



# Long-term successional dynamics of microbial association networks in anaerobic digestion processes



Linwei Wu<sup>a</sup>, Yunfeng Yang<sup>a, \*</sup>, Si Chen<sup>b</sup>, Mengxin Zhao<sup>a</sup>, Zhenwei Zhu<sup>b, 1</sup>, Sihang Yang<sup>a</sup>, Yuanyuan Qu<sup>c</sup>, Qiao Ma<sup>c</sup>, Zhili He<sup>d</sup>, Jizhong Zhou<sup>a, d, e, \*\*</sup>, Qiang He<sup>b, f, \*\*\*</sup>

<sup>a</sup> State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, China

<sup>b</sup> Department of Civil and Environmental Engineering, The University of Tennessee, Knoxville, TN 37996, USA

<sup>c</sup> Key Laboratory of Industrial Ecology and Environmental Engineering (Ministry of Education), School of Environmental Science and Technology, Dalian University of Technology, Dalian 116024, China

<sup>d</sup> Institute for Environmental Genomics and Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019, USA

<sup>e</sup> Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

<sup>f</sup> Institute for a Secure and Sustainable Environment, The University of Tennessee, Knoxville, TN 37996, USA

## ARTICLE INFO

### Article history:

Received 12 April 2016

Received in revised form

12 July 2016

Accepted 29 July 2016

Available online 2 August 2016

### Keywords:

Anaerobic digestion  
Microbial interactions  
Process stability

## ABSTRACT

It is of great interest to elucidate underlying mechanisms to maintain stability of anaerobic digestion, an important process in waste treatment. By operating triplicate anaerobic digesters continuously for two years, we found that microbial community composition shifted over time despite stable process performance. Using an association network analysis to evaluate microbial interactions, we detected a clear successional pattern, which exhibited increasing modularity but decreasing connectivity among microbial populations. Phylogenetic diversity was the most important factor associated with network topology, showing positive correlations with modularity but negative correlations with network complexity, suggesting induced niche differentiation over time. Positive, but not negative, correlation strength was significantly related ( $p < 0.05$ ) to phylogeny. Furthermore, among populations exhibiting consistent positive correlations across networks, close phylogenetic linkages were evident (e.g. *Clostridiales* organisms). *Clostridiales* organisms were also identified as keystone populations in the networks (i.e., they had large effects on other species), suggestive of an important role in maintaining process stability. We conclude that microbial interaction dynamics of anaerobic digesters evolves over time during stable process performance.

© 2016 Elsevier Ltd. All rights reserved.

## 1. Introduction

Widely applied in wastewater treatment and animal waste management, anaerobic digestion is an important microbial process in waste treatment and renewable energy recovery (Aydin et al., 2015a; Talbot et al., 2008; Zhang et al., 2011). Therefore, it

is crucial to understand the ecology and function of microbial communities involved in anaerobic digestion. With the development and application of molecular microbial ecology techniques, progress has recently been made to characterize microbial community compositions in anaerobic digestion processes (Narihiro and Sekiguchi, 2007). For example, Aydin et al. (2015b, 2016) showed that changes in microbial community composition led to altered biodegradation capacity of organic waste and antibiotics during anaerobic digestion, which linked microbial community compositions to the function of anaerobic digesters. However, more studies are needed that focus on the potential interactions among microbial populations at the whole community level, which is expected to contribute more to system functions than individual populations (Ma et al., 2016).

Microorganisms live within complicated networks through a multitude of interactions, such as mutualism and competition

\* Corresponding author.

\*\* Corresponding author. State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, China.

\*\*\* Corresponding author. Department of Civil and Environmental Engineering, The University of Tennessee, Knoxville, TN 37996, USA.

E-mail addresses: [yangyf@tsinghua.edu.cn](mailto:yangyf@tsinghua.edu.cn) (Y. Yang), [jzhou@ou.edu](mailto:jzhou@ou.edu) (J. Zhou), [qianghe@utk.edu](mailto:qianghe@utk.edu) (Q. He).

<sup>1</sup> Current affiliation: Loyola University Chicago, Institute of Environmental Sustainability, USA.

(Faust and Raes, 2012). However, most of those interactions among microbial populations cannot be directly observed, representing a great challenge for studying population interactions in microbial communities. Network analysis has been used to deduce potential interactions among microbial populations by uncovering strong, non-random associations (Faust et al., 2012). It has been applied to examine complex microbial communities in various habitats, such as oceans (Chow et al., 2014), soils (Barberán et al., 2012), human microbiomes (Faust et al., 2012) and bioreactors (Ju and Zhang, 2015). In addition, network analysis is capable of revealing changes in the topology of microbial networks (Deng et al., 2015; Zhou et al., 2010, 2011). Therefore, network analyses have been considered as powerful tools for studying population interactions in complex microbial communities (Lupatini et al., 2014).

Various approaches of network analyses have been developed and widely applied in functional genomics studies based on gene expression data, including differential equation-based network methods, Bayesian network analyses, and relevance/association network methods (Deng et al., 2012). Among them, the association network method based on co-occurrence/correlation is the most commonly used, owing to its computational simplicity and noise tolerance (Gardner and Faith, 2005). However, most studies employing association network analyses use arbitrary thresholds, thus compromising the constructed networks with subjectivity. To address this, a random matrix theory (RMT)-based approach was developed to objectively identify a threshold for network construction based on microarray data or high-throughput sequencing data (Luo et al., 2006, 2007). This approach was shown to be effective in identifying network interactions among microbial populations (Deng et al., 2012, 2015; Zhou et al., 2010, 2011).

Process stability is highly desirable during anaerobic digestion processes. Previous efforts have primarily focused on the roles of individual populations in process stability, especially on methanogens (Chen and He, 2015; Sekiguchi, 2006). Owing to the importance of microbial interactions in system functions (Ma et al., 2016), herein we evaluated microbial population interactions by performing network analysis of the microbial communities in anaerobic digesters operated continuously for two years. A clear successional pattern was identified, exhibiting increasing modularity but decreasing connectivity between populations over time. Furthermore, microbial phylogenetic diversity was found to be the most important factor associated with network topology, indicative of induced niche differentiation over time.

## 2. Material and methods

### 2.1. Anaerobic digester operation and biomass sampling

Triplicate mesophilic continuous anaerobic digesters, designated as C1, C2 and C3 hereafter, were established and operated with dairy waste as the substrate as previously described (Chen and He, 2015). All anaerobic digesters had a working volume of 3.6 L and were operated at a constant temperature of 35 °C. The hydraulic retention time was maintained at 20 days and the organic loading rate (OLR) was kept at 1.0 g volatile solids (VS)/L/day throughout the two-year study period. Process performance remained stable throughout the study period and biomass samples were collected from the digesters periodically, resulting in a total of 156 samples from 52 time points. All samples were stored at –80 °C before use. The detailed sampling points and process performance parameters are summarized in [Supplementary Table S1](#).

### 2.2. Acquisition and processing of 16S rRNA gene sequences

DNA was extracted from biomass samples using previously

described protocols (Ma et al., 2015). Briefly, biomass samples were suspended in 630 µL DNA-extraction buffer, followed by treatment with 60 µL of a lysozyme mixture (37 °C, 60 min), 60 µL of a protease mixture (37 °C, 30 min), and 80 µL of 20% sodium dodecyl sulfate (37 °C, 90 min). The treated sample suspension was subsequently extracted with phenol–chloroform–isoamyl alcohol (25:24:1) at 65 °C for 20 min and the supernatant was extracted using chloroform–isoamyl alcohol (24:1). DNA extract was then mixed with 0.6 volume of isopropanol and stored at 4 °C overnight. DNA was obtained by centrifugation followed by washing with 70% cold ethanol, drying and resuspension in nuclease-free water. DNA concentration and purity were analyzed with a NanoDrop spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA). The V4 region of microbial 16S rRNA genes was amplified by primer pairs (Wu et al., 2015), 515F (50-GTG CCA GCM GCC GCG GTA A-30) and 806R (50-GGA CTA CHV GGG TWT CTA AT-30). PCR was performed at 94 °C for 1 min; 35 cycles of 94 °C for 20 s, 53 °C for 25 s, and 68 °C for 45 s; and a final extension at 68 °C for 10 min using the AccuPrime High Fidelity Taq Polymerase (Invitrogen, Grand Island, NY, USA). PCR products were pooled and purified using the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA, USA) and amplicon sequencing was performed with the Miseq Illumina platform at the Institute for Environmental Genomics (IEG), University of Oklahoma.

For sequencing data analysis, the primer sequences were trimmed from the paired-end sequences, which were then merged using FLASH. Merged sequences were processed to generate operational taxonomic units (OTUs) by UPARSE at the 97% sequence similarity threshold. Taxonomy was assigned with a confidence cutoff of 50% using the RDP classifier. Phylogenetic trees were then constructed from all representative sequences using the FastTree algorithm (Price et al., 2009). The phylogenetic distance between OTUs was then determined by their relatedness in the phylogenetic tree with function *cophenetic* in R *picante* package. The rRNA gene copy number for each OTU was estimated with the rrnDB database (Stoddard et al., 2014). The OTU matrices were rarefied to 11,558 sequences per sample. The abundance-weighted average rRNA gene copy number was then calculated for each sample.

### 2.3. TaqMan qPCR analysis

TaqMan qPCR analyses were performed with triplicate biomass samples at 15 time points (Day 45, 73, 90, 111, 121, 132, 146, 167, 251, 289, 326, 347, 395, 453 and 501). Genus-specific TaqMan qPCR assays were used to quantify the populations of *Methanosarcina* and *Methanosaeta*. To determine the relative abundance of the both methanogens in the archaeal community, a domain-specific TaqMan qPCR assay was performed to quantify total archaeal populations. The characteristics of TaqMan primer/probe sets used in this study were summarized in [Table S2](#), and the qPCR procedure was performed with a CFX96 Real-Time PCR Detection System (Bio-Rad, Hercules, California, USA) as previously described (Chen and He, 2015). In brief, the qPCR assays were performed in 25 µL reaction volumes with 15 pmol primers, 5 pmol probe, and Brilliant II QPCR Master Mix (Agilent, Santa Clara, California, USA). The thermal cycling was started by an incubation at 50 °C for 2 min and an initial denaturation at 95 °C for 10 min, followed by up to 45 cycles at 95 °C for 30 s and 60 °C (for all primer/probe sets) for 45 s.

### 2.4. Network construction

To construct a time-lag network, it is desirable to use a minimum of 12 samples with consistent time intervals between samples. We fit our time-series data to this criteria by categorizing samples into 9 operational intervals according to sampling time: 1)

31–45 days; 2) 59–73 days; 3) 76–90 days; 4) 122–136 days; 5) 160–171 days; 6) 206–252 days; 7) 309–322 days; 8) 475–501 days; and 9) 735–763 days. Within each operational interval, triplicate 16S rRNA gene sequence datasets from 5 time points (except for operational interval 5, which had only 4 time points) were used to construct a network (Table S1). The 9 networks constructed in this study were thereafter named as N31, N59, N76, N122, N160, N206, N309, N475 and N735, corresponding to the 9 operational intervals, respectively. To minimize the impact of rare OTUs, OTUs detected in less than 80% of the samples (i.e. less than 12 out of 15) were not considered. Rarefied sequence data were log transformed before network construction. Each network was constructed with a time-lagging RMT-based approach (Deng et al., 2015). In brief, the molecular ecological network (MEN), as defined by Deng et al. (2012), was inferred primarily from the Pearson correlation matrix. Because microbial populations may have delayed response, one time point forward or backward was allowed when the correlation coefficient ( $r$ ) was calculated between OTUs. These two  $r$  values, as well as a third  $r$  value calculated without time delay, were then compared and the greatest of the three values was recorded as the correlation coefficient between the two OTUs for developing the correlation matrix. The correlation matrix was subsequently converted into a similarity matrix by taking the absolute values, and the cutoff for absolute  $r$  values was determined based on the RMT-based algorithms. In this study, the cutoff  $r$  value was determined to be 0.84 for all networks except N160, which was constructed with only 4 time points and therefore a cutoff  $r$  value of 0.92 was used for N160. The MEN construction and network topology characterization were processed using the network analysis pipeline at <http://ieg2.ou.edu/MENA>.

Correlation coefficients across all network topological features were also calculated. To remove redundant topological features and get a better representation of network topology, a feature set, without pairwise correlations greater than 0.95, was selected for further analysis. The feature set included number of nodes, number of edges, average degree (avgK), centralization of degree (CD), average cluster coefficient (avgCC), average geodesic distance (GD), centralization of betweenness (CB), centralization of stress centrality (CS), density, number of modules and modularity. Detailed definitions of these network topological features are described in Table 1.

### 2.5. Statistical analysis

A principal component analysis (PCA) was performed to evaluate the temporal dynamics of microbial community composition and network topology. A canonical correspondence analysis-based variation partitioning analysis (VPA) was used to determine the contribution of individual process variables and their interactive effects to the temporal variation in microbial community compositions. The importance of community and process variables for network topological features were estimated using multiple regression of distance matrices (MRM) with the R *ecodist* package. Network topological features as well as community and process variables (mean values during the corresponding time interval) were standardized with function *decostand* in R *vegan* package, and the Euclidean distance matrices for these two datasets were used in MRM models. The relationship between phylogenetic similarity and correlation strength was compared using a Mantel correlogram as calculated with function *mantel.correlog* in R *vegan* package. Pearson correlation coefficients between OTUs in relative abundance were classified into bins. Within each bin, the correlation between phylogenetic distance and Pearson correlation coefficients was determined by Mantel test with 999 randomizations. Subsequently, the Mantel's  $r$  value was plotted against the median of

correlation coefficient bins. Alpha significance values were progressively corrected among the correlation coefficient bins. The density line of phylogenetic distance between positively or negatively linked nodes within a network was determined by function *geom\_density* in R *ggplot2* package. To assign significance to the phylogenetic distances of linked nodes, a Wilcoxon rank-sum test was used to test the differences in phylogenetic distance of positively (or negatively) linked nodes and of nodes within the network.

## 3. Results

### 3.1. Microbial community composition in anaerobic digesters

The results of PCA showed that microbial community compositions in triplicate anaerobic digesters gradually changed with time (Fig. 1a). The difference in microbial community composition increased over time, showing a significant time-decay relationship with a correlation ( $r^2 = 0.42$ ,  $p < 0.001$ ) between microbial community dissimilarity and operation time of the anaerobic digesters (Fig. S1). The relative abundance of methanogen populations also changed with time (Fig. 1b). For example, the relative abundance of *Methanosaeta* between 289 and 501 days was significantly higher than that between 45 and 167 days ( $p < 0.001$ , two-tailed  $t$ -test), while the relative abundance of *Methanosarcina* during 289–501 days was significantly lower than that during 45–167 days ( $p < 0.001$ , two-tailed  $t$ -test). In sharp contrast, the PCA of measured process variables did not show a similar time-decay relationship, which was in accordance with stable operation conditions. The VPA showed that these process variables together could only explain 15.75%, 20.51% and 15.76% of microbial community variations in C1, C2 and C3, respectively. The most important variable was volatile solids concentration (VS), which could explain 3.03%, 5.53% and 3.88% of microbial community variations in C1, C2 and C3, respectively.

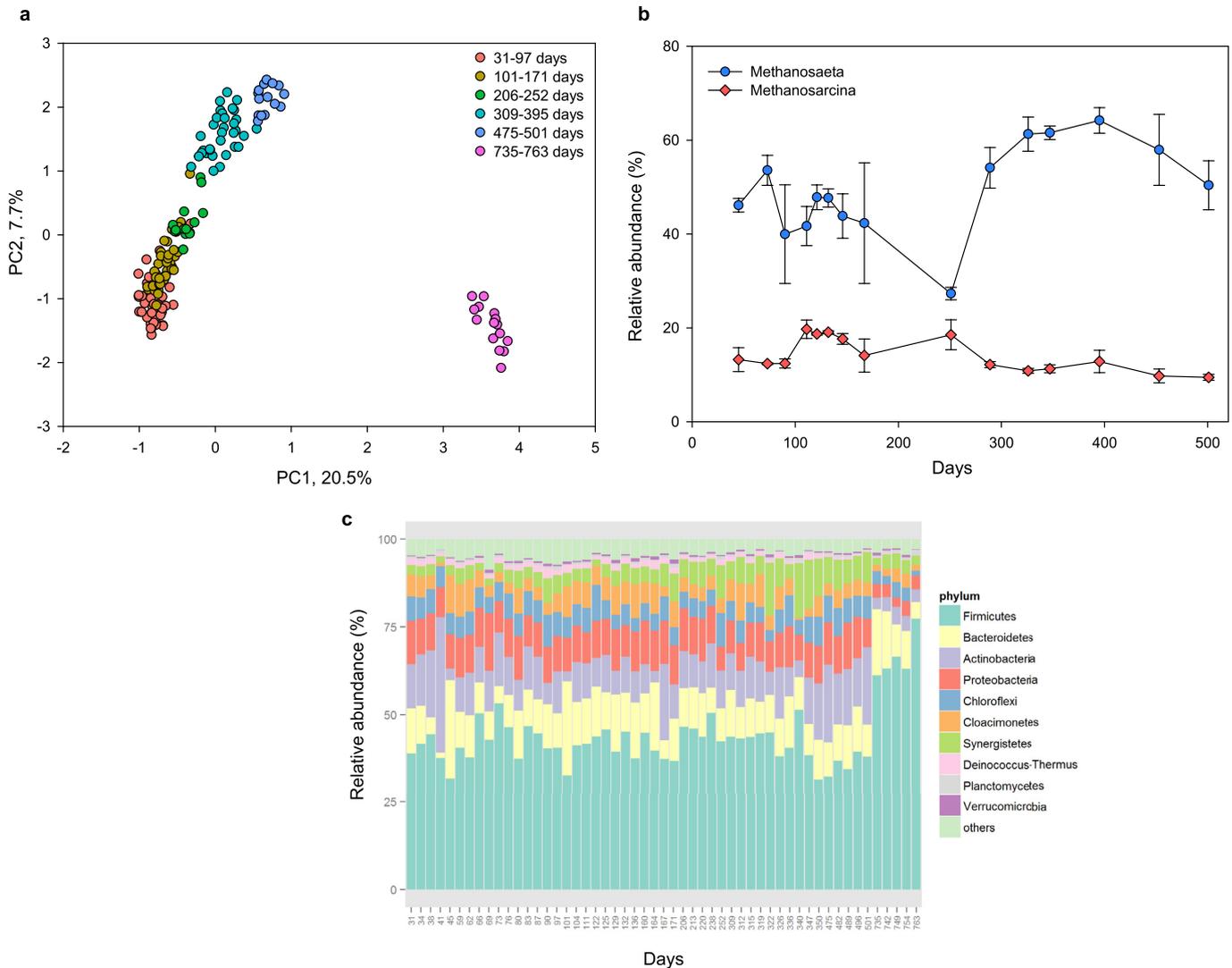
The most abundant bacterial phylum in the anaerobic digesters was *Firmicutes*, which accounted for about 44% of total sequences retrieved. Other abundant phyla included *Bacteroidetes* (12%), *Actinobacteria* (11%), *Proteobacteria* (10%), *Chloroflexi* (6%), *Cloacimonetes* (5%) and *Synergistetes* (5%). At the phylum level, the microbial community composition was relatively stable in the first 500 days (Fig. 1c). However, at the end of the study period (Operational interval 9), the relative abundance of *Firmicutes* rose considerably from 46% to 77% of the community in the anaerobic digesters, which appeared to be a non-random pattern since it was observed in all of the triplicate digesters (Fig. 1c & Fig. S2).

### 3.2. Succession of network topology

Samples were categorized into 9 operational intervals according to sampling time. Accordingly, 9 association networks were constructed for each of the 9 operational intervals, with network nodes representing OTUs and edges representing correlations between OTUs. Network topological features used in this study are summarized in Table 1. Consistent with previous studies (Deng et al., 2012, 2015; Zhou et al., 2010, 2011), the networks exhibited topological features such as scale free, small world and modular (Table S3). Specifically, the network topology fit the power law distribution very well ( $r^2 > 0.96$ ), indicating that some OTUs in the networks had numerous connections while most OTUs had only few connections (i.e., scale free). The average geodesic distance (GD) and average clustering coefficient (avgCC) were significantly different from corresponding randomized networks, which were observed in other networks displaying small-world behavior (Watts and Strogatz, 1998). Modularity values were significantly

**Table 1**  
The network topological features used in this study.

Features	Formula	Explanation	Note
Part I: topological features for individual nodes			
Degree	$k_i = \sum_{j \neq i} a_{ij}$	$a_{ij}$ is the connection strength between nodes $i$ and $j$ .	The number of direct association interactions for a specific OTU.
Stress centrality	$SC_i = \sum_{jk} \sigma(j, i, k)$	$\sigma(j, i, k)$ is the number of shortest paths between nodes $j$ and $k$ that pass through node $i$ .	The number of times a node acts as a bridge along the shortest path between two other nodes.
Betweenness	$B_i = \sum_{jk} \frac{\sigma(j, i, k)}{\sigma(j, k)}$	$\sigma(j, k)$ is the total number of shortest paths between $j$ and $k$ .	The ratio of paths that pass through the $i$ th node, which is a measure of others' dependence on a given node.
Clustering coefficient	$CC_i = \frac{2l_i}{k_i(k_i-1)}$	$l_i$ is the number of links between neighbors of node $i$ and $k_i'$ is the number of neighbors of node $i$ .	Representing how well a node is connected with its neighbors.
Part II: the overall network topological features			
Average degree	$avgK = \frac{\sum_{i=1}^n k_i}{n}$	$k_i$ is degree of node $i$ and $n$ is the number of nodes.	An index of complexity of networks.
Average geodesic distance	$GD = \frac{1}{n(n-1)} \sum_{j \neq i} d_{ij}$	$d_{ij}$ is the shortest path between node $i$ and $j$ .	A smaller $GD$ means that all the nodes in the network are closer.
Centralization of degree	$CD = \sum_{i=1}^n (\max(k) - k_i)$	$\max(k)$ is the maximal value of all degree values. Finally this value is normalized by the theoretical maximum centralization score.	It is close to 1 for a network with star topology and in contrast close to 0 for a network where each node has the same degree.
Centralization of betweenness	$CB = \sum_{i=1}^n (\max(B) - B_i)$	$\max(B)$ is the maximal value of all betweenness values. Finally this value is normalized by the theoretical maximum centralization score.	It is close to 0 for a network where each node has the same betweenness, and the bigger the more difference among all betweenness values.
Centralization of stress centrality	$CS = \sum_{i=1}^n (\max(SC) - SC_i)$	$\max(SC)$ is the maximal value of all stress centrality values. Finally this value is normalized by the theoretical maximum centralization score.	It is close to 0 for a network where each node has the same stress centrality, and the bigger the more difference among all stress centrality values.
Density	$D = \frac{l}{l_{exp}} = \frac{2l}{n(n-1)}$	$l$ is the sum of total links and $l_{exp}$ is the number of possible links	It's also used to describe the network complexity.
Average clustering coefficient	$avgCC = \frac{\sum_{i=1}^n CC_i}{n}$	$CC_i$ is the clustering coefficient of node $i$	It is used to measure the extent of hierarchical structure present in a network.
Modularity	$Q = \frac{1}{2l} \sum_{i \neq j} \left[ A_{ij} - \frac{k_i k_j}{2l} \right] \delta(m_i, m_j)$	$l$ is the sum of total links; $A_{ij}$ is 1 if node $i$ and $j$ are connected and 0 otherwise; $k_i$ is the degree of $i$ ; $m_i$ is the module that $i$ belong to; $\delta(m_i, m_j)$ is 1 if $m_i = m_j$ and 0 otherwise.	It demonstrates how well a network could be naturally divided into modules.



**Fig. 1.** The temporal dynamics of microbial community composition in anaerobic digesters. PCA showed that microbial community composition changed over time (a); the relative abundance of *Methanosaeta* and *Methanosarcina* changed with time (b) and fluctuation of major phyla (exemplified in C1) (c).

higher than those from the corresponding randomized networks. Therefore, these networks appeared to be modular.

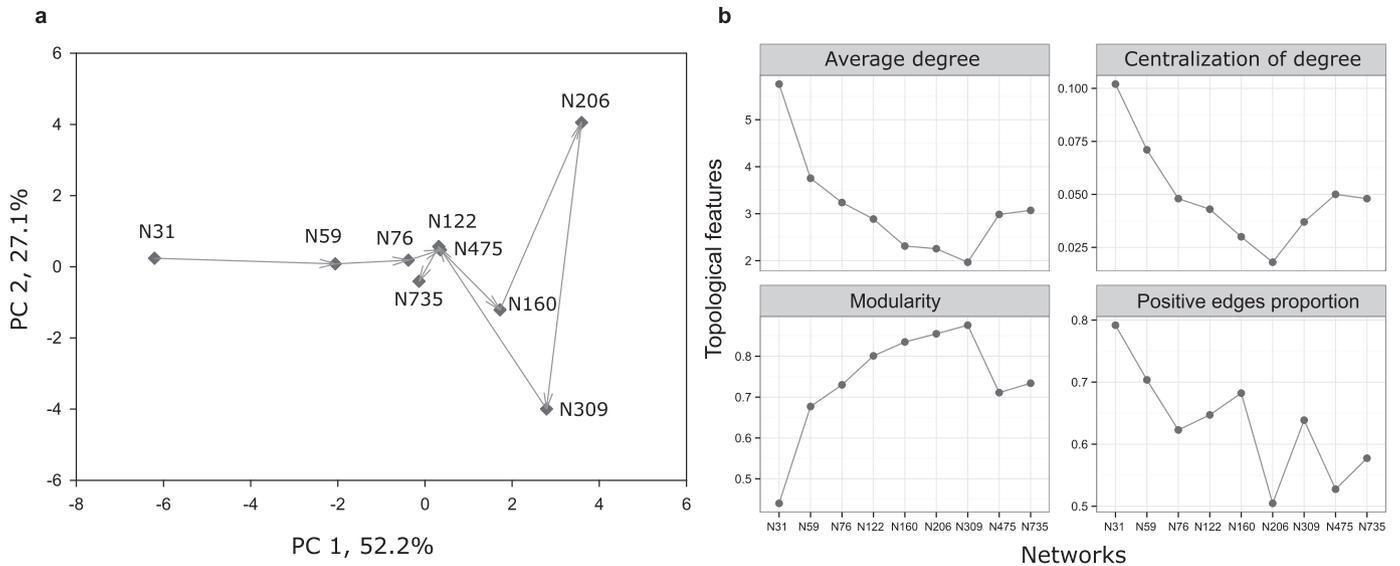
The PCA revealed a successional pattern of network topology among the 9 association networks constructed from each of operational intervals (Fig. 2a). The topology of the first 6 networks, corresponding to 31–252 days of operation, shifted gradually over time. In contrast, the last two networks (N475 and N735, corresponding to operational intervals 8 and 9) shared similar topological properties with that of N122 (network constructed for operational interval 4 during 122–136 days). The average degree and the centralization of degree decreased over time in the first 7 networks, and then increased slightly from N309 to N475 and N735 (Fig. 2b). The modularity of these networks increased in the first 7 networks and then decreased. By calculating the percentage of positive edges in all edges within a network, we found that the proportion of positive edges generally decreased over time.

The topological roles of nodes can be defined by two parameters, within-module connectivity ( $z_i$ ) and among-module connectivity ( $P_i$ ). According to values of  $z_i$  and  $P_i$ , the roles of nodes were classified into four categories: (i) peripherals, which have low  $z$  and  $P$  values, i.e., they have only a few links and almost always to the species within their modules; (ii) connectors, which have a low  $z$

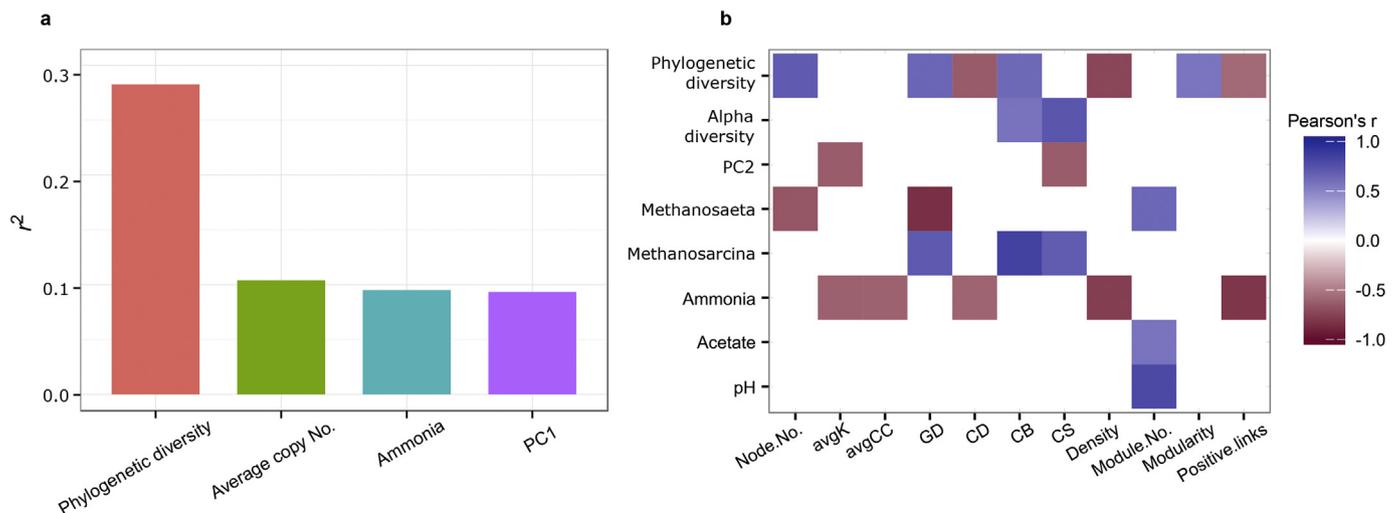
but a high  $P$  value, hence, these nodes are highly linked to several modules; (iii) module hubs, which have a high  $z$  but a low  $P$  value, thus, they are highly connected to many species in their own modules; and (iv) network hubs, which have high  $z$  and  $P$  values, acting as both module hubs and connectors (Deng et al., 2012). Across the 9 networks, only N31 had a network hub, which was OTU99 from genus *Blastopirellula*. A total of 38 OTUs were identified as module hubs; yet only three of them were classified as module hubs in two networks and no module hub was detected more than twice (Table S4), suggesting that hubs were intensively dynamic between different networks. Observations for connectors were similar, with only two of 47 OTUs detected twice as connectors (Table S5). The majority of the hub OTUs were from the phyla *Firmicutes* (69.0%), *Bacteroidetes* (11.9%) and unclassified bacteria (11.9%). OTUs for connectors were mainly from *Firmicutes* (51.1%), *Bacteroidetes* (14.9%) and *Actinobacteria* (10.6%).

### 3.3. Correlation between network topological features and community/process variables

A MRM analysis was performed to estimate the correlations of microbial community composition or process variables with



**Fig. 2.** The successional pattern of network topological features. PCA of the network topological feature set (a) and the change of network average degree, centralization of degree, modularity and the proportion of positive edges over networks (b).



**Fig. 3.** The contributions of the top four most important variables to network topological features (a), and Pearson correlation between community composition or process variables and network topological features (b). The  $r^2$  values in (a) were estimated with the MRM models using the distance matrices calculated from community/process variables and network topological features. In (b), the correlation matrix only keeps Pearson correlations with  $p < 0.1$ . PC1 (PC2) stands for the first (second) axis score of PCA for microbial community compositions; avgK stands for the average degree of network; avgCC stands for average clustering coefficient; GD stands for average geodesic distance; CD stands for centralization of degree; CB stands for centralization of betweenness and CS stands for centralization of stress centrality.

network topological features. Community composition variables included in the MRM model were microbial alpha-diversity (represented by Shannon's index), phylogenetic diversity (measured by Faith's index), the first axis score of PCA (PC1), the second axis score of PCA (PC2), and average rRNA gene copy number of community members. Process variables included were acetate concentration, ammonia concentration, pH, and VS. Collectively, these variables explained 72.8% ( $p = 0.036$ ) of the temporal variations in network topology. Further investigation showed that the most important variable correlated with network topology was phylogenetic diversity, which explained 28.3% ( $p = 0.01$ ) of the network topology variations (Fig. 3a). The next three important variables were average copy number, ammonia concentration and PC1 that explained 10.4%, 9.5% and 9.3% of the variations in network topology.

We further examined the correlations among network topological features, community composition, and process variables (Fig. 3b). The Pearson's correlation analysis showed that node number, GD, centralization of betweenness (CB) and modularity positively correlated with microbial phylogenetic diversity ( $r > 0.57$ ,  $p < 0.10$ ) (Fig. 3b). Centralization of degree (CD), density and the proportion of positive edges negatively correlated with microbial phylogenetic diversity ( $r < -0.57$ ,  $p < 0.10$ ) and ammonia concentration ( $r < -0.60$ ,  $p < 0.10$ ). In addition, average degree and average clustering coefficient negatively correlated with ammonia concentration ( $r < -0.60$ ,  $p < 0.10$ ). Also, CB and centralization of stress centrality (CS) positively correlated with microbial alpha diversity ( $r > 0.58$ ,  $p < 0.10$ ). The number of modules positively correlated with acetate concentration and pH ( $r > 0.57$ ,  $p < 0.10$ ). The network topology also correlated with dynamics of

methanogen populations (Fig. 3b). The node number and CD negatively correlated with the relative abundance of *Methanosaeta* ( $r < -0.65$ ,  $p < 0.10$ ), while GD, CB and CS positively correlated with *Methanosarcina* ( $r > 0.70$ ,  $p < 0.05$ ).

#### 3.4. Topological consistency between OTUs across networks

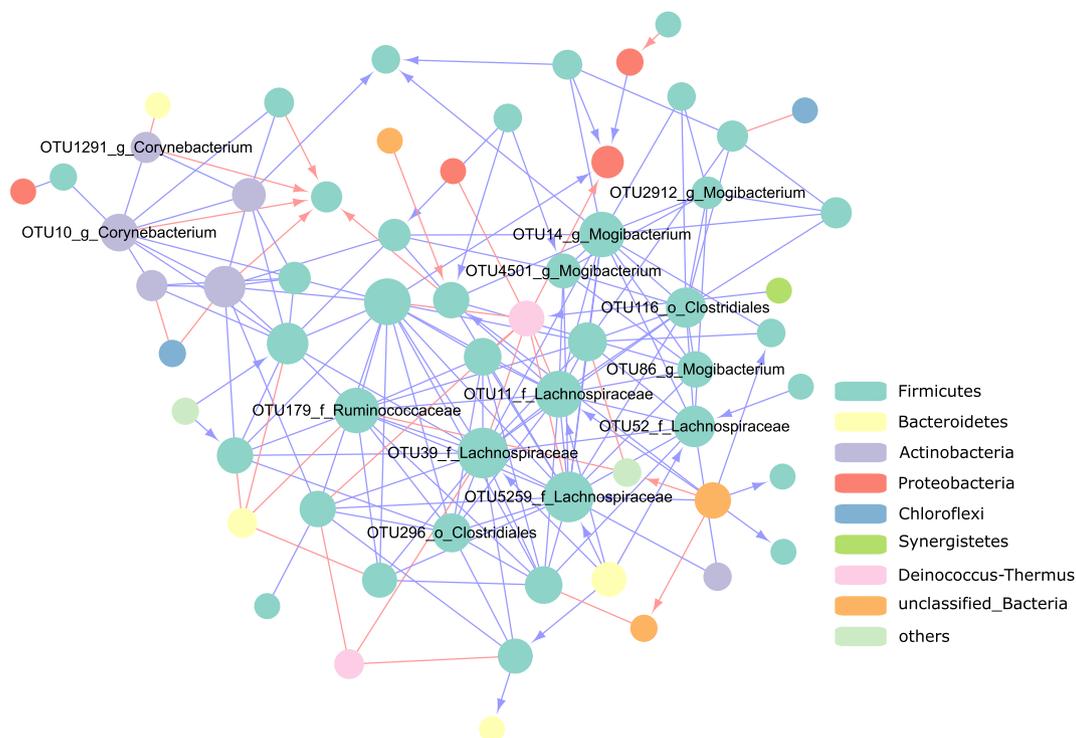
A total of 908 OTUs were present at least once in 9 networks, which resulted in 4724 edges. Only 275 edges were detected in two or more networks, with a majority (89%) being consistently positive or negative (i.e., only 29 edges had both positive and negative correlations in different networks). This topological consistency became stronger when the same edges were present in more networks. Only 24 pairs of correlations were detected in five or more networks, all of which were positive. In fact, we found that 40 pairs of OTUs had consistently negative correlations, but all of them were detected only twice.

The 15 OTUs involved in the 24 pairs of positive correlations identified in five or more networks are summarized in Table S6. Strikingly, these OTUs were closely related, phylogenetically. For example, OTU11 and OTU5259 were positively correlated in all 9 networks, and both belonged to the family *Lachnospiraceae*. The pairs of OTU14 and OTU86, OTU14 and OTU2912, as well as OTU14 and OTU4501, were identified to be positively correlated in 8 networks, and all of them were from the genus *Mogibacterium*. In fact, 13 out of the 15 OTUs were from the same order *Clostridiales*, and were strongly correlated. Most of these OTUs were co-located in the same module across networks (Fig. 4 & Table S6), indicative of stronger interactions than other pairs of OTUs.

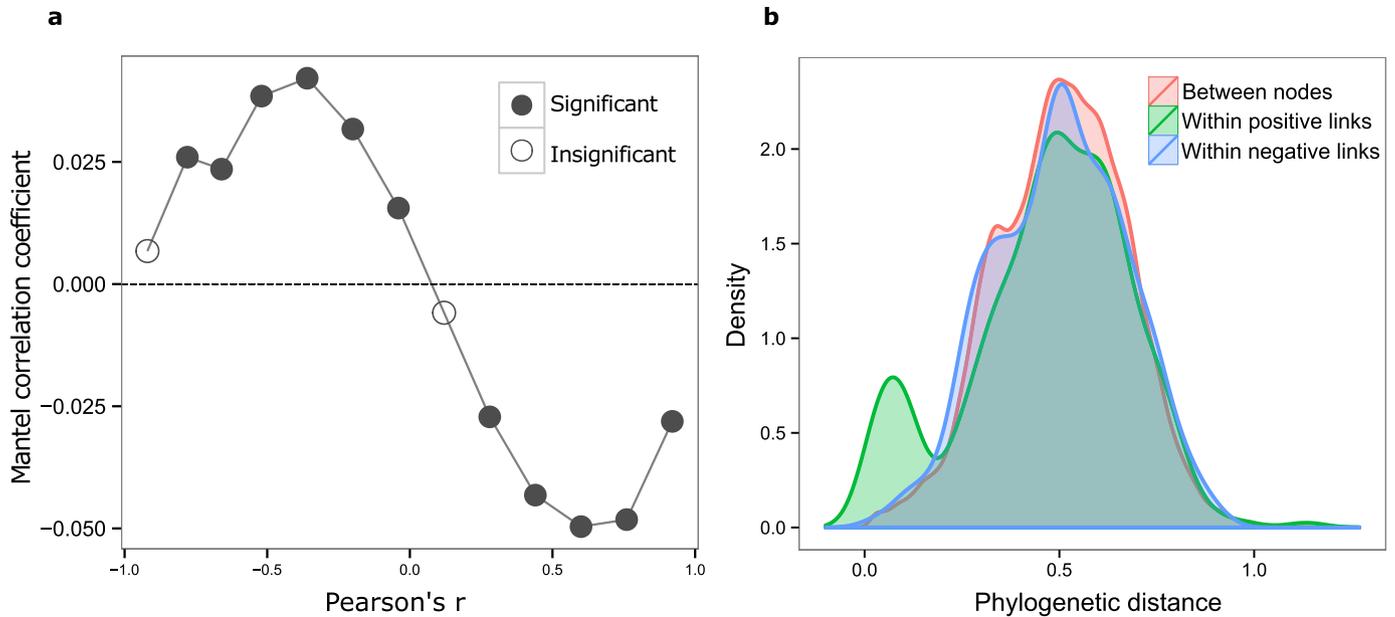
#### 3.5. Linkage between correlation and phylogeny

The Mantel test was performed to determine whether the edge strength (i.e. the correlation strength between nodes) was related to phylogenetic similarity (Fig. 5a & S3). For positively correlated OTUs ( $0.50 < r < 1.00$ ), the Pearson correlation coefficient  $r$  exhibited negative ( $p < 0.01$ ) correlation with the phylogenetic distance, a negative index of phylogenetic similarity. For OTUs with modest negative correlation ( $-0.84 < r < -0.50$ ), the Pearson's  $r$  was positively ( $p < 0.01$ ) correlated with phylogenetic distance, i.e. the negative correlation strength increased (Pearson's  $r$  decreased) with phylogenetic similarity (when phylogenetic distance decreased). In contrast, strong, negative correlations ( $-1.00 \leq r \leq -0.84$ ) were not correlated with phylogenetic relatedness. Analyses of population dynamics data from all operational intervals showed similar results (Fig. S3). Since the first and last bins of Pearson correlation coefficients (i.e., the strong negative and positive correlations) corresponded to the correlation coefficient threshold of networks, these results suggested that only positive, but not negative, correlations in the networks were phylogenetically related.

We examined the phylogenetic distance between positively or negatively linked OTUs within a network. Fig. 5b showed the distribution of phylogenetic distance within positive links, negative links and between any two nodes in network N735. While the distribution of phylogenetic distance between negatively linked nodes was similar to that between any two nodes with a peak around 0.50, there was an obvious peak around 0.10 in the distribution of phylogenetic distance between positively linked nodes. The phylogenetic distance distribution in other 8 networks showed similar patterns (Fig. S4), with the exception that a unique



**Fig. 4.** The subnetwork revealing intense interaction between OTUs from the order *Clostridiales*. The subnetwork was the second module of N76. The node color indicates the phylum of OTUs. Labeled nodes are those detected to be positively correlated in at least 5 networks. The label stands for the lowest taxonomic rank (o., f., and g. representing order, family and genus, respectively) of OTUs, and node size is proportional to its degree. The blue edges represent the positive associations, while the red edges represent the negative associations. The arrow direction is assigned from ahead OTU pointing to lagged one in time dynamics. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 5. The relationship between correlation and phylogeny.** The correlation between phylogenetic distance and correlation strength by Mantel test (a) and the distribution of phylogenetic distance between positively or negatively linked nodes (b). The correlation strength is represented by Pearson correlation coefficient (Pearson's  $r$ ) of OTUs in relative abundance. The Pearson's  $r$  between OTUs were classified into bins (such as  $[-1.0, -0.84]$ ,  $[-0.84, -0.78]$  etc) and Mantel test was applied within each bin to test the Mantel's correlation between phylogenetic distance and correlation strength (i.e., Pearson's  $r$  value). Then, Mantel's correlation coefficients were plotted against the Pearson's  $r$  (median of each bin) in (a). Plot (a) show Mantel test results using the time-series data of OTU dynamics from 31 to 45 days; Plot (b) show the phylogenetic distance distribution based on network N735.

distribution within short phylogenetic distance of positive links was not obvious in N206. In addition, the Wilcoxon rank-sum test showed that the mean of phylogenetic distance between positively linked nodes in all of the 9 networks were significantly smaller ( $p < 0.05$ ) than that of average nodes. In contrast, the mean of phylogenetic distance between negatively linked nodes was not different (Wilcoxon rank-sum test,  $p > 0.10$ ) from the average in most of the networks. Exceptions were N31 and N160, in which the mean of phylogenetic distance between negatively linked nodes were larger than the average (Wilcoxon rank-sum test,  $p < 0.05$ ).

#### 4. Discussion

In this study, we report a RMT-based network analysis to delineate the successional pattern of microbial interactions in anaerobic digesters. The RMT-based analysis is a reliable, sensitive and robust tool for identifying microbial associations with several advantages. First, this approach was developed based on the two universal laws of RMT, which were well characterized and hence reliable as a theoretical foundation. Second, the threshold for constructing a network is automatically defined based on the data structure rather than arbitrarily chosen, and hence no ambiguity exists in network construction. Third, RMT is able to remove noise from nonrandom, system-specific features (Luo et al., 2007), resulting in optimized networks. The association networks constructed by microbial populations in anaerobic digesters revealed topological features frequently observed in complex systems and were different from random expectations, which further verified our RMT-based approach.

Our results show that positive rather than negative correlations were related to phylogeny (Fig. 5). Interestingly, the distribution of phylogenetic distance between positively correlated nodes was nearly bimodal, with one peak close to random distribution and another one close phylogenetic distances. Positive associations could include cross-feeding, co-aggregation in biofilms, co-

colonization and niche overlap (Faust and Raes, 2012). The positive correlation between phylogenetically similar OTUs is usually attributed to niche overlap, as microorganisms similar in phylogeny are likely to behave similarly in niche adaptation (Tringe et al., 2005). Recently, a study based on evolution model simulations suggests that phylogeny could shape interaction networks, wherein more closely related species tend to have stronger mutualistic interactions when interactions are mediated by a mechanism of phenotype matching (Nuismer and Harmon, 2015). However, positive links between two closely related OTUs might be caused by mutual, i.e. false, cross-assignment because taxon assignment at lower taxonomic levels is often inaccurate (Faust and Raes, 2012). If so, consistent positive correlations between *Clostridiales* OTUs might imply that the current taxon classification within *Clostridiales* is too fine.

Negative associations could arise from amensalism, in which one population is harmed without any advantage to the other (Brenner et al., 2008), prey–predator relationships, competition, and/or differential niche adaptation, to name a few (Faust and Raes, 2012). Amensalism and predation were less likely to occur in microbial communities of anaerobic digesters. Rather, we expect negative associations to be prevalently found between closely related OTUs as they would compete for similar resources (Faust and Raes, 2012), and/or between distinct OTUs due to differential niche adaptation, i.e. different ability to adapt to changing environments (Faust et al., 2012). However, the distribution of phylogenetic distance between negatively correlated OTUs were not different from random events in most networks, suggesting that the negative associations likely resulted from stochastic processes.

We found that both microbial community compositions and network topological features changed over time despite stable operation performance. One possible explanation for the succession of the microbial communities is that the characters of sludge gradually changed during long-term operation. For example, the absorption of toxic compounds by anaerobic sludge, such as

antibiotics, could result in the accumulation of such chemicals in sludge (Aydin et al., 2015c). It has been established that network topology reflects interactions between microbial populations (Ma et al., 2016). Our results showed that network N31, representing the first operational interval, had the highest average degree and degree centralization (Fig. 2b). The average degree, representing microbial interaction frequency, decreased steadily from 31 to 322 days. A similar pattern of decrease in degree centralization is indicative of more 'even' networks over time in the anaerobic digesters.

The network N31 had the highest average degree and degree centralization (Fig. 2b). Consistently, N31 was the only network having a network hub (OTU99 from genus *Blastopirellula*). This is the first time that a network hub was identified by microbial network analysis (Deng et al., 2012, 2015; Zhou et al., 2010, 2011). From an ecological perspective, it has been proposed that peripherals in a network might represent specialists whereas module hubs/connectors were generalists and network hubs behave as super-generalists (Olesen et al., 2007). OTU99 was detected in all samples, suggesting that it was a generalist in anaerobic digesters, albeit it comprised only 0.09% of all sequences. A recent study showed that several clones affiliated with *Blastopirellula* were identified as putative anammox bacteria in wastewater treatment plants (Bae et al., 2010), implicating a potentially important role of OTU99 in anaerobic ammonium oxidation process. However, the specific function of this population in anaerobic digestion remains to be characterized.

The modularity of population networks exhibited an increasing trend, in contrast with the decreasing trend of the average degree of the networks (Fig. 2b). The modules could be perceived as functional units in the microbial communities (Luo et al., 2006). Previous studies have also interpreted modules as niches (Chaffron et al., 2010; Eiler et al., 2012). Thus, the increase in network modularity over time might be linked to the greater extent of segregation within the microbial community into finer niches and functional units (i.e., niche differentiation). As a consequence, microbial populations tended to cluster more in subunits (modules) and the interactions among modules were generally reduced over time, resulting in increased modularity and decreased average degree. Similarly, a previous study showed that strong niche differentiation resulted in weaker interactions between soil microbial populations (Faust and Raes, 2012). It could be postulated theoretically, that the number of niches cannot increase indefinitely, owing to the limited availability of resources. Thus, it will eventually reach a balanced or saturated state in which the network topological features would not remain constant, but show fluctuations, according to the theory of stable limit cycles of ecological systems (Holling, 1973). This postulation will be tested in these anaerobic digesters as the operation of the anaerobic digesters analyzed in this study is still ongoing.

We performed an MRM-based analysis to identify community composition and process variables attributable to topological variations over time. It appeared that microbial community composition and structure played more important roles than process variables (Fig. 3), which was not unexpected as the process variables remained relatively unchanged during stable operation. The most important community/process variable that influenced network topology was the phylogenetic diversity of the microbial community. The positive correlation between phylogenetic diversity and network modularity suggests that network modularity is related to niche differentiation as inferred above, since niche differentiation is essential in maintaining population diversity (Leibold and McPeck, 2006).

The interactions between populations, identified by network analysis, were dynamic. For example, 94% of the association

correlations between OTU pairs were detected only once across the 9 networks, which might result from temporary interactions between microorganisms, or just random variations in relative abundance that possibly arose from ecological drift (Hubbell, 2001). It must be noted that association networks were generated from Pearson correlations, which cannot be taken for as direct interactions. However, among OTU pairs identified with significant correlation at least twice, most of which (nearly 90%) were consistent (i.e., either positive or negative). Therefore, these association correlations were less likely to be random. For the OTU pairs that had 'switching' linkages (i.e., sometimes positive and sometimes negative), their interactions might be contingent on the environmental conditions, or a result of random effects. In addition, we found that 69% of hub OTUs belonged to the phylum *Firmicutes*, which was substantially higher than expected by random chance since *Firmicutes* comprised only 44% of total sequences. Most of these *Firmicutes* OTUs were from the order *Clostridiales*, suggesting that this bacterial group, known for capabilities in organic decomposition and fermentation (Desvaux, 2005), could play keystone roles in microbial interactions. It has been suggested that OTUs reflecting keystone taxa (e.g., hub OTUs) are important to maintain the functions of ecosystems and their extinction might lead to community fragmentation (Allesina and Bodini, 2004; González et al., 2010; Lupatini et al., 2014). Thus, the hub OTUs identified by network analysis could play important roles in maintaining the functional stability of digesters. For example, the family *Lachnospiraceae* of the order *Clostridiales* was identified as a major component of animal gastrointestinal tracts (Cotta and Forster, 2006) but has been rarely investigated in anaerobic digesters. However, an unidentified member of *Lachnospiraceae* was found to be very abundant in anaerobic digesters treating poultry litter and was associated with the breakdown of cellulosic biomass (Smith et al., 2014). The identification of *Lachnospiraceae* members as keystone populations in our study (Table S4 & S5) also suggested that the influent microbial community could affect the microbial communities in the digesters. Since the influent was actual dairy waste taken from dairy farms, the influent was expected to harbor a large microbial community which could potentially alter the resident microbial communities in the anaerobic digesters. It is also noted that most of the hub OTUs were not ranked high in the communities (i.e., less abundant) (Table S4 & S5), yet they were likely to play more important roles than peripheral OTUs with high abundance. Future work focusing on uncultured keystone species is crucial to gain a better understanding of the roles of these microorganisms.

## 5. Conclusions

In summary, our study of microbial association networks identified complex interactions among microbial populations and successional pattern of these interactions, exhibiting increasing modularity but decreasing connectivity among microbial populations over time. Phylogenetic diversity appeared to be an important factor associated with network topology since it showed positive correlations with modularity, suggestive of induced niche differentiation over time. Phylogeny could shape the positive, but not negative, interactions among microorganisms, as more closely related populations tended to have stronger positive correlations. Despite the dynamic network topology, there were consistent, positive associations among *Clostridiales* populations across the two-year experimental period, signifying the important roles of the microbial populations involved in lignocellulose degradation and fermentation during anaerobic digestion. These results provide valuable insight into key microbial population interactions potentially important for process stability in anaerobic digestion.

## Acknowledgement

The authors wish to thank Dr. Lauren Hale for language editing, and the anonymous reviewers for constructive comments. This work was partly supported by Major Science and Technology Program for Water Pollution Control and Treatment (2013ZX07315-001-03), a U.S. Environmental Protection Agency Grant XA-83539201, and the Science Alliance—Tennessee Center of Excellence. SC was partly supported by the Institute for a Secure and Sustainable Environment at the University of Tennessee, Knoxville. YY was partly supported from National Science Foundation of China (41471202) and the open funds of state key laboratory of urban and region ecology of China (No. SKLURE2015-2-3). JZ was partly supported from the National Science Foundation of China (41430856) and Collaborative Innovation Center for Regional Environmental Quality.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.watres.2016.07.072>.

## References

- Allesina, S., Bodini, A., 2004. Who dominates whom in the ecosystem? Energy flow bottlenecks and cascading extinctions. *J. Theor. Biol.* 230 (3), 351–358.
- Aydin, S., Ince, B., Cetecioglu, Z., Arikani, O., Ozbayram, E.G., Shahi, A., Ince, O., 2015a. Combined effect of erythromycin, tetracycline and sulfamethoxazole on performance of anaerobic sequencing batch reactors. *Bioresour. Technol.* 186, 207–214.
- Aydin, S., Ince, B., Ince, O., 2015b. Application of real-time PCR to determination of combined effect of antibiotics on Bacteria, Methanogenic Archaea, Archaea in anaerobic sequencing batch reactors. *Water Res.* 76, 88–98.
- Aydin, S., Ince, B., Ince, O., 2015c. Development of antibiotic resistance genes in microbial communities during long-term operation of anaerobic reactors in the treatment of pharmaceutical wastewater. *Water Res.* 83, 337–344.
- Aydin, S., Ince, B., Ince, O., 2016. Assessment of anaerobic bacterial diversity and its effects on anaerobic system stability and the occurrence of antibiotic resistance genes. *Bioresour. Technol.* 207, 332–338.
- Bae, H., Park, K.-S., Chung, Y.-C., Jung, J.-Y., 2010. Distribution of anammox bacteria in domestic WWTPs and their enrichments evaluated by real-time quantitative PCR. *Process Biochem.* 45 (3), 323–334.
- Barberán, A., Bates, S.T., Casamayor, E.O., Fierer, N., 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* 6 (2), 343–351.
- Brenner, K., You, L., Arnold, F.H., 2008. Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol.* 26 (9), 483–489.
- Chaffron, S., Rehrauer, H., Pernthaler, J., von Mering, C., 2010. A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res.* 20 (7), 947–959.
- Chen, S., He, Q., 2015. Persistence of Methanosaeta populations in anaerobic digestion during process instability. *J. Ind. Microbiol. Biotechnol.* 42 (8), 1129–1137.
- Chow, C.E., Kim, D.Y., Sachdeva, R., Caron, D.A., Fuhrman, J.A., 2014. Top-down controls on bacterial community structure: microbial network analysis of bacteria, T4-like viruses and protists. *ISME J.* 8 (4), 816–829.
- Cotta, M., Forster, R., 2006. *The Prokaryotes*. Springer, pp. 1002–1021.
- Deng, Y., Jiang, Y.-H., Yang, Y., He, Z., Luo, F., Zhou, J., 2012. Molecular ecological network analyses. *BMC Bioinforma.* 13 (1), 113.
- Deng, Y., Zhang, P., Qin, Y., Tu, Q., Yang, Y., He, Z., Schadt, C.W., Zhou, J., 2015. Network succession reveals the importance of competition in response to emulsified vegetable oil amendment for uranium bioremediation. *Environ. Microbiol.* 18 (1), 205–218.
- Desvaux, M., 2005. *Clostridium cellulolyticum*: model organism of mesophilic cellulolytic clostridia. *FEMS Microbiol. Rev.* 29 (4), 741–764.
- Eiler, A., Heinrich, F., Bertilsson, S., 2012. Coherent dynamics and association networks among lake bacterioplankton taxa. *ISME J.* 6 (2), 330–342.
- Faust, K., Raes, J., 2012. Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* 10 (8), 538–550.
- Faust, K., Sathirapongsasuti, J.F., Izard, J., Segata, N., Gevers, D., Raes, J., Huttenhower, C., 2012. Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.* 8 (7) e1002606–e1002606.
- Gardner, T.S., Faith, J.J., 2005. Reverse-engineering transcription control networks. *Phys. Life Rev.* 2 (1), 65–88.
- González, A.M.M., Dalsgaard, B., Olesen, J.M., 2010. Centrality measures and the importance of generalist species in pollination networks. *Ecol. Complex.* 7 (1), 36–43.
- Holling, C.S., 1973. Resilience and stability of ecological systems. *Annu. Rev. Ecol. Syst.* 1–23.
- Hubbell, S.P., 2001. *The Unified Neutral Theory of Biodiversity and Biogeography* (MPB-32). Princeton University Press.
- Ju, F., Zhang, T., 2015. Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. *ISME J.* 9 (3), 683–695.
- Leibold, M.A., McPeck, M.A., 2006. Coexistence of the niche and neutral perspectives in community ecology. *Ecology* 87 (6), 1399–1410.
- Luo, F., Yang, Y., Zhong, J., Gao, H., Khan, L., Thompson, D.K., Zhou, J., 2007. Constructing gene co-expression networks and predicting functions of unknown genes by random matrix theory. *BMC Bioinforma.* 8 (1), 1.
- Luo, F., Zhong, J., Yang, Y., Zhou, J., 2006. Application of random matrix theory to microarray data for discovering functional gene modules. *Phys. Rev. E* 73 (3), 031924.
- Lupatini, M., Suleiman, A.K., Jacques, R.J., Antonioli, Z.I., de Siqueira Ferreira, A., Kuramae, E.E., Roesch, L.F., 2014. Network topology reveals high connectance levels and few key microbial genera within soils. *Front. Environ. Sci.* 2, 10.
- Ma, B., Wang, H., Dsouza, M., Lou, J., He, Y., Dai, Z., Brookes, P.C., Xu, J., Gilbert, J.A., 2016. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME J.* 10, 1891–1901.
- Ma, Q., Qu, Y., Shen, W., Zhang, Z., Wang, J., Liu, Z., Li, D., Li, H., Zhou, J., 2015. Bacterial community compositions of coking wastewater treatment plants in steel industry revealed by Illumina high-throughput sequencing. *Bioresour. Technol.* 179, 436–443.
- Narihiro, T., Sekiguchi, Y., 2007. Microbial communities in anaerobic digestion processes for waste and wastewater treatment: a microbiological update. *Curr. Opin. Biotechnol.* 18 (3), 273–278.
- Nuismer, S.L., Harmon, L.J., 2015. Predicting rates of interspecific interaction from phylogenetic trees. *Ecol. Lett.* 18 (1), 17–27.
- Olesen, J.M., Bascompte, J., Dupont, Y.L., Jordano, P., 2007. The modularity of pollination networks. *Proc. Natl. Acad. Sci.* 104 (50), 19891–19896.
- Price, M.N., Dehal, P.S., Arkin, A.P., 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26 (7), 1641–1650.
- Sekiguchi, Y., 2006. Yet-to-be cultured microorganisms relevant to methane fermentation processes. *Microbes Environ.* 21 (1), 1–15.
- Smith, A.M., Sharma, D., Lappin-Scott, H., Burton, S., Huber, D.H., 2014. Microbial community structure of a pilot-scale thermophilic anaerobic digester treating poultry litter. *Appl. Microbiol. Biotechnol.* 98 (5), 2321–2334.
- Stoddard, S.F., Smith, B.J., Hein, R., Roller, B.R., Schmidt, T.M., 2014. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.* gku1201.
- Talbot, G., Topp, E., Palin, M., Masse, D., 2008. Evaluation of molecular methods used for establishing the interactions and functions of microorganisms in anaerobic bioreactors. *Water Res.* 42 (3), 513–537.
- Tringe, S.G., Von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C., 2005. Comparative metagenomics of microbial communities. *Science* 308 (5721), 554–557.
- Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. *Nature* 393 (6684), 440–442.
- Wu, L., Wen, C., Qin, Y., Yin, H., Tu, Q., Van Nostrand, J.D., Yuan, T., Yuan, M., Deng, Y., Zhou, J., 2015. Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. *BMC Microbiol.* 15 (1), 125.
- Zhang, Y., Cañas, E.M.Z., Zhu, Z., Linville, J.L., Chen, S., He, Q., 2011. Robustness of archaeal populations in anaerobic co-digestion of dairy and poultry wastes. *Bioresour. Technol.* 102 (2), 779–785.
- Zhou, J., Deng, Y., Luo, F., He, Z., Tu, Q., Zhi, X., 2010. Functional molecular ecological networks. *MBio* 1 (4) e00169–10.
- Zhou, J., Deng, Y., Luo, F., He, Z., Yang, Y., 2011. Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO<sub>2</sub>. *MBio* 2 (4) e00122–00111.