# Operation-driven heterogeneity and overlooked feed-associated populations in global anaerobic digester microbiome

Mei, Ran; Nobu, Masaru K.; Narihiro, Takashi; Kuroda, Kyohei; Muñoz Sierra, Julian; Wu, Zhuoying; Ye, Lin; Lee, Patrick K.H.; Lee, Po Heng; van Lier, Jules B.

**DOI**
10.1016/j.watres.2017.07.050

**Publication date**
2017

**Document Version**
Accepted author manuscript

**Published in**
Water Research

**Citation (APA)**
Mei, R., Nobu, M. K., Narihiro, T., Kuroda, K., Muñoz Sierra, J., Wu, Z., Ye, L., Lee, P. K. H., Lee, P. H., van Lier, J. B., McInerney, M. J., Kamagata, Y., & Liu, W. T. (2017). Operation-driven heterogeneity and overlooked feed-associated populations in global anaerobic digester microbiome. *Water Research*, *124*, 77-84. https://doi.org/10.1016/j.watres.2017.07.050

Postprint of Water Research
Volume 124, 1 November 2017, Pages 77-84
Link to formal publication (Elsevier): https://doi.org/10.1016/j.watres.2017.07.050

# Operation-driven Heterogeneity and Overlooked Feed-associated Populations in Global Anaerobic Digester Microbiome

Ran Mei[a], Masaru K. Nobu[a,b], Takashi Narihiro[b], Kyohei Kuroda[a,c], Julian Muñoz Sierra[d], Zhuoying Wu[e], Lin Ye[f], Patrick K. H. Lee[g], Po-Heng Lee[e], Jules B. van Lier[d], Michael J. McInerney[h], Yoichi Kamagata[b], Wen-Tso Liu[a*]

a. Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

b. Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan

c. Department of Environmental Systems Engineering, Nagaoka University of Technology, Kami-tomioka, Niigata, Japan

d. Section Sanitary Engineering, Department of Water Management, Delft University of Technology, Delft, The Netherlands

e. Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

f. School of the Environment, Nanjing University, Nanjing, Jiangsu, China

g. School of Energy and Environment, City University of Hong Kong, Kowloon, Hong Kong

h. Department of Botany and Microbiology, University of Oklahoma, Norman, OK, USA

*Corresponding author: Wen-Tso Liu

Email: wtliu@illinois.edu

**Abstract**

Anaerobic digester (AD) microbiomes harbor complex, interacting microbial populations to achieve biomass reduction and biogas production, however how they are influenced by operating conditions and feed sludge microorganisms remain unclear. These were addressed by analyzing the microbial communities of 90 full-scale digesters at 51 municipal wastewater treatment plants from five countries. Heterogeneity detected in community structures suggested that no single AD microbiome could be defined. Instead, the AD microbiomes were classified into eight clusters driven by operating conditions (*e.g.*, pretreatment, temperature range, and salinity), whereas geographic location of the digesters did not have significant impacts. Comparing digesters populations with those present in the corresponding feed sludge led to the identification of a hitherto overlooked feed-associated microbial group (*i.e.*, the residue populations). They accounted for up to 21.4% of total sequences in ADs operated at low temperature, presumably due to ineffective digestion, and as low as 0.8% in ADs with pretreatment. Within each cluster, a core microbiome was defined, including methanogens, syntrophic metabolizers, fermenters, and the newly described residue populations. Our work provides insights into the key factors shaping full-scale AD microbiomes in a global scale, and draws attentions to the overlooked residue populations.

**Keywords**

Anaerobic digester, microbiome, operation, feed sludge

## 1. Introduction

Wastewater treatment processes, including primary treatment for solids separation and secondary treatment for carbon and nutrients removal, produce substantial amount of waste sewage sludge. For example, the amount of waste sludge generated in European Union is estimated to exceed 13 million dry solid tons in 2020 (Kelessidis and Stasinakis 2012). Anaerobic digestion (AD) has been used worldwide to simultaneously degrade waste sludge and produce methane, and is an promising solution to treat the increasing global growth of organic solid waste (Appels et al. 2011). Meanwhile, the microbial community involved in AD is complex (Narihiro et al. 2015) and a better understanding of the AD ecosystem would optimize existing processes and enhance the engineering application (Vanwonterghem et al. 2014).

To identify critical populations responsible for the AD process, multiple researches have tried to define the core AD microbiome. Campanaro et al. (Campanaro et al. 2016) and Treu et al. (Treu et al. 2016) analyzed metagenomic sequences of mesophilic and thermophilic lab-scale digesters treating cattle manure, and concluded that 77 out of 265 genome bins could be considered as the core essential microbial groups in biogas production. Our recent study analyzed the microbial communities of three full-scale digesters in the a wastewater treatment plant and observed a core microbiome that accounted for 59% of the total 16S rRNA gene sequences (Mei et al. 2016a). Studies investigating multiple full-scale plants reported that core populations constituted 36.4% of the total 16S rRNA gene sequences in seven digesters from Seoul, South Korea (Lee et al. 2012), and 28% of the total 16S rRNA gene sequences in seven digesters from France, Germany, and Chile (Riviere et al. 2009). De Vrieze et al. (De Vrieze et al. 2015) evaluated the microbial communities of 29 AD installations whose locations were not specified, and reported that *Clostridiales* and *Bacteroidales* were part of the core microbiome as

78 they were shared by each sample with >0.1% abundance. So, if a large number of digesters are
79 sampled and multiple operating parameters are considered, such as temperature, ammonia
80 concentration, and system configuration that are known to influence AD community (De Vrieze
81 et al. 2015, Smith et al. 2017), would it be still possible to define a core AD microbiome?
82 Furthermore, geographical differences in microbiomes have been observed for waste-treating
83 ecosystems like activated sludge (Zhang et al. 2012) and solid waste landfill (Stamps et al. 2016).
84 Would a similar difference be observed with the AD microbiome?

85 A classic categorization of microorganisms in AD consists of fermenting bacteria
86 (fermenters), syntrophic metabolizers (syntrophs), and methanogenic archaea (methanogens)
87 (Schink and Stams 2006). However, it has been realized that AD microbiome embraces a large
88 proportion of prokaryotes with unrecognized ecophysiology (Narihiro 2016). For example, our
89 recent study (Mei et al. 2016a) revealed that 25% of the AD populations in one wastewater
90 treatment plant migrated from the upstream activated sludge process and remained as residue
91 populations in AD. The presence of those non-anaerobic residue populations has not been widely
92 examined to test whether it is a common phenomenon in all digesters under different operating
93 conditions from different geographical locations. Furthermore, the microbial populations in
94 activated sludge can vary considerably due to differences in process configuration and
95 geographical locations (Zhang et al. 2012). Thus, it is not clear whether such variations of
96 microbial populations in the feed sludge impacts the AD microbiome.

97 In this study, we used high-throughput sequencing technologies to characterize
98 microbiomes in digesters around the world by sampling 90 full-scale digesters with diverse
99 operating conditions and feed sludge characteristics from 51 municipal wastewater treatment
100 plants. The impacts of operating conditions and geographical locations on AD microbiome were

examined. Clustering of samples was performed and cluster-specific core populations were identified. Within the AD microbiome, feed-derived populations were investigated and the distribution in different digesters was characterized.

## 2. Materials and methods

### *2.1. Sample collection*

In total, 148 digester sludge samples were collected from 90 full-scale ADs in 51 municipal wastewater treatment plants. Feed sludge in 27 plants were collected prior to entering ADs, and feed sludge in the rest plants were not collected due to sampling difficulties. All operation-related information was provided by the plant operators. Besides the volatile solids reduction (VSR) provided by plant operators, we calculated VSR values using the Van Kleeck equation according to the USEPA regulation (Regulations 2003), which were further used in the downstream analyses. Most plants were operated with the conventional primary-secondary (activated sludge) treatment scheme, while three plants were only configured with primary treatment before AD (plant CAII, CALG, and USRA). Seven plants (JPHW, JPMU, JPNA, JPST, JPTB, JPYS, and USDV) used a two-stage anaerobic digestion process with similar sludge retention time (the first digester treating sludge from primary/secondary clarifiers and the second digester treating sludge from the first digester). Seven plants (JPHG, JPNA, JPNG, USST, USUR, NEAV, and USCA) introduced external solid wastes into digesters, such as food waste, green waste, and sludge from other sources. Wastewater to two Hong Kong plants (HKST and HKTP) had approximately 1/4 to 1/5 of seawater of high salinity. Due to its high saline nature with high sulfate content, these two AD digesters dosed ferric chloride ($FeCl_3$) to suppress sulfide production, leading to a chloride concentration of 4,000 to 6,000 mg/L (Koenig and Bari

2001, Zhang et al. 2012). Wastewater to another Hong Kong plant (HKYL) had effluent from the tannery industry and contained high concentrations of Zn and Cr (Wong et al. 2001). Digester NEAV1 had both high salinity influent (electrical conductivity about 30-35 mS/cm) and external food waste sludge simultaneously. Digesters from Hong Kong and US (except for USWA and USSF) were sampled at multiple time points with at least one-month interval. These multiple tome points samples were considered as different samples. Fifty milliliters of sludge were collected from the recirculation lines of digesters, transported to laboratory in UIUC on ice (including international samples), and stored at -80°C until DNA extraction.

*2.2. 16S rRNA gene sequencing*

Genomic DNA was extracted from 2 mL of well-mixed sludge using the FastDNA SPIN Kit for Soil (MP Biomedicals, Carlsbad, CA, USA), and quantified using a Nanodrop 2000c spectrophotometer. For PCR amplification, 60 ng of genomic DNA was added into a total reaction volume of 25 μL as template. With a dual-indexing approach (Kozich et al. 2013), a universal primer set 515F (5'-GTGCCAGCMGCCGCGGTAA-3')/909R(5'-CCCCGYCAATTCMTTTRAGT-3') targeting the V4-V5 region of both bacterial and archaeal 16S rRNA gene was used for PCR amplification. PCR was performed with the thermal cycling protocol consisting of initial denaturation (94°C, 3 min), 25 cycles of denaturation (94 °C, 30 s), annealing (55 °C, 45 s) and extension (72 °C, 1 min), and a final extension (72 °C, 10 min) (Mei et al. 2016b). The PCR amplicons were purified using the Wizard SV Gel and PCR Clean-Up system (Promega, Fichburg, WI, USA) and quantified by Qubit 2.0 Fluorometer. Library preparation and sequencing on Illumina Miseq Bulk 2 × 300 nt paired-end system was performed at the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign, IL, USA.

147 *2.3. Microbial community analyses*

148 Paired-end raw sequences were assembled, screened, and trimmed using Mothur 1.33.3
149 (Schloss et al. 2009) with a maximum sequence length of 400 bp and a quality score of 20. The
150 output data were analyzed using QIIME 1.9.1 (Caporaso et al. 2010b) for OTU (operational
151 taxonomic unit, 97% sequence similarity) picking with the *de novo* strategy, which included
152 OTU grouping by UCLUST (Edgar 2010), alignment by PyNAST (Caporaso et al. 2010a),
153 chimera identification by ChimeraSlayer (Haas et al. 2011), taxonomy assignment by BLAST
154 using reference sequences in the GreenGene 2013 database. After removing singletons (OTUs
155 that only had one sequence in the entire dataset), all samples were rarefied to an even depth of
156 20,957 sequences (determined by the sample with fewest sequences). Shannon index ($H =$
157 $-\sum p_i ln p_i$, $p_i$ is the relative abundance of an individual population) calculation, UniFrac
158 distance matrix calculation, Bray-Curtis distance matrix calculation, principal coordinate
159 analysis (PCoA), and unweighted pair group method with arithmetic mean (UPGMA) with 100
160 iterations were all performed using QIIME. Relative abundance was calculated from OTU table.
161 Phylogenetic trees was constructed using the methods of neighbor joining and parsimony
162 provided in ARB program (Ludwig et al. 2004).

163 Statistical differences of principal components between samples from different locations
164 were evaluated using Mann Whitney U test with Bonferroni correction with R (Ihaka and
165 Gentleman 1996). A p-value < 0.01 was considered as statistical significance. Correlations
166 between microbial groups, alpha diversity, and VSR were determined using the Spearman's Rank
167 Order Correlation test with R. Evaluation of normality of the data using Shapiro Wilk Normality
168 test, and preparation of box plot and histogram were also performed using R. Distance-based
169 linear model (DistLM) and analysis of similarity (ANOSIM) were performed with Primer 6

(Clarke 1993). Raw Illumina sequences obtained in this study have been deposited in DDBJ/NCBI/EMBL-EBI under the accession number DRA005150.

## 3. **Results**

### *3.1. Operation-driven heterogeneity of AD microbiome*

In total, over 7 million quality-filtered, non-chimeric sequences were obtained from 148 AD samples in 51 municipal wastewater treatment plants (Fig. S1, Table S1 in the Supplementary material). After removing singletons and subsampling to an even depth (20,957 sequences per sample, determined by the sample with fewest sequences), each AD sample on average contained 1,844 OTUs with a high standard deviation of 595 OTUs. The Shannon index that characterized both richness and evenness of a community showed large variations (Fig. S2), with the highest value being 2.5 times higher than the lowest value (9.12 vs. 3.68). Dissimilarity between AD communities was also reflected in the large variations in the relative abundance of major phyla (Fig. S3). For example, the abundance of *Bacteroidetes* varied from 5% to 71% in different samples, and the abundance of *Thermotogae* varied from 0 to 56%.

Principal coordinate analysis (PCoA) performed on beta-diversity (weighted UniFrac distance) showed that there were different types of AD communities (Fig. S4). However, the variance could not be explained by geographical locations, as only North America samples significantly differed from Hong Kong samples in PC1 and from Japan samples in PC2. In addition, only small portions of the variance could be explained by single environmental parameters such as temperature (9.63%), pH (3.22%), and sludge retention time (SRT) (1.63%) (Table S2A).

To identify shaping factors of the heterogeneous AD communities, the dissimilarity based on weighted Unifrac was further analyzed using unweighted pair group method with arithmetic mean (UPGMA), a clustering method that could fully reveal the variance in beta diversity. Eight clusters were observed (Fig. 1). Cluster A contained six samples from saline digesters in two Hong Kong plants due to flushing toilet with sea water. Cluster B contained two samples from digesters (one from the US and one from the Netherlands) that received feed sludge after pretreatment with thermal hydrolysis. Cluster C contained three samples from the digester treating wastewater partially from the tannery industry in a Hong Kong plant. Cluster D contained 14 samples from thermophilic digesters (>50°C) located in Japan, US, Canada, and the Netherlands. Cluster E contained seven samples from two Japanese plants and one USA plant that operated digesters at temperatures < 30°C for at least three months. Cluster F contained six samples from one non-saline Hong Kong plant (HKSW, digester temperature at 36.0°C) and one USA plant (USNO, digester temperature at 30.3°C), but the operating conditions that determined high community similarity of these two plants are still not clear. Cluster G contained 16 samples from seven Japanese plants, with slightly high operation temperatures between 38 and 42°C, except for plant JPSS at 36.5°C. The largest cluster (H) contained 91 samples of from 16 USA plants, six Japanese plants, two Canadian plants, and four Netherlandish plants, which operated digesters mainly under mesophilic conditions. Samples from plant USLA and NEAV were not assigned to any cluster due to lack of clear association with operating conditions. Within each cluster, samples that originated from the same plant generally clustered together, even though they might be collected from different reactors or on different dates. The clustering of the AD microbiomes into eight clusters was confirmed by ANOSIM, which gave global R-values close to 1, showing that the between-cluster distances were significantly larger than the within-cluster

distances (Table S2B). In contrast, the clustering solely based on the geographical location of the samples generated much smaller global R-values (less than 0.6) (Table S2C). A UPGMA-based clustering on Bray-Curtis distance matrix produced very similar results (Fig. S5C), where only two samples diverged from cluster G and three samples diverged from cluster H compared to the results based on weighted UniFrac.

*3.2. Characterization of feed-derived residue populations*

Our previous study revealed that, in a single wastewater treatment plant, AD microbial communities could contain exogenous populations (*i.e.*, residue populations) that migrated from the feed sludge, resisted to digestion, and not actively involved in anaerobic metabolism (Mei et al. 2016a). In the present study with a much broader sampling scale, we identified such residue populations by comparing the upstream feed sludge and the corresponding AD. To be stringent, we first defined an OTU as being more abundant in feed sludge in a plant only when its feed/AD abundance ratio was over 2, and, conversely, an OTU as being more abundant in AD when the feed/AD abundance ratio was below 0.5. Further, we defined OTUs as residue populations if they were frequently more abundant in feed sludge (minimum five plants) and rarely more abundant in AD (maximum five plants) (Fig. 2). Using these criteria, 1,464 OTUs were identified as residue populations. In agreement, only 172 of them were associated with known obligate anaerobic taxa based on family-level phylogeny (TableS3) obtained from literature (Rosenberg et al. 2014, Vos et al. 2011). In total, 704 residue OTUs were associated with *Proteobacteria* and 298 OTUs with *Bacteroidetes*, accounting for 20.8% and 13.4% of sequences in feed sludge, respectively (Fig. 3A). Abundances of these OTUs in the AD community decreased drastically to 3.6% and 1.6%, respectively. Other phyla including *Firmicutes*, *Planctomycetes*, and *Chloroflexi* also contained residues populations but were

presented by a small number of OTUs (<100) and low relative abundance. Detailed phylogenetic analysis of the top 50 abundant residue OTUs indicated that 21 of them were associated with *Proteobacteria* (excluding *Deltaproteobacteria*) and 18 OTUs with *Bacteroidetes* (excluding *Bacteroidales*) (Fig. S6), which were mostly known as aerobes or facultative anaerobes and were consistent with our previous study (Mei et al. 2016a). On the other hand, known anaerobic populations in AD were not assigned as residue in our analysis, although they were detected in the feed sludge. These populations included for example methanogens (*e.g.*, *Methanobacteriales*, *Methanomicrobiales*, and *Methanosarcinales*), fermenters (*e.g.*, *Anaerolineales*), and syntrophs (*e.g.*, *Syntrophobacterales*). Their abundance increased after entering AD, and no residue OTU was related to these taxa (Fig. S7).

We further observed that the presence of residue populations was a universal phenomenon in all the digesters sampled (Fig. 3B). The lowest relative abundance of residue populations in a sample was 0.02% in USSF1 that received feed sludge after pretreatment, and the majority (117 out of 148 AD samples) were less than 10%. High residue populations were less common, with 26 samples between 10%-20%, and five samples between 20-30%. The highest abundance was observed with JPYS1 (27.3%) that was operated below 20°C. Furthermore, we observed a clear positive correlation ($rho=0.846$, $p<0.01$) between residue populations and alpha diversity (Shannon index) of the AD community (Fig. 3C), indicating the migration of residue populations increased both species richness and evenness of the AD microbial community. In contrast, varying abundance of endogenous populations, such as methanogens or syntrophs, did not correlate with Shannon index of the community (small rho values, Fig. S8). Also a higher residue population abundance was observed to coincide with a

lower digestion efficiency (volatile solids reduction) (Fig. S9), but the correlation was weak as indicated by a low coefficient (rho=-0.361, $p<0.01$).

The presence of residue populations was also influenced by operating conditions (Fig. 3D). The highest residue population abundance was 21.4%, observed in cluster E (low operating temperature), followed by 13.9% in cluster F. Correspondingly, clusters E and F had the highest alpha diversity. The abundance of residue populations in cluster H, which represented most of the digesters studied, was 6.0%. In comparison, clusters B (pretreatment), D (thermophilic), and G (>40°C) contained residue population at relative abundances of 0.8%, 3.3%, and 1.6%, respectively. In addition, residue populations could be more abundant than syntrophs (1.2-7.1%) and methanogens (0.3-2.6%), such as in cluster E and F (Fig. S10). We also tested whether residue populations affected beta-diversity by removing residue OTUs from each community. Based on weighted UniFrac distance, clusters A to G remained intact. Seven samples that were originally in cluster H were separated from the cluster(Fig. S5A and B). Based on Bray-Curtis distance, samples in cluster E were split (Fig. S5C and D).

### *3.3. Identification of cluster-specific core populations*

The heterogeneity revealed by the occupancy distribution of OTUs among all 148 AD samples precluded the ability to define a universal core AD microbiome (Fig. S11A). No OTU was present in 147 or 148 samples. Only 14 OTUs were detected in more than 136 samples, and they only accounted for 4.8% of total sequences. In contrast, within each cluster, OTUs shared by all the samples accounted for a large portion of the total sequences (>50% in each cluster, Fig. S11B), indicating that samples in the same cluster tended to have highly similar microbiomes. Thus, we defined cluster-specific core populations (Fig. 4) by including OTUs that were both prevalent and abundant (top 15 abundant bacterial and top three abundant archaeal OTUs that

were detected in all samples of that cluster). The phylogeny of core OTUs was confirmed by building phylogenetic trees (Fig. S12).

In the core communities, OTUs related to known syntrophs were limited to *Smithella* and *Syntrophomonas* (Fig. S12A), known to syntrophically oxidize propionate. *Smithella* related OTUs were observed in clusters B, E, F, G, and H, whereas *Syntrophomonas* related OTUs were observed in clusters with high salinity (cluster A), industrial influent (cluster C), and high operating temperature (cluster D). For the methanogenic core populations, there was a similar trend that an OTU related to *Candidatus* Methanofastidiosa (hydrogenotrophic methanogen) and an OTU related to *Methanosaeta* (aceticlastic methanogen) were consistently observed in clusters B, E, F, G, and H. The high-temperature cluster D contained two unique core OTUs related to *Methanothermobacter* and *Methanoculleus*. The low-temperature cluster E contained one unique core OTU related to *Methanoregula*. Cluster C with industrial influent contained two core OTUs related to *Methanosarcina*, absent in the core communities of other clusters. Cluster A with high salinity contained an OTU related to *Methanolinea* but at low abundance (<0.05%) compared with other hydrogenotrophic methanogens. The core community of cluster A also contained an OTU related to *Methanosaeta*, but likely a different species from the one shared by other clusters based on phylogenetic analysis (Fig. S12B).

With regards to residue populations, the core communities of cluster B (plants with pretreatment) and cluster G (plants operated at ~ 40°C) did not contain any OTU identified as residue population. For the core communities of other clusters, *Proteobacteria* were the major taxa, and the core residue populations were generally related to *Zoogloea*, *Decholomonas*, *Azospira*, and *Acidovorax* (Fig. S12C). Cluster F contained residue populations mainly related to *Sphingobacteria* in *Bacteroidetes*, likely because the feed sludge of cluster F had highest

306 abundance of *Bacteroidetes* and lowest abundance *Proteobacteria* in comparison to other
307 clusters (Fig. S13).

308 The remaining core populations were classified as fermenters. *Bacteroidetes*, as the most
309 diverse, abundant, and ubiquitous phylum, contained 30 core OTUs, all related to the order
310 *Bacteroidales* (Fig. S12D). All clusters contained multiple *Bacteroidetes*-related core OTUs,
311 except for cluster B (plants operated at thermophilic conditions) with only one *Bacteroidetes*-
312 related core OTU. Other major phyla were *Firmicutes*, *Candidatus* Cloacimonetes (WWE1),
313 *Spirochaetes*, and *Thermotogae*. The majority of fermenters were only assigned to a taxonomic
314 level at order or phylum, as a few known closely isolates were available including *Mesotoga*,
315 *Defluviitoga*, *Anaerobaculum*, *Sedimentibacter*, and *Coprothermobacter*. Last, we observed core
316 populations related to phyla without cultivated representatives, including *Candidatus*
317 Aminicenantes (OP8), *Candidatus* Fermentibacteria (Hyd24-12), *Candidatus* Atribacteria (OP9)
318 and *Candidatus* Marinimicrobia (SAR406).

319

320 **4. Discussion**

321 Determining the core microbiome for an ecosystem is an effective approach to delineate
322 how microbes drive biochemical processes (Consortium 2012, Gilbert et al. 2014, Sunagawa et
323 al. 2015). This study demonstrated heterogeneity in AD microbial communities, and rejected the
324 possibility to define a universal core microbiome for all digesters that differed in operational
325 conditions. This was contradictory to studies using a small number of digesters (Campanaro et al.
326 2016, Lee et al. 2012, Mei et al. 2016a, Riviere et al. 2009), but consistent with the previous
327 report that when a relatively large number of digesters were sampled, different types of
328 communities appeared (De Vrieze et al. 2015). Such heterogeneity in AD microbial communities

329 was linked to diversity in operating conditions, which further led to the discovery of cluster-
330 specific core microbiomes. For example, in digesters operated at high temperature (those in
331 cluster D), core OTUs related to thermophiles, including *Methanoculleus* (Cheng et al. 2008),
332 *Methanothermobacter* (Cheng et al. 2011) *Defluviitoga* (Hania et al. 2012), *Coprothermobacter*
333 (Etchebehere et al. 1998), and *Anaerobaculum* (Rees et al. 1997) were uniquely detected. OTUs
334 related to zinc-tolerant *Sedimentibacter* (Burkhardt et al. 2011) were detected in digesters
335 (cluster C) receiving tannery industry wastewater that had high Zn concentration. OTUs related
336 to sulfur-utilizing *Mesotoga* (Nesbø et al. 2012) were detected in digesters (cluster D) receiving
337 sea water. These sulfur-utilizing microorganisms could compete for hydrogen and suppress
338 hydrogenotrophic methanogens in cluster D. An OTU related to *Methanoregula* that could grow
339 at 10°C was detected in digesters in cluster E operated under 30°C (Yashiro et al. 2011). It could
340 be expected that if more digesters with more diverse operating conditions are included, the
341 heterogeneity and the clustering complexity will keep increasing as niche diversity increases.

342 Although there was no shared population among all the eight clusters, some populations
343 were frequently observed in clusters B, E, F, G, and H. These populations included OTUs related
344 to the novel archaeal clade *Candidatus* Methanofastidiosa that is predicted to perform
345 hydrogenotrophic methanogenesis through methylated thiol reduction (Nobu et al. 2016), and
346 *Smithella* that syntrophically oxidize propionate (Liu et al. 1999). Possibly methylated thiol
347 compounds (*e*.*g*., methanethiol and dimethylsulfide) and propionate are critical intermediates
348 prevalent in most ADs. We also observed abundant and diverse OTUs affiliated with the phyla
349 *Bacteroidetes* and *Candidatus* Cloacimonetes, whose ecological functions in AD are still
350 difficult to discern. For example, isolates of *Bacteroidetes* from anaerobic reactors could be
351 saccharolytic (Su et al. 2014, Sun et al. 2016) or proteolytic (Abe et al. 2012, Chen and Dong

2005), but the vast majority of the members in this phylum remain uncultivated and, thus, their metabolism is unknown (Wu et al. 2011). *Candidatus* Cloacimonetes-related populations have been proposed to perform amino acids fermentation (Pelletier et al. 2008), syntrophic propionate oxidation (Nobu et al. 2015), or extracellular cellulose hydrolysis (Limam et al. 2014). Given that the core OTUs in this phylum were associated with distinct uncultivated phylogenetic clades (*e.g.*, W22, SHA-116, BHB21, and W5), one can only speculate about their metabolisms until more genomics information becomes available or until representatives of these clades are cultured.

Previous studies detected core AD populations related to known aerobic and facultative microorganisms including *Thauera*, *Brachymonas*, and *Rhodobacter* (Nelson et al. 2011, Riviere et al. 2009) that were reported as predominant microorganisms in activated sludge (Zhang et al. 2012). Their appearance as core populations in AD is likely due to incomplete digestion, in contrast to other core populations such as methanogens, syntrophs, and fermenters. It is known that activated sludge processes sometimes contain anaerobic zones supporting the growth of anaerobic microorganisms in (Kämpfer et al. 1996). Based on the change in abundance before and after entering AD, our analysis could effectively distinguish microorganisms in feed sludge as residue populations (*i.e.*, decreasing abundance) from those contributing to digestion (*i.e.*, increase in abundance) in AD. Thus, the residue populations we define here were unlikely to involve in the essential functions in AD, *i.e.*, waste degradation and biogas production. Further investigations are necessary to elucidate the exact survival mechanisms of the residue populations in AD. For example, some of them could survive on accumulated carbon reserve like polyhydroxyalkanoates (Liu et al. 2001) or carry out anaerobic metabolism with different

electron acceptors (*e.g.*, nitrate reduction by *Zoogloea* (Shao et al. 2009) and chlorate reduction by *Dechloromonas* (Achenbach et al. 2001)).

On the basis of our previous study of AD in a single plant (Mei et al. 2016a), we showed here that the presence of residue populations was a common phenomenon among all the sampled digesters. Residue populations could account for at least 6% of total sequences obtained from digesters under normal conditions (*i.e.*, cluster H community) and were more abundant than methanogens and syntrophs. Higher abundance of residue populations (*i.e.*, 21.4%) was observed with cluster E likely due to low operating temperature at <30°C. In addition, pretreatment such as thermal hydrolysis could successfully reduce residue populations in AD (*i.e.*, 0.8% abundance in cluster B community).

We observed that the presence of residue populations only contributed to the increase of alpha diversity of the AD microbiome. By removing residue populations from each community, we observed almost no change on the beta-diversity, and the topology of the clustering remained almost the same based on either weighted UniFrac (only seven samples split from cluster H) or Bray-Curtis distance (only cluster E split). This is likely due to the fact that most residue populations were affiliated with *Proteobacteria* and *Bacteroidetes*, which only represented a small fraction of the vast phylogenetic diversity of AD microbiome. Moreover, the abundances of residue populations were generally less than 10% in most digesters, thus their impacts on the beta diversity calculation were marginal. Only when the abundance of residue populations was high (*i.e.*, in cluster E), a major impact was observed. Finally, a very weak correlation was observed between the abundance of residue populations and overall digestion efficiency, likely because the presence of residue populations could only indicate inefficient cell lysis, the first step of AD process (Amani et al. 2010). The digestion efficiency of full-scale systems is

collectively influenced by other factors including compositions of the feed sludge, mixing condition of the reactor, and monitoring approaches.

## 5. Conclusion

The analyses of microbial communities of 90 full-scale anaerobic digesters around the world lead to the following conclusions:

- The differences of microbial community structures were determined by the operating conditions of digesters, whereas geographical location of the digesters did not have a significant impact.
- Residue populations associated with undigested feed sludge were commonly observed in all the AD samples, with the highest abundance observed in low-temperature digesters and lowest abundance in digesters with pretreatment.
- There was no population shared by all the sampled digesters due to the operation-driven heterogeneity. The cluster-specific core microbiome contained methanogens, syntrophs, fermenters, and residue populations.

## Acknowledgements

We are sincerely grateful to the numerous operation staff in the 51 wastewater treatment plants for their generous help on providing samples and sharing information on operating parameters and efficiencies. This research is part of an ongoing project supported by Joint Genomic Institute.

## References


Abe, K., Ueki, A., Ohtaki, Y., Kaku, N., Watanabe, K. and Ueki, K. (2012) Anaerocella delicata gen. nov., sp. nov., a strictly anaerobic bacterium in the phylum Bacteroidetes isolated from a methanogenic reactor of cattle farms. The Journal of General and Applied Microbiology 58(6), 405-412.

Amani, T., Nosrati, M. and Sreekrishnan, T. (2010) Anaerobic digestion from the viewpoint of microbiological, chemical, and operational aspects-a review. Environmental Reviews 18(NA), 255-278.

Appels, L., Lauwers, J., Degrève, J., Helsen, L., Lievens, B., Willems, K., Van Impe, J. and Dewil, R. (2011) Anaerobic digestion in global bio-energy production: potential and research challenges. Renewable and Sustainable Energy Reviews 15(9), 4295-4301.

Burkhardt, E.-M., Bischoff, S., Akob, D.M., Büchel, G. and Küsel, K. (2011) Heavy metal tolerance of Fe (III)-reducing microbial communities in contaminated creek bank soils. Applied and environmental microbiology 77(9), 3132-3136.

Campanaro, S., Treu, L., Kougias, P.G., Francisci, D., Valle, G. and Angelidaki, I. (2016) Metagenomic analysis and functional characterization of the biogas microbiome using high throughput shotgun sequencing and a novel binning strategy. Biotechnology for biofuels 9(1), 26.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L. and Knight, R. (2010a) PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26(2), 266-267.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K. and Gordon, J.I. (2010b) QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7(5), 335-336.

Chen, S. and Dong, X. (2005) Proteiniphilum acetatigenes gen. nov., sp. nov., from a UASB reactor treating brewery wastewater. International journal of systematic and evolutionary microbiology 55(6), 2257-2261.

Cheng, L., Dai, L., Li, X., Zhang, H. and Lu, Y. (2011) Isolation and characterization of Methanothermobacter crinale sp. nov., a novel hydrogenotrophic methanogen from the Shengli oil field. Applied and environmental microbiology 77(15), 5212-5219.

Cheng, L., Qiu, T.L., Li, X., Wang, W.D., Deng, Y., Yin, X.B. and Zhang, H. (2008) Isolation and characterization of Methanoculleus receptaculi sp. nov. from Shengli oil field, China. FEMS microbiology letters 285(1), 65-71.

Clarke, K.R. (1993) Non-parametric multivariate analyses of changes in community structure. Australian journal of ecology 18(1), 117-143.

Consortium, H.M.P. (2012) Structure, function and diversity of the healthy human microbiome. Nature 486(7402), 207-214.

De Vrieze, J., Saunders, A.M., He, Y., Fang, J., Nielsen, P.H., Verstraete, W. and Boon, N. (2015) Ammonia and temperature determine potential clustering in the anaerobic digestion microbiome. Water Research 75, 312-323.

455 Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26(19),
456 2460-2461.

457 Etchebehere, C., Pavan, M., Zorzopulos, J., Soubes, M. and Muxi, L. (1998) Coprothermobacter platensis
458 sp. nov., a new anaerobic proteolytic thermophilic bacterium isolated from an anaerobic mesophilic
459 sludge. International journal of systematic and evolutionary microbiology 48(4), 1297-1304.

460 Gilbert, J.A., Jansson, J.K. and Knight, R. (2014) The Earth Microbiome project: successes and
461 aspirations. BMC biology 12(1), 1.

462 Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D.,
463 Highlander, S.K. and Sodergren, E. (2011) Chimeric 16S rRNA sequence formation and detection in
464 Sanger and 454-pyrosequenced PCR amplicons. Genome research 21(3), 494-504.

465 Hania, W.B., Godbane, R., Postec, A., Hamdi, M., Ollivier, B. and Fardeau, M.-L. (2012) Defluviitoga
466 tunisiensis gen. nov., sp. nov., a thermophilic bacterium isolated from a mesothermic and anaerobic whey
467 digester. International journal of systematic and evolutionary microbiology 62(6), 1377-1382.

468 Ihaka, R. and Gentleman, R. (1996) R: a language for data analysis and graphics. Journal of
469 computational and graphical statistics 5(3), 299-314.

470 Kämpfer, P., Erhart, R., Beimfohr, C., Böhringer, J., Wagner, M. and Amann, R. (1996) Characterization
471 of bacterial communities from activated sludge: culture-dependent numerical identification versus in situ
472 identification using group-and genus-specific rRNA-targeted oligonucleotide probes. Microbial ecology
473 32(2), 101-121.

474 Kelessidis, A. and Stasinakis, A.S. (2012) Comparative study of the methods used for treatment and final
475 disposal of sewage sludge in European countries. Waste Management 32(6), 1186-1195.

476 Koenig, A. and Bari, Q. (2001) Vane shear strength of dewatered sludge from Hong Kong. Water science
477 and technology 44(2-3), 389-397.

478 Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K. and Schloss, P.D. (2013) Development of a
479 dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq
480 Illumina sequencing platform. Applied and environmental microbiology 79(17), 5112-5120.

481 Lee, S.-H., Kang, H.-J., Lee, Y.H., Lee, T.J., Han, K., Choi, Y. and Park, H.-D. (2012) Monitoring
482 bacterial community structure and variability in time scale in full-scale anaerobic digesters. Journal of
483 Environmental Monitoring 14(7), 1893-1905.

484 Limam, R.D., Chouari, R., Mazéas, L., Wu, T.D., Li, T., Grossin-Debattista, J., Guerquin-Kern, J.L.,
485 Saidi, M., Landoulsi, A. and Sghir, A. (2014) Members of the uncultured bacterial candidate division
486 WWE1 are implicated in anaerobic digestion of cellulose. MicrobiologyOpen 3(2), 157-167.

487 Liu, W.T., Nielsen, A.T., Wu, J.H., Tsai, C.S., Matsuo, Y. and Molin, S. (2001) In situ identification of
488 polyphosphate-and polyhydroxyalkanoate-accumulating traits for microbial populations in a biological
489 phosphorus removal process. Environmental Microbiology 3(2), 110-122.

490 Liu, Y., Balkwill, D.L., Aldrich, H.C., Drake, G.R. and Boone, D.R. (1999) Characterization of the
491 anaerobic propionate-degrading syntrophs Smithella propionica gen. nov., sp. nov. and Syntrophobacter
492 wolinii. International journal of systematic bacteriology 49(2), 545-556.

493 Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Buchner, A., Lai, T., Steppi, S., Jobb, G.
494 and Förster, W. (2004) ARB: a software environment for sequence data. Nucleic acids research 32(4),
495 1363-1371.

496 Mei, R., Narihiro, T., Nobu, M.K., Kuroda, K. and Liu, W.-T. (2016a) Evaluating digestion efficiency in
497 full-scale anaerobic digesters by identifying active microbial populations through the lens of microbial
498 activity. Scientific reports 6, 34090.

499 Mei, R., Narihiro, T., Nobu, M.K. and Liu, W.-T. (2016b) Effects of heat shocks on microbial community
500 structure and microbial activity of a methanogenic enrichment degrading benzoate. Letters in Applied
501 Microbiology 63(5), 356-362.

502 Narihiro, T. (2016) Microbes in the Water Infrastructure: Underpinning Our Society. Microbes and
503 Environments 31(2), 89-92.

504 Narihiro, T., Nobu, M.K., Kim, N.K., Kamagata, Y. and Liu, W.T. (2015) The nexus of syntrophy-
505 associated microbiota in anaerobic digestion revealed by long-term enrichment and community survey.
506 Environmental Microbiology 17(5), 1707-1720.

507 Nelson, M.C., Morrison, M. and Yu, Z. (2011) A meta-analysis of the microbial diversity observed in
508 anaerobic digesters. Bioresource Technology 102(4), 3730-3739.

509 Nesbø, C.L., Bradnan, D.M., Adebusuyi, A., Dlutek, M., Petrus, A.K., Foght, J., Doolittle, W.F. and Noll,
510 K.M. (2012) Mesotoga prima gen. nov., sp. nov., the first described mesophilic species of the
511 Thermotogales. Extremophiles 16(3), 387-393.

512 Nobu, M.K., Narihiro, T., Kuroda, K., Mei, R. and Liu, W.-T. (2016) Chasing the elusive Euryarchaeota
513 class WSA2: genomes reveal a uniquely fastidious methyl-reducing methanogen. ISME J 10(10), 2478-
514 2487.

515 Nobu, M.K., Narihiro, T., Rinke, C., Kamagata, Y., Tringe, S.G., Woyke, T. and Liu, W.T. (2015)
516 Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor.
517 ISME J 9(8), 1710-1722.

518 Pelletier, E., Kreimeyer, A., Bocs, S., Rouy, Z., Gyapay, G., Chouari, R., Rivière, D., Ganesan, A.,
519 Daegelen, P. and Sghir, A. (2008) "*Candidatus* Cloacamonas acidaminovorans": genome sequence
520 reconstruction provides a first glimpse of a new bacterial division. Journal of Bacteriology 190(7), 2572-
521 2579.

522 Rees, G.N., Patel, B.K., Grassia, G.S. and Sheehy, A.J. (1997) Anaerobaculum thermoterrenum gen. nov.,
523 sp. nov., a novel, thermophilic bacterium which ferments citrate. International journal of systematic and
524 evolutionary microbiology 47(1), 150-154.

525 Regulations, E. (2003) Technology: Control of Pathogens and Vector Attraction in Sewage Sludge.
526 USEPA, Office of Research and Development.

Riviere, D., Desvignes, V., Pelletier, E., Chaussonnerie, S., Guermazi, S., Weissenbach, J., Li, T., Camacho, P. and Sghir, A. (2009) Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. The ISME Journal 3(6), 700-714.

Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E. and Thompson, F. (2014) The Prokaryotes: Other Major Lineages of Bacteria and The Archaea, Springer Berlin Heidelberg.

Schink, B. and Stams, A.J. (2006) Syntrophism among prokaryotes, Springer.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H. and Robinson, C.J. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and environmental microbiology 75(23), 7537-7541.

Smith, A.L., Shimada, T. and Raskin, L. (2017) A comparative evaluation of community structure in full-scale digesters indicates that two-phase digesters exhibit greater microbial diversity than single-phase digesters. Environmental Science: Water Research & Technology.

Stamps, B.W., Lyles, C.N., Suflita, J.M., Masoner, J.R., Cozzarelli, I.M., Kolpin, D.W. and Stevenson, B.S. (2016) Municipal solid waste landfills harbor distinct microbiomes. Frontiers in microbiology 7.

Su, X.-L., Tian, Q., Zhang, J., Yuan, X.-Z., Shi, X.-S., Guo, R.-B. and Qiu, Y.-L. (2014) Acetobacteroides hydrogenigenes gen. nov., sp. nov., an anaerobic hydrogen-producing bacterium in the family Rikenellaceae isolated from a reed swamp. International journal of systematic and evolutionary microbiology 64(9), 2986-2991.

Sun, L., Toyonaga, M., Ohashi, A., Tourlousse, D.M., Matsuura, N., Meng, X.-Y., Tamaki, H., Hanada, S., Cruz, R. and Yamaguchi, T. (2016) Lentimicrobium saccharophilum gen. nov., sp. nov., a strictly anaerobic bacterium representing a new family in the phylum Bacteroidetes, and proposal of Lentimicrobiaceae fam. nov. International journal of systematic and evolutionary microbiology.

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R. and Alberti, A. (2015) Structure and function of the global ocean microbiome. Science 348(6237), 1261359.

Treu, L., Kougias, P.G., Campanaro, S., Bassani, I. and Angelidaki, I. (2016) Deeper insight into the structure of the anaerobic digestion microbial community; the biogas microbiome database is expanded with 157 new genomes. Bioresource Technology 216, 260-266.

Vanwonterghem, I., Jensen, P.D., Ho, D.P., Batstone, D.J. and Tyson, G.W. (2014) Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. Current opinion in biotechnology 27, 55-64.

Vos, P., Garrity, G., Jones, D., Krieg, N.R., Ludwig, W., Rainey, F.A., Schleifer, K.-H. and Whitman, W. (2011) Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes, Springer Science & Business Media.

Wong, J., Li, K., Fang, M. and Su, D. (2001) Toxicity evaluation of sewage sludges in Hong Kong. Environment International 27(5), 373-380.

564 Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S.A., Bewtra, M., Knights, D.,
565 Walters, W.A. and Knight, R. (2011) Linking long-term dietary patterns with gut microbial enterotypes.
566 Science 334(6052), 105-108.

567 Yashiro, Y., Sakai, S., Ehara, M., Miyazaki, M., Yamaguchi, T. and Imachi, H. (2011) Methanoregula
568 formicica sp. nov., a methane-producing archaeon isolated from methanogenic sludge. International
569 journal of systematic and evolutionary microbiology 61(1), 53-59.

570 Zhang, T., Shao, M.-F. and Ye, L. (2012) 454 Pyrosequencing reveals bacterial diversity of activated
571 sludge from 14 sewage treatment plants. The ISME Journal 6(6), 1137-1147.
572

**Figure 1**

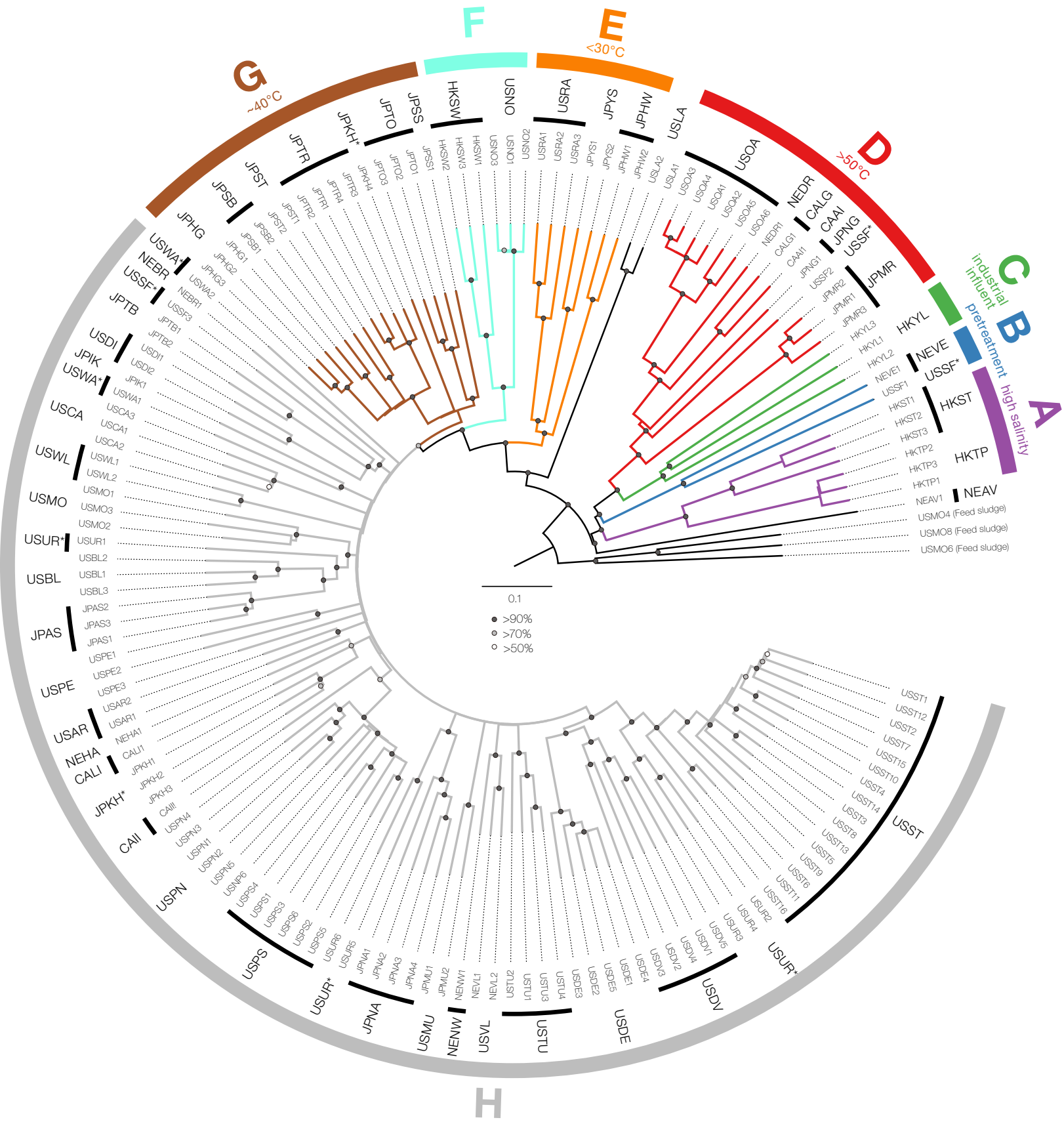


Figure 1. Clustering of digester microbial communities. UPGMA dendrogram was built using weighted UniFrac as distance matrix after jackknifed rarefaction to 20,957 sequences per sample with 100 iterations. Three feed sludge samples are used as outgroup to root the tree. Plants that have samples not clustered together are marked.

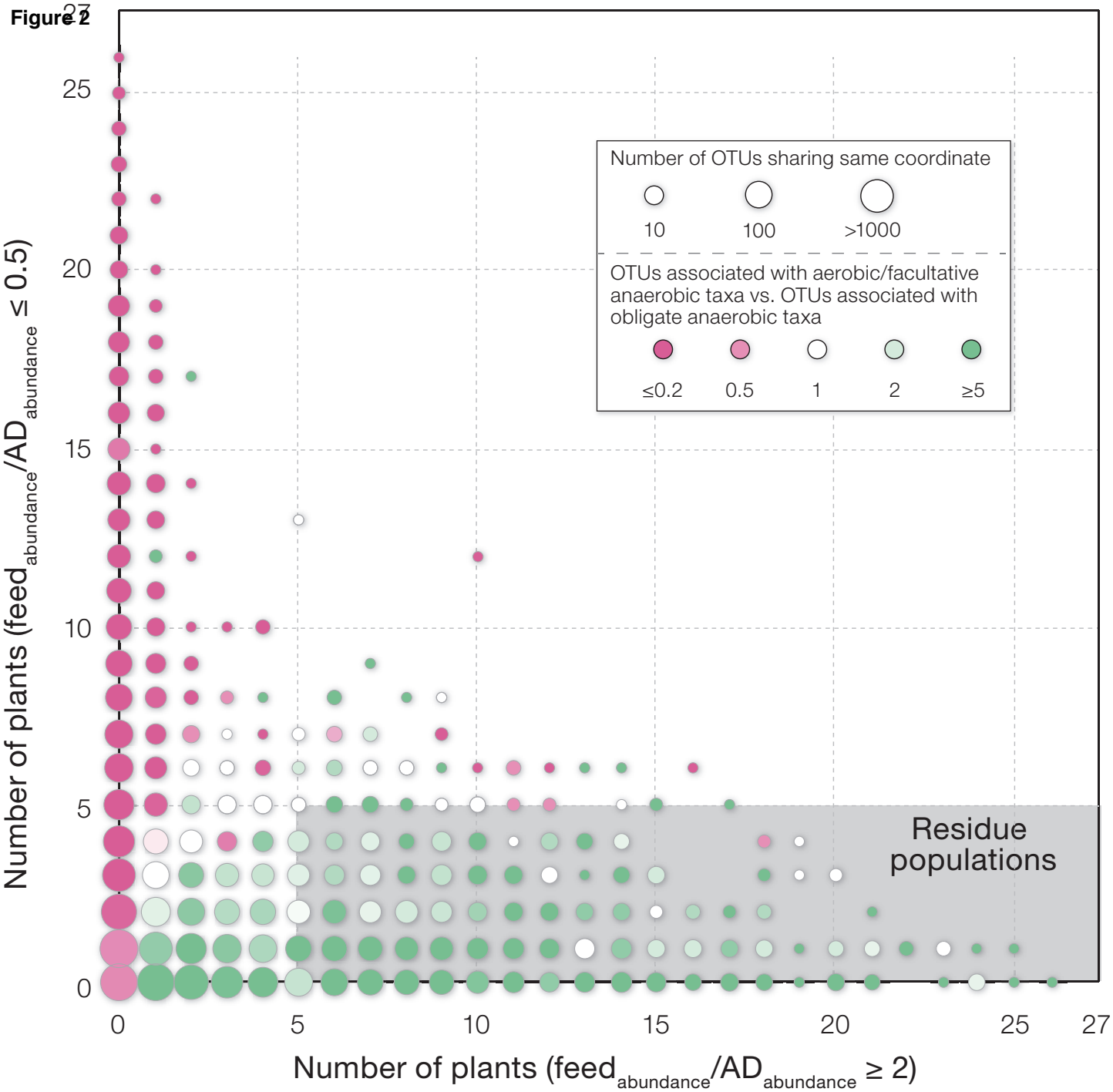


**Figure 2. Identification of OTUs related to residue populations.** For each OTU, the x value represents the number of plants where the OTU has more than double abundance in feed sludge than in AD. The y value represents the number of plants where the OTU has more than double abundance in AD than in feed sludge. The size of each bubble represents the number of OTUs (in log scale) sharing the same x-y coordinate. The color scale represents at a given coordinate the ratio of the number of OTUs associated with aerobic/facultative anaerobic taxa over the number of OTUs associated with obligate anaerobic taxa. Shadowed region ($x \geq 5$, $y \leq 5$) represents OTUs defined as residue populations in this study.

**Figure 3**

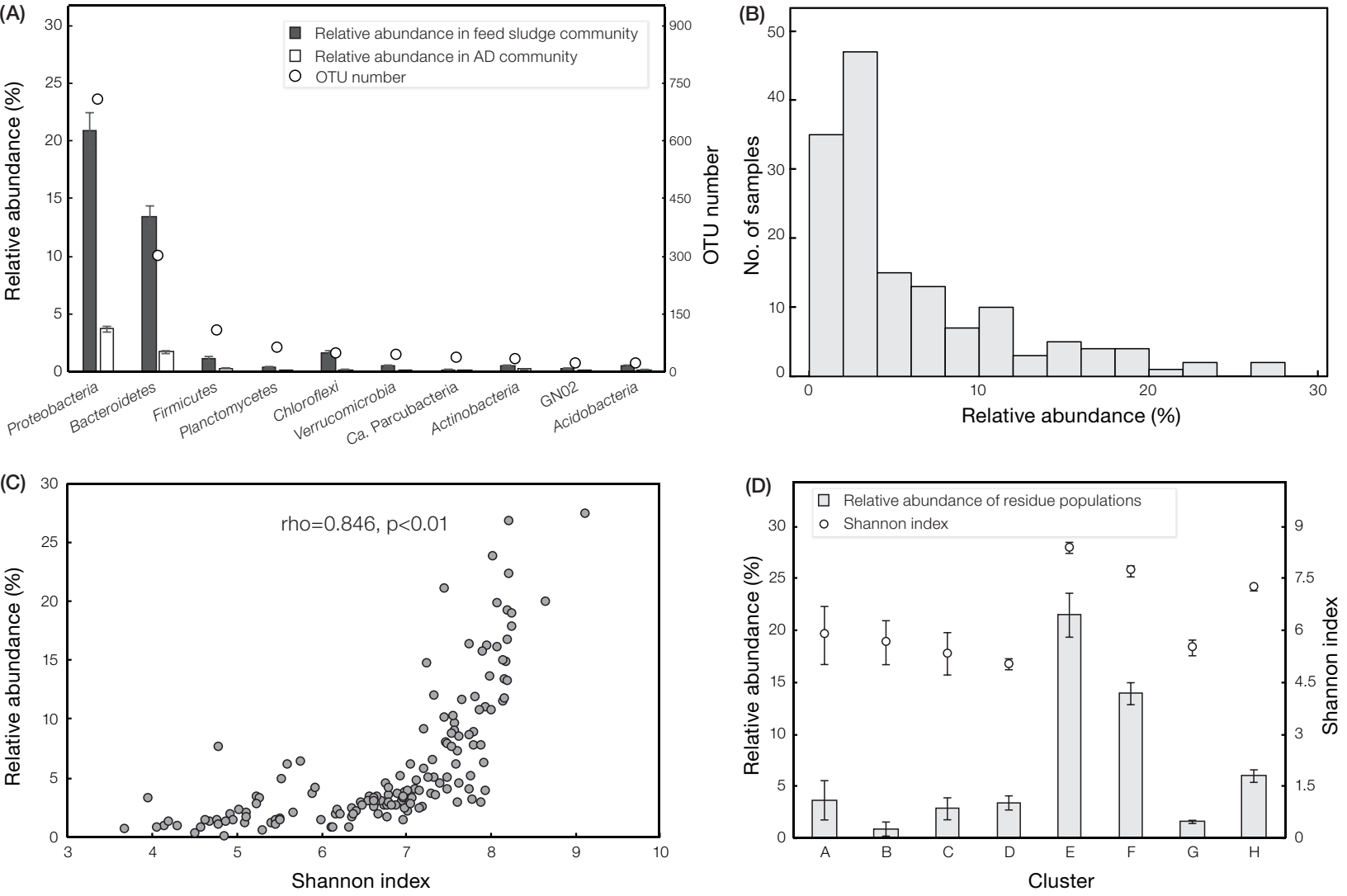


Figure 3. Distribution of OTUs identified as residue populations. Panel (A) shows the top ten phyla that contain high numbers of residue OTUs. Dots represent numbers of OTUs of this phylum (primary y axis). Solid bars represent abundance of residue populations of this phylum in feed sludge community and open bars represent abundance of residue populations of this phylum in AD community (secondary y axis). Panel (B) shows the distribution of residue abundance in 148 AD samples. Panel (C) shows correlation between Shannon index and residue abundance of each AD sample. Panel (D) shows residue abundance (dots, primary y axis) and Shannon index (bar, secondary y axis) of each cluster determined previously.

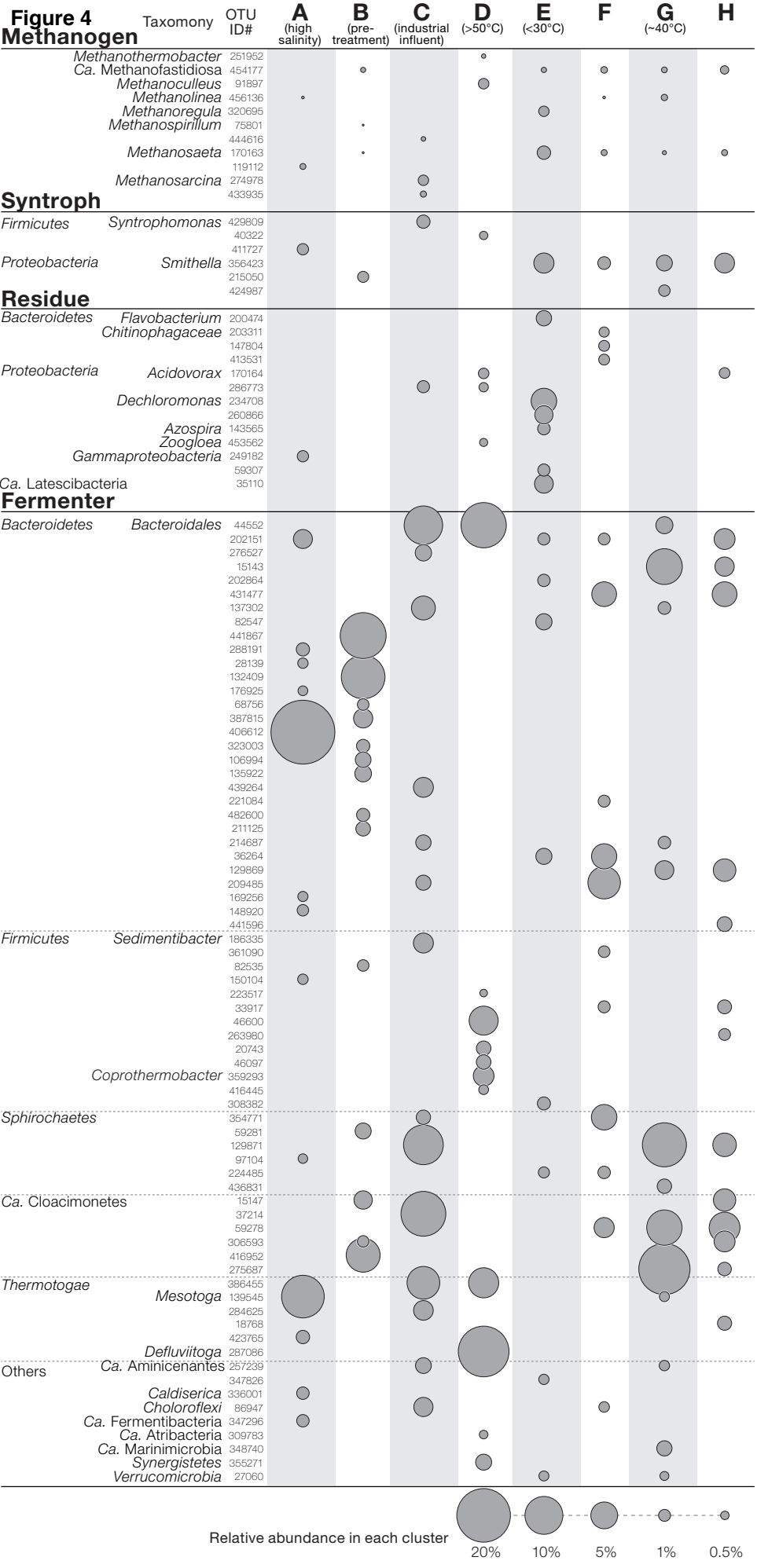


Figure 4. Core microbial community of each cluster.

**Electronic Supplementary Material (for online publication only)**

**Click here to download Electronic Supplementary Material (for online publication only): Supplementary.pdf**