

Research Article

Functional Metagenomics Characterization of an Anaerobic Saltwater Bioreactor

**D. Derilus,¹ A. Forestil,¹ J. Fortuné,¹ O. Polyanska,²
C. Louime,¹ G. Gervais,¹ and S. E. Massey ^{1,3}**

¹Department of Environmental Sciences, UPR - Rio Piedras, San Juan, PR 00931, USA

²High Performance Computing Facility, UPR - Rio Piedras, San Juan, PR 00931, USA

³Department of Biology, UPR - Rio Piedras, San Juan, PR 00931, USA

Correspondence should be addressed to S. E. Massey; stevenemassey@gmail.com

Received 24 October 2018; Revised 20 March 2019; Accepted 24 March 2019; Published 2 May 2019

Academic Editor: Nuri Azbar

Copyright © 2019 D. Derilus et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Methanogens are restricted to a few genera of Archaea, however they have great importance in the carbon cycle, impacting climatic considerations, and also find a role in renewable energy in the form of biogas. Here, we examine the microbial contribution to the production of methane in a sargassum fed anaerobic saltwater bioreactor, which are poorly characterized compared to fresh water bioreactors, using a comprehensive functional metagenomics approach. Despite abundant production of methane, we detected a low proportion of Archaea in the system using 16S rRNA community profile analyses. We address the low representation using an additional 16S rRNA analysis of shotgun data and a consideration of CO₂:CH₄ production. Using a novel network alignment and tree building approach, we measured similarity between the meta-metabolic capabilities of different anaerobic microbial communities. The saltwater bioreactor samples clustered together, validating the approach and providing a method of determining meta-metabolic similarity between microbial communities, with a range of potential applications. We also introduce a number of additional approaches for examining and interpreting meta-metabolic network topology. The low abundance of methanogens appears as a common property of such anaerobic systems and likely reflects the relatively poor energetics of methanogens, while examination of key enzymes confirms that hydrogen producing bacteria are the major fermentative guild. Our results indicate that the use of readily available seawater and marine macroalgae is a promising approach to the production of biogas as a source of renewable energy.

1. Introduction

Anaerobic bioreactors (or anaerobic digester, AD, systems) produce biogas from a variety of organic feedstocks through the process of methanogenesis. This microbial process is an important part of global carbon cycling as it represents the end stage of the breakdown of organic carbon compounds under anaerobic conditions [1]. The resulting biogas possesses a high proportion of methane, an important greenhouse gas, being several times more potent than carbon dioxide. Such is the climatic importance of methane; the evolution and subsequent population expansion of methanogens have been causally linked to the Permian extinction [2]. Capture of this gas via anaerobic digestion of organic waste material and subsequent use as biofuel

represents a potentially important manner of ameliorating greenhouse gas emission as it represents the conversion of methane to carbon dioxide. Therefore, this technique holds promise both as a source of renewable energy and in the bioremediation of organic waste. Most anaerobic bioreactors, however, use fresh water, which is a limited resource. In areas where fresh water is limited, seawater may be a viable alternative. In addition, on small islands and coastal regions where land for agriculture is scarce, a potential source of feedstock may be marine macroalgae, given a viable culturing methodology or sustainable harvesting regime.

Seven orders of euryarchaeal methanogens are known (*Methanococcales*, *Methanobacteriales*, *Methanomicrobiales*, *Methanopyrales*, *Methanosarcinales*, *Methanocellales* and *Methanoplasmatales*; [3]), but none have been described from

bacteria or eukaryotes. Hydrogenotrophic methanogens are able to reduce carbon dioxide to methane using hydrogen as an electron donor, whereas acetoclastic methanogens convert acetate to methane. A third group, methylotrophic methanogens, are able to reduce methyl groups to methane.

Acidogens, acetogens and hydrogen producing bacteria comprise fermentative guilds, in addition to the methanogens [4]. Acidogens convert monosaccharides and amino acids into organic acids, which form the substrate for acetogens. Most methanogenic species are hydrogenotrophic, forming syntrophic interactions with hydrogen producing bacteria, using hydrogen to reduce CO_2 to CH_4 . In sulfate free marine sediments CO_2 reduction is dominant, while acetoclastic methanogenesis is dominant in freshwater [5]. However, where sulfate is present, sulfate reducing bacteria (SRB) will compete with methanogens for the common substrates hydrogen and acetate [6], and H_2S can also preferentially inhibit some methanogens such as the *Methanosaetaceae* [7]. Acetogens are mostly bacterial and anaerobically produce acetate from CO_2 and an electron source such as H_2 , so they can also compete with hydrogenotrophic methanogens. However, as they produce acetate, they provide the substrate for acetoclastic methanogens. Most methanogenesis is acetoclastic in nature, however only two genera of methanogens have been reported as containing acetoclasts, *Methanosarcina* and *Methanosaeta* [8].

Metagenomics approaches allow a comprehensive characterization of microbial taxa and known microbial metabolic pathways present in a microhabitat and provide an opportunity to understand the function of bioreactors in terms of the specific metabolic pathways that are present in the metagenome. This represents the metabolic capacity of the microbial community given that in prokaryotes, most genes present in the genome are expressed at some point during the lifecycle, as pseudogenes are rare [9]. Metagenomic analyses of anaerobic bioreactors have shown that Archaea comprise only 8-11% [10], 12-15% [11], 10% [12], 1% [13], 4-7% [14] and 3-5% [15] of the prokaryotic population (in contrast, an anaerobic terephthalate degrading bioreactor displayed 61% Archaea [16]). Nonmetagenomic studies have also shown substantial variability in the proportion of Archaea; for example, a range of 0-20% was observed in a series of sewage and waste bioreactors [17], and 4-8% Archaea were observed in a lab-scale bioreactor [18]. The minor proportion of Archaea in the majority of these bioreactor systems has not been adequately addressed and is explored in detail here using a functional metagenomics approach and a range of novel bioinformatics approaches, tailored for this study but with potential applications elsewhere.

2. Methods

2.1. Bioreactor Setup. Two identical two-stage semicontinuous bioreactors were constructed from acrylic jars, polypropylene fittings, and tygon tubing. The operational liquid volumes of the first and the second stage were 0.5 L and 1.5 L, respectively.

The bioreactors were inoculated using anoxic sediment acquired from the bottom of the San José Lagoon in San Juan, Puerto Rico (PR). The sediment was extracted from an anoxic pit created when the lagoon was used as source of landfill material for construction in San Juan. A month before inoculation, grab samples were taken from the lagoon bottom to prepare a new inoculum. The sample was placed in a 20 L bucket of water previously deoxygenated and adjusted to 15 ppt with synthetic marine salts (Instant Ocean). This bottom mud sample had a final dissolved oxygen (DO) concentration and salinity of 0.22 ppm and 16 ppt, respectively. The bucket was held at controlled room temperature (25°C) until needed for inoculating the refurbished bioreactors. The inoculum was drawn on March 10, 2013 (day 0) at a specific point in San José lagoon that was selected due to its specific characteristics. The site had a depth of 6.70 m, a superficial salinity of 12.2 ppt and was located at 18° 25' 20.1" N and 066° 00' 45.2" W. A sediment grabber was used to take the mud from the bottom of the lagoon. The mud drawn from the bottom of the lagoon had a salinity of 25 ppt and a concentration of dissolved oxygen of 1 ppm.

The salinity, DO, pH, and temperature of the bioreactors were measured prior to inoculation. At the time of inoculation (day 1), the salinity was 15 ppt and the DO was 0.63ppm as measured by a YSI optical DO meter, the pH was 7.5, and the temperature of the stored inoculum was 23°C. Prior to inoculation, the clean and refurbished 2 L systems were filled with 15 ppt synthetic seawater to the appropriate level. Oxygen was stripped from the two 2 L systems with a nitrogen purge. Subsequently, the two bioreactors with a total of 4 chambers were connected in series to the inoculum bucket. Using a peristaltic pump and 1/2 i.d. tubing, approximately 40 L of the bottom mud from the inoculum bucket was pumped through the bioreactors in order to assure that all the chambers received a homogeneous and similar inoculum. The return flow from the bioreactors was fed back into the bucket through a submerged outlet to avoid agitating the surface of the sample and risking the entry of oxygen into the inoculum. The bucket was tightly lidded during the entire inoculation procedure and the bioreactor chambers were sealed, and their floating lid gas accumulators filled with pure nitrogen gas.

The bioreactors were gradually adapted to the target salinities using synthetic sea water to a salinity of 10 ppt (low salinity bioreactor) and 30 ppt (high salinity bioreactor). Each system was fed on a daily basis with a suspension of finely ground *Sargassum* spp. harvested from beaches in San Juan, PR as carbon source using the corresponding salt concentration in order to gently adapt the systems to the desired final target salinities of 10 ppt and 35 ppt for the low and high salinity bioreactors, respectively. Osmotic shocks to the microbial populations were avoided by this method of gently adapting the salinities from their initial value of approximately 25 ppt to the desired final salt concentrations during the regular feeding operations. The adaptation to the desired target salinities was completed in three weeks. Adjusting the salinity of the feed blend subsequently used a similar mechanism of salinity maintenance by adjusting the salinity of the feed blend to return the systems to their

target values any time a drift in the system salinity was observed. Both digesters were fed continuously using organic feed suspensions with increasing flow rates of 50, 61, 91, and 120 ml/d that theoretically correspond to hydraulic retention times (HRT) of 40, 32, 21, and 17 days respectively. These flow rates were calculated in order to increase the organic loading rate (OLR) in a step-wise fashion from 1.25 to 5 g of volatile solids/L bioreactor liquid volume per day ($VS \cdot L^{-1} \cdot d^{-1}$). Each incremental increase in OLR was initiated once the concentration of volatile organic acids (VOA) had peaked and then fallen to a steady-state level. After each increase of the OLR, it typically required about four weeks for VOA values to peak and then return to a steady state. Operational parameters including pH, total solids (TS), volatile solids (VS), volatile organic acids (VOA), and alkalinity were monitored twice a week in alternating bioreactor stages. Running average values of the test parameters were used to monitor bioreactor performance over time.

In addition, two 15 L 3 chamber bench-scale anaerobic bioreactors were established at low salinity (10 ppt) and high salinity (35 ppt) for comparative purposes. The 15 L bioreactors were inoculated using the sludge of the 2L anaerobic systems. Chemical and physical analyses of representative samples were conducted four times a week.

2.2. Measurement of Biogas. The biogas was collected in floating lid gas accumulators, and the daily volumetric production was measured. To measure the biogas quality, the biogas was sampled from the reactor and stocked in Fluorinated Ethylene Propylene (FEP) gas sampling bags. The biogas quality was characterized by Fourier Transformation Infrared Spectroscopy (FTIR) using a Thermo Nicolet NEXUS 470 FTIR. A volume of 500 μ l (in triplicate) of each biogas sample was drawn using a 1 ml syringe and subsequently injected into a 72 ml CaF_2 gas cell with a 100 mm path length from Pike Technologies, following a nitrogen purge of the cell, and then the final infrared spectra were preserved for further analysis. The relative concentrations of methane and carbon dioxide were calculated based on a standard calibration mixture from Linde Gas Inc., composed of 10% CO_2 , 10% CH_4 , 5% H_2S , and 75% N_2 .

2.3. Community Profile Metagenomic Sequencing and Analysis. Environmental DNA was extracted from bioreactor samples using the UltraClean Soil DNA Kit by MO Bio Laboratories, Inc. For polymerase chain reaction (PCR), the universal prokaryotic primers (F515: GTGCCAGCMGCCGCGGTAA and R806: TAATCTTWTGGVHCATCAGG [20]) were used in order to amplify the 16S rRNA gene V4 hypervariable region. PCR was conducted using the HotStarTaq Plus 2X Master Mix Kit (Qiagen, USA). Cycling conditions were: 94°C for 3 minutes, followed by 28 cycles at 94°C for 30 seconds, 53°C for 40 seconds, and 72°C for 1 minute, after which a final elongation step at 72°C for 5 minutes was performed. The forward primer contains a 454 sequencing primer and a multiplex identifier (MID) sequence, to identify the origin of each sample during computational analysis. The barcoded PCR products were sequenced at MrDNA Laboratories (www.mrdnalab.com, Shallowater, TX, USA)

on a Roche 454 GS-FLX+ Pyrosequencer. Sequence statistics are in Supplementary Tables 3 and 4. The sequence data used to support the findings of this study have been deposited in the NCBI Sequence Read Archive (SRA) under accession number SRR5864747.

The Qiime package [21] was used to determine the microbial community structure of the different bioreactor samples. Each dataset was filtered and demultiplexed separately by the following criteria: minimum quality score = 20, minimum length = 200, no ambiguous bases allowed, maximum homopolymer allowed = 5, barcodes, and adapters were removed. Then, the resulting sequences were binned in operational taxonomic units (OTUs) that are considered as cluster of related sequences that share at least 97% sequence identity. The OTU picking was conducted based on a minimum identity threshold of 97% to the sequences in Greengenes databases that are preclustered at 97% identity [22]. During OTU picking, formation of clusters that did not match the reference database was avoided. Subsequently, one representative sequence (longest) for each OTU was picked to be used in the downstream analysis. These representative sequences were aligned using the MUSCLE algorithm [23], and the alignment was used to build a phylogenetic tree and OTU table required for the phylogenetic α and β diversity analyses. The taxonomic assignment of the representative sequences was carried out using the Ribosomal Database Project (RDP) Classifier, a naive Bayesian classifier [24]. The OTUs were taxonomically classified against the most recent update of the Green Genes database under a statistical confidence of 80%. All OTUs with a confidence values lower than 80% (minimum at phylum level) were considered as unclassified or unknown.

The α diversity of the microbial communities was computed using the Shannon Index using Qiime. Ten subsampling repetitions were performed without replacement at each sampling depth with a maximum of 2500 sequences per sample. Samples with a number of reads lower than the maximum sequence depth were omitted. Rarefaction curves were generated by considering the number of OTUs observed at different sequence depths (Supplementary Figure 1). β diversity was computed using unweighted unifrac distances. Similar to the case of α diversity, the sequence depth could have an impact on the β diversity. To counteract this problem, a jackknifing was performed by resampling 100 times with a maximum of 2400 sequences per sample. Only one sample (S3 10ppt.Oct.13) with the number of sequences below this maximum sequence depth was excluded in the β diversity analysis. To compare the high and low salinity bioreactor categories, principal coordinate analysis (PCoA) was computed. ANOSIM (Analysis of Similarity) was used to test the significance of PCoA clustering.

A 16S rRNA community profile meta-analysis was conducted on a range of different anaerobic sediments from natural and artificial environments including municipal solid waste (MSW), industrial waste water treatment plants, lagoon sediments, slaughterhouse waste, codigested sludge, deep sea sediments, and lignocellulose materials. These were combined with the 20 datasets generated in this study. Most of the samples were obtained from NCBI based on the accession

number provided by the respective study authors. When the data was not available in the public database the authors were contacted via email to access the data. The barcodes for each sample were identified and a custom perl script was used to barcode the nonbarcoded samples. These samples were demultiplexed and filtered as mentioned previously. The reads from each dataset were then merged and considered as input for OTU picking, which followed the method outlined above, on all the datasets using the open reference method. As the same reference method was used for OTU picking, the OTU table generated could be used to compare microbial diversity between different samples independently of the 16S rRNA region sequenced. Given that different primer pairs and amplification conditions were used for each sample, this can potentially introduce an additional source of bias into the analysis.

2.4. Shotgun Metagenomic Sequencing and Analysis

(i) *Sequencing.* Metagenomic shotgun sequencing was conducted using the Illumina platform by MrDNA on samples taken from the high salinity bioreactor. A paired-end set of reads (R1 and R2) was produced for each sample. To avoid potential bias inherent to paired-end reads joining, including gaps and overlapping, only the forward reads file (R1) was used for the downstream analysis. The shotgun sequence data used to support this study have been deposited in the NCBI SRA, under accession numbers SRR5891520, SRR5865137 and SRR5891595.

(ii) *rRNA Taxonomic Profiling.* All sequences shorter than 100 base pairs were removed from the metagenomic datasets using Prinseq [25]. Sequence fragments that possessed significant similarity to 16S rRNA were identified using the metaxa2 software [26] by Blast searching against the 16S rRNA Greengenes reference database, with a sequence identity cutoff of 99%. The resulting 16S rRNA homologs from the different metagenomic datasets were barcoded and concatenated to be used as input into Qiime in order to generate diversity metrics for each shotgun dataset. The 16S rRNA reads were demultiplexed and filtered, and OTU picking was conducted using the same criteria, as for the 16S rRNA community profile analyses, described above. The OTU table generated was used to compare microbial diversity between different samples independently of the 16S rRNA region sequenced. This represents a more accurate measure of taxonomic diversity of shotgun data than measuring the taxonomic affiliation of sequence fragments derived from the taxonomic affiliation of the closest sequence match, which is generally not precise.

(iii) *Construction of Meta-Metabolic Networks.* Meta-metabolic networks show the relative abundances of enzymes catalyzing different metabolic pathways in a microbial community and were generated as follows. Individual reads were Blast searched against the nonredundant NCBI sequence database using Blastx [27] and an e-value threshold of e-15. The search was conducted on an 8 node computer cluster using a bash script to parallelize the task, utilizing

the Open Grid Scheduler batch queuing system. The GI numbers of significant hits were extracted and converted via ID mapping to the corresponding KEGG K numbers. This was accomplished using the idmapping.dat mapping file obtained from Uniprot (ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/idmapping/). Then, the relative redundancy of the K numbers was calculated using a custom perl script and superimposed onto a map of central metabolism using the iPATH tool [28].

For quantitative analysis, each meta-metabolic network was converted into an edge list, as follows. K numbers were converted into rn numbers by parsing all kgml files, obtained from the KEGG database (<https://www.genome.jp/kegg/>). rn numbers were converted into edges linking two or more cpd (compound) nodes using a mapping file generated from the kgml files using the KEGGgraph R package [29]. Relative abundance of enzymes responsible for different reactions is reflected in the edge widths between cpd nodes. The resulting graphs were visualized using Gephi [30], and topological characteristics were calculated using the iGraph R package [31].

(iv) *Network Tree Analysis.* The relative similarity between a group of networks can be presented as a tree, with branch length representing similarity. Pairs of meta-metabolic networks were aligned by identifying the number of common edges linking cpd nodes between the two networks under consideration. Pairwise distances between networks (D) were computed using a custom script by calculating $1 - \text{number of common edges in networks 1 and 2, divided by the number of unique edges}$, as follows:

$$D = 1 - \frac{|E_1 \cap E_2|}{|E_1| + |E_2| - |E_1 \cap E_2|} \quad (1)$$

where E_1 is the set of edges in network 1 and E_2 is the set of edges in network 2. Using this calculation, for identical networks, $D = 0$, and for networks with no edges in common $D = 1$; a larger value of D therefore reflects greater dissimilarity between networks. A pairwise distance matrix was computed by calculating D for every pair of meta-metabolic networks in the dataset. The resulting distance matrix was then used to calculate a neighbor joining tree [32], using Phylip [33].

(v) *Identification of Individual Methanogenic and Glycolytic Enzymes.* Homologs of methyl-coenzyme M reductase (*mcrA*), which catalyzes the last step in methanogenesis, were identified using tblastn using the *Methanosarcina acetivorans* sequence (WP_011024419.1) as a probe and an e-value threshold of e-15 and 40% sequence identity at the amino acid level. Homologs of bacterial phosphofructokinase (*pfk*), which catalyzes the committed step of glycolysis, universal to all fermentation pathways, were likewise identified using the *Escherichia coli* sequence (WP_085454157.1) as a probe. Homologs of formyltetrahydrofolate, which is a key enzyme in acetogenic bacteria, were identified using the *Caldanaerobacter subterraneus* sequence (AAM25530.1) as a probe. Acetate kinase, butyrate kinase and butyryl-CoA:acetate-CoA transferase are key enzymes in acidogenic bacteria. Homologs of acetate kinase were identified using *Bacteroides*

salanitronis sequence (ADY36292.1) as a probe, homologs of butyrate kinase were identified using the *Odoribacter splanchnicus* sequence (ADY31769.1) as a probe and homologs of butyryl-CoA:acetate-CoA transferase were identified using the *Clostridium tyrobutyricum* sequence (AND83913.1) as a probe. [FeFe] hydrogenase and [NiFe] hydrogenase are key enzymes in hydrogen producing bacteria. Homologs of [FeFe]-hydrogenase and [NiFe]-hydrogenase were identified using the *Clostridium butyricum* (ABO42543.2) and *Rhodobacter capsulatus* (AAA69668.1) sequences, respectively. To normalize the datasets, the total numbers of protein coding genes in each dataset was calculated using FragGeneScan [34].

2.5. Total CO₂ Calculation. The total amount of CO₂ produced by the system was estimated by measuring the gaseous CO₂ produced by the bioreactor and then calculating the total CO₂ by using a partition function that accounts for the solubility of CO₂ in water. Given the slow kinetics of precipitation and dissolution of solid carbonates in the solid fraction, especially within a matrix of inorganic and organic solids and biofilms, we cannot assume that the partition of CO₂ operates at chemical equilibrium during the study period [35]. However, rough estimates of the partition of CO₂ between the gas phase and liquid phase can be derived using experimentally derived Henry's law constants [36].

3. Results and Discussion

3.1. Behavior of the System over Time. The two 2L bioreactors were operated during a six month campaign that began in April 2013. During that time, they were fed with increasing amounts of ground sargassum suspension. The nominal OLR was 1.25 g VS*L⁻¹*d⁻¹ and by the end of the campaign the OLR was 5 g VS*L⁻¹*d⁻¹ based on a nominal value of 30% ash in the algal solids (Supplementary Table 1). The true values of the OLR, based on actual experimental measures of volatile solids in 7% suspensions of ground sargassum suggest that the initial OLR was 1.40 g VS*L⁻¹*d⁻¹, while the final value at the end of the campaign was approximately 5.7 g VS*L⁻¹*d⁻¹. During the metagenomics campaign, the system never achieved steady-state operation, although that was not critical for these experiments. Biogas yield in the high salinity system peaked at approximately 100 ml biogas per gram VS at an OLR of 1.5 g VS*L⁻¹*d⁻¹. For comparison, Bird et al. [37] reported a specific yield of biogas from fresh *Sargassum fluitans* of 200 ml per g VS*L⁻¹*d⁻¹, while [38] reported findings of more than 400 ml of CH₄ (~ 800 ml of biogas) per gram VS fed per L*d when fresh *Ulva* spp. or *Gracilaria* spp. were used as feedstock for anaerobic digestion. As expected, the use of sun-dried beach wrack as feedstock for AD, a very refractile feedstock, could not reproduce the much higher yields of biogas reported in the literature when bioreactors are fed with fresh ground algal slurries. The 2L bioreactors were fed continuously since the end of this campaign in the fall of 2013 until 2015, when an FTIR spectrometer was brought on line, and the composition of the biogas was characterized at an OLR of 1.5 g VS*L⁻¹*d⁻¹ (Supplementary Table 2).

3.2. Taxonomic Composition of the Bioreactors

(i) 16S rRNA Community Profile Metagenomics. Using 16S rRNA community profile metagenomics, the taxonomic composition of the bioreactors was assessed in order to better understand the function of the bioreactors. The archaeal composition of the initial inoculum (32.28%) shows a marked reduction after the first day, which continued to 6 months, for both bioreactors (0.39% Archaea low salinity bioreactor; 0.40% Archaea high salinity bioreactor; Figure 1). NGS statistics are displayed in Supplementary Tables 3 and 4, physical performance of the bioreactors when sequencing samples were taken are listed in Supplementary Table 5. The initial inoculum was dominated by methylotrophic methanogens, with the remainder of the microbial population mostly comprised of hydrogenotrophic methanogens. The dominance of methylotrophic methanogens is consistent with the predominance of methylated compounds in brackish environments [8]. The large drop in methanogen numbers between sampling (day 0) and inoculation (day 1) might be attributed to an accidental, transient oxygen exposure. In the lower salinity bioreactor, acetoclastic methanogens dominate after 6 months, while in the higher salinity bioreactor hydrogenotrophic methanogens dominate (Figure 1).

The presence of anaerobic methane-oxidizing archaea (ANME) may result in a diminution in the efficiency of the bioreactors, as they consume methane [39]. While a low proportion of ANME (5.5% and 2.9% of the total archaeal community on days 0 and 1, respectively) was observed in the initial inoculum, no ANME were observed during operation of the bioreactor (Figure 1), and so they are not expected to affect methane generating efficiency in our system. The low overall proportion of Archaea in the mature bioreactors appears counterintuitive, given the high levels of methane production, possible explanations are discussed below.

SRB in the initial inoculum are dominated by *Desulfobacteriales*, and the overall proportion of SRB is 10.33%. The proportion in both reactors decreases markedly (3.27% high salinity bioreactor, 1.54% low salinity bioreactor, after 6 months), accompanied by a substantial increase in the proportion of *Desulfovibrionales*, in both bioreactors (Figure 1). This difference in SRB was expected since the high salinity bioreactor has a higher concentration of sulfates which facilitates the proliferation of SRB. The presence of SRB is undesirable for the efficient production of methane, given that SRB compete with methanogens. Possible strategies for reducing the SRB component would comprise potential future work.

The overall α diversity of both of the bioreactors decreases over time and does not substantially diverge from each other (Supplementary Figure 2). The reduction may be explained by a process of habitat filtering, whereby the habitat "selects" the best adapted microbes from the initial inoculum, for the new environmental conditions of the bioreactor. This is likely a better explanation than the alternative view that diversity is mostly determined by species interactions [40], given that the species coexisted before inoculation. The results indicate that the higher salinity bioreactor does not constitute a more stringent filtering regime given that the final α diversity is

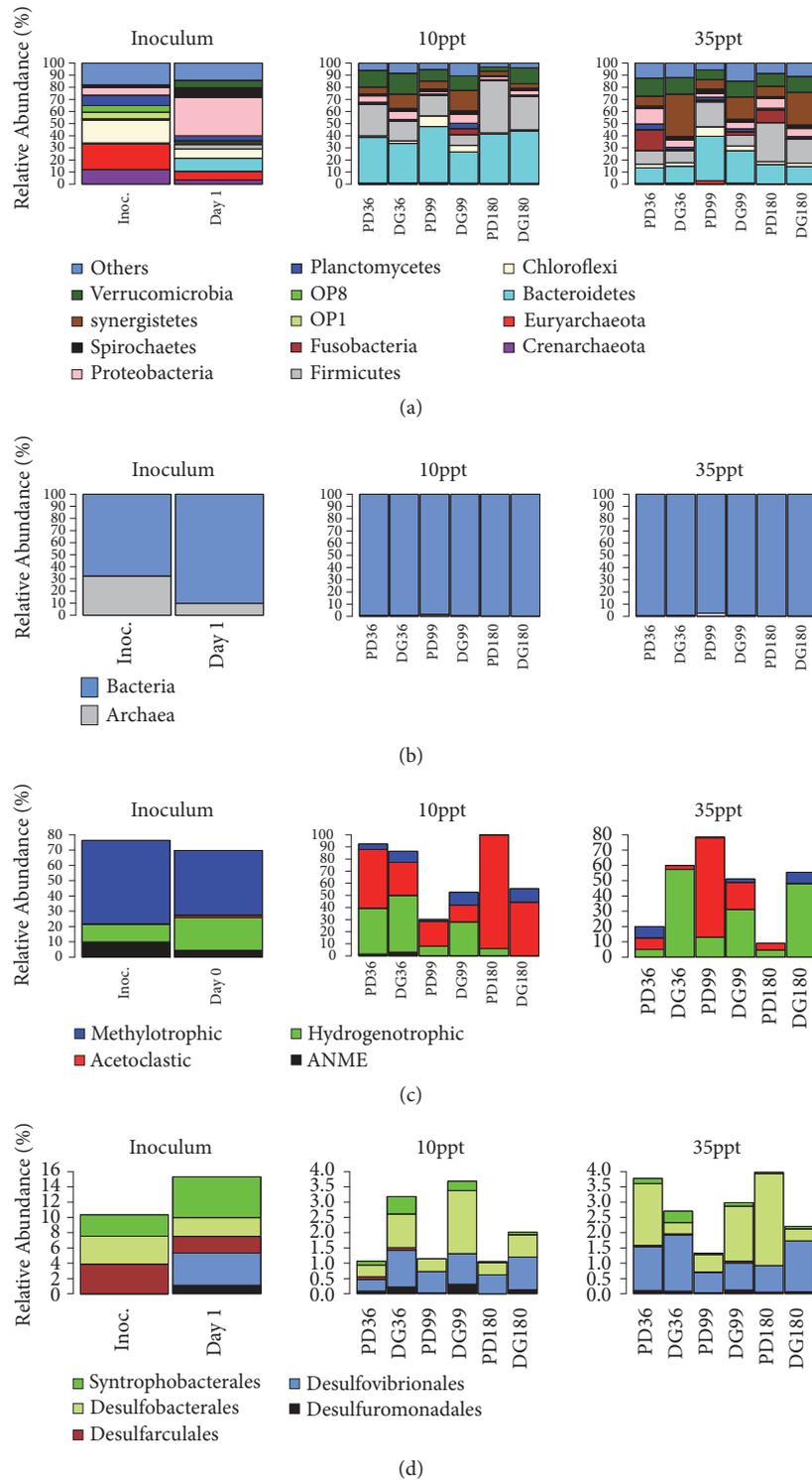


FIGURE 1: Change in the taxonomic composition of the bioreactors over time. 16S rRNA community profile metagenomics data were generated from (i) initial inoculum; (ii) a high salinity; and (iii) a low salinity bioreactor. The graphs show the change in the relative proportions over time of (a) overall taxonomic diversity in the bioreactors; (b) Archaea versus Bacteria; (c) hydrogenotrophic, acetoclastic, and methylophilic archaeal methanogens and anaerobic methane-oxidizing archaea (ANME); (d) sulfate reducing bacteria (SRB).

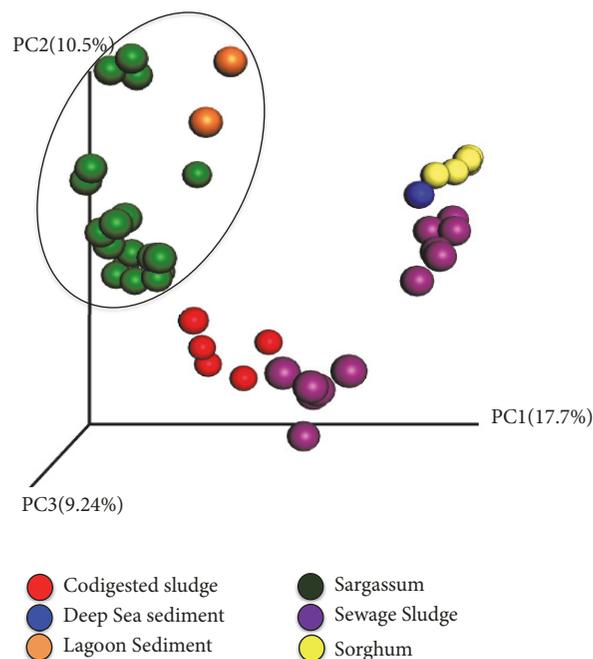


FIGURE 2: Meta-analysis of bioreactor microbial taxonomic diversity. Principal coordinate analysis was conducted on 16S rRNA community profile metagenomics data generated from the saltwater bioreactors and from the literature (Supplementary Table 3). The unweighted unifrac measure was used as a measure of β diversity, as described in Methods. Percentage values in parentheses represent the percentage variation in each dimension.

comparable to that of the lower salinity bioreactor (Shannon indices of 5.05 and 5.25, respectively), indicating that the species present are tolerant to a range of salt concentrations. This observation implies that it should be no more difficult to adapt AD systems to operate in full strength seawater compared to low salt or no-salt systems.

In order to compare the taxonomic composition of our bioreactors with those of other published studies, a 16S rRNA community profile meta-analysis of bioreactor microbial communities was conducted using the unifrac measure of β diversity [41] which incorporates phylogenetic affinity. This is useful when there is a lack of an exact species match, in contrast to a counting method such as Bray-Curtis. Such a lack of exact species matches is typical when comparing microbial communities from different habitats (as most taxa identified are “novel”), and this can give spurious results when using counting methods. The analysis revealed that the samples generated in the saltwater bioreactors and the original inoculum obtained from San José lagoon clustered together (Figure 2). This confirms that the San José sediment inoculum is the ultimate source for the majority of the microbes in the bioreactors and also indicates that the final composition of the bioreactors bears a significant similarity to the initial inoculum, when compared to samples from other bioreactor systems (Figure 2). The other bioreactor microbial communities also cluster together dependent on feedstock validating the community profile analysis methodology. The consistent clustering pattern of bioreactors samples based on feedstock is a further indication that biomass feedstock represents one of the important ecological factors that shapes microbial communities in artificial ecosystems.

(ii) *Shotgun Metagenomic Taxonomic Analysis.* Community profile metagenomics is semiquantitative as it is reliant on PCR amplification, which is subject to biases [42]. In contrast, shotgun metagenomics avoids this type of bias, thus significant differences may be observed when comparing 16S rRNA metagenomics community profile data with 16S rRNA sequences derived from shotgun metagenomics data [43]. Therefore, a taxonomic analysis of the shotgun data was used to corroborate the community profile results, which is particularly important given the low level of Archaea indicated by the community profile results. Sequence statistics of the shotgun datasets generated in the study are in Supplementary Table 6. Sequence data was generated from the 15 L high salinity bioreactor (B1) and for the 2L high salinity bioreactor three months after inoculation (B2 from the upper chamber; B3 from the lower chamber).

Principal coordinate analysis of β diversity shows that the microbial communities of the three sargassum fed saltwater bioreactor samples cluster with each other (Figure 3), consistent with the results for the community profile results (Figure 2) and validating our methodology. The ten shotgun datasets examined all had low levels of Archaea, < 10% of the total 16S rRNA taxonomic assignments in each case, with the exception of B5 which had 30.33% (Supplementary Table 6). Samples B1, B2, and B3, generated in this study, had 0.30%, 9.02%, and 7.71% Archaea, respectively. These observations are consistent with reports of low archaeal proportions using 16S rRNA community profile approaches (see Introduction) and also confirm the low proportion of Archaea observed in our study using community profile analysis.

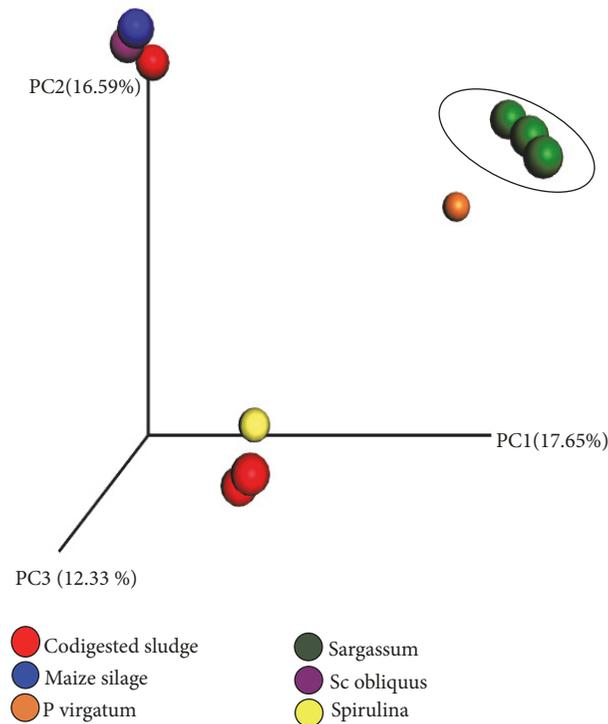


FIGURE 3: 16S rRNA analysis of bioreactor shotgun metagenomic datasets. Principal coordinate analysis was conducted on 16S rRNA homologs extracted from ten shotgun metagenomics datasets (Supplementary Table 6), using the unweighted unifracs β diversity measure, as described in Methods.

There is some discrepancy between the proportions of Archaea observed from the 16S rRNA community profile results and the 16S rRNA shotgun analysis. The community profile results indicate 0.66%, 2.43%, and 0.68% Archaea, for the shotgun samples B1, B2, and B3. Shotgun samples B2 and B3 show substantially higher values of Archaea, compared to the 16S rRNA community profile analysis. Potential explanations include amplification bias during the generation of sequences for the 16S rRNA community profile analyses or variations in 16S rRNA copy number between Bacteria and Archaea, which would also contribute to amplification bias, discussed further below. Interestingly, our shotgun 16S rRNA analysis provides more unique OTUs than the community profile 16S rRNA approach: there were 888, 2118, and 919 unique OTUs identified in the B1, B2, and B3 shotgun samples, respectively. In contrast, there were 895, 756, and 647 unique OTUs identified, respectively, in the corresponding 16S rRNA community profile analyses. Strikingly, there were substantially fewer reads corresponding to 16S rRNA extracted from the shotgun data, than were generated by the 16S rRNA community profile analyses (1309 versus 8085 for B1, 2683 versus 7798 for B2, and 1236 versus 5573 for B3). This indicates that, during PCR amplification, while the number of sequences may be amplified, diversity is lost. Based on these data the argument can be made that the shotgun approach is more data rich and less biased. However, this would require further study before generalities can be made.

3.3. Meta-Metabolic Network Analysis. Analysis of meta-metabolic networks allows an insight into the overall emergent metabolic characteristics of microbial communities. The overall metabolic (i.e., ‘meta-metabolic’) characteristics of a microbial community may be visualized using 2D and 3D network representations, which take into account the relative abundances of different enzymes specific for different metabolic pathways present in the metagenome (Figure 4). This view assumes that the gene copy number of a protein in a metagenomic dataset will correlate with the concentration of the translated protein.

One of the emergent features that may be inferred is the meta-metabolic similarity between datasets. We characterize this by aligning networks and calculating a measure of distance based on the proportion of common edges. We use methods derived from phylogenetics for the clustering analysis, but these are equally applicable to considerations of network similarity, as for species or gene relationships. The approach can be used to generate a network tree that clusters networks based on meta-metabolic similarity (Figure 5). The meta-metabolic networks derived from the saltwater bioreactors (samples B1, B2, and B3) cluster closely with each other, when compared with datasets derived from other bioreactors. This indicates that they are more closely related to each in terms of overall metabolic similarity of the microbial communities present in the other bioreactors. The high degree of lateral gene transfer in prokaryotes means that that phylogenetic signal from core genes such as 16S

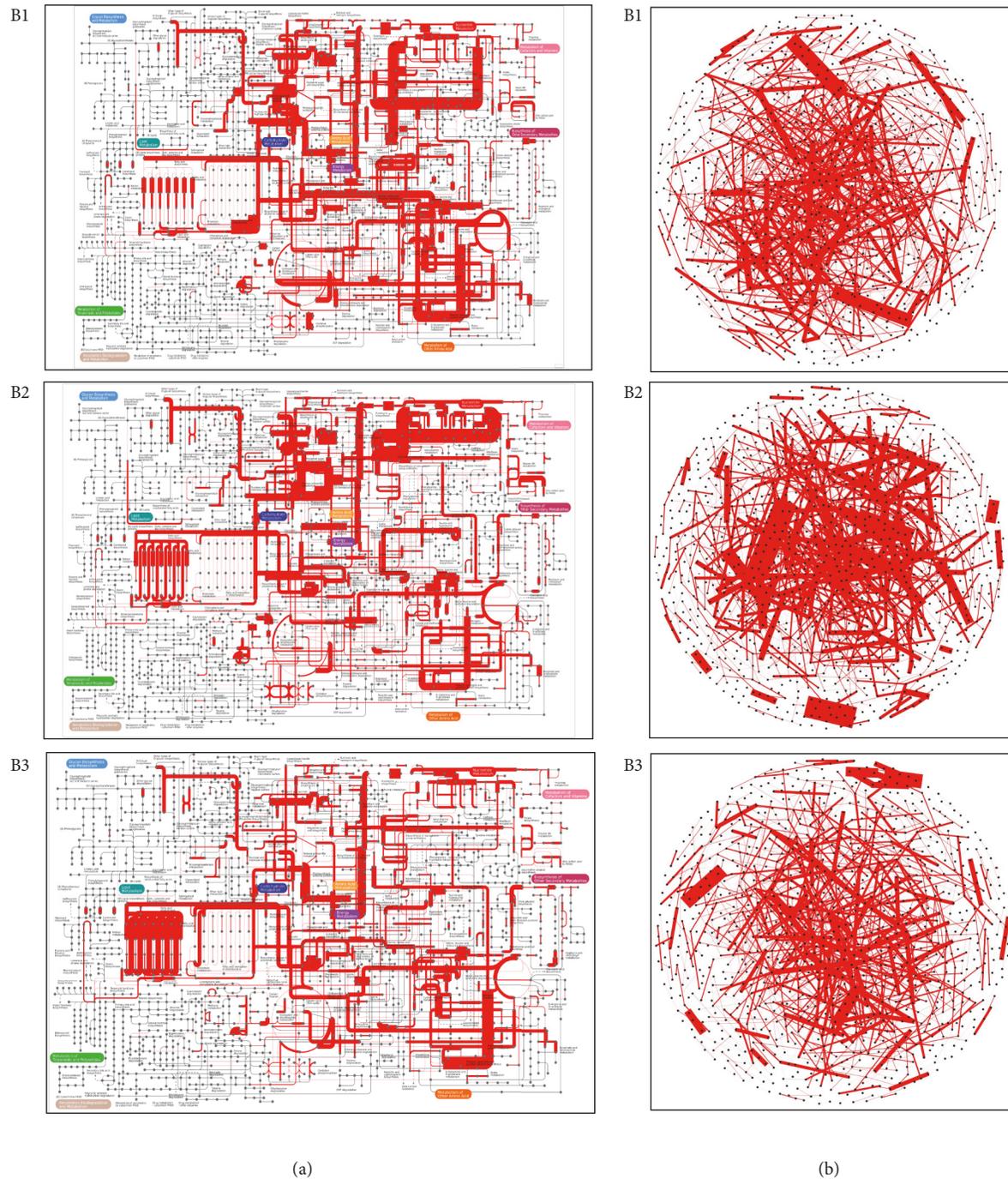


FIGURE 4: Bioreactor meta-metabolic networks. Meta-metabolic representations of the microbial communities of the three saltwater bioreactors fed with sargassum showing relative enzymatic abundances using (a) iPath 2.0; (b) 3D Fruchterman-Reingold [19] representations. The respective bioreactor samples are B1 (i), B2 (ii), and B3 (iii).

rRNA is not necessarily tied to metabolic capacity. Therefore, it is important to validate empirically, for example by using our method of constructing network trees. The results indicate that the taxonomic affinities between the bioreactor shotgun datasets revealed in the β diversity estimates of 16S rRNA (Figure 3), also reflects metabolic affinities. A potential weakness of the network tree approach is that of incomplete

and biased ID mapping, which might lead to artefactual clustering. We expect the problem of inefficient ID mapping to reduce in time as genomic and metabolic databases become more comprehensively annotated.

Additional features of meta-metabolic network topology were examined (Table 1). Firstly, the number of edges (representing KEGG metabolic pathways) is not much larger

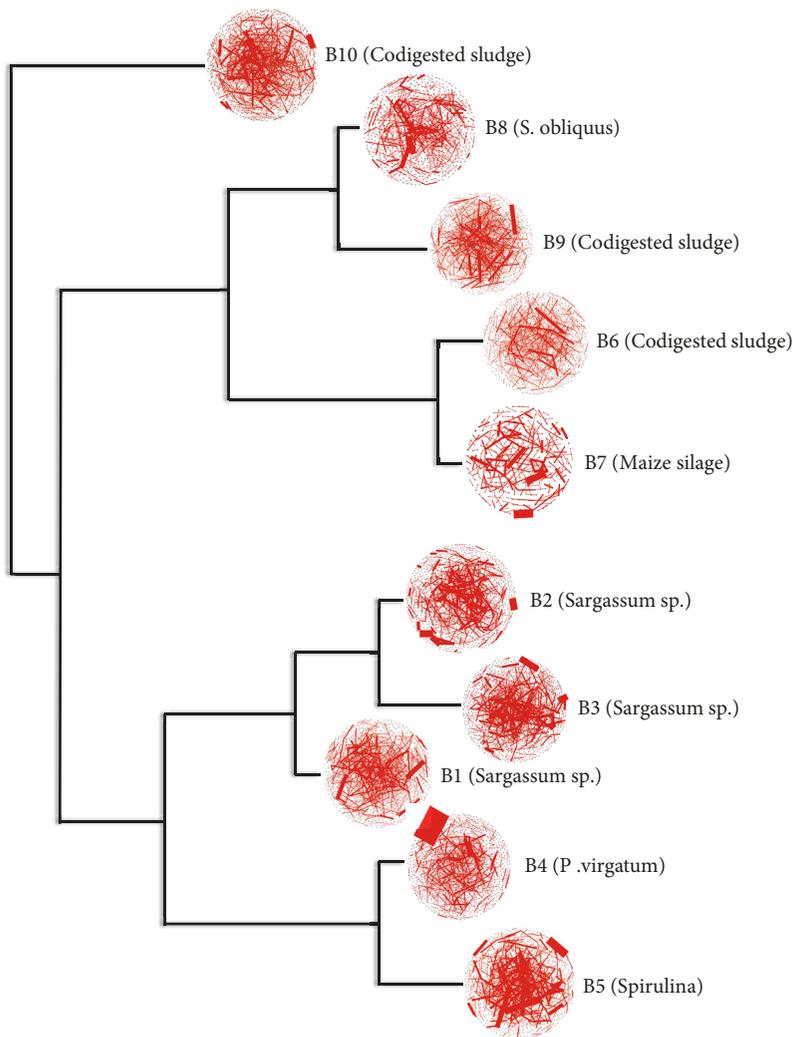


FIGURE 5: Neighbor joining tree of bioreactor meta-metabolic networks. A neighbor joining tree of ten shotgun metagenomic datasets obtained from anaerobic bioreactors (Supplementary Table 6) was constructed. First, individual sequence reads were Blast searched against the NCBI nonredundant database, and then the identified enzyme homologs were ID mapped to K numbers, which were then used to construct a meta-metabolic network. The network edges were aligned, and distance was calculated between each pair of networks. The resulting pairwise distance matrix was then used to construct a neighbor joining tree. The distance represents D . Further details are in Methods.

TABLE 1: Topological characteristics of the meta-metabolic networks generated from bioreactor shotgun metagenomic data.

Dataset	Number of nodes N	Number of edges m	Average degree \bar{k}	Average path length \bar{l}	Average clustering \bar{C}	Assortativity r	Giant component G_c
B1	1015	1042	1.95	10.60	0.026	0.12	151
B2	1130	1148	1.97	10.30	0.023	0.13	164
B3	1060	1092	1.94	10.05	0.024	0.13	164
B4	935	1040	1.80	10.19	0.017	0.12	194
B5	850	894	1.90	9.98	0.019	0.12	143
B6	700	752	1.86	10.13	0.016	0.12	120
B7	617	682	1.81	9.68	0.017	0.14	121
B8	666	720	1.85	9.81	0.016	0.17	118
B9	972	1024	1.90	9.70	0.020	0.11	168
B10	878	945	1.86	9.99	0.020	0.12	158

than the number of nodes (representing metabolites) for each network, ranging from 1.02 the number of nodes for network B2 to 1.11 the number of nodes for network B4. This is consistent with the low average degree, which ranges from 1.70 for B2 to 1.80 for B4, and reflects the observation that most metabolic pathways do not have a large number of branch points, which would lead to a greater proportion of edges in the networks. The average path length ranges from 9.63 (network B10) to 10.60 (network B1), in contrast with other real world networks, where the average path length is typically shorter [44]. This also derives in part from the low number of branch points in the networks, which contributes to the low average clustering, which ranges from 0.016 for networks B7 and B9 to 0.026 for network B1.

Secondly, all the networks are weakly assortative, with values ranging from 0.11 to 0.17. The property of assortativity has been attributed to similarly connected nodes being connected with each other and is typical of social networks, arising from a process of homophily [45]. The presence of assortativity is unusual for a biomolecular network, which is mostly disassortative, indicated by negative assortativity values [45]. The fact that the meta-metabolic networks possess the property of assortativity is interesting and warrants further study. Finally, the giant component was small compared with the total number of nodes in each network (ranging from 118 for network B9 to 194 for network B4). This has implications for the metabolic robustness of the meta-metabolic networks, given that a smaller giant component imposes a limit on the ability of metabolic interconversions to circumvent the absence or removal of a particular metabolic pathway. In addition, it may impact the interactions of individual microbes, limiting interconversion, exchange, and sequestering of metabolites within the microbial community. Those networks with a larger giant component are likely to be more flexible in terms of the ability to metabolize a wider range of substrates, which could be important to scrutinize when assessing the efficiency of a particular bioreactor system. For example, such metabolic flexibility appears to be a key ability of acetogens to compete in anaerobic environments [46]. Such considerations may help to better understand the resilience (or not) of a bioreactor to shocks or system changes.

3.4. A Surprising Lack of Methanogens? The data clearly show that methanogens are only a minor component of the bioreactor microbial communities. Methanogens might reasonably be expected to be dominant if the majority of the carbon waste product is methane, consequently the low proportion of known methanogens in the bioreactors revealed here and in other systems presents something of a puzzle. One potential explanation is that novel, previously uncharacterized methanogens are present. The presence of novel methanogens would have significant evolutionary and environmental implications, but would require confirmation at the genomic level. An alternative explanation is the poor energetics of methane production. The ultimate energy input in the bioreactor are carbohydrates derived from sargassum. While a considerable volume of methane may be produced, this may not result in the production of much cellular energy,

in the form of ATP. For instance, the conversion of acetate to methane by acetoclastic methanogens only produces -36 kJmol^{-1} , which is converted to ATP via an electrochemical gradient [47]. This is expected to translocate two Na^+ ions, four of which are estimated to produce one ATP molecule [47]. Likewise, the conversion of hydrogen and carbon dioxide to methane by hydrogenotrophic methanogens produces less than a mole of ATP per mole of methane produced [48]. In contrast, the energetics of homoacetate fermentation, which results in the production of acetate from glucose, produces up to 4.3 moles of ATP from 1 mole of glucose [46].

The poor energetics of methanogenesis may be reflected in the low growth yields of methanogens, with the highest of up to 7 g per 1 mole of methane produced [48]. This compares with a yield of 19 g per 1 mole glucose for *Escherichia coli* under anaerobic conditions [49]. However, the relationship between ATP production and growth yield is not straightforward, given that not all ATP molecules in the cell are dedicated to biosynthesis [50]. Therefore, it is unclear if the majority of the living biomass in the bioreactors would be expected to belong to those microbes that catalyze the most energetically favorable reactions.

Lastly, a more parochial explanation derives from the relative copy number of rRNA genes in Bacteria compared to Archaea. Differences in copy number can influence 16S rRNA community profile results [51] and the shotgun 16S rRNA analysis, albeit to a lesser extent due to the lack of a PCR amplification step. Archaeal genomes on average have a lower rRNA copy number than bacteria (1.62 ± 0.84 versus 3.82 ± 2.6 , respectively [52]), which might contribute to the apparent observation of a low proportion of Archaea in the bioreactors. The potential bias resulting from differences in rRNA copy numbers between Archaea and Bacteria can be addressed by examining the copy number of the enzymes involved in the different energy producing pathways, specific to autotrophic methanogens or bacterial fermenters. We determined that there was an excess in *pfk* copy number than *mcrA* copy number in all of the datasets (Figure 6, Supplementary Table 6), which is consistent with a minor archaeal component in the community, detected using 16S rRNA approaches. In addition, FTIR measurements show that in the gaseous phase, each bioreactor system produces a similar quantity of CO_2 and CH_4 , although the overall volumetric yield of biogas is one-third higher in the low salinity bioreactor (Supplementary Table 2). When using the partitioning approach described in Methods, it becomes apparent that there is substantially more CO_2 produced in the system than CH_4 ($\text{CO}_2:\text{CH}_4$ ratio 1.5:1.0 for the low salinity bioreactor and 1.6:1.0 for the high salinity bioreactor), although caution should be expressed assigning exact values. These observations are consistent with the observation of a low proportion of Archaea present in the total microbial populations in both bioreactors.

Additional fermentative guilds were examined, using the relative presence of key enzymes as described in Methods (Figure 6, Supplementary Table 6). This approach was utilized as taxonomy is not a good indicator of fermentative ability, in nonmethanogens. In the three shotgun datasets generated in this study, FeFe hydrogenase was the dominant enzyme

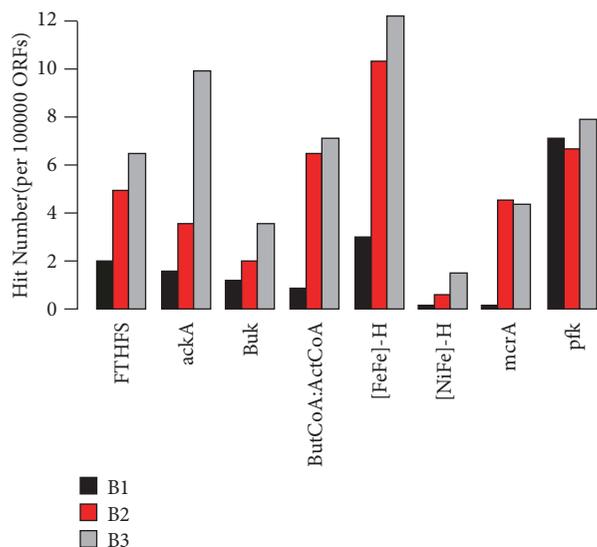


FIGURE 6: Comparison of key enzymes from different fermentative guilds. The relative proportion of key enzymes from the three shotgun datasets generated in this study are shown: formyltetrahydrofolate synthetase (FTHFS), acetate kinase (ackA), BuK, butyrate kinase (BuK), butyryl-CoA:acetate-CoA transferase (butyr-A-CoA), [FeFe]-hydrogenase ([FeFe]-H), [NiFe]-hydrogenase ([NiFe]-H), methyl-coenzyme M reductase (mcrA), and phosphofructokinase (pfk).

present, indicating the importance of hydrogen producing bacteria in the system. Acidogens were also highly represented, with the sum of the three key enzymes (acetate kinase, butyrate kinase, and butyryl-CoA: acetate-CoA transferase) representing the highest enzymatic component of all, however it is not clear if the individual genes frequently coexist within a single genome, and so it is difficult to directly infer the proportion of acidogens present from the data. Finally, the number of formyltetrahydrofolate synthetase homologs is higher than *mcrA* homologs in all three datasets, implying that acetogens outnumber methanogens. The bioinformatics approach may allow the optimization of bioreactor systems with regard to modifying the proportion of acetogens or hydrogen producing bacteria, which both produce substrates for methanogens.

3.5. Syntrophic Interactions. In the process of methanogenesis, a series of pairwise interactions occur between fermentative microbes with differing metabolic capabilities. These interact syntrophically, leading to the conversion of sugars to methane, with the waste product of one fermentative guild acting as a carbon source for another. The fermentative microbes appear to interact cooperatively, implying reciprocity, given that the recipient microbe gains a carbon and energy source, while the donor microbe benefits by the removal of waste products [53]. Such a cooperative interaction may be considered from a game theoretical perspective, with the metabolites representing a public good [54]. Freeloading (selfish behavior) is a problem that besets public goods both at the human [55] and at microbial [54]

levels. In microbes, one solution to the freeloading problem is to develop a close physical interaction between the two reciprocating syntrophic microbes [54]. This means that information regarding syntrophic interactions on a meta-metabolic network, may also contain some information regarding the spatial distribution of different taxonomic groups associated with different metabolic processes.

There are two main ways in which syntrophic interactions might be represented on a meta-metabolic network. Firstly, the taxonomic affiliations of enzymes could help to distinguish complementary enzymatic functions belonging to separate taxonomic groups. Second, information regarding the directionality of enzymes for each biochemical reaction (which is reversible) can ultimately only be generated experimentally, in terms of the enzyme's kinetic rate constants. This information may be incorporated into the network by consideration of sequence homology with enzymes that have had their rate constants experimentally determined. Considerations of syntrophy may be important when considering the metabolic flux through a microbial community [56] (which might be better understood using meta-metabolic networks) and constitute a topic for future investigation.

3.6. Potential Strategies to Improve Methane Production. Our results provide insights into potential approaches to improve methane production in the system. In particular, it is apparent that the proportion of SRB is lower in the high salinity bioreactor, despite the presence of significantly more sulfate in the high salinity bioreactor. Thus, a potential method of suppressing SRB numbers may be to modulate the salinity of the bioreactor, when saltwater is used. In addition, the evidence suggests that the high sulfate concentration in seawater is not nearly as problematic as the literature suggests, because the SRB cannot take full advantage of the extra sulfur. Lastly, our network topology analysis approaches may provide tools to identify or tailor the "ideal" bioreactor microbial community, which might be characterized by a high taxonomic and metabolic diversity leading to more efficient utilization of the complex compounds present in the sargassum feedstock. Such an ideal bioreactor community will have a particular meta-metabolic network which associates with it. Tailoring bioreactor systems would therefore involve modifying the conditions until a similar meta-metabolic network is derived. Clues as to which parameters need changing may be generated by the pathway and taxonomic composition of the respective meta-metabolic network. Macroecological research indicates that taxonomic diversity supports greater productivity of an ecosystem [57]. In our anaerobic bioreactors the metagenomics approach may allow us to observe directly the mechanistic effects of greater microbial taxonomic diversity. This should lead to greater metabolic diversity, in principle leading to more efficient utilization of the energy put into the system.

Data Availability

The sequence data used to support the findings of this study have been deposited in the NCBI Sequence Read Archive.

Additional Points

Highlights. (i) The study shows that bioreactors can run effectively in saltwater using macroalgae as feedstock. (ii) The bioreactor microbial community was examined using metagenomics and a low level of methanogens detected. (iii) Novel bioinformatic approaches were utilized to dissect the taxonomic and metabolic diversity of the bioreactors. (iv) A meta-analysis comparison with other bioreactor meta-metabolic networks was conducted.

Disclosure

The content of this paper is solely the responsibility of the authors and does not necessarily represent the official view of NIGMS or NIH.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We thank Dr. Kai Gribenow (Department of Chemistry, UPR, Rio Piedras) as well as the Office of the Dean of Natural Sciences, the Office of the Dean of Graduate Studies and Research, and the Department of Environmental Sciences of the University of Puerto Rico, Rio Piedras, for their support. This work was funded by DoD Grant W911NF-11-1-0218 and supported by an Institutional Development Award (IDeA) INBRE Grant no. P20GM103475 from the National Institute of General Medical Sciences (NIGMS), a component of the National Institutes of Health (NIH), and the Bioinformatics Research Core of INBRE.

Supplementary Materials

Figure 1: rarefaction curves of community profile metagenomic datasets generated in this study. *Figure 2:* change in α diversity of the bioreactors over time. *Table 1:* loading regime for 2L bioreactors. *Table 2:* peak specific yield and composition of biogas. *Table 3:* summary of information about each sample used in the 16S rRNA community profile meta-analysis. *Table 4:* 16S rRNA community profile datasets. *Table 5:* physical parameters of the bioreactors associated with metagenomics samples. *Table 6:* sequence statistics of shotgun metagenomic datasets. (*Supplementary Materials*)

References

- [1] J. G. Ferry, "How to make a living by exhaling methane," *Annual Review of Microbiology*, vol. 64, pp. 453–473, 2010.
- [2] D. H. Rothman, G. P. Fournier, K. L. French et al., "Methanogenic burst in the end-Permian carbon cycle," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 15, pp. 5462–5467, 2014.
- [3] G. Borrel, P. W. O'Toole, H. M. B. Harris, P. Peyret, J.-F. Brugère, and S. Gribaldo, "Phylogenomic data support a seventh order of methylotrophic methanogens and provide insights into the evolution of methanogenesis," *Genome Biology and Evolution*, vol. 5, no. 10, pp. 1769–1780, 2013.
- [4] K. Ziemiński and M. Frąc, "Methane fermentation process as anaerobic digestion of biomass: transformations, stages and microorganisms," *African Journal of Biotechnology*, vol. 11, no. 18, pp. 4127–4139, 2012.
- [5] M. J. Whitticar, E. Faber, and M. Schoell, "Biogenic methane formation in marine and freshwater environments: CO₂ reduction vs. acetate fermentation-Isotope evidence," *Geochimica et Cosmochimica Acta*, vol. 50, no. 5, pp. 693–709, 1986.
- [6] G. Muyzer and A. J. M. Stams, "The ecology and biotechnology of sulphate-reducing bacteria," *Nature Reviews Microbiology*, vol. 6, no. 6, pp. 441–454, 2008.
- [7] P. Scherer and B. Demirel, "The roles of acetotrophic and hydrogenotrophic methanogens during anaerobic conversion of biomass to methane: a review," *Reviews in Environmental Science and Bio/Technology*, vol. 7, no. 2, pp. 173–190, 2008.
- [8] U. Deppenmeier, "The unique biochemistry of methanogenesis," *Progress in Nucleic Acid Research and Molecular Biology*, vol. 71, pp. 223–283, 2002.
- [9] I. Goodhead and A. C. Darby, "Taking the pseudo out of pseudogenes," *Current Opinion in Microbiology*, vol. 23, pp. 102–109, 2015.
- [10] S. Jaenicke, C. Ander, T. Bekel et al., "Comparative and joint analysis of two metagenomic datasets from a biogas fermenter obtained by 454-pyrosequencing," *PLoS ONE*, vol. 6, no. 1, Article ID e14519, 2011.
- [11] M. T. Wong, D. Zhang, J. Li et al., "Towards a metagenomic understanding on enhanced biomethane production from waste activated sludge after pH 10 pretreatment," *Biotechnology for Biofuels*, vol. 6, no. 1, p. 38, 2013.
- [12] R. Wirth, E. Kovács, G. Maróti, Z. Bagi, G. Rákhely, and K. L. Kovács, "Characterization of a biogas-producing microbial community by short-read next generation DNA sequencing," *Biotechnology for Biofuels*, vol. 5, no. 1, p. 41, 2012.
- [13] L. Solli, O. E. Håvelsrud, S. J. Horn, and A. G. Rike, "A metagenomic study of the microbial communities in four parallel biogas reactors," *Biotechnology for Biofuels*, vol. 7, no. 1, p. 146, 2014.
- [14] M. Cai, D. Wilkins, J. Chen et al., "Metagenomic reconstruction of key anaerobic digestion pathways in municipal sludge and industrial wastewater biogas-producing systems," *Frontiers in Microbiology*, vol. 7, p. 778, 2016.
- [15] A. Li, Y. Chu, X. Wang et al., "A pyrosequencing-based metagenomic study of methane-producing microbial community in solid-state biogas reactor," *Biotechnology for Biofuels*, vol. 6, no. 1, p. 3, 2013.
- [16] M. K. Nobu, T. Narihiro, C. Rinke et al., "Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor," *The ISME Journal*, vol. 9, no. 8, pp. 1710–1722, 2015.
- [17] C. Sundberg, W. A. Al-Soud, M. Larsson et al., "454 pyrosequencing analyses of bacterial and archaeal richness in 21 full-scale biogas digesters," *FEMS Microbiology Ecology*, vol. 85, no. 3, pp. 612–626, 2013.
- [18] I. Sundh, H. Carlsson, Å. Nordberg, M. Hansson, and B. Mathisen, "Effects of glucose overloading on microbial community structure and biogas production in a laboratory-scale anaerobic digester," *Bioresource Technology*, vol. 89, no. 3, pp. 237–243, 2003.

- [19] T. M. J. Fruchterman and E. M. Reingold, "Graph drawing by force-directed placement," *Software: Practice and Experience*, vol. 21, no. 11, pp. 1129–1164, 1991.
- [20] J. G. Caporaso, C. L. Lauber, W. A. Walters et al., "Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 1, pp. 4516–4522, 2011.
- [21] J. G. Caporaso, J. Kuczynski, J. Stombaugh et al., "QIIME allows analysis of high-throughput community sequencing data," *Nature Methods*, vol. 7, no. 5, pp. 335–336, 2010.
- [22] T. Z. DeSantis, P. Hugenholtz, N. Larsen et al., "Greengenes, a chimeric checked 16S rRNA gene database and workbench compatible with ARB," *Applied and Environmental Microbiology*, vol. 72, no. 7, pp. 5069–5072, 2006.
- [23] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Research*, vol. 32, no. 5, pp. 1792–1797, 2004.
- [24] Q. Wang, G. M. Garrity, J. M. Tiedje, and J. R. Cole, "Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy," *Applied and Environmental Microbiology*, vol. 73, no. 16, pp. 5261–5267, 2007.
- [25] R. Schmieder and R. Edwards, "Quality control and preprocessing of metagenomic datasets," *Bioinformatics*, vol. 27, no. 6, pp. 863–864, 2011.
- [26] J. Bengtsson-Palme, M. Hartmann, K. M. Eriksson et al., "METAXA2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomics data," *Molecular Ecology Resources*, vol. 15, no. 6, pp. 1403–1414, 2015.
- [27] C. Camacho, G. Coulouris, V. Avagyan et al., "BLAST+: architecture and applications," *BMC Bioinformatics*, vol. 10, article 421, 2009.
- [28] T. Yamada, I. Letunic, S. Okuda, M. Kanehisa, and P. Bork, "IPATH2.0: Interactive pathway explorer," *Nucleic Acids Research*, vol. 39, no. 2, pp. W412–W415, 2011.
- [29] J. D. Zhang and S. Wiemann, "KEGGgraph: a graph approach to KEGG PATHWAY in R and bioconductor," *Bioinformatics*, vol. 25, no. 11, pp. 1470–1471, 2009.
- [30] M. Heymann, S. Jacomy, and M. Bastian, "Gephi: an open source software for exploring and manipulating networks," in *Proceedings of the International AAAI conference on weblogs and social media*, 2009.
- [31] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *Inter Journal Complex Systems*, p. 1695, 2006.
- [32] N. Saitou and M. Nei, "The neighbor-joining method: a new method for reconstructing phylogenetic trees," *Molecular Biology and Evolution*, vol. 4, no. 4, pp. 406–425, 1987.
- [33] J. Felsenstein, "PHYMLIP - phylogeny inference package (Version 3.2)," *Cladistics*, vol. 5, pp. 164–166, 1989.
- [34] M. Rho, H. Tang, and Y. Ye, "FragGeneScan: predicting genes in short and error-prone reads," *Nucleic Acids Research*, vol. 38, no. 20, p. e191, 2010.
- [35] K. Seuss and E. Chave, "Calcium carbonate supersaturation in seawater: effects of dissolved organic matter," *Limnology and Oceanography*, vol. 15, pp. 633–637, 1970.
- [36] F. J. Roy and R. N. Millero, "A chemical equilibrium model for the carbonate system in natural waters," *Croatica Chemica Acta*, vol. 70, pp. 1–38, 1997.
- [37] K. T. Bird, D. P. Chynoweth, and D. E. Jerger, "Effects of marine algal proximate composition on methane yields," *Journal of Applied Phycology*, vol. 2, no. 3, pp. 207–213, 1990.
- [38] C. Habig, D. A. Andrews, and J. H. Ryther, "Nitrogen recycling and methane production using *Gracilaria tikvahiae*: A closed system approach," *Resources, Conservation & Recycling*, vol. 10, no. 4, pp. 303–313, 1984.
- [39] S. E. McGlynn, "Energy metabolism during anaerobic methane oxidation in ANME archaea," *Microbes and Environments*, vol. 32, no. 1, pp. 5–13, 2017.
- [40] R. E. Ricklefs, "Community diversity: relative roles of local and regional processes," *Science*, vol. 235, no. 4785, pp. 167–171, 1987.
- [41] C. Lozupone, M. Hamady, and R. Knight, "UniFrac—an online tool for comparing microbial community diversity in a phylogenetic context," *BMC Bioinformatics*, vol. 7, article 371, 2006.
- [42] M. F. Polz and C. M. Cavanaugh, "Bias in template-to-product ratios in multitemplate PCR," *Applied and Environmental Microbiology*, vol. 64, no. 10, pp. 3724–3730, 1998.
- [43] N. Shah, H. Tang, H. G. Doak, and T. Ye, "Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics," in *Proceedings of the Pc Symp Bioomput*, pp. 165–176, 2011.
- [44] Q. K. Telesford, K. E. Joyce, S. Hayasaka, J. H. Burdette, and P. J. Laurienti, "The ubiquity of small-world networks," *Brain Connectivity*, vol. 1, no. 5, pp. 367–375, 2011.
- [45] M. E. J. Newman, "Assortative mixing in networks," *Physical Review Letters*, vol. 89, no. 20, Article ID 208701, 2002.
- [46] K. Schuchmann and V. Müller, "Energetics and application of heterotrophy in acetogenic bacteria," *Applied and Environmental Microbiology*, vol. 82, no. 14, pp. 4056–4069, 2016.
- [47] C. Welte and U. Deppenmeier, "Bioenergetics and anaerobic respiratory chains of aceticlastic methanogens," *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, vol. 1837, no. 7, pp. 1130–1147, 2014.
- [48] R. K. Thauer, A.-K. Kaster, H. Seedorf, W. Buckel, and R. Hedderich, "Methanogenic archaea: ecologically relevant differences in energy conservation," *Nature Reviews Microbiology*, vol. 6, no. 8, pp. 579–591, 2008.
- [49] M. Muir, L. Williams, and T. Ferenci, "Influence of transport energization on the growth yield of *Escherichia coli*," *Journal of Bacteriology*, vol. 163, no. 3, pp. 1237–1242, 1985.
- [50] J. B. Cook and G. F. M. Russell, "Energetics of bacterial growth: balance of anabolic and catabolic reactions," *Microbiological Reviews*, vol. 59, pp. 48–62, 1995.
- [51] T. Větrovský and P. Baldrian, "The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses," *PLoS ONE*, vol. 8, no. 2, Article ID e57923, 2013.
- [52] D.-L. Sun, X. Jiang, Q. L. Wu, and N.-Y. Zhou, "Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity," *Applied and Environmental Microbiology*, vol. 79, no. 19, pp. 5962–5969, 2013.
- [53] B. E. L. Morris, R. Henneberger, H. Huber, and C. Moissl-Eichinger, "Microbial syntrophy: Interaction for the common good," *FEMS Microbiology Reviews*, vol. 37, no. 3, pp. 384–406, 2013.
- [54] Ö. Özkaya, K. B. Xavier, F. Dionisio, and R. Balbontín, "Maintenance of microbial cooperation mediated by public goods in single- and multiple-trait scenarios," *Journal of Bacteriology*, vol. 199, no. 22, Article ID e00297-17, 2017.
- [55] G. Hardin, "The tragedy of the commons," *Science*, vol. 162, no. 3859, pp. 1243–1248, 1968.
- [56] M. Hanemaaijer, W. F. Röling, B. G. Olivier, R. A. Khandelwal, B. Teusink, and F. J. Bruggeman, "Systems modeling approaches

for microbial community studies: from metagenomics to inference of the community structure,” *Frontiers in Microbiology*, vol. 6, p. 213, 2015.

- [57] D. Tilman, P. B. Reich, and F. Isbell, “Biodiversity impacts ecosystem productivity as much as resources, disturbance, or herbivory,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 26, pp. 10394–10397, 2012.

