Integrating independent microbial studies to build predictive models of anaerobic digestion inhibition by ammonia and phenol

Simon Poirier, Sébastien Dejean, Cédric Midoux, Kim-Anh Lê Cao, Olivier Chapleur

1 **Title:**

2 Integrating independent microbial studies to build predictive models of anaerobic digestion
3 inhibition by ammonia and phenol

4 **Author names and affiliations**

5 Simon Poirier[a$], Sébastien Déjean[b], Cédric Midoux[a], Kim-Anh Lê Cao[c], Olivier Chapleur[a*],

6 [a] Université Paris-Saclay, INRAE, PRocédés biOtechnologiques au Service de
7 l'Environnement, 92761 Antony, France

8 [b] Toulouse Mathematics Institute, UMR 5219 CNRS, Toulouse University, Toulouse, France

9 [c] Melbourne Integrative Genomics, School of Mathematics and Statistics, The University of
10 Melbourne, Parkville, Victoria, Australia

11 $ Present address: Ecole Polytechnique Fédérale de Lausanne (EPFL), ENAC IIE Laboratory
12 for Environmental Biotechnology, 1015, Lausanne, Switzerland

13 simon.poirier@epfl.ch
14 sebastien.dejean@math.univ-toulouse.fr
15 cedric.midoux@inrae.fr
16 kimanh.lecao@unimelb.edu.au
17 olivier.chapleur@inrae.fr
18
19 **Corresponding author** (*)
20 Olivier Chapleur
21
22 Declarations of interest: "none".

Abstract

Anaerobic digestion (AD) is a process that can efficiently degrade organic waste into renewable energies. AD failure is however common as the underpinning microbial mechanisms are highly vulnerable to a wide range of inhibitory compounds. Sequencing technologies enable the identification of microbial indicators of digesters inhibition, but existing studies are limited. They used different inocula, substrates, sites and types of reactors and reported different or contradictory indicators. Our aim was to identify a robust signature of microbial indicators of phenol and ammonia inhibitions across four independent AD microbial studies. To identify such signature, we applied an original multivariate integrative method on two in-house studies, then validated our approach by predicting the inhibitory status of samples from two other studies with more than 90% accuracy. Our approach shows how we can efficiently leverage on existing studies to extract reproducible microbial community patterns and predict AD inhibition to improve AD microbial management.

Key words

Methane; Ammonia; Phenol; Microbial indicator; Inhibition prediction

1 Introduction

Anaerobic digestion (AD) is considered as the most efficient and sustainable technology for organic waste treatment. It has the ability to enable simultaneously waste stabilization and valorization through the production of methane rich biogas and of digestate used as an organic amendment. Encouraged by the renewable energy policies, biogas production with AD has increased in the European Union to reach 18 billion m^3 methane (654 PJ) in 2015 (Scarlat et al., 2018). However, The European Biomass Association (AEBIOM) estimates that AD still has a considerable potential for expansion with a biogas potential at about 78 billion m^3 methane. To reach this goal, the optimization of biogas production is essential to improve high process stability and efficiency and lower susceptibility to disturbances. Indeed, process failure reduces the economic and environmental performances of biogas technology as they lead to decreased methane yields and thus reduce revenues. Therefore, it is important that applied research on biogas technology improve robustness of these systems to stress factors, such as altered operating conditions or inhibitory compounds.

Among the broad range of inhibitors from AD substrates, high concentrations of ammonia and micro-pollutants such as phenol are considered as the primary cause of digester failure (Gonzalez-Gil et al., 2018; Rajagopal et al., 2013). Commonly used feedstock such as livestock manure, slaughterhouse byproducts and food industrial residues contain organic nitrogen such as urea and proteins, which readily release ammonia during their anaerobic degradation (Yenigün & Demirel, 2013). Previous studies addressing the topic of ammonia inhibition in AD have reported a large disparity in the inhibitory limits, which range from 27 to 1450 mg of NH_3-N/L and from 1.1 to 11.8 g of NH_4-N/L (Capson-Tojo et al., 2020).

In addition, various natural or anthropogenic phenolic compounds are detected in different types of effluents from coal gasification, coking, petroleum refining, petrochemical

63 manufacturing and paper (Rosenkranz et al., 2013). Phenols are also produced from
64 biodegradation of naturally occurring aromatic polymers such as humic acids and tannins or
65 from degradation of xenobiotic compounds such as pesticides (Veeresh et al., 2005). As a
66 result, contaminated sludge produced during the treatment of these various effluents can cause
67 digester imbalance. Phenol concentrations greater than 1000 mg/L were reported to severely
68 inhibit AD (Dong et al., 2019) and generally, half maximal inhibitory concentrations (IC_{50})
69 were reported to be between 1.1 and 1.8 g/L (Chapleur et al., 2016).The stability and
70 efficiency of the overall AD process relies on tightly coupled synergistic activities between an
71 intricate community of microorganisms. But the understanding of biological mechanisms of
72 AD is still hampered by the extreme complexity of the microbial ecosystem involved in this
73 process (Li et al., 2019). New knowledge is needed to unravel bioindicators of digesters
74 inhibition which have the potential to guide and optimize operation management during
75 unexpected onset of inhibitors and prevent biogas production.

76 Different methods for characterizing AD microbiome have been proposed and reviewed
77 by Lim et al. (Lim et al., 2020). New molecular biology techniques have revealed the great
78 diversity of the biogas-producing microbiomes. Thus, these omics technologies provide
79 unprecedented opportunities to characterize microbiomes' diversity, composition, gene
80 expression or metabolism and can be used to evidence bioindicators. For example, Hao et al.
81 used quantitative PCR and 16S rDNA amplicon sequencing to evidence associations between
82 process parameters and the abundance of specific microbial phylotypes in full scale digesters
83 (Hao et al., 2016). De Vrieze et al. evaluated the microbial community in full-scale AD plants
84 through amplicon sequencing of both the 16S rRNA gene and the 16S rRNA transcripts to
85 compare the total and active microbial community directly (De Vrieze et al., 2018).

86 High-throughput sequencing technologies, including 16S rRNA amplicon sequencing
87 have enabled many studies to monitor microbial changes during steady state or the inhibition

of anaerobic digesters, for example with phenolic compounds (Chapleur et al., 2016; Madigou et al., 2016; Poirier et al., 2016a) or ammonia (Lü et al., 2016; Poirier et al., 2016b). However, given a same inhibitor, independent studies regularly identified inconsistent microbial indicators, due to differences in inocula and substrates usage, geographical sites and/or at different times, and types of digesters.

The lack of reproducibility is further accentuated when studies focus on the identification of a single microbial bioindicator. Such univariate perspective is unlikely to shed light into the global and complex ecosystem of AD.

Our study aims at identifying a robust microbial signature, consisting in multiple microbial bioindicators, reflective of the interaction network within anaerobic digesters whilst leveraging on in-house and other existing studies. The analytical challenge was to combine such independent studies plagued by unwanted variation (e.g. different substrates and types of digesters) that outweigh the interesting biological variation of ammonia and phenol inhibitions. We applied a recently developed integration method MINT (Multivariate INTegrative) (Rohart et al., 2017a) that provides an integrated view of anaerobic digester microbiota subject to distinct types of inhibition. By integrating two independent in-house experiments, we identified reproducible microbial bioindicators that characterize ammonia inhibition, phenol inhibition and no inhibition. We evaluated this model by predicting AD status on two external studies assessing the influence of ammonia on AD. By doing so, we demonstrate the feasibility of detecting robust indicators evidencing the microbial symptoms of AD process dysfunction in independent studies which suggests promising applications in various biotechnologies thanks to the expansion of data deposited in public databases.

2 Materials and methods

2.1 Experimental data

Four studies assessing the influence of ammonia or phenol on AD were selected to build the predictive models (Lü et al., 2016; Peng et al., 2018; Poirier & Chapleur, 2018a; Poirier & Chapleur, 2018b). We selected these studies as they describe the effect of different levels of the previously mentioned inhibitors on both the performances of the digesters and the dynamics of the microbiome through 16S metabarcoding. Additionally, for each study, raw sequencing data had been deposited in publicly available databases and could be linked to the samples described in the original papers with no ambiguity.

In all studies, gas productions were measured to evaluate whether AD was inhibited. As described in supplementary material , studies 1 and 2 were conducted in our laboratory with the same type of substrate (biowaste) but with two different inocula collected one year apart from an industrial mesophilic digester treating wastewater treatment sludge. Samples were taken across time and under different inhibitory conditions. DNA was extracted and 16S rRNA gene was sequenced providing datasets of raw sequences associated to different inhibitory conditions. Study 1 aimed at assessing in parallel the effect of ten different levels of total ammonia nitrogen (TAN) (from 0 to 50 g TAN/L) and phenol (from 0 to 5 g/L) on the microbial community of batch anaerobic digester (Poirier & Chapleur, 2018b). Study 2 assessed the influence of support media addition to mitigate AD ecosystem inhibition in presence of two inhibitory conditions (19 g/L of TAN and 1.5 g/L of phenol, respectively) (Poirier & Chapleur, 2018a). Studies 3 and 4 were conducted in two distinct external laboratories. Study 3 was conducted by Lü *et al.* using as inoculum an anaerobic granular sludge collected from a plant-scale 35°C upflow anaerobic sludge bed reactor (Shanghai, China) with liquid internal recirculation that was treating paper mill wastewater. They

evaluated the effectiveness of biochar of different particle sizes in alleviating different ammonia inhibition levels (0, 3.5 and 7 g TAN/L) during AD of 6 g/L glucose (Lü et al., 2016). Study 4 by Peng et al. used an inoculum collected from a lab-scale, high-solids anaerobic digester and sought for microbial community changes during inhibition by ammonia (from 2.7 to 3.6 g TAN/L) in high solid AD of food waste in a continuous stirred-*tank* reactor (Peng et al., 2018).

In order to train an accurate and relevant MINT model, we removed some samples from studies 1 and 2 that were deemed non-representative of our analytical objectives (listed in supplementary material. In studies 1 and 2, only samples collected after at least 10 days of incubation were retained to ensure that the microbial community was representative of the inhibitory conditions. Samples taken after more than 60 days of incubation were removed as biogas production was completed. Moreover, for studies 1 and 2, methane cumulated production data were fitted to a modified Gompertz three-parameter model (Eq. (1)) where M(t) is the cumulative CH_4 production (mL) at time t (d); P is the ultimate CH_4 yield (mL); R_{max} is the maximum CH_4 production rate (mL/d); λ is the lag phase (d); e is the mathematical constant (also known as Euler number):

$$M(t) = P \times exp\left\{-exp\left[\frac{R_{max} \times e}{P} \times (\lambda - t) + 1\right]\right\} \quad \text{(Eq.1)}$$

Reactors were deemed inhibited when R_{max} was less than 80% of R_{max} in the controls without inhibitor, and not inhibited when R_{max} was greater than 90% of R_{max} in the controls without inhibitor. In study 1, samples from reactors incubated with 250 or 500 mg/L of phenol and 5.0 g/L of ammonium were discarded as the inhibition status (inhibited-non-inhibited) was not well defined. Samples collected from batch digesters incubated with 50 g/L of ammonium or 5 g/L of phenol were removed as these incubations were totally inhibited and microbial community did not evolve. In total 81 samples remained in studies 1 and 2, taken in

35 different digesters. All samples from study 3 and bacterial samples from study 4 were retained and their inhibition status predicted with the model (respectively 37 and 10, table S2).

2.2 Data processing

In these four studies, sequencing of the V3-V4 or V4-V5 region of the 16S rRNA gene was performed with three different approaches, as described in supplementary material . Data from external studies were downloaded from NCBI with fastq-dump 2.8.1. Paired-end reads from Lü *et al*., and Peng *et al.,* studies were merged with pear v0.9.11 (Zhang et al., 2014). Adapters from each study were specifically removed with cutadapt v1.12 (Martin, 2011). All sequences were imported into FROGS pipeline (Find, Rapidly, Otus with Galaxy Solution) (Escudié et al., 2018) . Samples from studies 1 and 2 were processed together while studies 3 and 4 were processed independently because of the differences in sequencing approaches. Taxonomic assignment of Operational Taxonomic Units (OTUs) was performed using Silva 132 SSU as reference database. OTUs were trimmed by keeping only those present more than 10 times in the whole dataset (resp. 1133, 399, 158 OTUs for studies 1 and 2, 3, 4). For joint analysis of data from studies 1 and 2, data was processed as obtained. For joint analysis of studies 1, 2, 3 and 4, the three distinct biom files were concatenated and data were discussed at the genus level. Sequences of interest were then assigned at the species level using the Blastn+ algorithm (Camacho et al., 2009).

2.3 Statistical analyses and predictive model

OTUs abundances were scaled with total sum scaling to account for uneven sequencing depth. OTUs that exceeded 3% in at least one sample were retained for the analysis. The total relative abundance of these minor OTUs represented 17% of the total number of sequences.

Data were then transformed with centered log ratio (CLR) transformation to account for compositional structure of the scaled data. All statistical analyses were implemented with mixOmics R package, as described in (Rohart et al., 2017b).

In order to obtain a first understanding of the major sources of variation in the training data (studies 1 and 2), and to obtain a first insight into the similarities between samples, we conducted principal component analyses (PCA) on the 16S rRNA tags datasets (pca function). A sparse Partial Least Squares Discriminant Analysis (Sparse PLS-DA) was then conducted to assess the potential to discriminate the samples according to the type of inhibition (Lê Cao et al., 2016) (sPLS-DA function) and identify microbial signatures characterizing inhibition type. Classification accuracy was calculated based on the microbial signature identified by the method, as described in (Rohart et al., 2017b). Finally, the MINT sPLS-DA method (referred to as MINT in the following), that generalizes sPLS-DA while accounting for study-specific effects was applied (Rohart et al., 2017a) (mint.splsda function). Parameters to tune in MINT included the number of PLS-DA components, and the number of variables to select, which was performed using 10-fold cross-validation. The final MINT model was then fitted on the data, and the classification performance was estimated using the perf function and 10-fold cross-validation repeated 10 times. Graphical display of the discriminative OTU signature identified by MINT were output using clustered image maps (cim function). The multivariate model not only identifies a microbial signature characterizing inhibition status, but it also enables to predict the groups of samples from external data sets as described in detail in (Rohart et al., 2017b). For prediction, we trained another MINT model on studies 1 and 2 for conditions ammonia/no inhibition and predicted the inhibition status of the test samples (studies 3 and 4) using the predict.mint.splsda function. The prediction area was visualized with a colored background on the sample plot, as described in (Rohart et al., 2017b). Code and functions used for data analysis are described in (Rohart et al., 2017a; Rohart et al.,

206 2017b) and available at http://mixomics.org/ and https://gitlab.irstea.fr/olivier.chapleur/mint-
207 bioindicators/.

3 Results and discussion

3.1 Integration of independent studies to identify microbial bioindicators

3.1.1 Inhibition status classification of digesters according to methane production performance

A total of 81 samples were selected from studies 1 and 2. These samples were collected in 35 distinct digesters (Poirier & Chapleur, 2018a; Poirier & Chapleur, 2018b). Prior to identifying potential bioindicators characteristic of both type of inhibition, inhibition status of each digester (non-inhibited, inhibited by phenol or inhibited by ammonia) has been characterized. For this purpose, maximum CH_4 production rate (mL CH_4/day) was chosen as the most informative performance indicator. These values were calculated for each digester using Grofit package of R software (version 3.1.2) in both previous studies(Poirier & Chapleur, 2018a; Poirier & Chapleur, 2018b). To integrate both studies, we decided that samples were inhibited as soon as maximum CH_4 production rate decreased by more than 20% compared to control and not inhibited if CH_4 production rate decreased by less than 10%. Figure 1 presents boxplots describing the distribution of the relative decrease of maximum CH_4 production rate of each digester according to their inhibition status.

According to this threshold, we determined that 29 samples were non-inhibited whereas 24 samples were inhibited by phenol and 28 samples by ammonia. In study 1, samples were non-inhibited as soon as initial inhibitor concentration remained lower than 0.1 g/L of phenol or 2.5 g/L of TAN. In digesters inhibited by ammonia and phenol a decrease by respectively 20 to 60% and 20 to 80% of methanogenic activity was observed. In study 2, regardless support addition, all digesters facing 19g/L of TAN were considered as inhibited (decrease of methanogenic activity by 60 to 90%). In presence of 1.5g/L of phenol, only digesters supplemented with activated carbons were considered as non-inhibited.

3.1.2 MINT modelling accounts for study effect

Considering, the inhibition status classification according to methane production performance, PCA was performed on the data (Fig. 2A), for a first exploration of the major sources of variation in the data. Sample distribution highlighted a strong study effect. Samples on the left part of the individual plot were related to study 1 conducted with the inoculum A while samples collected during study 2 conducted with inoculum B were on the right side of the individual plot.

However, a clear influence of the type of inhibition on microbial community could still be observed. For both studies, ecosystems facing ammonia inhibition were strongly discriminated from samples that were non-inhibited or inhibited by phenol. Similarly, within the study 1 conducted without support media, samples collected from batch digester inhibited by phenol were separated from non-inhibited samples.

A supervised PLS-DA model was then fitted on the data. Sparse version of the method was applied to select features and to identify discriminative OTUs that best described the difference between groups of samples. In order to conduct sPLS-DA, parameters such as the number of components, and the number of OTUs to select must be specified. We set these parameters based on the classification performance of sPLS-DA using cross-validation. Thirty-nine OTUs were thus selected by sPLS-DA and achieved a balanced error rate to 7.0%. Samples distribution based on the first two components is presented on Fig. 2B. As expected, sPLS-DA enabled to mitigate the study effect compared to the unsupervised PCA. However, within each condition, the study effect was still present: each sample collected in Study 1 was clearly separated from the ones collected in Study 2.

In order to counteract this bias, we applied MINT that combines independent studies measured on the same OTU predictors and identifies reproducible bioindicator signatures

across heterogeneous studies. As described above for sPLS-DA, we chose the optimal number of components and number of OTUs to select based on cross-validation, resulting in 45 OTUs and achieved a balanced error rate to 9.2% (2 components) (supplementary material). Samples representation from MINT is presented in Fig. 2C. It evidenced that the study effect was accounted for, with the strongest separation observed according to inhibiting condition rather than studies. The classification error rate of the final MINT model was 9.2% confirming the good performance of MINT to classify our samples and identify a microbial signature.

3.1.3 Analysis of microbial community

Microbial signatures identified with MINT were output in a clustered image map (or heatmap, 81 samples and 45 OTUs) in Fig. 3. This representation confirmed that, based on their microbial community composition, samples could be grouped by inhibition type (non-inhibited samples, samples inhibited by ammonia and samples inhibited by phenol). Moreover, the 45 OTUs selected by MINT were clustered into five different groups with a hierarchical clustering algorithm. The first group (Group A) was composed of 9 OTUs which were specifically correlated to digesters inhibited by phenol. Similarly, a second group of 17 OTUs (Group E) was associated to samples inhibited by ammonia. In group D, 6 OTUs were characteristic of both inhibitory conditions (phenol and ammonia). Group C included of 6 OTUs characterizing non-inhibited ecosystems while Group B was composed of 7 OTUs not recovered under ammonia inhibition. Interestingly, 6 of the 7 OTUs recovered in Group B were found in samples where phenol degradation was advanced. Consequently, the presence of these OTUs in this group may be explained by the variability of the inhibitory pressure throughout the incubation because of phenol degradation, and thus to their resilience capacity after phenol inhibition. Our following results are reported at the genus level, which was the most precise taxonomic level we could obtain with 16S rRNA sequencing.

3.1.3.1 Genera correlated to non-inhibitory conditions

The only discriminative archaeal OTU evidenced by the model, belonging to *Methanosarcina* genus, was negatively correlated to ammonia inhibition. Its belonging to Group B indicated that this genus was also inhibited by phenol but could still grow once phenol was degraded, confirming that this genus is characteristic of non-inhibitory conditions. *Methanosarcina* genus is known to play a key role in AD of many feedstocks (FitzGerald et al., 2015). However, although its robustness has been evidenced during different disruption situations (De Vrieze et al., 2012; Lins et al., 2014), it appeared that a drop of its relative abundance can reveal the presence of inhibitors such as ammonia or phenol in digesters. OTU_156 and OTU_6 assigned to *Clostridiaceae* 1 family and more precisely to *Caloramator* and *Clostridium butyricum*, strictly anaerobic acetogenic species described as butyrate producer (Cassir et al., 2016) were also correlated to non-inhibitory conditions. This result confirmed the observation that was made for studies 1 and 2 during which butyrate production was inhibited by phenol and ammonia (Poirier et al., 2016a; Poirier et al., 2016b). While butyrate, acetate and propionate are regularly mentioned as indicators of process imbalance (Ahring et al., 1995), our study tends to indicate that early variations in butyrate producers abundances could also be used as a bioindicator of inhibition. Interestingly, a single member of *Chloroflexi* phylum (OTU_26) kept for our model was also greatly associated to digesters for which phenol degradation had occurred. It was only assigned to the family *Anaerolinaceae,* which is widespread in full-scale digesters (Kirkegaard et al., 2017). Its filamentous form was suggested to favor synergistic relationship with archaeal populations such as *Methanosaeta* also known to be inhibited by phenol and ammonia (McIlroy et al., 2017). *Cloacimonetes* phylu*m* (formerly known as WWE1) was also specifically and strongly correlated to digesters where phenol degradation was advanced. It was represented by three distinct OTUs (OTU_11, OTU_35 and OTU_77). Only OTU_77 could be assigned at the

genus level as *Cloacomonas*. As *Anaerolinaceae*, this genus was recovered in various studies analyzing methanogenic sludge which suggested it was a syntrophic bacterium capable of degrading propionate, amino acids and cellulose (Regueiro et al., 2015). Another study indicated that this genus was sensitive to different inhibitions and notably to the antibiotic monensin which is released in cow manure and recovered in anaerobic digester (Spirito et al., 2018). The last phylum that was specifically correlated to digesters where phenol was partly degraded was *Thermotogaceae* with a single OTU (OTU_58) assigned to *Petrotogaceae,* which is known to be involved in phenol degradation (Na et al., 2016).

3.1.3.2 Genera correlated to a single type of inhibition

The model permitted the identification of genera specifically related to each type of inhibition. Interestingly, among the 17 OTUs that were correlated with ammonia inhibition, 9 of them belonged to 8 genera exclusively recovered in digesters inhibited by ammonia. They were assigned *Caldicoprobacter*, *Clostridium sensu stricto 15*, *Defluviitalea*, *Tepidimicrobium*, *Vagococcus*, *Tissierella, Lachnospiraceae NK4136* group and to an unknown genus of the *Family XI.* Among them, recent studies found that *Caldicoprobacter* populations were dominant in reactors exposed to high levels of ammonia (Poirier et al., 2016b). Similarly, Rui *et al*., evidenced that *Tissierella* genus was positively correlated to high concentration of TAN (Rui et al., 2015) while *Defluviitalea* were also dominant in digesters treating animal manure (Ma et al., 2017).

For the phenol, three OTUs belonging to *Clostridiales* order and assigned to genus *Sporoanaerobacter* (*Family XI*), to an unknown genus of the *vadinBB60* family and to an unassigned genus of the *Ruminococcaceae* family were evidenced. However, these cosmopolitan families are widely represented in anaerobic digesters.

On the other hand, we observed that some genera were similarly associated to digesters inhibited by phenol and to non-inhibited digesters suggesting that they were specifically sensitive to ammonia inhibition but not to phenol inhibition. It was notably the case for OTUs belonging to *Syntrophomonadaceae*, one of the major families of AD ecosystems. It consisted of five OTUs all assigned to *Syntrophomonas* genus but to unknown species. According to the model, three of them were related to non-inhibited digesters while two others were linked to phenol inhibition. *Syntrophomonas* are obligate anaerobic and syntrophic bacteria, which have the ability to oxidize saturated fatty acids, which is expected to enhance VFAs consumption. These functions are crucial in AD process. Thus, this result tended to confirm that under inhibitory conditions, reorganizations within the *Syntrophomonas* populations which carry these functions occur, thus confirming that the plasticity of the ecosystem is directly responsible for its resistance and resilience capacities (Shade et al., 2012). Notably, the high functional redundancy among *Syntrophomonas* species seemed to allow the preservation of global metabolic chain and methane production.

3.1.3.3 Genera correlated to both inhibitions

A few bioindicators were also associated to both types of inhibition. It was notably the case for the four OTUs belonging to genus *Bacteroidetes*, as well as for both couples of OTUs belonging to family *Porphyromonadaceae* and assigned to *Petrimonas* and *Proteiniphilum*. These three genera have been acknowledged to play an important role in degrading complex carbohydrates and proteins and catalyzing the production of VFAs and CO_2. Furthermore, the maintenance of important percentages of *Bacteroidetes* within a digester has already been suggested to be responsible for the ability of the anaerobic microbiota to counteract disturbances such as shock loadings (Regueiro et al., 2015). Similarly, within *Clostridiales* order, genera *Anaerosalibacter*, *Mobilitalea*, *Peptostreptococcus* and two unknown genera of *Lachnospiraceae* family were correlated to both inhibitions. *Lachnospiraceae* and particularly

353 to genus *Mobilitalea* as well as *Peptostreptococcus* and *Anaerosalibacter* were described as
354 resistant to phenol and ammonia inhibition and have been suggested to play important roles in
355 protein hydrolysis (Biddle et al.,2013). They were also reported to hydrolyze a variety of
356 polysaccharides by different mechanisms.

357 3.1.3.4 Clostridiales are key bioindicators of inhibition

358 Interestingly, 25 out of the 45 discriminant OTUs selected in our model belonged to
359 *Clostridiales* order. Among these 25 OTUs, 20 were related to inhibited ecosystems, which
360 tended to indicate that inhibitory pressure by phenol or by ammonia would preferentially
361 select more resistant bacteria affiliated to this order. A dominance of the *Clostridiales* order
362 has been reported by many studies at suboptimal conditions for methanogenesis (increased
363 ammonia or salt concentrations) (De Vrieze et al., 2015). Moreover, the importance of this
364 class is regularly mentioned as crucial in AD process (Joyce et al., 2018). However, the
365 majority of the OTUs belonging to this class could not be specifically correlated with only
366 one type of inhibition. It could be hypothesized that it may be due to the high functional
367 redundancy within these genera and diverse populations as mentioned previously. This
368 limitation was also observed for other phyla and notably for *Bacteroidetes* and *Spirochaetae*.
369 Furthermore, the lack of sequencing precision associated to the short length of 16S regions
370 amplified prevented the affiliation at taxonomic rank such as species or subspecies. Despite
371 the fact that we could not conclude about the correlation between these OTUs and the type of
372 inhibition, the model still highlighted that the emergence of these genera were associated to a
373 selection pressure caused by phenol or ammonia. Notably, it confirmed the negative
374 correlation observed by Heyer *et al*., between *Methanosarcinales*, which is a marker of steady
375 state and bacterial orders such as *Clostridiales* and *Spirochaetales* (Heyer et al., 2016). It also
376 reinforced the link found by Lee *et al*. between *Spirochaetales* and inhibited ecosystems (Lee
377 et al., 2013).

3.2 A predictive model for ammonia inhibition validated in two external studies

Samples from external studies 3 and 4 were analyzed with distinct sequencing techniques. Thus, different bioinformatic treatments were applied to raw sequences of each dataset due to differences in sequencing primers and targeted regions. OTUs were subsequently aggregated at the genus level (called 'clusters') in order to merge the three datasets. Since both external studied were focused on inhibition by ammonia only, we trained a new MINT model from our in-house samples where we removed the phenol condition. A total of 55 samples collected during studies 1 and 2 were retained, including 29 samples considered as non-inhibited and 26 as inhibited by ammonia.

3.2.1 Inhibition status prediction of external samples

In order to be consistent with the experimental strategy of the authors, samples from Lü et al., study were categorized into four groups depending on the sampling time and on the inhibitory pressure. Samples collected from digesters non-inhibited with ammonia were gathered in the "No inhibition" group. Samples inhibited with 3g/L of TAN were clustered in the "Ammonia moderate concentration" group. Samples inhibited with 7g/L of TAN were divided into two groups: "Ammonia inhibition, early days" for samples collected at the end of the lag phase and "Ammonia inhibition, final days" for samples collected near the end of the methane-production phase. Similarly, samples from Peng et al., study were categorized into four groups depending on the operational time when they were collected. Samples from day 0 to day 127, where methane yield remained stable circa 0.5 mL CH_4/g VS, were clustered in the "No inhibition" group. Samples from day 139 to day 152 were clustered into the "Ammonia inhibition start" group. During this phase, methane yield began to strongly fluctuate and slightly decrease down to 0.4 mL CH_4/g VS. Samples from day 172 to day 212 during which methane yield dropped down to 0.25 mL CH_4/g VS, were clustered in the

"Ammonia inhibition" group. Both samples from day 223 to day 232 were clustered in the "Ammonia inhibition decrease" group where biogas production restarted while methane yield remained below 0.25 mL CH_4/g VS.

3.2.2 Ammonia inhibition model

As previously, PCA and sPLSDA were performed on the data (Fig. 4A and Fig. 4B). An optimal number of 17 clusters were selected by the model, leading to a balanced error rate of 1.7% for sPLD-DA Sample distribution before applying MINT confirmed the strong study effect. The efficiency of MINT to discriminate the in-house samples into inhibition / no-inhibition conditions is depicted in Fig. 4C for the first two MINT components, with the first (horizontal) component highly discriminative of the inhibitory status of the digester. Yet, three of the inhibited samples were located on the left hand side of the plot. They corresponded to the most inhibited samples of study 1 with the highest concentration of TAN (25 g/L).

The OTUs selected in this second MINT model (supplementary material) were consistent with the observation made with the first model in our previous section (Figure 5). Among them, 10 clusters were positively correlated to non-inhibited digesters. Notably, it confirmed that *Clostridium butyricum* (Cluster_6) could be considered as a robust bioindicator of non-inhibitory conditions. Furthermore, clusters assigned to *Cloacimonetes* phylum (Cluster_11, Cluster_43 and Cluster_49) and particularly to genus *Cloacomonas* clearly related to non-inhibited digesters were re-evidenced. Nevertheless, due to the novelty of this phylum, missing reference genomes most probably hampered a deeper taxonomic classification of *Cloacimonetes* species. The ecological function of *Cloacimonetes* is thus not established, but this group is suggested to be only present in mesophilic conditions, involved in amino acid fermentation, syntrophic propionate oxidation and extracellular hydrolysis

(Muller et al., 2016). It was also evidenced that its abundance decreased with increasing ammonia levels (Westerholm et al., 2018). We also confirmed that *Syntrophomonas* (Cluster_17) was more represented in absence of ammonia. Moreover, the second model also revealed three new genera considered as discriminant of non-inhibited digesters conditions. These genera were assigned to two families belonging to *Clostridiales* order: *Ruminococcaceae* (Cluster_34 and Cluster_69), and *Christensenellaceae* (Cluster_42). Nevertheless, genera belonging to *Clostridiales* order are known to be involved in various metabolic activities which prevent us from elucidating their specific role under non-inhibiting conditions.

On the other hand, seven clusters were strongly linked to ammonia inhibition. Six of them belonged to *Clostridiales* order and notably to genera that were previously identified in the first model as specific markers of this type of inhibition (*Caldicoprobacter*, *Defluviitalea*, *Anaerosalibacter*, *Tepidimicrobium* and *Tissierella*). Another unknown genus belonging to *Family XI* was also discriminant confirming the great resistance of this family to ammonia inhibitory pressure. The last OTU was assigned to *Sphaerochaeta*, which belongs to *Spirochaetales* order. This order was associated to both types of inhibitions in the previous model. Interestingly, the heatmap presented in Fig. 5 emphasized that six of these seven clusters were significantly more abundant in samples collected in digesters inhibited with the highest ammonia concentration, thus reinforcing the robustness of these bioindicators.

The bioindicators highlighted by the second model were consistent with those evidenced by the model integrating phenol inhibition, thus confirming the robustness of this statistical analysis.

3.2.3 Prediction of the inhibitory status of samples analyzed by external studies

This second model was built to predict the inhibitory status of samples collected in two external studies 3 and 4. The prediction results are indicated in supplementary material . In order to visualize external samples distribution in the model, Fig. 6 presents the test samples from studies 3 and 4 projected on the two first components of the trained model, as well as prediction areas (Rohart et al., 2017b).

As expected, inhibited samples were separated against the non-inhibited samples on the first component. Two samples (11 and 39) were wrongly predicted as 'inhibited'. Samples of reactors that just started inhibition ("early days" in Lü et al. and 'start of the inhibition' in Peng et al.) were mostly predicted at an intermediary position and predicted as either inhibited or non-inhibited. We hypothesized that microbial community had started to change but was not yet totally characteristic of inhibited reactors. Interestingly, sample 45 (Peng et al., day 223, just after inhibitory pressure was lowered) was predicted as inhibited while sample 46 (Peng et al. day 232, several days after inhibitory pressure was lowered) was predicted as non-inhibited. This result may illustrate the progressive resilience of the microbial community after the inhibition. Sample 14 (moderate ammonia, final days, no addition of activated charcoal) was predicted as inhibited while samples 13, 15, 16 (moderate ammonia, but early days or addition of activated charcoal) were predicted as non-inhibited, in agreement with the conclusions of the authors. Finally, taking into account all the samples from digesters clearly non-inhibited (15) or clearly inhibited by ammonia (13) we estimated that the model predicted the inhibitory status of external samples with an accuracy of 93% as only two samples were incorrectly predicted.

3.3 Further perspectives for anaerobic digesters management

16S rRNA gene sequencing have revolutionized environmental biotechnology research and helped to progressively unravel the complexity of AD inhibition, but we still need to improve the resilience of AD systems and promote its implementation at a larger scale to avoid system failure. Our study focused on identifying robust microbial indicators from 16S sequencing data by integrating two in-house studies. We built a multivariate model that was highly predictive of inhibition status, as demonstrated in two external studies.

Despite the complexity and the functional redundancy of the microbial community within digesters, our model revealed the feasibility of detecting key indicators evidencing the state of the AD process whilst addressing the challenge of study-specific effects. The microbial indicators we identified were separate from cosmopolitan OTUs that tended to co-occur in all conditions. Thus, they can be considered as bioindicators to announce early signs of process dysfunction in anaerobic digesters, or, on the contrary, indicate steady functioning. Additionally these bioindicators are specific of one type of inhibitor, phenol or ammonia. Our study also emphasized on the benefit of using multivariate methods to identify a microbial signature from a training dataset (here our two in-house studies) then predict the inhibitory status of external studies despite differences in sequencing primers and targeted regions.

Future investigation will include a finer definition of the role of the identified microorganisms in AD process. As 16S reference databases are currently still incomplete, taxonomic assignment at the genus level results in a substantial lack of data interpretation. Consequently, the use of multiple marker genes DNA gyrase subunit B (*gyrB*) (Poirier et al., 2018), RNA polymerase subunit B (*rpoB*) (Case et al., 2007), DNA recombinase protein (*recA*), protein synthesis elongation factor-G (fusA), and dinitrogenase protein subunit D (*nifD*) (Holmes et al., 2004) could improve the assignation and abundance estimation for crucial taxonomic groups. Moreover, the use of shotgun metagenomic and metatranscriptomic

tools will also be useful to obtain different type of information such as functional bioindicators.

When successful, integrative multi-studies analysis not only enables to increase sample size and statistical power, but also to share and re-use data deposited in public databases. Our proposed approach will benefit the research community interested in identifying reproducible microbial signatures. Given the uncertainty related to inhibition thresholds in individual reactors, identifying robust microbial biomarkers can improve digester management. An increasing number of 16S sequences databases is becoming available to build predictive models of different types of inhibition, irrespective of the digesters operating conditions. Leveraging on these data will allow to optimize the prediction quality by recursively updating monitoring model. We anticipate that a single metagenetic sequencing will reduce the number and the complexity of analyses targeting different inhibitors, as bioindicators identified could be correlated to different types of inhibitions. Moreover, miniaturization and portability of sequencers will soon allow high frequency on-line measurements of microbial dynamics, involving low workload, and limited sampling issues (Shaffer, 2019). Such tools could be very useful in the daily management of AD systems and complement the existing physicochemical-based management. They can provide support to detect early warning indicators of process failure but also enhance the process performance by implementing the most suitable operating conditions for microbial communities.

4 Conclusion

A robust microbial signature of AD inhibition by ammonia and phenol was identified by integrating two independent16S metabarcoding studies with the multivariate approach MINT. The model based on the identified biomarkers was successful in predicting ammonia inhibition in independent digesters. This outcome highlights the potential of our approach to for other inhibitors and studies to discover new warning bioindicators. These data could thereafter be aggregated and used to build robust statistical models of AD inhibition to detect instabilities from various sources at early stages in industrial plants, and ultimately improve management of AD.

Data availability: The sequencing data of study 1, 2, 3 and 4 are respectively available in the bioprojects PRJNA450311, PRJNA450313, PRJNA253784 and PRJNA324313 in the NCBI BioProject database.

Competing Interests: There are no competing financial interests in relation to the work described.

Funding: KALC and OC scientific travels were supported in part by the France-Australia Science Innovation Collaboration (FASIC) Program Early Career Fellowships from the Australian Academy of Science.

Acknowledgement: This work has an electronic preprint posted on biorxiv preprint server. Link for this preprint is https://www.biorxiv.org/content/10.1101/2020.03.16.993220v2.full

References

1. Ahring, B.K., Sandberg, M., Angelidaki, I. 1995. Volatile fatty acids as indicators of process imbalance in anaerobic digestors. Applied Microbiology and Biotechnology, 43, 559-565.

2. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L. 2009. BLAST+: architecture and applications. BMC Bioinformatics, 10, 421.

3. Capson-Tojo, G., Moscoviz, R., Astals, S., Robles, Á., Steyer, J.P. 2020. Unraveling the literature chaos around free ammonia inhibition in anaerobic digestion. Renewable Sustainable Energy Rev, 117, 109487.

4. Case, R.J., Boucher, Y., Dahllof, I., Holmstrom, C., Doolittle, W.F., Kjelleberg, S. 2007. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. Appl Environ Microbiol, 73, 278-88.

5. Cassir, N., Benamar, S., La Scola, B. 2016. Clostridium butyricum: from beneficial to a new emerging pathogen. Clinical Microbiology and Infection, 22, 37-45.

6. Chapleur, O., Madigou, C., Civade, R., Rodolphe, Y., Mazéas, L., Bouchez, T. 2016. Increasing concentrations of phenol progressively affect anaerobic digestion of cellulose and associated microbial communities. Biodegradation, 27, 15-27.

7. De Vrieze, J., Hennebel, T., Boon, N., Verstraete, W. 2012. Methanosarcina: the rediscovered methanogen for heavy duty biomethanation. Bioresour Technol, 112, 1-9.

8. De Vrieze, J., Pinto, A.J., Sloan, W.T., Ijaz, U.Z. 2018. The active microbial community more accurately reflects the anaerobic digestion process: 16S rRNA (gene) sequencing as a predictive tool. Microbiome, 6, 63.

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

10. Dong, D., Wang, R., Geng, P., Li, C., Zhao, Z. 2019. Enhancing effects of activated carbon supported nano zero-valent iron on anaerobic digestion of phenol-containing organic wastewater. J Environ Manag, 244, 1-12.

11. 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.

12. FitzGerald, J.A., Allen, E., Wall, D.M., Jackson, S.A., Murphy, J.D., Dobson, A.D.W. 2015. Methanosarcina Play an Important Role in Anaerobic Co-Digestion of the Seaweed Ulva lactuca: Taxonomy and Predicted Metabolism of Functional Microbial Communities. PloS one, 10, e0142603-e0142603.

13. Gonzalez-Gil, L., Mauricio-Iglesias, M., Serrano, D., Lema, J.M., Carballa, M. 2018. Role of methanogenesis on the biotransformation of organic micropollutants during anaerobic digestion. The Science of the total environment, 622-623, 459-466.

14. Hao, L., Bize, A., Conteau, D., Chapleur, O., Courtois, S., Kroff, P., Desmond-Le Quéméner, E., Bouchez, T., Mazéas, L. 2016. New insights into the key microbial

578 phylotypes of anaerobic sludge digesters under different operational conditions. Water
579 Res, 102, 158-169.

580 15. Heyer, R., Benndorf, D., Kohrs, F., De Vrieze, J., Boon, N., Hoffmann, M., Rapp, E.,
581 Schlüter, A., Sczyrba, A., Reichl, U. 2016. Proteotyping of biogas plant microbiomes
582 separates biogas plants according to process temperature and reactor type.
583 Biotechnology for biofuels, 9, 155-155.

584 16. Holmes, D.E., Nevin, K.P., Lovley, D.R. 2004. Comparison of 16S rRNA, nifD, recA,
585 gyrB, rpoB and fusA genes within the family Geobacteraceae fam. nov. Int J Syst Evol
586 Microbiol, 54, 1591-9.

587 17. Joyce, A., Ijaz, U.Z., Nzeteu, C., Vaughan, A., Shirran, S.L., Botting, C.H., Quince, C.,
588 O'Flaherty, V., Abram, F. 2018. Linking Microbial Community Structure and
589 Function During the Acidified Anaerobic Digestion of Grass. Frontiers in
590 Microbiology, 9.

591 18. Kirkegaard, R.H., McIlroy, S.J., Kristensen, J.M., Nierychlo, M., Karst, S.M., Dueholm,
592 M.S., Albertsen, M., Nielsen, P.H. 2017. Identifying the abundant and active
593 microorganisms common to full scale anaerobic digesters. bioRxiv.

594 19. Lê Cao, K.A., Costello, M.E., Lakis, V.A., Bartolo, F., Chua, X.Y., Brazeilles, R.,
595 Rondeau, P. 2016. MixMC: A multivariate statistical framework to gain insight into
596 microbial communities. PLoS ONE, 11.

597 20. Lee, S.-H., Park, J.-H., Kang, H.-J., Lee, Y.H., Lee, T.J., Park, H.-D. 2013. Distribution
598 and abundance of Spirochaetes in full-scale anaerobic digesters. Bioresource
599 Technology, 145, 25-32.

600 21. Li, Y., Chen, Y., Wu, J. 2019. Enhancement of methane production in anaerobic digestion
601 process: A review. Applied Energy, 240, 120-137.

602 22. Lim, J.W., Park, T., Tong, Y.W., Yu, Z. 2020. The microbiome driving anaerobic
603 digestion and microbial analysis. Advances in Bioenergy, 5, 1-61.

604 23. Lins, P., Reitschuler, C., Illmer, P. 2014. Methanosarcina spp., the key to relieve the start-
605 up of a thermophilic anaerobic digestion suffering from high acetic acid loads.
606 Bioresour Technol, 152, 347-54.

607 24. Lü, F., Luo, C., Shao, L., He, P. 2016. Biochar alleviates combined stress of ammonium
608 and acids by firstly enriching Methanosaeta and then Methanosarcina. Water
609 Research, 90, 34-43.

610 25. Ma, S., Huang, Y., Wang, C., Fan, H., Dai, L., Zhou, Z., Liu, X., Deng, Y. 2017.
611 Defluviitalea raffinosedens sp. nov., a thermophilic, anaerobic, saccharolytic
612 bacterium isolated from an anaerobic batch digester treating animal manure and rice
613 straw. International journal of systematic and evolutionary microbiology, 67, 1607-
614 1612.

615 26. Madigou, C., Poirier, S., Bureau, C., Chapleur, O. 2016. Acclimation strategy to increase
616 phenol tolerance of an anaerobic microbiota. Bioresour Technol, 216, 77-86.

617 27. Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing
618 reads. 2011, 17, 3.

619 28. McIlroy, S.J., Kirkegaard, R.H., Dueholm, M.S., Fernando, E., Karst, S.M., Albertsen,
620 M., Nielsen, P.H. 2017. Culture-Independent Analyses Reveal Novel Anaerolineaceae

621 as Abundant Primary Fermenters in Anaerobic Digesters Treating Waste Activated
622 Sludge. Frontiers in microbiology, 8, 1134-1134.

623 29. Muller, B., Sun, L., Westerholm, M., Schnurer, A. 2016. Bacterial community
624 composition and fhs profiles of low- and high-ammonia biogas digesters reveal novel
625 syntrophic acetate-oxidising bacteria. Biotechnology for biofuels, 9, 48.

626 30. Na, J.-G., Lee, M.-K., Yun, Y.-M., Moon, C., Kim, M.-S., Kim, D.-H. 2016. Microbial
627 community analysis of anaerobic granules in phenol-degrading UASB by next
628 generation sequencing. Biochem Eng J, 112, 241-248.

629 31. Peng, X., Zhang, S., Li, L., Zhao, X., Ma, Y., Shi, D. 2018. Long-term high-solids
630 anaerobic digestion of food waste: Effects of ammonia on process performance and
631 microbial community. Bioresource technology, 262, 148-158.

632 32. Poirier, S., Bize, A., Bureau, C., Bouchez, T., Chapleur, O. 2016a. Community shifts
633 within anaerobic digestion microbiota facing phenol inhibition: Towards early
634 warning microbial indicators? Water research, 100, 296-305.

635 33. Poirier, S., Chapleur, O. 2018a. Influence of support media supplementation to reduce the
636 inhibition of anaerobic digestion by phenol and ammonia: Effect on degradation
637 performances and microbial dynamics. Data in Brief, 19, 1733-1754.

638 34. Poirier, S., Chapleur, O. 2018b. Inhibition of anaerobic digestion by phenol and ammonia:
639 Effect on degradation performances and microbial dynamics. Data in Brief, 2235-
640 2239.

641 35. Poirier, S., Desmond-Le Quemener, E., Madigou, C., Bouchez, T., Chapleur, O. 2016b.
642 Anaerobic digestion of biowaste under extreme ammonia concentration: Identification
643 of key microbial phylotypes. Bioresour Technol, 207, 92-101.

644 36. Poirier, S., Rué, O., Peguilhan, R., Coeuret, G., Zagorec, M., Champomier-Vergès, M.-C.,
645 Loux, V., Chaillou, S. 2018. Deciphering intra-species bacterial diversity of meat and
646 seafood spoilage microbiota using gyrB amplicon sequencing: A comparative analysis
647 with 16S rDNA V3-V4 amplicon sequencing. PLOS ONE, 13, e0204629.

648 37. Rajagopal, R., Masse, D.I., Singh, G. 2013. A critical review on inhibition of anaerobic
649 digestion process by excess ammonia. Bioresour Technol, 143, 632-41.

650 38. Regueiro, L., Spirito, C.M., Usack, J.G., Hospodsky, D., Werner, J.J., Angenent, L.T.
651 2015. Comparing the inhibitory thresholds of dairy manure co-digesters after
652 prolonged acclimation periods: Part 2--correlations between microbiomes and
653 environment. Water research, 87, 458-66.

654 39. Rohart, F., Eslami, A., Matigian, N., Bougeard, S., Lê Cao, K.A. 2017a. MINT: A
655 multivariate integrative method to identify reproducible molecular signatures across
656 independent experiments and platforms. BMC Bioinformatics, 18.

657 40. Rohart, F., Gautier, B., Singh, A., Lê Cao, K.-A. 2017b. mixOmics: An R package for
658 'omics feature selection and multiple data integration. PLOS Computational Biology,
659 13, e1005752.

660 41. Rosenkranz, F., Cabrol, L., Carballa, M., Donoso-Bravo, A., Cruz, L., Ruiz-Filippi, G.,
661 Chamy, R., Lema, J.M. 2013. Relationship between phenol degradation efficiency and
662 microbial community structure in an anaerobic SBR. Water research, 47, 6739-49.

42. Rui, J., Li, J., Zhang, S., Yan, X., Wang, Y., Li, X. 2015. The core populations and co-occurrence patterns of prokaryotic communities in household biogas digesters. Biotechnology for biofuels, 8, 158-158.

43. Scarlat, N., Dallemand, J.-F., Fahl, F. 2018. Biogas: Developments and perspectives in Europe. Renewable Energy, 129, 457-472.

44. Shade, A., Peter, H., Allison, S.D., Baho, D.L., Berga, M., Burgmann, H., Huber, D.H., Langenheder, S., Lennon, J.T., Martiny, J.B., Matulich, K.L., Schmidt, T.M., Handelsman, J. 2012. Fundamentals of microbial community resistance and resilience. Front Microbiol, 3, 417.

45. Shaffer, L. 2019. Inner Workings: Portable DNA sequencer helps farmers stymie devastating viruses. Proceedings of the National Academy of Sciences, 116, 3351.

46. Spirito, C.M., Daly, S.E., Werner, J.J., Angenent, L.T. 2018. Redundancy in Anaerobic Digestion Microbiomes during Disturbances by the Antibiotic Monensin. Applied and Environmental Microbiology, 84.

47. Veeresh, G.S., Kumar, P., Mehrotra, I. 2005. Treatment of phenol and cresols in upflow anaerobic sludge blanket (UASB) process: a review. Water research, 39, 154-70.

48. Westerholm, M., Isaksson, S., Karlsson Lindsjö, O., Schnürer, A. 2018. Microbial community adaptability to altered temperature conditions determines the potential for process optimisation in biogas production. Applied Energy, 226, 838-848.

49. Yenigün, O., Demirel, B. 2013. Ammonia inhibition in anaerobic digestion: A review. Process Biochem, 48, 901-911.

50. Zhang, J., Kobert, K., Flouri, T., Stamatakis, A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics (Oxford, England), 30, 614-20.

689 Figures legends:

690 Figure 1: Relative maximal specific methanogenic activity in the digesters of studies 1 and 2.
691 The different boxplots correspond to the different groups of digesters in each study. Relative
692 maximal specific methanogenic activity was calculated as the ratio of maximal specific
693 methanogenic activity in a digester divided by the maximal specific methanogenic activity in
694 controls without inhibitor.

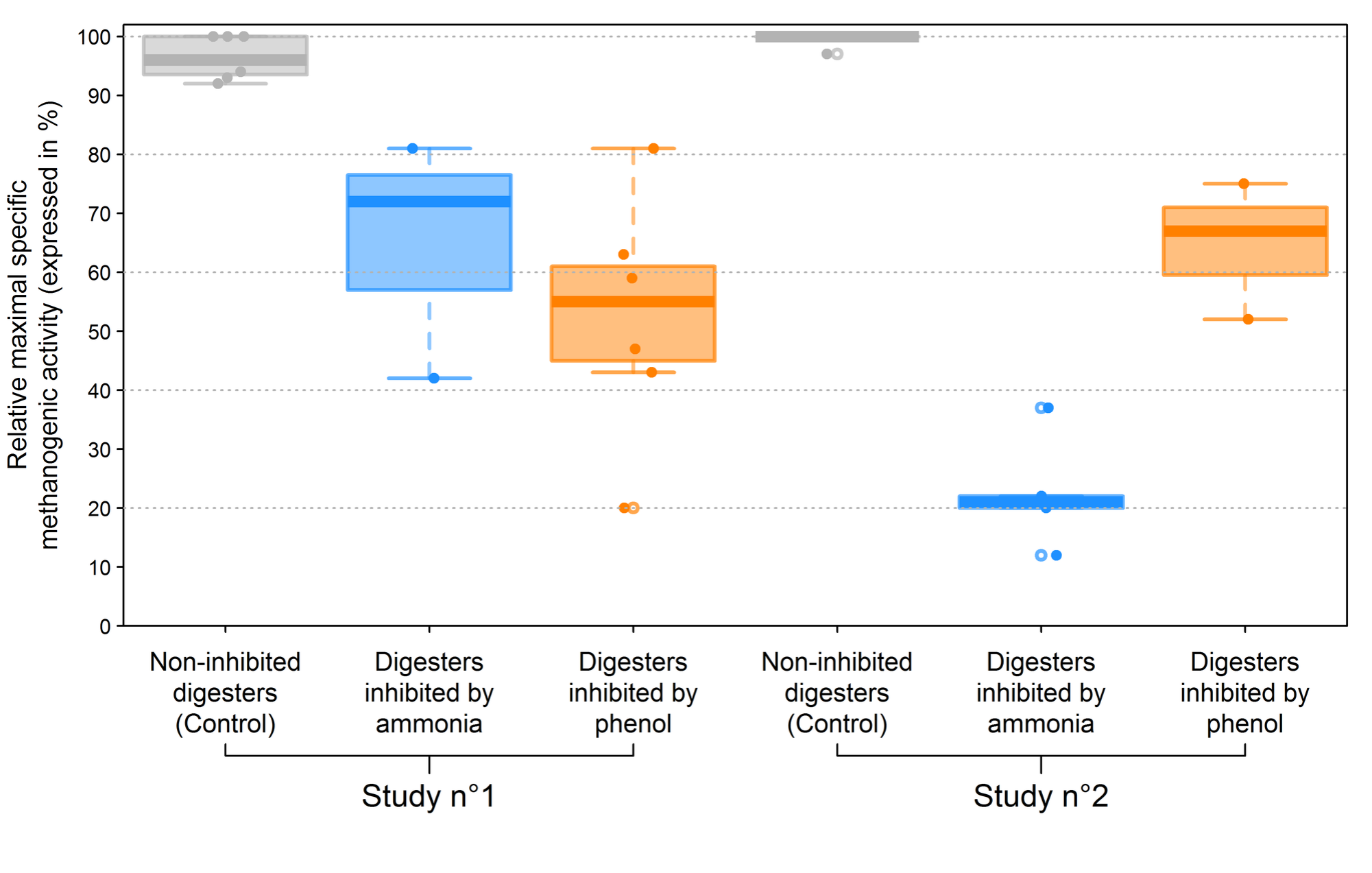
695 Figure 2: (A) Principal Component Analyses (PCA), (B) Sparse Partial Least Squares
696 Discriminant Analysis, (C) Multivariate Integrative Sparse Partial Least Squares Discriminant
697 Analysis of OTUs distribution in samples from studies 1 and 2, inhibited by phenol, ammonia
698 or not inhibited. On the factorial maps each sample is represented with a coloured marker.
699 The colour scale represents the type of inhibitor. The type of marker represents the study.
700 OTU data was generated by 16S rRNA gene sequencing.

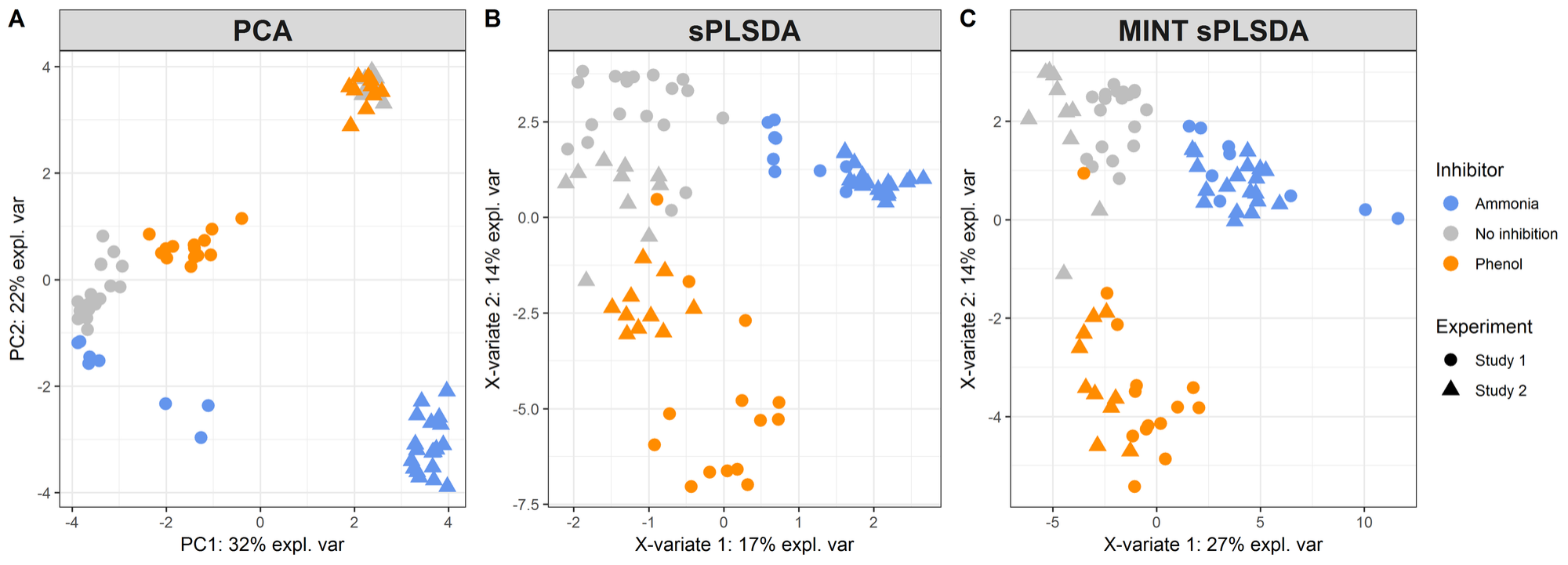
701 Figure 3: Heatmap of the most discriminant OTUs. Heatmap was built after selection of the
702 most discriminant OTUs with Multivariate Integrative Sparse Partial Least Squares
703 Discriminant Analysis of all OTUs generated by 16S rRNA gene sequencing for the different
704 samples of studies 1 and 2, inhibited by phenol, ammonia or not inhibited. Name of the OTUs
705 is indicated at the bottom. The color scale on the left represents the type of inhibitor. The
706 color key of the heatmap shows the abundance of the OTUs after CLR transformation (from
707 blue = low abundance to brown red = high abundance).

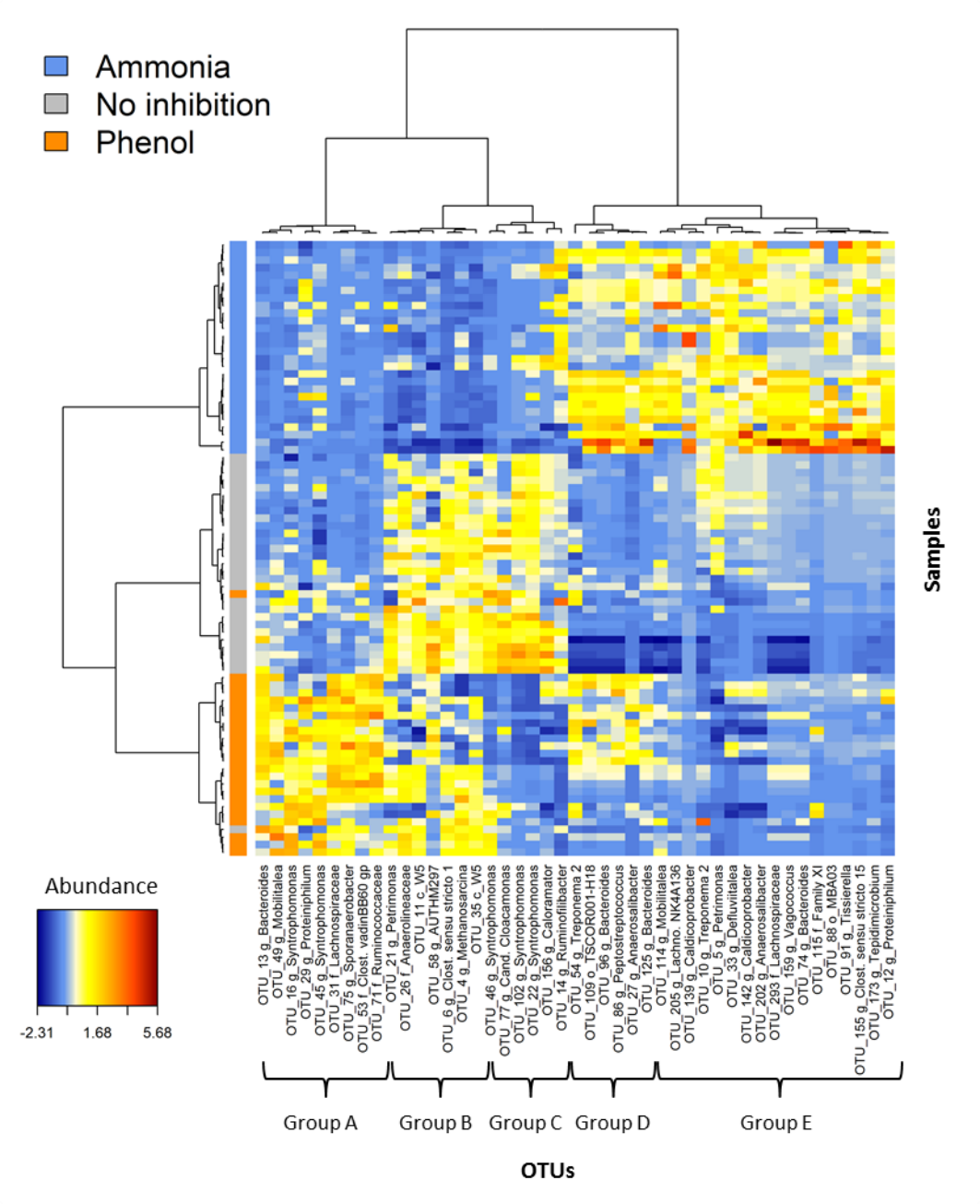
708 Figure 4: (A) Principal Component Analyses (PCA), (B) Sparse Partial Least Squares
709 Discriminant Analysis, (C) Multivariate Integrative Sparse Partial Least Squares Discriminant
710 Analysis of taxonomic distribution (genus level) in samples from studies 1 and 2, inhibited by
711 ammonia or not inhibited. On the factorial maps each sample is represented with a coloured
712 marker. The colour scale represents the type of inhibitors. The type of marker represents the
713 study. Taxonomic data was generated by 16S rRNA gene sequencing and aggregation of the
714 data at the genus level.

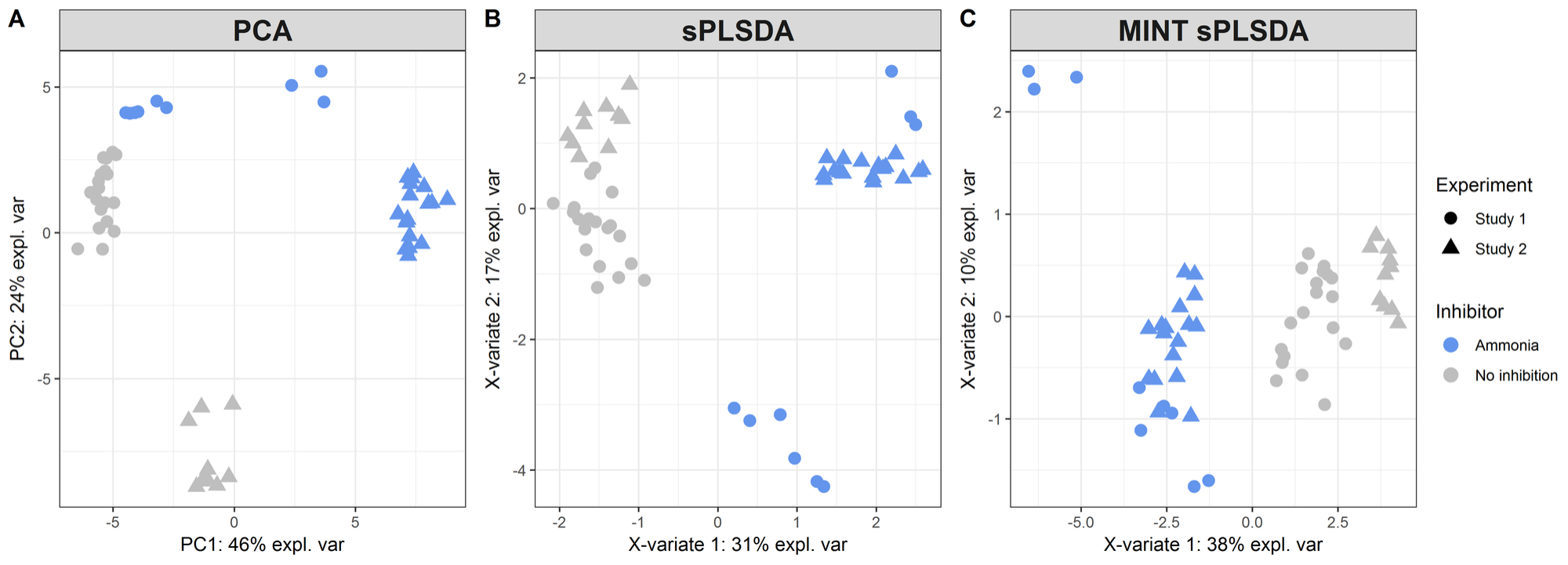
715 Figure 5: Heatmap of the most discriminant clusters of OTUs. Heatmap was built after
716 selection of the most discriminant clusters with Multivariate Integrative Sparse Partial Least
717 Squares Discriminant Analysis of all clusters generated after OTUs aggregation at the genus
718 level for the different samples of studies 1 and 2, inhibited ammonia or not inhibited. Name of
719 the clusters is indicated at the bottom. The color scale on the left represents the type of
720 inhibitor. The color key of the heatmap shows the abundance of the clusters after CLR
721 transformation (from blue = low abundance to orange = high abundance).

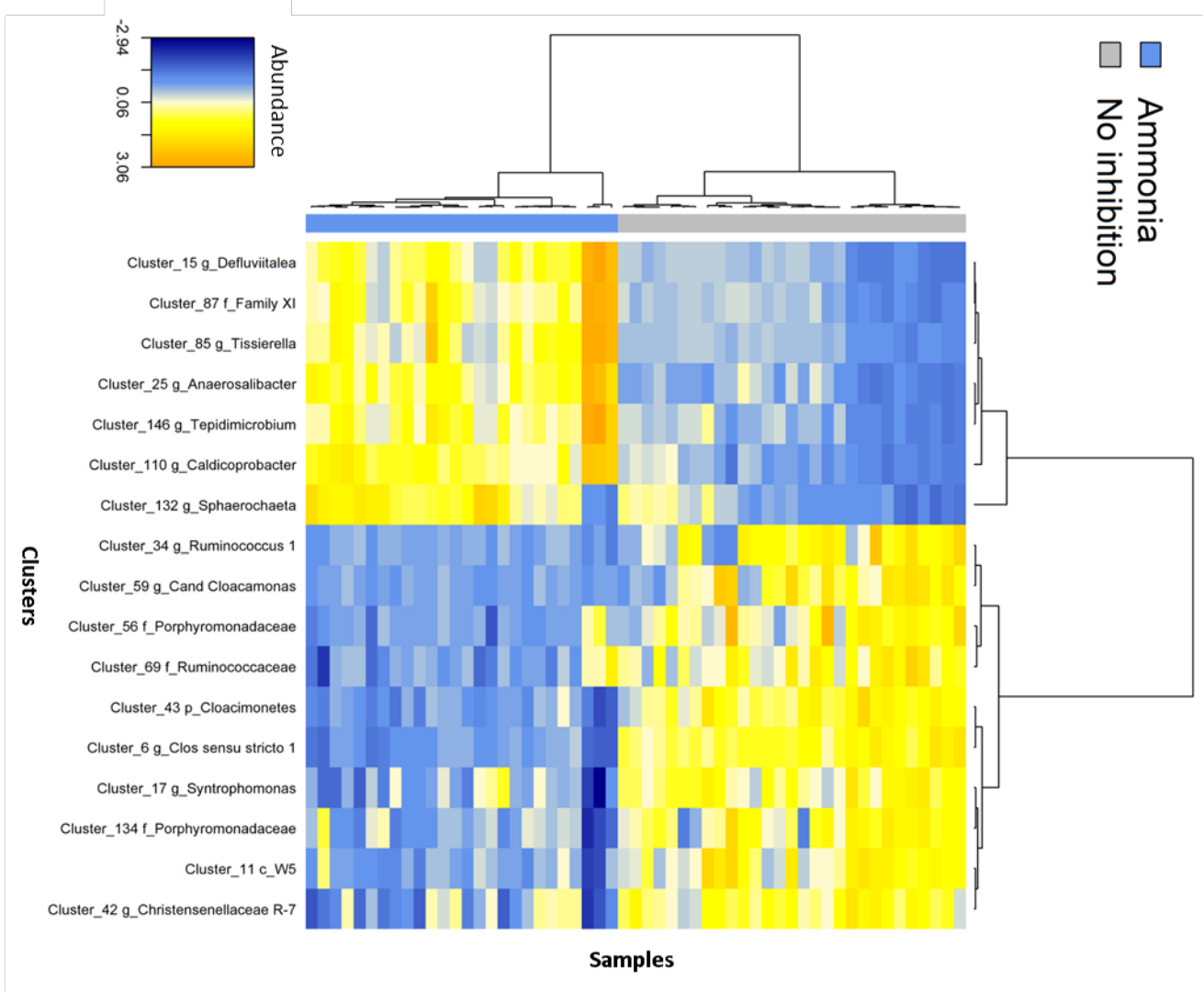
722 Figure 6: Projection of samples from studies 3 and 4 in the individual plot determined after
723 Multivariate Integrative Sparse Partial Least Squares Discriminant Analysis of samples from
724 studies 1 and 2. Each sample from studies 3 and 4 is represented by a marker. Type of marker
725 indicates the study. Color of the marker indicates the inhibition status in the reactor where the
726 sample was taken. A prediction area, based on studies 1 and 2 was calculated and is plotted on
727 the graph. The different figures indicate remarkable samples.

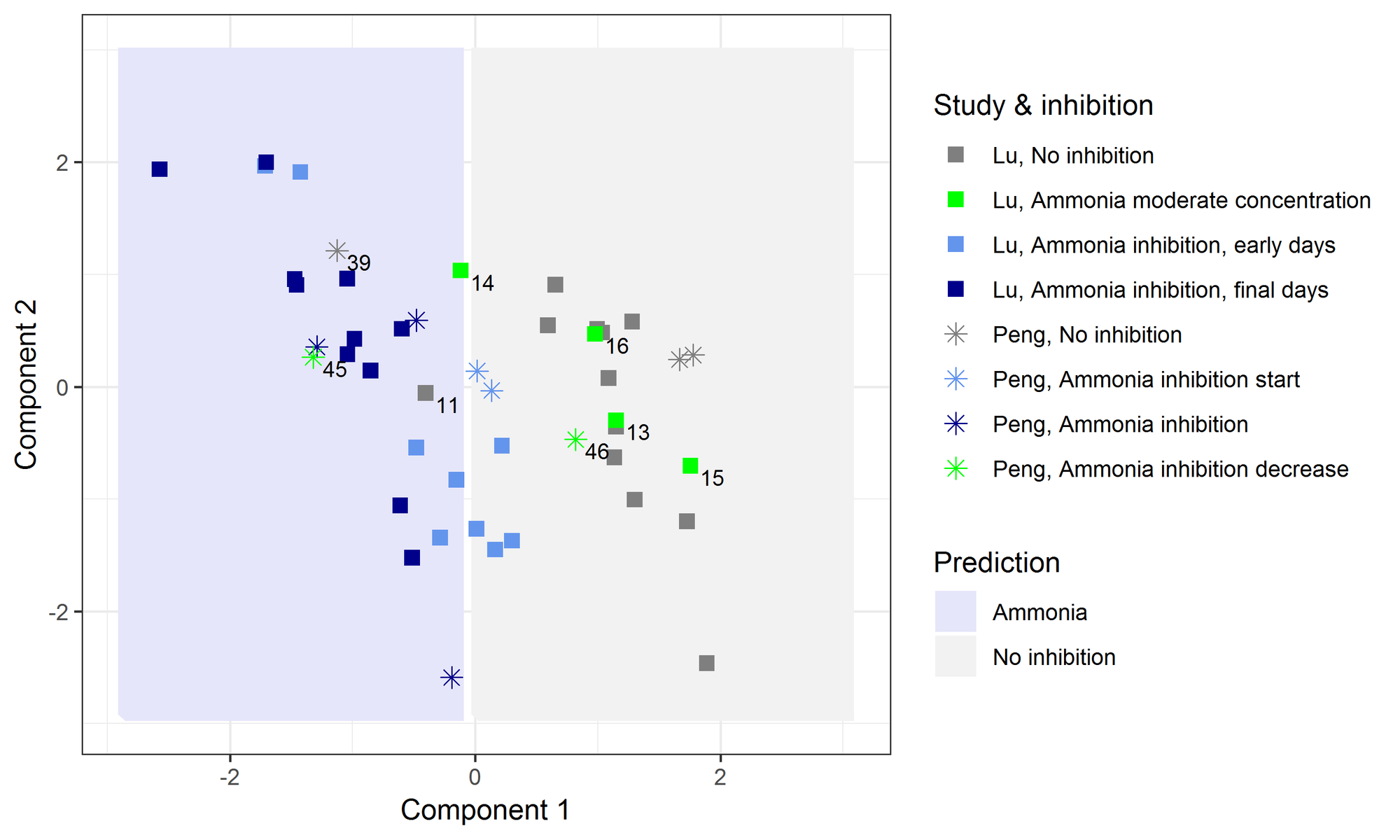

Relative maximal specific methanogenic activity (expressed in %)
100
90
80
70
60
50
40
30
20
10
0
Non-inhibited digesters (Control)
Digesters inhibited by ammonia
Digesters inhibited by phenol
Non-inhibited digesters (Control)
Digesters inhibited by ammonia
Digesters inhibited by phenol
Study n°1
Study n°2

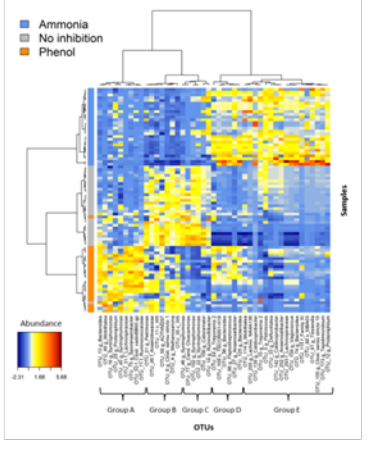

A
PCA
PC2: 22% expl. var
PC1: 32% expl. var
B
sPLSDA
X-variate 2: 14% expl. var
X-variate 1: 17% expl. var
C
MINT sPLSDA
X-variate 2: 14% expl. var
X-variate 1: 27% expl. var
Inhibitor
Ammonia
No inhibition
Phenol
Experiment
Study 1
Study 2


Ammonia
No inhibition
Phenol
Abundance
-2.31
1.68
5.68
OTU_13 g_Bacteroides
OTU_49 g_Mobilitalea
OTU_16 g_Syntrophomonas
OTU_29 g_Proteiniphilum
OTU_45 g_Syntrophomonas
OTU_31 f_Lachnospiraceae
OTU_75 g_Sporanaerobacter
OTU_53 f_Clost. vadinBB60 gp
OTU_71 f_Ruminococcaceae
OTU_21 g_Petrimonas
OTU_26 f_Anaerolineaceae
OTU_11 c_W5
OTU_58 g_AUTHM297
OTU_6 g_Clost. sensu stricto 1
OTU_4 g_Methanosarcina
OTU_35 c_W5
OTU_46 g_Syntrophomonas
OTU_77 g_Cand. Cloacamonas
OTU_102 g_Syntrophomonas
OTU_22 g_Syntrophomonas
OTU_156 g_Caloramator
OTU_14 g_Ruminofilibacter
OTU_54 g_Treponema 2
OTU_109 o_TSCOR001-H18
OTU_96 g_Bacteroides
OTU_86 g_Peptostreptococcus
OTU_27 g_Anaerosalibacter
OTU_125 g_Bacteroides
OTU_114 g_Mobilitalea
OTU_205 g_Lachno. NK4A136
OTU_139 g_Caldicoprobacter
OTU_10 g_Treponema 2
OTU_5 g_Petrimonas
OTU_33 g_Defluviitalea
OTU_142 g_Caldicoprobacter
OTU_202 g_Anaerosalibacter
OTU_293 f_Lachnospiraceae
OTU_159 g_Vagococcus
OTU_74 g_Bacteroides
OTU_115 f_Family XI
OTU_88 o_MBA03
OTU_91 g_Tissierella
OTU_155 g_Clost. sensu stricto 15
OTU_173 g_Tepidimicrobium
OTU_12 g_Proteiniphilum
Group A
Group B
Group C
Group D
Group E
OTUs
Samples

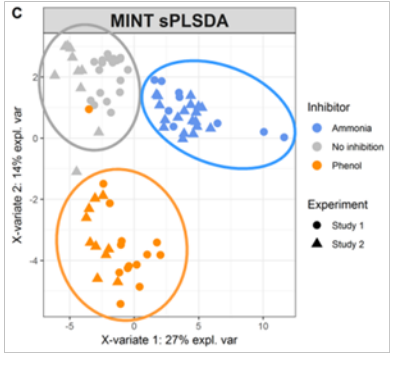

A
PCA
PC2: 24% expl. var
PC1: 46% expl. var
B
sPLSDA
X-variate 2: 17% expl. var
X-variate 1: 31% expl. var
C
MINT sPLSDA
X-variate 2: 10% expl. var
X-variate 1: 38% expl. var
Experiment
Study 1
Study 2
Inhibitor
Ammonia
No inhibition


-2.94
0.06
3.06
Abundance
Ammonia
No inhibition
Cluster_15 g_Defluviitalea
Cluster_87 f_Family XI
Cluster_85 g_Tissierella
Cluster_25 g_Anaerosalibacter
Cluster_146 g_Tepidimicrobium
Cluster_110 g_Caldicoprobacter
Cluster_132 g_Sphaerochaeta
Cluster_34 g_Ruminococcus 1
Cluster_59 g_Cand Cloacamonas
Cluster_56 f_Porphyromonadaceae
Cluster_69 f_Ruminococcaceae
Cluster_43 p_Cloacimonetes
Cluster_6 g_Clos sensu stricto 1
Cluster_17 g_Syntrophomonas
Cluster_134 f_Porphyromonadaceae
Cluster_11 c_W5
Cluster_42 g_Christensenellaceae R-7
Clusters
Samples

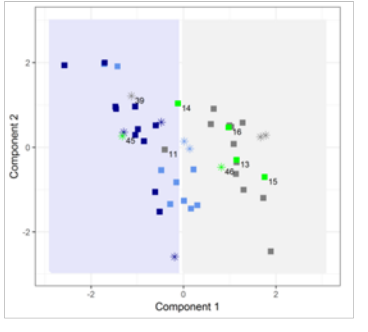

Component 2
Component 1
2
0
-2
-2
0
2
39
14
16
45
11
13
46
15
Study & inhibition
Lu, No inhibition
Lu, Ammonia moderate concentration
Lu, Ammonia inhibition, early days
Lu, Ammonia inhibition, final days
Peng, No inhibition
Peng, Ammonia inhibition start
Peng, Ammonia inhibition
Peng, Ammonia inhibition decrease
Prediction
Ammonia
No inhibition



Identification of biomarkers of inhibition



Integrative analysis of data from independent studies



Prediction of the inhibition in external digesters