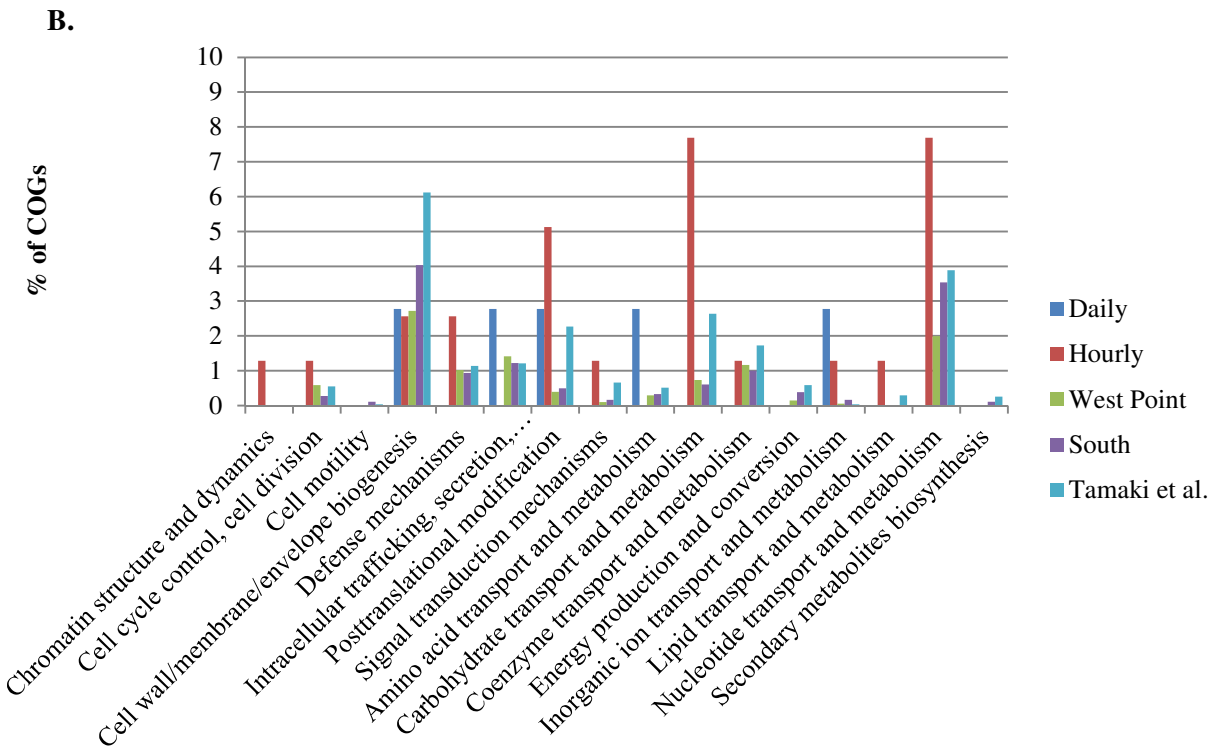


Figure 5. Comparison of relatively less abundant functional profile among the samples from this study and the anaerobic digester of Tamaki *et al*. (2011) based on level two COG assignments (E-value $\leq 10^{-5}$).

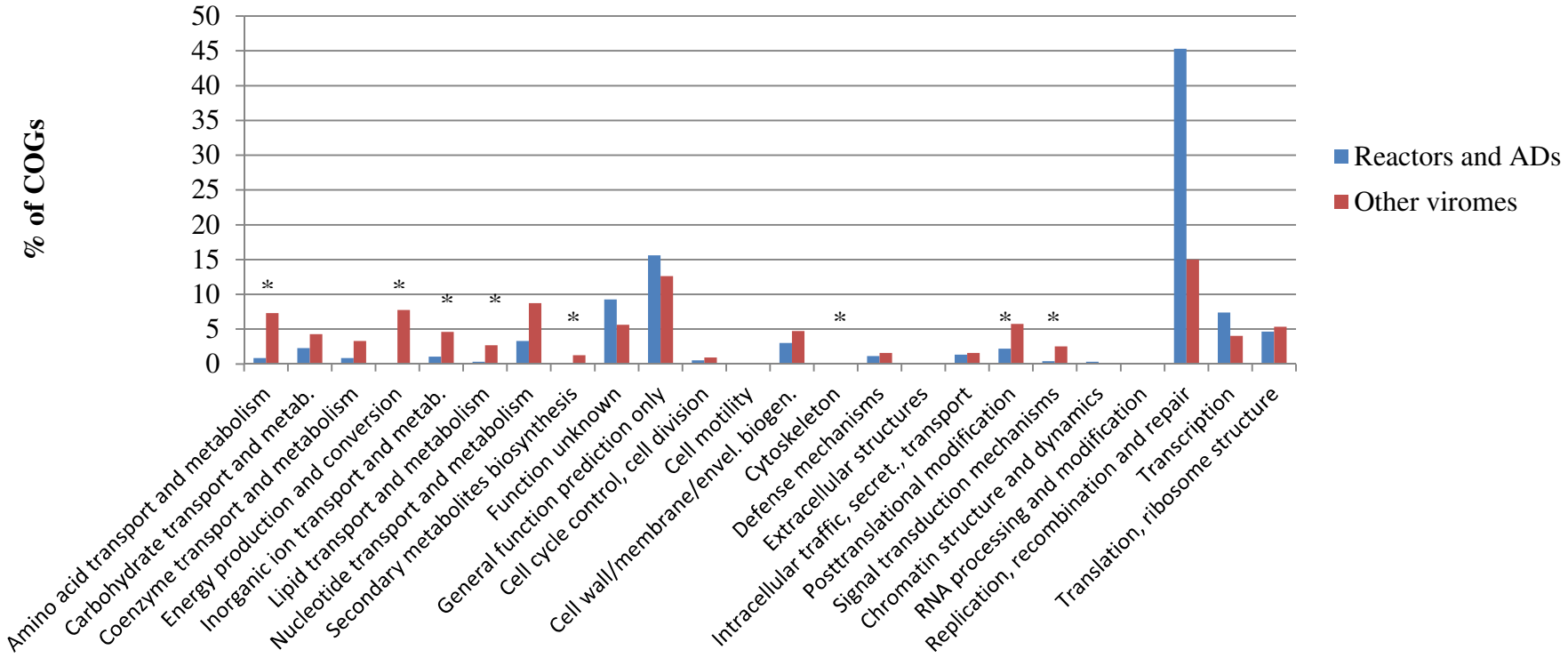


Figure 6. Comparison of functional profile among the samples from this study and a collection of publicly available metagenomes based on level two COG assignments (E-value $\leq 10^{-5}$). Asterisk denotes statistical significance on normalized data (P-value ≤ 0.002).

3.4 CRISPR Spacer Matches to Viromes

CRISPRs in archaea and bacteria serve as indications of viral infection history. A BLASTN comparison (E-value ≤ 0.001) of the anaerobic reactor viromes to the CRISPRdb, containing 2669 identified CRISPRs in 1700 archaea and bacteria, was conducted in order to identify unknown viruses by matches to spacers from prokaryotic species either known or likely to be in the reactors. The significance criteria selected appeared to be sufficiently selective to omit most matches to species unlikely to be in the system while still retaining most matches to species likely to be in the system. The hourly-fed reactor had 14 significant matches to spacers. Of these, three were to *Methanosaeta concilii*, three were to syntrophic bacteria, and two were to *Methanococcales* found in marine geothermal sediment. The daily-fed reactor had 27 significant matches to spacers. Of these, twenty were to *M. concilii*, two of which had 100% identity across the entire spacer; one was to syntrophic bacteria; and two were to other sludge species (Table 3). To ensure that these matches were not derived from contaminating sequences, the viromes were compared to the direct repeats using the same parameters. The hourly-fed and daily-fed viromes had one and two matches, respectively, none of which were plausible.

Table 3. Significant matches to spacers in CRISPRdb by BLASTN (E-value ≤ 0.001)

Hourly Fed	
Species Hit	**Number of Hits**
Methanosaeta concilii GP-6	3
Syntrophic species	3
Methanothermococcus okinawensis IH1	1
Methanocaldococcus vulcanius M7	1
Other	6
Daily Fed	
Species Hit	**Number of Hits**
Methanosaeta concilii GP-6	20
Syntrophic species	1
Other sludge species	2
Other	4

4. DISCUSSION

The alpha diversity estimates were greater for the anaerobic digesters than for the reactors. This difference in diversity reflects the differences in energy and nutrient availability between these systems. The anaerobic digesters have substantial energy and nutrient sources, enabling the coexistence of a wide range of microbes. The reactors, fed on acetate medium, were designed to select for a subset of the anaerobic digester community. The lower viral diversity in the reactors reflects the restricted number of host species. The alpha diversity estimates may be artificially high, based on comparison of the results from the MG-RAST calculation to those generated by PHACCS (Angly *et al*., 2005), a tool specifically designed for viral metagenome diversity analysis, for the WWTP viromes of Tamaki *et al*. (2011). However, the relative estimated diversity among the samples should be similar using either method. The rarefaction curves of the assembled reads (Figure 2) indicate that more sequence data may be beneficial to more fully characterize the viral communities in these systems, particularly for the reactors. Similarly, the coverage depth across all samples was relatively low, ranging from 3.4 to 13.1. Increasing the coverage depth by gathering more data could improve the quality and length of the assemblies by providing stronger evidence for the correct bases where differences in base reads arise, reducing the probability of forming chimaeras, and providing more sequences with which to extend contigs.

The majority of sequences from the anaerobic digesters and laboratory reactors had no significant similarity to M5NR database sequences. This result is consistent with other virome studies, which typically demonstrate ~70% or more of sequences having no significant matches to database sequences, and may reflect uncharacterized viral diversity and demonstrates the paucity of current knowledge of viruses in the environment. This unknown fraction may predominantly be of viral origin (Edwards and Rohwer, 2005). Virus-specific annotation pipelines have been shown to yield more functional annotations for viromes compared to MG-RAST, which bases its functional classification on bacterial and archaeal sequences (Lorenzi *et al*., 2011). This result supports the idea that a portion of the unknown fractions likely contain viral sequences.

The smaller proportion of sequences in the reactor viromes, compared to the anaerobic digester viromes, that had significant taxonomic matches to database sequences reflects that the reactors have been enriched for a relatively unexplored community. The two anaerobic digesters

sampled and that of Tamaki *et al*. (2011) had similar percentages of taxonomic assignment (Table 2) and were comparable to many other virome studies (Mokili *et al*., 2012). This similarity suggests that the anaerobic digester systems harbor viral communities that are relatively similar to many other environments. The isolation of the numerically non-dominant component of the broader community may present a useful means of assessing more fully the uniqueness of viruses in different environments. These species are minimally represented in current viral metagenomes, which may mask their actual diversity and relative abundances. Further investigation of systems similar to the reactors would be of interest in assessing the uniqueness of this relatively unknown community.

The taxonomic classification of all viromes was dominated by matches to bacteria. Similarly, all viromes had significant matches to archaea, with the reactor samples having relatively large percentages of archaeal matches. Sequences classified as viral made up only a small percentage of each sample (Table 2). These results are consistent with prior studies (Tamaki *et al*., 2011; Anderson *et al*., 2011; Berg Miller *et al*., 2012). Analysis of the 16S rRNA gene content indicated no detectable contamination in the reactor samples and minimal contamination in the anaerobic digester samples. Thus, some of the sequences in the anaerobic digester samples may reflect actual prokaryotic sequences. Though no contamination was detected in either reactor sample, the imperfect nature of the viral purification process may have resulted in undetected prokaryotic DNA in these viromes. However, the abundance of matches to prokaryotes may also reflect the bias of many current databases due to having many more non-viral than viral sequences, unidentified prophages in these databases, and horizontal gene transfer between viruses and their hosts (Lorenzi *et al*., 2011; Krupovic *et al*., 2010; Dinsdale *et al*., 2008). As the contamination of the reactor samples appears to be minimal, the relatively large percentage of matches to archaea may reflect some combination of unidentified prophages and horizontal gene transfer, suggesting that these systems may have enriched for archaeoviruses.

Sequences identified as viral were classified predominantly into families of the order *Caudovirales*. *Siphoviridae* was the most frequently matched family in the daily fed reactor and both WWTP samples, while *Myoviridae* was the most frequently matched family for the hourly fed reactor. These classifications are supported by TEM images (Figure 1). If the classification is representative of the unclassified majority of the viromes, then these systems are dominated by *Caudovirales*. Family-level abundances in the reactors should be interpreted with caution given

the small number of sequences classified as viral for the daily fed and hourly fed reactors (8 and 14, respectively). The profiles of both WWTP samples appear similar to the anaerobic digester virome of Tamaki *et al*. (2011). The classification of viruses into families with hosts that are not present in WWTP systems could reflect biases in current databases due to differences in the extents to which virus groups have been studied, or could reflect viruses that have entered and persisted in the WWTPs. The identification of *Mimiviridae* suggests database bias as a more likely cause, as these viruses should have been filtered out during sample processing.

Further investigation of the reactor samples specifically for phages of methanogens by comparison to the ACLAME database yielded several matches. The only matched gene was the large terminase subunit (TerL) of phage psiM2 of *Methanothermobacterium autotrophicus*, or the largely similar prophage psiM100 of *Methanothermobacter wolfeii*, thermophilic anaerobic digester hydrogenotrophic methanogens (Liu and Whitman, 2008). The temperature of the reactors (31-34°C) is outside of the optimum range for *Methanothermobacter* spp., suggesting that these species are unlikely to be found in the reactors. The TerL is relatively conserved in the order *Caudovirales* and is used as a marker gene for phylogenetic classification (Roux *et al*., 2011). As such, matches to this TerL gene suggest the presence of at least one virus population similar to psiM2 and psiM100. Tailed archaeal proviruses appear to group based on gene content in a manner similar to the hosts' taxonomic grouping, indicating possible coevolution (Krupovic *et al*., 2010). Thus, the presence of these similar viruses suggests the corresponding presence of a host similar to *Methanothermobacter* spp. These matches indicate the possible presence of viruses of a methanogen in the reactor systems.

A level two COG comparison of the viromes in this study to each other and to the anaerobic digester of Tamaki *et al*. (2011) indicated that the distribution of functional abundance is similar across these viromes, though the hourly fed reactor appears to be somewhat divergent (Figure 4, Figure 5). This divergence may be an artifact of the relatively small number of sequences assigned to COGs. Metagenomes, both microbial and viral, from different environments have been shown to have metabolic profiles that reflect the perceived relative importance of various functions in each environment. These profiles have been proposed to be independent of differences in taxonomic profile at different locations within a given environment (Dinsdale *et al*., 2008). The similar functional profiles observed across the anaerobic digesters and daily fed reactor suggests that these systems exert similar demands on the viral communities. This

agreement may indicate that the daily fed reactor has successfully maintained conditions that demand functionality that is similar to what was required in the original anaerobic digester system, suggesting that the reactor community may be a realistic representation of the initial community that was isolated. The apparent divergence of the hourly fed reactor, if real, could indicate that this system has become relatively artificial in comparison to its source environment. The larger proportion of COGs related to carbohydrate and nucleotide transport and metabolism and smaller proportion related to replication, recombination, and repair in this reactor suggests that it may present different functional requirements on the microbial community. It has been hypothesized that decreased abundance of metabolic genes and increased abundance of replication, recombination, and repair genes may be reflective of the viral survival strategy in environments with abundant nutrient and carbon supplies (Tamaki *et al*., 2011). This hypothesis is supported in this study by the statistical comparison of the metabolic profiles of the reactor and anaerobic digester viromes to a collection of 42 other viromes, which showed significantly lower abundances of several categories of metabolic genes (Figure 6) in the reactor and digester viromes. Further support was generated in a study of cow rumen viromes (Berg Miller *et al*., 2012), which showed similar patterns of significantly lowered metabolic gene abundances. The hourly fed reactor receives aliquots of acetate medium that result in relatively small concentration increases. This less concentrated food supply may exert stronger selective pressure related to microbial metabolism, leading to a corresponding decrease in replication, recombination, are repair genes, compared to the other systems.

The most abundant COG functions were generally similar across samples, and showed an abundance of DNA methylase genes. The five most abundant COGs were identical and in the same order between the South and West Point viromes and coincided closely with those of the Tamaki *et al*. (2011) anaerobic digester, further supporting that there is a high degree of functional similarity among anaerobic digesters. As with the level two COG profiles, the daily fed reactor shared a similar profile of most abundant COGs, while the hourly fed reactor appeared to be relatively divergent, only sharing the observed high proportion of DNA methylases. The abundance of methylase genes agrees with prior studies of anaerobic digestion environments (Tamaki *et al*., 2011; Berg Miller *et al*., 2012). The high proportion of these genes was found in all examined compartments of the WWTP of Tamaki *et al*. (2011). Methylation of phage DNA may act as a defense mechanism against restriction digestion by prokaryotic

endonucleases (Kruger and Bickle, 1983). Alternatively, these DNA methylation genes could be acting upon the host DNA, acting to manipulate gene expression to enhance their fitness. Phages in an enhanced biological phosphorus removal system have been found to contain a gene encoding H-NS, a repressor protein. It was proposed that the phages were using this protein to repress host defense mechanisms, such as CRISPRs (Skennerton *et al*., 2011). The DNA methylase genes may be acting in a similar manner.

The presence of multiple CRISPR sequences in many methanogens points to a history of infection by viruses. The majority of spacer matches in the daily fed virome and several in the hourly fed virome were to spacers from *Methanosaeta concilii* (Table 3). These matches provide evidence for the presence of one or more viruses targeting this acetoclastic methanogen. The dominance of *Methanosaeta* in these reactors further supports this possibility. The large number of matches for the daily fed virome to spacers from *M. concilii* may reflect that these viral sequences dominate the system along with their potential host. This possibility is supported by the fact that several of the contigs that matched the *M. concilii* were among the longest and most deeply covered contigs (data not shown). Analysis of the species richness and evenness in these samples would be beneficial in further assessing this possibility. However, the associated viruses could have other hosts as well. For instance, the virus phiF1 has been shown to target several *Methanobacterium* species (Nolling and Groffen, 1993). If an alternative host is in this system, then it would likely be similar to *M. concilii* (Krupovic *et al*., 2010). Alternatively, the spacers may have been transferred to another species (Brodt *et al*., 2011), though this event rarely occurs across genera. Analysis of the bacterial 16S rRNA genes isolated from these reactors yielded few matches to identified species, none of which are present in the CRISPRdb (data not shown). The smaller number of matches to CRISPR sequences in the hourly fed reactor could indicate that this viral community targets a larger proportion of these undescribed hosts. The presence of a relatively small proportion of sequences matching *M. concilii* spacers, in spite of being dominated by this methanogen like the daily fed reactor, could reflect that this system has greater species evenness, or could be an artifact of community dynamics, reflecting a relatively low point in the abundance of these viral sequences. Alternatively, it could reflect a higher rate of genomic coevolution with host species due to intense selective pressure from CRISPRs (Weinberger *et al*., 2012).

The presence of sequences matching spacers from syntrophic bacteria suggests that the reactors have maintained this component of the original anaerobic digester community. The presence of syntrophic bacteria is supported by 16S rRNA gene analysis (data not shown). The presence of syntrophic bacteria suggests the possible presence of hydrogenotrophic methanogens or homoacetogens as well. The presence of hydrogenotrophic methanogens may explain the matches to the TerL gene of psiM2 as well as the matches to spacers of *Methanothermococcus* and *Methanocaldococcus* the hourly fed reactor. However, as these species inhabit marine geothermal sediment, these matches may be false positives. The observed diversity of the viral community and the apparent diversity of hosts indicates that multiple functional groups from the original anaerobic digester system have established stable populations in the reactors.

This study demonstrated that phages are an abundant component of the acetate-fed anaerobic community. Evidence was generated for the presence of phages of acetoclastic methanogens in anaerobic digesters. The reason for the apparent disparity in functional and taxonomic profile of the hourly fed reactor is uncertain. Further investigation of anaerobic digesters is needed in order to better explain these differences. The relatively small proportion of reactor virome sequences that matched characterized genes demonstrates the need for further study of phages that target less common species like the archaea. The ability to link unidentified viruses to hosts is an important aspect of describing viral communities that is currently limited in scope. Isolating selected portions of natural microbial communities facilitates the connection of viruses to their respective hosts. As the microbial and viral communities of more environments are explored, patterns in functional profiles and the abundance of specific gene groups, such as DNA methylases, will become more apparent. The samples analyzed in this study represent non-replicated communities and a viral community that for the digesters was a grab sample and for the reactors was averaged over the course of several days, providing no indication of the possible variability or dynamics of these communities. More work is needed in order to understand these aspects of the relationship between viruses and both the methanogenic community and the overall anaerobic digester. There is still much that remains to be described about the viruses in anaerobic digester systems. Understanding the function of these viruses may lead to future improvements in the stability of anaerobic digesters.

5. REFERENCES

Alperovitch-Lavy, A., Sharon, I., Rohwer, F., Aro, E.M., Glaser, F., Milo, R., *et al*. (2011) Reconstructing a puzzle: existence of cyanophages containing both photosystem-I and photosystem-II gene suites inferred from oceanic metagenomic datasets. *Environ Microbiol* **13**: 24-32.

Anderson, R.E., Brazelton, W.J., and Baross, J.A. (2011) Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *FEMS Microbiol Ecol* **77**: 120-133.

Angly, F., Rodriguez-Brito, B., Bangor, D., McNairnie, P., Breitbart, M., Salamon, P., *et al*. (2005) PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. *BMC Bioinformatics* **6**: 41.

Appels L., Baeyens, J., Degreve, J., and Dewil, R. (2008) Principles and potential of the anaerobic digestion of waste-activated sludge. *Prog Energy Combust Sci* **34**: 755-781.

Ariesyady, H.D., Ito, T., and Okabe, S. 2007. Functional bacterial and archaeal community structures of major trophic groups in a full-scale anaerobic sludge digester. *Was Res* **41**: 1554-1568.

Berg Miller, M.E., Yeoman, C.J., Chia, N., Tringe, S.G., Angly, F.E., Edwards, R.A., *et al*. (2012) Phage-bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. *Environ Microbiol* **14**: 207-227.

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., *et al.* (2002) Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci USA* **99**: 14250-14255.

Breitbart, M., and Rohwer, F. (2005) Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. *BioTechniques* **39**: 729-736.

Briones, A., and Raskin, L. (2003) Diversity and dynamics of microbial communities in engineered environments and their implications for process stability. *Curr Opin Biotechnol* **14**: 270-276.

Brodt, A., Lurie-Weinberger, M.R., and Gophna, U. (2011) CRISPR loci reveal networks of gene exchange in archaea. *Biol Direct* **6**: 65.

Chen Y., Cheng, J.J., and Creamer, K.S. (2007) Inhibition of anaerobic digestion process: a review. *Bioresource Tech* **99**: 4044-4064.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris R.J., *et al*. (2009) The ribosomal database project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**: D141-D145.

Conklin, A., Stensel, H.D., and Ferguson, J. (2006) Growth kinetics and competition between *Methanosarcina* and *Methanosaeta* in mesophilic anaerobic digestion. *Wat Environ Res* **78**: 486-496.

Delbes, C., Moletta, R., and Godon, J-J. (2001) Bacterial and archaeal 16S rDNA and 16sS rRNA dynamics during an acetate crisis in an anaerobic digestor ecosystem. *FEMS Microbiol Ecol* **35**: 19-26.

Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., *et al*. (2008) Functional metagenomic profiling of nine biomes. *Nature* **453**: 629-632.

Dupla, M., Conte, T., Buoview, J.C., Bernet, N., and Steyer, J.P. (2004) Dynamic evaluation of a fixed bed anaerobic-digestion process in response to organic overloads and toxicant shock loads. *Wat Sci Technol* **49**: 61-68.

Edwards, R.A., and Rohwer, F. (2005) Viral metagenomics. *Nat Rev Microbiol* **3**: 504-510.
Eiserling, F., Pushkin, A., Gingery, M., and Bertani, G. (1999) Bacteriophage-like particles associated with the gene transfer agent of *Methanococcus voltae* PS. *J Gen Virol* **80**: 3305-3308.
Fernandez, A., Huang, S., Seston, S., Zing, J., Hickey, R., Criddle, C., and Tiedge, J. (1999) How stable is stable? Function versus community composition. *Appl Environ MIcrobiol* **65**: 3697-3704.
Garrett, R.A., Prangishvili, D., Shah, S.A., Reuter, M., Stetter, K.O., and Peng, X. (2010) Metagenomic analyses of novel viruses and plasmids from a cultured environmental sample of hyperthermophilic neutrophiles. *Environ Microbiol* **12**: 2918-2930.
Gladman, S., and Seemann, T. (2011) VelvetOptimizer. Victorian Bioinformatics Consortium, Monash University
Grissa, I., Vergnaud, G., and Pourcel, C. (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**: 172.
Harrison, P., and Seemann, T. (2011) From high-throughput sequencing read alignments to confident, biologically relevant conclusions with Nesoni. Victorian Bioinformatics Consortium, Monash University.
Hashsham, S.A., Fernandes, A.S., Dollhopf, S.L., Dazzo, F.B., Hickey, R.F., Tiedje, J.M., and Criddle, C.S. (2000) Parallel processing of substrate correlates with greater functional stability in methanogenic bioreactor communities perturbed by glucose. *Appl Environ Microbiol* **66**: 4050-4057.
Hori, T., Haruta, S., Ueno, Y., Ishii, M., and Igarashi, Y. (2006) Dynamic transition of a methanogenic population in response to the concentration of volatile fatty acids in a thermophilic anaerobic digester. *Appl Environ Microbiol* **72**: 1623-1630.
Jeris, J.S., and McCarty, P.L. (1965) The biochemistry of methane fermentation using C-14 tracers. *Wat Environ Fed* **37**: 178-192.
Kruger, D.H., and Bickle, T.A. (1983) Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol Rev* **47**: 345-360.
Krupovic, M., Forterre, P., and Bamford, D.H. (2010) Comparative analysis of the mosaic genomes of tailed archaeal viruses and proviruses suggests common themes of virion architecture and assembly with tailed viruses of bacteria. *J Mol Biol* **39**: 144-160.
Leclerc, J., Delgenes, J-P., and Godon, J-J. (2004) Diversity of the archaeal community in 44 anaerobic digesters as determined by single strand conformation polymorphism analysis and 16S rDNA sequencing. *Environ Microbiol* **6**: 809-819.
Leitao, R.C., van Haandel, A.C., Zeeman, G., and Lettinga, G. (2006) The effects of operational and environmental variations on anaerobic wastewater treatment systems: a review. *Bioresource Technol* **97**: 1105-1118.
Leplae, R., Lima-Mendez, G., and Toussaint, A. (2009) ACLAME: A classification of mobile genetic elements, update 2010. *Nucleic Acid Res* **38**: D57-D61.
Liu, Y., and Whitman, W.B. (2008) Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann NY Acad Sci* **1125**: 171-189.
Lorenzi, H.A., Hoover, J., Inman, J., Safford, T., Murphy, S,. Kagan, L., and Williamson, S.J. (2011) Theviral metagenome annotation pipeline (VMGAP): an automated tool for the functional annotation of viral metagenomic shotgun sequencing data. *Standards Genomic Sci* **4**: 418-429.

Luo Y., Pfister, P., Leisinger, T., and Wasserfallen, A. (2001) The genome o archaeal prophage psiM100 encodes the lytic enzyme responsible for autolysis of *Methanothermobacter wolfeii*. *J Bacteriol* **183**: 5788-5792.

Meile L., Jenal, U., Studer, D., Jordan, M., and Leisinger, T. (1989) Characterization of psiM1, a virulent phage of *Methanobacterium thermoautotrophicum* Marburg. *Arch Microbiol* **152**: 105-110.

Meile L., Abendschein, P., and Leisinger, T. (1990) Transduction in the archaebacterium *Methanobacterium thermoautotrophicum* Marburg. *J Bacteriol* **172**: 3507-3508.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M. *et al*. (2008) The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.

Nelson, M.C., Morrison, M., and Yu, Z. (2011) A meta-analysis of the microbial diversity observed in anaerobic digesters. *Bioresource Technol* **102**: 3730-3739.

Noble, R.T., and Fuhrman, J.A. (1998) Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquat Microb Ecol* **14**: 113-118.

Nolling, J., Groffen, A., and de Vos, W.M. (1993) phiF1 and phiF3, two novel virulent, archaeal phages infecting different thermophilic strains of the genus *Methanobacterium*. *J Gen Microbiol* **139**: 2511-2516.

Park, E-J., Kim, K-H., Abell, G.C.J., Kim, M-S., Roh, S.W., and Bae, J-W. (2011) Metagenomic analysis of the viral communities in fermented foods. *Appl Environ Microbiol* **77**: 1284-1291.

Park, M-O., Ikenaga, J., and Watanabe, K. (2007) Phage diversity in a methanogenic digester. *Microbial Ecology* **53**: 98-103.

Pfister P., Wasserfallen A., Stettler, F., and Leisinger, T. (1998) Molecular analysis of Methanobacterium phage psiM2. *Molec Microbiol* **30**: 233-244.

Pina, M., Bize, A., Forterre, P., and Prangishvili, D. (2011) The archaeoviruses. *FEMS Microbiol Rev* **35**: 1035-1054.

Prangishvili D., Garrett, R.A., and Koonin, E.V. (2006) Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. *Virus Res* **117**: 52-67.

Ramirez, I., and Steyer, J.P. (2008) Modeling microbial diversity in anaerobic digestion. *Wat Sci Technol* **57**: 265-270.

Rohwer, F., Prangishvili, D., and Lindell, D. (2009) Roles of viruses in the environment. *Environ Microbiol* **11**: 2771-2774.

Roux, S., Enault, F., Robin, A., Ravet, V., Personnic, S., Theil, S., *et al*. (2012) Assessing the diversity and specificity of two freshwater viral communities through metagenomics. *PLoS ONE* **7**: e33641.

Schink, B. (1997) Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol Molec Biol* **61**: 262-280.

Shapiro, O.H., Kushmaro, A., and Brenner, A. (2010) Bacteriophage predation regulates microbial abundance and diversity in a full-scale bioreactor treating industrial wastewater. *ISME J* **4**: 327-336.

Shelton D.R., and Tiedje, J.M. (1984) General method for determining anaerobic biodegradation potential. *Appl Environ Microbiol* **47**: 850-857.

Skennerton, C.T., Angly, F.E., Breitbart, M., Bragg, L., He, S., McMahon, K.D., *et al*. (2011) Phage encoded H-NS: a potential Achilles heel in the bacterial defence system. *PLoS ONE* **6**: e20095.

Smith, P.H., and Mah, R.A. (1966) Kinetics of acetate metabolism during sludge digestion. *Appl Microbiol* **14**: 368-371.
Tamaki, H., Zhang, R., Angly, F.E., Nakamura, S., Hong, P-Y., Yasunaga, T., *et al*. (2011) Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate. *Environ Microbiol* **14**: 441-452.
Thurber, R.V., Haynes, M., Breitbart, M., Wegley, L., and Rohwer, F. (2009) Laboratory procedures to generate viral metagenomes. *Nature Protocols* **4**: 470-483.
Vogelsang-Wenke, H., and Oesterhelt, D. (1988) Isolation of a halobacterial phage with a fully cytosine-methylated genome. *Mol Gen Genet* **211**: 407-414.
Weinberger, A.D., Sun, C.L., Plucinski, M.M., Denef, V.J., Thomas, B.C., Horvath, P., *et al*. (2012) Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS ONE* **8**: e1002475.
Wu, Q., and Liu, W-T. (2009) Determination of virus abundance, diversity and distribution in a municipal wastewater treatment plant. *Wat Res* **43**: 1101-1109.
Zerbino, D.R., and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**: 821-829.
Zhang, H., Banaszak, J.E., Parameswaran, P., Alder, J., Krajmalnik-Brown, R., and Rittmann, B.E. (2009) Focused-pulsed sludge pre-treatment increases the bacterial diversity and relative abundance of acetoclastic methanogens in a full-scale anaerobic digester. *Wat Res* **43**: 4517-4526.
Zumstein, E., Moletta, R., and Godon, J-J. (2000) Examination of two years of community dynamics in an anaerobic bioreactor using fluorescence polymerase chain reaction (PCR) single-strand conformation polymorphism analysis. *Environ Microbiol* **2**: 69-78.