Bioresource Technology 102 (2011) 3730–3739

Contents lists available at ScienceDirect

Bioresource Technology

journal homepage: www.elsevier.com/locate/biortech

A meta-analysis of the microbial diversity observed in anaerobic digesters

Michael C. Nelson [a], Mark Morrison [a,b], Zhongtang Yu [a,b,*]

[a] Environmental Science Graduate Program, The Ohio State University, Columbus, OH 43210, USA
[b] Department of Animal Sciences, The Ohio State University, Columbus, OH 43210, USA

ARTICLE INFO

Article history:
Received 31 August 2010
Received in revised form 23 November 2010
Accepted 25 November 2010
Available online 3 December 2010

Keywords:
Anaerobic digestion
Methane
Biogas
Biomethanation
Microbial diversity

ABSTRACT

In this study, the collective microbial diversity in anaerobic digesters was examined using a meta-analysis approach. All 16S rRNA gene sequences recovered from anaerobic digesters available in public databases were retrieved and subjected to phylogenetic and statistical analyses. As of May 2010, 16,519 bacterial and 2869 archaeal sequences were found in GenBank. The bacterial sequences were assigned to 5926 operational taxonomic units (OTUs, based on ⩾97% sequence identity) representing 28 known bacterial phyla, with *Proteobacteria* (1590 OTUs), *Firmicutes* (1352 OTUs), *Bacteroidetes* (705 OTUs), and *Chloroflexi* (693 OTUs) being predominant. Archaeal sequences were assigned to 296 OTUs, primarily *Methanosaeta* and the uncharacterized WSA2 group. Nearly 60% of all sequences could not be classified to any established genus. Rarefaction analysis indicates that approximately 60% of bacterial and 90% of archaeal diversity in anaerobic digesters has been sampled. This analysis of the global bacterial and archaeal diversity in AD systems can guide future studies to further examine the microbial diversity involved in AD and development of comprehensive analytical tools.

1. Introduction

Anaerobic digestion (AD) of organic wastes is a microbially mediated process whereby complex organic wastes are ultimately converted into methane biogas, a potential renewable energy source. The overall AD process can be conceptually divided into four phases defined by the primary catabolic reactions that occur at each phase: hydrolysis of complex polymers (I, hydrolysis), fermentation of the hydrolysis end-products to short chain fatty acids (SCFAs) (II, acidogenesis), conversion of SCFAs of three or more carbons to primarily acetate (III, syntrophic acetogenesis), and the production of methane (IV, methanogenesis) (Yu et al., 2010). The guilds of microbes involved in each phase of AD are interdependent through cross-feeding and/or maintenance of chemothermodynamic gradients. As a result the AD process can quickly and easily breakdown when one of the four phases is out of balance, such as an accumulation of SCFAs that can lead to acidification of the entire system (Chen et al., 2008). Failures of the AD process in bioreactors treating high strength organic wastes from industrial or agricultural operations can lead to damaging economic losses. As AD is increasingly looked upon as a source of bioenergy, the reliability and stability of AD systems becomes critical to ensuring both reliable energy supplies and uninterrupted core business operations.

Numerous studies have been conducted to gain a better understanding of the microbiomes present in AD reactors and their influence on the efficiency and stability of the AD processes (e.g. reviewed in Chen et al., 2008). While initial studies employed traditional cultivation-based methods, the primary methods in current use are DNA-based molecular biology methods such as cloning and sequencing of either functional or 16S ribosomal RNA (rRNA) genes, FISH, DGGE, single-stranded conformation polymorphisms (SSCP), and quantitative PCR (Leclerc et al., 2004; Malin and Illmer, 2008; Sousa et al., 2007). Because it allows for identification of potential known and unknown microbes present in AD reactors, cloning and sequencing of 16S rRNA genes has been generally favored over other methods.

Most studies to date, however, have focused on a single specific AD system (e.g. upflow anaerobic sludge bed (UASB) reactors or continuous stirred tank reactors (CSTRs) processing a single waste stream (e.g. municipal sewage, brewery wastewater). Many of the datasets published contain a small numbers of cloned sequences (generally <100), thus revealing only a small portion of the full diversity present in anaerobic digesters (e.g. Lefebvre et al., 2007). Some of these studies were further limited by a narrow focus on one particular microbial group such as the Archaea or a particular phylum (e.g. Chouari et al., 2003; Hori et al., 2006). Additionally, many sequences recovered from AD systems were deposited into GenBank but have not yet been reported in the literature, contributing little to no additional information on the microbial diversity and its function. As a result, the understanding of the microbiomes involved in AD is fragmented and likely biased,

* Corresponding author at: Department of Animal Sciences, The Ohio State University, Columbus, OH 43210, USA. Tel.: +1 614 292 3057; fax: +1 614 292 2929.
E-mail address: yu.226@osu.edu (Z. Yu).

0960-8524/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.biortech.2010.11.119

exemplified by these microbiomes still being regarded as a "black box". This knowledge gap limits the understanding of how these complex microbiomes either hamper or enhance the efficiency and stability of AD systems.

A few studies have examined the microbial diversity of anaerobic digesters using relatively large (>200 sequences) 16S rRNA clone libraries (Chouari et al., 2005; Figuerola and Erijman, 2007; Godon et al., 1997; Rivière et al., 2009). Additionally, a few studies have used 454-pyrosequencing, either alone or in combination with the Sanger sequencing technology, to analyze the microbiomes in anaerobic digesters, producing large datasets of short, difficult to classify sequence reads (Krause et al., 2008; Kröber et al., 2009; Schlüter et al., 2008; Zhang et al., 2009). To date, however, there has been no collective overview of the microbial diversity generally found in AD systems. In this study, a meta-analysis was performed on all publicly available 16S rRNA gene sequences generated by Sanger sequencing from various anaerobic digesters in an effort to provide a collective appraisal of the microbial diversity in AD systems. Estimates of the current coverage of the microbial diversity already identified in anaerobic digesters were made and particular gaps in the knowledge and understanding of the microbial populations involved in AD were identified.

2. Methods

2.1. Sequence data collection

Initial sequence sets were obtained from the GenBank (http://www.ncbi.nlm.nih.gov) and RDP (Release 10, http://rdp.cme.msu.edu) databases using the search terms 'anaerobic digester', 'bioreactor', 'CSTR', and 'UASB' in the months through May 01, 2010. All non-16S rRNA sequences were removed and the resulting composite dataset was de-replicated to remove identical records based on Accession Number. Sequences not recovered from methanogenic AD systems, particularly those corresponding only with heavy metal and chlorinated solvent remediation, were manually removed according to the annotation provided in the GenBank sequence records. Published datasets that were not automatically retrieved using the search terms were manually added. Sequences with vector nucleotides were trimmed to leave only nucleotides confirmed as rRNA after alignment against the 16S reference sequences from *Escherichia coli* (Accession Number: U00096) for bacteria or *Methanothermobacter thermoautotrophicus* (Accession Number: AE000666) for archaea. Sequences shorter than 250 bp were removed from the dataset to avoid uncertainties in comparing and classifying short sequences that have little or no sequence overlap. The remaining sequences comprised the redacted composite dataset used in this study.

2.2. Phylogenetic analyses

Sequences were grouped into batches of roughly 5000 sequences by size such that the shortest sequence was no more than 20% shorter than the longest sequence in each batch. Batches were submitted for NAST alignment with the minimum alignment length set to 80% of the shortest sequence in each batch and all other criteria using default values (DeSantis et al., 2006). The resulting aligned sequences were imported into ARB and inserted into the Greengenes database ARB tree using the positional variance by parsimony method (Ludwig et al., 2004). Unaligned sequences were classified *en masse* to taxonomic ranks with the Classifier program implemented as part of the RDP database using default parameters (Wang et al., 2007). Based on the classifications determined with the Classifier program, distance matrices were computed within ARB using Jukes-Cantor correction for the following groups: *Archaea*, *Bacteria*, *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, *Chloroflexi*, and the collected "minor phyla" of *Bacteria* that comprised sequences not assigned to any of the aforementioned phyla. Individual distance matrices were analyzed using MOTHUR to cluster OTUs, generate rarefaction curves, and determine the nonparametric ACE and Chao1 richness estimates (Schloss et al., 2009). A parametric estimation of expected maximum number of OTUs was conducted using the non-linear models procedure (PROC NLIN) of SAS (V9.1, SAS Inst. Inc., Cary, NC). This method fits the monomolecular function to the rarefaction output generated by MOTHUR to determine the asymptote that serves as the upper bound of the curves as previously described (Larue et al., 2005). The value defined by the asymptote is an estimate of the expected maximum species richness complementary to the ACE and Chao1 estimates and has been used previously to estimate maximum species richness (Larue et al., 2005; Youssef and Elshahed, 2008). Unless otherwise stated, the term OTU was defined as a grouping of sequences that share $\leqslant$0.03 sequence dissimilarity and is taken to represent the species taxonomic rank. The following dissimilarity cut-offs were used to approximate other taxonomic ranks: 0.05, genus; 0.10, family; 0.15, class/order; 0.20, phylum (Schloss and Handelsman, 2004). A treemap based on the output from the RDP Classifier was constructed using version 4.1.1 of the program Treemap (http://www.cs.umd.edu/hcil/treemap).

2.3. Nucleotide Accession Numbers

The Accession Numbers for all sequences analyzed in this study are available from the corresponding author. The sequences are currently maintained in an in-house ARB database of anaerobic digester sequences. A copy of this database and the sequence alignment are also available by request from the corresponding author.

3. Results and discussion

This study was conducted as a naïve meta-analysis of all publicly available 16S rRNA gene sequences recovered from AD reactors worldwide. The term naïve is used here to imply that sequences were collected and analyzed irrespective of their previously determined taxonomic associations or other analyses. As AD becomes an increasingly engineered process for waste management and biogas production, it becomes necessary to understand the totality of microorganisms that are able to

Table 1
Diversity satistics for Archaea, Bacteria, and 'Major' phylum groups.

Group	Total sequences	% Unclassified to phylum	# of OTUs[a]	ACE[a]	Chao1[a]	Rarefaction estimation[a]	Current coverage[b] (%)
Archaea	2869	2.15	296	362	336	327	90
Bacteria	16519	16.28	5926	20538	11717	9646	61
Chloroflexi	3744	–	693	3238	1858	1157	60
Proteobacteria	3585	–	1590	6548	3498	2658	60
Firmicutes	2549	–	1352	3184	2674	2298	59
Bacteroidetes	2436	–	705	1494	1221	1076	66

[a] Values were calculated using a 0.03 dissimilarity cut-off.
[b] Coverage = # OTUs / Rarefaction Estimate.