Metagenomic binning reveals the functional roles of core abundant microorganisms in twelve full-scale biogas plants

Campanaro, Stefano; Treu, Laura; Kougias, Panagiotis ; Luo, Gang; Angelidaki, Irini

Published in:
Water Research

Link to article, DOI:
[10.1016/j.watres.2018.04.043](https://doi.org/10.1016/j.watres.2018.04.043)

Publication date:
2018

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Campanaro, S., Treu, L., Kougias, P., Luo, G., & Angelidaki, I. (2018). Metagenomic binning reveals the functional roles of core abundant microorganisms in twelve full-scale biogas plants. *Water Research*, 140, 123-134. <https://doi.org/10.1016/j.watres.2018.04.043>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Accepted Manuscript

Metagenomic binning reveals the functional roles of core abundant microorganisms in twelve full-scale biogas plants

Stefano Campanaro, Laura Treu, Panagiotis G. Kougias, Gang Luo, Irini Angelidaki

PII: S0043-1354(18)30337-3

DOI: [10.1016/j.watres.2018.04.043](https://doi.org/10.1016/j.watres.2018.04.043)

Reference: WR 13740

To appear in: *Water Research*

Received Date: 13 January 2018

Revised Date: 22 March 2018

Accepted Date: 16 April 2018

Please cite this article as: Campanaro, S., Treu, L., Kougias, P.G., Luo, G., Angelidaki, I., Metagenomic binning reveals the functional roles of core abundant microorganisms in twelve full-scale biogas plants, *Water Research* (2018), doi: 10.1016/j.watres.2018.04.043.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Abstract Graphic

**Metagenomic binning reveals the functional roles of core abundant
microorganisms in twelve full-scale biogas plants**

Stefano Campanaro^{b#}, Laura Treu^{c#}, Panagiotis G Kougias^{c*}, Gang Luo^{a,d*} and Irini
Angelidaki^c

^a Shanghai Key Laboratory of Atmospheric Particle Pollution and Prevention (LAP3),
Department of Environmental Science and Engineering, Fudan University, 200433,
Shanghai, China

^b Department of Biology, University of Padua, Via U. Bassi 58/b, 35121 Padova, Italy

^c Department of Environmental Engineering, Technical University of Denmark, Kgs.
Lyngby DK-2800, Denmark

^d Shanghai Institute of Pollution Control and Ecological Security, Shanghai 200092,
P.R. China

Corresponding Authors

*E-mail: panak@env.dtu.dk; tel.: +45 45251429; fax: +45 45932850 (P.G.K.).

*E-mail: gangl@fudan.edu.cn; tel./fax: +86 021-65642297 (G.L.).

These two authors contributed equally to the work.

Abstract: The aim of this work was to elucidate the microbial ecology in twelve mesophilic and thermophilic full-scale biogas plants using a genome-centric metagenomic approach. In this study both biogas plants treating manure and those treating sludge from waste water treatment plants were considered. The identification of 132 Metagenome-Assembled Genomes (MAGs) and analysis of their abundance profile in different samples allowed the identification of the most abundant core members of the anaerobic digestion microbiome. Canonical correspondence analysis was used to determine the influence of biotic and environmental factors on MAGs abundance and to investigate the methanogenic performance of the biogas plants. Prediction of the functional properties of MAGs was obtained analyzing their KEGG pathways and their carbohydrate active domains. Network analysis allowed investigation of species-species associations and shed light on syntrophic interactions between members belonging to the anaerobic digestion dark matter (phylum *Fermentibacteria*). By stratifying and comparing different levels of information, it was predicted that some MAGs have a crucial role in the manure-supplemented thermophilic biogas plants and it was highlighted the importance of the glycine cleavage system in complementing the “truncated” Wood-Ljungdahl pathway.

Key words: Anaerobic digestion; Metagenomic binning; Abundant Metagenome-Assembled Genomes; Functional analysis

1. Introduction

Biogas production from organic wastes via anaerobic digestion (AD) has been used worldwide, allowing renewable energy recovery from organic wastes. In some countries biogas production has been proposed as a basis of circular economy, allowing for bio-waste valorization in the form of biogas; it is also relevant in nutrient recycling, in the decrease of the green-house gas emissions and in the reduction of nitrogen leakage to the soil and groundwater. AD includes four sequential metabolic steps namely hydrolysis, acidogenesis, acetogenesis, and methanogenesis (Sundberg et al., 2013). A number of different microorganisms are involved in the AD process, and different species have distinct roles in a collective organization, which forms an extremely specialized and complex microbiome. Therefore, microbial communities are an important resource for improved anaerobic conversion of organic wastes. In order to optimally exploit this resource it is required to understand not only which microorganisms are present, but also their metabolic potentials (Luo et al., 2016). It was previously reported that more than 99% of prokaryotes in the environment cannot be cultured in the laboratory (Tringe and Rubin, 2005), which limits our understanding of microbial physiology, genetics and community ecology. Culture-independent molecular methods based on 16S rRNA genes and on shotgun sequencing of total DNA (metagenomic sequencing) have been developed to characterize the microbial communities in biogas reactors and their metabolic potential (Luo et al., 2016; Treu et al., 2016a). Indeed, our understanding of these complex systems has been greatly improved by investigating the microbial

community compositions and their correlations with the environmental variables. Previous studies characterized the microbial community using short unassembled reads (Bassani et al., 2015; Luo et al., 2016; Schlüter et al., 2008) or metagenomic assembly (Bremges et al., 2015a; Wirth et al., 2012). However, the identification of species in the biogas community by sequencing of both 16S rRNA genes and total DNA were mainly performed with similarity search for sequences deposited in public databases (Campanaro et al., 2016). Since these reference sequences belong to species frequently isolated from natural environments unrelated to engineered AD process, they might have different functional properties even if they belong to phylogenetically related groups. Therefore, the taxonomic assignment and the functional characterization of dominant microorganisms in biogas reactors are still challenging. The genomes of individual species reconstructed from mixed cultures are named Metagenome Assembled Genomes (MAGs) and can be successfully reconstructed by binning the scaffolds obtained from metagenomic assembly. This genome-centric approach provides a powerful method to understand the phylogenetic and metabolic diversity in anaerobic digestion without relying on culture-dependent techniques. In order to reconstruct MAGs, different binning strategies were proposed (Imelfort et al., 2014; Kang et al., 2015), including one specifically applied to characterize the microbial communities in lab-scale biogas reactors fed with cattle manure (Campanaro et al., 2016). Recent studies recovered MAGs present in the AD system, focusing the attention on laboratory-scale reactors fed with specific substrates (e.g. cellulose as the sole carbon and energy source) (Vanwonterghem et al., 2016), or

supplemented with lipids or H₂ (Campanaro et al., 2016; Treu et al., 2016b). Other studies were focused on biogas plants, but using 16S rRNA gene-based methods or investigating only a limited number of samples with high sequencing depth (Stolze et al., 2016)(Rivière et al., 2009; Werner et al., 2011). However, detailed information about the core abundant microorganisms and their functional properties in full-scale biogas reactors is still missing. In Denmark, there are more than 40 centralized biogas plants co-digesting manure and industrial wastes as feedstock. Moreover, most of the wastewater treatment plants (WWTPs) have full-scale biogas reactors treating the primary and secondary sludge. Therefore, the aim of this study was to elucidate the microbiome of full-scale biogas reactors operating under various operational conditions, to correlate the environmental parameters applied on the reactors with microbial compositions and to propose functional roles for poorly-characterized species and for the most abundant MAGs.

2. Materials and methods

2.1 Sampling and DNA extraction

The collected samples were obtained from Danish full-scale biogas reactors. Only biogas plants that had been running for more than two years under similar operational conditions were considered. The detailed information about operational conditions and performance of each biogas reactor is shown in Table S1. All biogas reactors had sampling points in the effluent lines close to the reactors ensuring good representative samples of the reactor biomass. The sampling valve was opened for 5 min before sample acquisition to flush the sampling valve and tube (Luo et al., 2016). “MT” and

“MM” represent the biogas reactors fed with manure and operated respectively under thermophilic and mesophilic conditions, while “SM” represents the mesophilic biogas plants fed with sludge from WWTPs. Samples MT2a and MT2b were collected from Blåhøj, and samples MT3a and MT3b were obtained from Lemvig biogas plants. Both Blåhøj and Lemvig have two biogas reactors connected in series, and thus samples were collected from both steps of the serial process. All the biogas plants in this study were operating at stable conditions at the time of sampling, which ensured the representativeness of these microbial communities. The samples for DNA extractions were collected in sterile tubes (15 mL) and frozen immediately in a cooler with dry ice. QIAamp DNA Stool Mini Kit (QIAGEN) was used to extract total genomic DNA from the pellet obtained after centrifugation (10 min at 16000xg). NanoDrop 1000 (ThermoFisher) was used to measure the DNA concentration and purity.

2.2 Metagenomic sequencing and binning analysis

The extracted DNA was sent to Beijing Genomics Institute for library preparation (101 bp paired-end strategy) and sequenced using Illumina Hiseq 2000 platform obtaining between 9.6 and 30 million paired-end sequences depending on sample. All the metagenomic datasets were uploaded to MG-RAST (Rapid Annotation using Subsystems Technology for Metagenomes) with the project ID 6474.

Sequence reads with low quality or ambiguous bases were filtered with Trimmomatic (v0.33) (Bolger et al., 2014), and the pair-end sequence reads were merged with Flash (Salzberg, 2011). These high quality reads were imported to CLC Genomics workbench v. 5.1 (CLC Bio, QIAGEN, USA) and assembled using de novo assembly

algorithm (kmer 63, bubble size 60 and minimum scaffold length 500 bp). Protein-encoding genes were predicted using Prodigal (v2.6.2) in metagenome mode (Hyatt et al., 2012). Scaffolds coverage was determined by aligning to the assembly the reads of each sample with bowtie2 (v2.2.4) (Langmead and Salzberg, 2012) and calculating values with the BEDTools package (v2.17.0) (Quinlan and Hall, 2010). Metagenomic binning was performed using a strategy based on hierarchical clustering followed by canopy profile selection (Campanaro et al., 2016). Completeness, contamination and taxonomic assignment of the MAGs were determined using CheckM (Parks et al., 2015). The 16S rRNA genes were identified for each MAG with in-house developed perl script using Hidden Markov Models obtained from RNAmmer (Lagesen et al., 2007) and taxonomy was determined using RDP classifier trained on GreenGenes 13_5 (Wang et al., 2007) and on SILVA 132 (Quast et al., 2013). These databases were selected considering previous comparative taxonomic investigations performed on the AD microbiome (Campanaro et al., 2018). MAGs taxonomic assignment was further validated using PhyloPhlAn (Segata et al., 2013). Average Nucleotide Identity (ANI) values were calculated as previously described (Campanaro et al., 2017) considering genomes deposited in the NCBI microbial database representative for all the microbial species. ANI was also calculated with the MAGs identified in previous binning projects (Campanaro et al., 2016; Jing et al., 2017; Mosbæk et al., 2016; Ruiz-Sánchez et al., 2017; Stolze et al., 2016; Treu et al., 2016b; Vanwonterghem et al., 2016). Two genomes were considered as belonging to the same species if ANI was higher than 95% and they have at least 70% of genes in

common (Varghese et al., 2015). An additional investigation was performed on four MAGs (DTU308, DTU309, DTU310, DTU320) by selecting 5278 representative genomes from NCBI microbial genomes database (details regarding selection is provided in the Supporting Information file), building a tree using PhloPhlAn (Segata et al., 2013) (126 h total CPU computation time) and performing a manual inspection supported by Dendroscope (Huson and Scornavacca, 2012). The taxonomic assignments obtained with different methods were manually compared and a name representative of the taxonomic assignment for each MAG was proposed. Abundance of the MAGs in different samples was calculated considering the coverage values and length of their scaffolds as previously described (Campanaro et al., 2017). The coverage of the MAGs in different samples was clustered based on Pearson using MeV software in order to identify behavior of MAGs in different samples (Saeed et al., 2003). This result was subsequently converted in a network representation by computing the similarity matrix using the Pearson Correlation Coefficient with the “Expression Correlation network” Plugin tool and visualizing the result with Cytoscape (v2.8.3) (Shannon et al., 2003). Protein-encoding genes for each MAG were predicted using Prodigal (v2.6.2) in normal mode and annotated using MAPLE to verify completeness of the biochemical pathways (Takami et al., 2016). Protein encoding genes were also analyzed with FOAM software (Prestat et al., 2014); subsequently the number of genes associated with each pathway and module were calculated using self-written perl scripts available in SourceForge (<https://sourceforge.net/projects/perl-scripts-kegg/>). MAGs were analyzed using the

RAST server (Aziz et al., 2008) and Carbohydrate Active Enzymes were identified using dbCAN (v4.0) (Yin et al., 2012).

The differentiation among biogas plants was determined by calculating β -diversity with ExpressBetaDiversity (EBD) software (v1.0.7) (Parks and Beiko, 2013). Canonical correspondence analysis (CCA) was performed using the R functions implemented in VEGAN v2.4-4 (Dixon, 2009), while correspondence analysis (CA) based on Pearson calculation and Non-metric Multidimensional Scaling (NMDS) were performed with R following the procedure described by Torondel and colleagues (Torondel et al., 2016).

The MAGs genome sequences are freely available at the biogasmicrobiome web site (<https://biogasmicrobiome.env.dtu.dk/>).

3. Results and discussion

3.1 Overview of the reconstructed microbiome

Assembly generated 666,024 scaffolds with length > 500 bp (N50 1,776) that were binned leading to the extraction of 132 MAGs. Despite that the number of reads obtained in the present study was lower than previous projects (Campanaro et al., 2016; Vanwonterghem et al., 2016), the number of MAGs recovered was higher. This result was probably due to the higher number of samples investigated (biogas plants) and also to their heterogeneous microbial composition. This diversity in the microbiome can be due to the different parameters characterizing the biogas plants, particularly in terms of temperature (“mesophilic” and “thermophilic”) and feedstock composition (“manure” or “sludge”) (Table S1). Quality evaluation results indicated

that 103 of MAGs had estimated completeness higher than 70%, while 43 of them had completeness higher than 90% (Data set S1 WS1).

The degree of novelty of the genomes was assessed by calculating the Average Nucleotide Similarity level (ANI) with the MAGs extracted from previous genome-centric investigations (Data set S1 WS1) and with genomes deposited at the NCBI microbial genome resource. Fifty-seven MAGs were previously found in AD systems and eleven of these belonged to species with known genomes deposited at NCBI. In particular, eight out of the eleven genomes derived from previous binning approaches, while three genomes (*Methanosaeta concilii*, *Methanosarcina flavescentis* and *Defluviitoga tunisiensis*) were sequenced from axenic cultures (Data set S1 WS1). Additional taxonomic investigations performed using PhyloPhlAN and the 16S rRNA gene sequences extracted from the MAGs, were not able to increase the assignment to known microbial species. The high degree of novelty of these MAGs indicates that species belonging to unexplored taxa are frequently present in AD systems.

As shown in Data set S1 WS1 and in Figure 1, the vast majority of MAGs belong to the three phyla more often identified in biogas plants (Bremges et al., 2015b; Fontana et al., 2016), *Firmicutes* (63 MAGs), *Bacteroidetes* (20 MAGs) and *Proteobacteria* (11 MAGs). By checking the taxa found at low frequency in the present study, interesting insights on poorly investigated phyla were provided. In fact two MAGs were assigned to *Verrucomicrobia*, one to *Spirochaetes* and one to *Acidobacteria*. It should be noted that four MAGs were not assigned to phylum level using taxonomic informative proteins, further indicating the presence of MAGs belonging to unknown

taxa in biogas reactors. To gain more insights and additional investigation was performed running PhloPhlAn on 5278 genomes selected from NCBI database and assigning the four MAGs to Bacteria candidate phyla (Figure S1, Data set S2). More specifically, Unclassified bacterium DTU309 was tentatively assigned to Candidate Shapirobacteria division (Bacteria candidate phyla, *Patescibacteria* group) and Unclassified bacterium DTU320 to the candidate phylum Hyd24-12 (*Fermentibacteria*) (both results were supported by 16S rRNA analysis). Both Unclassified bacterium DTU310 and Unclassified bacterium DTU308 were assigned to Parcubacteria group (Bacteria candidate phyla, *Patescibacteria* group) respectively as *Candidatus Moranbacteria* and *Candidatus Falkowbacteria*. As expected, microbial composition and abundance of different phyla were strongly dependent on the operational conditions (i.e. temperature and influent feedstock); for example all the seven MAGs assigned to *Chloroflexi* were identified only in the sludge reactors (Figure 1). Also the phyla identified at low frequency (*Verrucomicrobia*, *Spirochaetes* and Hyd24-12) were mainly restricted to the sludge reactors, suggesting that in this environment species belonging to poorly characterized phyla could be more frequent or, alternatively, they could be identified more easily than in the manure-supplemented reactors where they are less abundant.

3.2 Microbial diversity in biogas plants

The cluster analysis reported in Figure 1 evidenced that MAGs detected in manure-supplemented samples were generally not present or in low abundant in sludge-supplemented samples. As a consequence, hierarchical clustering reported on

top of Figure 1 divided the samples into three well-defined groups, those derived from Manure-supplemented Mesophilic biogas plants (MM1, MM2 and MM3), those from Manure-supplemented Thermophilic biogas plants (MT1, MT2a, MT2b, MT3a, MT3b and MT4), and those from Sludge-supplemented Mesophilic biogas plants (SM1, SM2, SM3, SM4, and SM5). NMDS further supported the classification of the samples obtained from Pearson correlation (Figure 2A). This analysis was confirmed with a calculation of β -diversity (Figure S2) and evidenced that the microbial composition similarity among MT samples is higher compared with MM and SM samples, which are more diverse. These results indicated that both temperature and feedstock play an important role in shaping the microbial community in biogas reactors, confirming the findings of previous studies (Amani et al., 2010; Sundberg et al., 2013). Additionally, the higher similarity among the MT samples suggests that high temperature poses a strong selection to the microbial community.

By considering abundance of each MAG in the samples, it was possible to classify them according to the specific reactor type (i.e. MT, MM or SM) (Data set S1). The classification was performed according to the following rule: if the average coverage of one MAG in a specific reactor type (e.g. MT) was higher than 90% of the average coverage calculated across all the reactors, the MAG was assigned to that specific group (i.e. "MT"). According to this rule, 28 MAGs were classified as MT, 21 as MM and 49 as SM. The remaining microbes did not show a clear predominance in a specific type of reactor. This finding revealed that highly different microbiomes can support methanation process in biogas plants with distinct results in terms of

biogas-production rate and methane content (Table S1). Moreover, MAGs consistently abundant in each group were identified (Figure 1). Fir056_BP, Fir040_BP, Fir057_BP and Bac024_BP (three assigned to *Clostridia* and one to *Rikenellaceae*) were shared by all the MT samples. Fir056_BP (*Clostridiales* sp. DTU010) and Fir051_BP (*Clostridia* sp. DTU204) were abundant in some of MM and MT samples, which indicated that some microorganisms can well adapt to both mesophilic and thermophilic anaerobic digestion processes. However, there was only one abundant MAG shared by all the MM (Fir051_BP, *Clostridia* sp. DTU204) or SM (Bac008_BP, *Bacteroidales* sp. DTU326) samples. Previous studies demonstrated that the diversity in microbial community composition in mesophilic digestion was higher than thermophilic digestion (Bassani et al., 2015), and therefore the functional redundancy in mesophilic conditions is higher compared to the thermophilic ones.

3.3 Network analysis and species-species associations

The co-occurrence of MAGs in different samples can reveal possible syntrophic relations (obligate mutualistic metabolism) or the presence of other facultative mutualistic associations. These potential associations were investigated by converting the species abundance in different reactors into a network where nodes represent MAGs and edge thickness represents similarity between abundance vectors (Figure 3). MAGs having similar abundance profiles are represented by nodes from which many edges are departing (defined as “hubs”). Representing MAGs abundance with circle sizes (Figure 3) provides a clear picture of the part of the AD microbiome that was captured. The network is dominated by two “clusters” mainly representing the

thermophilic MAGs and the sludge MAGs (Figure 3). Interestingly, the cluster depicted in the right part of the figure is composed of four dominant MAGs (Fir056_BP, Fir040_BP, Bac024_BP, Eur033_BP), while the cluster to the left represents a microbial community with high evenness. As mentioned previously, Fir056_BP (*Clostridiales* sp. DTU010) covers a crucial role in the manure-supplemented biogas plants since the correspondence analysis indicated that it is strictly correlated with VFA.

Thirteen nodes had more than twenty edges and were considered as “hubs” in the network. These “hubs” are represented by two different classes of MAGs: those identified in both MT and MM samples (Fir044_BP, Fir065_BP and Fir066_BP) and those identified in sludge (Chl027_BP, Chl029_BP, Chl030_BP, Chl032_BP and Pro104_BP). The first class of MAGs (MT/MM) is able to grow in a wide range of temperatures and to “interact” with many other species. This property is shared by different taxa.

As shown in Figure 3, archaeal MAGs have few connecting edges and this suggests that they specifically interact with selected microbes. By taking into account the strength of the interactions between MAGs (represented by lines thickness), it was found that Eur033_BP (*Candidatus Methanoculleus thermohydrogenotrophicum*) (Kougias et al., 2017) is associated with The122_BP (which was found to be 98.8% identical to *Defluviitoga tunisiensis* (Dataset S1). From the investigation of *D. tunisienis* L3 genome it was found that H₂ and CO₂ are possible end-products of fermentation process (Maus et al., 2016). Since these compounds represent growth

substrates for the hydrogenotrophic methanogen, it is reasonable to explain their co-occurrence in the examined reactors. However, the ability of *D. tunisiensis* to grow in pure culture, suggests that the syntrophic interaction is not obligatory. Another point of particular interest is the strong connection between two MAGs present in the SM samples, Eur034_BP (*Methanomicrobiales* sp. DTU283) and Unc130_BP, possibly associated with the candidate division Hyd24-12 (Dataset S1). From the genome annotation and metabolic reconstruction of some recently identified MAGs belonging to this underexplored phylum, it was proposed that they are involved in acidogenesis, producing acetate and hydrogen from sugars fermentation (Kirkegaard et al., 2016). Their ability to produce these compounds is a possible explanation for their strict relation with methanogens like Eur034_BP (tentatively assigned to the *Methanolinea* genus based on 16S rRNA analysis). Additionally, since the high H₂ concentration can inhibit glycolysis and acidogenesis (Huang et al., 2015) the excess H₂ removal performed by Eur034_BP can be beneficial for both partners. Interestingly, bioinformatic investigation of the pathways present in Eur034_BP, suggests its ability to produce as end product formate, which is a known substrate used by members of the *Methanolinea* taxa (Imachi et al., 2008).

3.4 Correlation between MAGs and process parameters

The connections among reactors parameters (operational and physicochemical), the methanogenic performance and the abundance of the MAGs were explored by means of Canonical Correspondence Analysis (CCA) (Figure 2 B-D) and correlation analysis (Figure S3). Due to the marked difference in the composition of the microbial

community present among the three reactor types, correlation analyses were performed independently for MT, MM and SM.

The overall results obtained for MAGs belonging to the same phylum provided a general overview of the characteristics in the three different microcosms. It is worth noticing that MAGs belonging to *Bacteroidetes* are inversely correlated with ammonia. This finding is also supported by previous studies (Alsouleman et al., 2016), which suggested that the inverse correlation between *Bacteroidetes* and NH_4^+ level was probably due to their inability to perform SAO, a process efficient at high NH_4^+ concentrations and usually performed by Firmicutes in syntrophic cooperation with hydrogenotrophic methanogens. More specifically, inverse correlation with NH_4^+ is more evident for Bac019_BP (*Bacteroidetes* sp. DTU134) (predominant in MM) and for Bac017_BP (*Bacteroidales* sp. DTU337) (predominant in SM). Moreover, correspondence analysis indicated that, particularly in MM reactors, *Bacteroidetes* were inversely correlated with HRT suggesting that at mesophilic conditions they grow slower than other species. Correspondence analysis indicated that *Chloroflexi* MAGs have a quite similar behavior to *Bacteroidetes* and frequently are inversely correlated with free ammonia concentration (Park et al., 2016). Regarding *Actinobacteria*, MAGs were found to be inversely correlated with VFA. As previously mentioned, most of the MAGs identified in the biogas plants under investigation belonged to *Firmicutes*. The high taxonomic and functional variability harbored in this taxonomic group, probably allows the different members of this taxon to colonize reactors with very different environmental parameters. Two MAGs of particular

351 interest, Fir092_BP and Fir097_BP (*Pelotomaculum* sp. DTU098,
 352 *Syntrophomonadaceae* sp. DTU295), were present in low abundance, but showed a
 353 high correlation with biogas production. Fir092_BP (*Pelotomaculum* sp. DTU098) is
 354 the only species widespread in all the investigated reactors. Analysis of the metabolic
 355 reaction network performed using SEED evidenced its ability to oxidize propionate, a
 356 characteristic probably supporting its syntrophic growth with methanogens as
 357 reported for *Pelotomaculum thermopropionicum* (Ohashi et al., 2002) (Figure 1 and
 358 Data set S1). Some MAGs were directly correlated with VFA concentration, such as
 359 Fir077_BP, Fir078_BP (*Clostridiales Family XI Incertae Sedis* sp. DTU351-DTU352)
 360 and Fir056_BP (*Clostridiales* sp. DTU010); both CCA and CA supported this finding.
 361 VFA are status indicators for process imbalances, in fact accumulation of VFA can
 362 lead to deterioration of the process. Results obtained for Fir056_BP indicated that it is
 363 extremely abundant in MT and MM. High abundance of such a species (along with its
 364 genomic characteristics, described more in detail in the next section) suggested its
 365 involvement in the degradation of such intermediate compounds, supporting its role in
 366 maintaining the process well balanced and avoiding process inhibition. The same
 367 MAG was also identified in previous binning projects focused on the adaptation of the
 368 AD microbiome in reactors exposed to high lipid concentration (Campanaro et al.,
 369 2016), on the investigation of H₂ addition to the reactor (Treu et al., 2016b), and on
 370 the investigation of the acetate accumulation conditions in biogas reactors (Mosbæk et
 371 al., 2016). The independent identification of this MAG in multiple analyses is also
 372 supporting its central role in manure-supplemented biogas reactors.

By inspecting the phyla represented by very few MAGs, one of the most interesting was Pro111_BP (*Erysipelotrichaceae* sp. DTU294); it belongs to *Tenericutes* and it is correlated to VFA in MM samples. Abundances of Pro111_BP and some other members of the microbiome present at low abundance correlate with important biochemical parameters (Figure 2 and S1), and this suggests that the rare members of the microbiome can cover important roles in the AD process.

3.5 Functional roles of the microbial genomes in MT, MM and SM reactors

The function of proteins encoded by each MAG was derived by the exploration of databases such as KEGG and CAZy (Data set S1 WS3, WS4) and by assigning proteins in the context of the KEGG metabolic pathways. The prediction of the putative functional roles of the MAGs in the microbial community was determined only for those having completeness higher than 70% (Figure 4). In Figure 4 the completeness of the biochemical pathways identified on each MAG is reported providing a functional representation of the microbiomes. As expected this analysis evidenced that the functional capabilities of the three microcosms (TM, MM and SM) are different. Regarding the “energy metabolism” related to carbon fixation, the conversion of acetyl-CoA to acetate (phosphate acetyltransferase-acetate kinase pathway) and the reductive pentose phosphate cycle (glyceraldehyde-3P to ribulose-5P), are more frequent in the species associated with MT and MM. Thus, it is demonstrated that different pathways are involved in metabolizing VFA in sludge-supplemented samples compared to manure-supplemented samples.

Moreover, it is also interesting that most of the MAGs encoding a “nearly complete”

Wood-Ljungdahl (WL) pathway were assigned to manure-supplemented reactors (e.g. Fir042_BP, Fir050_BP, Fir051_BP, Fir100_BP and The126_BP). This is due to the higher ammonia concentration in MT and MM biogas plants which inhibits the acetoclastic methanogenic archaea. In particular Eur038_BP (*Methanosarcina* sp. DTU009) has low abundance in MM2 and MT2a, which are the reactors with the highest concentration of free ammonia (Table S1). The accumulation of acetate in the reactor is avoided by its conversion into hydrogen and carbon dioxide via the so-called syntrophic acetate oxidation (SAO) pathway. Accumulation of hydrogen and carbon dioxide was probably prevented by hydrogenotrophic archaea in order to maintain SAO energetically favorable. Interestingly, by analyzing the interaction network in Figure 3, it is evident that two of the MAGs presenting the WL pathway, namely Fir050_BP (*Clostridia* sp. DTU196) and Fir100_BP (*Tepidanaerobacter* sp. DTU063), are hubs in the network and probably work concertedly with other MAGs, but (strangely) not with methanogenic archaea. In particular, both Fir050_BP and Fir100_BP behave like Fir056_BP, which is the most abundant species in the microbiome and strongly correlated with VFA concentration. Under acetate accumulation conditions, that are probably similar to those experienced in reactors MT2a/MT2b and MT4, Fir056_BP is present at high abundance and, as previously demonstrated (Mosbæk et al., 2016), it is able to convert acetate to CO₂. Due to the presence in its genome of the formyltetrahydrofolate synthetase gene (*fhs*), it was previously proposed that this species is a syntrophic acetate-oxidizing (SAO) bacterium (Mosbæk et al., 2016). Despite these evidences, the WL pathway of

417 Fir056_BP is “truncated” and some key genes such as methylenetetrahydrofolate
 418 reductase (NADPH) (*metF*), 5-methyltetrahydrofolate corrinoid/iron sulfur protein
 419 methyltransferase (*acsE*) and acetyl-CoA synthase (*acsB*) were missing (Figure S4,
 420 Table S2). Comparative analysis of five SAO bacteria (*Clostridium ultunense*
 421 DSM10521, *Pseudothermotoga lettingae* TMO, *Syntrophaceticus schinkii*,
 422 *Tepidanaerobacter acetatoxydans* Re1, *Thermacetogenium phaeum* DSM12270)
 423 evidenced that this observation is not infrequent, and a very similar composition of
 424 the WL pathway is present for example in *Pseudothermotoga lettingae* TMO
 425 (Zhaxybayeva et al., 2009) (Table S2). It was previously reported that, the presence of
 426 “truncated” but active WL-pathway is present in the organohalide-respiring species
 427 *Dehalococcoides mccartyi* (Zhuang et al., 2014). However, a remarkable difference
 428 between *D. mccartyi* and Fir056_BP is the absence in the former species of the
 429 formate dehydrogenase alpha subunit (*fdhA*) gene which converts formate to CO₂ in
 430 the methyl branch of the WL pathway. The absence of *acsB* in Fir056_BP (as well as
 431 in *C. ultunense* DSM10521 and *P. lettingae* TMO) can be explained by the presence
 432 of the alternative glycine cleavage system (GCS) as depicted in Figure S4 (Nobu et al.,
 433 2015). The conversion of acetyl-CoA to pyruvate, which can enter the GCS system, is
 434 performed by the pyruvate-flavodoxin oxidoreductase (EC 1.2.7.-) (Figure S4). As for
 435 other syntrophs, Fir056_BP relies on the combination of the Rhodobacter nitrogen
 436 fixation (Rnf) complex (Rnf; NADH:Fd oxidoreductase) to generate reduced
 437 ferredoxin, a process working in combination with H₂ production through
 438 electron-confurcating hydrogenase (Nobu et al., 2015) (Table S3).

Regarding methane pathways, it was shown they are more represented in MT samples (Figure 4). This difference is mainly due to the higher average abundance of Eur033_BP (*Candidatus Methanoculleus thermohydrogenotrophicum*) and Eur038_BP (*Methanosarcina* sp. DTU009) in the MT samples in comparison to MM and SM. In SM the lower abundance is compensated by the higher level of MAGs associated with the *Methanosaeta* taxonomic group (Eur035_BP, 036 and 037_BP) and to the *Methanomicrobiales* (Eur034_BP).

Finally, the presence of fatty acids in the sludge-supplemented reactors (Zhu et al., 2017) supported the growth of microbial species encoding the β -oxidation pathway such as Pro103_BP, Pro104_BP, Pro_106_BP and Syn118_BP. This functional role is even more evident by checking the very high fraction of MAGs encoding the acyl-CoA synthesis pathways in SM samples in comparison to MT/MM samples.

3.6 Specific activities in hydrolytic degradation of polymers

A very crucial step in the AD process in manure-supplemented reactors is the degradation of complex fibers derived from residuals of animal nutrition. This process requires a complex set of enzymes working in a concerted way (Hess et al., 2011). In order to identify the MAGs more directly involved in this process, their proteins were analyzed using a specific software for automated annotation of carbohydrate-active enzymes (CAZy) (Yin et al., 2012). By classifying the proteins according to the classes defined in the CAZy database it was possible to shed light on the roles of the MAGs in polysaccharides degradation (Figure 5, Data set S1 WS4).

Forty MAGs with 100 or more CAZy were identified (Figure 5) and were similarly

distributed between MT, MM and SM samples. However, these MAGs display a very different gene composition, considering the aforementioned CAZy categories (Figure 5). By checking the results obtained for the most complete MAGs, it is evident that four of them have a markedly high number of CAZy domains encoded in their genomes (Fir050_BP, Bac009_BP, Bac021_BP, Ver132_BP). They differ both in terms of growth temperature and in terms of CAZy composition. Interestingly, only Fir050_BP (*Clostridia* sp. DTU196) belong to *Clostridia*, a well-known class of species with high activity in polysaccharide degradation. Regarding the other MAGs, Bac009_BP was assigned to *Bacteroidales*, Bac021_BP to *Marinilabilia* and Ver132_BP to *Verrucomicrobiales*. This finding suggested that the role of some phyla in carbohydrates degradation in AD system is far from being elucidated. Fir050_BP encodes the cellulosome structure and it has also a large number of glycoside hydrolases (Figure 5). Interestingly, Fir050_BP is one of the MAGs encoding the WL-pathway and it can be considered as the MAGs possessing multifunctional roles in the microbial community (Campanaro et al., 2016). The presence of these functional pathways in the same MAG is very interesting and it suggests that many steps of the AD food chain can be performed by the same microbe, able to degrade complex carbohydrates directly into the substrates used by methanogens. Fir050_BP was identified also in previous binning projects (Campanaro et al., 2016; Treu et al., 2016b), suggesting the importance of multifunctional MAGs in the AD microbiome. Their presence in multiple biogas reactors assesses their relevance in the “thermophilic core AD microbiome”. The other three MAGs (Bac009_BP,

Bac021_BP, Ver132_BP) have a similar CAZy composition with many glycoside hydrolases and carbohydrate esterases. This clearly suggests their role in the hydrolytic step of the AD and their variable abundance partially reflects differences in poly/oligosaccharides in the influent feedstocks. Finally, despite in bioreactors the cellulosome is usually associated with the thermophilic members of *Clostridiales* (Tsapekos et al., 2017), a mesophilic *Bacteroidales* (Bac010_BP) had 7 dockerins and 9 cohesins localized in 14 proteins. The taxonomy of this MAG was confirmed both considering the 16S rRNA gene sequence and by selected taxonomic informative proteins (PhyloPhlAn). Despite the presence of dockerins in MAGs associated with this phylum has been recently demonstrated (Bensoussan et al., 2017; Svartström et al., 2017), the simultaneous presence of dockerins and cohesins in this taxa is a completely new finding. In the present study it was shown that mesophilic AD system harbors *Bacteroidetes* species with remarkable novel properties in polysaccharides degradation.

4. Conclusions

The present study revealed the plasticity of the AD microbiome in response to different environment parameters existing in biogas plants. Investigation of the complete set of biochemical pathways in multiple samples and its connection with different layers of information (MAGs abundance profile, predicted functional pathways and environmental factors) allowed the identification of species associated to crucial operational parameters such as temperature, feedstock, VFA concentration and biogas production. This approach revealed important relationships existing

between high-abundant members of the thermophilic reactors (e.g. acetate utilizers), their metabolic pathways and biochemical data. Additionally, it promises to become a widely used strategy in the future to predict the functional role of microbes participating in anaerobic decomposition of organic matter.

Acknowledgments

The study was funded by National Key Research and Development Program of China (2015BAD15B06, 2017YFC0212205), National Natural Science Foundation of China (51408133, 51750110498), State Key Laboratory of Pollution Control and Resource Reuse Foundation (PCRRF16009), and Science and Technology Commission of Shanghai Municipality (17230740900).

Supporting Information

Details regarding MAGs taxonomy, genome properties, abundance in different reactors, KEGG modules/pathways and CAZy composition are reported in Data set S1 (xlsx). Parameters of the reactors, selected genes belonging to the WL pathway and some associated metabolic functions, graphics for Pearson correlation and graphics for the WL-pathway in Fir056_BP are reported in Table S1-S3, Fig S1-S4 in supporting information.

References

Alsouleman, K., Linke, B., Klang, J., Klocke, M., Krakat, N., Theuerl, S., 2016. Reorganisation of a mesophilic biogas microbiome as response to a stepwise increase of ammonium nitrogen induced by poultry manure supply. *Bioresour.*

- Technol. 208, 200–204.
- Amani, T., Nosrati, M., Sreekrishnan, T.R., 2010. Anaerobic digestion from the viewpoint of microbiological, chemical, and operational aspects — a review. Environ. Rev. 18, 255–278.
- Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9, 75.
- Bassani, I., Kougias, P.G., Treu, L., Angelidaki, I., 2015. Biogas Upgrading via Hydrogenotrophic Methanogenesis in Two-Stage Continuous Stirred Tank Reactors at Mesophilic and Thermophilic Conditions. Environ. Sci. Technol. 49, 12585–93.
- Bensoussan, L., Morais, S., Dassa, B., Friedman, N., Henrissat, B., Lombard, V., Bayer, E.A., Mizrahi, I., 2017. Broad phylogeny and functionality of cellulosomal components in the bovine rumen microbiome. Environ. Microbiol. 19, 185–197.
- Bolger, A., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics.
- Bremges, A., Maus, I., Belmann, P., Eikmeyer, F., Winkler, A., Albersmeier, A., Pühler, A., Schlüter, A., Sczyrba, A., 2015a. Deeply sequenced metagenome and

- 549 metatranscriptome of a biogas-producing microbial community from an
550 agricultural production-scale biogas plant. *Gigascience* 4, 33.
- 551 Campanaro, S., Treu, L., Cattani, M., Kougias, P.G., Vendramin, V., Schiavon, S.,
552 Tagliapietra, F., Giacomini, A., Corich, V., 2017. In vitro fermentation of key
553 dietary compounds with rumen fluid: A genome-centric perspective. *Sci. Total*
554 *Environ.* 584–585.
- 555 Campanaro, S., Treu, L., Kougias, P.G., De Francisci, D., Valle, G., Angelidaki, I.,
556 2016. Metagenomic analysis and functional characterization of the biogas
557 microbiome using high throughput shotgun sequencing and a novel binning
558 strategy. *Biotechnol. Biofuels* 9.
- 559 Campanaro, S., Treu, L., Kougias, P.G., Zhu, X., Angelidaki, I., 2018. Taxonomy of
560 anaerobic digestion microbiome reveals biases associated with the applied high
561 throughput sequencing strategies. *Sci. Rep.* 8, 1926.
- 562 Dixon, P., 2009. VEGAN, a package of R functions for community ecology.
- 563 Fontana, A., Patrone, V., Puglisi, E., Morelli, L., Bassi, D., Garuti, M., Rossi, L.,
564 Cappa, F., 2016. Effects of geographic area, feedstock, temperature, and
565 operating time on microbial communities of six full-scale biogas plants.
566 *Bioresour. Technol.* 218, 980–990.
- 567 Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., Luo, S.,
568 Clark, D.S., Chen, F., Zhang, T., Mackie, R.I., Pennacchio, L.A., Tringe, S.G.,
569 Visel, A., Woyke, T., Wang, Z., Rubin, E.M., 2011. Metagenomic Discovery of
570 Biomass-Degrading Genes and Genomes from Cow Rumen. *Science* (80-.). 331,

- 571 463–467.
- 572 Hollister, E.B., Forrest, A.K., Wilkinson, H.H., Ebbole, D.J., Tringe, S.G., Malfatti,
 573 S.A., Holtzapple, M.T., Gentry, T.J., 2012. Mesophilic and thermophilic
 574 conditions select for unique but highly parallel microbial communities to
 575 perform carboxylate platform biomass conversion. PLoS One 7, e39689.
- 576 Huang, W., Wang, Z., Zhou, Y., Ng, W.J., 2015. The role of hydrogenotrophic
 577 methanogens in an acidogenic reactor. Chemosphere 140, 40–6.
- 578 Huson, D.H., Scornavacca, C., 2012. Dendroscope 3: an interactive tool for rooted
 579 phylogenetic trees and networks. Syst. Biol. 61, 1061–7.
- 580 Hyatt, D., LoCascio, P.F., Hauser, L.J., Uberbacher, E.C., 2012. Gene and translation
 581 initiation site prediction in metagenomic sequences. Bioinformatics 28, 2223–
 582 2230.
- 583 Imachi, H., Sakai, S., Sekiguchi, Y., Hanada, S., Kamagata, Y., Ohashi, A., Harada, H.,
 584 2008. *Methanolinea tarda* gen. nov., sp. nov., a methane-producing archaeon
 585 isolated from a methanogenic digester sludge. Int. J. Syst. Evol. Microbiol. 58,
 586 294–301.
- 587 Imelfort, M., Parks, D., Woodcroft, B.J., Dennis, P., Hugenholtz, P., Tyson, G.W.,
 588 2014. GroopM: an automated tool for the recovery of population genomes from
 589 related metagenomes. PeerJ 2, e603.
- 590 Jing, Y., Campanaro, S., Kougias, P., Treu, L., Angelidaki, I., Zhang, S., Luo, G., 2017.
 591 Anaerobic granular sludge for simultaneous biomethanation of synthetic
 592 wastewater and CO with focus on the identification of CO-converting

- 593 microorganisms. Water Res. 126, 19-28.
- 594 Kang, D.D., Froula, J., Egan, R., Wang, Z., 2015. MetaBAT, an efficient tool for
595 accurately reconstructing single genomes from complex microbial communities.
596 PeerJ 3, e1165.
- 597 Kirkegaard, R.H., Dueholm, M.S., McIlroy, S.J., Nierychlo, M., Karst, S.M.,
598 Albertsen, M., Nielsen, P.H., 2016. Genomic insights into members of the
599 candidate phylum Hyd24-12 common in mesophilic anaerobic digesters. ISME J.
600 10, 2352–64.
- 601 Kougias, P.G., Campanaro, S., Treu, L., Zhu, X., Angelidaki, I., 2017. A novel
602 archaeal species belonging to *Methanoculleus* genus identified via de-novo
603 assembly and metagenomic binning process in biogas reactors. Anaerobe 46.
- 604 Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.-H., Rognes, T., Ussery, D.W.,
605 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes.
606 Nucleic Acids Res. 35, 3100–8.
- 607 Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat.
608 Methods 9, 357–9.
- 609 Luo, G., Fotidis, I.A., Angelidaki, I., 2016. Comparative analysis of taxonomic,
610 functional, and metabolic patterns of microbiomes from 14 full-scale biogas
611 reactors by metagenomic sequencing and radioisotopic analysis. Biotechnol.
612 Biofuels 9, 51.
- 613 Maus, I., Cibis, K.G., Bremges, A., Stolze, Y., Wibberg, D., Tomazetto, G., Blom, J.,
614 Sczyrba, A., König, H., Pühler, A., Schlüter, A., 2016. Genomic characterization

- 615 of *Defluviitoga tunisiensis* L3, a key hydrolytic bacterium in a thermophilic
 616 biogas plant and its abundance as determined by metagenome fragment
 617 recruitment. J. Biotechnol. 232, 50–60.
- 618 Mosbæk, F., Kjeldal, H., Mulat, D.G., Albertsen, M., Ward, A.J., Feilberg, A., Nielsen,
 619 J.L., 2016. Identification of syntrophic acetate-oxidizing bacteria in anaerobic
 620 digesters by combined protein-based stable isotope probing and metagenomics.
 621 ISME J. 10, 2405–18.
- 622 Nobu, M.K., Narihiro, T., Rinke, C., Kamagata, Y., Tringe, S.G., Woyke, T., Liu,
 623 W.-T., 2015. Microbial dark matter ecogenomics reveals complex synergistic
 624 networks in a methanogenic bioreactor. ISME J. 9, 1710–1722.
- 625 Ohashi, A., Sekiguchi, Y., Hanada, S., Harada, H., Kamagata, Y., Imachi, H., 2002.
 626 *Pelotomaculum thermopropionicum* gen. nov., sp. nov., an anaerobic,
 627 thermophilic, syntrophic propionate-oxidizing bacterium. Int. J. Syst. Evol.
 628 Microbiol. 52, 1729–1735.
- 629 Park, S., Cui, F., Mo, K., Kim, M., 2016. Mathematical models and bacterial
 630 communities for ammonia toxicity in mesophilic anaerobes not acclimated to
 631 high concentrations of ammonia. Water Sci. Technol. 74, 935–942.
- 632 Parks, D.H., Beiko, R.G., 2013. Measures of phylogenetic differentiation provide
 633 robust and complementary insights into microbial communities. ISME J. 7, 173–
 634 83.
- 635 Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015.
 636 CheckM: assessing the quality of microbial genomes recovered from isolates,

- single cells, and metagenomes. *Genome Res.* 25, 1043–55.
- Prestat, E., David, M.M., Hultman, J., Taş, N., Lamendella, R., Dvornik, J., Mackelprang, R., Myrold, D.D., Jumpponen, A., Tringe, S.G., Holman, E., Mavromatis, K., Jansson, J.K., 2014. FOAM (Functional Ontology Assignments for Metagenomes): a Hidden Markov Model (HMM) database with environmental focus. *Nucleic Acids Res.* 42, e145.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590-6.
- Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–2.
- Rivière, D., Desvignes, V., Pelletier, E., Chaussonnerie, S., Guermazi, S., Weissenbach, J., Li, T., Camacho, P., Sghir, A., 2009. Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. *ISME J.* 3, 700–14.
- Ruiz-Sánchez, J., Campanaro, S., Guivernau, M., ernández, B., Prenafeta-Boldú, F., 2017. Effect of ammonia on the active microbiome and metagenome and from stable full-scale digesters. *Bioresour. Technol.* 250, 513–522.
- Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V., Quackenbush, J., 2003. TM4: a free, open-source system for microarray data

- management and analysis. *Biotechniques* 34, 374–8.
- Salzberg, S.L., 2011. FLASH: Fast Length Adjustment of Short Reads to Improve Genome Assemblies Tanja Mago 1–8.
- Schlüter, A., Bekel, T., Diaz, N.N., Dondrup, M., Eichenlaub, R., Gartemann, K.-H., Krahn, I., Krause, L., Krömeke, H., Kruse, O., Mussnug, J.H., Neuweiger, H., Niehaus, K., Pühler, A., Runte, K.J., Szczepanowski, R., Tauch, A., Tilker, A., Viehöver, P., Goesmann, A., 2008. The metagenome of a biogas-producing microbial community of a production-scale biogas plant fermenter analysed by the 454-pyrosequencing technology. *J. Biotechnol.* 136, 77–90.
- Segata, N., Börnigen, D., Morgan, X.C., Huttenhower, C., 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* 4, 2304.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13, 2498–2504.
- Stolze, Y., Bremges, A., Ruming, M., Henke, C., Maus, I., Pühler, A., Sczyrba, A., Schlüter, A., 2016. Identification and genome reconstruction of abundant distinct taxa in microbiomes from one thermophilic and three mesophilic production-scale biogas plants. *Biotechnol. Biofuels* 9, 156.
- Sundberg, C., Al-Soud, W.A., Larsson, M., Alm, E., Yekta, S.S., Svensson, B.H., Sørensen, S.J., Karlsson, A., 2013. 454 pyrosequencing analyses of bacterial and

- archaeal richness in 21 full-scale biogas digesters. FEMS Microbiol. Ecol. 85, 612–26.
- Svartström, O., Alneberg, J., Terrapon, N., Lombard, V., de Bruijn, I., Malmsten, J., Dalin, A.-M., El Muller, E., Shah, P., Wilmes, P., Henrissat, B., Aspeborg, H., Andersson, A.F., 2017. Ninety-nine de novo assembled genomes from the moose (*Alces alces*) rumen microbiome provide new insights into microbial plant biomass degradation. ISME J. 11, 2538–2551.
- Takami, H., Taniguchi, T., Arai, W., Takemoto, K., Moriya, Y., Goto, S., 2016. An automated system for evaluation of the potential functionome: MAPLE version 2.1.0. DNA Res. 23, 467–475.
- Torondel, B., Ensink, J.H.J., Gundogdu, O., Ijaz, U.Z., Parkhill, J., Abdelahi, F., Nguyen, V.A., Sudgen, S., Gibson, W., Walker, A.W., Quince, C., 2016. Assessment of the influence of intrinsic environmental and geographical factors on the bacterial ecology of pit latrines. Microb. Biotechnol. 9, 209–23.
- Treu, L., Campanaro, S., Kougias, P.G., Zhu, X., Angelidaki, I., 2016a. Untangling the Effect of Fatty Acid Addition at Species Level Revealed Different Transcriptional Responses of the Biogas Microbial Community Members. Environ. Sci. Technol. 50.
- Treu, L., Kougias, P.G., Campanaro, S., Bassani, I., Angelidaki, I., 2016b. Deeper insight into the structure of the anaerobic digestion microbial community; The biogas microbiome database is expanded with 157 new genomes. Bioresour. Technol. 216.

- 703 Tringe, S.G., Rubin, E.M., 2005. Metagenomics: DNA sequencing of environmental
704 samples. *Nat. Rev. Genet.* 6, 805–814.
- 705 Tsapekos, P., Kougias, P.G., Treu, L., Campanaro, S., Angelidaki, I., 2017. Process
706 performance and comparative metagenomic analysis during co-digestion of
707 manure and lignocellulosic biomass for biogas production. *Appl. Energy* 185,
708 126–135.
- 709 Vanwonterghem, I., Jensen, P.D., Rabaey, K., Tyson, G.W., 2016. Genome-centric
710 resolution of microbial diversity, metabolism and interactions in anaerobic
711 digestion. *Environ. Microbiol.* 18, 3144–58.
- 712 Varghese, N.J., Mukherjee, S., Ivanova, N., Konstantinidis, K.T., Mavrommatis, K.,
713 Kyrpides, N.C., Pati, A., 2015. Microbial species delineation using whole
714 genome sequences. *Nucleic Acids Res.* 43, 6761–6771.
- 715 Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naive Bayesian classifier for
716 rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl.*
717 *Environ. Microbiol.* 73, 5261–7.
- 718 Werner, J.J., Knights, D., Garcia, M.L., Scalfone, N.B., Smith, S., Yarasheski, K.,
719 Cummings, T.A., Beers, A.R., Knight, R., Angenent, L.T., 2011. Bacterial
720 community structures are unique and resilient in full-scale bioenergy systems.
721 *Proc. Natl. Acad. Sci. U. S. A.* 108, 4158–63.
- 722 Wirth, R., Kovács, E., Maróti, G., Bagi, Z., Rákhely, G., Kovács, K.L., 2012.
723 Characterization of a biogas-producing microbial community by short-read next
724 generation DNA sequencing. *Biotechnol. Biofuels* 5, 41.

- 725 Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., Xu, Y., 2012. dbCAN: a web resource
726 for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 40,
727 W445–W451.
- 728 Zhaxybayeva, O., Swithers, K.S., Lapierre, P., Fournier, G.P., Bickhart, D.M., DeBoy,
729 R.T., Nelson, K.E., Nesbo, C.L., Doolittle, W.F., Gogarten, J.P., Noll, K.M., 2009.
730 On the chimeric nature, thermophilic origin, and phylogenetic placement of the
731 Thermotogales. *Proc. Natl. Acad. Sci.* 106, 5865–5870.
- 732 Zhu, F., Wu, X., Zhao, L., Liu, X., Qi, J., Wang, X., Wang, J., 2017. Lipid profiling in
733 sewage sludge. *Water Res.* 116, 149–158.
- 734 Zhuang, W.Q., Yi, S., Bill, M., Brisson, V.L., Feng, X., Men, Y., Conrad, M.E., Tang,
735 Y.J., Alvarez-Cohen, L., 2014. Incomplete Wood-Ljungdahl pathway facilitates
736 one-carbon metabolism in organohalide-respiring *Dehalococcoides mccartyi*.
737 *Proc. Natl. Acad. Sci. U. S. A.* 111, 6419–24.
- 738

Figure 1. Abundance of MAGs in the reactors under investigation and their taxonomy.

Abundance of MAGs is represented as a heatmap and data were clustered according to Pearson to highlight similar abundance profiles. In the right part of the figure it is reported the taxonomic assignment at phylum level of MAGs (see inset at the top of

the figure). On top of the figure the names of the biogas plants are reported and colored boxes are representative of the manure-based thermophilic reactors (MT/red), manure-based mesophilic reactors (MM/orange) and sludge-based mesophilic reactors (SM/green).

Figure 2. Non-metric Multidimensional Scaling (NMDS) and Canonical Correspondence Analysis (CCA). (A) The dissimilarities in the samples matrix were analyzed and plotted in a low-dimensional space. MT (Thermophilic Manure samples), MM (Mesophilic Manure samples), SM (Mesophilic Sludge samples). In (B) the results obtained from CCA performed on the MAGs assigned to MT and MT/MM, in (C) results obtained for the MM and MT/MM and in (D) those obtained for SM. The MAGs identified as significantly correlated to environmental parameters using correlation analysis are highlighted in red, while names of the samples are reported in black.

Figure 3. Network analysis of the AD microbiome. The co-occurrence of MAGs in different samples is depicted as a network. MAGs are represented by circles with area related to the highest abundance value determined in the samples examined. Colors are representative of the phyla, while edges (i.e. connecting lines) thickness is proportional to the Pearson correlation coefficient calculated between each couple of MAGs. Unconnected circles are not reported in figure.

Figure 4. In-silico functional analysis of MAGs. The completeness level of the KEGG pathways identified in the MAGs is reported on top-left. The prevalence of the MAGs in MT/MM/SM samples is reported on the left part of the figure using colors (red/orange/ green).

Figure 5. Analysis of CAZY in MAGs. Number of CAZY domains identified in 40 selected MAGs.

In the bottom part of the figure MAGs are highlighted considering their prevalence in MT (red),

MM (orange) and MS (green). CAZY domains are classified according to five main categories.

Highlights:

- Metagenomic binning identified core MAGs of the anaerobic digestion microbiome
- CCA revealed influence of biotic and environmental factors on MAGs abundance
- Functional properties of core MAGs were analyzed
- Network analysis revealed the syntrophic interactions between MAGs