3-1-2017

Correlating Methane Production to Microbiota in Anaerobic Digesters Fed Synthetic Wastewater

Kaushik Venkiteshwaran
Marquette University, kaushik.venkiteshwaran@marquette.edu

K. Milferstedt
INRA

M. Fujimoto
Marquette University

M. Johnson
University of Kentucky

Daniel Zitomer
Marquette University, daniel.zitomer@marquette.edu

Correlating Methane Production to Microbiota in Anaerobic Digesters Fed Synthetic Wastewater

K. Venkiteshwaran
Department of Civil, Construction and Environmental Engineering, Marquette University, Milwaukee, WI
K. Milferstedt
LBE, INRA, 102 Avenue des Etangs, Narbonne, F-11100, France
J. Hamelin
LBE, INRA, 102 Avenue des Etangs, Narbonne, F-11100, France
M. Fujimoto
Department of Civil, Construction and Environmental Engineering, Marquette University, Milwaukee, WI
M. Johnson
Department of Electrical and Computer Engineering, University of Kentucky, Lexington, KY
D. H. Zitomer
Department of Civil, Construction and Environmental Engineering, Marquette University, Milwaukee, WI

Water Research, Vol 110 (March 2017): pg. 161-169. DOI.

Abstract: A quantitative structure activity relationship (QSAR) between relative abundance values and digester methane production rate was developed. For this, 50 triplicate anaerobic digester sets (150 total digesters) were each seeded with different methanogenic biomass samples obtained from full-scale, engineered methanogenic systems. Although all digesters were operated identically for at least 5 solids retention times (SRTs), their quasi steady-state function varied significantly, with average daily methane production rates ranging from 0.09 ± 0.004 to 1 ± 0.05 L-CH_4/L_R-day (L_R = Liter of reactor volume) (average ± standard deviation). Digester microbial community structure was analyzed using more than 4.1 million partial 16S rRNA gene sequences of *Archaea* and *Bacteria*. At the genus level, 1300 operational taxonomic units (OTUs) were observed across all digesters, whereas each digester contained 158 ± 27 OTUs. Digester function did not correlate with typical biomass descriptors such as volatile suspended solids (VSS) concentration, microbial richness, diversity or evenness indices. However, methane production rate did correlate notably with relative abundances of one *Archaeal* and nine *Bacterial* OTUs. These relative abundances were used as descriptors to develop a multiple linear regression (MLR) QSAR equation to predict methane production rates solely based on microbial community data. The model explained over 66% of the variance in the experimental data set based on 149 anaerobic digesters with a standard error of 0.12 L-CH_4/L_R-day. This study provides a framework to relate engineered process function and microbial community composition which can be further expanded to include different feed stocks and digester operating conditions in order to develop a more robust QSAR model.

Keywords: Amplicon sequencing, Anaerobic digestion, Bioindicator, Microbial community composition, Multiple linear regression, Quantitative structure activity relationship

1. Introduction

There is an increasing emphasis among industries and municipalities to achieve sustainability goals by shifting from wastewater treatment to energy generation and resource recovery using anaerobic biotechnology (Angenent et al., 2004 and Novotny et al., 2010; van Loosdrecht and Brdjanovic, 2014). However, challenges still remain regarding anaerobic biotechnology implementation, as much is undetermined about the microbial factors that distinguish between a healthy and unhealthy digester (Leitão et al., 2006). Current mathematical models used for designing anaerobic treatment plants such as ADM1 (Batstone et al., 2002) typically do not include microbial diversity information and rely on the assumption that each trophic group in the anaerobic digestion (AD) process is composed of a single taxon (Ramirez et al., 2009). In

ADM1, the seven trophic groups correspond to the degradation of sugars, amino acids, long chain fatty acids, acetate, propionate, butyrate-valerate and hydrogen. ADM1 requires the input of 24 variables, of which seven relate to microbial function associated with these seven trophic groups. One of the major reasons that microbial community parameters are not included in models is because inadequate microbiological data exist; specifically community structure-function relationships and kinetic data are missing. Therefore, in order to improve the predictability of current models, understanding regarding how the microbial community structure relate to process function, such as methane generation, must be deepened (Curtis et al., 2003).

Microbial biomass concentration, along with microbial community descriptors such as; microbial richness, diversity and evenness have been described in previous studies to correlate with anaerobic digester function and stability when operating under transient conditions such as variable influent organic strength (Hashsham et al., 2000 and Fernandez et al., 2000). However, the relationships established are qualitative, not quantitative nor predictive. A few studies using multiple linear regression (MLR) modeling have reported quantitative linear relationships between *Archaeal* (i.e, methanogen) descriptors and their activity (Venkiteshwaran et al., 2015). Tale et al. (2011) applied MLR to anaerobic digester data, relating specific methanogenic activities to community structures, as defined by DGGE banding patterns targeting the methyl coenzyme A gene (*mcrA*) from methanogens. The abundance of *mcrA* was also shown to be linearly correlated with specific methanogenic activity of four H_2/CO_2 enrichment cultures (Morris et al., 2014). Bocher et al. (2015) used the *mcrA* DGGE banding pattern from a large set of 49 distinct biomass samples to develop two MLR equations to predict specific methanogenic activity (SMA) against propionate and glucose, respectively. Taken together, these studies applied MLR modeling, targeting the methanogen population only (i.e, *mcrA*), to predict specific methanogenic activity values in batch experiments, and not at steady state in a continuous operation mode.

However, since AD involves both *Archaea* and *Bacteria*, this study investigated whether both *Archaeal* and *Bacterial*

microorganisms are good descriptors in an MLR model that can predict digester operation. This report describes a relationship between digester operation after a period of more than five SRTs and the relative abundance of *Archaea* as well as *Bacteria*. This is in contrast to the previous work in which a standard bioassay parameter, SMA, was related to relative abundance values of only methanogens. The SMA value is determined in a batch test and is not similar to the standard operational conditions of full-scale plants.

In this study we use high throughput sequencing of partial 16S rRNA gene amplicons from both *Archaeal* and *Bacterial* populations. To include a large data set and diversity of anaerobic microorganisms, multiple digester sets (i.e, 50), each containing triplicate digester, started with different seed biomass were operated under identical conditions. Digesters were acclimated for a minimum of 5 hydraulic retention times (HRTs) before functional data and microbial community samples were collected. Subsequently, a predictive, quantitative structure activity relationship (QSAR) between anaerobic microbial community descriptors and digester methane production rate was developed.

2. Materials and methods

2.1. Seed inocula

Biomass samples were obtained from 50 full-scale, engineered methanogenic systems that were geographically diverse (from 49 states within the United States) and used to inoculate lab-scale digesters (Table S1). No anaerobic systems were found in Rhode Island, and two samples were obtained from different anaerobic systems (Systems A and B) in Wisconsin (WI). One sample was obtained from each of the remaining 48 states. Methanogenic biomass was from 25 anaerobic systems treating industrial waste (food, dairy and brewery industries) and 25 digesters stabilizing municipal wastewater sludge. One sample was from an AnMBR (TX) and six biomass samples were from upflow anaerobic sludge blanket (UASB) reactors (from Alabama (AL), Arkansas (AR), Colorado (CO), Idaho (ID), Kansas (KS) and Wisconsin A (WI A) with granular biomass; all other samples were flocculent biomass from continuous stirred-tank

reactors (CSTRs). With the exception of a thermophilic digester in Michigan (MI), biomass samples were from mesophilic systems.

2.2. Laboratory digester operation

Each biomass sample was used to inoculate a set of three laboratory digesters that were 160 mL serum bottles with 50 mL of active volume incubated at 35 ± 2 °C on a shaker table. Inocula containing granular biomass from UASB systems were blended using a bench-top blender for 10 s prior to seeding the digesters to disrupt the granules. Effluent removal and feeding was done manually by inserting a needle with a plastic syringe through serum bottle septa (Tale et al., 2011, Tale et al., 2015 and Carey et al., 2016). A 10-day hydraulic retention time (HRT) was maintained by removing 5 mL of effluent and adding an equal volume of synthetic industrial wastewater every day (Tale et al., 2011, Tale et al., 2015 and Venkiteshwaran et al., 2016). Synthetic industrial wastewater was a mixture of non-fat dry milk (Roundy's Supermarkets, Inc., Milwaukee, WI USA) containing 52% w/w sugars and 35% w/w proteins, 10 g/L $NaHCO_3$ and nutrient medium. The nutrient medium, as described by Speece (2008), contained the following [mg/l]: NH_4Cl [400]; $MgSO_4 \cdot 6H_2O$ [250]; KCl [400]; $CaCl_2 \cdot 2H_2O$ [120]; $(NH_4)_2HPO_4$ [80]; $FeCl_3 \cdot 6H_2O$ [55]; $CoCl_2 \cdot 6H_2O$ [10]; KI [10]; the salts $MnCl_2 \cdot 4H_2O$, NH_4VO_3, $CuCl_2 \cdot 2H_2O$, $Zn(C_2H_3O_2)_2 \cdot 2H_2O$, $AlCl_3 \cdot 6H_2O$, $Na_2MoO_4 \cdot 2H_2O$, H_3BO_3, $NiCl_2 \cdot 6H_2O$, $NaWO_4 \cdot 2H_2O$, and Na_2SeO_3) [each at 0.5]; yeast extract [100]; and resazurin [1]. Resazurin was used as an indicator of dissolved oxygen in the digesters.

All digesters were seeded at an initial volatile suspended solids (VSS) concentration of 8 g/L and operated at an organic loading rate (OLR) of 3 g COD/L-day (COD = Chemical oxygen demand). This OLR was identified after a preliminary investigation in which five inocula were tested at OLR values of 2, 3, 4 and 5 g COD/L-day. The purpose of the preliminary investigation was to identify a sustainable OLR that did not result in digester failure (i.e., digester pH < 6.5 and methane production less than 20% of theoretical maximum), but challenged the system with a high OLR to observe a wide range of COD removal and methane production rates. An OLR of 2 g COD/L-day resulted in >98 ± 0.1% COD removal for all digesters. In contrast, all preliminary

digesters failed at OLR values of 4 and 5 g COD/L-day. Therefore, the 3 g COD/L-day OLR was used for subsequent testing since it did not cause failure, but resulted in 60–90% COD removal for the inocula tested.

Biogas production volume was measured daily from day 0 by inserting a needle with a wetted glass barrel syringe through serum bottle septa. After 50 days of operation, the systems were assumed to be at a quasi-steady state based on less than 20% variation (standard deviation) in daily biogas production (Venkiteshwaran et al., 2016). Digester biogas and effluent samples were then collected for quasi steady state functional analysis over seven consecutive days. The functional analyses of all digesters were conducted between Day 50–65 of their respective operation. Functional parameters measured included biogas methane concentration, effluent volatile fatty acids (VFAs) and soluble COD (SCOD) concentrations. In addition, volatile suspended solids (VSS) and pH were measured on day seven of the analysis period. Biogas methane concentration was quantified by gas chromatography (GC System 7890A, Agilent Technologies, Irving, TX, USA) using a thermal conductivity detector. SCOD was measured by filtering the sample through a 0.45 μm membrane syringe filter and determining the filtrate COD by standard methods (APHA et al., 1998). VFA concentrations were measured by gas chromatography (GC System 7890A, Agilent Technologies, Irving, TX, USA) using a flame ionization detector. Digester VSS were determined by standard methods (APHA et al., 1998). The pH was measured using a pH meter (Orion 4 Star, Thermo, Waltham, MA, USA).

2.3. Microbial community analyses

Digester effluent samples (1 mL) were collected for DNA extraction on six consecutive days when digester functional analyses were performed and pooled. Each day, the effluent samples were centrifuged at 10,000 RPM (9400 g) for 10 min. Centrifuged solids for consecutive days were combined at equal proportions and DNA was extracted using a commercial kit according to manufacturer instructions (PowerSoil™ DNA Isolation Sample Kit, MoBio Laboratories, Inc., Carlsbad, CA). Biomass samples were subjected to an initial bead beating step for 10 min on a vortexer (Model 58,816-

121, VWR International, Radnor, PA, USA). DNA extraction from one digester inoculated with biomass from North Dakota (ND3) failed. Therefore, data from this digester were excluded from further analysis and data from the remaining 149 digesters were subsequently employed.

Forward and reverse primers 515–532U and 909–928U (Wang and Qian, 2009) were used to amplify the V4-V5 variable region of the *Bacterial* and *Archaeal* 16S rRNA genes. Previous studies have used these primer pairs for studying methanogenic communities. (Braun et al., 2015, Resende et al., 2015 and Venkiteshwaran et al., 2016). The DNA sample and primers with their respective linkers were amplified over 30 cycles at an annealing temperature of 65 °C. An index sequence was added in a second PCR reaction of 12 cycles, and the resulting products were purified and loaded onto the Illumina MiSeq cartridge for sequencing of paired 300 bp reads following manufacturer instructions (Reagent Kit v3, Illumina, Inc., San Diego, CA USA). Sequencing work was performed at the Genotoul Lifescience Network Genome and Transcriptome Core Facility in Toulouse, France (get.genotoul.fr).

Forward and reverse sequences were assembled and quality checked using a modified version of the standard operation procedure by Kozich et al. (2013) in Mothur version 1.33.0, including chimera detection by sample using the uchime implementation in Mothur. Zero ambiguous bases and zero unknown bases (N) were allowed. Preclustering of sequences was done to minimize noise because of random sequencing errors. Sequences were allowed to cluster with up to four bases difference over an average length of 375 bp. Singletons and sequences found only twice were removed from the analysis. Sequence alignment and taxonomic outlining was accomplished using SILVA SSURef NR99, release 119, as provided by Schloss et al. (2009). The same database was used in Mothur's classify. seqs() command to assign taxonomic affiliation using a cutoff value of 80%. Final sequence data were clustered in 1300 operational taxonomic units (OTUs) at 97% similarity and rarified to the lowest sequence reads per sample (27,315 reads).

2.4. Initial screening to select significant OTUs

Initial screening was performed to select OTUs with relative abundances that were highly correlated with average methane production rates. Initial screening was done by performing 50 iterations to build 50 Spearman's rank correlation matrices. In each iteration, the Spearman's rank correlation coefficients for the relative abundance of each of the 1300 OTUs and the average methane production rates were calculated using data from 75 (of the total 149) randomly selected digesters to obtain 50 unique Spearman's matrices. Multiple, unique Spearman's matrices were desired so that OTUs that consistently showed strong correlation with methane production could be identified. The Spearman's rank matrices were checked to ensure that all 149 digesters were included at least once among the 50 iterations. Spearman's rank was employed as a measure of monotonic statistical dependence because of its robustness since it does not require underlying assumptions regarding the frequency of distribution of variables (e.g., normal distribution, uniformly distributed etc.) or the existence of a linear relationship between variables (Zuur et al., 2007).

The 10 OTUs with relative abundance values most positively related (i.e., having the highest Spearman rank scores) and the 10 OTUs most negatively related (i.e., having the lowest Spearman rank scores) to methane production rates were selected during each of the initial 50 screening iterations. The OTUs which were repeatedly selected in more than 75% of the 50 initial screening iterations were deemed to be highly significant. Relative abundance values of these highly significant OTUs were subsequently used to develop the QSAR linear regression model.

2.5. Linear model and QSAR equation

A MLR leave group out (LGO) approach was employed to validate a quantitative relationship between relative abundance values of the 10 highly significant OTUs identified during initial screening. Digesters were randomly partitioned into 10 subsets of 14 or 15 digesters each. Subsequently, ten validation tests were performed. In each test, 9 of these 10 subsets were combined and used as a training

set to develop an MLR equation, whereas the remaining subset was used as a validation set to test the equation predictability. This was repeated until all 10 digester subsets were used once as a validation subset.

The predictability of the MLR equation was deemed good if the following four criteria were met (Golbraikh and Tropsha, 2000 and Konovalov et al., 2008): (1) $q^2 > 0.5$, (2) $R^2 > 0.6$, (3) $(R^2-R_O^2)/R^2 < 0.1$ and (4) $0.85 \leq K \leq 1.15$. For this, q^2 is the chi square value calculated using the observed versus predicted methane production values described by Schüürmann et al. (2008) and R^2 is the coefficient of determination for the linear regressions of predicted versus observed methane production rates. Additionally, R_O^2 and K are the coefficient of determination and the slope for the test set linear regression equation of predicted versus observed methane production rates forced through the origin, respectively.

After confirming that MLR equations demonstrated good predictability, all 10 digester subsets were combined and used to determine a final, QSAR linear regression model.

2.6. Analytical methods

Average, standard deviation, variance and *t*-test calculations were performed using Excel 2010 (Version 14.3.2 – Microsoft, USA) built in functions. Richness (*S*), Shannon diversity (*H*), and evenness (*E*) indices were calculated from the abundance tables. Richness was calculated as the number of OTUs identified at the genus level. The Shannon-Weaver diversity index values were determined as described by Briones et al. (2007). Evenness was calculated as described by Falk et al. (2009). Principal Coordinates analysis (PCoA) was performed using the R software environment (R Core Team, 2015). ANOSIM using Bray-Curtis dissimilarity was performed to assess the relationship between methane production and relative abundance values microbial community using the vegan package in R (Oksanen et al., 2016). The Bray-Curtis dissimilarity, Spearman rank correlation and MLR analyses were performed using Excel 2010 (Version 14.3.2 – Microsoft, USA) with the added statistical software package XLStat Pro 2014 (Addinsoft, USA). Blast search of representative sequences was conducted using default settings and excluding uncultured sequences

on the browser based blastn tool (https://blast.ncbi.nlm.nih.gov/) (Altschul et al., 1990).

3. Results and discussion

3.1. Digester function

All digester sets were operated identically, but were seeded with different biomass. The seed biomass origin ostensibly had a significant influence on functional performance, both initially as well as after 80 days. Significant variability was observed in methane production rates (0.09 ± 0.004 to 0.98 ± 0.05 L CH_4/L_R-day) (Table S2) as well as effluent SCOD concentration (2.6 ± 0.30 to 25 ± 1.1 g/L), total VFA concentration (1.6 ± 3.8 to 19 ± 1.3 g/L as acetic acid) and pH (5.8–7.6) (Fig. 1). VFAs constituted 56 ± 12% of the effluent SCOD.

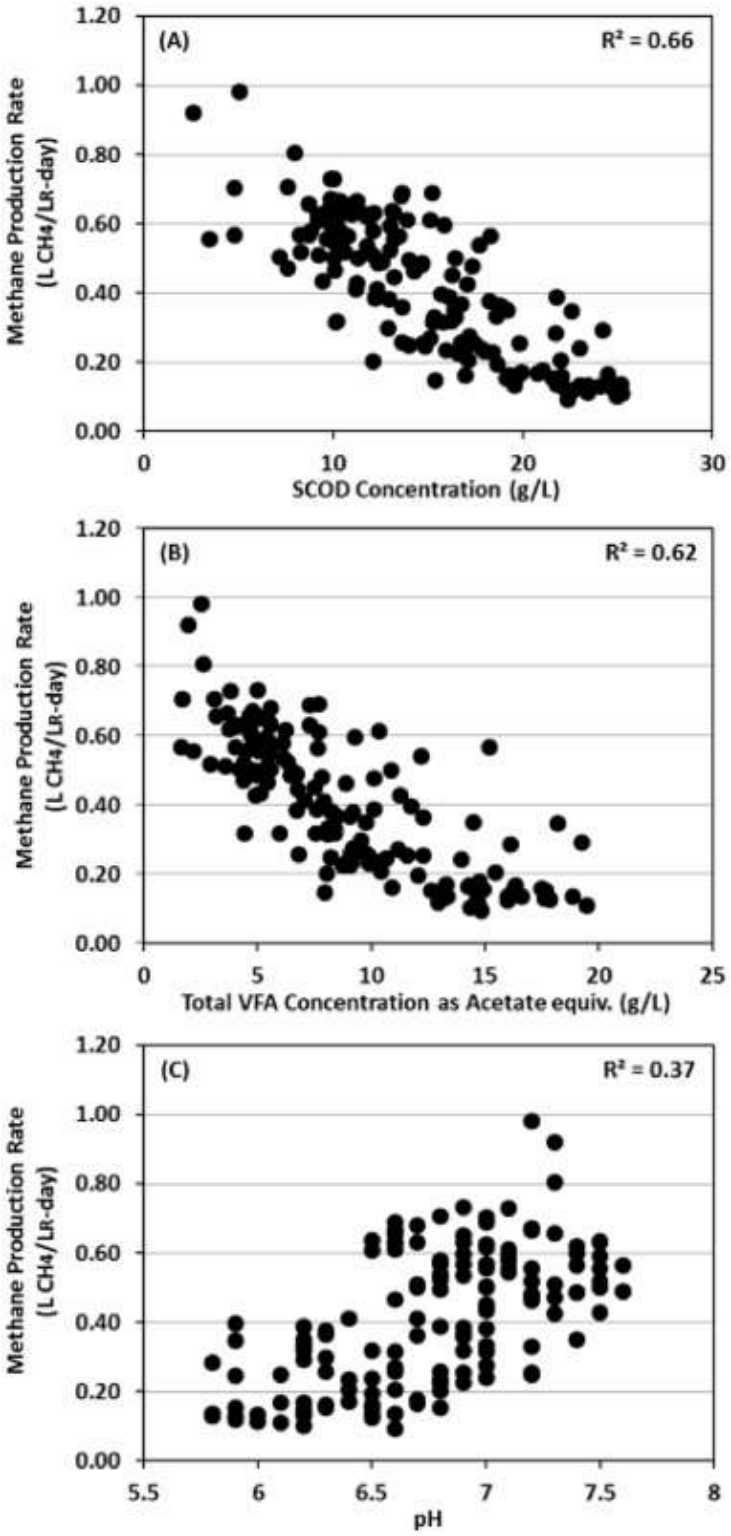


Fig. 1. Average daily methane production versus effluent parameters. Average daily methane production (L-CH4/LR-day) versus (A) SCOD concentration (g/L), (B) total VFA concentration as acetic acid (g/L) and (C) pH.

Relationships between the methane production rates and effluent parameters were as expected: methane production was inversely correlated to effluent SCOD concentrations, effluent total VFA concentrations and pH (Fig. 1). Acetic acid and propionic acid contributed 55 ± 18% and 40 ± 20%, respectively, of the total VFA equivalents (Fig. S1). Among all VFAs, the acetic acid concentrations showed the strongest linear correlation with methane production rates (Fig. S1A).

3.2. Microbial diversity analysis

Illumina sequencing yielded a total of 14.5 million raw sequence reads after making contigs, with 90,324 ± 32,901 raw reads per digester sample. After filtering, quality control and chimera removal using the procedure described in section 2.3, 10 million total sequence reads with 62,943 ± 24,560 reads per sample were obtained. A sequence-based rarefaction analysis was performed to test for efficient OTU coverage, as shown in Fig. S2. After 27,315 sequence reads (lowest sequence reads per sample), the number of OTUs was saturated, as revealed by the asymptotic nature of the 149 rarefaction curves. Therefore, a total of 4.1 million sequence reads from all 149 digesters were analyzed with 27,315 rarified sequence reads per sample. Based on 97% similarity, 1300 microbial OTUs were observed with an average of 158 ± 27 observed OTUs per digester. The microbial communities of replicate (seeded from the same source) and non-replicate (seeded from different sources) digesters were compared using Bray-Curtis dissimilarity analyses. The average Bray-Curtis dissimilarity value of all replicate digester pairs (0.19 ± 0.1, n = 148) was significantly smaller than that for non-replicate digester pairs (0.52 ± 0.15, n = 10,878). Therefore, replicate digester microbial communities were more similar than communities of digesters seeded with biomass from different sources.

3.2.1. Archaeal community

The relative abundance of *Archaeal* sequences ranged from <0.01% to 3%. Fig. S3 shows *Archaeal* community profile at the class level across 149 digesters. Two *Archaeal* class *Methanobacteria* and *Methanomicrobia* were highly prevalent and were observed in 98% and

89% of the digesters. *Methanobacteria* and *Methanomicrobia* were also the most dominant among the *Archaeal* classes, with an average relative abundance of 50 ± 39% and 38 ± 18% across all digesters. At genus level, methanogens of the genera *Methanosarcina* and *Methanobacterium* showed the highest average relative abundance. These two methanogens were detected in 67% and 81% of the digesters, respectively, and their combined relative abundance was 80 ± 19% of the total *Archaeal* OTUs observed. *Methanosarcina* can perform both aceticlastic (acetate utilizing) and hydrogenotrophic (hydrogen utilizing) methanogenesis whereas; members of genus *Methanobacterium* are only known to perform hydrogenotrophic methanogenesis (Liu and Whitman, 2008).

Methanosarcina have a higher growth rate and lower affinity for acetate than the only other known aceticlastic methanogen (*Methanosaeta*). They typically outcompete *Methanosaeta* in digesters with high acetate concentration (>500 mg/L) (Hori et al., 2006, Westermann et al., 1989 and Conklin et al., 2006). Since 99% of the digesters in this study had an acetic acid concentration of more than 500 mg/L, the presence of *Methanosarcina* as the dominant aceticlastic methanogen is reasonable.

Hydrogenotrophic methanogens including *Methanobacterium* are typically more tolerant than aceticlastic methanogensto stress conditions such as low pH and high VFA concentrations (Liu and Whitman, 2008). The relative abundance of *Methanobacterium* was higher than that of *Methanosarcina* in 66% of the digesters (Fig. S4), with the average pH and VFA concentration of these digesters being 6.4 ± 0.3 and 10.1 ± 4.5 g/L, respectively. Therefore, in these digesters, the typically higher relative abundance of *Methanobacterium* was probably due to the inhibition of *Methanosarcina* by low pH and high VFA concentration.

3.2.2. Bacterial community

Bacterial communities were dominated by the phyla *Bacteroidetes*, *Firmicutes* and *Synergistes*, contributing 59 ± 17%, 22 ± 17% and 9 ± 5% of the total *Bacterial* sequences, respectively. More than 99% of these three phyla were dominated by members of the class *Bacteroidia, Clostridia* and *Synergistia*, respectively, and their

presence were observed across all 149 digesters (Fig. S5). It was shown previously that *Proteobacteria*, *Firmicutes*, *Bacteroidetes* and *Chloroflexi* were the four major phyla in the *Bacterial* domain in a survey of 21 full-scale anaerobic digesters (Sundberg et al., 2013). The combined relative abundance of the three major *Bacterial* phyla found in this study was 91 ± 8% of the total *Bacterial* sequences. Members of the phyla *Bacteroidetes* and *Firmicutes* are functionally diverse. These phyla contain hydrolytic bacteria as well as acidogenic, fermentative bacteria (Noike et al., 1985, Mata-Alvarez et al., 2000, Vidal, 2000, Ariunbaatar et al., 2014 and Stiles and Holzapfel, 1997). However, metagenomic analyses have shown that a majority of species involved in anaerobic digestion still remain unclassified (Rivière et al., 2009 and Treu et al., 2016).

The most abundant genera of the phylum *Bacteroidetes* observed across digesters were *Bacteroides*, *Petrimonas*, *Paludibacter*, *Porphyromonas*, *VadinBC27 wastewater sludge group*, *unclassified M2PB4*-65 *termite group* and *unclassified Prevotellaceae* (present in >95% digesters; combined abundance = 87 ± 12% of total *Bacteroidetes*). Similarly for *Firmicutes*, unclassified *Family XI*, *Family XIII* and *Ruminococcaceae* were detected in >95% of the digesters, contributing 40 ± 17% of the total *Firmicutes* sequences. *Synergistes* were dominated by the genus *Aminobacterium*. The synthetic wastewater carbon source was non-fat dry milk that contained 16% proteins by mass. Species of the genera *Aminobacterium* such as; *Aminobacterium colombiense*, *Aminobacterium mobile* and *Aminobacterium thunnarium* have been identified as amino acid fermenting bacterium (Baena et al., 1998, Baena et al., 2000 and Hamdi et al., 2015). Therefore, detection of genus *Aminobacterium* in systems fed with protein is reasonable.

Microbial biomass concentration (measured as VSS concentration), microbial richness, Shannon-Weaver diversity and evenness indices did not correlate with digester methane production rates (Fig. 2). Although it is generally assumed that digesters with higher biomass VSS concentration achieve higher biogas production rates compared to similar digesters with lower biomass concentration, results indicate that having a higher VSS concentration cannot be universally considered to yield better function.

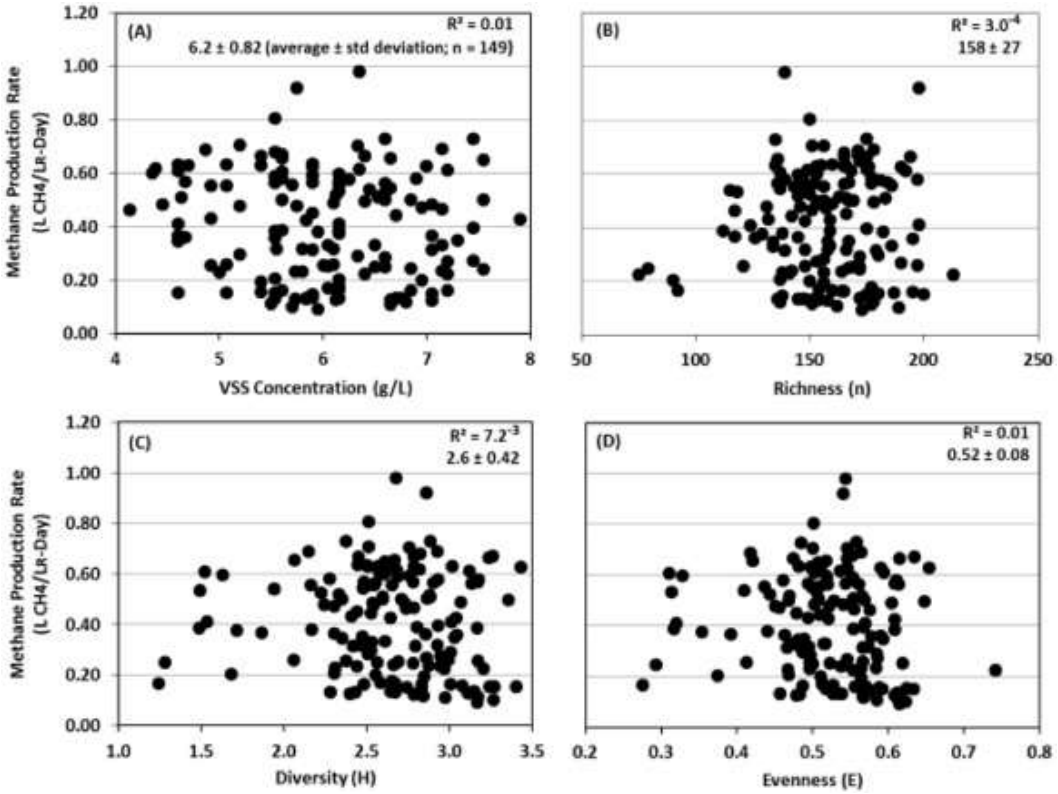


Fig. 2. Average methane production (L-CH4/LR-day) versus (A) VSS Concentration (B) Richness (C) Diversity (D) Evenness.

Others have reported that microbial community descriptors such as diversity and evenness indices relate to anaerobic digester function (Fernandez et al., 2000, Hashsham et al., 2000, Carballa et al., 2011 and Werner et al., 2011). Increased microbial diversity and evenness relate to increased functional resistance and resilience when conditions are not steady and influent characteristics such as flow rate, organic strength, feedstock composition and temperature vary and cause perturbations. Higher diversity results in a higher probability of functional redundancy and, thus, functional stability during and after perturbation (Fernandez et al., 2000, Hashsham et al., 2000, Carballa et al., 2011 and Werner et al., 2011).

3.3. Initial screening and quantitative structure activity relationship (QSAR)

Although methane production did not correlate with overall biomass concentration, richness, microbial diversity or evenness, it did

correlate with the relative abundance values of 10 OTUs identified during initial screening (Table S3).

Digesters were grouped in three roughly equally sized classes according to their methane production rate ("High", "Medium" and "Low") (Table S2). Triplicate digesters were not always grouped together in this partitioning. The methane production rates in the three categories were 0.63 ± 0.09 (n = 50), 0.41 ± 0.08 (n = 50) and 0.18 ± 0.05 (n = 49) L-CH_4/LR-day, respectively. Microbial communities in high methane production digesters were different from those associated with medium (ANOSIM, p = 0.002, R^2 = 0.072) and low (ANOSIM, p = 0.001, R^2 = 0.368) methane production rates (Fig. S6A). Reducing the number of OTUs from 1300 to the 10 highly significant OTUs resulted in greater observable differences among microbial communities, increasing the observed community variation by PCoA axis-1 from 21.4 to 42.3% (Fig. S6). In addition, using only the 10 highly significant OTUs resulted in greater observable variation between high and medium (ANOSIM, p = 0.001, R^2 = 0.138) as well as high and low (ANOSIM, p = 0.001, R^2 = 0.493) digester groups (Fig. S6B).

Collinearity between pairs of the 10 highly significant relative abundance values was tested, since in MLR analysis, intercorrelation between any two (collinearity) or many variables (polycollinearity) can produce false models (Nirmalakhandan and Speece, 1988). Intercorrelation R^2 values were relatively low, averaging 0.07 ± 0.07 (n = 45) and ranging from 0.001 to 0.34. Therefore, collinearity problems did not exist among the 10 highly significant OTU relative abundance values (Table S4), allowing these OTUs to be used as descriptor variables for subsequent MLR analysis.

3.3.1. Linear model validation and QSAR equation

Ten MLR validation tests using the 10 highly significant OTUs were conducted by randomly dividing the 149 digesters into 10 groups (Table S5, Fig. S7). All four criteria for good predictability were satisfied in nine of the 10 validation iterations, indicating that the MLR approach resulted in equations with good predictability (Table S6). Therefore, the final QSAR MLR equation was generated by combining data from all 149 digesters:

Water Research, Vol 110 (March 2017): pg. 161-169. DOI.

equation(1)

$$MPR = 0.4 + {}^{2-04}\cdot\boldsymbol{OTU1} + 1.3^{-01}*\boldsymbol{OTU2} + 2.6^{-01}*\boldsymbol{OTU3} + 6.0^{-03}*\boldsymbol{OTU4} + 4.5^{-04}*\boldsymbol{OTU5} + 2.1^{-01}*\boldsymbol{OTU6} + 9.1^{-03}*\boldsymbol{OTU7} - 1.5^{-03}*\boldsymbol{OTU8} - 5.8^{-02}*\boldsymbol{OTU9} - 2.5^{-01}\boldsymbol{OTU10}$$

$$n = 149;\ R^2 = 0.66;\ SE = 0.12\ (L - CH_4/L_R - day)$$

where MPR is the methane production rate ($L\text{-}CH_4/L_R$-day) and OTUn is the relative abundance for taxon n (%). A plot of observed methane production rates versus rates predicted using Equation (1) for all 149 digesters is shown in Fig. 3.

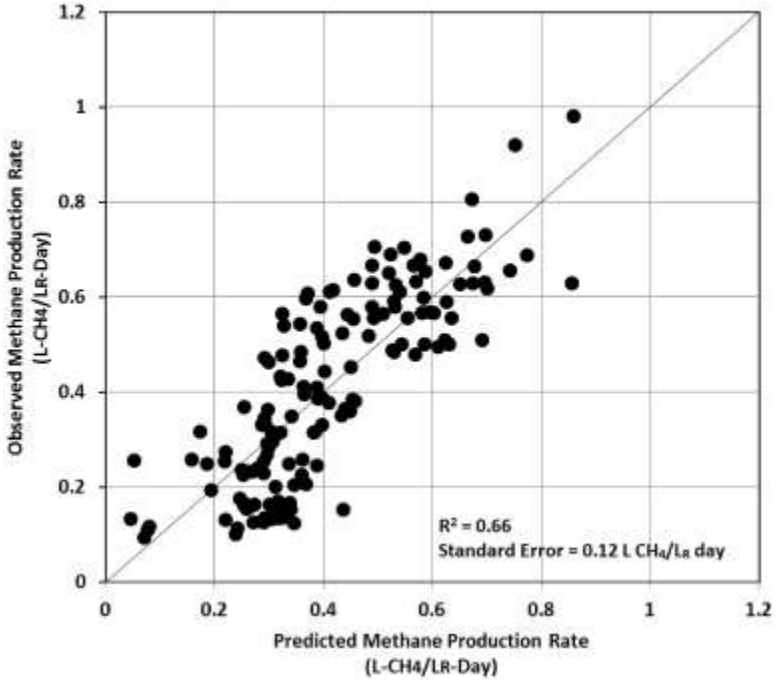


Fig. 3. Observed versus predicted methane production rate. The predicted rate was calculated using equation.[1]

The prevalence and range of relative abundance values for the 10 highly significant OTUs varied across all the digesters. The OTUs were prevalent in 52% (OTU7) to 100% (OTU1) of the 149 digesters, whereas the average relative abundance across all the digesters ranged from 0.2% (OTU6) to 8.9% (OTU8) (Table 1).

Water Research, Vol 110 (March 2017): pg. 161-169. DOI.

Table 1. Highly significant OTUs.

OTU	Order	[a]Prevalence (%)	Relative abundance range & (average) (%)	Eq. (1) coefficient	Average contribution value (Absolute value (coefficient average relative abundance))[a]100
1	*Bacteroidales*	100	<0.01 to 36 (6.5)	2^{-04}	0.13
2	*Bacteroidales*	77	<0.01 to 1.3 (0.09)	1.3^{-01}	1.1
3	*Spirochaetales*	67	<0.01 to 0.7 (0.05)	2.6^{-01}	1.3
4	*Bacteroidales*	98	<0.01 to 54 (6.1)	6^{-03}	3.7
5	*Clostridiales*	97	<0.01 to 3.7 (0.3)	4.5^{-04}	0.013
6	*Methanosarcinales*	54	<0.01 to 2.6 (0.2)	2.1^{-01}	4.2
7	*Clostridiales*	52	<0.01 to 5.2 (0.08)	9.1^{-03}	0.072
8	*Bacteroidales*	99	<0.01 to 58 (8.9)	-1.5^{-03}	1.4
9	*Clostridiales*	97	<0.01 to 4.4 (0.36)	-5.8^{-02}	2.1
10	*Clostridiales*	78	<0.01 to 1.0 (0.08)	-2.5^{-01}	2.0

[a]Prevalence = (n/149)×100; where n = number of the digesters in which an OTU was observed at >0.01% relative abundance.

The coefficients of the MLR equation (e.g., coefficient value 2^{-04} for *OTU1*) could not be used directly as indicators of the relative contribution of independent variables since the OTU average relative abundance values were different. Therefore, an average contribution value was calculated for each OTU as the absolute value of the product of the MLR coefficient and the corresponding average relative abundance (Table 1). Based on average contribution values, OTU6 was the independent variable that most significantly contributed to the predicted methane production rate, followed by OTU4 and OTU9. A blast search was conducted using the reference sequences of the 10 highly significant OTUs (Table 2). OTUs 6, 4 and 9 were most similar to *Methanosarcina*, unclassified *Bacteroidales* and unclassified *Intestinimonas sp.*, respectively. Therefore, high relative abundance values of *Methanosarcina* and the detected unclassified *Bacteroidales* are ostensibly beneficial when high methane production rate is desired under the conditions studied. Similarly, low relative abundance of the detected unclassified *Intestinimonas sp.* is ostensibly beneficial. In this way, QSAR models can be employed as research tools to identify

Water Research, Vol 110 (March 2017): pg. 161-169. DOI.

potentially desirable and undesirable taxa for further consideration. For example, bioaugmenting low methane producing digesters with specific taxa identified as beneficial by QSAR modeling may be promising, but more research is warranted to explore this approach.

Table 2. BLAST search results for highly significant OTUs.

OTU	Accession #	Name	Query length (bp)	Query cover (%)	Identity (%)	E value
1	LT558828	*Petrimonas sulfuriphila strain Marseille-P1901*	372	100	97	3^{-174}
2	KF282390	*Cytophagaceae bacterium GUDS1294*	371	100	89	8^{-134}
3	GU196244.1	*Bacterium enrichment culture clone R4-82B*	376	100	100	0.0
4	LC049960	*Bacteroidales bacterium TBC1*	372	100	86	6^{-112}
5	NR122058	*Syntrophomonas wolfei strain Goettingen G311*	376	100	97	9^{-180}
6	CP008746	*Methanosarcina barkeri CM1*	380	100	99	0.0
7	NR041236	*Lutispora thermophila*	376	100	95	1^{-170}
8	FJ848568	*Porphyromonas sp. 2192 16S ribosomal RNA gene*	372	99	93	3^{-150}
9	KP114242	*Intestinimonas sp. FSAA-17*	375	100	99	0.0
10	AB910747	*Clostridium scindens*	375	100	100	0.0

Apart from a methanogen (*Methanosarcina*) and a syntrophic acetogen (*Syntrophomonas*), the 10 descriptors also include fermenters (acidogens) (Table 2). Also, the fermenters were both positively and negatively correlated with digester methane production rate. Of the fermenters identified at the genus level, members of *Petrimonas* and *Porphyromonas* are known to ferment sugars whereas *Lutispora* are amino acid fermenters (Grabowski et al., 2005, Shah and Collins, 1988 and Shiratori et al., 2008). The positive correlation of higher methane production with high relative abundance of OTU7 (most similar to *Lutispora*) is reasonable since the synthetic wastewater contained protein.

Digesters exhibiting high methane production also had higher pH values (Fig. 1). Also, it is possible that different digester pH values selected for different fermenters. Studies have reported fluctuation in anaerobic digester *Bacterial* populations in response to variations in environmental parameters, including pH (Lü et al., 2009 and Ogbonna et al., 2015). Digesters with near-neutral pH may have supported higher growth rates of *Petrimonas* and *Lutispora* that exhibited relative

abundances positively correlated with methane production (Table 2). Representative species of these genera (i.e., *Petrimonas sulfuriphila* and *Lutispora thermophile*) show optimal growth rates at neutral pH (Grabowski et al., 2005 and Shiratori et al., 2008). Relative abundance of *Ruminococcaceae*, which was negatively correlated with methane production, has been observed to increase in digesters undergoing perturbation and with low pH (Tian et al., 2014).

The relationship observed in this study may be based on correlation rather than causation. Additional research should be performed to determine if the microbial community composition as described by relative abundance causes digester methane production to vary. Given the many factors influencing microbial community, including wastewater composition, digester operation, environmental parameters (pH, temperature, salt, VFA concentration etc.) and optimal growth range of various *Archaea* and *Bacteria* (Chen et al., 2008, Enright et al., 2009 and Karakashev et al., 2005), developing a more general, robust QSAR may require extensive research using a large number of environmental conditions. This would be a worthwhile endeavor to help improve modeling and functional performance of anaerobic digesters and other engineered bioprocesses.

4. Conclusions

The study investigated whether microbial community descriptors can be used in a QSAR model to predict digester methane production rate. To include a large data set and diversity of anaerobic microorganisms, 50 distinct biomass samples were used to seed triplicate lab-scale digesters. Although all digesters were operated identically for a minimum of 5 retention cycles, their quasi steady-state function varied significantly. The most dominant *Archaeal* OTUs were *Methanosarcina* and *Methanobacterium*. The *Bacterial* community was dominated by the phyla *Bacteroidetes*, *Firmicutes* and *Synergistes*.

No correlation was observed between methane production rate and the common biomass descriptors of digester biomass concentrations (VSS), microbial richness, Shannon Weaver diversity and evenness indices. However, the relative abundance values of 10 OTUs including one *Archaeal* and nine *Bacterial* taxa were found to

significantly correlate with digester methane production rate. Seven OTUs positively correlated and the remaining three negatively correlated to digester methane production rate. The relative abundance values of the 10 OTUs were used as descriptors to develop a MLR equation demonstrating good predictability of digester methane production rate. Apart from a methanogen (*Methanosarcina*) and a syntrophic acetogen (*Syntrophomonas*), the 10 descriptors also included fermenters (acidogens). To the author's knowledge, this is the first report of a quantitative, predictive correlation between digester quasi steady state methane production rate and microbial community descriptors.

Future research with multiple biomass samples subjected to factors that are known to influence digester microbial community and their activity, such as different wastewater composition, pH, temperature, digester configuration and OLR, in combination with high-throughput sequencing and conventional approaches for kinetic study, could advance development of more robust QSAR models that could be incorporated into existing AD models to improve predictability.

Acknowledgements

The authors thank Mike Dollhopf (Water Quality Center, Marquette University, Milwaukee USA) and Dr. Jean-Philippe Steyer (Laboratory of Environmental Biotechnology, Narbonne, France) for their contributions to this work.

References

Altschul et al., 1990. S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman. Basic local alignment search tool. *J. Mol. Biol.,* 215 (3) (1990), pp. 403–410 http://dx.doi.org/10.1016/S0022-2836(05)80360-2

APHA et al., 1998. American Public Health Association (APHA), American Water Works Association (AWWA), Water Environment Federation (WEF), *et al. Standard Methods for the Examination of Water and Wastewater*. (twentieth ed.) American Public Health Association (1998) ISBN: 0875532357

Angenent et al., 2004. L.T. Angenent, K. Karim, M.H. Al-Dahhan, B.A. Wrenn, R. Domíguez-Espinosa. Production of bioenergy and biochemicals from

industrial and agricultural wastewater. *Trends Biotechnol.,* 22 (9) (2004), pp. 477–485 http://dx.doi.org/10.1016/j.tibtech.2004.07.001

Ariunbaatar et al., 2014. J. Ariunbaatar, A. Panico, G. Esposito, F. Pirozzi, P.N.L. Lens. Pretreatment methods to enhance anaerobic digestion of organic solid waste. *Appl. Energy,* 123 (2014), pp. 143–156 http://dx.doi.org/10.1016/j.apenergy.2014.02.035

Baena et al., 1998. S. Baena, M.L. Fardeau, M. Labat, B. Ollivier, P. Thomas, J.L. Garcia, B.K.C. Patel. Aminobacterium colombiense gen. nov. sp. nov., an amino acid-degrading anaerobe isolated from anaerobic sludge. *Anaerobe*, 4 (1998), pp. 241–250 http://dx.doi.org/10.1006/anae.1998.0170

Baena et al., 2000. S. Baena, M.L. Fardeau, M. Labat, B. Ollivier, J.L. Garcia, B.K. Patel. Aminobacterium mobile sp. Nov., a new anaerobic amino-acid-degrading bacterium. *Int. J. Syst. Evol. Microbiol.,* 50 (1) (2000), pp. 259–264 http://dx.doi.org/10.1099/00207713-50-1-259

Batstone et al., 2002. D.J. Batstone, J. Keller, I. Angelidaki, S.V. Kalyuzhnyi, S.G. Pavlostathis, A. Rozzi, W.T. Sanders, H. Siegrist, V.A. Vavilin. The IWA anaerobic digestion model No 1 (ADM1). *Water Sci. Technol.,* 45 (10) (2002), pp. 65–73 ISBN: 1900222787

Bocher et al., 2015. B.T.W. Bocher, K. Cherukuri, J.S. Maki, M. Johnson, D.H. Zitomer. Relating methanogen community structure and anaerobic digester function. *Water Res.,* 70 (2015), pp. 425–435 http://dx.doi.org/10.1016/j.watres.2014.12.018

Braun et al., 2015. F. Braun, J. Hamelin, A. Bonnafous, N. Delgenès, J.P. Steyer, D. Patureau. Similar PAH fate in anaerobic digesters inoculated with three microbial communities accumulating either volatile fatty acids or methane. *PLoS One,* 10 (4) (2015), pp. 1–20 http://dx.doi.org/10.1371/journal.pone.0125552

Briones et al., 2007. A.M. Briones, B.J. Daugherty, L.T. Angenent, K.D. Rausch, M.E. Tumbleson, L. Raskin. Microbial diversity and dynamics in multi and single-compartment anaerobic bioreactors processing sulfate rich waste streams. *Environ. Microbiol.,* 9 (1) (2007), pp. 93–106 http://dx.doi.org/10.1111/j.1462-2920.2006.01119.x

Carballa et al., 2011. M. Carballa, M. Smits, C. Etchebehere, N. Boon, W. Verstraete. Correlations between molecular and operational parameters in continuous lab-scale anaerobic reactors. *Appl. Microbiol. Biotechnol.,* 89 (2) (2011), pp. 303–314 http://dx.doi.org/10.1007/s00253-010-2858-y

Carey et al., 2016. D.E. Carey, D.H. Zitomer, K.R. Hristova, A.D. Kappell, P.J. McNamara. Triclocarban influences antibiotic resistance and alters anaerobic digester microbial community structure. *American Chemical Society Environ. Sci. Technol*., 50 (1) (2016), pp. 126–134 http://dx.doi.org/10.1021/acs.est.5b03080

Chen et al., 2008. Y. Chen, J.J. Cheng, K.S. Creamer. Inhibition of anaerobic digestion process: a review. *Bioresour. Technol.,* 99 (10) (2008), pp. 4044–4064 http://dx.doi.org/10.1016/j.biortech.2007.01.057

Conklin et al., 2006. A. Conklin, H.D. Stensel, J. Ferguson. Growth kinetics and competition between *Methanosarcina* and *Methanosaeta* in mesophilic anaerobic digestion. *Water Environ. Res.,* 78 (5) (2006), pp. 486–496 ISSN: 1061-4303

Curtis et al., 2003. T.P. Curtis, I.M. Head, D.W. Graham. Theoretical engineering: ecology for biology. *Environ. Sci. Technol.* (2003), pp. 64–70 http://dx.doi.org/10.1021/es0323493

Enright et al., 2009. A.M. Enright, V. McGrath, D. Gill, G. Collins, V. O'Flaherty. Effect of seed sludge and operation conditions on performance and archaeal community structure of low-temperature anaerobic solvent-degrading bioreactors. *Syst. Appl. Microbiol.,* 32 (1) (2009), pp. 65–79 http://dx.doi.org/10.1016/j.syapm.2008.10.003

Falk et al., 2009. M.W. Falk, K.G. Song, M.G. Matiasek, S. Wuertz. Microbial community dynamics in replicate membrane bioreactors - Natural reproducible fluctuations. *Water Res.,* 43 (3) (2009), pp. 842–852 http://dx.doi.org/10.1016/j.watres.2008.11.021

Fernandez et al., 2000. A.S. Fernandez, S.A. Hashsham, S.L. Dollhopf, L. Raskin, O. Glagoleva, F.B. Dazzo, R.F. Hickey, C.S. Criddle, J.M. Tiedje. Flexible community structure correlates with stable community function in methanogenic bioreactor communities perturbed by glucose. *Appl. Environ. Microbiol.,* 66 (9) (2000), pp. 4058–4067 http://dx.doi.org/10.1128/AEM.66.9.4058-4067.2000

Golbraikh and Tropsha, 2000. A. Golbraikh, A. Tropsha. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *Mol. Divers.,* 5 (4) (2000), pp. 231–243 http://dx.doi.org/10.1023/A:1021372108686

Grabowski et al., 2005. A. Grabowski, B.J. Tindall, V. Bardin, D. Blanchet, C. Jeanthon. Petrimonas sulfuriphila gen. nov., sp. nov., a mesophilic fermentative bacterium isolated from a biodegraded oil reservoir. *Int. J. Syst. Evol. Microbiol.,* 55 (3) (2005), pp. 1113–1121 http://dx.doi.org/10.1099/ijs.0.63426-0

Hamdi et al., 2015. O. Hamdi, W.B. Hania, A. Postec, H. Bouallagui, M. Hamdi, P. Bonin, B. Ollivier, M.L. Fardeau. Aminobacterium thunnarium sp. nov., a mesophilic, amino acid-degrading bacterium isolated from an anaerobic sludge digester, pertaining to the phylum synergistetes. *Int. J. Syst. Evol. Microbiol.,* 65 (2015), pp. 609–614 http://dx.doi.org/10.1099/ijs.0.068965-0

Hashsham et al., 2000. S.A. Hashsham, A.S. Fernandez, S.L. Dollhopf, F.B. Dazzo, R.F. Hickey, J.M. Tiedje, C.S. Criddle. Parallel processing of substrate correlates with greater functional stability in methanogenic

bioreactor communities perturbed by glucose. *Appl. Environ. Microbiol.,* 66 (9) (2000), pp. 4050–4057 http://dx.doi.org/10.1128/AEM.66.9.4050-4057.2000

Hori et al., 2006. T. Hori, S. Haruta, Y. Ueno, M. Ishii, Y. Igarashi. Dynamic transition of a methanogenic population in response to the concentration of volatile fatty acids in a thermophilic anaerobic digester dynamic transition of a methanogenic population in response to the concentration of volatile fatty acids in a thermophilic anaerobic digester. *Appl. Environ. Microbiol.,* 72 (2) (2006), pp. 1623–1630 http://dx.doi.org/10.1128/AEM.72.2.1623

Karakashev et al., 2005 D. Karakashev, D.J. Batstone, I. Angelidaki. Influence of environmental conditions on methanogenic compositions in anaerobic biogas reactors. *Appl. Environ. Microbiol.,* 71 (1) (2005), pp. 331–338 http://dx.doi.org/10.1128/AEM.71.1.331-338.2005

Konovalov et al., 2008. D.A. Konovalov, L.E. Llewellyn, Y.V. Heyden, D. Coomans. Robust cross-validation of linear regression QSAR models robust cross-validation of linear regression QSAR models. *J. Chem. Inf. Model.,* 48 (2008), pp. 2081–2094 http://dx.doi.org/10.1021/ci800209k

Kozich et al., 2013. J.J. Kozich, S.L. Westcott, N.T. Baxter, S.K. Highlander, P.D. Schloss. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.,* 79 (17) (2013), pp. 5112–5120 http://dx.doi.org/10.1128/AEM.01043-13

Leitão et al., 2006. R.C. Leitão, A.C. van Haandel, G. Zeeman, G. Lettinga. The effects of operational and environmental variations on anaerobic wastewater treatment systems: a review. *Bioresour. Technol.,* 97 (9) (2006), pp. 1105–1118 http://dx.doi.org/10.1016/j.biortech.2004.12.007

Liu and Whitman, 2008. Y. Liu, W.B. Whitman. Metabolic, phylogenetic, and ecological diversity of the methanogenic *Archaea. Ann. N. Y. Acad. Sci.,* 1125 (2008), pp. 171–189 http://dx.doi.org/10.1196/annals.1419.019

Lü et al., 2009. F. Lü, L.M. Shao, V. Bru, J.J. Godon, P.J. He. Synergetic effect of pH and biochemical components on bacterial diversity during mesophilic anaerobic fermentation of biomass-origin waste. *J. Appl. Microbiol.,* 106 (2) (2009), pp. 580–591 http://dx.doi.org/10.1111/j.1365-2672.2008.04029.x

Mata-Alvarez et al., 2000. J. Mata-Alvarez, S. Macé, P. Llabrés. Anaerobic digestion of organic solid wastes. An overview of research achievements and perspectives. *Bioresour. Technol.,* 74 (1) (2000), pp. 3–16 http://dx.doi.org/10.1016/S0960-8524(00)00023-7

Morris et al., 2014. R. Morris, A.S. Gimenez, U. Bhattad, C. Kearney, C.A. Struble, D.H. Zitomer, J.S. Maki. Methyl coenzyme M reductase (mcrA) gene abundance correlates with activity measurements of methanogenic H_2/CO_2 enriched anaerobic biomass. *Microb. Biotechnol.,* 7 (1) (2014), pp. 77–84 http://dx.doi.org/10.1111/1751-7915.12094

Nirmalakhandan and Speece, 1988. N.N. Nirmalakhandan, R.E. Speece. Structure-activity relationships Quantitative techniques for predicting the behavior of chemicals in the ecosystem. *Environ. Sci. Technol.,* 22 (6) (1988), pp. 606–615 http://dx.doi.org/10.1021/es00171a002

Noike et al., 1985. T. Noike, G. Endo, J.E. Chang, J. Yaguchi, J. Matsumoto. Characteristics of carbohydrate degradation and the rate-limiting step in anaerobic digestion. *Biotechnol. Bioeng.,* 27 (10) (1985), pp. 1482–1489 http://dx.doi.org/10.1002/bit.260271013

Novotny et al., 2010. V. Novotny, J. Ahern, P. Brown. *Water Centric Sustainable Communities: Planning, Retrofitting and Building the Next Urban Environment*. John Wiley & Sons, Hoboken, NJ (2010) ISBN: 047064284X

Ogbonna et al., 2015. C.B. Ogbonna, D.P. Berebon, E.K. Onwuegbu. Relationship between temperature, pH and population of selected microbial indicators during anaerobic digestion of Guinea grass (*Panicum Maximum*). *Am. J. Microbiol. Res.,* 3 (1) (2015), pp. 14–24 http://dx.doi.org/10.12691/ajmr-3-1-3

Oksanen et al., 2016. J. Oksanen, F.G. Blanchet, R. Kindt, P. Legendre, P.R. Minchin, R.B. O'Hara, G.L. Simpson, P. Solymos, M. Henry, H. Stevens, H. Wagner. *Vegan: Community Ecology Package. R Package Version 2.3-4* (2016)

R Core Team, 2015. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2015) URL https://www.R-project.org/

Ramirez et al., 2009. I. Ramirez, E.I.P. Volcke, R. Rajinikanth, J.P. Steyer. Modeling microbial diversity in anaerobic digestion through an extended ADM1 model. *Water Res.,* 43 (11) (2009), pp. 2787–2800 http://dx.doi.org/10.1016/j.watres.2009.03.034

Resende et al., 2015. J.A. Resende, J.J. Godon, A. Bonnafous, P.B. Arcuri, V.L. Silva, M.H. Otenio, C.G. Diniz. Seasonal variation on microbial community and methane production during anaerobic digestion of cattle manure in Brazil. *Microb. Ecol.* (2015), pp. 735–746 http://dx.doi.org/10.1007/s00248-015-0647-y

Rivière et al., 2009. D. Rivière, V. Desvignes, E. Pelletier, S. Chaussonnerie, S. Guermazi, J. Weissenbach, T. Li, P. Camacho, A. Sghir. Towards the definition of a Core of microorganisms involved in anaerobic digestion of sludge. *ISME J.,* 3 (6) (2009), pp. 700–714 http://dx.doi.org/10.1038/ismej.2009.2

Schloss et al., 2009. P.D. Schloss, S.L. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, C.F. Weber. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.,* 75 (23) (2009), pp. 7537–7541 http://dx.doi.org/10.1128/AEM.01541-09

Schüürmann et al., 2008. G. Schüürmann, R.U. Ebert, J. Chen, B. Wang, R. Kühne. External validation and prediction employing the predictive squared correlation coefficient test set activity mean vs training set activity mean. *American Chemical Society J. Chem. Inf. Model.,* 48 (11) (2008), pp. 2140–2145 http://dx.doi.org/10.1021/ci800253u

Shah and Collins, 1988. H.N. Shah, M.D. Collins. Proposal for reclassification of Bacteroides asaccharolyticus, Bacteroides gingivalis, and Bacteroides endodontalis in a new genus, Porphyromonas. *Int. J. Syst. Bacteriol.,* 38 (1) (1988), pp. 128–131 http://dx.doi.org/10.1099/00207713-38-1-128

Shiratori et al., 2008. H. Shiratori, H. Ohiwa, H. Ikeno, S. Ayame, N. Kataoka, A. Miya, T. Beppu, K. Ueda. Lutispora thermophila gen. nov., sp. nov., a thermophilic, spore-forming bacterium isolated from a thermophilic methanogenic bioreactor digesting municipal solid wastes. *Int. J. Syst. Evol. Microbiol.,* 58 (4) (2008), pp. 964–969 http://dx.doi.org/10.1099/ijs.0.65490-0

Speece, 2008. R.E. Speece. *Anaerobic biotechnology and odor/corrosion control for municipalities and industries*. Fields Publishing, Inc, Nashville, TN (2008) ISBN: 1-57843-052-9

Stiles and Holzapfel, 1997. M.E. Stiles, W.H. Holzapfel. Lactic acid bacteria of foods and their current taxonomy. *Int. J. Food Microbiol.,* 36 (1) (1997), pp. 1–29 http://dx.doi.org/10.1016/S0168-1605(96)01233-0

Sundberg et al., 2013. C. Sundberg, W.A. Al-Soud, M. Larsson, E. Alm, S.S. Yekta, B.H. Svensson, S.J. Sørensen, A. Karlsson. 454 pyrosequencing analyses of *bacterial* and *archaeal* richness in 21 full-scale biogas digesters. *FEMS Microbiol. Ecol.,* 85 (3) (2013), pp. 612–626 http://dx.doi.org/10.1111/1574-6941.12148

Tale et al., 2011. V.P. Tale, J.S. Maki, C.A. Struble, D.H. Zitomer. Methanogen community structure-activity relationship and bioaugmentation of overloaded anaerobic digesters. *Water Res.,* 45 (16) (2011), pp. 5249–5256 http://dx.doi.org/10.1016/j.watres.2011.07.035

Tale et al., 2015. V.P. Tale, J.S. Maki, D.H. Zitomer. Bioaugmentation of overloaded anaerobic digesters restores function and archaeal community. *Water Res.,* 70 (2015), pp. 138–147 http://dx.doi.org/10.1016/j.watres.2014.11.037

Tian et al., 2014. Z. Tian, L. Cabrol, G. Ruiz-Filippi, P. Pullammanappallil. Microbial ecology in anaerobic digestion at agitated and non-agitated conditions. *PloS One,* 9 (10) (2014), p. e109769 http://dx.doi.org/10.1371/journal.pone.0109769

Treu et al., 2016. L. Treu, P.G. Kougias, S. Campanaro, I. Bassani, I. Angelidaki. Deeper insight into the structure of the anaerobic digestion microbial community; the biogas microbiome database is expanded with 157 new genomes. *Bioresour. Technol.,* 216 (2016), pp. 260–266 http://dx.doi.org/10.1016/j.biortech.2016.05.081 September

van Loosdrecht and Brdjanovic, 2014. M.C.M. van Loosdrecht, D. Brdjanovic. Water treatment. Anticipating the next century of wastewater treatment. *Science*, 344 (6191) (2014), pp. 1452–1453 http://dx.doi.org/10.1126/science.1255183

Venkiteshwaran et al., 2015. K. Venkiteshwaran, B. Bocher, J. Maki, D. Zitomer. Relating anaerobic digestion microbial community and process function. *Libertas Academica Microbiol. Insights*, 8 (Suppl. 2) (2015), pp. 37–44 http://dx.doi.org/10.4137/MBI.S33593

Venkiteshwaran et al., 2016. K. Venkiteshwaran, K. Milferstedt, J. Hamelin, D. Zitomer. Anaerobic digester bioaugmentation influences quasi steady state performance and microbial community. *Water Res.,* 104 (2016), pp. 128–136 http://dx.doi.org/10.1016/j.watres.2016.08.012

Vidal, 2000. G. Vidal. Influence of the content in fats and proteins on the anaerobic biodegradability of dairy wastewaters. *Bioresour. Technol.,* 74 (3) (2000), pp. 231–239 http://dx.doi.org/10.1016/S0960-8524(00)00015-8

Wang and Qian, 2009. Y. Wang, P.Y. Qian. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PloS One,* 4 (10) (2009), p. e7401 http://dx.doi.org/10.1371/journal.pone.0007401

Werner et al., 2011. J.J. Werner, D. Knights, M.L. Garcia, N.B. Scalfone, Samual Smith, Kevin Yarasheski, Theresa a Cummings, Allen R. Beers, Rob Knight, Largus T. Angenent. Bacterial community structures are unique and resilient in full-scale bioenergy systems. *Proc. Natl. Acad. Sci. U. S. A.,* 108 (10) (2011), pp. 4158–4163 http://dx.doi.org/10.1073/pnas.1015676108

Westermann et al., 1989. P. Westermann, B.K. Ahring, R.A. Mah. Threshold acetate concentrations for acetate catabolism by aceticlastic methanogenic bacteria. *Appl. Environ. Microbiol.,* 55 (2) (1989), pp. 514–515 ISSN: 0099-2240

Zuur et al., 2007. A. Zuur, E.N. Ieno, G.M. Smith. *Analysing Ecological Data*. Springer Science & Business Media, New York, NY (2007) ISBN: 0387459723

Appendix A. Supplementary data

The following are the supplementary data related to this article:

Table S1: Seed biomass source data

Digester ID	Waste treated at seed source	Digester configuration*
Alaska (AK)	Municipal sludge	CSTR
Alabama (AL)	Petrochemical	UASB
Arkansas (AR)	Food waste	UASB
Arizona (AZ)	Municipal sludge	CSTR
California (CA)	Winery	CSTR
Colorado (CO)	Brewery	UASB
Connecticut (CT)	Municipal sludge	CSTR
Delaware (DE)	Municipal & industrial WW mix	CSTR
Florida (FL)	Municipal sludge	CSTR
Georgia (GA)	Municipal sludge	CSTR
Hawaii (HI)	Municipal sludge	CSTR
Iowa (IA)	Confections manufacture	CSTR
Idaho (ID)	Ethanol	UASB
Illinois (IL)	Food & beverage	CSTR
Indiana (IN)	Corn mill	CSTR
Kansas (KS)	Soda bottling	UASB
Kentucky (KY)	Cracker & cereal	CSTR
Louisiana (LA)	Food waste	CSTR
Massachusetts (MA)	Food waste	CSTR
Maryland (MD)	Yeast	CSTR
Maine (ME)	Municipal sludge & industrial WW mix	CSTR
Michigan (MI)	Municipal sludge & paper	CSTR
Minnesota (MN)	Paper	CSTR
Missouri (MO)	Food waste	CSTR
Mississippi (MS)	Municipal sludge	CSTR
Montana (MT)	Municipal sludge	CSTR
North Carolina (NC)	Municipal sludge	CSTR
North Dakota (ND)	Beet sugar & yeast	CSTR
Nebraska (NE)	Municipal sludge	CSTR
New Hampshire (NH)	Dairy	CSTR
New Jersey (NJ)	Food waste	CSTR
New Mexico (NM)	Dairy	CSTR
Nevada (NV)	Municipal sludge	CSTR
New York (NY)	Dairy	CSTR
Ohio (OH)	Municipal sludge	CSTR
Oklahoma (OK)	Soybean process waste	CSTR
Oregon (OR)	Municipal sludge	CSTR
Pennsylvania (PA)	Dairy	CSTR
South Carolina (SC)	Municipal sludge & fruit juice	CSTR

South Dakota (SD)	Municipal sludge	CSTR
Tennessee (TN)	Municipal sludge	CSTR
Texas (TX)	Cheese whey	AnMBR
Utah (UT)	Municipal sludge	CSTR
Virginia (VA)	Municipal sludge	CSTR
Vermont (VT)	Brewery	CSTR
Washington (WA)	Municipal sludge	CSTR
Wisconsin A (WI A)	Brewery	UASB
Wisconsin B (WI B)	Municipal sludge	CSTR
West Virginia (WV)	Municipal sludge	CSTR
Wyoming (WY)	Municipal sludge	CSTR

*CSTR – Completely stirred type reactor, UASB – Upflow anaerobic sludge blanket, AnMBR – Anaerobic membrane reactor

Table S2: Observed methane production (L-CH4/LR-day) rate in 149 digesters. The functional parameters of the digesters were analyzed for 7 consecutive days and the table reports the average and standard deviation of those 7 data points. The digesters were classified in to three equally sized group (High, Medium and Low) based on their methane production rate. 50 highest methane producing digesters were classified as "High", the next 50 digesters were classified as "Medium" and the remaining lowest methane producing digesters were classified as "Low".

Digester ID	Methane production rate (L-CH4/LR-day)		Classification	Digester ID	Methane production rate (L-CH4/LR-day)		Classification
	Average	Std. dev			Average	Std. dev	
AK1	0.32	0.02	Medium	MT1	0.59	0.01	High
AK2	0.33	0.03	Medium	MT2	0.70	0.08	High
AK3	0.39	0.03	Medium	MT3	0.56	0.03	High
AL1	0.43	0.02	Medium	NC1	0.37	0.02	Medium
AL2	0.49	0.03	Medium	NC2	0.21	0.01	Low
AL3	0.54	0.03	High	NC3	0.64	0.03	High
AR1	0.24	0.03	Low	ND2	0.55	0.02	High
AR2	0.24	0.03	Low	ND3	0.52	0.02	Medium
AR3	0.27	0.03	Medium	NE1	0.43	0.01	Medium
AZ1	0.50	0.03	Medium	NE2	0.46	0.02	Medium
AZ2	0.40	0.04	Medium	NE3	0.36	0.01	Medium
AZ3	0.35	0.05	Medium	NH1	0.15	0.01	Low
CA1	0.65	0.06	High	NH2	0.15	0.01	Low
CA2	0.73	0.06	High	NH3	0.14	0.02	Low
CA3	0.66	0.06	High	NJ1	0.12	0.01	Low
CO1	0.63	0.05	High	NJ2	0.13	0.01	Low
CO2	0.60	0.04	High	NJ3	0.11	0.01	Low
CO3	0.48	0.04	Medium	NM1	0.50	0.07	Medium
CT1	0.17	0.02	Low	NM2	0.58	0.03	High
CT2	0.28	0.12	Medium	NM3	0.63	0.05	High
CT3	0.48	0.12	Medium	NV1	0.13	0.02	Low
DE1	0.15	0.01	Low	NV2	0.13	0.01	Low
DE2	0.16	0.01	Low	NV3	0.13	0.01	Low
DE3	0.29	0.02	Medium	NY1	0.17	0.02	Low
FL1	0.69	0.05	High	NY2	0.16	0.02	Low
FL2	0.67	0.01	High	NY3	0.18	0.01	Low
FL3	0.67	0.01	High	OH1	0.41	0.03	Medium
GA1	0.32	0.04	Medium	OH2	0.39	0.05	Medium
GA2	0.13	0.02	Low	OH3	0.38	0.04	Medium
GA3	0.12	0.01	Low	OK1	0.66	0.04	High
HI1	0.38	0.02	Medium	OK2	0.73	0.06	High
HI2	0.61	0.04	High	OK3	0.81	0.05	High
HI3	0.36	0.02	Medium	OR1	0.26	0.02	Low

IA1	0.57	0.04	High	OR2	0.58	0.04	High
IA2	0.58	0.04	High	OR3	0.41	0.03	Medium
IA3	0.62	0.06	High	PA1	0.63	0.08	High
ID1	0.30	0.02	Medium	PA2	0.52	0.03	Medium
ID2	0.33	0.03	Medium	PA3	0.71	0.10	High
ID3	0.32	0.02	Medium	SC1	0.50	0.09	Medium
IL1	0.63	0.07	High	SC2	0.51	0.02	Medium
IL2	0.67	0.07	High	SC3	0.68	0.06	High
IL3	0.67	0.06	High	SD1	0.20	0.03	Low
IN1	0.56	0.04	High	SD2	0.17	0.02	Low
IN2	0.49	0.04	Medium	SD3	0.25	0.02	Low
IN3	0.52	0.04	Medium	TNS1	0.48	0.05	Medium
KS1	0.98	0.05	High	TNS2	0.63	0.12	High
KS2	0.92	0.06	High	TNS3	0.63	0.04	High
KS3	0.56	0.04	High	TX1	0.50	0.04	Medium
KY1	0.69	0.07	High	TX2	0.35	0.01	Medium
KY2	0.57	0.07	High	TX3	0.45	0.02	Medium
KY3	0.61	0.05	High	UT1	0.09	0.004	Low
LA1	0.20	0.02	Low	UT2	0.10	0.01	Low
LA2	0.43	0.02	Medium	UT3	0.11	0.01	Low
LA3	0.57	0.01	High	VA1	0.25	0.01	Low
MA1	0.35	0.04	Medium	VA2	0.13	0.01	Low
MA2	0.13	0.03	Low	VA3	0.13	0.01	Low
MA3	0.15	0.02	Low	VT1	0.25	0.03	Low
MD1	0.54	0.07	High	VT2	0.31	0.05	Medium
MD2	0.61	0.08	High	VT3	0.27	0.03	Low
MD3	0.60	0.05	High	WA1	0.63	0.08	High
ME1	0.16	0.02	Low	WA2	0.57	0.08	High
ME2	0.15	0.01	Low	WA3	0.59	0.06	High
ME3	0.16	0.02	Low	WI A1	0.38	0.02	Medium
MI1	0.23	0.02	Low	WI A2	0.56	0.04	High
MI2	0.23	0.03	Low	WI A3	0.50	0.04	Medium
MI3	0.23	0.02	Low	WI B1	0.51	0.02	Medium
MN1	0.25	0.03	Low	WI B2	0.57	0.01	High
MN2	0.26	0.03	Low	WI B3	0.62	0.04	High
MN3	0.25	0.02	Low	WV1	0.23	0.02	Low
MO1	0.37	0.05	Medium	WV2	0.23	0.02	Low
MO2	0.33	0.02	Medium	WV3	0.19	0.01	Low
MO3	0.47	0.05	Medium	WY1	0.26	0.01	Low
MS1	0.54	0.05	High	WY2	0.32	0.04	Medium
MS2	0.47	0.12	Medium	WY3	0.13	0.01	Low

MS3	0.44	0.03	Medium					

Table S3: Highly significant OTUs determined by initial screening.
50 iterations of Spearman's rank analysis were performed, where 75 out of 149 digesters where randomly selected and correlated with digester methane production rate. The value column "*N*" represents the number of times the OTU was observed as the top ten positively or negatively correlated out of the total 50 iterations. *N* = 50 represents 100% observation. The 7 OTUs in Table S3A and 3 OTUs in Table S3B, were observed to be positively and negatively correlated to methane production rate in 38 out of the 50 iterations (>75% of the iterations). These 10 OTUs were selected for the subsequent MLR analysis.

(A) Relative abundance positively correlated to methane production					
OTU ID	**Class**	**Order**	**Family**	**Genus**	***N***
1	*Bacteroidia*	*Bacteroidales*	*Porphyromonadaceae*	*Petrimonas*	50
2	*Bacteroidia*	*Bacteroidales*	*Marinilabiaceae*	*unclassified*	50
3	*Spirochaetes*	*Spirochaetales*	*PL-11B10*	*unclassified*	50
4	*Bacteroidia*	*Bacteroidales*	*M2PB4-65 termite group*	*unclassified*	46
5	*Clostridia*	*Clostridiales*	*Syntrophomonadaceae*	*Syntrophomonas*	46
6	*Methanomicrobia*	*Methanosarcinales*	*Methanosarcinaceae*	*Methanosarcina*	46
7	*Clostridia*	*Clostridiales*	*Gracilibacteraceae*	*Lutispora*	40

(B) Relative abundance negatively correlated to methane production					
OTU ID	**Class**	**Order**	**Family**	**Genus**	**N**
8	*Bacteroidia*	*Bacteroidales*	*Porphyromonadaceae*	*Porphyromonas*	50
9	*Clostridia*	*Clostridiales*	*Ruminococcaceae*	*unclassified*	50
10	*Clostridia*	*Clostridiales*	*Lachnospiraceae*	*unclassified*	50

Table S4: **R square values from the cross correlation of the highly significant OTUs.**

	OTU1	OTU2	OTU3	OTU4	OTU5	OTU6	OTU7	OTU8	OTU9	OTU10
OTU1	**1**	0.19	0.06	0.096	0.087	0.23	0.02	0.18	0.077	0.093
OTU2	0.19	**1**	0.04	0.003	0.03	0.17	0.012	0.069	0.033	0.044
OTU3	0.06	0.04	**1**	0.022	0.059	0.17	0.038	0.078	0.042	0.045
OTU4	0.096	0.003	0.022	**1**	0.03	0.02	0.001	0.07	0.047	0.03
OTU5	0.087	0.03	0.059	0.03	**1**	0.17	0.34	0.044	0.03	0.05
OTU6	0.23	0.17	0.17	0.02	0.17	**1**	0.012	0.13	0.07	0.07
OTU7	0.02	0.012	0.038	0.001	0.34	0.012	**1**	0.013	0.007	0.0067
OTU8	0.18	0.069	0.078	0.07	0.044	0.13	0.013	**1**	0.047	0.003
OTU9	0.077	0.033	0.042	0.047	0.03	0.07	0.007	0.047	**1**	0.088
OTU10	0.093	0.044	0.045	0.03	0.05	0.07	0.0067	0.003	0.088	**1**

Table S5: Test and training groups for the 10 validation tests.
Validation tests indicating (A) test and training groups employed and (B) identities of digesters employed for each iteration.

(A)		
Validation test number	**Test set group number**	**Training set group numbers**
1	10	1 to 9
2	9	1 to 8 & 10
3	8	1 to 7,9 & 10
4	7	1 to 6 & 8 to 10
5	6	1 to 5 & 7 to 10
6	5	1 to 4 & 6 to 10
7	4	1 to 3 & 5 to 10
8	3	1, 2 & 4 to 10
9	2	1 & 3 to 10
10	1	2 to 10

(B)									
Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**	**Group 6**	**Group 7**	**Group 8**	**Group 9**	**Group 10**
Digester ID									
AL3	ID2	AL1	CT1	AK2	AK1	CO2	AL2	NJ1	AR1
CT2	ID3	FL2	GA3	CT3	AK3	HI3	AR3	NY3	AR2
FL1	MD3	ID1	HI2	DE2	AZ2	IL3	IN3	OH2	AZ1
IL1	MO1	KS3	IA3	IN1	CA2	KY2	KS2	OH3	AZ3
IL2	MO2	KY1	KS1	IN2	CA3	LA3	KY3	OK1	CA1
MI3	NE2	ME2	LA1	LA2	CO1	MD1	MA1	OK3	DE1
NC2	NE3	MI2	MI1	MA3	CO3	MS3	ME1	OR1	DE3
NE1	NH1	MT3	MN2	MD2	GA1	NC1	MT1	SD2	FL3
OH1	OK2	NJ3	NC3	MN1	GA2	ND2	MT2	VA2	HI1
SC2	OR2	NY1	ND3	MO3	MA2	NV2	NM3	VA3	IA1
TX3	TNS1	SC3	NH3	MS1	MS2	NY2	PA2	VT2	IA2
UT3	TNS3	TNS2	NM2	NJ2	NM1	PA1	SD1	WA1	ME3
VT1	UT1	WI A2	NV3	NV1	TX1	PA3	SD3	WI B1	MN3
WV2	WA2	WI B3	SC1	OR3	TX2	UT2	VA1	WV1	NH2
WY3	WI B2	WY1	WY2	WI A3	VT3	WA3	WI A1	WV3	-

Table S6: **Summary table of the 10 validation tests with the results of the four validation criteria.**

Validation Test no.	q^2	R^2	$(R^2-R_0^2)/R^2$	K
1	0.65	0.68	0.04	1.0
2	0.69	0.83	0.02	0.86
3	0.68	0.68	0.0	1.03
4	0.52	0.62	0.05	1.11
5	0.58	0.67	0.01	0.90
6	0.22	0.35	0.26	1.09
7	0.57	0.68	0.01	1.01
8	0.64	0.74	0.09	0.93
9	0.54	0.65	0.09	1.09
10	0.65	0.66	0.0	0.97

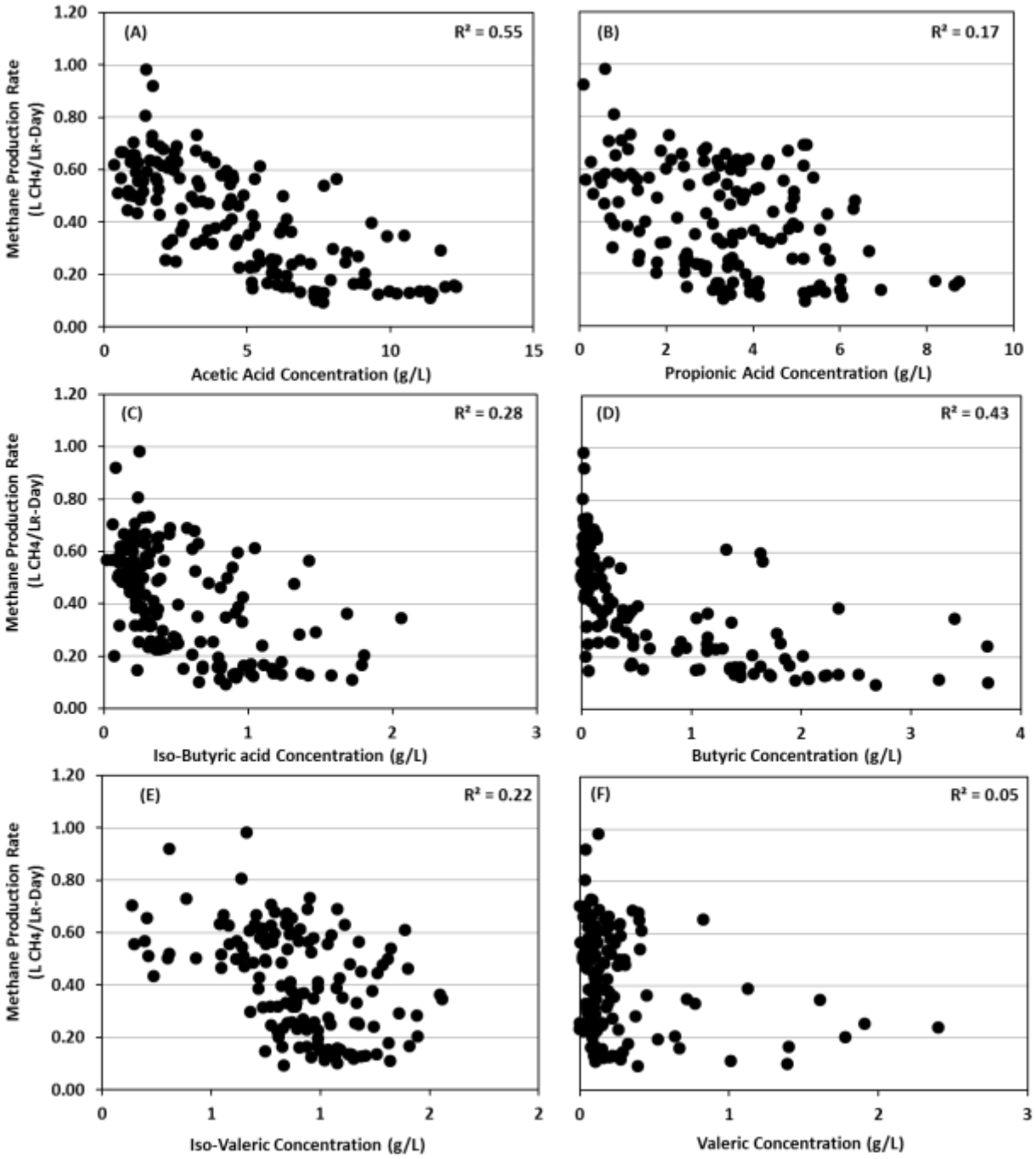


Figure S1: Average daily methane production versus individual VFA concentrations. Average daily methane production (L-CH_4/L_R-day) versus (A) acetic acid, (B) propionic acid, (C) iso-butyric acid, (D) butyric acid, (E) iso-valeric acid and (F) valeric acid concentration (g/L). Error bars are not included.

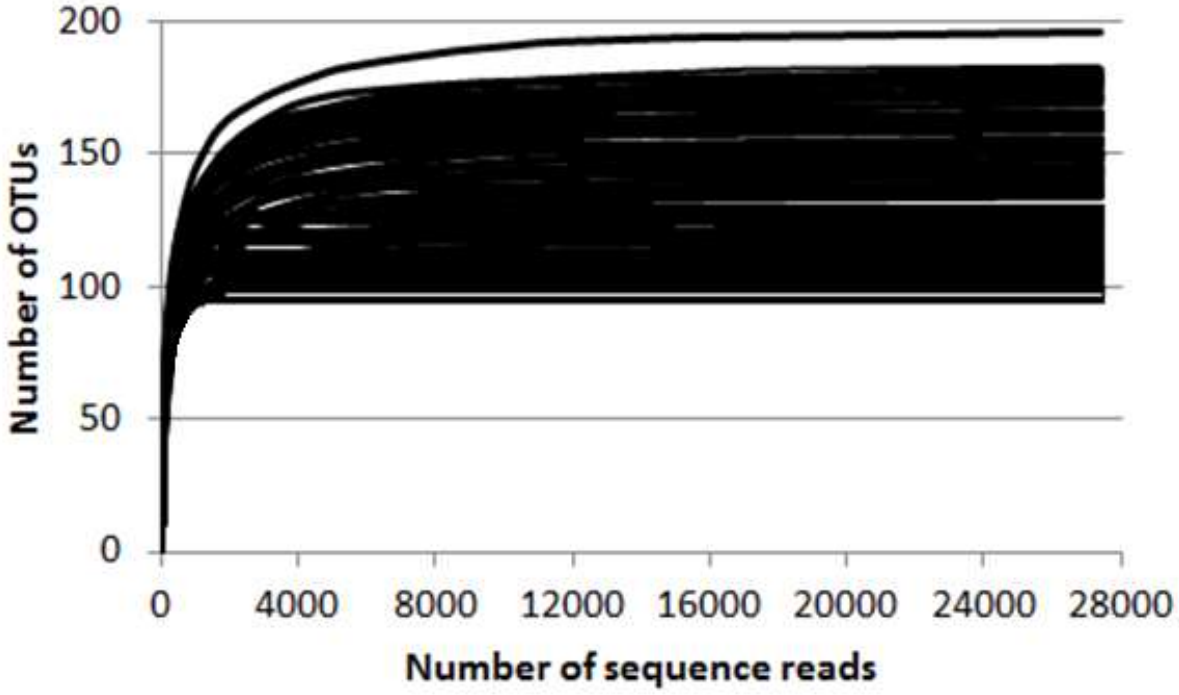
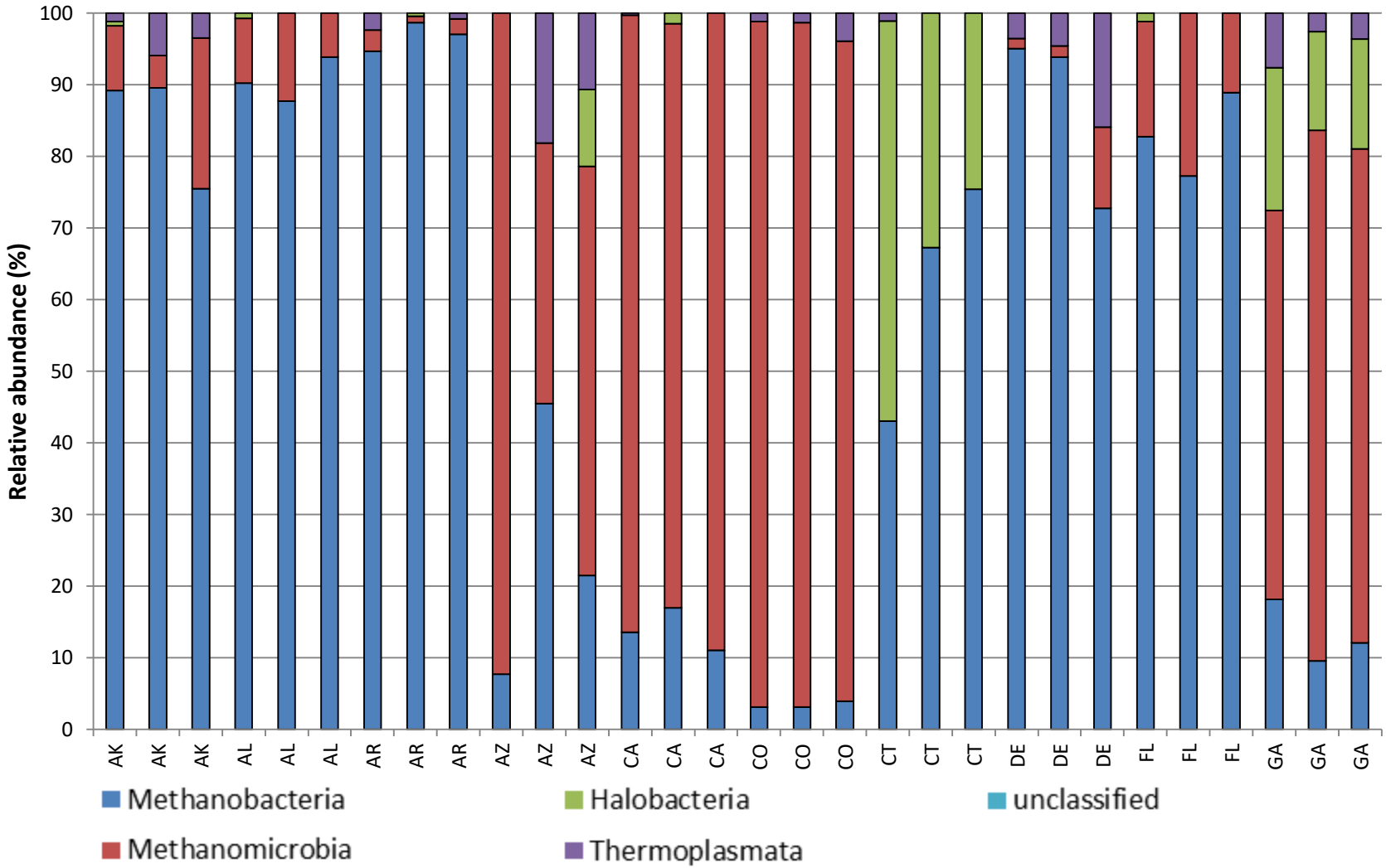
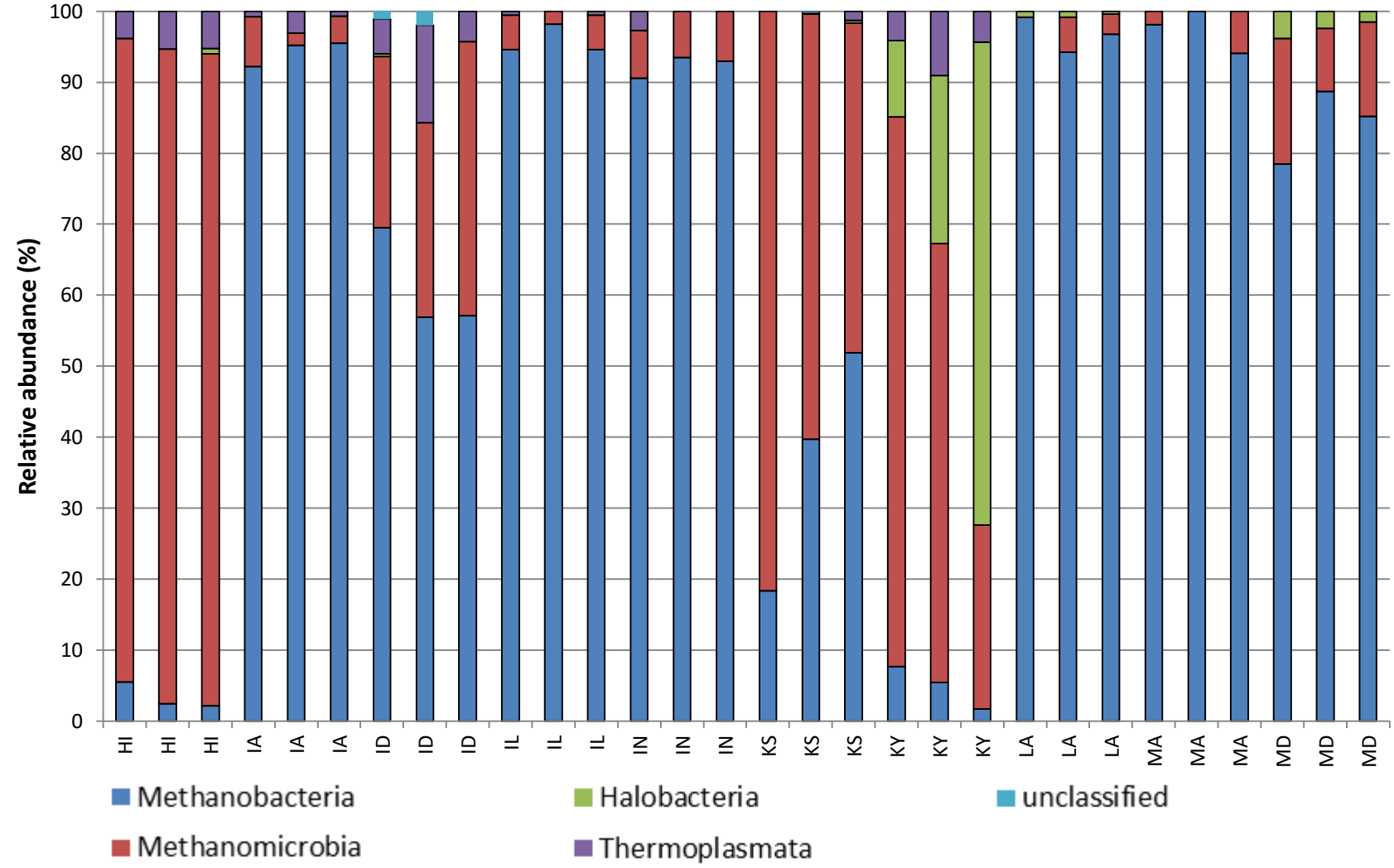


Figure S2: Rarefaction analysis for the assessment of OTU coverage.
Plot shows 149 rarefaction curves showing the increase in OTU numbers (Y axis) as a function of the number of sequence reads (X axis). The number of sequence reads go up to 27, 315, which is the lowest observed among the 149 digester samples The curve becomes asymptotic as the OTU number saturates, and increasing the reads adds an increasingly smaller number of new OTUs, indicating adequate coverage for the samples tested.

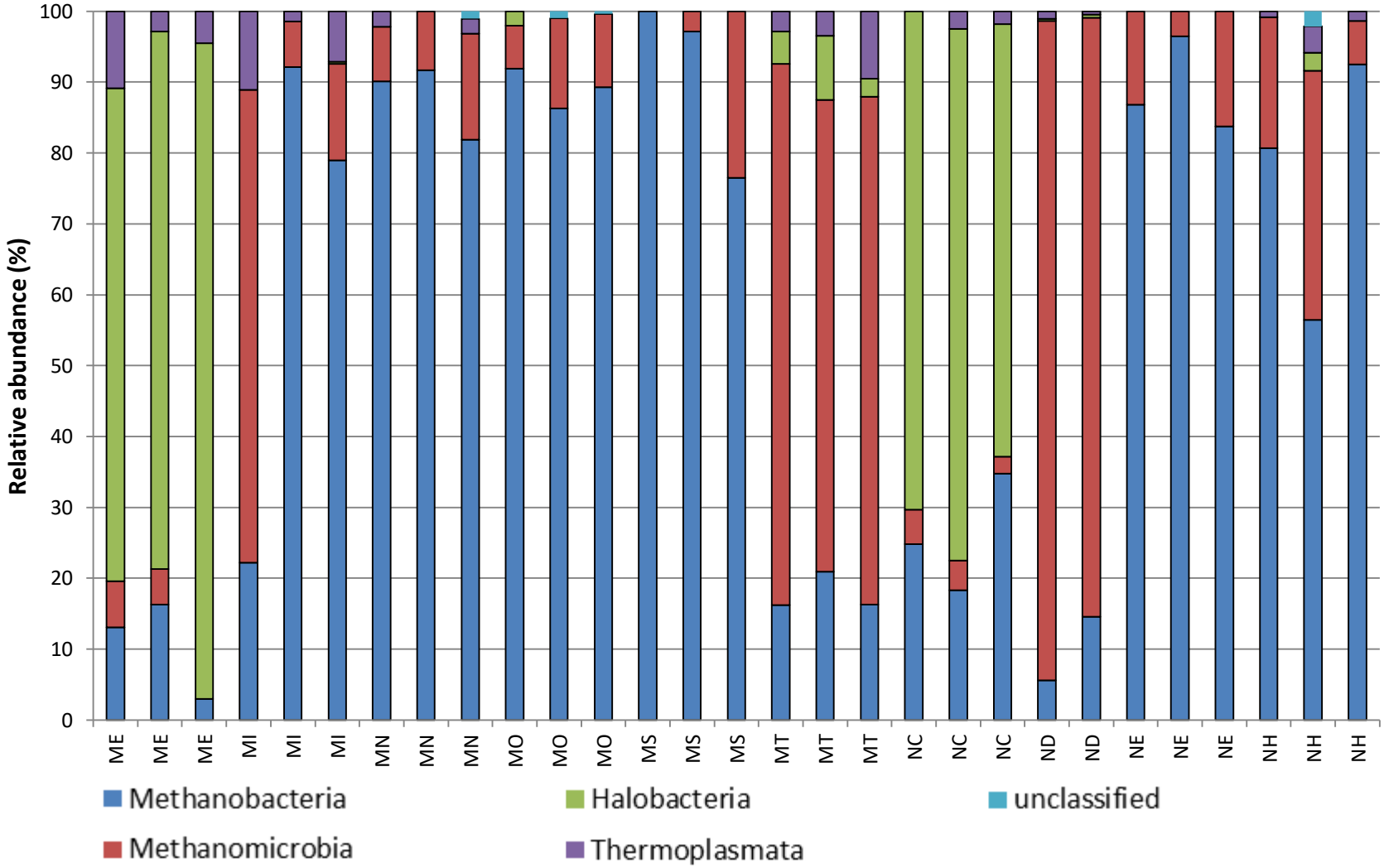
(A)


Relative abundance (%)
100
90
80
70
60
50
40
30
20
10
0
AK AK AK AL AL AL AR AR AR AZ AZ AZ CA CA CA CO CO CO CT CT CT DE DE DE FL FL FL GA GA GA
Methanobacteria
Halobacteria
unclassified
Methanomicrobia
Thermoplasmata

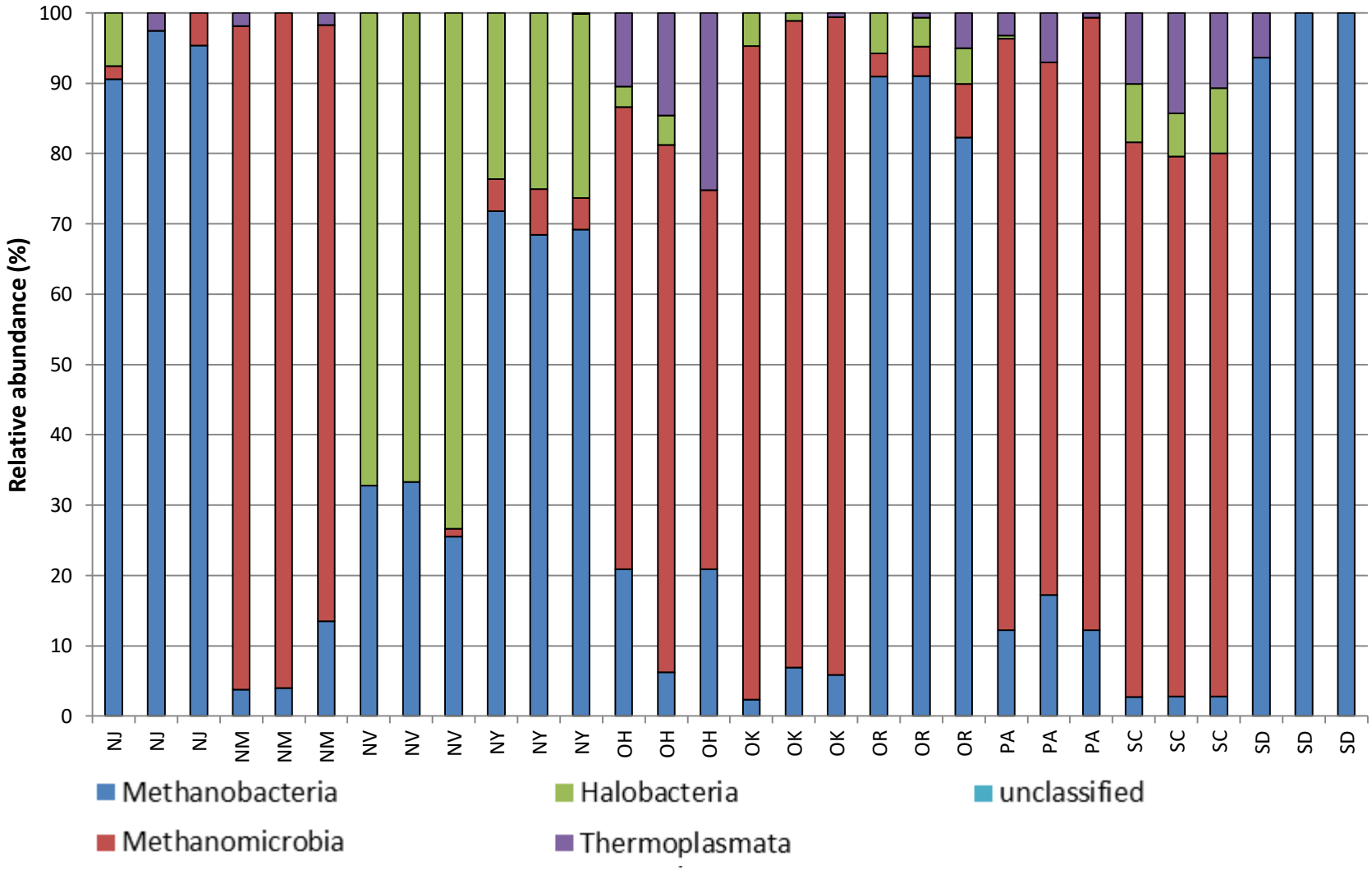
(B)


Relative abundance (%)
100
90
80
70
60
50
40
30
20
10
0
HI
HI
HI
IA
IA
IA
ID
ID
ID
IL
IL
IL
IN
IN
IN
KS
KS
KS
KY
KY
KY
LA
LA
LA
MA
MA
MA
MD
MD
MD
Methanobacteria
Halobacteria
unclassified
Methanomicrobia
Thermoplasmata

(C)


Relative abundance (%)
100
90
80
70
60
50
40
30
20
10
0
ME ME ME MI MI MI MN MN MN MO MO MO MS MS MS MT MT MT NC NC NC ND ND NE NE NE NH NH NH
Methanobacteria
Halobacteria
unclassified
Methanomicrobia
Thermoplasmata

(D)


Relative abundance (%)
100
90
80
70
60
50
40
30
20
10
0
NJ
NJ
NJ
NM
NM
NM
NV
NV
NV
NY
NY
NY
OH
OH
OH
OK
OK
OK
OR
OR
OR
PA
PA
PA
SC
SC
SC
SD
SD
SD
Methanobacteria
Halobacteria
unclassified
Methanomicrobia
Thermoplasmata

(E)

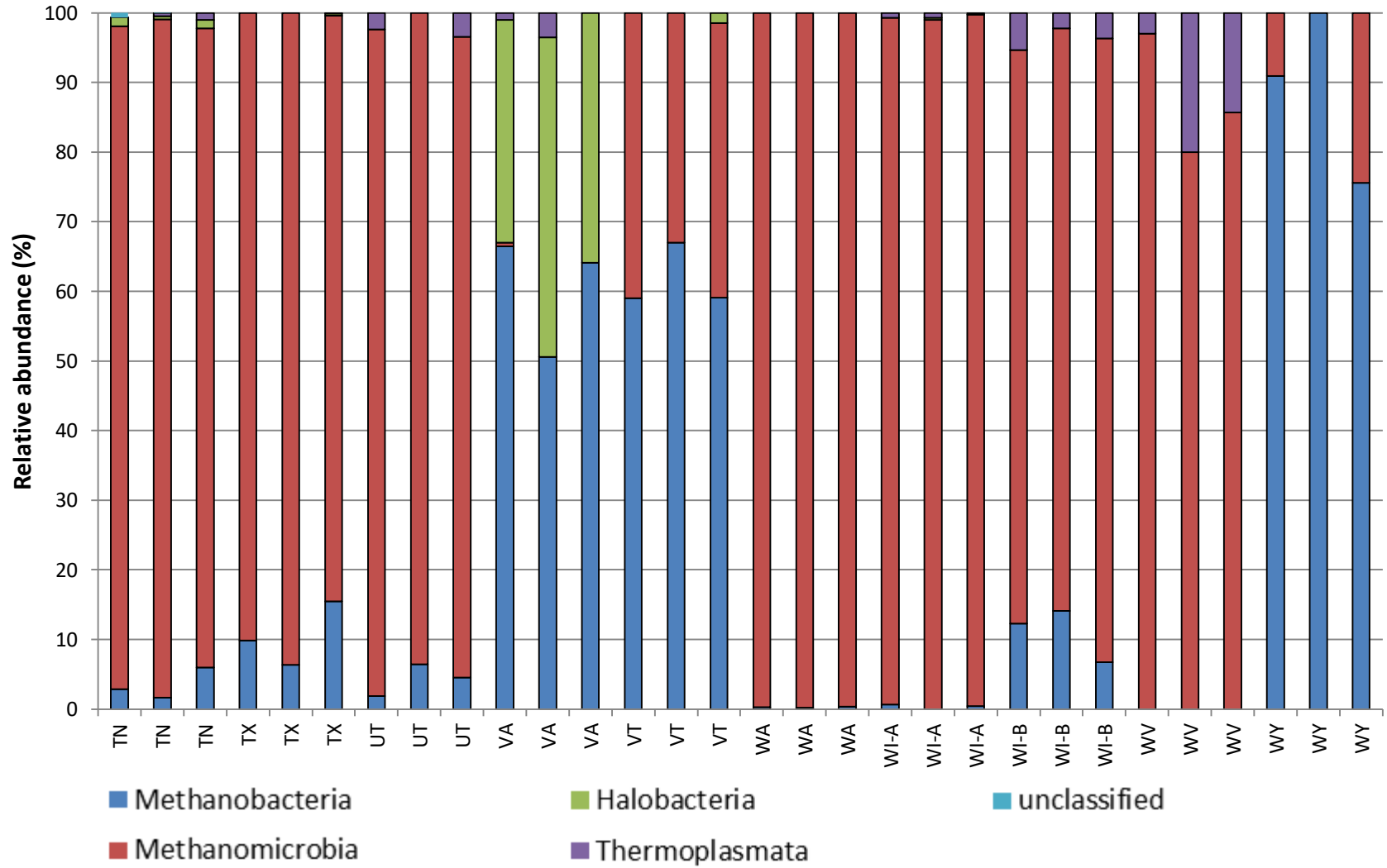

Relative abundance (%)
100
90
80
70
60
50
40
30
20
10
0
TN
TN
TN
TX
TX
TX
UT
UT
UT
VA
VA
VA
VT
VT
VT
WA
WA
WA
WI-A
WI-A
WI-A
WI-B
WI-B
WI-B
WV
WV
WV
WY
WY
WY
Methanobacteria
Halobacteria
unclassified
Methanomicrobia
Thermoplasmata

Figure S3: *Archaeal* community profile of 149 digesters at the class level. The profiles of 149 digesters are divided in to five stacked bar graphs A, B, C, D and E, respectively. The Y axis represents the relative abundance (%) of *Archaeal* classes observed of the total *Archaeal* sequences in a digester sample. The digester sample on the X axis are labelled based on the US state(s) from where the seed source was obtained (i.e. AK = Alaska, HI = Hawaii etc.). Replicate digesters seeded from the same seed are given the same labels.

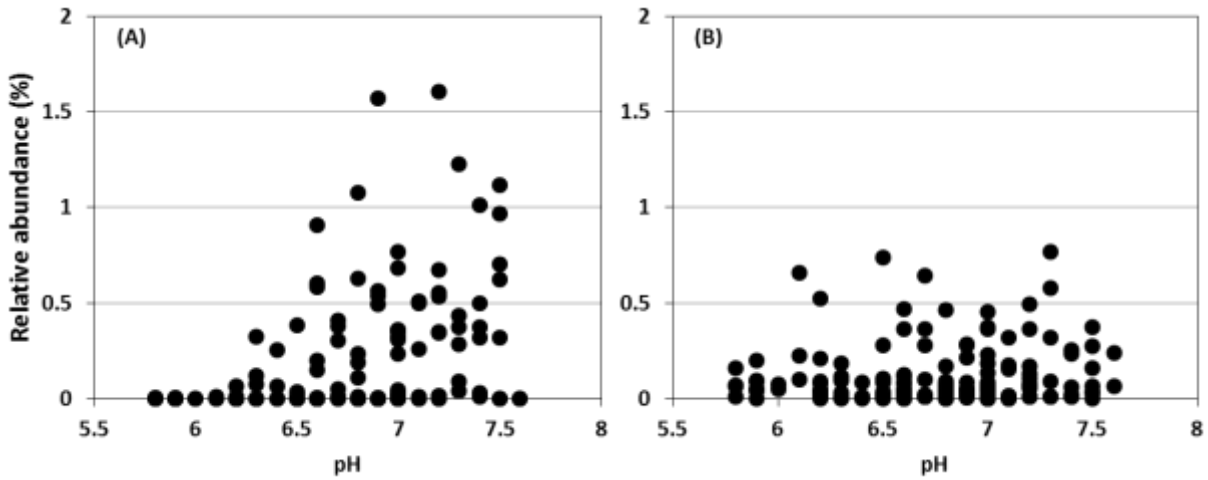
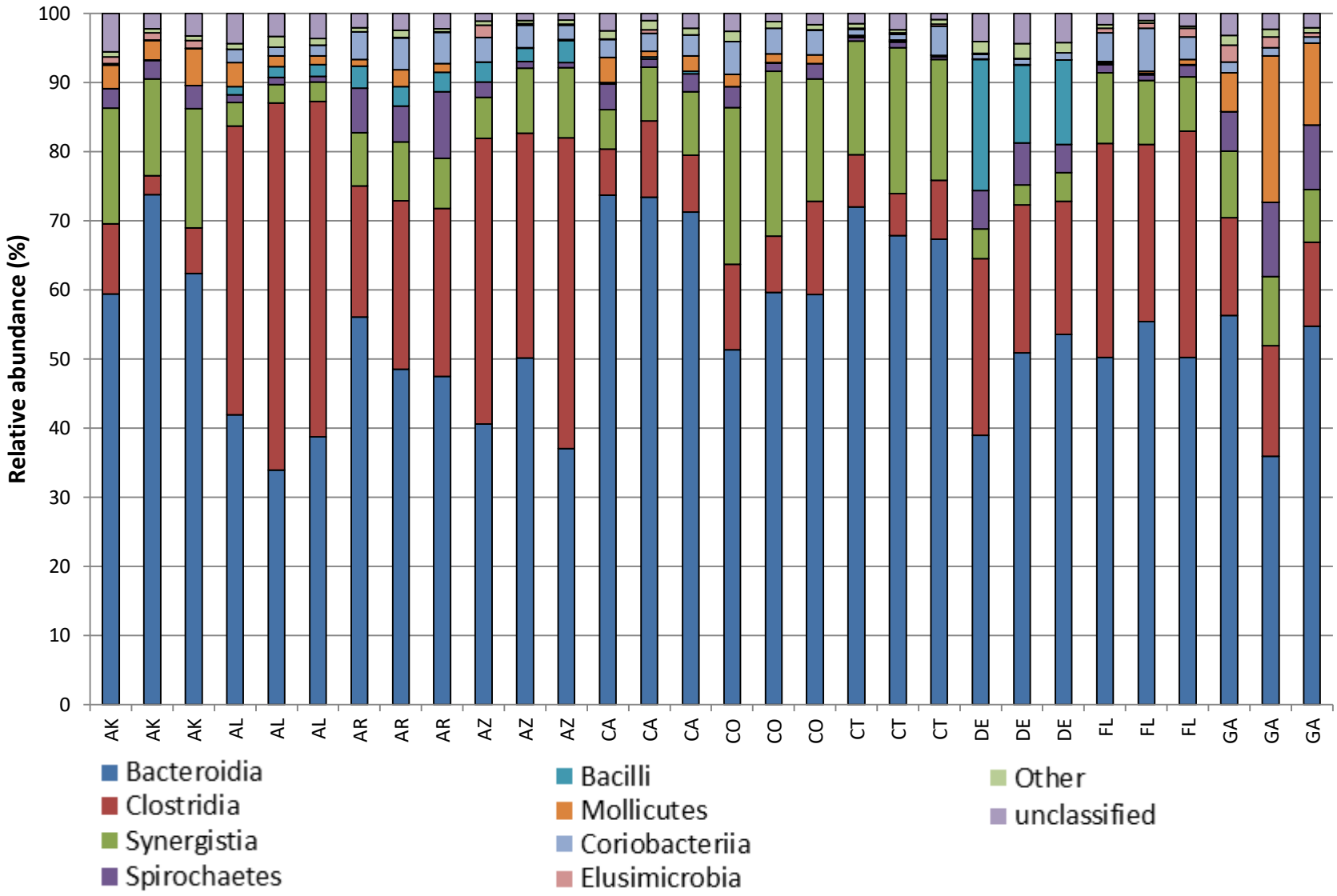
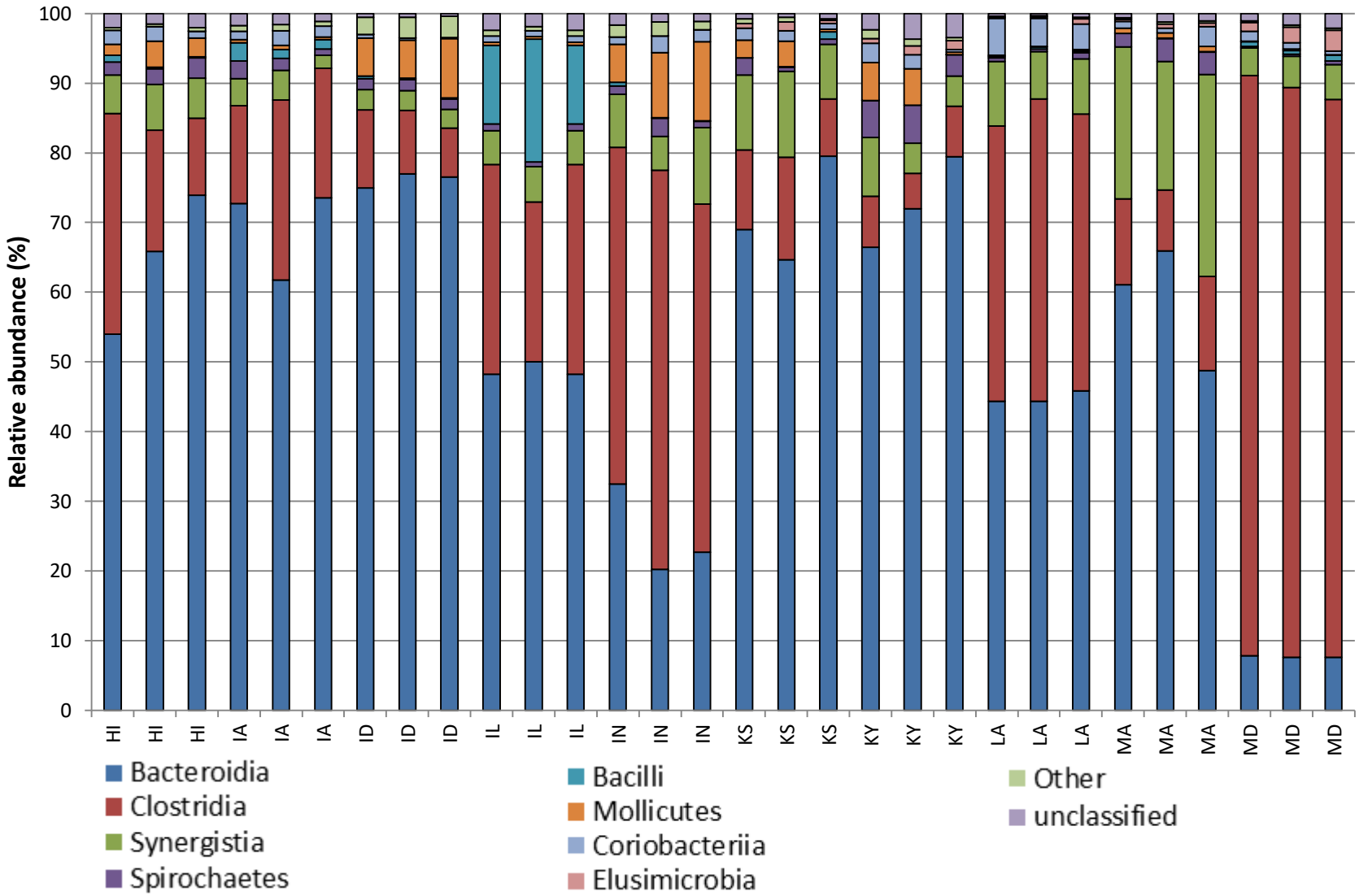


Figure S4: Percent relative abundance of dominant methanogens versus digester pH.
Percent relative abundance of (A) *Methanosarcina* and (B) *Methanobacterium* versus digester pH.

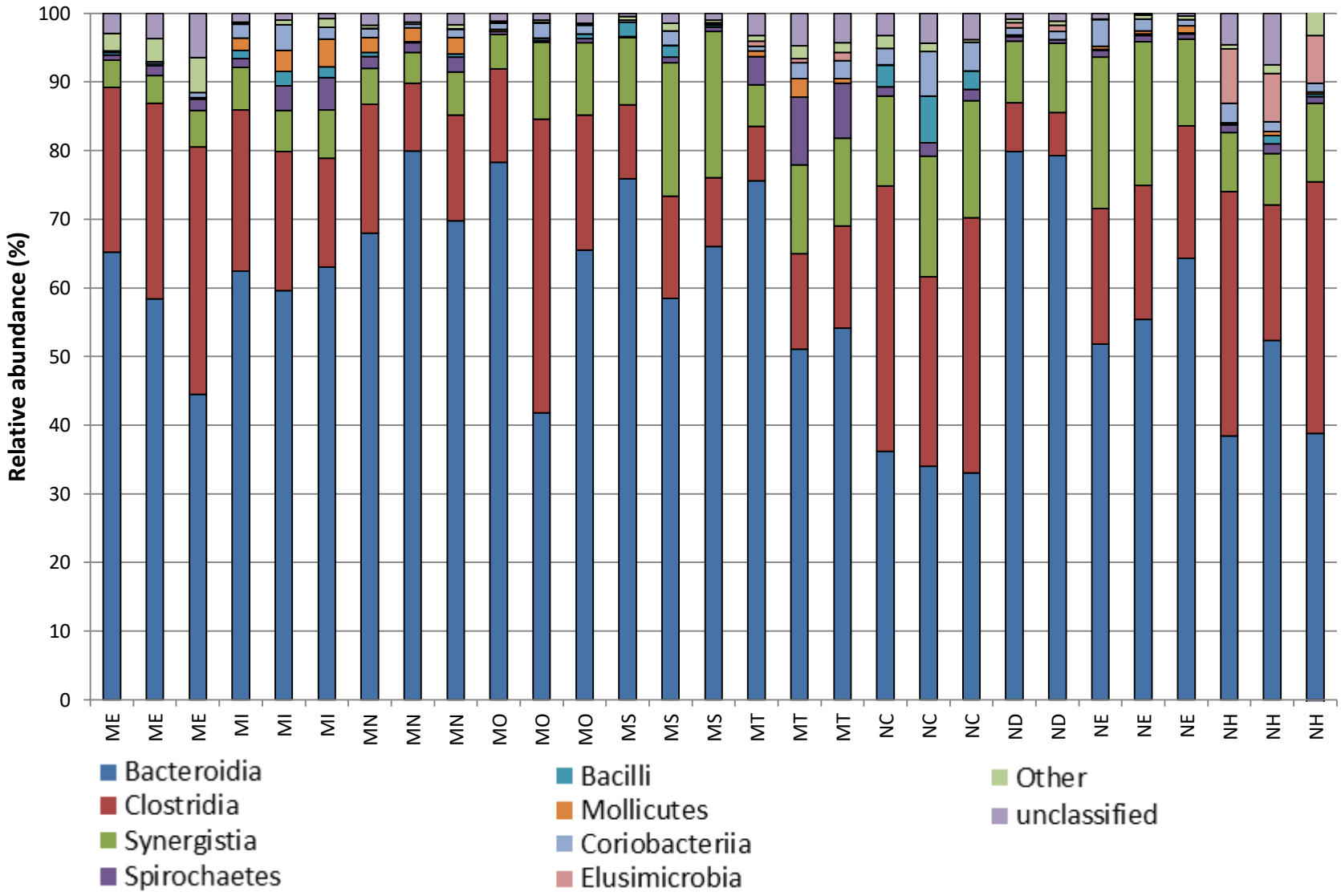
(A)


Relative abundance (%)
100
90
80
70
60
50
40
30
20
10
0
AK AK AK AL AL AL AR AR AR AZ AZ AZ CA CA CA CO CO CO CT CT CT DE DE DE FL FL FL GA GA GA
Bacteroidia
Clostridia
Synergistia
Spirochaetes
Bacilli
Mollicutes
Coriobacteriia
Elusimicrobia
Other
unclassified

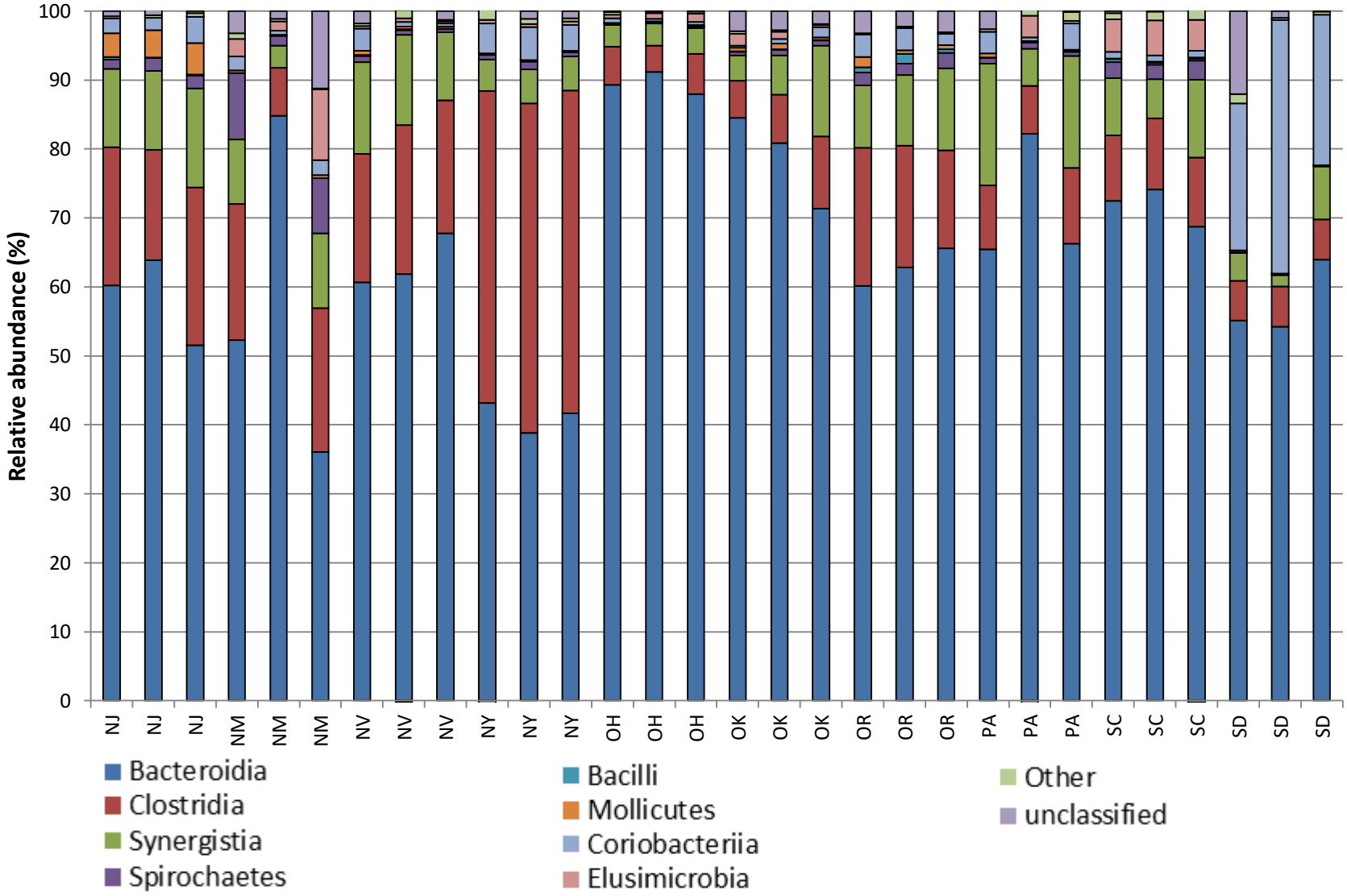
(B)


Relative abundance (%)
100
90
80
70
60
50
40
30
20
10
0
HI HI HI IA IA IA ID ID ID IL IL IL IN IN IN KS KS KS KY KY KY LA LA LA MA MA MA MD MD MD
Bacteroidia
Clostridia
Synergistia
Spirochaetes
Bacilli
Mollicutes
Coriobacteriia
Elusimicrobia
Other
unclassified

(C)


Relative abundance (%)
100
90
80
70
60
50
40
30
20
10
0
ME ME ME MI MI MI MN MN MN MO MO MO MS MS MS MT MT MT NC NC NC ND ND NE NE NE NH NH NH
Bacteroidia
Clostridia
Synergistia
Spirochaetes
Bacilli
Mollicutes
Coriobacteriia
Elusimicrobia
Other
unclassified

(D)


Relative abundance (%)
100
90
80
70
60
50
40
30
20
10
0
NJ
NJ
NJ
NM
NM
NM
NV
NV
NV
NY
NY
NY
OH
OH
OH
OK
OK
OK
OR
OR
OR
PA
PA
PA
SC
SC
SC
SD
SD
SD
Bacteroidia
Clostridia
Synergistia
Spirochaetes
Bacilli
Mollicutes
Coriobacteriia
Elusimicrobia
Other
unclassified

(E)

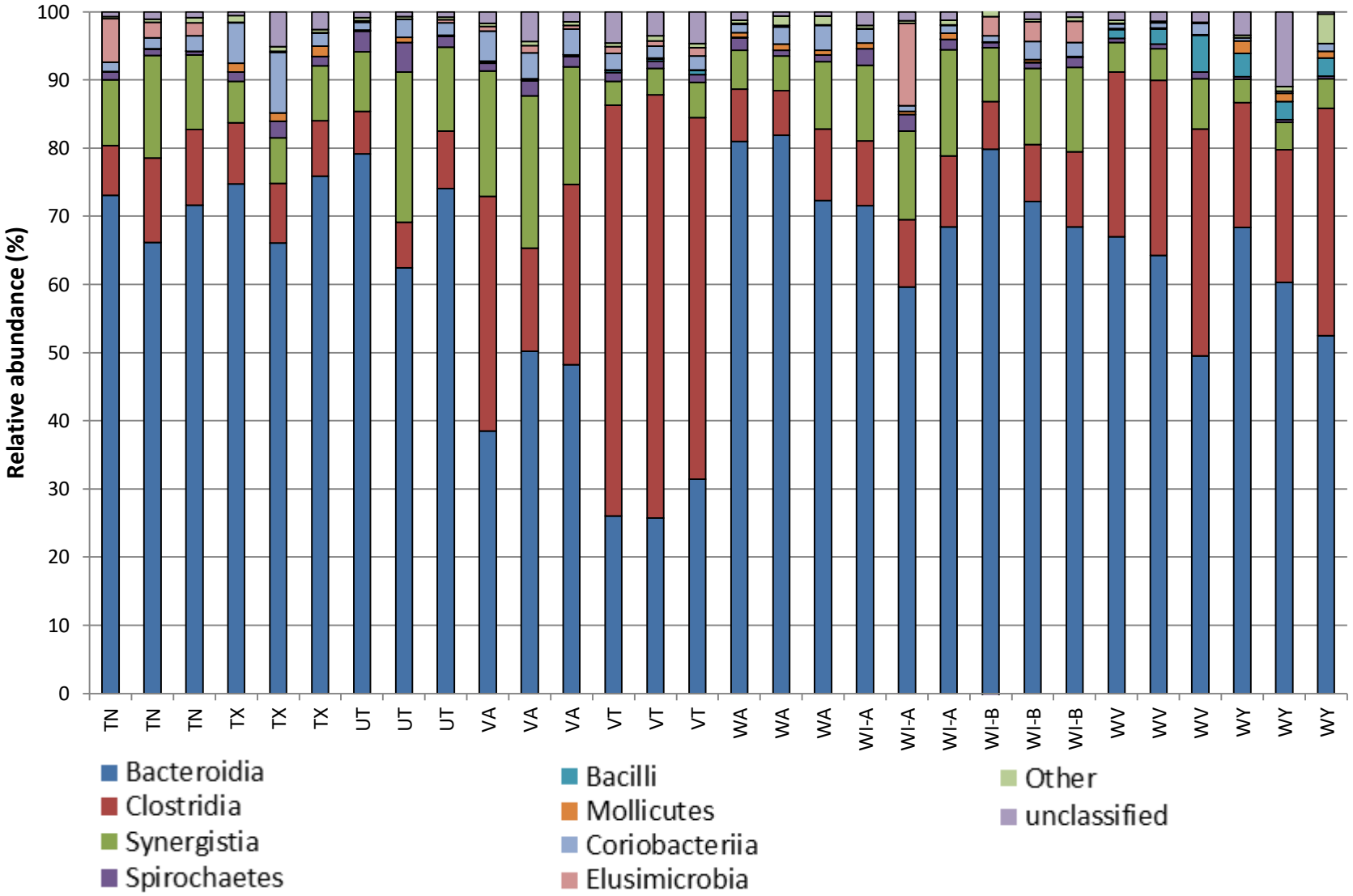

Relative abundance (%)
100
90
80
70
60
50
40
30
20
10
0
TN
TN
TN
TX
TX
TX
UT
UT
UT
VA
VA
VA
VT
VT
VT
WA
WA
WA
WI-A
WI-A
WI-A
WI-B
WI-B
WI-B
WV
WV
WV
WY
WY
WY
Bacteroidia
Clostridia
Synergistia
Spirochaetes
Bacilli
Mollicutes
Coriobacteriia
Elusimicrobia
Other
unclassified

Figure S5: *Bacterial* community profile of 149 digesters at the class level. The profiles of 149 digesters are divided in to five stacked bar graphs A, B, C, D and E, respectively. The Y axis represents the relative abundance (%) of *Bacterial* classes observed of the total *Bacterial* sequences in a digester sample. The digester sample on the X axis are labelled based on the US state(s) from where the seed source was obtained (i.e. AK = Alaska, HI = Hawaii etc.). Replicate digesters seeded from the same seed are given the same labels.

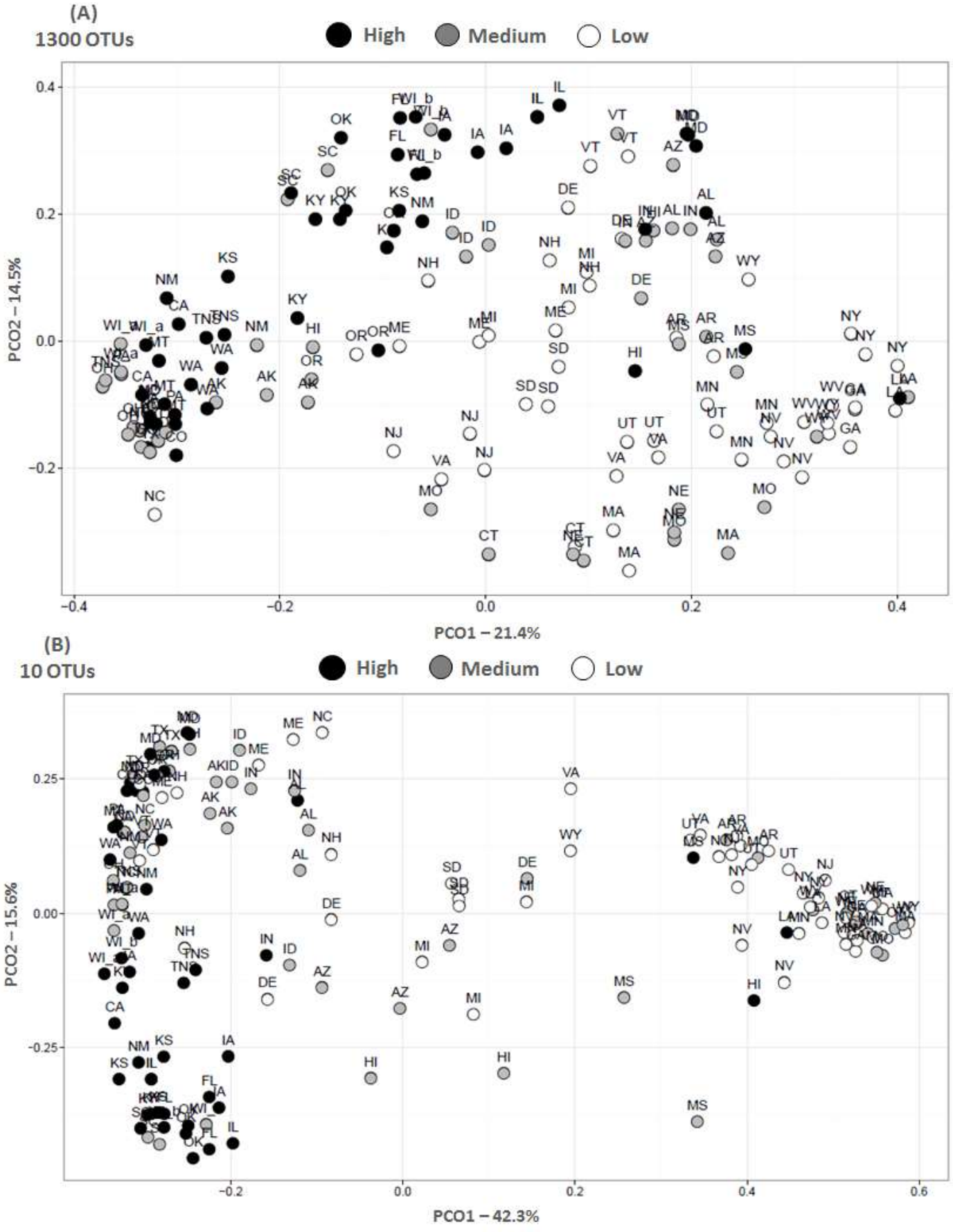

(A)
1300 OTUs
High
Medium
Low
PCO2 – 14.5%
PCO1 – 21.4%
(B)
10 OTUs
High
Medium
Low
PCO2 – 15.6%
PCO1 – 42.3%

Figure S6: Microbial community principal component analysis (PCOA).
PCoA plots using (A) all 1300 OTUs and (B) 10 highly significant OTUs. Methane production rate classifications are shown as High (black), Medium (grey) and Low (white) symbols. 149 data points are show in both PCoA plots and labelled based on the US state(s) from where the seed source was obtained (i.e. AK = Alaska, HI = Hawaii etc.). Replicate digesters seeded from the same seed are given the same labels.

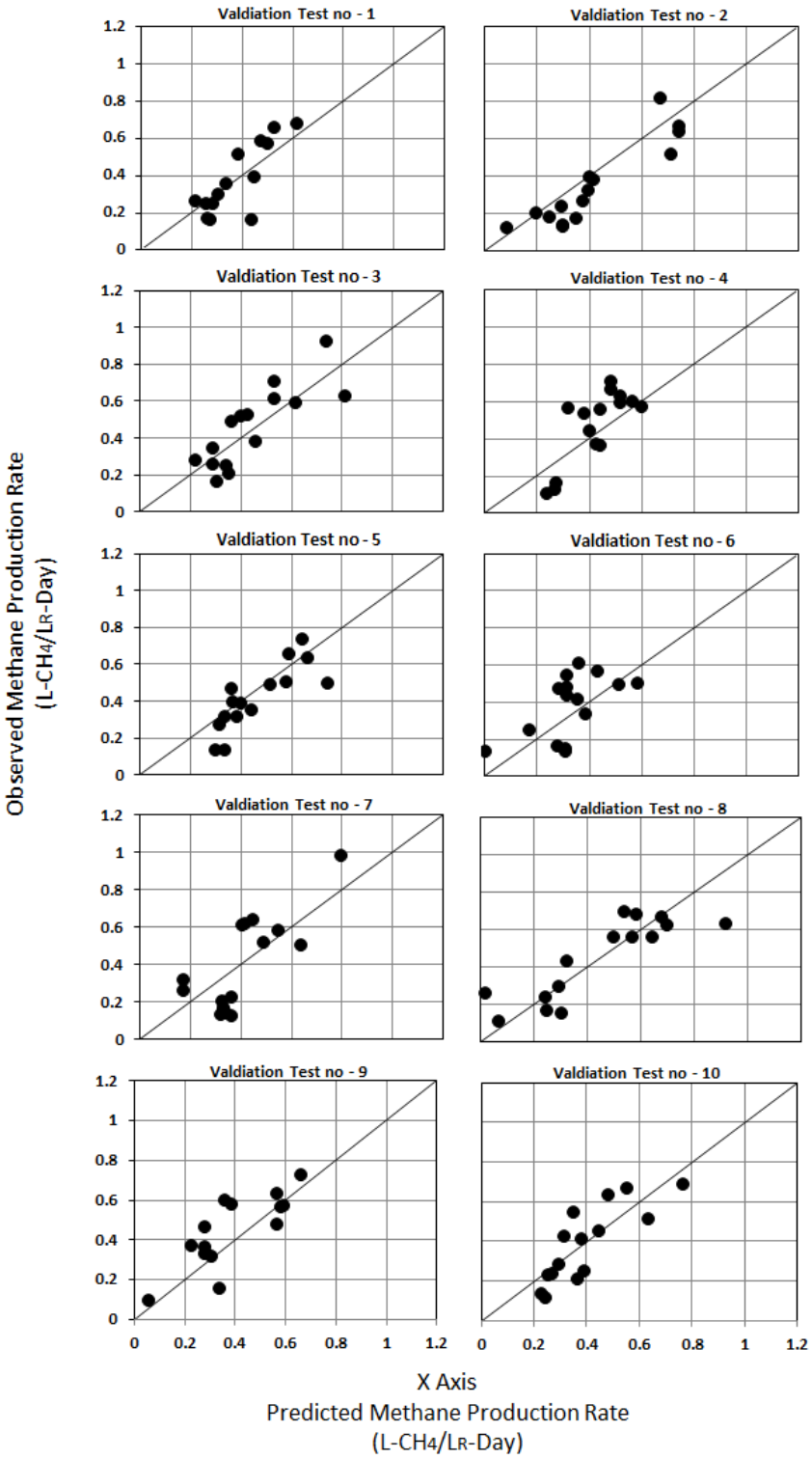


Figure S7: Summary plots of the 10 validation tests.
Results of validation tests using the highly significant OTUs. The data points are for digesters in the test set for each validation test. Values of the validation criteria (q^2, R^2, R^2-R_0^2/R^2 and K) are shown in Table S6. The line in each plot represents the regression line with slope equal to one and intercept equal to zero.