Title

Integrative analyses to investigate the link between microbial activity and metabolites 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, School of Mathematics and Statistics, The University of Melbourne, Parkville VIC 3010, 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

olivier.chapleur@irstea.fr

https://orcid.org/0000-0001-9460-921X

Abstract

Anaerobic digestion (AD) is a promising biological process which converts waste into sustainable energy. To fully exploit AD's capability, we need to deepen our knowledge of the microbiota involved in this complex bioprocess. High-throughput methodologies open new perspectives to investigate AD process at the molecular level, supported by recent data integration methodologies to extract relevant information. In this study, we investigated the link between microbial activity and substrate degradation in a lab-scale anaerobic co-digestion experiment, where bioreactors were fed with 9 different mixtures of three co-substrates (fish

waste, sewage sludge, and grass). Samples were profiled using 16S rRNA sequencing and untargeted metabolomics. In this article, we propose a suite of multivariate tools to statistically integrate these data and identify coordinated patterns between groups of microbial and metabolic profiles specific of each co-substrate. Five main groups of features were successfully evidenced, including cadaverine degradation found to be associated with the activity of microorganisms from the order *Clostridiales* and the genus *Methanosarcina*. 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; Partial Least Squares regression

Introduction

Deciphering the microbial communities in diverse domains, such as health, food safety and environment, has been widely addressed by the development of high-throughput technologies. In particular, Anaerobic Digestion (AD) bioprocess is constituted by extremely complex microbial communities, mainly composed of bacteria and archaea, with high functional redundancies. While AD is a promising bioprocess for the production of biogas, a sustainable energy, from the degradation of organic waste, the key microorganisms driving AD remain largely unknown. AD bioprocess involves four principal steps: hydrolysis, acidogenesis, acetogenesis and methanogenesis. The optimal functioning of AD relies on the efficient network between the microbes carrying successively these steps. However, the microbial diversity and interactions between microorganisms are highly sensitive and depend on multiple technical parameters such as temperature (Madigou et al., 2019; Noll et al., 2010), presence of inhibitors (Li et al., 2017; Poirier et al., 2016; Sousa et al., 2013), or feedstock composition (Zamanzadeh et al., 2017).

Emerging omics and high-throughput approaches can monitor microbial complexity across multiple functional levels (Vanwonterghem et al., 2014). Metagenomics, metatranscriptomics, metaproteomics and metabolomics provide necessary information related to a community's genes, gene expression, proteins and metabolite production. They reveal potential and activated functions of microbial communities. In our context, these omics can unravel the intricate networks of functional processes of AD, especially when comparing bioreactors under different operational conditions to elucidate their functioning (Carballa et al., 2015; De Vrieze and Verstraete, 2016; Lü 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 analysis and visualisation of these omics results (Bouhlel et al., 2018; Callahan et al., 2016; Rohart et al., 2017; Singh et al., 2019). As the statistical analysis of single omics is not sufficient to decipher complex microbial relationships, the combination of information from several data sources has become mandatory. Computational analytical methods have a rising potential to capitalise on this rich data but they are still at their infancy and are not broadly used for this type of problems. Classical multivariate methods, including Principal Component Analysis (PCA) or Principal Coordinate Analysis (PCoA), have been widely used for single omics datasets, such as 16S or metabolomic data (Poirier et al., 2016; Wang et al., 2017). However, such methods are exploratory and do not fully characterised the microbial communities playing a role in the biological systems being studied. Moreover, these methods are not currently used to integrate information from different omics experiments. Novel computational and statistical methods are becoming available to fully harvest the large amount of data generated in multi-omics experiments (Lê Cao et al., 2009; Singh et al., 2019; Straube et al., 2017). These methods aim to extract complementary information from several

datasets for a holistic understanding of the interplay between the different levels of information.

Several AD studies have already shown the utility of using omics technologies to decipher the anaerobic microbial population (Amha et al., 2018; Hassa et al., 2018). Analyses of a single omics are routinely carried out (Bize et al., 2015; Cai et al., 2016) but the use of several paired omics on the same samples, and their statistical integration are rare currently. 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) but no data integration was conducted between the two approaches. However, correlation between transcriptomics and metabolomics has already been used to identify important metabolites and genes for example in fungal or plants (Askenazi et al., 2003; Hoefgen and Nikiforova, 2008).

Our study is the first of its kind in the anaerobic bioprocess field to investigate the link between microbial activity from 16S rRNA sequencing data and patterns of substrate degradation from metabolomics data. We propose a statistical framework to integrate these datasets to identify associations, or correlations, between microbial activity and metabolites specific of substrates degradation. To that end, 27 anaerobic bioreactors were fed with binary mixtures of three substrates of different chemical composition (sludge and grass, or sludge and fish) at different proportions. The microbial community was analysed through the sequencing of the 16S rRNA-to characterise metabolically active microorganisms (De Vrieze et al., 2018). The patterns of the substrate degradation were studied with untargeted metabolomics using LC-MS, to determine the molecular fingerprints of waste degradation (Villas-Bôas et al., 2006). Our statistical analyses revealed subsets of active microorganisms highly associated to dynamics of substrate degradation over time, and enabled to posit novel hypotheses regarding the capacity of microorganisms to degrade specific metabolites.

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 during 10 days in order to degrade the organic matter in excess before carrying 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 INRAE institute. Fish waste and garden grass were crushed and kept at 4°C before carrying out the experiments.

Bioreactors experimental set-up

Binary mixtures of sludge with linearly decreasing percentages of either fish or grass were prepared (0-100, 25-75, 50-50, 75-25, 100-0 respectively, see 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 bioreactors, and the ratio of substrate/inoculum used to feed and inoculate all the bioreactors was fixed at 12 gCOD/1.2 gCOD (Table S1). All 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 flushed with N_2 (purity >99.99%, Linde gas SA). In total 27 anaerobic bioreactors were set-up.

In every reactor, 6 mL of liquid phase were sampled through the septum using a syringe at a weekly frequency (days 0, 14, 21, 28). The collected samples were centrifuged at 10 000g for 10 minutes to separate the supernatants from the pellets. Supernatants and pellets were snap frozen using liquid nitrogen. Supernatants were kept at -20°C for metabolomics analysis and pellets kept at -80°C for microbial analysis.

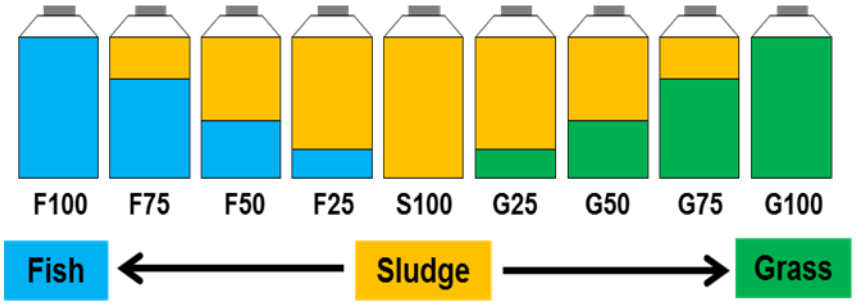


Figure 1. Scheme of the batch experimental design. S100 stands for bioreactors fed with wastewater sludge alone, F25, F50, F75, F100 stands for bioreactors fed with 25, 50, 75 or 100% of fish (F) respectively, in co-digestion with sludge. G25, G50, G75, G100 stands for bioreactors fed with 25, 50, 75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

RNA extraction and 16S rRNA sequencing

Based on the biogas production, a total of 22 samples were selected (corresponding to the highest biogas production as indicated in Figure 2). Since fish-fed bioreactors showed a higher delay than grass-fed bioreactors before biogas production started, the sampling time-points used were different for the two substrates (days 14 and 21 for grass-fed bioreactors, and days 21 and 28 for fish-fed bioreactors, Figure 2).

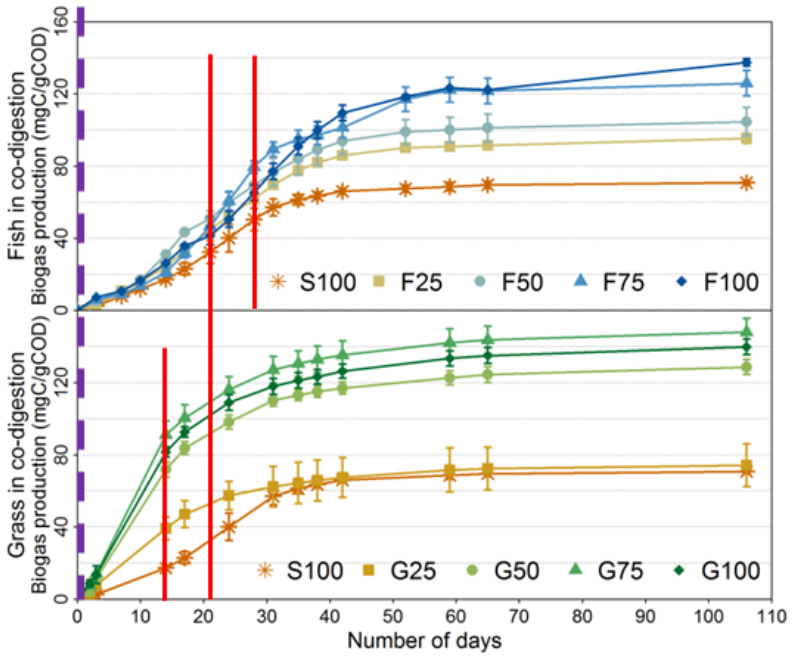


Figure 2. Cumulated biogas production (mgC/gCOD) over time (Days) for the different bioreactors. Mean values of the triplicate bioreactors are indicated with error bars representing standard deviations within triplicates. Co-digestion percentage is described in Figure 1, with either Fish (top) or Grass (bottom) co-digestion. Red solid lines indicate 16S rRNA sequencing and metabolomics sampling dates. Purple dashed line corresponds to the start of the incubation where an additional metabolomics 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 then removed. 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 2200 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). The raw sequences were deposited in the NCBI databases under the project accession number PRJNA562430.

FROGS (Find Rapidly OTU with Galaxy Solution), a galaxy/CLI workflow (Escudié et al., 2018), was used to generate an OTU count matrix. R CRAN software (version 3.5.1) was used for alpha diversity analysis with Shannon method using phyloseq R package (version 1.20.0). Proportions of archaeal and bacterial OTUs abundances were scaled per sample and OTUs that exceeded 1% in terms of relative abundance in at least one sample, were selected for the remainder of the analysis and square-root transformed.

Metabolomics analysis

Metabolomics 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 MilliQ 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^{TM} C18 column (50x2.1 mm, 1.7µm, Thermo Scientific) during 23 minutes, followed by a

stabilisation 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 of 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 (region of interest - ROIs) with a *m/z* error of 10 ppm and a peakwidth between 20 and 50 seconds. The ROIs found in different samples were grouped using the group method with a bandwidth of 30. 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 and data integration

Multivariate data analysis was carried out using the mixOmics R package, using a suite of component-based methods to identify groups of microorganisms and metabolites degradation with coordinated patterns.

First, each dataset was analysed independently using an exploratory (unsupervised) approach, or a supervised approach to identify key OTUs and metabolites prior to data integration as follows.

Unsupervised analysis on the 16S data. Sparse Principal Component Analysis (sPCA, (Shen and Huang, 2008)) was applied on all the samples of the microbial dataset (16S rRNA, 22 samples) to evidence main sources of variation and identify OTUs that most contribute to each principal component. Briefly, sPCA is based on PCA that includes a LASSO penalty for variable selection. In our context, the main source of variation correspond to specific feeding conditions (see Figure 3), and thus variable selection from sPCA enables us to identify microorganisms characteristic of co-substrate.

Supervised analysis on the metabolomic day 0 data. Sparse Partial Least Squares Discriminant Analysis (sPLS-DA, (Lê Cao et al., 2011)), a supervised classification method, was applied on the mono-digestion samples (S100, F100, G100, 9 samples in total) at day 0 in order to determine the molecules specific to each substrate before any degradation. PLS-DA classifies samples into groups (sludge, fish, grass) and the sparse variant identifies metabolites discriminating each co-substrate at the start of the experiment. This method has already been used in metabolomics data analysis and successful in removing noise by selecting the most informative variables (e.g. Jiang et al., 2014).-Classification performance was assessed based on the mean classification error rate using leave-one-out cross-validation, and Mahalanobis distance for prediction of the class of the samples (Rohart et al., 2017).

Clustering. Once selected, the abundances of the OTUs and the metabolites selected with sPCA and sPLS-DA respectively were clustered using hierarchical clustering (using Ward's method and Manhattan and Euclidean distances respectively for each data type), then represented in heatmaps (heatmap.2 function from gplots R package, version 3.0.1).

Data integration of the selected OTUs and metabolites. The aim of this analysis is to identify patterns of different variables (microbes and metabolites) characteristics of the different conditions. Thus, this analysis included all 22 samples representative of the different feeding conditions. We calculated the metabolite degradation rate data as the ratio between the metabolites abundances at days 14, 21 or 28, relative to day 0. Both datasets were then integrated using Partial Least Square (PLS) regression method with a canonical mode (Lê Cao et al., 2009) to identify groups of OTUs and degraded metabolites that are highly correlated. Briefly, PLS weights variables from both datasets optimally using loading vectors, so that their linear combination (called PLS components) is maximally correlated and thus extract similar patterns across datasets (see supplemental figure S2 A and B). Loading vectors correspond to the weight assigned to each variable to define the PLS components. We subsequently applied hierarchical clustering on the loadings vectors to identify clusters of metabolites and OTUs variables with similar patterns of degradation rate or activity across the different types of substrate mixtures. As the correlation structure between high-throughput datasets can be highly spurious, we then further assessed the strength of the association between the variables assigned to each cluster using the proportionality distance (R package *propr* (Quinn et al., 2017)). We compared the distance between the profiles of each variable within one cluster and to the distance between the profiles outside this cluster. A small distance between two variables indicates a strong association.

Results and discussion

Influence of the feeding composition on the biogas production

We examined the cumulated biogas production in the different mixtures. Three main results can be observed from Figure 2. Firstly, the sludge mono-digestion produced a lower biogas quantity compared to the fish and grass mono-digestion. Secondly, the grass mono-

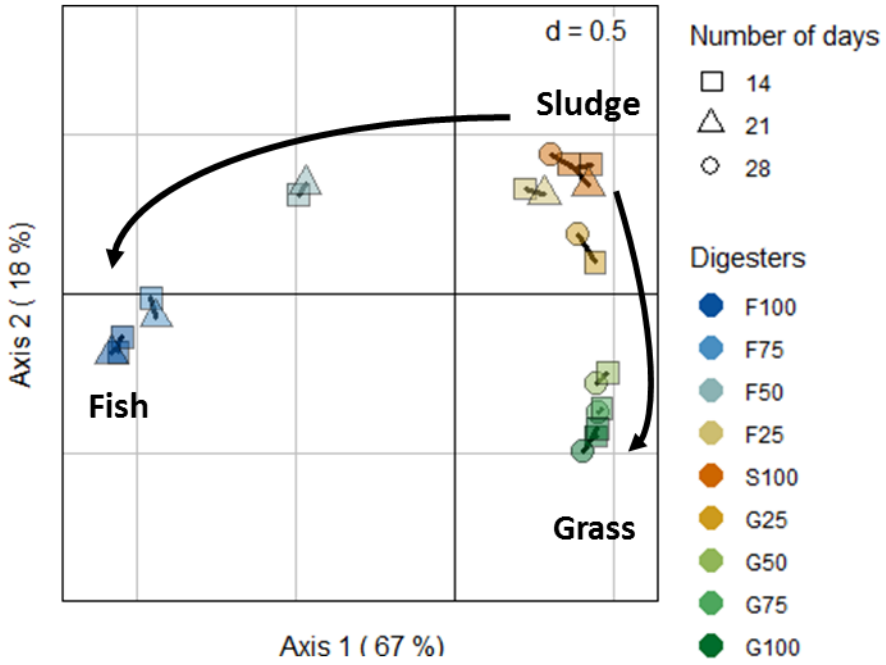
digestion provided the best biogas production performances, as the biogas production rate and the final biogas production were higher than in fish mono-digestion. Thirdly, compared to sludge mono-digestion, the biogas production in bioreactors with mixtures of waste presented two patterns of improvement according to the type of co-substrate (fish or grass). In co-digestion with fish, the biogas production from the sludge improved linearly with the increase of fish quantity in the feedstock composition. In co-digestion with grass, the final biogas production was improved only when at least 50% of grass was present in the feedstock composition. A comparison of the major chemical parameters of anaerobic digestion has been already described in a precedent study (Cardona et al., 2019). In the present study, we focus on the influence of the feedstock composition on microbial activity and its relation with degradation of specific metabolites.

Influence of the feeding composition on the microbial community

The influence of the feeding composition on the microbial diversity was evaluated using the Shannon index calculated on archaeal and bacterial communities (supplemental Figure S1). The bacterial diversity was inversely proportional to the amount of fish present in the bioreactors, while the level of the bacterial diversity remained stable when grass was present in co-digestion with sludge. The archaeal diversity decreased sharply with the presence of fish in the feedstock, or with high amount of grass. The low microbial diversity induced by the presence of fish can be explained by the composition of a simpler substrate (in terms of molecular variety) compared to the other substrates, or by a lower functional redundancy of the microorganisms that can degrade fish substrate compared to those that can degrade grass or sewage sludge.

Sparse PCA highlighted the main sources of variability in the 16S rRNA data set, while identifying key OTUs driving this variation. We selected 43 OTUs, which in combination explained 67 (first component) to 85% (two components) of the total microbial community

289 variance in the data. The sample plot in Figure 3 showed the strong influence of the feeding
290 composition on the microbial community. The samples were regrouped according to the
291 major co-substrate, and regardless of the sample collection time. This result suggests that the
292 variance due to time within conditions is low, and active microbial community over time is
293 relatively stable.



294

295 **Figure 3. Microbial dynamics over time with respect to the different feeding**
296 **compositions.** Sample plot from the sPCA from the 16S rRNA dataset. Bioreactors are
297 represented by colours and number of days by symbols.
298

299 The relative abundance of the selected microorganisms was compared between the
300 different mixtures (Figure 4). We observed differences in the active archaeal community
301 depending on the feeding type. Fish mono-digestion (F100) was mostly driven by the archaea
302 *Methanosarcina* OTU_6 and *Methanoculleus* OTU_18; grass mono-digestion (G100) by
303 *Methanospirullum* OTU_24, *Methanosarcina* OTU_1, *Methanofollis* OTU_64 and sludge
304 mono-digestion (S100) by *Methanosarcina* OTU_1, *Methanobacterium* OTU_63 and 204 and
305 two OTUs of *Methanoculleus* (18 and 50). With the exception of *Methanosarcina* which can

use different methanogenesis pathways, all the identified archaea only use the hydrogenotrophic pathway to produce methane. The archaea composition is likely to depend on the environmental conditions in the bioreactors induced by the different feeding conditions and from bacterial dynamics. The fish substrate induced the activity of a specific OTU of *Methanosarcina* that was not found in the other substrates. As *Methanosarcina* presents a versatile methanogenesis metabolism, the modification of the OTU between the substrate could reflect the use of different methanogenesis metabolism.

The diversity in the 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*, grouping cellulolytic degraders and fermentative members, represented more than 90% of the bacterial community while the abundance reached 25% and 50% for respectively sludge and grass mono-digestion. In grass mono-digestion, the abundance of *Spirochaetales*, *Fibrobacterales, Lactobacillales* and *Enterobacteriales* was higher than in fish and sludge mono-digestion. They were probably 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*, *Cloacimonadales*, *Coprothermobacteriaceae* and *Anaerolineales* are known or suspected to be able to form syntrophic interaction with hydrogenotrophic methanogens (Calusinska et al., 2018; Gagliano et al., 2015; Ito et al., 2011; Sekiguchi et al., 2001). *Hydrogenedentiales* are H_2-utilising bacteria and associations with the H_2-producers *Coprothermobacteriaceae* were evidenced (Nobu et al., 2015). 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 current literature, the type of substrate used to feed the bioreactors has been suggested to contribute to the development of an adapted microbial community of degraders (De Francisci et al., 2015; Lee et al., 2018). Specifically, in our study, fish substrate was found to induce not only a low microbial diversity, but also the growth of a specific community compared to what was observed in grass and in sludge substrates. Indeed, only 5% of the selected OTUs were common between fish- and sludge-bioreactors, 9% between both fish- and grass-bioreactors, while 17% of the selected OTUs were common between both grass- and sludge-bioreactors.

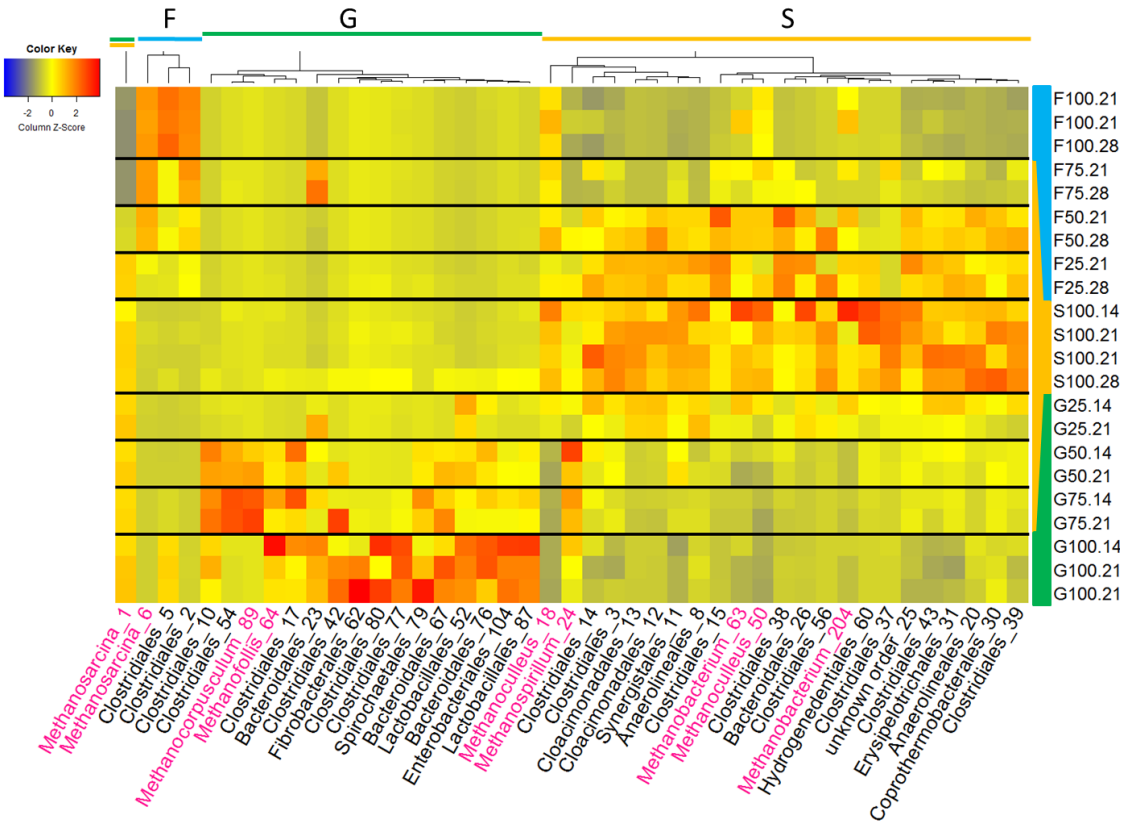


Figure 4. Heatmap of the microorganisms whose abundance contributes most to variance due to feeding composition selected with sPCA. The type of substrate used, day, duplicates were carried out on the bioreactors containing only fish, grass or sludge at day 21.

Taxonomy completed by the OTU number is indicated at the genus level for archaea (pink) and at the order level for bacteria (black). Abundance is represented in the heatmap ranging from blue (low) to high (red).

Within the mixtures, the abundance of the selected microorganisms changed with feeding composition. However, microbial activity did not necessarily increase with the amount of mixture proportions. For example, the relative abundances of the OTUs identified as active in the fish:sludge mixture at 75:25 (F75) were similar to the relative abundances of the same active OTUs in the fish mono-digestion (F100). As the proportion of sludge increased, the active microorganisms more characteristic of sludge increased as well. However, in grass mixtures, the active microbial community from sludge remained dominant regardless the proportion of grass used (e.g down to 25% in the mixture with 75% of sludge, G25). However, some microorganisms were found in specific proportions of grass:sludge. For example, *Methanocorpusculum* and specific OTUs of *Clostridiales* were found in the mixes 50:50 and 75:25 of grass:sludge.

Substrates degradation dynamics

We studied the temporal dynamics of degradation of the different substrate mixtures by analysing the metabolic fingerprint of the degradation in the bioreactors using an HPLC-ESI-HRMS instrument. After data examination with XCMS, a total of 267 regions of interest (ROI) were detected. ROI designate chromatographic features (m/z) in the spectrum. Sparse PLS-DA identified 70 ROIs out of the 267 ROIs specific to each substrate (sludge, grass, fish) initially present at day 0. The classification error rate from the sPLS-DA was 0, indicating that all samples could be perfectly assigned to their respective groups using leave-one-out cross validation based on these 70 ROIs. The ROIs were then clustered into four groups according to their intensities at day 0. The degradation pattern of the selected ROIs was then examined for every mixture across all time points. Four main degradation patterns were observed

(Figure 5). Three of these groups include the ROIs whose relative intensities were high only in sludge-, grass- or fish-fed bioreactors (indicated as S, G, and F, respectively, in Figure 5) at day 0. The other group (G and F) include the ROIs whose relative intensities were high in both fish- and grass- but low in sludge-containing bioreactors at day 0.

Whilst the ROIs were selected to discriminate only mono-digestion at day 0, we observed that the intensity of the ROIs differed across all feeding types. This result was expected given the differences in the molecular composition of the substrates. In general, the intensity of the ROIs representative of sludge decreased when fish or grass was mixed with sludge. We also observed some temporal dynamics, as that the intensity of some ROIs decreased over time, while some other either increased or remained stable during the experiment. This can be explained as some metabolites may not be easily degraded, or may be a product of degradation of other metabolites.

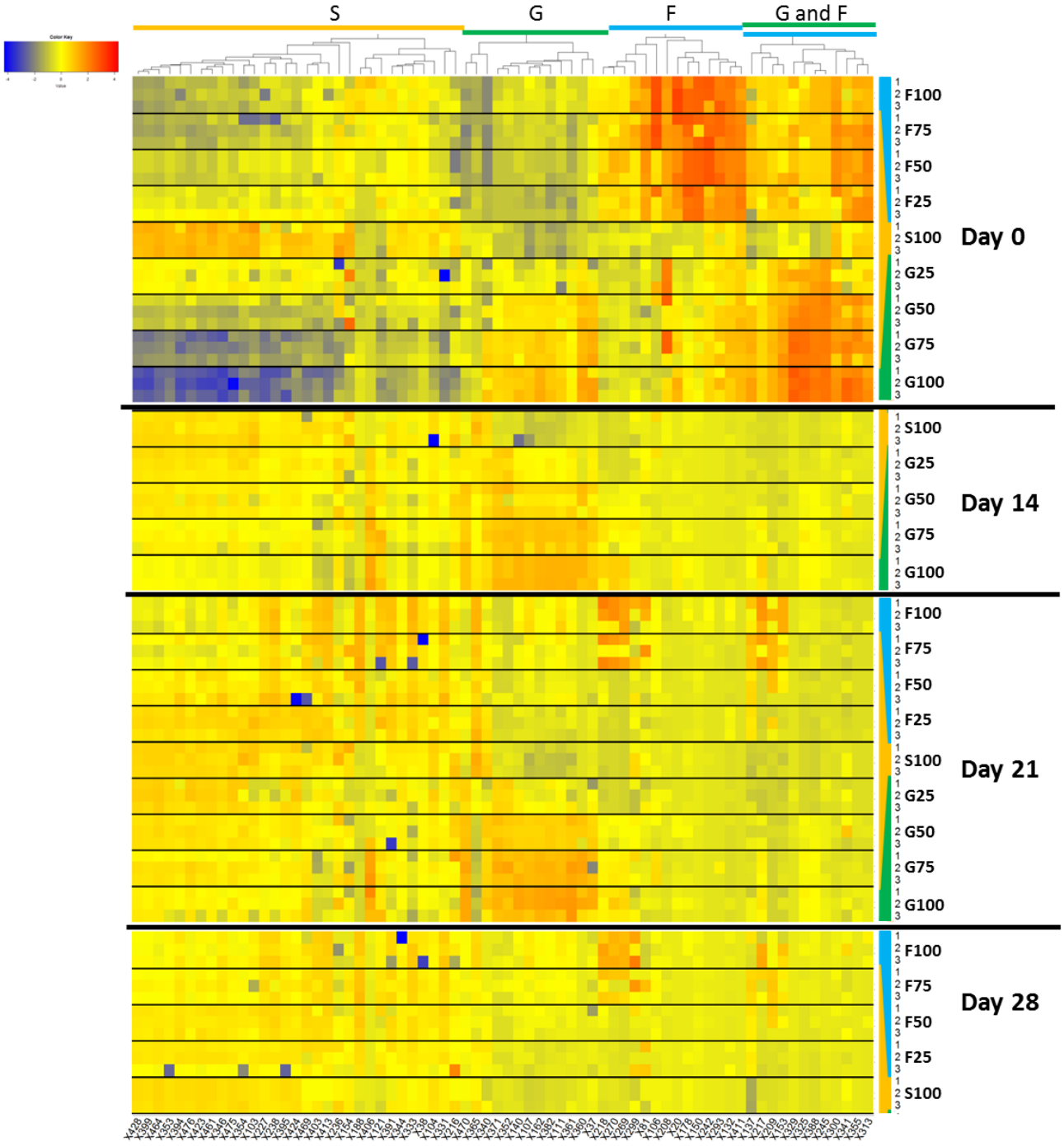


Figure 5. Metabolites dynamics within bioreactors and over time. The most discriminant ROIs for the different co-substrates were selected by sPLS-DA at day 0, then clustered with hierarchical clustering according to day 0. The profiles for the remaining days were then represented in the heatmap based on the intensities. The different panels represent 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 numbered 1, 2 and 3 are indicated. Intensity ranges from blue (low) to red (high).

394 The putative identification of the ROIs based on their *m/z* is given in the Table S2. Out
395 of the 70 selected ROIs, 19 metabolites were identified, whilst the rest could not be assigned.
396 The metabolites assignment step in MS metabolomics analysis is a traditional bottleneck
397 (Longnecker et al., 2015). Our study was further hampered by the high structural diversity of
398 the compounds found in bioreactors, the lack of databases specific for anaerobic digesters,
399 and the absence of relevant literature of metabolomics analysis on anaerobic bioreactors.
400 However, the metabolites identified in the different clusters were biologically consistent with
401 the corresponding substrates. For example, metabolites identified in sludge-containing
402 samples were diethylthiophosphate (X340) and 6-methylquinoline (X218 and X270).
403 Diethylthiophosphate is a common degradation product of organophosphorus pesticides,
404 while 6-methylquinoline is a flavouring ingredient found in tea. Metabolites characteristics of
405 grass-fed bioreactors were plant constituents (betaine, X245), metabolites obtained from
406 lignin degradation (trans-ferulic acid and p-coumaric acid, X388 and X365, respectively), and
407 metabolites from sugar metabolism (galactitol, X329). In fish-fed bioreactors, most of the
408 metabolites corresponded to organic compounds resulting from amino acids degradation
409 (cadaverine and histamine, X132 and X150, respectively).
410
411 *Correlation between microbial activity and substrates degradation pattern*

412 Statistical data integration analysis was carried out between the microbial and
413 metabolites datasets to identify potential degraders of specific metabolites. As far as we know,
414 this type of analysis has never been applied in the anaerobic digestion field. Correlation
415 studies have been used to relate metabolite profiles to microbiome in gut studies (Han et al.,
416 2019; McHardy et al., 2013). For example, Han et al (2019) estimated the correlation between
417 cecal compounds and bacteria profiles acquired respectively by LC/MS and 16S RNA
418 sequencing analyses using Spearman's correlation. However, in datasets containing a large

number of variables, such approach is highly likely to result in spurious correlations. One novelty in our study is to use component-based approaches, such as multivariate PLS method that are well suited for a high correlation structure within and between datasets, whilst being able to highlight key associations between variables from each dataset.

As described in the methods section, we considered the metabolites degradation rates relative to day 0 in our analysis. The sample (ordination) plots from PLS (Figure S2 A-C) show similar patterns for microbial and metabolites datasets, suggesting that an underlying correlation structure between the two datasets exists and can be extract with such method. The correlation circle plot (Figure S2 D) projects the variables into the space spanned by the PLS components to visualise simultaneously the active microorganisms and the metabolites responsible for this ordination of the samples (González et al., 2012). In particular, active microorganisms and metabolites with a high degradation rate projected close to each other on this graph are correlated, as they contribute equally to define the PLS components. A hierarchical clustering was performed on the PLS loadings of the microbial and metabolites variables. It helped to further identify groups of microorganisms potentially associated with the metabolites degradation. In total, five clusters were evidenced.

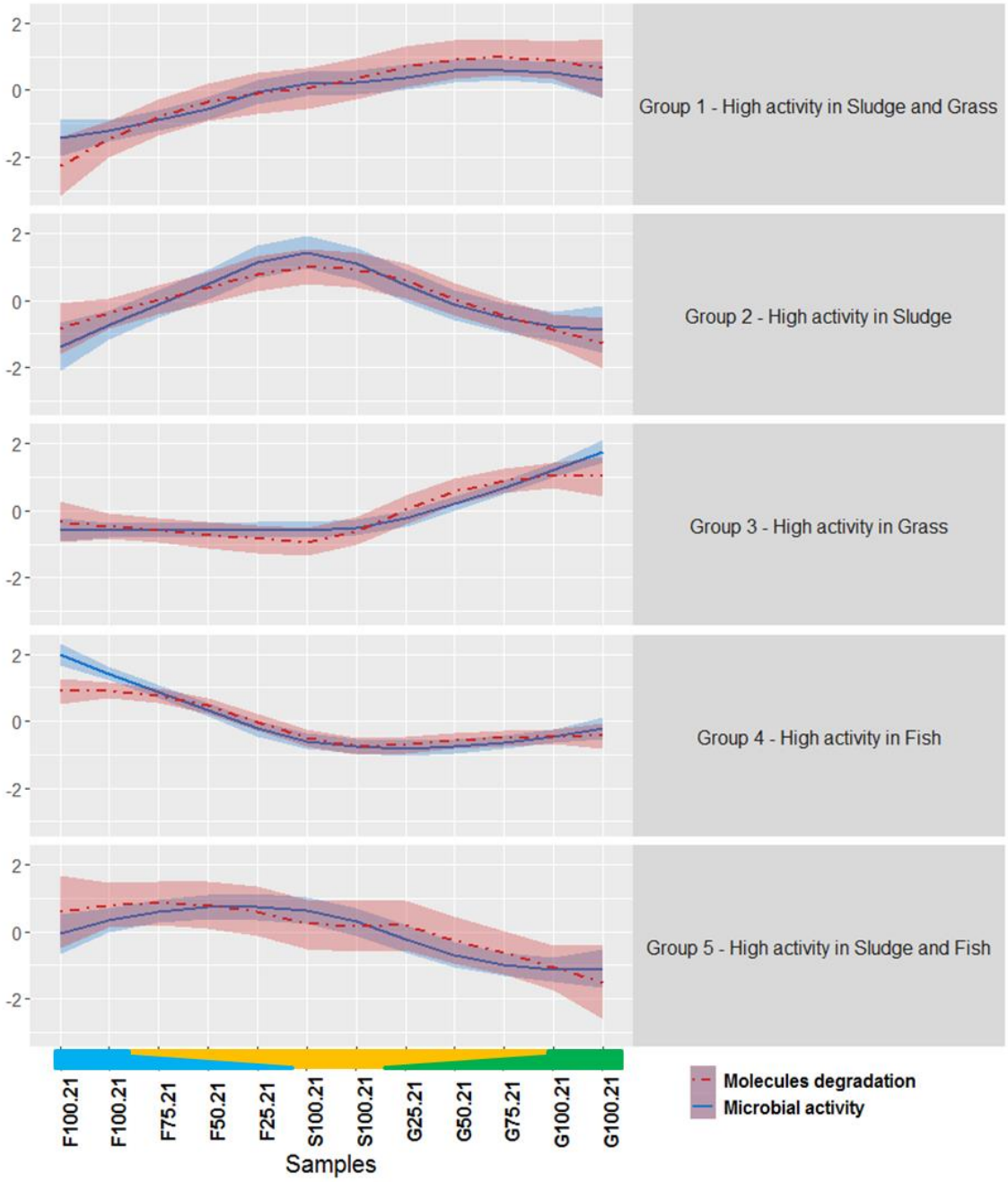


Figure 6. Correlated dynamics of the active microorganisms and metabolites degradation across all samples at day 21. Lines represent the mean values of the different microbial activity (solid blue line) or metabolites degradation rates (dashed red line) for each of the clusters identified using hierarchical clustering performed on the PLS loadings. Shadows represent standard deviation. 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.

To illustrate the associations, the Figure 6 depicts the mean values of the microbial activity and metabolite degradation rate according to the feeding types at day 21. Microorganisms and metabolites showing the same evolution in the different feeding mixtures are clustered together. Similar trends were observed at the other dates (see supplemental figure S4). The quality of the clustering was further assessed by comparing the proportionality distances between profiles (Bodein et al., 2019; Quinn et al., 2017). The distances were lower between profiles grouped within the same cluster, compared to the distances with the profiles assigned to other clusters (supplemental Figure S3 and Table S3). This result confirmed a strong association between the profiles of the active microorganisms and metabolites across all feeding types, and their links where further investigated.

The microorganisms and metabolites assigned to the five clusters are described in the supplemental Tables S3 and S4. Group 1 included microorganisms and metabolites with a high microbial activity and high metabolite degradation rate during the digestion of sludge and grass. Groups 2 to 4 included microorganisms and metabolites that were specific of either sludge, grass, or fish, respectively. Finally, group 5 included the microorganisms and metabolites that were highly active and highly degraded, respectively, in fish and sludge bioreactors.

Group 1 included two genera of archaea, *Methanosarcina* and *Methanospirullum*, and the metabolites diethylthiophosphate and N-(3-methylbutyl)acetamide (X153). As mentioned before, diethylthiophosphate is a pesticide degradation product found in urine (Nomura et al., 2014) while acetamides are produced in the fermentation of vegetal matter. One hypothesis explaining the association between these archaea and metabolites could be an indirect role of the archaea in the metabolites degradation through a syntrophic interaction with bacteria. This is the first time such association is reported, and further investigations are required.

Group 2 included 14 OTUs from the orders *Cloacimonadales*, *Clostridiales*, *Anaerolineales*, *Synergistales*, *Bacteroidales*, *Hydrogenedentiales* and *Coprothermobacterales*, and 8 metabolites 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 identified in this group as we would have expected to find 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 (groups 2 and 3) as shown in Figure 4.

In group 3, 12 OTUs mostly from the orders *Clostridiales*, *Lactobacillales*, *Bacteroidales* and *Spirochaetales* and the archaea *Methanofollis* were found associated with 7 metabolites from sugar degradation, (galactitol), lignin compounds (p-coumaric, trans-ferulic acid), plant constituent (betaine) or histidine degradation (Imidazolepropionic acid). From this group, it is worth noting the association 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* and metabolites classified as amino acids degradation products were identified. 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

494 *Clostridium aminobutyricum*, able to degrade cadaverine, in co-culture with the archaea
495 *Methanospirullum* (Roeder and Schink, 2009). Some *Clostridium* were also identified to be
496 involved in the histamine degradation (Pugin et al., 2017).

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

502 Thus, the refinement of our clusters has highlighted potentially relevant associations
503 between molecules degradation and microbial activity across different feeding substrates. This
504 is the first time that such extensive data integration analysis is performed in anaerobic
505 digestion. Previous multi-omics studies only considered single omics analysis (Beale et al.,
506 2016; Bize et al., 2015; Joyce et al., 2018).

507 Our analytical workflow suggests relationships between the degradation of metabolites
508 by different microorganisms. These relationships were consistent with the literature and future
509 work will be carried out to validate them. For instance, the degradation of the metabolites by
510 the identified microorganisms could be confirmed with either microbial cultures using these
511 specific metabolites as substrates, or by performing experiments with labelled metabolites,
512 such as stable isotope probing (Chapleur et al., 2016).

513

514 **Conclusion**

515 This study proposed a suite of statistical tools to unravel associations between
516 molecules degradation and microbial activity, whilst carefully preventing spurious
517 correlations between them. We applied component-based approaches for variable selection
518 and data integration, then further evaluated the clusters of variables using proportionality

519 distances. Our analyses show the existence of links between the anaerobic digester feeding
520 composition and microbiota activity. The associations identified between the microorganisms'
521 activity and the degradation of metabolites were biologically relevant and consistent with
522 previous literature. The statistical framework we developed in this study are promising for the
523 the screening analysis of potential metabolites degraders in different complex ecosystems
524 such as other bioprocesses, or human microbiota.
525

526 **Disclosures**

527 The authors declare no competing financial interest

528

529 **Funding**

530 This work was supported by the National Research Agency (grant number ANR-16-
531 CE05-0014) as part of the Digestomic project. Kim Anh Lê Cao, Olivier Chapleur and
532 Laëtitia Cardona scientific travels were supported in part by the France-Australia Science
533 Innovation Collaboration (FASIC) Program Early Career Fellowships from the Australian
534 Academy of Science (grant number 39417TM). Kim Anh Lê Cao was supported in part by the
535 National Health and Medical Research Council (NHMRC) Career Development fellowship
536 (grant number GNT1159458).
537

538 **Acknowledgements**

539 We thank Nadine Derlet from the Irstea PROSE analytical division for her technical
540 support. We acknowledge SUEZ Environment for providing us the access to the wastewater
541 treatment plant of Valenton.
542

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

Askenazi, M., Driggers, E.M., Holtzman, D.A., Norman, T.C., Iverson, S., Zimmer, D.P., Boers, M.-E., Blomquist, P.R., Martinez, E.J., Monreal, A.W., Feibelman, T.P., Mayorga, M.E., Maxon, M.E., Sykes, K., Tobin, J.V., Cordero, E., Salama, S.R., Trueheart, J., Royer, J.C., Madden, K.T., 2003. Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains. Nat. Biotechnol. 21, 150–156. https://doi.org/10.1038/nbt781

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 sewage sludge. Proteomics 15, 3532–3543. https://doi.org/10.1002/pmic.201500041

Bodein, A., Chapleur, O., Droit, A., Lê Cao, K.A., 2019. A generic multivariate framework for the integration of microbiome longitudinal studies with other data types. bioRxiv 34, 585802.

Bouhlel, J., Jouan-Rimbaud Bouveresse, D., Abouelkaram, S., Baéza, E., Jondreville, C., Travel, A., Ratel, J., Engel, E., Rutledge, D.N., 2018. Comparison of common components analysis with principal components analysis and independent components analysis: Application to SPME-GC-MS volatolomic signatures. Talanta 178, 854–863.

https://doi.org/10.1016/j.talanta.2017.10.025

Cai, M., Wilkins, D., Chen, J., Ng, S.K., Lu, H., Jia, Y., Lee, P.K.H., 2016. Metagenomic reconstruction of key anaerobic digestion pathways in municipal sludge and industrial wastewater biogas-producing systems. Front. Microbiol. 7. https://doi.org/10.3389/fmicb.2016.00778

Callahan, B.J., Sankaran, K., Fukuyama, J.A., McMurdie, P.J., Holmes, S.P., 2016. Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. F1000Research 5, 1492. https://doi.org/10.12688/f1000research.8986.2

Calusinska, M., Goux, X., Fossépré, M., Muller, E.E.L., Wilmes, P., Delfosse, P., 2018. A year of monitoring 20 mesophilic full‐scale bioreactors reveals the existence of stable but different core microbiomes in bio‐waste and wastewater anaerobic digestion systems. Biotechnol. Biofuels 11. https://doi.org/https://doi.org/10.1186/s13068-018-1195-8

Carballa, M., Regueiro, L., Lema, J.M., 2015. Microbial management of anaerobic digestion: Exploiting the microbiome-functionality nexus. Curr. Opin. Biotechnol. 33, 103–111. https://doi.org/10.1016/j.copbio.2015.01.008

Cardona, L., Levrard, C., Guenne, A., Chapleur, O., Mazéas, L., 2019. Co-digestion of wastewater sludge: choosing the optimal blend. Chem. Eng. J. 87, 772–781. https://doi.org/10.1016/j.wasman.2019.03.016

Chapleur, O., Mazeas, L., Godon, J.J., Bouchez, T., 2016. Asymmetrical response of anaerobic digestion microbiota to temperature changes. Appl. Microbiol. Biotechnol. 100, 1445–1457. https://doi.org/10.1007/s00253-015-7046-7

De Francisci, D., Kougias, P.G., Treu, L., Campanaro, S., Angelidaki, I., 2015. Microbial diversity and dynamicity of biogas reactors due to radical changes of feedstock composition. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2014.10.126

De Vrieze, J., Pinto, A.J., Sloan, W.T., Boon, N., Ijaz, U.Z., 2018. The active microbial community more accurately reflects the anaerobic digestion process : 16S rRNA (gene) sequencing as a predictive tool. Microbiome 1–13. https://doi.org/doi.org/10.1186/s40168-018-0449-9

De Vrieze, J., Verstraete, W., 2016. Perspectives for microbial community composition in anaerobic digestion: from abundance and activity to connectivity. Environ. Microbiol. 18, 2797–2809. https://doi.org/10.1111/1462-2920.13437

Escudié, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., Maman, S., Hernandez-Raquet, G., Combes, S., Pascal, G., 2018. FROGS: Find, Rapidly, OTUs with Galaxy Solution. Bioinformatics 34, 1287–1294.

Fahy, E., Sud, M., Cotter, D., Subramaniam, S., 2007. LIPID MAPS online tools for lipid research. Nucleic Acids Res. 35, 606–612. https://doi.org/10.1093/nar/gkm324

Fessard, A., Remize, F., 2017. Why Are Weissella spp. Not Used as Commercial Starter Cultures for Food Fermentation? Fermentation 3, 38. https://doi.org/10.3390/fermentation3030038

Filannino, P., Gobbetti, M., Angelis, M. De, Cagno, R. Di, 2014. Hydroxycinnamic Acids Used as External Acceptors of Electrons : an Energetic Advantage for Strictly Heterofermentative Lactic Acid Bacteria 80, 7574–7582. https://doi.org/10.1128/AEM.02413-14

Gagliano, M.C., Braguglia, C.M., Petruccioli, M., Rossetti, S., 2015. Ecology and biotechnological potential of the thermophilic fermentative Coprothermobacter spp. FEMS Microbiol. Ecol. 91, 1–12. https://doi.org/10.1093/femsec/fiv018

González, I., Lê Cao, K.-A., Davis, M.J., Déjean, S., 2012. Visualising associations between paired ‘omics’ data sets. BioData Min. 5, 19. https://doi.org/10.1186/1756-0381-5-19

Han, J., Meng, J., Chen, S., Li, C., 2019. Integrative analysis of the gut microbiota and

618 metabolome in rats treated with rice straw biochar by 16S rRNA gene sequencing and
619 LC/MS-based metabolomics. Sci. Rep. 9, 1–9. https://doi.org/10.1038/s41598-019-
620 54467-6
621 Hassa, J., Maus, I., Off, S., Pühler, A., Scherer, P., Klocke, M., Schlüter, A., 2018.
622 Metagenome, metatranscriptome, and metaproteome approaches unraveled compositions
623 and functional relationships of microbial communities residing in biogas plants. Appl.
624 Microbiol. Biotechnol. 102, 5045–5063. https://doi.org/10.1007/s00253-018-8976-7
625 Hoefgen, R., Nikiforova, V.J., 2008. Metabolomics integrated with transcriptomics: Assessing
626 systems response to sulfur-deficiency stress. Physiol. Plant. 132, 190–198.
627 https://doi.org/10.1111/j.1399-3054.2007.01012.x
628 Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, Kenichi,
629 Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F.,
630 Sawada, Y., Hirai, M.Y., Nakanishi, H., Ikeda, K., Akimoto, N., Maoka, T., Takahashi,
631 H., Ara, T., Sakurai, N., Suzuki, H., Shibata, D., Neumann, S., Iida, T., Tanaka, Ken,
632 Funatsu, K., Matsuura, F., Soga, T., Taguchi, R., Saito, K., Nishioka, T., 2010.
633 MassBank: A public repository for sharing mass spectral data for life sciences. J. Mass
634 Spectrom. 45, 703–714. https://doi.org/10.1002/jms.1777
635 Ito, T., Yoshiguchi, K., Ariesyady, H.D., Okabe, S., 2011. Identification of a novel acetate-
636 utilizing bacterium belonging to Synergistes group 4 in anaerobic digester sludge. ISME
637 J. 5, 1844–1856. https://doi.org/10.1038/ismej.2011.59
638 Jiang, M., Wang, C., Zhang, Y., Feng, Y., Wang, Y., Zhu, Y., 2014. Sparse Partial-least-
639 squares Discriminant Analysis for Different Geographical Origins of Salvia miltiorrhiza
640 by 1 H-NMR-based Metabolomics. Phytochem. Anal. 25, 50–58.
641 https://doi.org/10.1002/pca.2461
642 Joyce, A., Ijaz, U.Z., Nzeteu, C., Vaughan, A., Shirran, S.L., Botting, C.H., Quince, C.,

643 O'Flaherty, V., Abram, F., 2018. Linking microbial community structure and function
644 during the acidified anaerobic digestion of grass. Front. Microbiol. 9, 1–13.
645 https://doi.org/10.3389/fmicb.2018.00540
646 Lê Cao, K.-A., Boitard, S., Besse, P., 2011. Sparse PLS discriminant analysis: biologically
647 relevant feature selection and graphical displays for multiclass problems. BMC
648 Bioinformatics 12, 253. https://doi.org/10.1186/1471-2105-12-253
649 Lê Cao, K.-A., Martin, P.G., Robert-Granié, C., Besse, P., 2009. Sparse canonical methods
650 for biological data integration: application to a cross-platform study. BMC
651 Bioinformatics 10, 34. https://doi.org/10.1186/1471-2105-10-34
652 Lee, J., Kim, E., Han, G., Tongco, J.V., Shin, S.G., Hwang, S., 2018. Microbial communities
653 underpinning mesophilic anaerobic digesters treating food wastewater or sewage sludge:
654 A full-scale study. Bioresour. Technol. 259, 388–397.
655 https://doi.org/10.1016/j.biortech.2018.03.052
656 Li, N., He, J., Yan, H., Chen, S., Dai, X., 2017. Pathways in bacterial and archaeal
657 communities dictated by ammonium stress in a high solid anaerobic digester with
658 dewatered sludge. Bioresour. Technol. 241, 95–102.
659 https://doi.org/10.1016/j.biortech.2017.05.094
660 Longnecker, K., Futrelle, J., Coburn, E., Kido, M.C., Kujawinski, E.B., 2015. Environmental
661 metabolomics : Databases and tools for data analysis. Mar. Chem. 177, 366–373.
662 https://doi.org/10.1016/j.marchem.2015.06.012
663 Lü, F., Bize, A., Guillot, A., Monnet, V., Madigou, C., Chapleur, O., Mazéas, L., He, P.,
664 Bouchez, T., 2014. Metaproteomics of cellulose methanisation under thermophilic
665 conditions reveals a surprisingly high proteolytic activity. ISME J. 8, 88–102.
666 https://doi.org/10.1038/ismej.2013.120
667 Madigou, C., Lê Cao, K.-A., Bureau, C., Mazéas, L., Déjean, S., Chapleur, O., 2019.

668 Ecological consequences of abrupt temperature changes in anaerobic digesters. Chem.
669 Eng. J. 361, 266–277. https://doi.org/10.1016/J.CEJ.2018.12.003
670 McHardy, I.H., Goudarzi, M., Tong, M., Ruegger, P.M., Schwager, E., Weger, J.R., Graeber,
671 T.G., Sonnenburg, J.L., Horvath, S., Huttenhower, C., McGovern, D.P., Fornace, A.J.,
672 Borneman, J., Braun, J., 2013. Integrative analysis of the microbiome and metabolome of
673 the human intestinal mucosal surface reveals exquisite inter-relationships. Microbiome 1,
674 17. https://doi.org/10.1186/2049-2618-1-17
675 Mcinerney, M.J., Bryant, M.P., Hespell, R.B., Costerton, J.W., 1981. Syntrophomonas wolfei
676 gen. nov. sp. nov., an Anaerobic, Syntrophic, Fatty Acid-Oxidizing Bacterium 41, 1029–
677 1039.
678 Nobu, M.K., Narihiro, T., Rinke, C., Kamagata, Y., Tringe, S.G., Woyke, T., Liu, W., 2015.
679 Microbial dark matter ecogenomics reveals complex synergistic networks in a
680 methanogenic bioreactor. ISME J. 9, 1710–1722.
681 https://doi.org/https://doi.org/10.1038/ismej.2014.256
682 Noll, M., Klose, M., Conrad, R., 2010. Effect of temperature change on the composition of
683 the bacterial and archaeal community potentially involved in the turnover of acetate and
684 propionate in methanogenic rice field soil. FEMS Microbiol. Ecol. 73, 215–25.
685 https://doi.org/10.1111/j.1574-6941.2010.00883.x
686 Nomura, H., Ueyama, J., Sugiura, Y., Takaishi, A., Hayashi, Y., Kondo, T., Inuzuka, K.,
687 Kamijima, M., Ogi, H., Osaka, A., Inoue, M., Wakusawa, S., Saito, I., 2014. A revised
688 method for determination of dialkylphosphate levels in human urine by solid-phase
689 extraction and liquid chromatography with tandem mass spectrometry: application to
690 human urine samples from Japanese children. Environ. Health Prev. Med. 19, 405–413.
691 https://doi.org/10.1007/s12199-014-0407-5
692 Poirier, S., Bize, A., Bureau, C., Bouchez, T., Chapleur, O., 2016. Community shifts within

693 anaerobic digestion microbiota facing phenol inhibition: Towards early warning
694 microbial indicators? Water Res. 100, 296–305.
695 https://doi.org/10.1016/j.watres.2016.05.041
696 Pugin, B., O'Mahony, L., Westermann, P., Hellings, P., Wawrzyniak, M., Heider, A., Akdis,
697 C.A., Barcik, W., 2017. A wide diversity of bacteria from the human gut produces and
698 degrades biogenic amines. Microb. Ecol. Health Dis. 28, 1353881.
699 https://doi.org/10.1080/16512235.2017.1353881
700 Quinn, T.P., Richardson, M.F., Lovell, D., Crowley, T.M., 2017. Propr: An R-package for
701 Identifying Proportionally Abundant Features Using Compositional Data Analysis. Sci.
702 Rep. 7, 1–9. https://doi.org/10.1038/s41598-017-16520-0
703 Roeder, J., Schink, B., 2009. Syntrophic degradation of cadaverine by a defined methanogenic
704 coculture. Appl. Environ. Microbiol. 75, 4821–4828.
705 https://doi.org/10.1128/AEM.00342-09
706 Rohart, F., Gautier, B., Singh, A., Cao, K.-A. Le, 2017. mixOmics: an R package for 'omics
707 feature selection and multiple data integration. bioRxiv 108597.
708 https://doi.org/10.1101/108597
709 Sekiguchi, Y., Takahashi, H., Kamagata, Y., Ohashi, A., Harada, H., 2001. In Situ Detection,
710 Isolation, and Physiological Properties of a Thin Filamentous Microorganism Abundant
711 in Methanogenic Granular Sludges: a Novel Isolate Affiliated with a Clone Cluster, the
712 Green Non-Sulfur Bacteria, Subdivision I. Appl. Environ. Microbiol. 67, 5740–5749.
713 https://doi.org/10.1128/AEM.67.12.5740
714 Shen, H., Huang, J.Z., 2008. Sparse principal component analysis via regularized low rank
715 matrix approximation. J. Multivar. Anal. 99, 1015–1034.
716 https://doi.org/10.1016/j.jmva.2007.06.007
717 Singh, A., Benoît, G., Shannon, C.P., Vacher, M., Rohart, F., Tebbutt, S.J., Lê Cao, K.-A.,

718 2019. DIABLO: an integrative, multi-omics, multivariate method for multi-group
719 classification. Bioinformatics 1–50.
720 https://doi.org/https://doi.org/10.1093/bioinformatics/bty1054
721 Smith, C.A., Want, E.J., Maille, G.O., Abagyan, R., Siuzdak, G., 2006. XCMS : Processing
722 Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment ,
723 Matching , and Identification. Anal. chemestry 78, 779–787.
724 Sousa, D.Z., Salvador, A.F., Ramos, J., Guedes, A.P., Barbosa, S., Stams, A.J.M., Alves,
725 M.M., Pereira, M.A., 2013. Activity and viability of methanogens in anaerobic digestion
726 of unsaturated and saturated long-chain fatty acids. Appl. Environ. Microbiol. 79, 4239–
727 4245. https://doi.org/10.1128/AEM.00035-13
728 Straube, J., Huang, B.E., Lê Cao, K.-A., 2017. DynOmics to identify delays and co-
729 expression patterns across time course experiments. Sci. Rep. 7, 1–11.
730 https://doi.org/10.1038/srep40131
731 Swiatczak, P., Cydzik-Kwiatkowska, A., Rusanowska, P., 2017. Microbiota of anaerobic
732 digesters in a full-scale wastewater treatment plant. Arch. Environ. Prot. 43, 53–60.
733 https://doi.org/10.1515/aep-2017-0033
734 Vanwonterghem, I., Jensen, P.D., Ho, D.P., Batstone, D.J., Tyson, G.W., 2014. Linking
735 microbial community structure, interactions and function in anaerobic digesters using
736 new molecular techniques. Curr. Opin. Biotechnol. 27, 55–64.
737 https://doi.org/10.1016/j.copbio.2013.11.004
738 Villas-Bôas, S.G., Noel, S., Lane, G.A., Attwood, G., Cookson, A., 2006. Extracellular
739 metabolomics: A metabolic footprinting approach to assess fiber degradation in complex
740 media. Anal. Biochem. 349, 297–305. https://doi.org/10.1016/j.ab.2005.11.019
741 Wang, M., Zhang, X., Zhou, J., Yuan, Y., Dai, Y., Li, D., Li, Z., Liu, X., Yan, Z., 2017. The
742 dynamic changes and interactional networks of prokaryotic community between co-

743 digestion and mono-digestions of corn stalk and pig manure. Bioresour. Technol. 225,
744 23–33. https://doi.org/10.1016/j.biortech.2016.11.008
745 Wishart, D.S., Feunang, Y.D., Marcu, A., Guo, A.C., Liang, K., Vázquez-Fresno, R., Sajed,
746 T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M.,
747 Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., Serra-Cayuela, A., Liu, Y.,
748 Mandal, R., Neveu, V., Pon, A., Knox, C., Wilson, M., Manach, C., Scalbert, A., 2018.
749 HMDB 4.0: The human metabolome database for 2018. Nucleic Acids Res. 46, D608–
750 D617. https://doi.org/10.1093/nar/gkx1089
751 Zamanzadeh, M., Hagen, L.H., Svensson, K., Linjordet, R., Horn, S.J., 2017. Biogas
752 production from food waste via co-digestion and digestion- effects on performance and
753 microbial ecology. Sci. Rep. 7, 1–12. https://doi.org/10.1038/s41598-017-15784-w
754