Title

Integrative analyses to investigate the link between microbial activity and molecules degradation during anaerobic digestion

Author names + Affiliations

Laetitia Cardona[1], Kim Anh Lê Cao[2], Francesc Puig-Castellví[1,3], Chrystelle Bureau[1], Céline Madigou[1], Laurent Mazéas[1], Olivier Chapleur[1*]

[1]Université Paris-Saclay, INRAE, PROSE,
1 rue Pierre-Gilles de Gennes, CS 10030, 92761 Antony Cedex, France
[2]Melbourne Integrative Genomics research unit, Melbourne University,
Building 184/30 Royal Parade, Parkville VIC 3052, Australia
[3]Ingénierie, Procédés, Aliments research unit, AgroParisTech, INRA, Université Paris-Saclay,
1 rue des Olympiades, 91300 Massy, France

***Corresponding author**
Tel + 33 1 40 96 65 06
Fax + 33 1 40 96 61 99
olivier.chapleur@irstea.fr
https://orcid.org/0000-0001-9460-921X

Abstract

Anaerobic digestion (AD) is a promising biological process to convert waste into sustainable energy. However, the microbiota involved in this bioprocess is complex and additional knowledge is still needed to fully exploit its capability. High throughput methodologies open new perspectives, but innovative data integration methodologies are required for extracting relevant information from these rich data. We analysed the association between microbial activity and the patterns of substrate degradation during a lab-scale co-digestion experiment. These parameters were longitudinally monitored using 16S rRNA sequencing and untargeted metabolomics. In this experiment, samples were collected from digesters fed with 9 different mixtures of fish, sewage sludge, and grass. Our objective was to identify microorganisms

responsible for the degradation of molecules specific of each co-substrate. Five main groups of correlated features were successfully evidenced. For example, the degradation of cadaverine was found to be correlated with microorganisms from the order *Clostridiales* and the genus *Methanosarcina*, and the degradation of lignin compounds was correlated with cellulolytic degraders *Lactobacillales*. This study highlights the potential of data integration towards a comprehensive understanding of AD microbiota.

Keywords

16S RNA sequencing; metabolomic; data integration; methanisation; co-digestion; PLS canonical

Introduction

Deciphering the microbial communities from diversified domains such as health, food safety and environment has been widely addressed by the development of the high-throughput technologies and adapted computational statistical methods. It is now possible to study the structure, activity, interaction and function of complex microbial communities using new methods from the genomics, transcriptomics, proteomics and metabolomics fields (Vanwonterghem et al., 2014). In parallel to the important advancements in these fields, the recent development of specific statistical methods and user-friendly workflows has substantially improved the visualisation, analysis and integration of the omics results (Bouhlel et al., 2018; Callahan et al., 2016; Rohart et al., 2017; Singh et al., 2019).

Anaerobic digestion (AD) is a promising bioprocess to provide sustainable energy. Indeed, driven by a complex microbial community, AD allows to transform organic waste into biogas. This process involves four different steps: hydrolysis, acidogenesis, acetogenesis and methanogenesis managed by the interaction of bacteria and archaea. Nonetheless, this

microbial diversity and its interaction are highly sensitive and depend on multiple parameters such as temperature (Madigou et al., 2019; Noll et al., 2010), presence of inhibitors such as phenol (Poirier et al., 2016), ammonia (Li et al., 2017), long chain fatty acids (Sousa et al., 2013), and feeding composition (Zamanzadeh et al., 2017). Several studies have already shown the utility of using omics technologies to decipher the anaerobic microbial population (Amha et al., 2018; Hassa et al., 2018). The level of information differs according to the type of used omics technology. Analyses of a single omics data type are routinely carried out (Bize et al., 2015; Cai et al., 2016); but the use of several omics at the same time is rare. For example Beale *et al.* applied metagenomic and metabolomic approaches to obtain new insights on the diversity and activity of the anaerobic population after stress (Beale et al., 2016). However, no direct correlation was assessed between the two approaches. More generally, in the literature omics data integration methodologies are still sparse.

The aim of this study was to evaluate a methodology to correlate the microbial activity to the metabolites specific of the substrates degradation. For that purpose, 3 different substrates (sewage sludge, grass, and fish) of different chemical composition were used. In this study, 27 anaerobic bioreactors fed with binary mixtures of the substrates (sludge and grass, or sludge and fish) at different proportions (0-100, 25-75, 50-50, 75-25, 100-0) were employed. By using these binary mixtures, it is possible to evaluate the evolution of the substrate-specific microorganisms and molecules. The active microbial community was analysed through the sequencing of the 16S rRNA. While DNA can inform us about the total diversity of a community (that includes dead, dormant and living microorganisms), RNA sequencing allow us to only characterise microorganisms active in a given environment (Lin et al., 2016) and to study therefore the metabolically active microorganisms (De Vrieze et al., 2018). Untargeted metabolomics using LC-MS was performed to study the pattern of the substrate degradation. Metabolomic experiments can be potentially used to determine the

molecular fingerprints of waste degradation and to monitor the patterns of substrate degradation (Villas-Bôas et al., 2006).

To our knowledge, this is the first study that studies the correlation between microbial activity from 16S rRNA sequencing data and pattern of substrate degradation from metabolomics data. Our computational analyses revealed subsets of active microorganisms highly associated to dynamics of substrate degradation over time, and posit novel hypotheses regarding the capacity of microorganisms to degrade specific molecules. For instance, cadaverine degradation was correlated to *Clostridiales* and *Methanosarcina,* which suggests a possible syntrophic relationship between these two microorganisms. Such association was already highlighted with a different anaerobic syntrophy between a *Clostridium*-like bacteria and *Methanospirullum* archaea (Roeder and Schink, 2009).

Methods

Feedstock preparation

The inoculum used in the digestion experiments was sampled from a mesophilic full scale industrial anaerobic bioreactor treating primary sludge from a wastewater treatment plant (Valenton, France). The inoculum was incubated in anaerobic condition at 35°C without feeding in order to degrade the organic matter in excess before to carry out the experiments.

Substrates used in the experiments were wastewater sludge collected from an industrial wastewater treatment plant (Valenton, France), fish waste obtained from a fish shop, and garden grass mowed from IRSTEA institute. Fish and grass wastes were crushed and kept at 4°C until the incubation experiments.

Bioreactors experimental set-up

Binary mixtures of sludge with linearly increasing (0-100, 25-75, 50-50, 75-25, 100-0) percentages of either fish or grass were prepared (Figure 1). Experiments were carried out in

1L glass bottles (700 mL working volume) at 35°C in the dark without agitation. The same quantity of carbon was added in all the digesters, and the ratio of substrate/inoculum used to feed and inoculate all the digesters was fixed at 12 gCOD/1.2 gCOD (Table S1). All the bioreactors were complemented with a biochemical potential buffer (International Standard ISO 11734 (1995)) to reach a final working volume of 700 mL. All incubations were performed in triplicate. The bioreactors were then sealed with a screw cap and a rubber septum and the headspaces were then flushed with N_2 (purity >99.99%, Linde gas SA). In total 27 anaerobic bioreactors were set-up.

Weekly (at days 0, 7, 14, 21), for every reactor, 6 mL of liquid phase were sampled through the septum using a syringe. The collected samples were centrifuged at 10 000g for 10 minutes to separate the supernatants from the pellets. Supernatant were snap frozen using liquid nitrogen and kept at -20°C for metabolomic analysis and pellets kept at -80°C for microbial analysis.

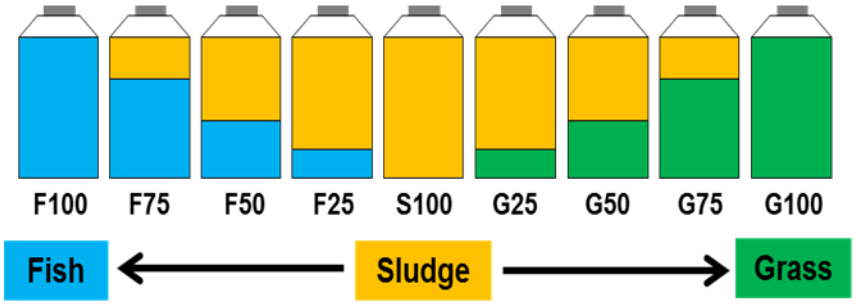


Figure 1. Scheme of the batch experimental design. S100 stands for wastewater sludge alone, F25, F50, F75, F100 stands for respectively 25, 50, 75 or 100% of fish (F) in co-digestion with sludge, G25, G50, G75, G100 stands for respectively 25, 50, 75 or 100% of Grass (G) in co-digestion with sludge

RNA extraction and 16S rRNA sequencing

Based on the biogas production, a total of 22 samples were selected (active biogas production - Figure 2). Since fish-containing digesters showed a higher delay in the biogas production than grass-containing digesters, the used sampling time-points were different for the two substrates (day 14 for grass-containing digesters, and day 28 for fish-containing digesters). In addition, the microbial analysis was also performed for samples from the same digesters collected at day 21.

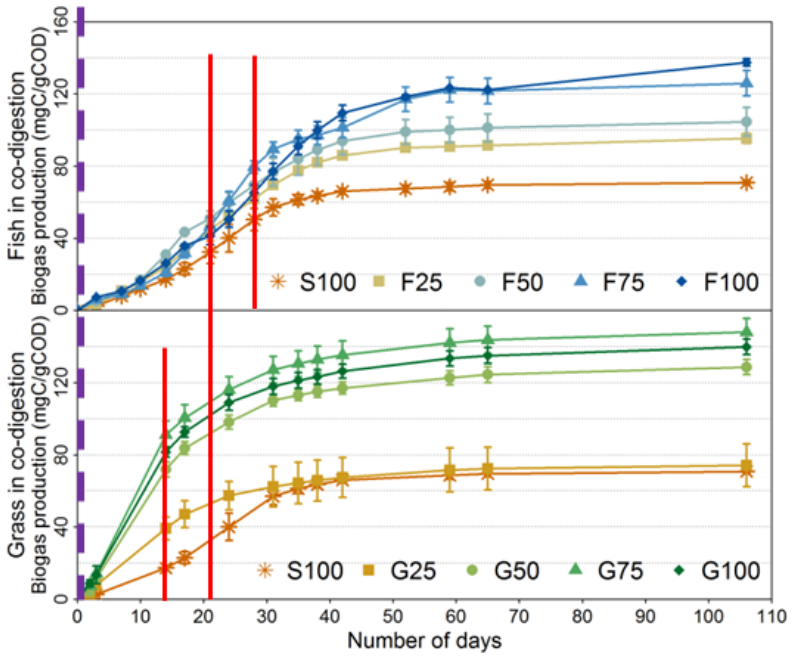


Figure 2. Cumulated biogas production (mgC/gCOD) over time (Days) for the different bioreactors. Mean values of the triplicate bioreactors, error bars represent standard deviations within triplicates. S100 stands for wastewater sludge alone; F25, F50, F75, F100 stands for respectively 25, 50, 75 or 100% of fish (F) in co-digestion with sludge; G25, G50, G75, G100 stands for respectively 25, 50, 75 or 100% of Grass (G) in co-digestion with sludge. Red solid lines correspond to the 16S rRNA sequencing points and metabolomic points. Purple dashed line corresponds to the point where only metabolomic analysis was carried out.

The commercial kit FastRNA Pro™ Soil-Direct (MP Biomedicals) was used to extract the total RNA following the manufacturer's specifications. TURBO™ DNase (Ambion) kit following the manufacturer's instructions allowed to remove DNA co-extracted. The RNA was denaturated by 2 min at 85°C in a dry bath and was then stored on ice. RNAClean XP magnetic beads purification system (Beckman Coulter) was used to RNA purification by adding 1.8 volumes of beads by volume of RNA. After mixing by pipetting and 5 min of incubation, beads were captured using a magnetic rack on one side of the tube and then washed by adding 500 µL of 70% cold ethanol (diluted in DEPC-water). Tubes were incubated during 30 seconds at room temperature and ethanol was removed then. This washing step was repeated 3 times. Once ethanol finally evaporated, beads were resuspended with DEPC-water to eluted RNA from the beads. Finally, beads were removed using the

magnetic rack and RNA was recovered in the supernatant. The integrity and quantity of the RNA was evaluated using High Sensitivity RNA ScreenTape and 4200 TapeStation (Agilent Technologies) following the manufacturer's protocol.

A reverse transcription polymerase chain reaction (RT-PCR) was carried out on the RNA using the mix iScript Reverse Transcription Supermix (Biorad) and the following thermocycler program: 5 min at 25°C, 30 min at 42°C and 5 min at 85°C. The cDNA was quantified using Qubit 2.0 fluorometer (ssDNA assay kit, Invitrogen, Life Technologies).

Archaeal and bacterial hyper variable region V4-V5 of the 16S rRNA gene were amplified as cDNA, and these amplicons were then sequenced according to the protocol described by Madigou *et al.* (Madigou et al., 2019).

FROGS (Find Rapidly OTU with Galaxy Solution), a galaxy/CLI workflow (Escudié et al., 2018), was used to design an OTU count matrix. R CRAN software (version 3.5.1) was used to examine the OTUs abundances. Alpha diversity was analysed using Shannon method using phyloseq R package (version 1.20.0). Considering the dispersion in the total number of reads identified in each sample, archaeal and bacterial OTUs abundances were scaled with total sum. Only OTUs that exceeded 1% in terms of relative abundance in at least one sample, were selected for the analysis and square-root transformed.

Metabolomic analysis

Metabolomic analysis was performed on all collected supernatants. Samples were analysed using reverse phase liquid chromatography coupled to high resolution mass spectrometry (HPLC-ESI-HRMS) using a LTQ-Orbitrap XL instrument (Thermo Scientific). Samples were diluted at 1/10 in water and 10 µL of the solution was injected into the analytical system. Chromatographic separation was performed on Accela 1250 pump at 400 µL/min with a linear gradient from 10 to 80% of mobile phase A (acetonitrile + 0.05% formic acid) and 90 to 20 % of mobile phase B (water + 0.05% formic acid) into a syncronis C18

column (50x2.1 mm, 1.7µm, Thermo Scientific) during 23 minutes, followed by a stabilization phase of 5 minutes to return at the initial condition. After chromatographic separation, the sample was ionized by electrospray ionization (ESI) on positive mode. The detection was performed in full scan over an m/z range from 50 to 500 at a resolution of 100 000. A sample consisting on the supernatant from the digestion of anaerobic sludge was used as a quality control and injected every 10 experiment samples, blank samples were injected every 10 samples, and an equimolar mix of the samples was injected every 5 samples.

The raw data obtained from the LC-HRMS analyses were transformed into mzXML files using MSConvert (ProteoWizard 3.0). The XCMS R package (version 1.52.0) was used to process the data (Smith et al., 2006). The method *centWave* was used to determine chromatographic peaks (ROIs) with a *m/z* error of 10 ppm and a peakwidth between 20 and 50 seconds. The ROIs found in different sample were grouped using the group method with a bandwidth of 30. ROIs retention times from the same ROI groups were unified across samples using the *orbiwarp* method. A second grouping was carried out using a bandwidth of 25. Finally, missing ROIs in the samples were filled using the *fillPeaks* method. Initial metabolite identification was performed based on the comparison of the accurate molecular mass measured by LC-HRMS with the corresponding values found in the online databases HMDB, LipidMaps, and PubChem (Fahy et al., 2007; Lee et al., 2018; Wishart et al., 2018). In addition, for selected compounds, a confirmation of the metabolite identification was performed by MS/MS fragmentation, and followed by the comparison of the acquired MS/MS spectra with the theoretical spectra from the online databases HMDB and MassBank (Horai et al., 2010).

Statistical analysis

Statistical methods from mixOmics R package (Rohart et al., 2017) were used to highlight the relationship between the key microorganisms and the molecules degradation pattern.

Firstly, in order to highlight the most active microorganisms involved in the substrate degradation independently of the mixture composition, the method sparse Principal Component Analysis (sPCA, (Shen and Huang, 2008)) was used on the microbial dataset. This method allows to select the OTUs that highly contributed to explain the main source of variance in the data.

To study the dynamics of the metabolites between the different feeding types, we used sparse Partial Least Squares Discriminant Analysis (sPLS-DA, (Lê Cao et al., 2011)) on the mono-digestion samples (S100, F100, G100) at day 0 to select the most discriminative molecules. This analysis enabled to identify the molecules specific of each substrate at the beginning of the experiment. These most discriminant molecules were selected based on the error rate of classification obtained for each component when adding progressively the number of molecules.

The patterns of abundance of the selected OTUs and molecules within the mixtures and over time were observed with heatmaps (heatmap.2 function from gplots R package, version 3.0.1) using the Ward method and Manhattan Euclidean distance for respectively OTUs and molecules abundances.

Secondly, to study the correlation between microbial activities and molecules degradation over time, the selected molecules data were transformed using a ratio between the molecules abundances at Day 0 and days 14, 21 or 28. The two sets of data (selected microorganisms and molecules in the first part) were statistically integrated using a PLS canonical mode (Lê Cao et al., 2009) and their variabilities through the different mixtures

were studied. Microorganisms and molecules with a similar pattern of evolution were grouped together using a hierarchical clustering using complete linkage method to identify the microorganisms involved in the degradation of the molecules.

Results and discussion

A comparison of the major chemical parameters of anaerobic digestion is described in a precedent study (Cardona et al., 2019). This study describes the influence of adding a second substrate to improve the digestion performances. Because the present study focuses on the correlation of microbial activity and molecule degradation, the performance results will not be described in details here.

Influence of the feeding composition on the microbial community

The alpha diversity using Shannon index (supplemental Figure S1) was calculated on archaeal and bacterial communities in order to evaluate the influence of the feeding composition on the microbial diversity. We observed that the archaeal and the bacterial diversity is inversely proportional to the amount of fish added in the digester. The low microbial diversity induced by the presence of fish substrate might be explained by a simpler substrate composition (in terms of molecular variety) in this substrate, or by a higher functional redundancy of the microorganisms that can degrade fish substrate if compared to those that can degrade grass or sewage sludge.

The sPCA sample plot highlighted the influence of the feeding composition on the microorganisms abundance (Figure 3). We identified 43 OTUs representing 65 to 85% of the total microbial community that explained most of the variability in the 16S rRNA data. Samples were grouped in three clusters, according to the major co-substrate, and regardless of their sample collection time. This result reveals low time variability within the conditions and a stable microbial community over time.

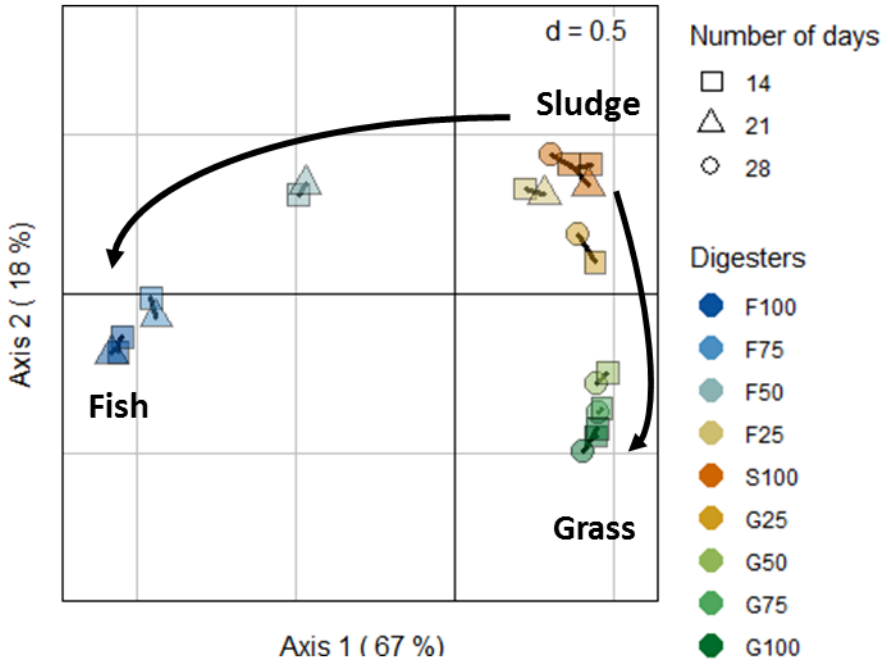
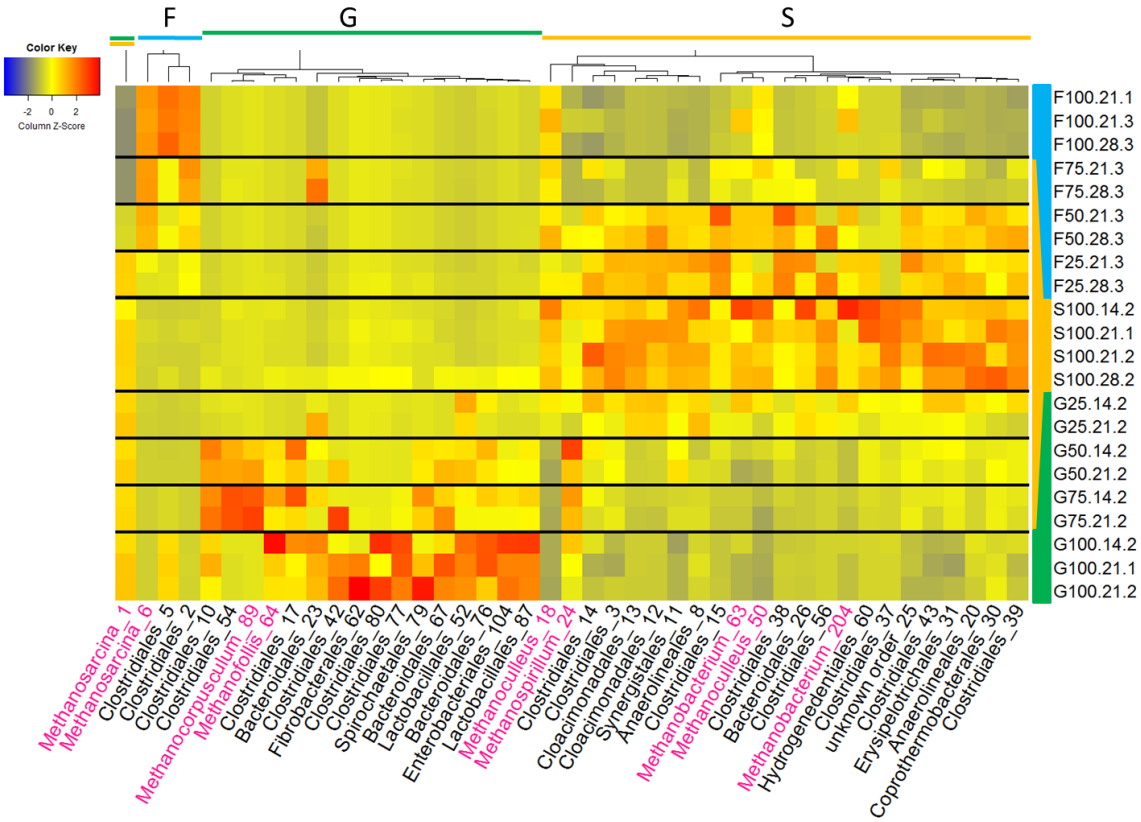


Figure 3. Microbial dynamics over time with respect to the different feeding compositions. Sample plot from the sPCA from the 16S rRNA dataset. Digesters are represented by colours and number of days by symbols.

The individual abundances of the selected microorganisms were compared between the different mixtures (Figure 4). The active archaeal community was different according to the feeding type. Fish mono-digestion (F100) was mostly driven by the archaea *Methanosarcina* OTU_6 and *Methanoculleus* OTU_18; grass mono-digestion (G100) by *Methanospirullum* OTU_24, *Methanosarcina* OTU_1, *Methanofollis* OTU_64 and sludge mono-digestion (S100) by *Methanosarcina* OTU_1, *Methanobacterium* OTU_63 and 204 and two OTUs of *Methanoculleus* (18 and 50). Except *Methanosarcina* which has a versatile methanogenesis metabolism, all the identified archaea use only the hydrogenotrophic pathway to produce methane. Regarding *Methanosarcina*, it may be noteworthy to highlight that fish substrate specifically induced the activity of *Methanosarcina* OTU_6, since it was not found in the other substrates.

The diversity in bacteria community also differed between feeding types. As observed in the alpha diversity analysis, the fish substrate induced a lower microbial diversity than grass or sludge. In fish mono-digestion, the order *Clostridiales* represented more than 90% of the community. In grass mono-digestion, the abundance of *Spirochaetales*, *Fibrobacterales*, *Lactobacillales* and *Enterobacteriales* was higher than in fish and sludge mono-digestion, favoured by their ability to degrade cellulolytic substrates such as grass. Bacterial community of sludge mono-digestion was mostly composed by *Cloacimonadales*, *Synergistales*, *Anaerolineales*, *Hydrogenedentiales*, *Erysipelotrichales* and *Coprothermobateriales*. Members from the orders *Synergistales* and *Anaerolineales* are known or suspected to be able to form syntrophic interaction with hydrogenotrophic methanogens (Ito et al., 2011; Sekiguchi et al., 2001). In sludge, the abundance of potential syntrophic partners in presence with hydrogenotrophic methanogens suggests that methane production was mainly produced from the hydrogenotrophic methanogenesis pathway, in line with the archaeal community described in this incubation.

In light of the literature, the type of substrate used to feed the digesters leads to the development of an adapted microbial community of degraders (De Francisci et al., 2015; Lee et al., 2018). Specifically, in addition to a lower microbial diversity, fish substrate induced the growth of a more specific community than that observed in grass and in sludge substrates. Indeed, only 5% of the OTUs were found in both samples collected from fish- and sludge-digesters, and 9% were found in both samples collected from fish- and grass-digesters, while 17% of the OTUs were common between samples collected from grass- and sludge-digesters.



<u>Figure 4.</u> Heatmap of the most discriminant active microbial community composition. The heatmap shows the abundances of the most discriminant microorganisms selected by the sPCA The substrate used, day, and replicate number are indicated in the row labels.. Duplicates were carried out on the bioreactors containing only fish, grass or sludge at day 21. Taxonomy is indicated at the genus level for archaea (pink) and order level for bacteria (black) completed by the OTU number. Heatmap color goes from blue to red in accordance to the abundance increase.

Within the mixtures, the proportion of the selected microorganisms evolved according to the feeding composition. However, the evolution of the microbial response was not completely linear. For example, the relative abundances of the active OTUs in the fish:sludge mixture at 75:25 (F75) was close to the relative abundances of the active microorganisms in the fish mono-digestion. On the other hand, the more sludge was added in the feeding, more active microorganisms characteristic of sludge were recovered. On the contrary, the active microbial community in grass remained dominant even down to 25% of grass (in the mixture with 75% of sludge, G25). Despite the progressive evolution for most of the microorganisms was observed across all the samples, some microorganisms were only found in specific

conditions such as *Methanocorpusculum* and specific OTUs of *Clostridiales* in the mixes 50:50 and 75:25 of grass:sludge.

Biogas can be regarded as the final outcome of the active microbial composition, and therefore, an association between microorganisms' dynamics and the biogas production performances can be drawn (Figure 2). For example, the similarity of the microbial composition between F100 and F75, or between G25 and S100, induced a similar biogas production. Thus, the bioreactors performances depend on the ability of the microbial community to adapt to the added substrates.

Substrates degradation dynamics

The temporal dynamics of the degradation of the different substrate mixtures in the bioreactors were studied. For that purpose, the metabolic fingerprint of the degradation in the digesters were analysed using an HPLC-ESI-HRMS instrument. After data examination with XCMS, a total of 267 ROIs were detected. To identify the molecules (ROIs) specific to each substrate, and that were initially present, sPLS-DA was performed on the metabolomics data from samples relative to the mono-digestion bioreactors (F100, G100 and S100) at day 0. From this analysis, a total of 70 molecules were determined to be specific of any of the 3 substrates.

The degradation pattern of the selected molecules within the mixtures and over time was evaluated by comparing their intensities (Figure 5). The molecules were grouped according to their initial intensities at day 0. Four major groups can be distinguished from this clustering. 3 of these groups include the ROIs whose relative intensities were significant only in sludge-, grass- or fish-enriched digesters, respectively (S, G, and F, respectively, in Figure 5). The other group (FG in Figure 5) include the ROIs whose relative intensities were high in both fish- and grass- but low in sludge-containing digesters.

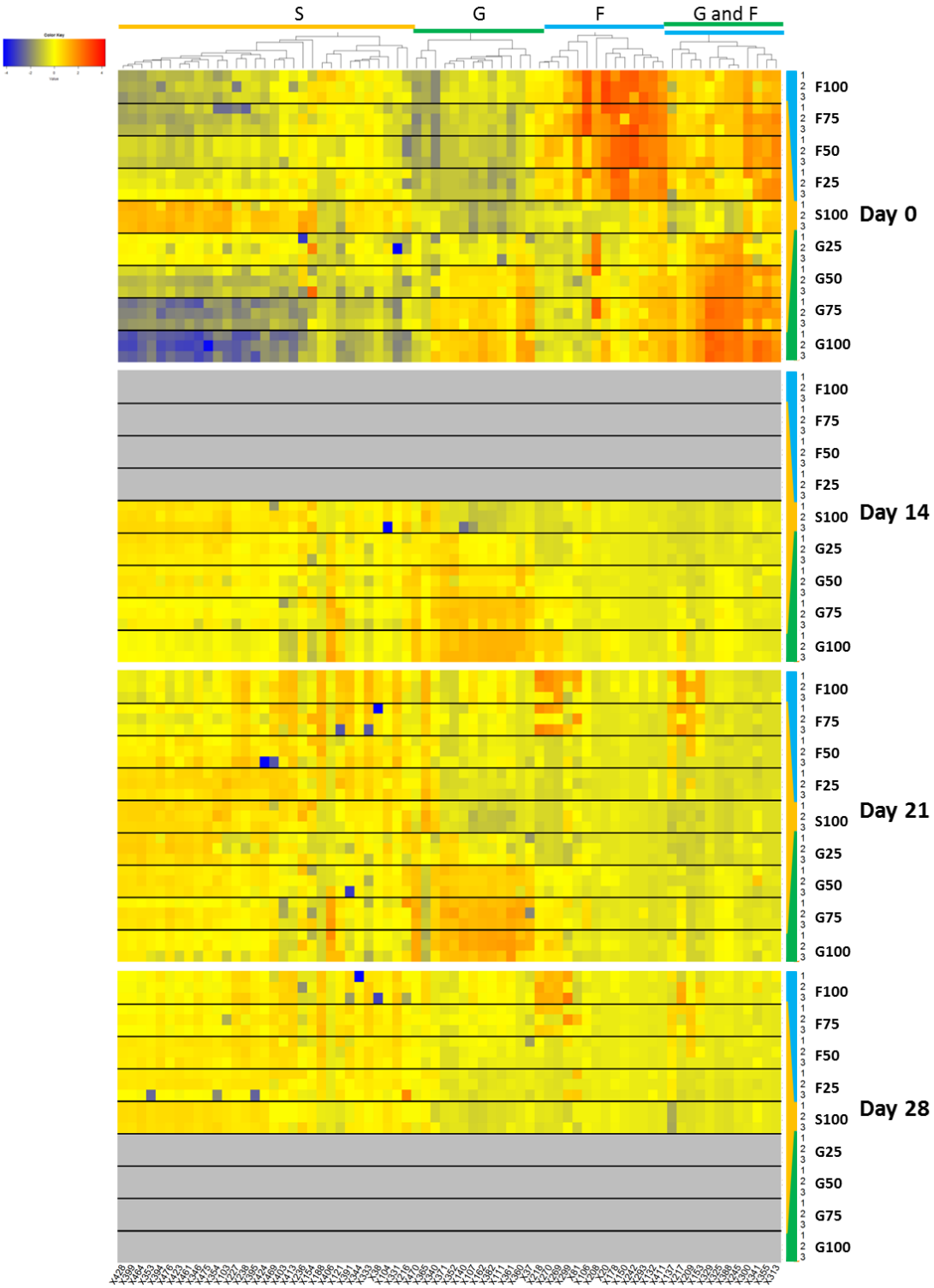


Figure 5. Molecules dynamics within bioreactors and over time. The heatmap shows the evolution among the samples and over time of the intensity of the most discriminant ROIs selected by the sPLS-DA in the different feeding type fish, grass and sludge at day 0. Analyses were carried out at days 21-28 for bioreactors containing fish, 14-21 for bioreactors containing grass and 14-21-28 for bioreactors containing sludge. For each waste mixture and date, triplicates bioreactors called 1, 2 and 3 were analysed. Heatmap color goes from blue to red in accordance with the abundance increase.

342 At day 0, the intensity of the ROIs representative of each feeding type differed within
343 the samples. This result was expected due to the differences in the substrate molecular
344 composition. Moreover, the intensity of the ROIs representative of sludge decreased when
345 fish or grass was mixed with sludge. On the other hand, the intensities of some ROIs
346 representative of fish or grass were similar within all the digesters containing fish or grass.
347 As expected, the intensity of some ROIs decreased over time, while some other either
348 increased or remained stable during the experiment. These molecules with stable intensity
349 across time may not be easily degraded, or they may be a product of degradation of other
350 molecules.
351 The putative identification of the ROIs based on the comparison of their theoretical
352 formula and molecular weight to databases is given in the Table S2. Examples of metabolite
353 biomarkers of sludge are diethylthiophosphate (X340) and 6-methylquinoline (X218 and
354 X270). Diethylthiophosphate is a common degradation product of organophosphorus
355 pesticides, while 6-methylquinoline is a flavouring ingredient found in tea.
356 In grass-fed bioreactors, grass biomarkers were classified as plant constituents (betaine,
357 X245), metabolites obtained in lignin degradation (trans-ferulic acid and p-coumaric acid,
358 X388 and X365, respectively), and metabolites from sugar metabolism (galactitol, X329).
359 In fish bioreactors, most of the metabolites corresponded to organic compounds
360 resulting from amino acids degradation (cadaverine and histamine, X132 and X150,
361 respectively). Therefore, in all cases, the substrate-specific metabolites were biologically
362 consistent with the corresponding substrates, as these metabolites were constituents of the
363 substrates or degradation products from these constituents.
364
365 *Correlation between microbial activity and substrates degradation pattern*

A PLS analysis was performed to integrate the data relative to the microorganism activity with the data of the molecules degradation. The rate of molecules degradation in the bioreactors was estimated by dividing the molecules relative intensity at day 0 by their intensity at days of interest.

The ordination plots from PLS (Figure S2 A to C) show similar patterns for the two datasets, suggesting that an underlying correlation structure between the two datasets exist. The correlation circle plot (Figure S2-D) allows visualizing at the same time the groups of active microorganisms correlated to the molecules degradation rate. To further identify the microorganisms potentially responsible of the molecules degradation, a hierarchical clustering based on the loadings of the microbial and metabolic from the PLS was performed. Five groups of correlated microorganisms and molecules were identified according to their substrate specificity (Tables S2 and S3). Figure 6 depicts the mean values of the microbial activity and molecule degradation rate according to the feeding types. Group 1 included microorganisms and molecules with a high microbial activity and high molecule degradation rate during the digestion of sludge and grass. Groups 2-4 included microorganisms and molecules that are specific of either sludge, grass, or fish, respectively. Finally, group 5 included the microorganisms and molecules that were highly active and highly degraded, respectively, in fish and sludge bioreactors.

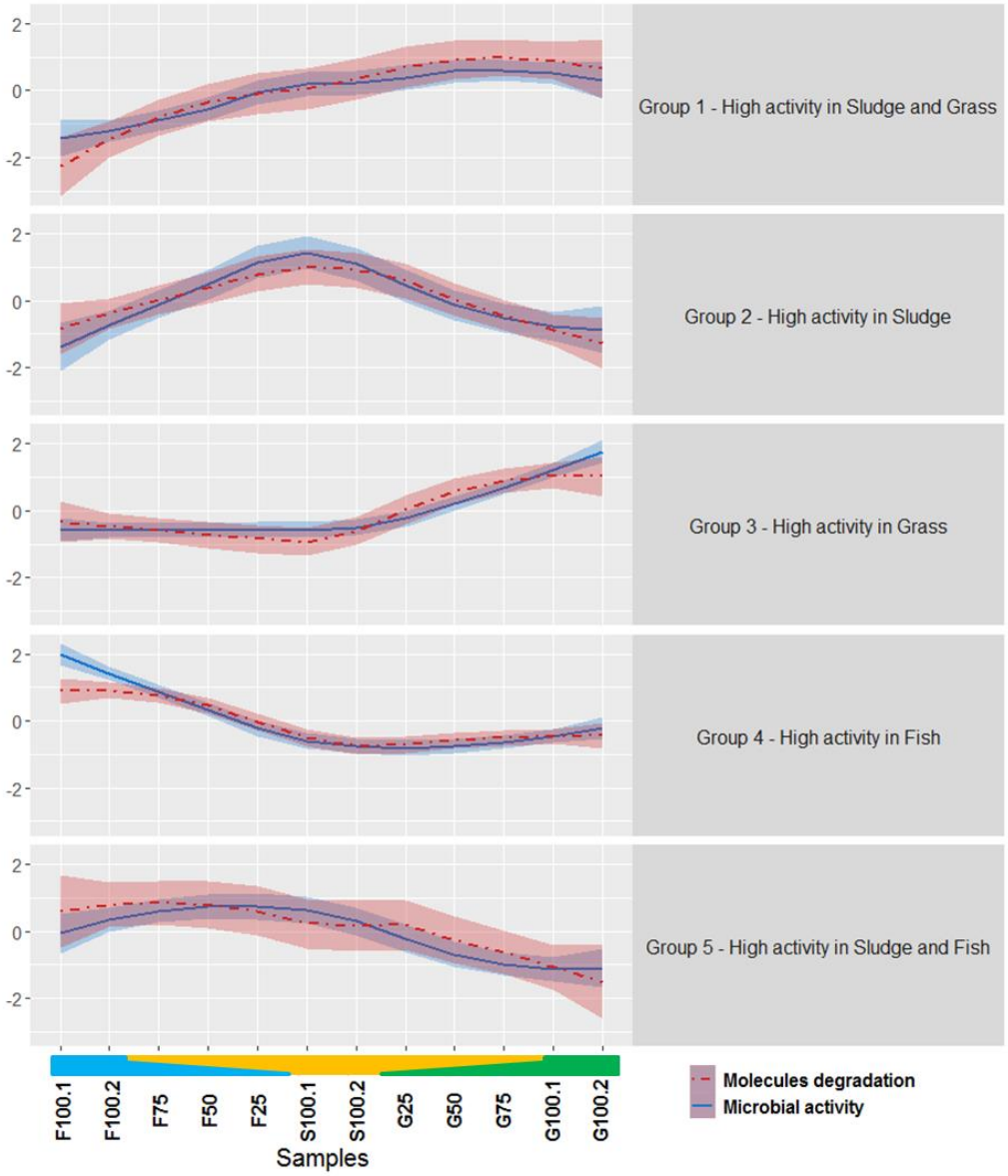


Figure 6. Dynamics of the active microorganisms correlated to the molecules degradation among the samples at day 21. Duplicates were carried out on the bioreactors containing only fish, grass or sludge. Line represents the mean values of the different microbial activity (solid blue line) or molecules degradation rates (dashed red line) included in every clusters, the shadowing traces represent their standard deviation. Duplicates were carried out on the bioreactors containing only fish, grass, or sludge. S100 stands for wastewater sludge alone, F25, F50, F75, F100 stands for respectively 25, 50, 75 or 100% of fish (F) in co-digestion with sludge, G25, G50, G75, G100 stands for respectively 25, 50, 75 or 100% of Grass (G) in co-digestion with sludge.

Group 1 included two genera of archaea, *Methanosarcina* and *Methanospirullum*, and the molecules diethylthiophosphate and N-(3-methylbutyl)acetamide (X153). As stated before, diethylthiophosphate is pesticide degradation product and a urine metabolite (Nomura

et al., 2014). Conversely, N-(3-methylbutyl)acetamide is a metabolite found in alcoholic beverages obtained by the fermentation of vegetal such as beer and wine. One hypothesis explaining the correlation of these archaea and molecules could be due to an indirect role of the archaea in the molecules degradation through a syntrophic interaction with bacteria. However, no further explanation can be done at this stage regarding their correlation, since such association had not been reported before and therefore further investigations are required.

Group 2 correlated 14 OTUs from the orders *Cloacimonadales*, *Clostridiales*, *Anaerolineales*, *Synergistales*, *Bacteroidales*, *Hydrogenedentiales* and *Coprothermobacterales* to 8 molecules including compounds from tryptophan degradation (L-tryptophanol and tryptamine, X217 and X269, respectively), 6-methylquinoline and thioxoacetic acid, among others. In agreement with these metabolites, *Anaerolineales* and *Synergistales* are known for their ability to degrade amino acids (Swiatczak et al., 2017). Thus, their correlation with L-tryptophanol and tryptamine is consistent with the literature. Surprisingly, no methanogen was clustered in this group despite they were expected to be highly correlated with these syntrophic bacteria. One reason could be that the methanogens were not specific partners of these bacteria. Indeed, most of the methanogens were found ubiquitously in bioreactors fed partly with either sludge or grass as shown in the figure 4.

In group 3, 12 OTUs mostly from the orders *Clostridiales*, *Lactobacillales*, *Bacteroidales* and *Spirochaetales* and the archaea *Methanofollis* were correlated to 7 molecules including sugar and lignin compounds, plant constituents, and protein degradation products. From this group, it is worth to highlight the correlation between the microbial activity of the lactic acid bacteria *Lactobacillales* and the degradation of lignin compounds such as trans-ferulic acid (X388) and p-coumaric acid (X365). Indeed, such microorganisms are lignin degraders (Fessard and Remize, 2017; Filannino et al., 2014).

In group 4, *Methanosarcina* and 2 OTUs from the order *Clostridiales* were correlated to molecules that can be classified as amino acids degradation products. Cadaverine (X132) and 5-aminopentanoic acid (X208) are obtained from L-lysine degradation, histamine (X150) from L-histidine degradation, and phenylpyruvic acid (X300) from L-phenylalanine degradation. The presumed role of these microorganisms in the degradation of cadaverine and L-histidine can be supported by previous studies. Roeder and Schink described a new strain close to *Clostridium aminobutyricum,* able to degrade cadaverine, in co-culture with the archaea *Methanospirullum* (Roeder and Schink, 2009). On the other hand, some *Clostridium* were also identified to be involved in the histamine degradation (Pugin et al., 2017).

In Group 5, heptane-1,2,3-triol and hexadecandiperoxoic acid were found to be descriptive of this cluster. The role of these molecules in this cluster is unknown and it must be further investigated. *Methanoculleus* and *Methanobacterium* were correlated to the genus *Syntrophomonas*. Some species of this bacterium are known to growth in syntrophy with H_2-consumer as methanogens (Mcinerney et al., 1981).

In the different groups, there were some ROIs that could not be assigned. In MS metabolomics analysis, the step of *m/z* assignment is a traditional bottleneck (Longnecker et al., 2015). In this study, in addition, the assignment is even more hampered due to the higher structural diversity of the compounds found in digesters, the lack of specific databases, and the absence of precedent literature of metabolomics analysis on anaerobic digesters. Despite all these drawbacks, some correlations between microorganisms and metabolite degradations could still be pointed out. Therefore, if the limitations in the metabolite assignment can be addressed, a better insight of the correlations between microbial and metabolic data will be drawn. .

The statistical method developed in this study allows to posit hypotheses on the degradation of molecules by different microorganisms. These hypotheses were consistent in

448 regards with the literature. However, in order to go further in the interpretation,
449 complementary analyses must be performed. Firstly, the molecules identification needs to be
450 reinforced by using MS/MS fragmentation and by comparison of the acquired MS/MS spectra
451 with the spectra of standards. Secondly, the ultimate proof of the degradation of the molecules
452 by the identified microorganisms can be obtained by performing microbial cultures using
453 these molecules as substrates, or by performing experiments with labelled molecules (stable
454 isotope probing, (Chapleur et al., 2016)). Finally, a much broader metabolic coverage can be
455 obtained by using different extraction protocols, by combining different instrumental
456 implementations (i.e., by analysing the samples using different ionisation modes), by using
457 different chromatographic columns (i.e., reverse phase and HILIC), and even by using
458 different high-throughput metabolomics techniques (i.e, LC-MS, GC-MS, and NMR
459 spectroscopy).
460

461 **Conclusion**

462 This study demonstrates the existence of links between the anaerobic digester feeding
463 composition and the microbiota development. Our method allowed the extraction of the
464 correlation patterns between the microorganisms' activity and the degradation of molecules,
465 characteristic of each substrate. We identified a subset of active microorganisms highly
466 correlated with molecules degradation patterns. The highlighted microbial and metabolic
467 correlations were biologically relevant and consistent with previous literature. The
468 development of new omics methodologies and associated databases focused on anaerobic
469 digesters metabolic composition will open new approaches to study and improve the
470 functioning mechanisms of these bioprocesses.
471

Disclosures

The authors declare no competing financial interest

Funding

This work was supported by the National Research Agency (grant number NR-16-CE05-0014) as part of the Digestomic project. Kim Anh Lê Cao, Olivier Chapleur and Laëtitia Cardona scientific travels were supported in part by the France-Australia Science Innovation Collaboration (FASIC) Program Early Career Fellowships from the Australian Academy of Science (grant number 39417TM). Kim Anh Lê Cao was supported in part by the National Health and Medical Research Council (NHMRC) Career Development fellowship (grant number GNT1159458).

Acknowledgements

We thank Nadine Derlet from the Irstea PROSE analytical division for their technical support. We acknowledge SUEZ Environment for providing us the access to the wastewater treatment plant of Valenton.

References

Amha, Y.M., Anwar, M.Z., Brower, A., Jacobsen, C.S., Stadler, L.B., Webster, T.M., Smith, A.L., 2018. Inhibition of anaerobic digestion processes: Applications of molecular tools. Bioresour. Technol. 247, 999–1014. https://doi.org/10.1016/j.biortech.2017.08.210

Beale, D.J., Karpe, A. V., McLeod, J.D., Gondalia, S. V., Muster, T.H., Othman, M.Z., Palombo, E.A., Joshi, D., 2016. An "omics" approach towards the characterisation of laboratory scale anaerobic digesters treating municipal sewage sludge. Water Res. https://doi.org/10.1016/j.watres.2015.10.029

Bize, A., Cardona, L., Desmond-Le Quéméner, E., Battimelli, A., Badalato, N., Bureau, C., Madigou, C., Chevret, D., Guillot, A., Monnet, V., Godon, J.J., Bouchez, T., 2015. Shotgun metaproteomic profiling of biomimetic anaerobic digestion processes treating

500 sewage sludge. Proteomics 15, 3532–3543. https://doi.org/10.1002/pmic.201500041
501 Bouhlel, J., Jouan-Rimbaud Bouveresse, D., Abouelkaram, S., Baéza, E., Jondreville, C.,
502 Travel, A., Ratel, J., Engel, E., Rutledge, D.N., 2018. Comparison of common
503 components analysis with principal components analysis and independent components
504 analysis: Application to SPME-GC-MS volatolomic signatures. Talanta 178, 854–863.
505 https://doi.org/10.1016/j.talanta.2017.10.025
506 Cai, M., Wilkins, D., Chen, J., Ng, S.K., Lu, H., Jia, Y., Lee, P.K.H., 2016. Metagenomic
507 reconstruction of key anaerobic digestion pathways in municipal sludge and industrial
508 wastewater biogas-producing systems. Front. Microbiol. 7.
509 https://doi.org/10.3389/fmicb.2016.00778
510 Callahan, B.J., Sankaran, K., Fukuyama, J.A., McMurdie, P.J., Holmes, S.P., 2016.
511 Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community
512 analyses. F1000Research 5, 1492. https://doi.org/10.12688/f1000research.8986.2
513 Cardona, L., Levrard, C., Guenne, A., Chapleur, O., Mazéas, L., 2019. Co-digestion of
514 wastewater sludge: choosing the optimal blend. Chem. Eng. J. 87, 772–781.
515 https://doi.org/10.1016/j.wasman.2019.03.016
516 Chapleur, O., Mazeas, L., Godon, J.J., Bouchez, T., 2016. Asymmetrical response of
517 anaerobic digestion microbiota to temperature changes. Appl. Microbiol. Biotechnol.
518 100, 1445–1457. https://doi.org/10.1007/s00253-015-7046-7
519 De Francisci, D., Kougias, P.G., Treu, L., Campanaro, S., Angelidaki, I., 2015. Microbial
520 diversity and dynamicity of biogas reactors due to radical changes of feedstock
521 composition. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2014.10.126
522 De Vrieze, J., Pinto, A.J., Sloan, W.T., Boon, N., Ijaz, U.Z., 2018. The active microbial
523 community more accurately reflects the anaerobic digestion process : 16S rRNA (gene)
524 sequencing as a predictive tool. Microbiome 1–13.
525 https://doi.org/doi.org/10.1186/s40168-018-0449-9
526 Escudié, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., Maman, S.,
527 Hernandez-Raquet, G., Combes, S., Pascal, G., 2018. FROGS: Find, Rapidly, OTUs
528 with Galaxy Solution. Bioinformatics 34, 1287–1294.
529 Fahy, E., Sud, M., Cotter, D., Subramaniam, S., 2007. LIPID MAPS online tools for lipid
530 research. Nucleic Acids Res. 35, 606–612. https://doi.org/10.1093/nar/gkm324
531 Fessard, A., Remize, F., 2017. Why Are Weissella spp. Not Used as Commercial Starter
532 Cultures for Food Fermentation? Fermentation 3, 38.
533 https://doi.org/10.3390/fermentation3030038

534 Filannino, P., Gobbetti, M., Angelis, M. De, Cagno, R. Di, 2014. Hydroxycinnamic Acids
535 Used as External Acceptors of Electrons : an Energetic Advantage for Strictly
536 Heterofermentative Lactic Acid Bacteria 80, 7574–7582.
537 https://doi.org/10.1128/AEM.02413-14
538 Hassa, J., Maus, I., Off, S., Pühler, A., Scherer, P., Klocke, M., Schlüter, A., 2018.
539 Metagenome, metatranscriptome, and metaproteome approaches unraveled compositions
540 and functional relationships of microbial communities residing in biogas plants. Appl.
541 Microbiol. Biotechnol. 102, 5045–5063. https://doi.org/10.1007/s00253-018-8976-7
542 Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K.,
543 Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F.,
544 Sawada, Y., Hirai, M.Y., Nakanishi, H., Ikeda, K., Akimoto, N., Maoka, T., Takahashi,
545 H., Ara, T., Sakurai, N., Suzuki, H., Shibata, D., Neumann, S., Iida, T., Tanaka, K.,
546 Funatsu, K., Matsuura, F., Soga, T., Taguchi, R., Saito, K., Nishioka, T., 2010.
547 MassBank: A public repository for sharing mass spectral data for life sciences. J. Mass
548 Spectrom. 45, 703–714. https://doi.org/10.1002/jms.1777
549 Ito, T., Yoshiguchi, K., Ariesyady, H.D., Okabe, S., 2011. Identification of a novel acetate-
550 utilizing bacterium belonging to Synergistes group 4 in anaerobic digester sludge. ISME
551 J. 5, 1844–1856. https://doi.org/10.1038/ismej.2011.59
552 Lê Cao, K.-A., Boitard, S., Besse, P., 2011. Sparse PLS discriminant analysis: biologically
553 relevant feature selection and graphical displays for multiclass problems. BMC
554 Bioinformatics 12, 253. https://doi.org/10.1186/1471-2105-12-253
555 Lê Cao, K.-A., Martin, P.G., Robert-Granié, C., Besse, P., 2009. Sparse canonical methods
556 for biological data integration: application to a cross-platform study. BMC
557 Bioinformatics 10, 34. https://doi.org/10.1186/1471-2105-10-34
558 Lee, J., Kim, E., Han, G., Tongco, J.V., Shin, S.G., Hwang, S., 2018. Microbial communities
559 underpinning mesophilic anaerobic digesters treating food wastewater or sewage sludge:
560 A full-scale study. Bioresour. Technol. 259, 388–397.
561 https://doi.org/10.1016/j.biortech.2018.03.052
562 Li, N., He, J., Yan, H., Chen, S., Dai, X., 2017. Pathways in bacterial and archaeal
563 communities dictated by ammonium stress in a high solid anaerobic digester with
564 dewatered sludge. Bioresour. Technol. 241, 95–102.
565 https://doi.org/10.1016/j.biortech.2017.05.094
566 Lin, Q., De Vrieze, J., Li, J., Li, X., 2016. Temperature affects microbial abundance, activity
567 and interactions in anaerobic digestion. Bioresour. Technol. 209, 228–236.

https://doi.org/10.1016/j.biortech.2016.02.132

Longnecker, K., Futrelle, J., Coburn, E., Kido, M.C., Kujawinski, E.B., 2015. Environmental metabolomics : Databases and tools for data analysis. Mar. Chem. 177, 366–373. https://doi.org/10.1016/j.marchem.2015.06.012

Madigou, C., Lê Cao, K.-A., Bureau, C., Mazéas, L., Déjean, S., Chapleur, O., 2019. Ecological consequences of abrupt temperature changes in anaerobic digesters. Chem. Eng. J. 361, 266–277. https://doi.org/10.1016/J.CEJ.2018.12.003

Mcinerney, M.J., Bryant, M.P., Hespell, R.B., Costerton, J.W., 1981. Syntrophomonas wolfei gen. nov. sp. nov., an Anaerobic, Syntrophic, Fatty Acid-Oxidizing Bacterium 41, 1029–1039.

Noll, M., Klose, M., Conrad, R., 2010. Effect of temperature change on the composition of the bacterial and archaeal community potentially involved in the turnover of acetate and propionate in methanogenic rice field soil. FEMS Microbiol. Ecol. 73, 215–25. https://doi.org/10.1111/j.1574-6941.2010.00883.x

Nomura, H., Ueyama, J., Sugiura, Y., Takaishi, A., Hayashi, Y., Kondo, T., Inuzuka, K., Kamijima, M., Ogi, H., Osaka, A., Inoue, M., Wakusawa, S., Saito, I., 2014. A revised method for determination of dialkylphosphate levels in human urine by solid-phase extraction and liquid chromatography with tandem mass spectrometry: application to human urine samples from Japanese children. Environ. Health Prev. Med. 19, 405–413. https://doi.org/10.1007/s12199-014-0407-5

Poirier, S., Bize, A., Bureau, C., Bouchez, T., Chapleur, O., 2016. Community shifts within anaerobic digestion microbiota facing phenol inhibition: Towards early warning microbial indicators? Water Res. 100, 296–305. https://doi.org/10.1016/j.watres.2016.05.041

Pugin, B., O'Mahony, L., Westermann, P., Hellings, P., Wawrzyniak, M., Heider, A., Akdis, C.A., Barcik, W., 2017. A wide diversity of bacteria from the human gut produces and degrades biogenic amines. Microb. Ecol. Health Dis. 28, 1353881. https://doi.org/10.1080/16512235.2017.1353881

Roeder, J., Schink, B., 2009. Syntrophic degradation of cadaverine by a defined methanogenic coculture. Appl. Environ. Microbiol. 75, 4821–4828. https://doi.org/10.1128/AEM.00342-09

Rohart, F., Gautier, B., Singh, A., Cao, K.-A. Le, 2017. mixOmics: an R package for 'omics feature selection and multiple data integration. bioRxiv 108597. https://doi.org/10.1101/108597

602 Sekiguchi, Y., Takahashi, H., Kamagata, Y., Ohashi, A., Harada, H., 2001. In Situ Detection,
603 Isolation, and Physiological Properties of a Thin Filamentous Microorganism Abundant
604 in Methanogenic Granular Sludges: a Novel Isolate Affiliated with a Clone Cluster, the
605 Green Non-Sulfur Bacteria, Subdivision I. Appl. Environ. Microbiol. 67, 5740–5749.
606 https://doi.org/10.1128/AEM.67.12.5740
607 Shen, H., Huang, J.Z., 2008. Sparse principal component analysis via regularized low rank
608 matrix approximation. J. Multivar. Anal. 99, 1015–1034.
609 https://doi.org/10.1016/j.jmva.2007.06.007
610 Singh, A., Benoît, G., Shannon, C.P., Vacher, M., Rohart, F., Tebbutt, S.J., Lê Cao, K.-A.,
611 2019. DIABLO: an integrative, multi-omics, multivariate method for multi-group
612 classification. Bioinformatics 1–50.
613 https://doi.org/https://doi.org/10.1093/bioinformatics/bty1054
614 Smith, C.A., Want, E.J., Maille, G.O., Abagyan, R., Siuzdak, G., 2006. XCMS : Processing
615 Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment ,
616 Matching , and Identification. Anal. chemestry 78, 779–787.
617 Sousa, D.Z., Salvador, A.F., Ramos, J., Guedes, A.P., Barbosa, S., Stams, A.J.M., Alves,
618 M.M., Pereira, M.A., 2013. Activity and viability of methanogens in anaerobic digestion
619 of unsaturated and saturated long-chain fatty acids. Appl. Environ. Microbiol. 79, 4239–
620 4245. https://doi.org/10.1128/AEM.00035-13
621 Swiatczak, P., Cydzik-Kwiatkowska, A., Rusanowska, P., 2017. Microbiota of anaerobic
622 digesters in a full-scale wastewater treatment plant. Arch. Environ. Prot. 43, 53–60.
623 https://doi.org/10.1515/aep-2017-0033
624 Vanwonterghem, I., Jensen, P.D., Ho, D.P., Batstone, D.J., Tyson, G.W., 2014. Linking
625 microbial community structure, interactions and function in anaerobic digesters using
626 new molecular techniques. Curr. Opin. Biotechnol. 27, 55–64.
627 https://doi.org/10.1016/j.copbio.2013.11.004
628 Villas-Bôas, S.G., Noel, S., Lane, G.A., Attwood, G., Cookson, A., 2006. Extracellular
629 metabolomics: A metabolic footprinting approach to assess fiber degradation in complex
630 media. Anal. Biochem. 349, 297–305. https://doi.org/10.1016/j.ab.2005.11.019
631 Wishart, D.S., Feunang, Y.D., Marcu, A., Guo, A.C., Liang, K., Vázquez-Fresno, R., Sajed,
632 T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M.,
633 Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., Serra-Cayuela, A., Liu, Y.,
634 Mandal, R., Neveu, V., Pon, A., Knox, C., Wilson, M., Manach, C., Scalbert, A., 2018.
635 HMDB 4.0: The human metabolome database for 2018. Nucleic Acids Res. 46, D608–

636 D617. https://doi.org/10.1093/nar/gkx1089
637 Zamanzadeh, M., Hagen, L.H., Svensson, K., Linjordet, R., Horn, S.J., 2017. Biogas
638 production from food waste via co-digestion and digestion- effects on performance and
639 microbial ecology. Sci. Rep. 7, 1–12. https://doi.org/10.1038/s41598-017-15784-w
640
641