Report on Metataxonomic Analysis of the Anaerobic Digestor Organic Overload Experiment

Introduction

Anaerobic digestors (AD) are enclosed systems which contain a microbiome capable of digesting/breaking down organic components (organic polymers) in the absence of oxygen. These systems yield biogas, such as methane, and nutrient rich digestate. An AD are of importance as an efficient method of extracting energy and other valuable products from organic waste and aid the circular economy[1,2].

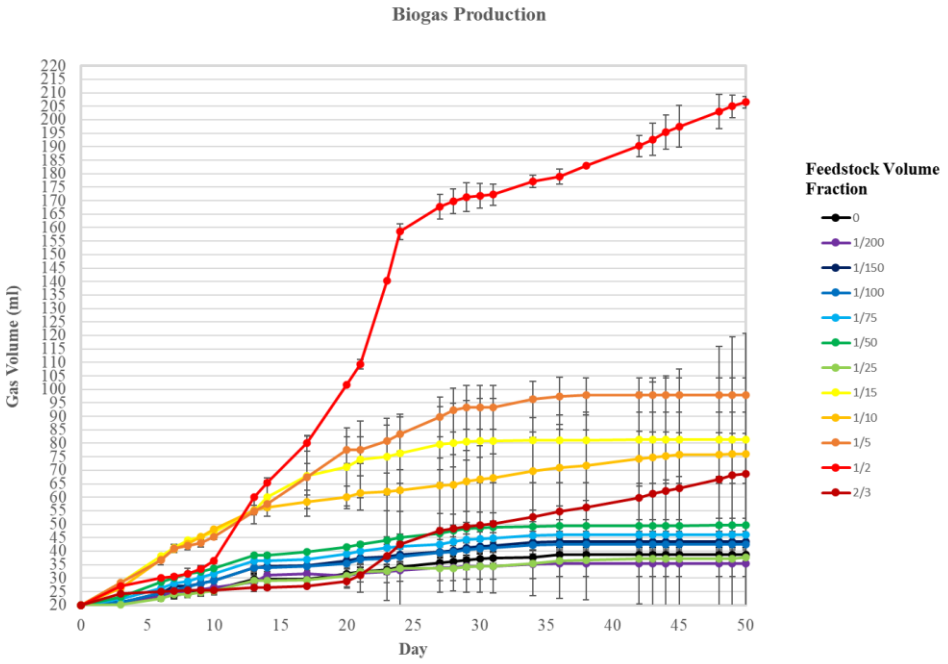
The microbiome responsible for the biochemical process within the AD is composed of a complex mixture of eukaryotes, archaea and, predominantly, prokaryotes. The dynamic syntrophic (metabolic interdependency) interactions in the AD microbiome is complex. As its yield is sensitive to the microbial diversity, process parameters and organic matter loaded into the reactor [2,3]. AD optimization can be achieved when the dynamics governing the production of the outputs are understood[3]. This analysis report examines the metataxonomic of ADs with changing organic overloads and the effect it has on biogas production and population differences among the samples. Additionally, the report will suggest taxonomic biomarkers indicative of the health of an AD.

The16Sr RNA was sequenced from digestates obtained from industrial-scale AD reactors at the Sea Wastewater Treatment Plant (WWTP) by the Edinburgh Genome Foundry using an Illumia MiSeq System. The AD samples were feed varying ratios of feed stock and sequenced. The experiment was repeated on the same samples and sequenced once again at a later date (4-5 weeks). The experimental sample set up is shown in Supplementary Table 1. The methane production of the varying feed-stock samples is measured. The QIIME2 workflow was employed to analyse the sequence data and ultimately produce taxonomic classification using the Silver 138 database, alpha and beta diversity analysis, and ANCOM analysis. Detailed information about the workflow can be found at QIIME 2.org and in Supplementary Fig1.

Results

Methane Production & Initial System's Conditions.

Methane measurements were obtained from glass syringes with varying feedstock samples. The results are displayed in Figure 1(A). Samples with low ratios slowly increase their biogas production, they then stabilize at approximately day 27. Samples 1/15 - 1/5 also stabilizes but they have a quicker increase.1/2 and 2/3 samples experience some initial inhibition in their production, then there is a rapid increase in the rate of methane production. 1/2 reaches a maximum of 167 ml by day 50, while the other samples have a range of 27-89ml.



A

Figure 1- (A) ***Biogas production of varying feedstock ratios*** *of experimental AD reactors.*

Sequencing and Taxonomic Assignment using Qiime2

There are 80 samples with a total sum of 10,070,536 raw reads and an average of 125,882 forward and reverse reads, and a median of 95,979 reads per sample. The minimum read count is 807 and the maximum is 428,460. The 5' end read profiles are of high quality. As expected, the reverse read's qualities are lower towards the 3'-end. Overall, the data is of high quality.

100% of the 80 sample reads entered the de-noising step. 83.5% of reads remained after the removal of nosey sequences. 82.8% of reads were left after the application of the error model on to the dataset. 79.5% of reads remained after the forward and reverse sequences were overlapped and the unreliable pairs were removed. Finally, chimeric sequences were filtered out, 72.9% of readings remained for further analysis. The percentage remaining reads is large, highlighting the quality of the dataset.

The DADA2 denoising step produced 3,417 amplicon sequence variants (ASV) from the remaining 7,343,926 reads (from the original 10,070,536). The new mean per sample is 91,799 and the median is 66,826. The negative control pool for the whole-data has 876 reads, indicating the effectiveness and reliability of the outcome. The feature table was then produced, and the taxa commonly known to contaminate metataxonomic procedures and barcode bleeds were removed.

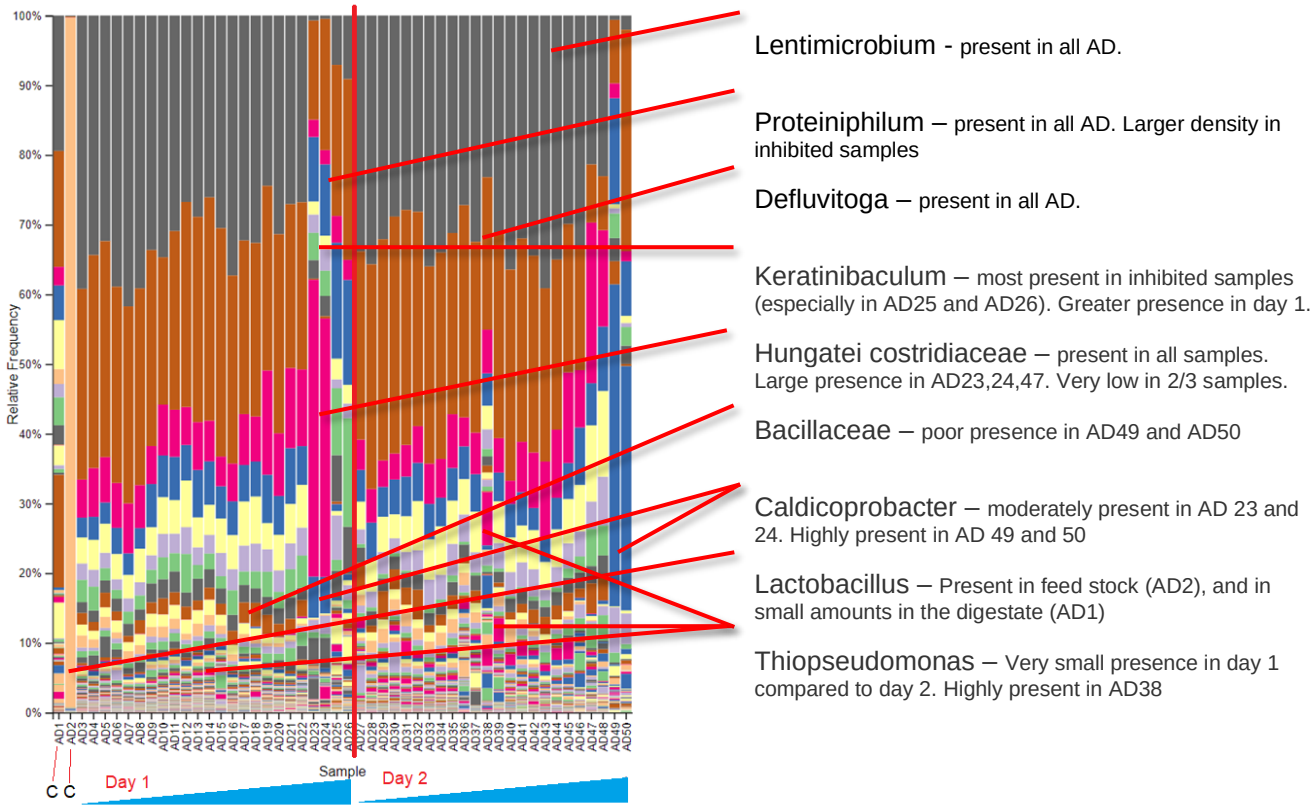


Figure 2 - Species-level taxonomic summary of AD samples. Visually interesting taxa are labelled and their noteworthy observations. C – control samples.

Non-inhibited samples from day 1 (AD3-AD22) and day 2 (AD27- AD46), show similar taxa patterns from a visual inspection of Figure 2. AD38 seems to be a possible outlier. Inhibited 1/2 and 2/3 feedstock ratios from day 1 (AD23-AD26) and day2 (AD47-AD50) appear to have (1) new/absent taxa, (2) different taxa density, (3) different patterns compared to the non-inhibited samples.

There are changes in the microbial distributions at these high levels of feed stock, which corelate with the reactor's functionality as seen in Figure 1(A). Table 1 provides additional information of the interesting taxa. The ANCOM test will later statistically evaluate the taxa which could be used as a biomarker for inhibited or un-inhibited reactors.

Taxa	Information
Lentimicrobium saccharophilum[4]	- major end products: acetate, malate, propionate, formate and hydrogen
Proteiniphilum Defluviitoga [5]	- responsible for hydrolysis and acid production - metabolizes sugar & generates hydrogen
Keratinibaculum paraultunense[6]	- metabolizes keratinase
Hungatei Clostridiaceae[7]	- nitrogen fixing - cellulose, xylan, cellobiose, cellodextrins, D-glucose, D-xylose, D-fructose, D-mannose and gentiobiose fermenting
Bacillaceae[8]	- Some species have cellulolytic and hemicellulolytic activities - Associated with the increased rate of hydrolysis in digestors
Caldicoprobacter[9]	- Capable of fermenting various saccharides - Produces lactate and acetate, ethanol, co2 and H2 - in the presence of yeast it utilized xylose, glucose, galactose, cellobiose, raffinose and xylan as carbon and energy sources
Lactobacillus delbrueckii [10,11]	- homofermentative (fermentation resulting wholly or principally in a single end product) producer of d-lactic acid
Thiopseudomonas[12]	- Denitrifying - Oxidizes sulfide anaerobically, nitrate is used as the electron acceptor

Table 1 - Summary of metabolic processes of interesting taxa derived from Figure 2

Diversity Analysis

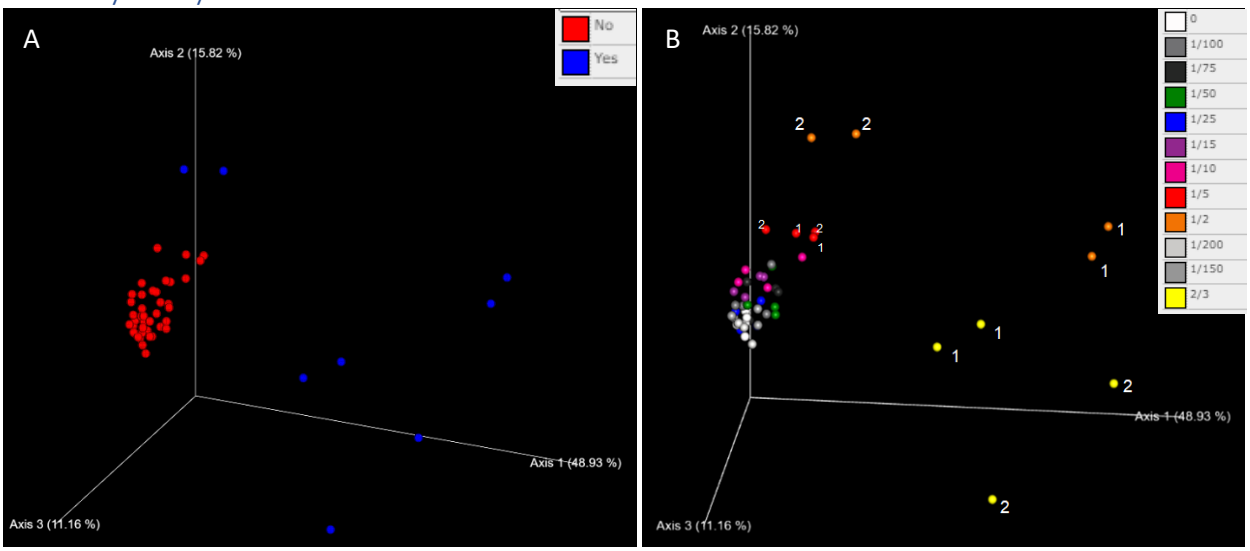


*Figure 3 - Beta diversity, Principal coordinates analysis (PCoA) derived from the **Bray-Curtis** distances among samples of the 47 populations (p = 0.001 by PERMANOVA, p = 0.001 by PERMDISP). (A) Coloured dots representative of the inhibited (Yes) and un-inhibited (No) samples. (B) Coloured dots representative of the various feedstock fractions. Inhibited samples have been labelled by the day they were sequenced.*

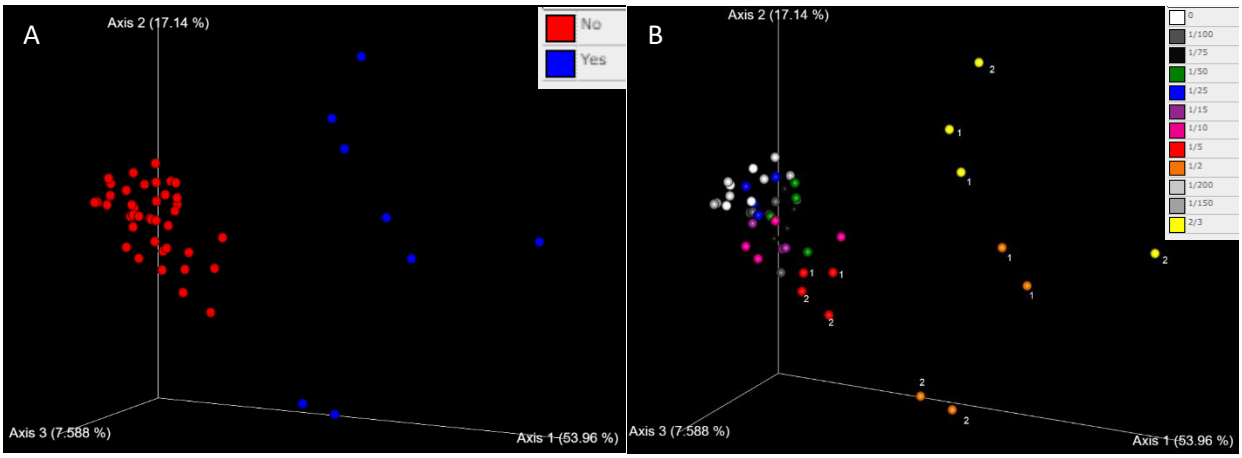


*Figure 4 - Figure 2 - Beta diversity, Principal coordinates analysis (PCoA) derived from the **Weighted Unifrac** distances among samples of the 47 populations. (A) Coloured dots representative of the inhibited (Yes) and un-inhibited (No) samples. (B) Coloured dots representative of the various feedstock fractions. Inhibited samples have been labelled by the day they were sequenced.*

Figure 3 represents 75% of the total variation among samples. Most of the differences (48.93%) lay on a single axis. Visually, there is a distinction between the microbiome of the inhibited and the uninhibited samples (Fig3A). The Inhibited samples feedstock ratio replicate pairs are near each other, indicating that their populations are similar (Fig3B). The 1/2 samples' populations seem to become more similar to the uninhibited sample populations as their day 2 replicates near the un-inhibited cluster. The 2/3 samples seem to become more distinct from the uninhibited cluster and from each other.

The PERMANOVA statistical analysis indicates that the spread of the samples in the inhibited and uninhibited reactors are different. The test signals that the uninhibited sample's populations are more similar to each other than the populations of the inhibited samples ($p = 0.001$; Pseudo-F = 30.76). The PERMDISP test indicates a statistically significant difference of the displacement metric between the inhibited and uninhibited sample ($p = 0.001$; F-Value = 175), showing that there is a difference between the populations of the two reactor groups.

The Weighted Unifrac graph, shown in Fig.4, represents 79% of the total variation among the samples. Most of the differences (53.96%) lays on a single axis. As the Bray-Curtis plot (Fig.3), the Weighted Unifrac plot suggests that the populations among the un-inhibited samples are more similar to each other than in the inhibited samples. Additionally, there is a noticeable overall population difference between the inhibited and un-inhibited reactors. Looking at Fig4.B it can be argued that the 1/2 feedstock day 2 samples do not drift to become more similar to the uninhibited samples, unlike the Bray-Curtis plot (Fig3.B). No statistical analysis was conducted on this PCoA plot (Fig4).

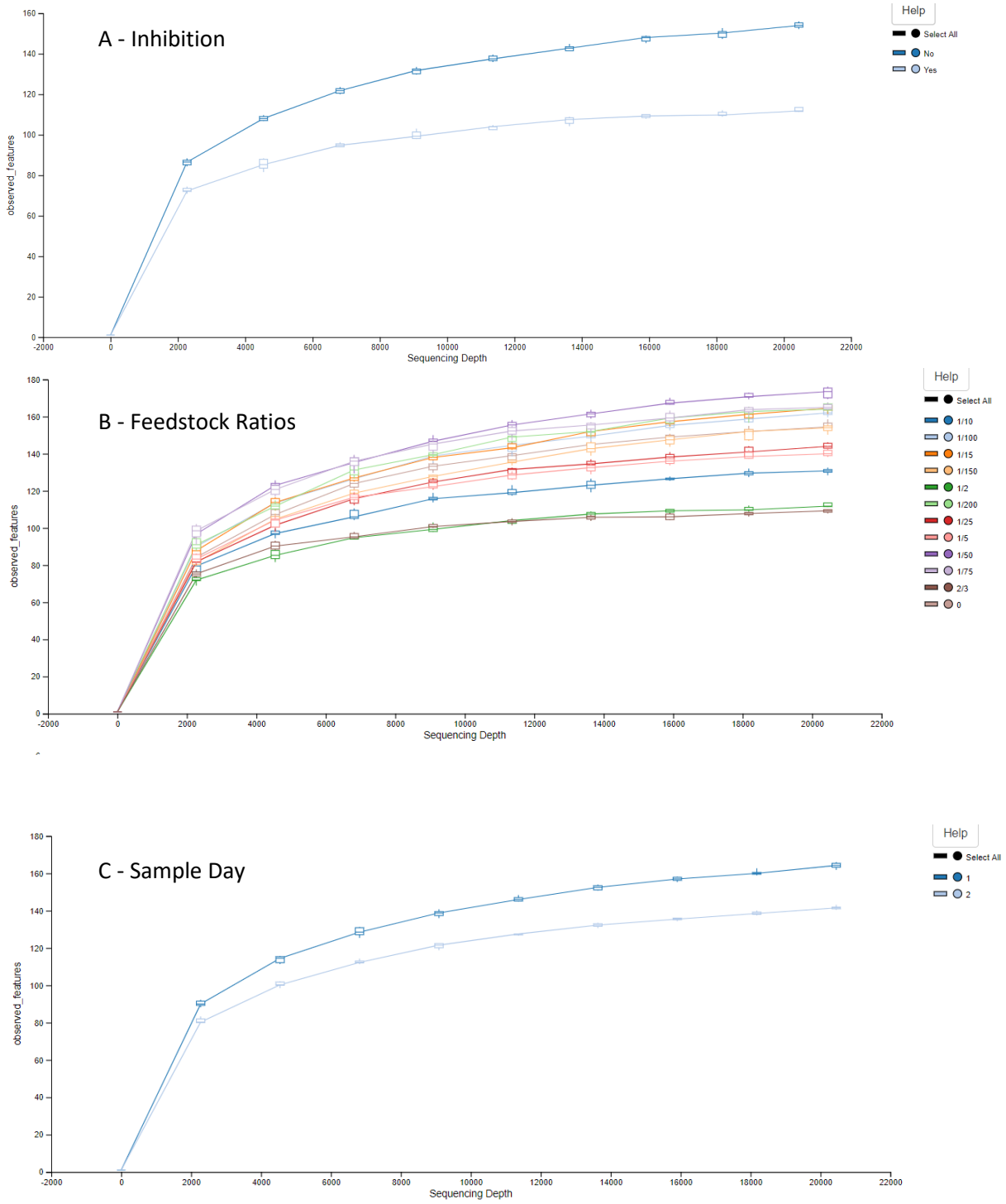


Figure 5 - Alpha diversity across (A)inhibition, (B) feedstock ratios, (C) sample day. Rarefaction curve shows, a commonly used measure of alpha diversity, observed features. (A) splits the samples into inhibited (Yes) and uninhibited (No) samples.

Figure 5 shows the rarefaction curves as a measure of alpha diversity across (A)inhibition, (B) feedstock ratios, and (C) sample day communities.

A Kruskal-Wallis test confirms that the difference of the inhibited and the un-inhibited sample' medians are statistically significant (p = 0.00006, H=16.00). Thus, the inhibited reactors have less diversity than the un-inhibited. The Kruskal-Wallis test (all groups) demonstrated that the difference of all the feedstock ratio medians are statistically significant (p = 0.013, H=23.76). Fig5B indicates that the most diverse feedstock ratio is1/50, and then 1/200, 1/75,1/100, 1/15 in no particular order. The least diverse sample is 2/3, followed by 1/2 (there is no statistically significant difference between them, p = 1); both of which are inhibited. The least diverse uninhibited group is 1/10. There is a statistical difference between sample day 1 and day 2 medians (p = 0.0004, H=12.66). Fig5C suggests that day 2 sample's diversity decreased.

ANCOM Analysis

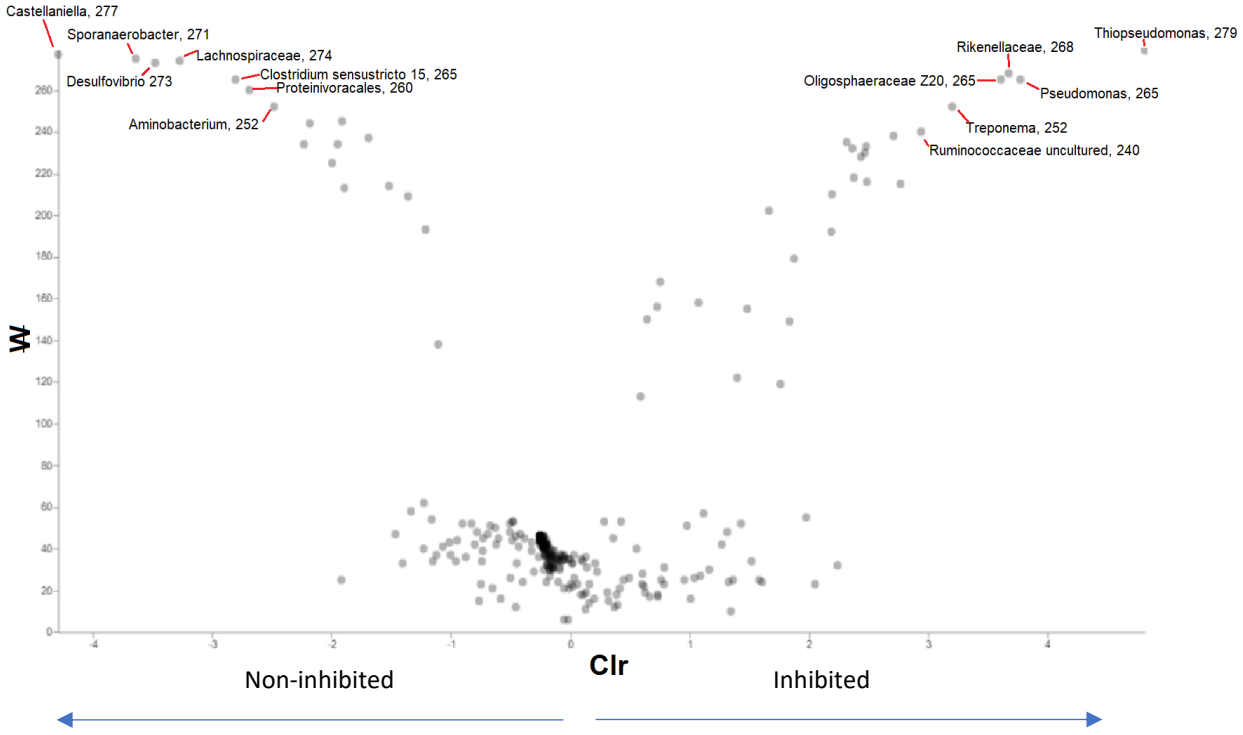


Figure 1 - ANCOM Volcano Plot. The clr (centered log ratio) transformed feature table at the genus level was used. The W value represents the number of times the null-hypothesis (the average abundance of a given species in a group is equal to that in the other group) was rejected for a given species. The x-axis value represents the clr transformed mean difference in abundance of a given species between the inhibited and non-inhibited reactors. Potential biomarker taxa samples are labelled along with their W value.

The ANCOM test was conducted to reduce the relative abundance problem associated with the taxonomic bar graph summary (Fig.2). Fig.5 shows there are genera which are statistically significant and solely abundant in inhibited or un-inhibited AD. Table 2 is a summary of the top 15 taxa with the highest W scores. These taxas could be used as potential biomarkers to indicate the health of the AD. The only taxa which was visually hypothesised to be significant from Fig.2 was *P. Thiopseudomonas*.

Taxa	₩	Inhibited
Pseudomonadaceae Thiopseudomonas	279	yes
Rikenellaceae	268	yes
Oligosphaeraceae Z20	265	yes
Pseudomonadaceae Pseudomonas	265	yes
Spirochaetaceae Treponema	252	yes
Ruminococcaceae uncultured	240	yes
Alcaligenaceae Castellaniella	277	no
Peptostreptococcales- Tissierellales Sporanaerobacter	275	no
Lachnospiraceae	274	no
Desulfovibrionaceae Desulfovibrio	273	no
Clostridiaceae Clostridium sensu stricto 15	265	no
Pseudomonadaceae Proteinivoracales	260	no
Synergistaceae Aminobacterium	252	no
Peptostreptococcales-Tissierellales uncultured	245	no
Clostridiaceae Clostridium sensu stricto 1	244	no

Table 2 - Biomarkers of Inhibited and Un-inhibited reactors obtained from the ANCOM test.

Conclusion

This paper demonstrates the results of a metataxanomic analysis of industrial-scale AD reactor samples from the WWTP that was loaded with increasing ratios of feedstock.

The alpha diversity study of the two groups of reactors, uninhibited and inhibited, shows that the uninhibited samples are more diverse than the inhibited samples. The decreased production of biogas in inhibited reactors may be associated with the decrease of diversity of its microbial community. A prolonged exposure to high levels of feedstock may decreased the diversity of reactors, thus decreasing biogas production as seen in the decrease diversity from day 1 to day 2 groups. Large inhibited feedstock ratios (1/2 and 2/3) had the lowest diversity compared to the other feedstock ratios. Overloading a reactor over a long period of time will cause inhibition and lower production of biogas.

There are clear taxonomic differences between the inhibited and uninhibited reactors as shown in the beta diversity. Not only are inhibited distinct from the uninhibited reactors, but they are also quite different among themselves, this highlights the instability caused by high-feedstock ratios. Based on observation, there seems to be a feedstock threshold which permits inhibited reactors to become more like the un-inhibited reactors as time progresses, this would be an interesting area for further investigation.

Various taxonomic markers were identified. These taxa, if present in reactors, could indicate the health of the reactor. Species such as (but not limited to) *Pseudomonadaceae Thiopseudomonas, Rikenellaceae, Oligosphaeraceae Z20, Pseudomonadaceae Pseudomonas, Spirochaetaceae Treponema and Ruminococcaceae uncultured* can indicate that a reactor may be overloaded. Other taxa were associated with uninhibited reactors but were not statistically significant. Similarly, there are microbial species which are associated with healthy AD reactors, these are found in the result section.

AD reactors are sensitive to organic overloading. Reactor's alpha and beta diversity are shown to change with various feedstock loads, and essentially cause changes in biogas production. Microbial populations can be analysed and compared to the biomarkers identified within this paper to determine the health of the AD reactors.

Bibliography

1. NNFCC. Anaerobic Digestion | The Official Information Portal on Anaerobic Digestion. https://www.biogas-info.co.uk/ (2021).
2. Campanaro, S. *et al.* The anaerobic digestion microbiome: a collection of 1600 metagenome-assembled genomes shows high species diversity related to methane production. *bioRxiv* 680553 (2019) doi:10.1101/680553.
3. Li, L., Peng, X., Wang, X. & Wu, D. Anaerobic digestion of food waste: A review focusing on process stability. *Bioresource Technology* **248**, 20–28 (2018).
4. Sun, L. *et al.* Lentimicrobium saccharophilum gen. nov., sp. nov., a strictly anaerobic bacterium representing a new family in the phylum Bacteroidetes, and proposal of Lentimicrobiaceae fam. nov. *International Journal of Systematic and Evolutionary Microbiology,* **66**, 2635–2642 (2016).
5. Kim, E., Lee, J., Han, G. & Hwang, S. Comprehensive analysis of microbial communities in full-scale mesophilic and thermophilic anaerobic digesters treating food waste-recycling wastewater. *Bioresource Technology* **259**, 442–450 (2018).
6. Huang, Y. *et al.* Isolation and characterization of Keratinibaculum paraultunense gen. nov., sp. nov., a novel thermophilic, anaerobic bacterium with keratinolytic activity. *FEMS Microbiology Letters* **345**, 56–63 (2013).
7. Monserrate, E., Leschine, S. B. & Canale-Parola, E. Clostridium hungatei sp. nov., a mesophilic, N2-fixing cellulolytic bacterium isolated from soil. *International Journal of Systematic and Evolutionary Microbiology,* **51**, 123–132 (2001).
8. Ince, O., Akyol, Ç., Ozbayram, E. G., Tutal, B. & Ince, B. Enhancing methane production from anaerobic co-digestion of cow manure and barley: Link between process parameters and microbial community dynamics. *Environmental Progress & Sustainable Energy* **39**, 13292 (2020).
9. Yokoyama, H., Wagner, I. D. & Wiegel, J. Caldicoprobacter oshimai gen. nov., sp. nov., an anaerobic, xylanolytic, extremely thermophilic bacterium isolated from sheep faeces, and proposal of Caldicoprobacteraceae fam. nov. *International Journal of Systematic and Evolutionary Microbiology,* **60**, 67–71 (2010).
10. de la Torre, I., Ladero, M., Link to external site, this link will open in a new window & Santos, V. E. Production of d-lactic acid by Lactobacillus delbrueckii ssp. delbrueckii from orange peel waste:

techno-economical assessment of nitrogen sources. *Applied Microbiology and Biotechnology* **102**, 10511–10521 (2018).
11. Crow, V. & Curry, B. LACTOBACILLUS spp. | Lactobacillus delbrueckii Group. in *Encyclopedia of Dairy Sciences* (ed. Roginski, H.) 1494–1497 (Elsevier, 2002). doi:10.1016/B0-12-227235-8/00243-1.
12. Tan, W.-B. *et al.* Thiopseudomonas denitrificans gen. nov., sp. nov., isolated from anaerobic activated sludge. *Int J Syst Evol Microbiol* **65**, 225–229 (2015).
13. Knight, R. *et al.* Best practices for analysing microbiomes. *Nat Rev Microbiol* **16**, 410–422 (2018).

Supplementary Data (NOT COUNTED FOR WORD COUNT)

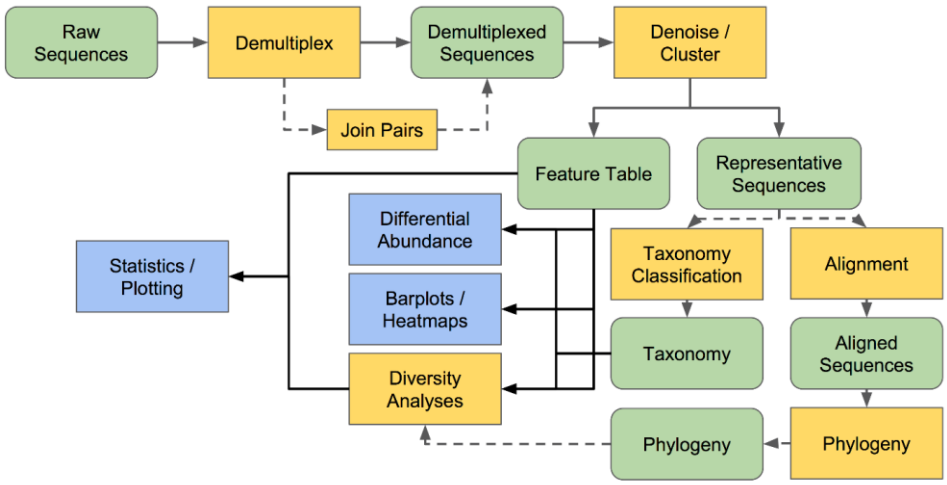


Figure 1- Workflow for Qiime2 for examining amplicon sequence data from the 16S rRNA gene[13].

Sample ID	SampleDay	FeedstockRatio	Inhibition	Sample ID	SampleDay	FeedstockRatio	Inhibition
#q2:types	categorical	categorical	categorical	#q2:types	categorical	categorical	categorical
AD1	0	NA	NA	AD27	2	0	No
AD2	0	NA	NA	AD28	2	0	No
AD3	1	0	No	AD29	2	1/200	No
AD4	1	0	No	AD30	2	1/200	No
AD5	1	1/200	No	AD31	2	1/150	No
AD6	1	1/200	No	AD32	2	1/150	No
AD7	1	1/150	No	AD33	2	1/100	No
AD8	1	1/150	No	AD34	2	1/100	No
AD9	1	1/100	No	AD35	2	1/75	No
AD10	1	1/100	No	AD36	2	1/75	No
AD11	1	1/75	No	AD37	2	1/50	No
AD12	1	1/75	No	AD38	2	1/50	No
AD13	1	1/50	No	AD39	2	1/25	No
AD14	1	1/50	No	AD40	2	1/25	No
AD15	1	1/25	No	AD41	2	1/15	No
AD16	1	1/25	No	AD42	2	1/15	No
AD17	1	1/15	No	AD43	2	1/10	No
AD18	1	1/15	No	AD44	2	1/10	No
AD19	1	1/10	No	AD45	2	1/5	No
AD20	1	1/10	No	AD46	2	1/5	No
AD21	1	1/5	No	AD47	2	1/2	Yes
AD22	1	1/5	No	AD48	2	1/2	Yes
AD23	1	1/2	Yes	AD49	2	2/3	Yes
AD24	1	1/2	Yes	AD50	2	2/3	Yes
AD25	1	2/3	Yes				
AD26	1	2/3	Yes				

Table 1 - Experimental set up of samples from anaerobic digestors with varying ratios of feed stock. AD1 is the Seafield full-scale digestate sample control and AD2 us the feedstock samples control. The feedstock is an organic matter from sewage sludge.