Sequence based analysis of microbial communities in anaerobic digestion reveal key organisms and rate limiting hydrolysis pathways in the methanisation of whisky distillation waste

James Robson

Doctor of Philosophy

University of York

Biology

September 2019

Abstract

Anaerobic digestion is the breakdown of an organic material into a methane rich biogas through the action of a complex microbial community. This research primarily uses sequence based technologies such as gene amplicon sequencing, qPCR, metagenomics and metatranscriptomics to investigate population composition, dynamics and functionality in these communities when fed a liquid waste product from malt whisky distillation termed pot ale.

Initial batch tests complemented with gene amplicon sequencing explored the potential of used wastewater dewatering polymer, a cationic polyacrylamide, and a community inoculum. Communities produced from this material showed production of biogas and methane and increased abundance in carbohydrate degrading organisms such as Bacterodia. Poor methane yields could be attributed to the inhibitory effects of the acrylamide containing inoculum so subsequent digestion of pot ale used an alternative inoculum. To quantitatively determine how the microbial community was changing, a microbial spike-in method was developed. This spike-in method involved characterisation of an organism not typically found in anaerobic digestion, *Sulfolobus solfataricus*, and showed to normalise DNA extraction procedures in addition to population quantification.

Further pot ale anaerobic digestion was performed in a custom-built lab scale system over a period of six months. This revealed that methane production from pot ale occurs in two distinct stages, the first fast and the second slow. DNA samples were collected over this time and metagenomic sequencing was used to reconstruct key metabolic pathways which revealed a functionally diverse, and functionally robust community. Genes specifically related to the two stage degradation of pot ale digestion were investigated using metatranscriptomics which revealed initial methane production was caused by hydrolysis of malto-oligosaccharides by Clostridia followed by the second, rate limiting, hydrolysis of beta-glucans by Bacterodia. By identifying this rate limiting enzymatic hydrolysis step, this opens the door to feedstock specific enzyme supplementation to increase digestion efficiency.

Table of contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisor Professor James Chong for allowing me the freedom to explore the microbial communities in anaerobic digestion as I saw fit and introduce me into the promising area of renewable biogas. The support and advice offered over the four year PhD was invaluable.

I would like to thank the lab group of my co-supervisor, Professor Gavin Thomas, for enjoyable social evenings and events.

The assistance of Dr Kelly Redeker was essential in setting up the gas chromatography. It was a long bumpy road, but we got there eventually.

Many thanks to Dr Daniela Barilla for providing the *Sulfolobus solfataricus* used in the development of a spike and emotional support when needed.

The fantastic help from Dr Sally James and Dr Lesley Gilbert at the Genomics Facility, University of York in preparing sequencing libraries and MinION sequencing.

The two summer students Monica Patel and Luna Yuan very kindly screened primers used in the microbial spike development.

My bestest friend Aritha Dornau for also having the terrible idea of working with waste material, without that we would never have met and York would have been a lot more boring. I also want to thank her three really cute snakes and her making sure lunch is eaten and gainz received. Also for knitting me a very nice hat. And bringing me some nice rocks. And being a beer snob.

Lots of love to my family; Alison, Martin, Sophia & Islay for always being a steadfast rock.

Lots of love to all my friends, especially the Edinburgh Gribble Squad. Their constant high energy mountain rules ensured that I finished all my whisky and cigarettes in an athletic, dynamic and decisive manner.

Thank you to the greatest lab mates ever; Kim Barnes, Annie Cansdale and Nathan Innard for sometimes great, but usually terrible chat.

My final and most grateful thanks to Armin van Buuren, Giuseppe Ottaviani and Aly & Fila for pushing my energy higher and focusing my mind into a free flowing productive trance state.

Author's declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1

Introduction

1.1 Current energy landscape

1.1.1 Fossil fuels and their relationship with Earth.

The industrial revolution brought unprecedented levels of productivity and manufactured goods to countries all over the world. Quality of living increased as goods previously considered luxury were now available for all to buy at a fraction of the cost. Steam trains allowed the transportation of food and goods to remote locations, and a well-deserved holiday to new places. This revolution was powered entirely by the burning of coal, a combustible carbon rich rock from the fossilised remains of plant material. Technology at the time meant that coal was only partially combusted, producing large quantities of soot, airborne amorphous carbon containing polycyclic aromatic hydrocarbons (PAHs) which are known mutagens. In addition to toxic carbon compounds, characteristic heavy metals in soot such as mercury and arsenic pose danger to human health.

Nowadays, although coal is still typically used in stationary power stations, global focus has shifted from coal towards oil, which is becoming the fossil fuel icon of the 21st century. Like coal, oil is created from organic remains of plants and organisms only under extreme pressure which creates the hydrocarbon rich liquid. The widespread popularity can be mainly attributed to the rise of plastic and automobiles, with many families owning more than one car. Just like with coal, the new fossil fuel increased living standards once again, and cheap polymers provided everyday objects and garments while the fuel fraction allowed personal travel on an unparalleled scale with the spread of the liquid petroleum engine.

The third major fossil fuel is natural gas, thought to have been used and transported as early as 500BCE in China, which is primarily methane and other alkanes derived from organic material under extreme temperature and pressure. For personal use, the natural gas is doped with an odourant as in its pure state is odourless. This doped gas is then transported directly into homes to be used for cooking and heating. Without a doubt, fossil fuels have transformed and advanced the world at an incredible speed. However, the effects of fossil fuel associated natural disasters have had devastating effects on cities, countries and the entire world.

Coal caused the Great Smog of London. Lasting from 5th-9th December 1952, this resulted in 32,000 tonnes of sulfuric acid and 4,000 tonnes of soot being condensed in London resulting in the direct deaths of 4,000 individuals and many more long term health problems. More recently, China fulfils 70% of its energy demand using coal based power stations leading to major pollution in mega-cities such as Beijing. Pollution associated with coal usage is now the leading cause of death in China (metoffice.gov.uk, 2017; Zhang et al., 2008).

On the 20th April 2010 the Deepwater Horizon disaster resulted in catastrophic marine pollution. An explosion occurred on the drilling rig following a high pressure release of gas which engulfed the entire platform. An

estimated 794 million litres of crude oil was released from the reserve causing major ecosystem devastation. Post spill water was found to have 40 times more polyaromatic hydrocarbons (PAHs) , in 2012 50% of shrimp in the area lacked eyes, and lesions in fish increased from 0.1% to 50%. The lack of accident planning was highlighted by 7 million litres of mutagenic, toxic and endocrine-disrupting experimental dispersant Corexit being used to remediate the spill. The addition of this untested chemical in an unplanned setting increased the toxicity of the oil by more than 52 fold (Jamail, 2012; Rico-Martínez et al., 2013).

In the year 2016 another lack of accident planning was highlighted by the world's largest natural gas leak. In the mountains above Los Angeles an estimated 100,000 tonnes of natural gas was released over a 3-month period presenting a larger carbon footprint than the Deepwater Horizon disaster. Although natural gas contains less pollutants than other fossil fuels, methane is a greenhouse gas 25 times more potent than carbon dioxide (Appels et al., 2008). Although the ecological effects of this disaster are not as severe as the previously mentioned ones, this event represents the most concentrated pollution event in history.

Because of these terrible fossil fuel disasters, in conjunction to their high carbon and pollution footprints (Table 1.1), countries around the world are striving to increase their usage of renewable energy (Figure 1) and distance themselves from "dirty" fossil fuels such as oil and coal in preference of cleaner natural gas. Ranked by global investment, the top three main sources of renewable energy in the year 2015 were solar ($161 BN), wind ($110 BN) and biomass waste to energy (WTE) ($6 BN) (FS-UNEP, 2016). WTE plants also prevents waste material being sent to costly, ugly and space wasting landfill.

Pollutant (Kg)	**Natural Gas**	**Oil**	**Coal**
Carbon Dioxide	15,795	22,140	28,080
Carbon Monoxide	5.4	4.55	28.08
Nitrogen Oxides	12.42	60.48	61.695
Sulfur Dioxide	0.135	151.5	349.8
Particulates	0.945	11.34	370.5
Mercury	0	0.000945	0.00216

Table 1.1 Fossil Fuel Emission Levels. Kg of pollutant per Watt. Modified from (Energy Information Administration 1999)

Emissions from different types of fossil fuels are provided which aim to highlight that natural gas is the least polluting out of the three with oil placing second in terms of pollutants. The least environmentally friendly fossil fuel is coal.

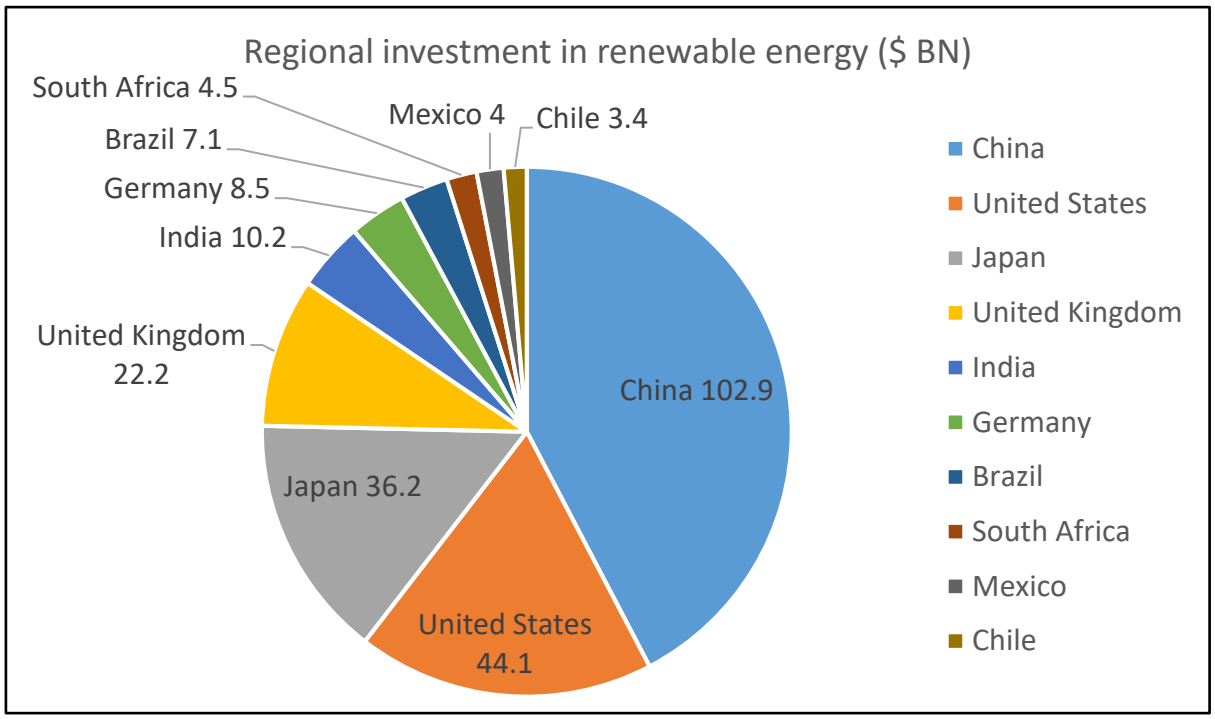


Figure 1.1 Annual regional investment in renewable energy for the year 2015. Modified from (FS-UNEP 2016)

In the year 2015 over $200 BN was invested into renewable energy sources, projects and infrastructure with the majority coming from China, the United States and Japan.

1.1.2 Renewable Energy in the UK

To catalyse a more rapid transition from fossil fuels to renewable energy, the European Commission set forth 2020 energy targets for member states. Article 4 of the Renewable Energy Directive (RED 4) placed targets in multiple sectors for the UK. Targets of 12, 30 and 10% were set for the fields of heat, electricity and transport. Being the centre of the industrial revolution led the United Kingdom to invest heavily into using the natural fossil fuel resources, thus any policy change will be a regression from the mean. Before the implementation of RED 4 in 2009 the UK generated far less from renewable sources compared to other EU member states. In 2005 only 1.5% of the total energy consumed came from renewable sources. However, this rose to 2.5% in 2009 and further still in the years following the prime directive on EU energy policy (DECC, 2009).

With the Industrial Revolution having roots in the UK, energy and novel energy innovation has always been a major part of both the past and the present. Recently the UK generated its energy through various sources; coal (22.6%), natural gas (29.5%), nuclear (20.8 %) and 24.7% of its energy through renewable sources in the year 2015. The sector of renewable energy generation grew 5.6% from 19.1% to 24.7% between the year 2014 and 2015 with the majority coming from wind farms (DECC (Department of Energy & Climate Change),

2016). Within the UK, Scotland particularly has a considerable opportunity to expand the renewable energy production harbouring 23% of the total European Union (EU) wind energy resources (Street, 2006).

RED 4 set a target for renewable energy to provide a total of 15% of total energy consumption in the year 2020. This meant a six-fold increase from the 2009 levels, a challenging target with many hurdles. Included within this total level additional sectors had thresholds that needed to be met; 30% of electricity demand, 12% of heat demand and 10% of transportation demand to be met with renewables. With 24.7% of total energy consumption being powered by renewable sources, the UK is satisfying this sector through wide scale implementation of wind power. The UK is set to achieve, or exceed, that total 30% goal being currently three-quarters of the way there. However, it is not without major developments will the UK fulfil the subsections of the RED 4 agreement, mainly heat and transport since current levels are less than half towards the target. Whilst the public transport sector is seeing major investment into fleets of electric vehicles, the biggest challenge is the decarbonisation of the heating sector; the reliance on natural gas (DECC, 2009; The Energy and Climate Change Committee, 2016).

Domestic total fuel consumption (solid, electrical and gas) has increased steadily by approximately 2.5% per year (Department for Bussiness Energy & Industrial Stratergy, 2016). Part of this is due to steady population increase but can equally be attributed to what is considered an acceptable level of warmth within the home. Although modern homes have carbon efficiency measures such as; insulation, double glazing and advanced boilers to keep homes warmer, 66% of domestic fuel consumption is gas heating since it is cheaper vs electricity (Department for Bussiness Energy & Industrial Stratergy, 2016). The gas boiler is expected to be an integral part of 75% of homes by 2020 with extensive infrastructure to support them (Kavanagh, 2013). For these reasons, gas will remain the primary fuel source for the near future and efforts must focus on creating a renewable supply of fuel gas. Anaerobic digestion is the most promising industry for increasing this supply of renewable gas and works by producing a methane rich bio-gas from biological catalysed decomposition of organic waste. It is this specific waste to energy technology that will reduce dependence on fossil fuels.

1.1.3 The green energy lie

One rapidly expanding “renewable” energy source is the bioenergy sector which showed an increase of 28% between 2014 and 2016. The term bioenergy covers technologies such as fuel crops, biomass plants and anaerobic digesters.

The concept of fuel crops involves production of ethanol or biodiesel, from crops such as sugar beer, wheat and sugar cane or rapeseed, soya or palm oil respectively. The effects of using agricultural land for the

cultivation of fuel crops have been intensely debated and there is general rejection of the concept of exclusive fuel crop cultivation, due to slightly increased food prices significantly effecting low-income families, in favour of co-growth or rotational growth (Comittee, 2013; Tenenbaum, 2008). In addition, successful fuel crops in other countries, such as corn ethanol in the US or sugarcane from Brazil, is due to excessive production on surplus land allowing for export and alternative usage (Tenenbaum, 2008). The UK imports much of its food, so does not have this excess to handle, and as such bioenergy from energy crops is unrealistic and unwise.

Biomass has historically been used as a fuel by mankind in the form of wood. Today, the European Parliament defined biomass in terms of bioenergy to include material such as vegetable matter, forestry or animal products (Wielgosiński et al., 2015). Biomass incineration plants use biomass as a substitute for coal to be considered renewable with zero CO_2 emissions under this definition. However the burning of "clean" or "urban" wood has shown to emit heavy metals in addition to other pollutants such as particulate matter, nitrogen oxides, carbon monoxide and other hazardous pollutants (PFPI Partnership for Policy Integrety, 2011). Emission guidelines for biomass plants are typically more relaxed compared to their fossil fuel counterparts as they fall under the "renewable" category. Although there is severe lack of data regarding biomass plant emissions, one such example of this exists for the Gainesville Renewable Energy Centre (GREC) in Florida, a modern facility constructed in 2013. By comparing the emissions of this facility to a nearby gas/diesel plant, the Pioneer Valley Energy Center (PVEC) in Massachusetts, in terms of emissions per MWh the results show that biomass plants are far from zero emission and indeed a health hazard (Table 1.2).

Pollutant	GREC biomass **kg/MWh**	PVEC gas/diesel **Kg/MWh**	Fold increase of **biomass pollutant**
Nitrogen oxides	0.431	0.023	19
Carbon monoxide	0.739	0.014	54
Particulate matter	0.259	0.014	19
Sulfur dioxide	0.254	0.005	56
Volatile organic compounds	0.082	0.005	18
Hazardous air pollutants	0.027	0.001	20
Carbon dioxide	1276	344	4

Table 1.2 Emission comparison of a biomass and gas/diesel plant. (Modified from (PFPI Partnership for Policy Integrety 2011)

Emissions from existing fossil fuel plants located in the United States highlight the high degree of pollution output by biomass plants when compared to gas and diesel alternatives in the year 2013.

Emission data of seven common biomass sources have also been compared to coal in a laboratory setting (Wielgosiński et al., 2015). Wielgosiński confirmed the that the levels of organic compound (nitrogen oxide, carbon monoxide and total organic carbon) emissions from combustion of major biomass feedstocks were

consitstantly significantly higher under varying combustion temperatures and air flow rates (Wielgosiński et al., 2015). These studies conclude that the combustion of biomass is far more polluting than traditional fossil fuels such as coal, natural gas or oil and their standing within the term bioenergy is dubious at best. From a thermodynamic point of view, combustion of complex organic material into gaseous waste products bleeds the energy that was input to create these complex biological polymers. Furthermore, combustion turns these biologically valuable elements such as nitrogen and sulfur into atmospheric pollutants. Ideally, at the end of the energy generation process, carbon would be transformed into a fuel and elements such as nitrogen and sulfur would remain in bioavailable forms to "close the loop" and return unwanted by-products back into the environment. Anaerobic digestion addresses both these challenges and creates both a methane rich fuel gas and fertiliser by-product from the biological decomposition of organic matter. Moreover, atmospheric pollutants like nitrogen oxides, airborne heavy metals and soot particles are not created from this process.

1.1.4 Waste and landfill in the UK

In addition to biomass crops used for combustion in biomass plants, domestic households in the UK throw away over 7.2 million tonnes of food and waste every year (Department for Environment Food and Rural Affairs, 2015). Preventing this waste in the first place would be ideal, but vast amounts from both the domestic and commercial sector will realistically be produced. Careful choice of how this biomass is treated will reduce reliance of landfill, where methane, a powerful greenhouse gas, is vented into the atmosphere unless capped. Government led initiatives to collect organic waste include introduction of designated bins for garden and food waste which is collected to be converted into compost, although composting sites can be the root of public complaints associated with the unpleasant odours generated by the degrading biological material primarily caused by terpenes, alcohols, ketones, sulphur compounds and amines (Lichtfouse et al., 2015). Apart from public distress due to released vapours, composting may not be as eco-friendly as initially thought since large amounts of CO_2 and other greenhouse gases such as NO_x are released into the atmosphere. This is especially the case when waste has a low carbon to nitrogen ratio, a high water content and does not have air pumped through the material (Lichtfouse et al., 2015). In a study of four industrial scale composting facilities, they were shown to emit between 0.16 and 4.37 kg CH_4 and between 0.075 and 0.251 kg N_2O per tonne of organic waste (Colón et al., 2012). These pollutants and emissions are due to the open nature of composting where degradation products are allowed to escape into the surrounding atmosphere.

The Landfill Directive (1999/31/EC) lay guidance for European member states to not only reduce the volume of waste sent to landfill but set guidelines for % tonnage of biodegradable municipal waste (BMW) sent to landfill. These requirements state that in 2010 no greater than 75% of BMW should be sent to landfill and no greater than 35% in 2020 (Defra, 2016). In addition to overall municipal waste to landfill decreasing from 38,001 tonnes (12,983 of which BMW) in 2010 to 22,940 (7,682 of which BMW) in 2015 the fraction of BMW

to landfill has been gradually decreasing at a rate of 3% per year to the 22% BMW to landfill level in 2015 (Defra, 2016). This represents a major achievement in that potential greenhouse gas polluting materials are no longer being sent to landfill and are instead utilised in a more controlled way. However one of the major reductions in BMW sent to landfill in Wales was caused by a major energy-from-waste combustion plant becoming fully operational. Their status as truly green energy has previously been discussed.

Landfill is the last resort for dealing with waste as it represents the final stage for any waste upgrading or recycling measures. A total of 202.8 million tonnes of waste was generated by the UK in 2014 with over half (59%) coming from the construction, demolition and excavation industries (Defra, 2016). Of that 202.8 million tonnes, a large proportion (44.5%) was recovered either through recycling or energy recovery measures with only 23% sent to landfill (Defra, 2016). Biological waste sent to landfill will emit methane into the atmosphere, thus efforts to reduce the volume of biodegradable municipal waste (BMW) have resulted in a decrease of 13,00 tonnes in 2010 to 7,500 in 2015 (Defra, 2016). This major change has been facilitated through government efforts to reduce waste in the food and drink sectors in addition to waste treatments such as composting becoming a popular method to add value to BMW (Department for Environment Food and Rural Affairs, 2015).

To reduce burden on landfill the national government and regional councils have promoted composting technologies to recycle domestic food waste, either at home or at a central facility (Leeds Council, 2019; North Yorkshire Council, 2019; Northern Ireland Environment Agency, 2012). This process creates a compost suitable for agriculture, horticulture, forestry and land restoration efforts. However, many aspects of composting contribute negatively to the environment such as energy input required, greenhouse gas emissions (including methane, nitrous oxides and methane) and unpleasant odours. Although gasses can be scrubbed to remove odours and capture carbon this can increase energy consumption by almost 50% (Colón et al., 2012). The two dominant methods of composting exist which are; confined in-vessel composting, which allows capture of gaseous products, and exposed windrow composting where material is piled into rows. Table 1.3 highlights the comparison between commercial composting practices, home practices and AD in terms of energy input/outputs (Colón et al., 2012). Although overall less compost is produced by anaerobic digestion the nitrogen content of digestate (from AD) is between 3-14% compared to the 1-3% of composts (Bolland and Gilkes, 1990; Möller and Müller, 2012).

Facility		In-vessel composting	Windrow composting	Home composting	Anaerobic digestion
Inputs	Total MJ (electricity + diesel)	871.9	579.24	33.77	472.26
	Total m^3 water required	0.56	0	0.0051	0.12
Outputs	m^3 leachate	n/a	n/a	n/a	0.03
	m^3 biogas condensates	n/a	n/a	n/a	0.05
	kg NH_3	0.11	2	0.84	0.23
	kg VOC	0.36	6.22	0.56	0.86
	kg N_2O	0.075	0.076	0.676	0.035
	kg CH_4	0.34	1.68	0.16	2.39
	t compost	0.1	0.52	0.25	0.03
	t refuse	0.13	0	0	0.41
	m^3 biogas	n/a	n/a	n/a	98.9
	Electricity MJ	n/a	n/a	n/a	717.12

Table 1.3 Greenhouse gas production and energy balance of composting and anaerobic digestion practices

Total energy expenditure, pollutants and energy yield of multiple compositing strategies compared to anaerobic digestion

1.1.5 Anaerobic digestion as a solution

By the year 2025 it is predicted that urban residents in the world will produce solid waste in excess of 2.2 billion tonnes per year (Mundial 2014). Limited landfill space means alternative disposal methods are needed to cope with excessive volume of waste. Anaerobic Digestion (AD) offers an opportunity to convert the 53% biodegradable fraction into a source of heat, electricity and money which saves space in expensive landfill (Themelis, 2002). The AD process involves the combined actions of a microbial community to degrade solid and liquid organic waste into methane rich gas and nitrogen rich solid used as fertiliser. Methane production occurs naturally in wet oxygen free environments such as marshes, lake sediments, coal seams and in the guts of ruminants (Beckmann et al., 2015; Bräuer et al., 2006; Highfield et al., 2009; Kotsyurbenko et al., 2007; Sizova et al., 2003). Humans have long artificially created such environments to harness these gas producing organisms by sealing cattle dung and organic waste together to generate a gas that could be captured and harnessed. In the 10th century BC, Assyrians used this gas to heat water baths and modern rural off-grid communities such as in India and China use AD to process waste and generate fuel and electricity (Ostrem, 2004).

Green energy subsidies have helped the biogas industry rapidly expand to provide a renewable energy source and reduce the volume of waste going to landfill. Biogas produced from AD is a mixture of 50-75% methane (CH_4) and carbon dioxide (CO_2) with contaminants such as hydrogen sulfide (H_2S) and water vapour (Ostrem 2004). By removing corrosive H_2S and unwanted CO_2, a methane rich gas (>95% CH_4) is produced in a process termed biogas upgrading. Upgraded biogas can either be used on-site for heat and electricity or injected directly into the natural gas grid. As infrastructure already exists for transport and usage of natural gas, biogas

is an attractive source of renewable energy since it avoids problems of storage, transport and usage associated with competing alternative fuels such as hydrogen (H_2) and methanol. However poor understanding of the biological processes and lack of operator training means system failure is commonplace and presents a hurdle to wide scale acceptance. This failure rate is exacerbated as modern focus areas of anaerobic digestion are moving away from typical biomatter degradation and transitioning into an industry using anaerobic digestion as a means to neutralise and detoxify challenging feedstocks such as pharmaceutical, high protein and metal rich waste (Holm-Nielsen et al. 2009). To make anaerobic digestion a more competitive industry, research must be undertaken to understand the composition, function and relationships within of the AD microbial population and investigate the value of alternative compounds that could be produced such as rubber precursors.

1.2 What is anaerobic digestion

1.2.1 Stages of anaerobic digestion

Methane production is facilitated by the action of a complex microbial community which collectively break down organic matter into a methane rich gas though their combined enzyme activity. A wide variety of feedstocks can be used such as vegetable foliage, pig slurry or corn husks thus making anaerobic digestion an inexpensive solution for waste remediation. The fundamental principle of anaerobic digestion is the anoxic incubation of organic matter with a bacterial community source containing methane producing methanogens (such as cow manure). Using cows for a source of methanogens is used worldwide especially in rural villages in countries like India and China for providing a source of gas for cooking. In this rural scenario the digester is typically a subterranean metal drum with an inlet for organic matter and a gas outlet. Larger more complex industrial systems can consist of multiple vessels that are heated, stirred and closely monitored for optimum operation. Waste particles are colonised by bacteria which secrete enzymes nearby to degrade long biological polymers into smaller products such as acetate, H_2 and CO_2. Methanogenic *Archaea* use these three molecules to generate methane, with the majority coming from acetate based (acetoclastic) methanogenesis and the remainder from combining CO_2 and H_2 (hydrogenotrophic methanogenesis). The sequential degradation of biological matter can be divided into three main stages; hydrolysis, acidogenesis and methanogenesis (Figure 1.2).

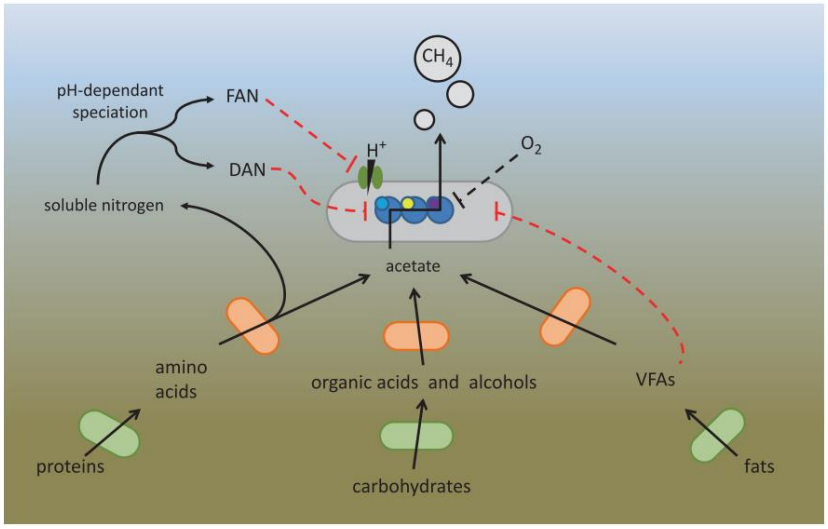


Figure 1.2 Stages in anaerobic digestion

Anaerobic digestion microbial communities are comprised of hydrolytic bacteria (green), acidogenic bacteria (orange) and methanogenic archaea (grey) which, in combination, degrade bio-polymers such as proteins, fats and carbohydrates to produce methane. Enzyme complexes used in methanogenesis (blue) require trace metals such as cobalt, nickel and molybdenum (coloured dots) and are sensitive to inhibition by by-products such as volatile fatty/organic acids (VFAs) or ammonium species from protein degradation. Ammonia (FAN) can inhibit enzyme sites directly or in the form of ammonium (DAN) which interferes with homeostasis. Above all else, methanogens are severely inhibited by oxygen as they lack superoxide dis-mutase genes. Figure taken from (Robson et al., 2016).

1.2.1.1 Hydrolysis

Initial degradation steps are performed by bacteria (Figure 1.3, green cells) which thrive in more acidic conditions that aids the solubilisation of organic matter and metals (Eastman and Ferguson, 1981; Ostrem, 2004). This hydrolysis step is considered to be the rate-limiting stage in the process as it typically involves liberation of sugars from fibrous cellulose rich substrates such as maize and rice straw which are noticeably more difficult to digest than fats and carbohydrates (Xu et al., 2019; Ziganshin et al., 2013). Hydrolytic organisms such as Clostridia are found throughout AD systems and excrete a variety of extracellular lipases, proteases, cellulases, amylases, lipases, etc to enzymatically digest bio-polymers into fatty acids, sugars and amino acids (Bermingham et al., 2017). Performing hydrolysis in acidic and elevated temperature conditions has been shown to improve biopolymer degradation (Cysneiros et al., 2012; Vavilin et al., 2008; Xia et al., 2014; Zhang et al., 2012).

1.2.1.2 Acidogenesis

Monomers produced from the previous step such as amino acids and sugars are further broken down into volatile fatty acids (VFAs) such as butyric, propanoic and acetic acid by acidogenic organisms (Figure 1.3, orange cells) which includes organisms belonging to *Bacilli*, *Clostridia* and *Bacterodia* classes (Z. Dai et al., 2014; de Vladar, 2012; Díaz et al., 2007; Maus et al., 2016b; Schröder et al., 1994). Eventually VFAs get broken down into acetate (sometimes called acetogenesis and considered a separate stage) where they can be used

by methanogenic archaea for acetoclastic methanogenesis. Gaseous by-products may also be produced at this stage including hydrogen and carbon dioxide which can be converted directly into methane via hydrogenotrophic methanogens, more favourable under acidic conditions, by organisms such as *Methanobacteria* (Kotsyurbenko et al., 2007).

1.2.1.3 Methanogenesis

Methane producing organisms (Figure 1.3, grey) prefer neutral conditions and are inhibited by acidic conditions, favourable for hydrolysis, reducing methane production at pH below six (Chen et al., 2008). To allow co-growth of both populations a typical anaerobic digester operates at a pH of around seven, a compromise between the two population optima. Community diversity of these two groups, methanogens and bacteria, is vast with the entire population comprising of hundreds of different species (Guo et al., 2015). Members of the methanogenic populations represent only 5-10% of the population abundance but this small fraction of methanogens are predominantly comprised of two genera; either *Methanosaeta*, which uses acetate exclusively, or *Methanosarcina* utilising substrates including methanol, methylamine, acetate, H_2 and CO_2 for methane production (Guo et al., 2015).

1.2.1.4 Syntrophy

Although not strictly a stage of anaerobic digestion, syntrophic organisms such as *Syntrophomonadaceae* facilitate the flow of electrons between Bacteria and Archaea via indirect methods, such as hydrogen or formate exchange, or direct methods using conductive minerals or cellular structures (Cruz Viggi et al., 2014; Sieber et al., 2012; Yamada et al., 2015). In addition to known high abundance organisms, many uncultured, newly discovered and syntrophic bacteria exist in anaerobic digestion communities adding additional layers of complexity to the already diverse network of microbial interactions (Treu et al., 2016).

1.2.2 Digester designs

To optimise operation, each stage of anaerobic digestion (hydrolysis, acidogenesis and methanogenesis) can be considered a separate process optimised for certain chemical environments. By using separating tanks conditions such as pH, oxygenation and temperature can be controlled independantly to allow optimum hydrolysis conditions.

The simplest digesters are constructed from a single tank where waste is added to stagnant sludge and gas collected after an anaerobic period of two to three weeks, this is common in rural communities in countries such as India (Gowda et al., 1995). One negative aspect regarding this simple design is settling of biomass that prevents proper contact between organisms and feed. Preventing organisms from settling by agitation

ensures thorough interface between waste and sludge community and prevents the formation of an acidic layer and so a stirred tank (Figure 1.3a) provides improvement at an affordable cost and is the most common form of AD system (Hansen et al., 1998; Hou et al., 2013; Kieu et al., 2011; Stitt, 2002; Ziganshin et al., 2011). Increasing stirred tank numbers to a multiple tank systems offer the advantage of each tank maintained at different pH levels to separate the acidiphilic hydrolytic bacteria from the more sensitive methanogenic species. For some easy to digest substrates this division of stages is not necessary but for difficult to hydrolyse material such as lignocellulose this allows a harsh rapid hydrolysis stage and increased gas yield (Kirtane et al., 2010; Yu et al., 2002).

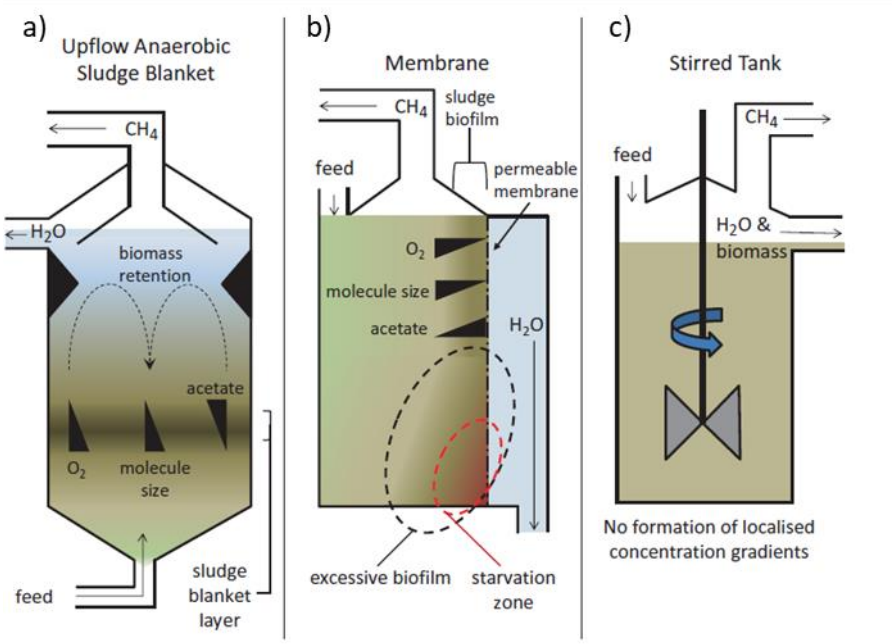


Figure 1.3 Designs of three common anaerobic digester reactor designs

a) Feedstock is fed into an upflow anaerobic sludge blanket reactor (UASBR) from the base to create a localised concentration gradient of nutrients within the system. The sludge blanket contains spatially organised microbes in either a 2D blanket or a 3D sphere with hydrolytic organisms situated on the feedstock-side surface. This spatial arrangement prevents methanogens being exposed to trace amounts of oxygen present in the incoming feed. As methane bubbles form on the blanket surface, pieces of sludge, termed flocs, may rise and are captured by an inverted funnel structure before returning downwards. This channelling structure passively ensures biomass is retained in the reactor. b) Similar in principle to the UASBR, the membrane bioreactor has a biofilm grown on a membrane to provide a degree of spatial separation. Membranes need to be cleaned regularly since accumulation of excessive biofilm may form starvation zones and lead to methanogenic population collapse. c) Stirred tank reactors provide the simplest design where feedstock and community are continually mixed. A single tank systems is operated under conditions which inhibit neither hydrolytic organisms nor methanogens. Separation of hydrolytic and methanogenic processes can be achieved with the additional tanks.

1.2.2.1 Upflow Anaerobic Sludge Blanket Reactors

One major improvement to basic tank design was the development of the Upflow Anaerobic Sludge Blanket Reactor (UASBR, as seen in Figure 1.3a). This system allowed the treatment of substrates previously considered too potent for traditional anaerobic digestion and works by focusing to achieve solid:liquid phase

separation with a catalytic microbial layer distinct to the liquid. Influent is dispersed into the base of the reactor and rises through a microbial sludge blanket considered “solid”. Microbes aggregated around the produced gas bubbles rise and are captured an array of inverted troughs. Biomass aggregates then settle back into the blanket area of the reactor allowing a higher level of microbial retention compared to traditional reactors where effluent containing microbes is discharged. High performance UASBRs operating for a length of time report change in the nature of the solid:liquid phase separation with sludge forming discrete granules between 1-5 mm in size termed granulation. Granulation facilitates loading rates and performance metrics far beyond conventional systems. Increased density of the granule enhances the settling characteristics from 2 m/h to 60 m/h compared to the regular sludge blanket and the ability to accommodate high loading rates of over 50 kg COD/m^3/day promote granulation as a desirable characteristic for AD systems treating high strength liquid waste (Hulshoff Pol et al., 2004). Organisation and functional specialisation of the microbial community within granules also occur with hydrolytic organisms located in the outermost granule layers and methanogenic archaea within the centre. By spatially separating the different communities, similar in principle to a multi-tank systems, communities are exposed to different chemical environments allowing functional specialisation and increased methane production (Show et al., 2004).

1.2.2.2 Membrane Bioreactors

By further expanding on the ideas of phase separation behind the UASBR design, the membrane bioreactor (MBR) and fixed bed membrane reactor (FBMR) were developed. Both designs utilise a membrane to retain the majority of solids and microorganisms in the system, although in mechanistically different ways. The membrane in MBRs is used to filter bioreactor liquid and concentrate solids allowing for a smaller reactor footprint with a more concentrated microbial community. A concentrated community is also achieved in FBMRs by having liquid waste trickle over a microbial community grown on an inert surface or film. Furthermore, specialised membranes can be integrated to remove specific toxins in difficult to digest feedstocks such as ammonium, a by-product of protein degradation in animal products, preventing inhibition of the ammonium sensitive methanogenic community (Awobusuyi, 2016).

1.2.3 Pre-treatment Methods

Like varied reactor designs, multiple methods of feedstock pre-treatment exist to enhance digestibility and methane production for difficult to degrade feedstocks. Solid feedstocks prevent microbial and enzymatic access and significantly reduce their ability to be digested with increasing particle size. Feedstocks pre-treatment can use physical, chemical, thermal or a combination of methods to achieve faster digestion rates. As such, simple mechanical treatments like crushing, milling or slicing offer a rapid treatment for quicker digestion periods and increased biogas yields between 20-50% (Montgomery and Bochmann, 2014). ultrasonic treatment is an alternative physical pre-treatment step and utilises destructive power of cavitation

and sound waves in the 15-20 kHz range to create collapsing microbubbles with localised temperatures and pressures of up to 4000^{o}C and 180 Mpa, respectively this treatment has been used to increase levels of reducing sugars from corn slurry almost three fold (Montalbo-Lomboy et al., 2010). These physical treatments do not chemically alter the substrate so application is limited for feedstocks containing hydrolysis resistant biological polymers such as ligninocellulose. Heated chemical treatments such as alkali, acid and oxidative offer improved ligninocellulose solubility and hydrolysis rates but tend to produce inhibitory cyclic by-products such as phenolics and furfurals (Jih-Gaw et al., 1997; Pakarinen et al., 2012). An effective combination of physical and chemical treatments stages is steam or ammonia explosion whereby a heated mixture is subject to intense pressures of around 200 bar then suddenly released to rupture cells and fibres (Ostrem, 2004). Explosive decompression treatments are effective at both reducing particle size, solubilising lignocellulose and improving enzymatic access to fibres.

Gentler methods such as biological pre-treatments avoid production of these inhibiter molecules and can even act to remove toxic compounds. Aerobic fungal pre-treatment using white rot fungus is able to remove harmful cyclic compounds and toxins and increase methane production (Blika et al., 2009). Microaeration of the bacterial community has also been shown to provide a more effective hydrolysis treatment for difficult to digest materials such as polymers and straw (Fu et al., 2016; Wei and Wang, 2013). Direct addition of enzymes is also a viable choice to accelerate hydrolysis speed and can be performed under anaerobic conditions such as lipase addition to soap stock (Cherif et al., 2014). Optimal pre-treatment method is entirely dependent on feedstock composition and there is no universally beneficial choice when considering costs, even for simple treatments such as shredding.

1.2.4 Optimisation of Process Operation

Production of inhibitory intermediates is a natural part of the AD process but these intermediates should be monitored to avoid accumulation and toxicity. Advances in plant monitoring and feeding technology will decrease risks of system failure and subsequent loss of revenue. Usage of training software and part automated plant operation ensures stability of the AD process in an industry plagued by failure rates as high as 80%, in the United States, preventing mainstream process acceptance (Themelis, 2002). Most failures arise from a single event of excess feeding, or increased organic loading rate (OLR), which creates localised highly acidic pockets killing methanogenic species. Whilst it is tempting to ever increase the organic loading rate, careful monitoring of metabolic intermediates such as volatile fatty acids (VFAs) must be done to ensure no accumulation as once the community starts to die it is often irreversible. By installing an in-line VFA measurement system, feed rate can be altered in real time to provide the highest OLR without the risk of VFA accumulation. Many commercially available software tools provides instant feedback and guidance for plant operators thus minimising the chances of operator induced error and maximising biogas output. By

inputting operational parameters such as feeding rate, biogas quality and ammonia levels the software can provide dynamic risk assessment whilst maintaining plant operational efficiency. Moreover, it informs the operator what actions are needed to either reduce toxin levels or further increase gas yield. Having instant responsive feedback and remedial measures for operators would further increase the global robustness of AD. Rapid detection of catastrophic toxin levels would allow reactive treatments such as addition of granulated carbon or EDTA, for negating organic and inorganic molecule inhibition respectively, and been proven to prevent a costly system crash (Thanh et al., 2015; Xiao et al., 2015).

1.2.5 Inhibition

Anaerobic Digestion, as the name suggests, is performed under strict oxygen free conditions. Methanogens are extremely sensitive to oxygen concentrations as low as 10 ppm, in addition to an array of inhibitory metabolic intermediates such as VFA and ammonia (Botheju et al., 2010; Kato et al., 1993). Although low levels of VFAs indicate a healthy turn over and degradation of bio-polymers, gradual accumulation indicate an over active or over fed bacterial community and usually foreshadow reactor crashes. The effects of VFA inhibition are amplified as carbon chain length increases as inhibition occurs via hydrophobic interactions with cellular membranes leading to collapse of membrane potential (Angelidaki and Ahring, 1992; Cirne et al., 2007; Wang et al., 1999).

1.2.5.1 Ammonia inhibition

Digestion of protein rich substrates leads to high levels of ammonia which inhibit sensitive methanogens. Various methods are able to limit or reverse this accumulation such as struvite precipitation but the exact biological effect of excess ammonium is unknown. Ideally to offset ammonia accumulation, high protein wastes such as animal products and slaughterhouse waste can be co-digested with carbon rich sources to maintain a C:N ratio of 25:1, shown to be optimal for microbial growth (Puyuelo et al., 2011). Whilst low levels of ammonia are beneficial, excess concentrations (above 1500mg/L) are especially toxic to methanogens (Chen et al., 2008; Hansen et al., 1998; Orzi et al., 2015; Rajagopal et al., 2013). Toxicity is twofold due to mechanisms of both the dissolved ammonium nitrogen (DAN) and free ammonia nitrogen (FAN) as visualised in Figure 1.2. DAN inhibits the methanogenic enzyme complexes such as McrA directly by binding the nickel containing active site but hydrophobic FAN is thought to be the main cause of inhibition by diffusing passively across cell membranes (Rajagopal et al., 2013; Wongnate and Ragsdale, 2015). Once inside the cell, FAN immediately dissolves to alter the pH causing stress responses and increase the energy required for homeostasis. In general acetoclastic species are more sensitive than hydrogenotrophic species to FAN (Campos et al., 2008). Additional factors such as cell wall composition may play a role in determining FAN permeability but effective methods exist to control ammonia toxicity include pH reduction, addition of

zeolites, community acclimatisation and use of ammonium rejecting membranes (Awobusuyi, 2016; Montalvo et al., 2012; Zheng et al., 2015).

1.2.5.2 Alternative electron acceptors

Metals are used biologically in catalytic sites within enzymes and as electron carriers. The methane producing enzyme pathway in archaea contains enzymes requiring atypical metals like molybdenum, tungsten, cobalt and nickel (Thanh et al. 2015). Iron is involved in common electron transferring FeS clusters and is universally required by organisms. However under anaerobic conditions, if present in high enough concentrations, microbes will also use these metals as a terminal electron acceptor precipitating the metal out of solution and diverting electrons away from methanogenesis. If the aim of the AD or remediation process is to recover metals this can be beneficial. However, if the goal is methane production from high iron products such as animal blood electron diversion can present a hindrance. Addition of metal chelating agents such as EDTA can decrease the bioavailability metals such as cobalt but acts to increase availability of other such as nickel, thus understanding exact metal composition and concentration in a feedstock can provide a basis for tailored supplementation (Thanh et al. 2015).

Organisms living in low energy environments, such as AD, use a wide variety of electron acceptors such as CO_2, metals and sulfur in an extremely energy limited system. Sulfur is an element commonly in structural proteins such as hair, feathers or skin, where a strong disulfide bridge is used to link together two chains of amino acids. Oxidised sulphur compounds are present in saltwater algal biomass and their anaerobic digestion typically results in production of hydrogen sulphide which, similar to ammonia, is also more toxic in its uncharged form with methanogen inhibition occurring at concentrations above 200 mg/L (Ward et al., 2014). Indirect methanogen inhibition also occurs from sulphide by precipitating essential cobalt or nickel trace metals as metal sulphides (Paulo et al., 2015). Methane is also depleted, converted into soluble formate, in anaerobic oxidation of methane (AOM) when coupled with sulfate reduction (Knab et al., 2008).

1.2.6 Adding value to AD

Continued research into novel anaerobic digestion methods ensure the industry of converting organic waste into biogas will remain a competitive and profitable one. Currently optimising production of methane is the focus of anaerobic digestion research. However, additional value can be added to the process if bio-products such as antimicrobials, esters or other useful organic molecules are produced and purified alongside methane production. Multi-carbon molecules such as VFAs, containing between two and eight carbon atoms, provide a platform for the production of butadiene, a synthetic rubber precursor, or 1,3 propandiol (Baek et al., 2014; Biebl et al., 1999). Alternatively, biogas could be used as the platform for high value molecule production as methanotrophic organisms are able to utilise methane as their sole source of carbon producing an array of

valuable products such as protein, vitamins and pigments (Rasouli et al., 2018; Ritala et al., 2017). Methanotrophs have also been shown to produce lactate, a chemical building block used as a reactant in various industrial processes (Henard et al., 2016). Recent developments in metagenomics sequencing aid identification of novel pathways and organisms since the majority of organisms in environmental samples are unculturable under normal lab conditions (Lee et al., 2015). Production of high value molecules from AD or biogas would cement AD into the pipeline of green chemical production.

1.2.7 Summary

Anaerobic digestion is a growing industry for treating the ever increasing volume of waste that humanity produces. However there are still multiple areas to improve the technology. The increasing number of anaerobic digesters has led to the development of novel reactor designs such as the UASBR and membrane reactor. By performing computational fluid:particle simulations, further novel reactor geometries could be identified which may increase performance (Kodama et al., 2004; Ladd and Verberg, 2001; Liang, 2013). The growing industry has also facilitated the anaerobic digestion of previously indigestible material, made possible by numerous pretreatment options such as thermal hydrolysis, fungal broths and acidic incubation (Blika et al., 2009; Kirtane et al., 2010; Ward et al., 2008). Both reactor design and pretreatment method have led to increased methane production for digesters. Combining this improved performance with modern technologies such as computer assisted feeding control and in-line indicator molecule monitoring would ensure stable digestion whilst avoiding inhibiting levels of ammonium, VFAs and metals. Due to the increase in renewable energy from sources such as wind and solar, it is likely the future the value in AD lies not in methane production, but in high value molecule production. Transforming metabolic intermediates or methane into valuable molecule such as lactate, proteins or esters will allow the AD industry to disconnect from fluctuating biogas prices.

Chapter 2

Biogas potential tests using pot ale with used flocculant acrylamide polymer as the inoculum

2.1 Abstract

In this study we examine alternative uses for used water absorbing cationic polyacrylamide (cPAM) produced by the wastewater industry, if a microbial community can be cultured from this material and what potential this community may have in the field of anaerobic digestion. This inoculum is of specific interest to the industry of waste treatment as it possesses an ability to be stored between seasons and will produce a reliable community once fed. The inoculum was also subjected to storage of up to 90 days and cycled through freezing and thawing conditions to investigate its robustness in storage or transport. Alongside cPAM, we also investigated the potential of whisky distillation waste, pot ale, as a feedstock. To study the inoculum activity, samples were incubated anaerobically in biomethane potential (BMP) tests and 16S rRNA DNA analysis performed to investigate microbial community structures that arose from the material precipitated by used cPAM before and after treatment. Pairwise phylogenetic tree branch node distances showed clustering of samples in relation to their initial feeding concentration in addition to temporal changes in the microbial community.

All communities treated under different storage conditions tended towards a phylogenetically similar structure and showed methanogenic activity. We confirm that a methanogenic microbiome developed from the material affixed to the used cPAM, and that used cPAM can be used as an alternative inoculum source for seeding AD reactors and is stable under chilled and ambient temperatures for durations up to 90 days. Pot ale was successfully anaerobically digested although the poor methane yield can be attributed to the inhibitory effects of cPAM.

2.2 Keywords

cPAM, flocculant, polymer, anaerobic digestion, microbial ecology

2.3 Introduction

2.3.1 Anaerobic digestion

Anaerobic digestion (AD) is the microbial catalysed conversion of waste organic material into a methane-rich gas for use as a renewable alternative to natural gas. Methane is produced by a complex microbial community, which degrades biological polymers such as protein, fats and carbohydrates to their respective monomers which are then fermented to acetate, with additional by-products such as ammonium. The complex microbial nature of AD is highlighted by the fact that methanogenic archaea make up a minority of the community, with many of the remaining species being unculturable. This makes studying the community difficult using traditional molecular biology techniques (Guo et al., 2015). Using DNA sequencing, it is possible to i) identify species present in the AD community ii) determine their abundance and iii) map changes of specific taxa over time. By understanding which members of the microbial community thrive under specific conditions a "core community" can be identified and used to monitor reactor health (Carballa et al., 2015).

2.3.2 Cationic polyacrylamide (cPAM)

In this study we evaluate the use of used cPAM as a seed material for AD under varying feeding concentrations and inoculum storage conditions. The polymer is typically a cationic acrylamide-based material that precipitates organic matter used commercially in the dewatering of waste water or anaerobic sludge (Garcia et al., 2008). cPAM attracts microbial biomass with an unknown composition, possibly having potential for use as source of microbial inoculum due to increased stability compared to liquids and being derived from a highly diverse community containing active methanogenic organisms. Previous studies on the anaerobic digestion of cPAM has shown to inhibit methanogenesis through production of polyacrylic acid, acrylic acid and acrylamide so it is unknown if a methanogenic community can start to form from used cPAM under these toxic conditions, especially if it has been in contact with the potential toxin for an extended duration (Wang et al., 2018).

This study therefore aims to investigate the possibility of culturing a methanogenic community from toxic cPAM, either fresh or stored, by focusing on changes in the microbial ecology and methane output. Therefore, several techniques were employed in this study to mimic environmental conditions, measure gas production and assess community composition. Firstly, the anaerobic

digestion of different feedstock dosage in the presence of cPAM was investigated and methane production confirmed. Secondly, the cPAM was stored for 30, 60 or 90 days at ambient and chilled temperatures. Thirdly, cPAM was subjected to varying rounds of freeze/thaw cycles to mimic winter storage. Community structure and dynamics were investigated using 16S rRNA sequencing to identify changes in the microbial community. By identifying enriched species and confirming methanogenic activity we are the first, to the best of our knowledge, to present used flocculant polymer as an alternative seed material with improved transport and storage stability for use in the AD industry.

2.3.3 Effects of Dormancy

Dormancy is present throughout the natural environment from bears to bacteria. While the specific mechanisms are different they accomplish the same objective; by entering a decreased metabolic state, their chance of survival in harsh conditions increases. However this period of inactivity is not without cost to the organism. While bears have to stockpile fat reserves in anticipation for hibernation, microorganisms have to prepare themselves in a variety of ways. Microbial species including cyanobacteria, fungi and bacteria have distinct mechanisms for coping with environmental stresses. Cyanobacteria, for instance, form structures called akinetes which have thick cell walls, whereas fungi produce spores which can maintain viability for years and some microbes synthesise endospores in response to adverse conditions (Lennon and Jones, 2011). The transition from activity to dormancy and back can be divided into three distinct stages. Initiation starts with the detection of unfavourable conditions in the environment which can be both biotic and abiotic such as toxins or temperature. A state of dormancy is reached when a reduced level of metabolic state is reached and recovery occurs to transition into an active state.

Dormancy in bacteria can occur through two very different mechanisms, either through forming a protective spore or by reaching a lowered metabolic state with or without secretion of a protectant (Díaz-gonzález et al., 2015). While both methods can be successful to protect individuals in a community, the biological mechanisms and survival rates for each case are very different. In the case of spore formation, organisms belonging to the genus of *Baccilus* are prime examples due to their high resistance against chemical, physical and radiological attacks in addition to being considered ideal bioweapons in the case of *B. anthracis* (Driks, 2002). Essential for their high resistance is the protective coat created during sporulation called the exosporium in which the dormant cell resides until favourable environmental conditions trigger the emergences of the newly active cell (Brunt et al., 2015).

The effects of starvation have been investigated previously for aerobic waste water treatment, anaerobic ammonium oxidation (anammox) granules and for the anaerobic digestion of various wastes such as swine wastewater and tequila vinasses (Hwang et al., 2010; Jáuregui-Jáuregui et al., 2014; Xing et al., 2016). They all concluded that reactor reactivation was feasible although coupled with a slight activity decrease due to extended microbial lag times. No community structure changes were observed, although the lag-phase growth of methanogens was more obvious due to their slow growing nature (Jáuregui-Jáuregui et al. 2014). It is therefore desirable to identify an inoculum that is compatible with seasonal industries and robust enough to producing an active methanogenic community following storage.

2.3.4 Pot ale

Whisky is a vital export for the UK economy since it accounts for over 20% of all food and drink exports equating to over £4.3 bn in revenue and growing in value by 8% annually ("HMRC data shows Scotch exports hit record high in 2018," 2019, "Scotch whisky exports hit record high," 2018). With this growth comes new opportunities for sourcing novel feedstocks for Anaerobic Digestion. One such feedstock is pot ale, a combination of yeast residue and spent grain from the whisky distillation process previously considered difficult to digest it comprises of distillation residues including alcohols, spent grain and dead yeast. Like other distillation residues such as rice, agave and grape-wine, pot ale is high in biologically available carbohydrates and proteins making it an attractive feedstock for anaerobic digestion (Jáuregui-Jáuregui et al., 2014). One of the concerns from industry is the presence of copper and its effect on the microbial community, especially the methanogens. Although previous research regarding copper concentrations in pot ale revealed non-toxic levels present, between 0 and 6 ppm, the study was confined to an individual distillery and thus feedstocks must be assessed on a case-by-case basis as each distillery is a bespoke construction (Goodwin and Stuart, 1994). The whisky manufacturing process from barley and subsequent pot ale anaerobic digestion is presented in Figure 2.1.

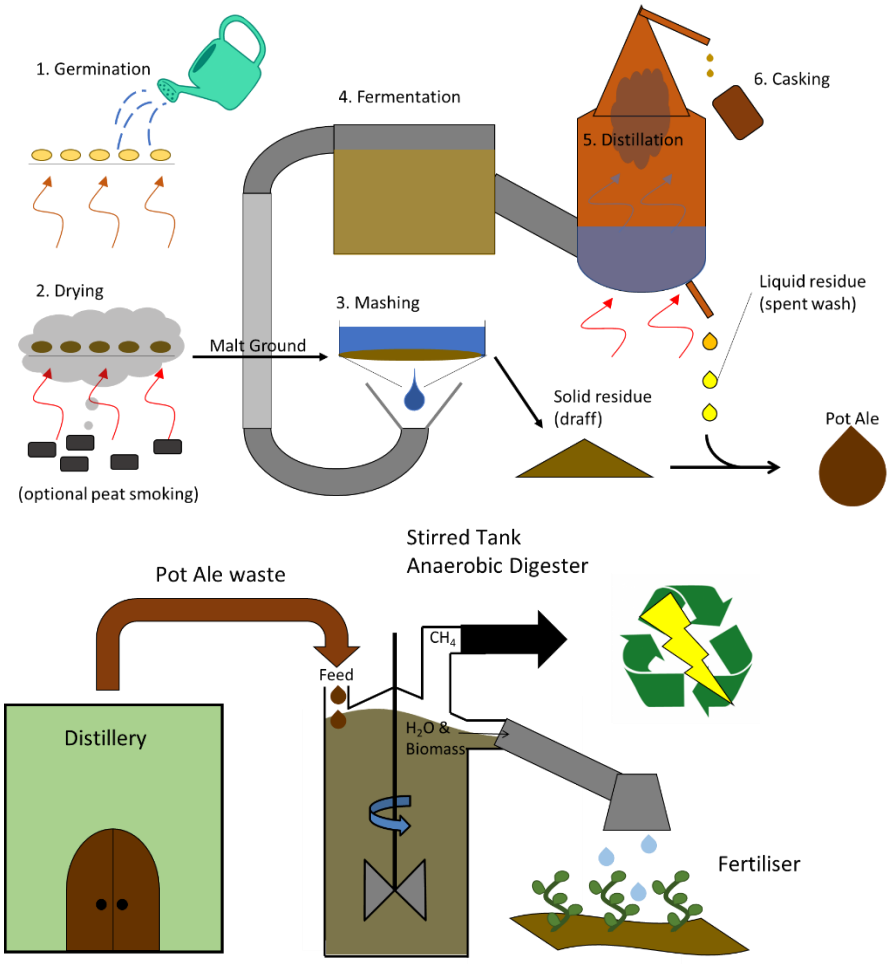


Figure 2.1 Whisky manufacturing process and subsequent pot ale anaerobic digestion

1) Barley is first soaked in water and allowed to germinate in a warm barn heated by a kiln. 2) The sugar containing malt is then dried in the kiln, no hotter than 70°C to prevent enzyme denaturation, which may be burning peat to add a smoked flavour. 3) Dried malt is ground into a course flour and mixed with batches of hot water, each hotter than the last, between 40 and 70°C to sequentially break down beta-glucans, proteins and starch forming a sugary liquid called wash and residual solids called draff. 4) Once cool, wash is transferred to a vessel where yeast is added and fermentation continuous for about 3 days to produce wash containing ~5% alcohol solution. 5) Two rounds of distillation occur; the first to separate the alcohol from the yeast and residual solids, which raises the alcohol content to ~20%, and the second to raise the alcohol content to ~65%. In this second round of distillation the initial and final fractions are typically discarded and mixed with residual yeast and solids to form pot ale. 6) Middle distillation fractions are stored in wooden casks for a minimum of three years before bottling. In this scenario, pot ale can be anaerobically digested on site to provide electricity and heating for the distillery in addition to providing digestate fertilizer for future barley growth as part of a renewable business model.

2.4 Materials and Methods

2.4.1 Inoculum sourcing, storage and seeding

cPAM was added to wastewater from a full-scale anaerobic digester fed with confectionary waste located at 55°00'44.1"N 1°39'17.7"W to bind organic matter and flocculate it from solution. Excess liquid was removed by mechanically pressing the material (now aerobic with organic material and wastewater community). To investigate effects of environmental storage, 10g aliquots were placed in 50ml falcon tubes. Spring/Summer was mimicked by storage at 4°C or room temperature. Winter conditions were simulated by freezing aliquots overnight at -20°C followed by, if applicable, defrosting for 24 hrs at room temperature. cPAM was resuspended with water at a w/w ratio of 1:3 to return it to the parental digester solids concentration, from 14% to 4% (v/v) before 25 ml aliquots were transferred to 50ml culture bottles and stoppered. Bottle headspaces were made anaerobic by flushing with 80% nitrogen and 20% carbon dioxide through a 21G needle for 30 minutes. Bottles were incubated at 35°C connected by silicon tubing (C-flex, Cole Parmer) to the gas collection apparatus described in section 7.3. Following a one-week anaerobic enrichment, after which daily gas production was zero, the microbial community was sampled and a 5 ml pot ale feed solution pot ale was added to yield total pot ale concentrations of either 5, 10 or 15% (v/v). These loading rates correspond to 3527.3, 7054.5 and 10058.2 mg COD/L respectively.

2.4.2 Feedstock and inoculum characterisation

C:N ratio of pot ale was calculated using Total Organic Carbon (#2760445) and Total Nitrogen (LCK338) cuvette test kits and COD (LCK014) commercially available from Hach Lange according to instructions. Concentrations of metals relevant to the methanogenesis enzyme pathway were quantified using an Agilent 7700x ICP-MS inductively coupled plasma mass spectrometer following digestion of 500 mg sample in 8:2 Nitric acid: H_2O_2. The Agilent 7700x ICP-MS was calibrated using serial dilutions of Agilent calibration standard 5183-4688. Liquid samples were injected into the ICP following filtration through a 22µm membrane and 8g of solid samples were first digested in 5ml HNO_3 and 5ml H_2SO_4 (Trace metal analysis grade) and heated at 200°C for one hour at 30 bar pressure until the sample was completely digested.

2.4.3 Gas collection and composition

Gas produced was captured using water displacement in a graduated 50 ml serological pipette (Sarstedt) sealed with a latex membrane (taken from a deconstructed injectable membrane male luer

lock, Vygon #891.00) glued (Loctite 401) in place to yield total volume which was recorded weekly at noon. Gas samples were taken, from the cultures fed varying feed concentrations only, through this latex barrier at atmospheric temperature and pressure using 12ml vacutainers (Labco) and double ended needles (BD Diagnostics). Gas samples were first pressurised three fold by injection of 2 x 12ml air through a 21G needle using a 20ml syringe. A 1ml sample of this pressurised gas was injected into an SRI 8610 GC system equipped with a Thermal Conductivity Detector (TCD) and a 3' x 1/8" Molecular Sieve 5A (8600-PK2A, SRI). The GC was calibrated using serially diluted 40% methane calibrant gas (CK Special Gases LTD) to produce a linear calibration curve (R^2 = 0.99) from peak area/peak height. An air blank and a 20% reference standard were run after every nine samples.

2.4.4 DNA sampling and amplification

1 ml digester samples were taken from biological triplicates through rubber bungs to avoid the introduction of oxygen. Digester samples were centrifuged at 10,000 RCF for 10 minutes at room temperature to yield a microbial pellet and clear liquid filtrate. Genomic DNA (gDNA) extractions were performed on the microbial pellet (200 mg) using the commercially available MoBio Powersoil DNA extraction kit to yield approximately 25 µg genomic DNA (gDNA) per gram of biomass. Universal prokaryotic primers A519F and 802R officially known as S-D-Arch-0519-a-S-15 and S-D-Bact-0785-b-A-18, specific to the V4 region of the ribosomal RNA 16S subunit were used to amplify the V4 16S rRNA region of DNA. Sequences of primers with illumina adapters A519F and 802R are TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGCMGCCGCGGTAA and GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACNVGGGTATCTAATCC respectively. These primers have previously been verified for joint amplification of bacteria and archaea showing high coverage for both phyla (Klindworth et al., 2013). A master-mix was prepared and dispensed into prepared DNA solutions that contained 50 ng and 1 ng of sample and *E. coli* gDNA respectively. 25 µl PCR reactions contained 1.25 µl of 10 mM primers, 5 µl Q5 buffer, 0.5 µl of 10 mM dNTPs, 0.25 µl (0.5 units) of Q5 polymerase and the rest ddH_2O. PCR conditions were as follows; initial denaturation for 3 minutes at 94°C followed by 30 cycles of 45 seconds denaturing at 94°C, 60 seconds annealing at 50°C and 90 seconds extension at 72°C with a final extension time of five minutes. Amplicon size was checked first for all samples using gel electrophoresis and then using a Bioanalyser DNA 1000 chip to verify the fragment size (284 bp) in 12 random samples before being purified using AMPure XP beads per manufacturer protocol. Purified 16S DNA samples were then barcoded using Nextera indices; I7 Index set B and I5 Index set C per manufacturer instructions. Remaining clean-up, normalisation and

denaturing were all done according to manufacturer protocol then sequenced using the MiSeq platform (2 x 300 cycles) (Illumina, 2013).

2.4.5 Bioinformatics Pipelines

QIIME (Quantitative Insights into Microbial Ecology, v.1.7.0; http://www.qiime.org) workflow scripts hosted on the Illumina BaseSpace sequence hub (https://basespace.illumina.com) were used to perform bioinformatics analysis. OTU clusters showing >97% identity were assigned using the Greengenes (May 2013) reference database (Caporaso et al., 2010; DeSantis et al., 2006). Taxonomy was assigned using the RDP classifier with the assign_taxonomy.py script. Alpha diversity in samples was calculated using alpha_diversity.py to compare different alpha diversity measurements from between 10 to 17200 sequences (using Chao 1, Shannon Index, total species, and phylogenetic distance metrics) across sequencing depth.

By increasing the number of sequences an ever-increasing number of species was observed in all samples, regardless of time point or treatment, which showed no signs of plateauing. With over 1500 species identified, many of the species were present at such low levels as 10-50 reads representing a minority of the population (below 0.1%). This makes the total number of species present in each sample an unsuitable measure of diversity as the rarefaction measure does not stabilise. Beta diversity distances between samples was calculated using beta_diversity.py and principle coordinate plots created using the weighted UniFrac distance metric (Lozupone et al., 2011). A reproducible measure of community alpha diversity was calculated when using the Shannon Index at a rarefied sequencing depth above 1729 reads per sample and the moving three-point average plotted as the trend line. The moving averages metric was chosen above other methods such as polynomial or logarithmic to identify the point after which the average becomes stable and thus an accurate metric of microbial diversity is reached.

Diversity between samples was then calculated using a weighted UniFrac distance metric from the comparison of phylogenetic tree profiles which considered the total number of shared species (nodes) in addition to the relative OTU abundance (weight) (Lozupone et al., 2011). By comparing samples within a dataset, a distance matrix is created using the values for phylogenetic dissimilarity. This type of principal component analysis (PCoA) is recommended over PCA when there are lots of missing data

(i.e missing days between sample collection) and when there are fewer individuals (two replicates) than characters (three variables) (Lozupone and Knight, 2005). Clustered genera heatmaps were generated from the *E. coli* normalised reads using the ClustVis, a web tool for visualising closeting of multivariate data (http://biit.cs.ut.ee/clustvis/) using unit variance scaling, to normalise abundance differences between genera, with both rows and columns being clustered using correlation distance and average linkage (Metsalu and Vilo, 2015). Stacked plots with dendogram linkage trees were generated by hierarchically clustering a reduced dataset containing the 19 most abundant genera names in addition to "other", a sum of the remaining genera, using the SciPy machine learning library in the custom script "Plotter.py" (Bauckhage, 2015).

2.5 Results

2.5.1 Properties of used cationic polyacrylamide

The flocculent polymers typically used within the waste water industry are cationic acrylamide derivatives that can be degraded anaerobically via the cleavage of the amide bond yielding carboxylic acids and inhibitory ammonium (Xu. Dai et al., 2014a). Pot ale, a waste material from the growing whisky industry comprised of dead yeast and spent grain, was used to feed the communities derived from the cPAM. Table 2.1 shows the chemical properties of the feedstock pot ale and inoculum from Chemical Oxygen Demand, Total Suspended Solids, Carbon:Nitrogen ratio (COD, TSS and C:N respectively) and metal analysis. Neat pot ale possessed a carbon to nitrogen ratio of 8:1 which is higher than cPAM's ratio of 3:1. Biomethane potential (BMP) tests fed with 5, 10 and 15% v/v pot ale contained 3527.3, 7054.5 and 10582 mg COD/L respectively. Trace metals such as cobalt, nickel and molybdenum are essential for multiple catalytic sites in methanogenesis related enzymes such as methyltransferases and methylreductases (Louis, 2016). Pot ale was relatively cobalt and nickel poor but copper and iron rich compared to the flocculant polymer with methanogenesis-related metals such as cobalt, nickel and molybdenum below concentrations recommended by Louis (2016) at cobalt concentrations of 4.99ug/L compared to the flocculant polymers concentration of 58.85ug/L, nickel concentrations of 96.23ug/L verses levels of 323.51ug/L in pot ale and molybdenum concentrations at 17.80 and 81.47 ug/L for pot ale and polymer respectively.

Measurement	Pot ale values	Resuspended polymer
Chemical Oxygen Demand (COD)	70545 mg/L	N/A
Total Suspended Solids	6566 mg/L	N/A
C:N ratio	8:1	3:1*
Calcium	88.24 mg/L	14.95 g/L
Copper	4484 µg /L	1713.54 µg /L
Iron	1715 µg /L	581.65 mg /L
Cobalt	4.99 µg /L	58.85 µg /L
Nickel	96.23 µg /L	323.51 µg /L
Molybdenum	17.80 µg /L	81.47 µg /L
*The empirical formula for acrylamides was used as insolubility prevented analysis		

Table 2.1. Pot ale and cPAM composition

Chemical properties of the feedstock pot ale and the used flocculant polymer inoculum. C:N ratio analysis using Hach Lang total C and total N kits showed carbon deficiency for both pot ale and the flocculant polymer indicating that additional carbon would need to be supplemented for any long-term digestion, but this was not considered an issue for a biomethane potential (BMP) test. ICP-MS results showed trace metals were above recommended stimulatory levels and below inhibitory levels considered essential for the enzymes involved in the methanogenesis pathway

2.5.2 Gas production kinetics of stored cPAM seed

The performance of stored used cPAM was assessed using a BMP test following storage at; either room temperature or 4°C between 30 and 90 days, or stored under cycles of freeze and thawing, to assess its suitability as a microbial seed material. Following resuspension and a short starvation period of seven days seed material was fed and gas production monitored. Ultimately, all samples were able to produce gas regardless of storage conditions (Figure 2.2). The BGP for samples subjected to storage at room temperature for 30, 60 and 90 days was 48.67 ± 1.31 ml, 25.17 ± 7.58 and 25.67 ± 0.62 ml respectively. Chilled samples across identical storage durations produced a total of 52.1 ± 1.11, 34.67 ± 1.93 and 31.50 ± 0.41 ml respectively. cPAM put through freeze-thaw cycles (1-3) produced 27.3 ± 0.12, 22.3 ± 0.1 and 20.0 ± 0.1ml over the course of the incubation. The BGP tests on stored inoculum showed no major decrease in gas yields after 30 days storage when compared to the day 0 reference sample (Figure 2.2, 48.67 ± 1.31 compared to 45.67 ± 0.24 ml). However, inoculum stored for 60 days or more at room temperature showed a decrease in gas production by almost half (Figure 2.2, 25.17 ± 7.58 and 25.67 ± 0.62 ml for 60 and 90 day timepoints respectively) whereas inoculum stored at 4°C showed a decrease by a third or more (Figure 1, 34.67 ± 1.93 and 31.50 ± 0.41 ml for 60 and 90 day timepoints respectively). This data is in agreement with Hagen who previously showed that samples

stored for up to one month show no significant decrease in activity until thereafter (Hagen et al., 2015).

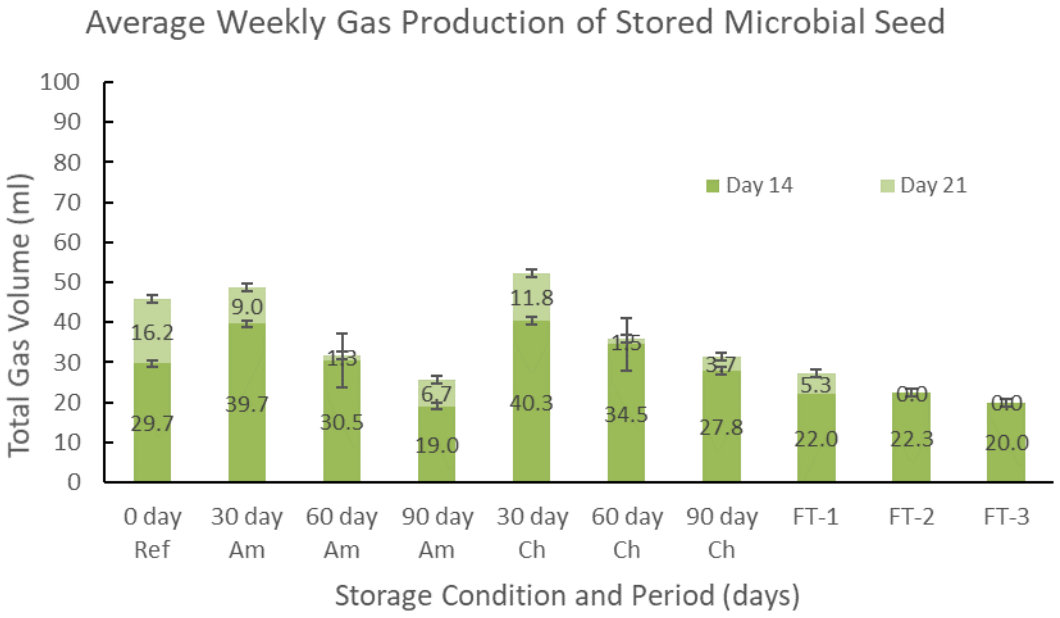


Figure 2.2 Weekly gas production of stored cPAM

The BMP tests (n = 3, error bars = standard deviation) on stored inoculum showed no major decrease in gas yields after a storage period of 30 days when compared to the day 0 reference sample (48.67 ± 1.31 compared to 45.67 ± 0.24 ml). However, inoculum stored for 60 days or more at room temperature showed a decrease in gas production by almost half (25.17 ± 7.58 and 25.67 ± 0.62 ml for 60 and 90 day timepoints respectively) whereas inoculum stored at 4 showed a decrease by a third or more (34.67 ± 1.93 and 31.50 ± 0.41 ml for 60 and 90 day timepoints respectively). This data is in agreement with Hagen who previously showed that samples stored for up to one month show no significant decrease in activity until thereafter (Hagen et al., 2015).

2.5.3 Gas and methane production of varying loading rates

Total gas volume and methane concentration were measured to confirm presence of a microbial community (Figure 2.3). Total gas production increased significantly (P value <0.0001), from 20 ± 1.4 ml to 45.7 ± 0.2 ml by increasing pot ale concentration from 5% to 10% v/v. When pot ale concentration was increased to 15% v/v total gas production was 35.0 ± 2.9 ml, a decrease of 23% compared to 10% v/v pot ale fed system. Trends observed in total gas production were mirrored in methane yields with 5, 10 and 15% v/v pot ale fed cultures produced 1.3, 8.0 and 4.7 ml methane respectively which shows community overloading and methane yield reduction by pot ale concentrations above 10% (Figure 2.3).

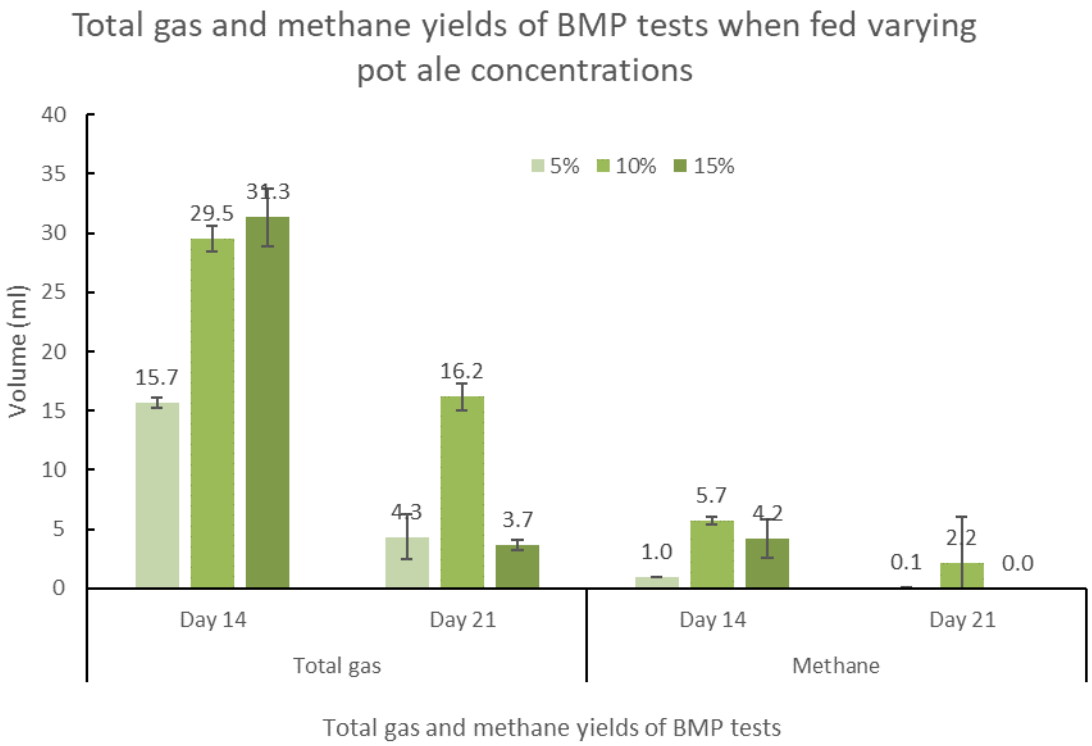


Figure 2.3 Weekly gas and methane production from cPAM fed varying concentrations of pot ale

Total gas produced by BMP tests (n = 3, error bars = standard deviation) was measured using a graduated inverted pipette 7 days after feeding with pot ale. BMP tests fed with 15% pot ale produced the most gas (31.3 ± 2.5 ml) followed by 10% (29.5 ± 1.1 ml) and 5% (15.6 ± 0.5 ml). Less methane was produced by 15% pot ale (4.2 ± 1.6 ml) than 10% (5.7 ± 0.3 ml). By day 14 gas production had almost stopped for BMP tests fed with 5% and 15%, producing 4.3 ± 1.9 ml and 3.7 ± 0.5 ml total respectively. This trend was not seen in the BMP test fed with 10% pot ale which showed a day 14 gas yield of 16.2 ± 1.2 ml with 2.2 ± 3.8 ml methane. Comparing the maximum methane yield of 19.4% (Day 7 for 10% pot ale) values to a functional AD plant operating at above 60% methane yield we can see that these BMP tests produce significantly lower quality gas.

2.5.4 Ammonium accumulation

Ammonium accumulation is not typically measured in BMP tests, but an early understanding of inhibitor production rates would provide warning signs to look for during full scale digestion (Holliger et al., 2016). Due to the high yeast content in pot ale, ammonia production and inhibition was primary concern for subsequent digestion. Initial nitrogen content and production of ammonium were quantified to determine if ammonium is responsible for community inhibition when fed increasing concentrations of pot ale. Indeed, the high nitrogen ratio was discovered to be C:N 8:1 as shown in Table 2.1 previously. By monitoring the levels of ammonium over the course of the loading rate experiment it was found that a non-inhibitory level of approximately 476.9 ± 49.9 mg/L was present following the initial starvation phase across the nine replicates (Rajagopal et al., 2013). Following

feeding with the pot ale, and thus introduction of the putative inhibiting agent, ammonium concentrations rose to final levels of 630 ± 43, 824 ± 56 and 795 ± 79 mg/L for BMP tests fed with 5, 10 and 15% respectively (Figure 2.4).

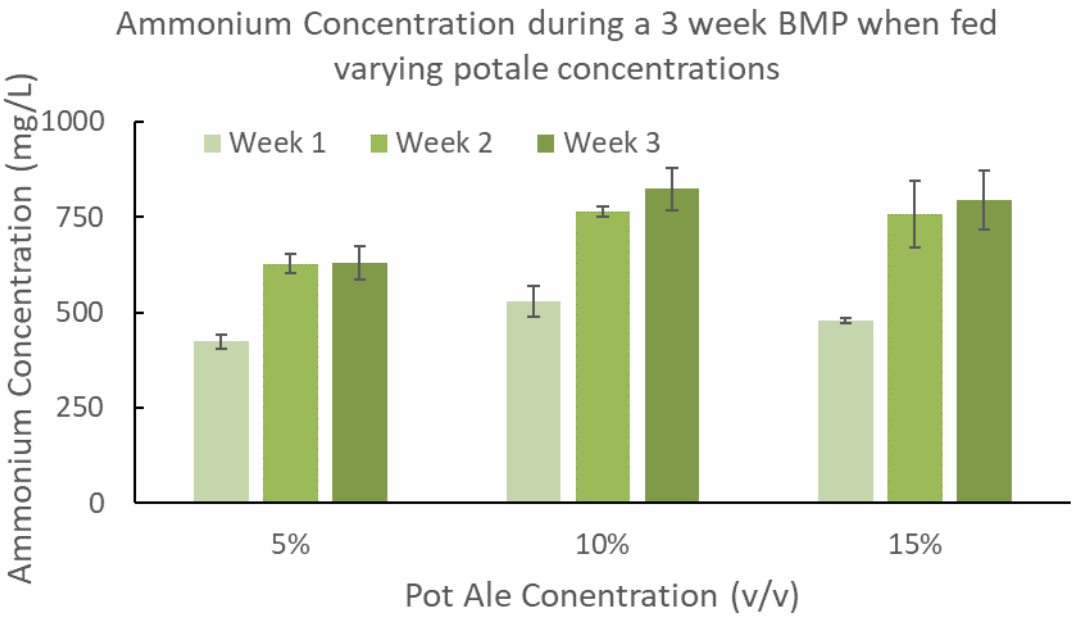


Figure 2.4 Ammonium levels when increasing loading rate

Both the feedstock and the inoculum contained suboptimal C:N ratios of 8:1 and 3:1 respectively. These incubation conditions contain an excess of nitrogen which is degraded into toxic ammonium and so toxicity was the main concern. Following the starvation phase, ammonium levels in samples (n = 3, error bars = standard deviation) were comparable at 476.9 ± 49.9 mg/L. In the first week following feeding these levels increased to 627 ± 25, 765 ± 13 and 758 ± 86 for BMP tests fed with 5, 10 and 15% v/v pot ale respectively. Final ammonium concentrations for the cultures were 630 ± 43, 824 ± 56 and 795 ± 79 mg/L with no major change observed between the second and third week.

2.5.5 Rarefaction plots of sequencing depth

Alpha diversity is a measurement of diversity within a single sample. 95 samples were sequenced during the course of this study investigation with an average of 47,046 ± 11,503 reads generated per sample. Over 6,000 unique OTUs representing more than 2,500 genera were detected across all samples. Before meaningful community analysis can occur, it must be verified that samples have been sequenced to a sufficient depth. There are multiple ways to measure alpha diversity such as raw total number of species, predicted total number of species (Chao 1 index), minimum branch node distance for a given set of taxa (Phylogenetic diversity (PD)) or Shannon Index, a measure of both species diversity and abundance (Figure 2.5a-d). The effect of sequencing depth on measure of diversity was plotted for these metrics between 0 and 17290 sequences, in steps of 1729, using the previously

mentioned metrics. Shannon Index was the only method which produced a stable measure of diversity (between 4 and 5.5 across all samples) which was achieved at any sequencing depth above 1729 (Figure 2.5d). Using the Shannon Index the community diversity was plotted over the course of the experiment and showed an overall decrease and bottleneck effect in response to varying feeding concentrations from between 4.35-5.38 to 4.41-5.00 (Supplementary Figure 2.1).

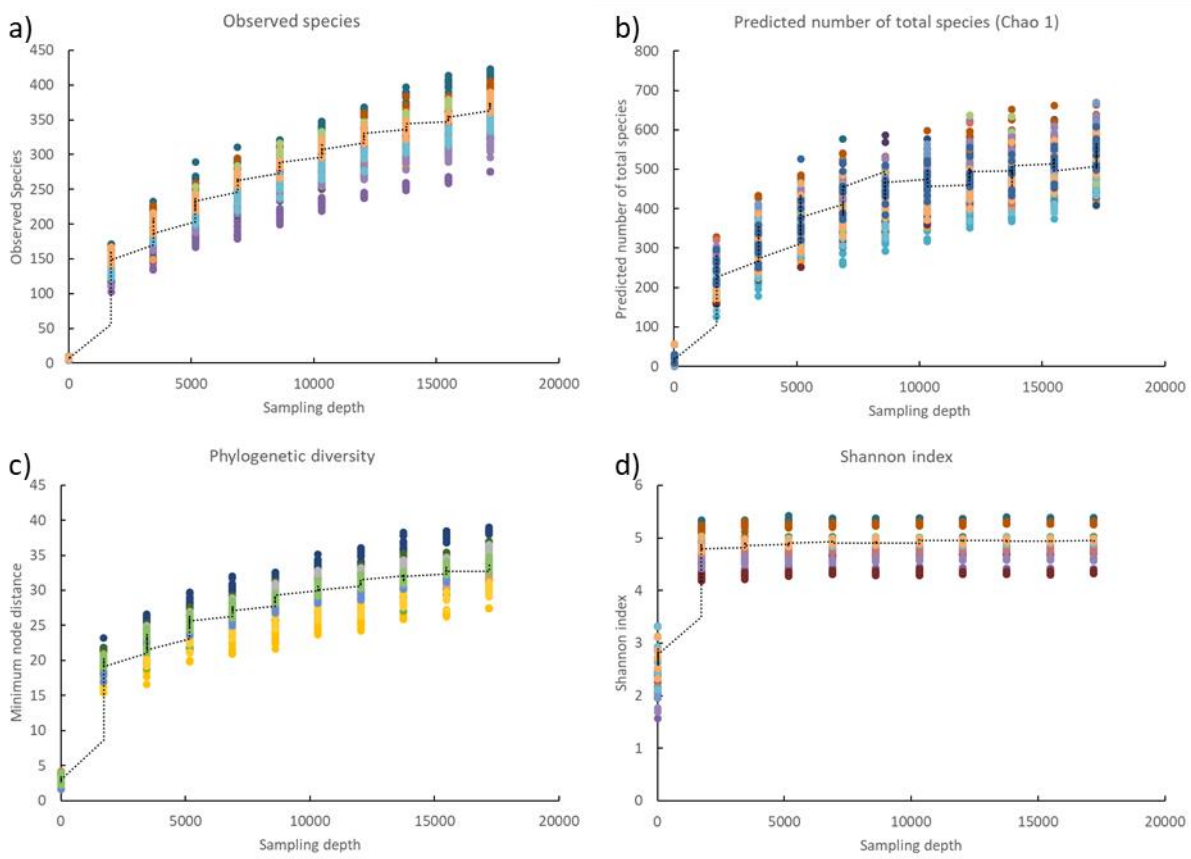


Figure 2.5 Comparison of community diversity metrics

a) Total observed species as increasing sequencing depth shows presence of a large number of scarcely abundant microbial species due to the gradual increase. b) Predicted number of species based upon discovery rate when increasing the sequencing depth reveals a lack of stability and linearity. c) Phylogenetic clustering of OTUs producing an increasingly complex linkage tree (measured by node distance) when increasing the sequencing depth. d) The Shannon index is a measure of sample diversity which considers both presence and relative abundance within samples. By increasing the sampling depth the Shannon index reaches a plateau after 1729 reads with no further increase of diversity when further increasing sampling depth. This indicates that the Shannon index is able to estimate an accurate measurement of species diversity using a subset of reads and more importantly does not change with increased sampling depth. Different colours represent different samples and replicates of colours indicate multiple sub-samples of the same sample.

2.5.6 The effect of storage conditions on community composition

2.5.6.1 Unweighted heatmap analysis

Storage temperature was the major driving force in determining the community structure for samples stored under anaerobic conditions in the absence of additional nutrients. At the initial timepoint, day 0, samples clustered according to their storage conditions (FT, am or ch in Figure 2.6). However, these initial differences soon disappear. By the end of the short batch tests when comparing community profiles on day 0 and day 21 the main factor differentiating samples is the timepoint (Figure 2.7) highlighting the convergent nature of the inoculum community when subjected to different storage.

These results show that dewatered sludge inoculum is a suitable inoculum for seeding anaerobic digesters following a period of storage at ambient, chilled, or freezing temperature as the community quickly recovers in response to feeding to produce a consistent community. This is advantageous for seasonal industries such as crop-based waste streams where biomass is not available for certain seasons such as olive mill waste (Blika et al., 2009).

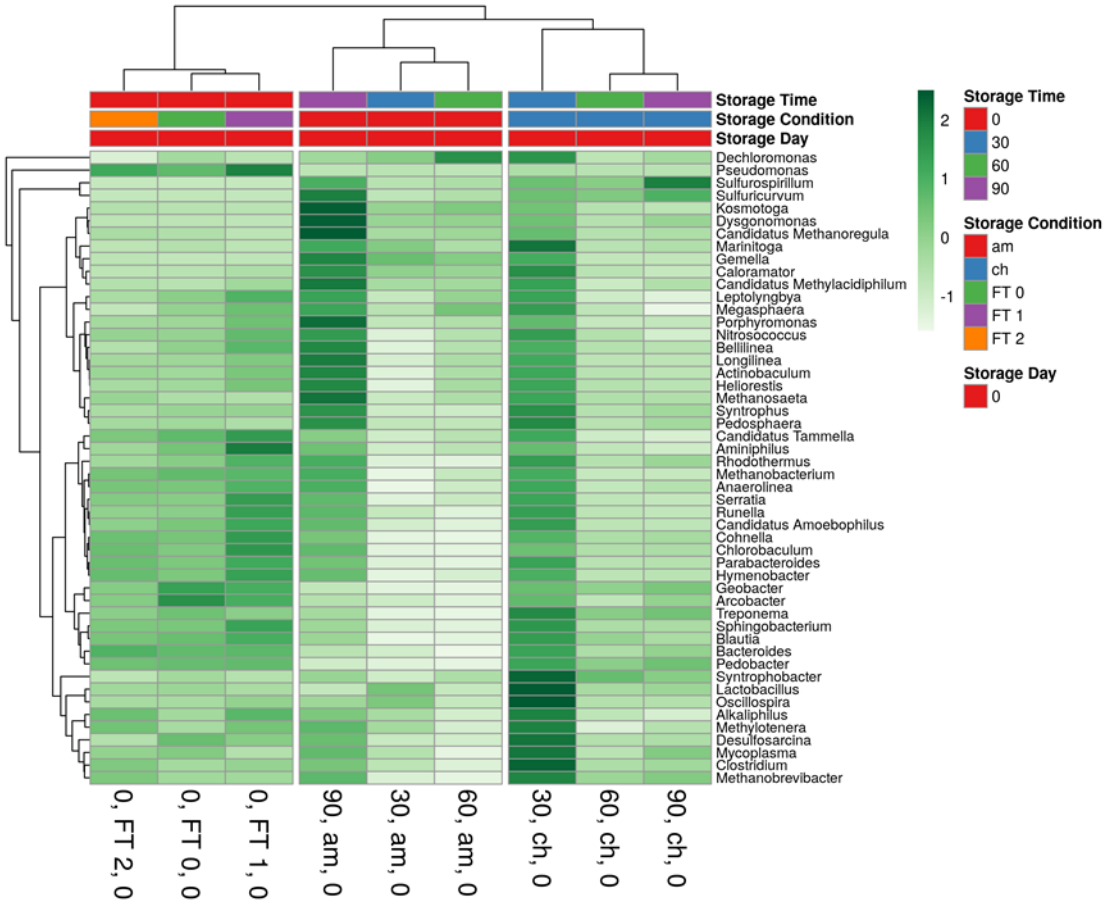


Figure 2.6 Unweighted clustering of starting communities following storage

The top 50 most abundant genera belonging to microbial communities immediately following storage time (days) under ambient, chilled or freezing conditions can be clustered according to their respective treatment. Unweighted abundance of individual microbes are visualised from least (white) to most (dark green). Duplicate samples are combined. Storage conditions are abbreviated as follows; at ambient temperature (am), chilled at 4°C (ch) or subjected to cycles of freezing (FT N, where N = 0 to 2).

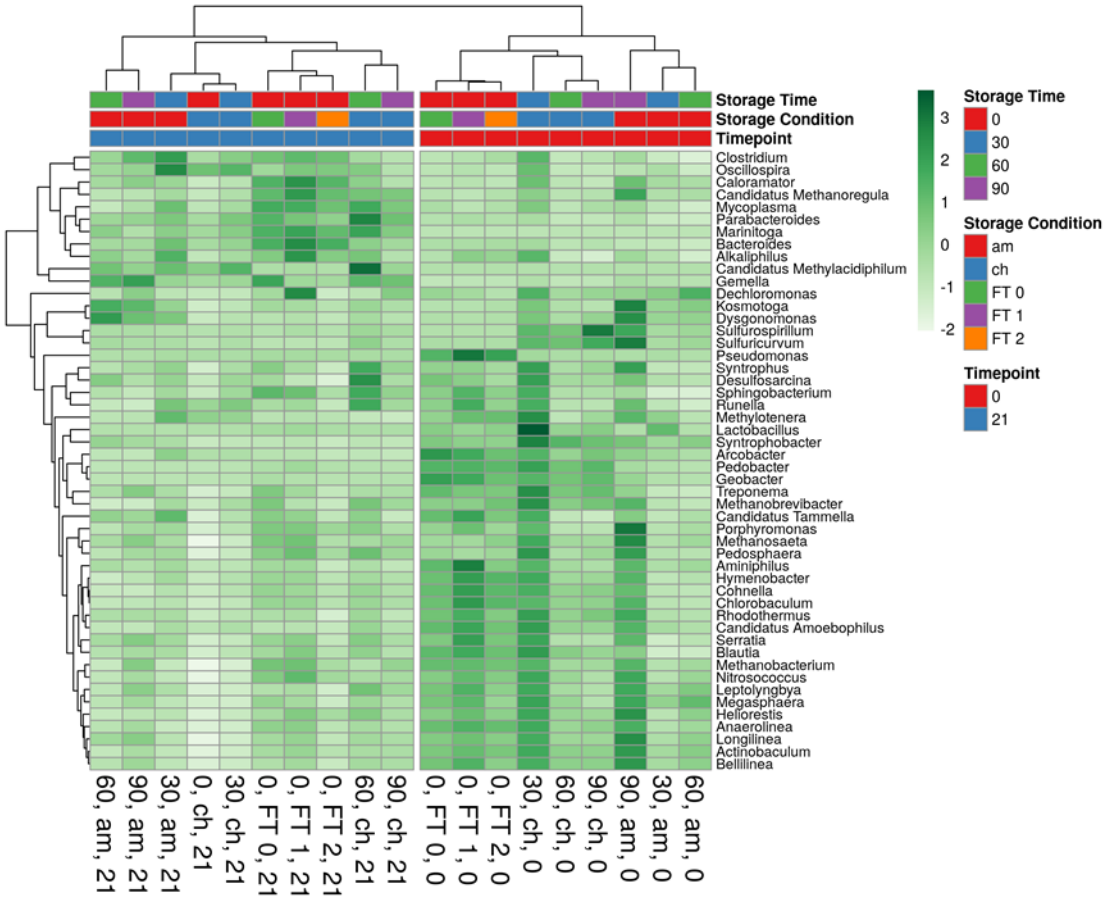


Figure 2.7 Unweighted clustering of initial and final communities

Comparison between start (d0) and end (d21) of genera in batch fed microbial communities inoculated using material stored under different environmental conditions show major community distinction is solely down to timepoint rather than storage condition. This reveals a robust and highly convergent nature of the cPAM inoculum. Duplicate samples are combined. Storage conditions are abbreviated as follows; at ambient temperature (am), chilled at 4°C (ch) or subjected to cycles of freezing (FT N, where N = 0 to 2).

2.5.6.2 Weighted relative abundance analysis

Community relative abundance plots for different storage durations (Figure 2.8) and temperatures or freeze-thawed (Figure 2.9) show the respective composition of the 19 most abundant classes of organism present with the 20th group is a combination of all others. In descending order, all communities are comprised of *Gammaproteobacteria, Bacteroidia, Anaerolineae, Clostridia, Sphingobacteria, Bacilli, Deltaproteobacteria, Methanobacteria, Flavobacteria, Thermotogae, Methanomicrobia, Spitochaetes, Synergistia, Actinobacteria, Epsilonproteobacteria, Betaprotobacteria, Synechoccocphycidea, Fusobacteria* and *Methyacidiphilae*.

Weighted correlation was used to cluster samples ,and assigning a cut-off value of 0.1 split all samples into pre (day 0 and 7) and post feeding (day 14 and 21) clusters (Figure 2.8 & 2.9). Within the two major clusters, the majority of replicate temperatures and timepoints clustered in a pairwise fashion and similar time point and conditions within a close distance e.g 2 or 3 nodes such as the chilled day 0 samples stored for 90 or 60 days. Individual timepoint clustering using the same method for only day 0 and day 21 showed samples clustering due to storage temperature or timepoint, respectively (Supplementary Figure 2.2 and 2.3). Just as in the unweighted correlation analysis previously discussed, the defining feature which so distinctly splits the samples when using weighted correlation analysis is the pre and post feeding conditions.

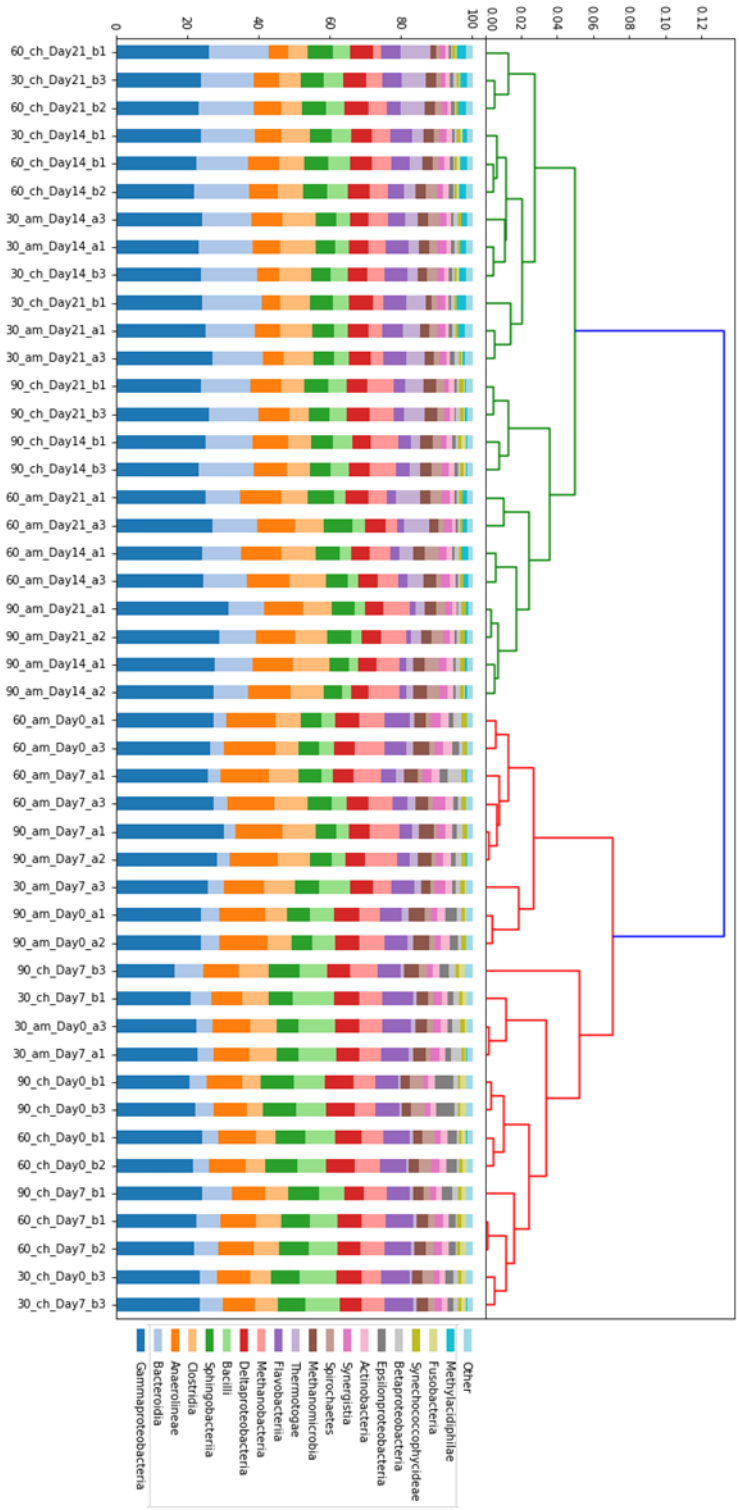


Figure 2.8 Weighted relative abundance plots for samples stored in ambient or chilled conditions

Weighted hierarchal clustering of relative abundance microbial community plots reveal main distinction due to pre (d0 and d7) and post (d14 and d21) feeding conditions. Microbial communities were inoculated from used cPAM stored at either ambient (am) or chilled (ch) for 30, 60 or 90 days.

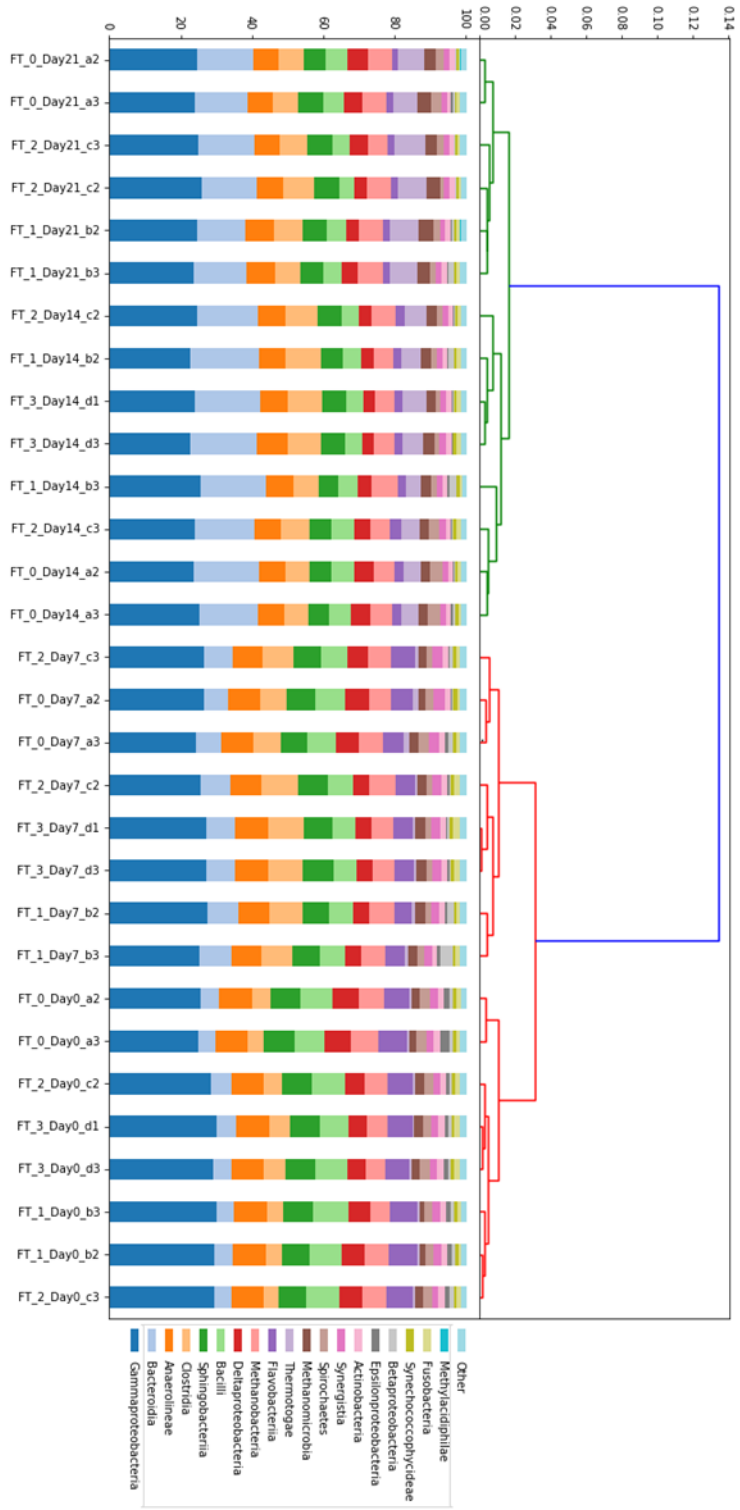


Figure 2.9 Weighted relative abundance plots for communities subjected to freeze-thaw cycles

Weighted hierarchal clustering of relative abundance microbial community plots reveal main distinction due to pre (d0 and d7) and post (d14 and d21) feeding conditions. Microbial communities were inoculated from used cPAM subjected to varying rounds of freeze-thawing (between one and three).

The community on the whole remains unchanged apart from two classes of organism; *Bacterodia* and *Anaerolineae*. Pre-feeding stored cPAM communites typically had Bacterodia relative abundances of 5.05% which increased post feeding to above 13.50% and *Anaerolineae* showed the opposite; initially 11.25% and then decreasing to 8.86%. Further statistical analysis of the relationships between storage condition, storage duration and pre and post feeding timepoints were performed using the relative abundance values for the three most abundant genera; *Gammaproteobacteria*, *Bacterodia* and *Anaerolineae* as they represented a large proportion (43.7%) of the entire community across all samples. The analysis performed investigate the relationship between a) storage duration b) storage conditions and c) pre and post feeding conditions.

Storage duration seemed to show no significant impact on overall *Gammaproteobacteria* abundances over the duration of the experiment (P = 0.32, n = 15, 16 and 16) although an average relative abundance decrease of 2.93% was observed when comparing ambient to chilled conditions with a 95% confidence interval between 1.62 and 4.24% (P = <0.0001, n = 24 and 23). Pre and post feeding conditions also did not seem to significantly affect the population abundance (P = 0.10). The Bacterodia population was also unaffected by storage duration (P = 0.63) although was slightly impacted by storage condition which showed an average increase of 2.69% when comparing ambient and chilled samples (P value = 0.047) with a 95% confidence interval between 0.04 and 5.33%. Feeding increased the relative abundance of the Bacterodia population on average by 8.46% with a 95% confidence interval between 7.35 and 9.56% (P = <0.0001). *Anaerolina* did seem to show an overall increase in abundance when increasing storage duration from 8.44 ± 2.06 % to 10.74 ±2.40 and 10.80 ± 1.69%, for storage durations of 60 and 90 days respectively. Keeping samples in chilled conditions reduced the abundance by 2.69% with a 95% confidence interval between 1.59 and 3.79% (P = <0.0001) as did the act of feeding the communities which reduced the population abundance by 2.38% from 11.25 to 8.86%.

Samples subjected to increasing rounds of freeze-thaw cycles also showed a similar split when comparing the pre and post feeding. Microbial communities showed no significant *Gammaproteobacteria* change when varying the number of cycles (P = 0.39) although a statistically significant decrease was observed between pre (27 ± 1.92 %) and post (24 ± 0.92%) feeding conditions (P = < 0.0001). Similarly, no significant differences were found between cycles of freeze thawing for Bacterodia (P = 0.98) although an increase from 6.6 ± 1.54% to 16.7 ± 1.67% post feeding was observed (P = <0.0001). Likewise, the *Anaerolina* population showed similar trends with no statistical

differences across freeze thaw cycles (P = 0.24) although the population did decrease between pre and post feeding timepoints from 16 ± 0.35% to 14 ± 0.46% (P = < 0.001).

A more inclusive and condensed version of the weighted correlation analysis, a PcA plot, was created using all genera taxonomic data points which showed that samples predominantly clustered based upon their time point above all else (Figure 2.10a & b for stored and FT samples respectively).

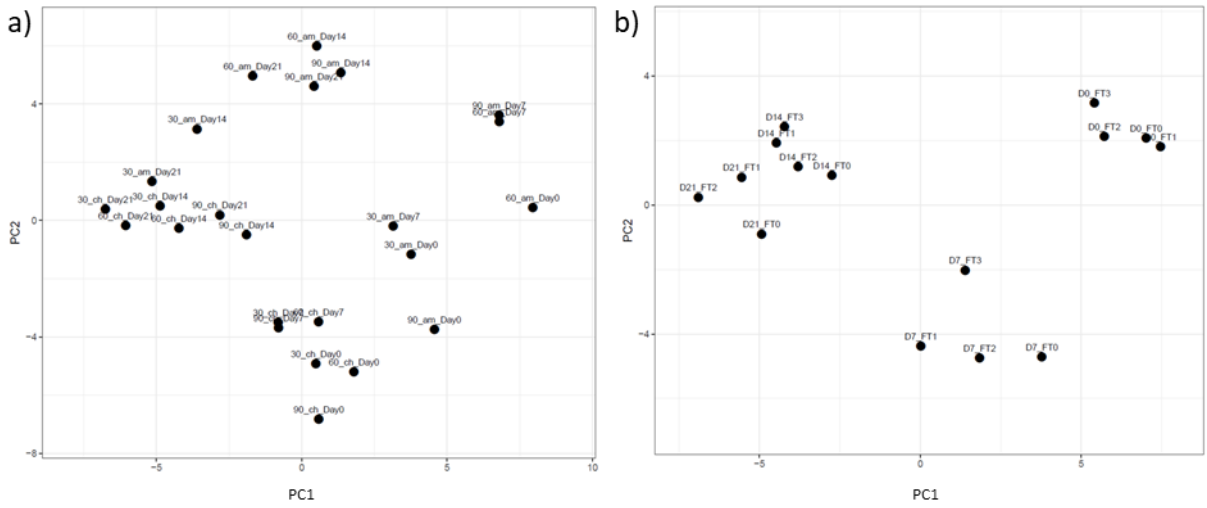


Figure 2.10 Principle component analysis of stored and frozen communities

a) Principle component analysis identified maximal variance of the top 50 most abundant genera present in microbial communities across storage temperature and duration experimental timepoints segregate communities based upon pre (d0 and d7, bottom right) and post (d14 and d21, top left) feeding conditions. b) Microbial communities across freeze-thaw cycle experimental timepoints segregate communities based upon pre (d0 and d7, top right) and post (d14 and d21, top left) feeding conditions.

2.5.7 The effect of increased feeding rate on community composition

2.5.7.1 Unweighted heatmap analysis

The top 50 most abundant genera (which represent 92.23 ± 0.43% of the community) were clustered based upon 16S rRNA gene phylogenetic similarity across all samples (Figure 2.11). This revealed three main trends across the experiment when using a K means clustering value of 3; 1) genera which show a universal decrease regardless of feeding rate and duration (box A) including *Lactobacillus* (facultative

anaerobe), 2) genera which show an increase in BMP tests fed with 10% (v/v) and 15% (v/v) pot ale (box B), such as *Bacteroides* and *Parabacteroides* (carbohydrate degraders) and 3) genera which are elevated only on day 14 or 21 when fed with 5% pot ale (box C), including; *Syntrophus* (methanogen symbiont), *Gemella* (capnophilic) and *Kosmotoga* (hydrogen producer), *Alkaliphilus* (alkaliphilic), *Aminiphilus* (an acetate producer previously isolated from brewery waste) and *Arcobacter* (Elshahed and McInerney, 2001; Cao, Liu and Dong, 2003; Takai *et al.*, 2003; Holt, Gahrn-Hansen and Bruun, 2005; O'Mahony *et al.*, 2005; Díaz *et al.*, 2007; Wexler, 2007; Pianta *et al.*, 2007; Sakamoto, Kitahara and Benno, 2007; DiPippo *et al.*, 2009; Ulger-Toprak *et al.*, 2010; Szabó *et al.*, 2011). These results indicate the presence of a core set of microbes which reproducibly converge under anaerobic conditions. Members identified include those required for long term anaerobic digestion of pot ale such as methanogens and carbohydrate degraders. BMP tests fed with 10 and 15% pot ale show similar communities, implying that community change is reproducible and will result in similar, if not identical, microbial communities.

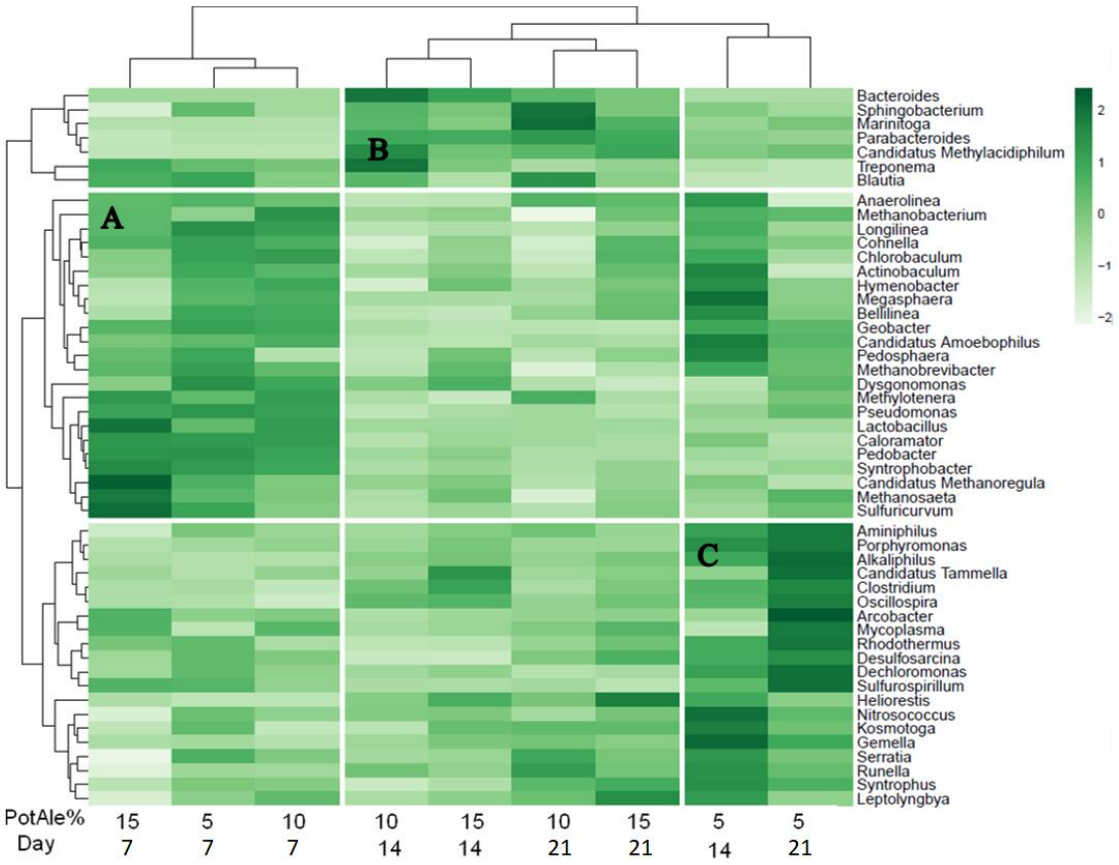


Figure 2.11 Unweighted heatmap analysis of communities fed varying concentrations of pot ale

An initial community selection event is observed in all samples (Box A) with a decrease in various community members following feeding. A second group of organisms show an increase in abundance on days 14 and 21 when fed the relatively higher concentrations of 10% and 15% pot ale (Box B). This group included carbohydrate degraders such as *Bacteroides* and *Parabacteroides*. Minor day specific abundances of select genera were observed for samples fed 5% pot ale on day 14 and 21 (Box C), revealing a possible community transition when fed lower concentrations of pot ale.

2.5.7.2 Weighted relative abundance analysis

Weighted correlation clustering of samples showed the majority (16/18) cluster in accordance with the pre and post feeding dogma previously described for samples stored under different environmental conditions using a tree cut-off of 0.125 (Figure 2.12). Reducing the cut height to 0.75 produces three main clusters of (from left to right) a) pre feeding samples b) post feeding 10 and 15% v/v pot ale samples and c) the two pre-feeding samples that did not cluster with their corresponding partners in addition to samples fed with the lowest concentration of pot ale. Note that for investigating feeding concentration the pre-feeding datapoints were not taken into consideration like in the previous stored samples (n = 4). Differences in levels of Gammaproteobacteria were unaffected by either increasing the feeding concentration of pot ale between 5 and 15% v/v (P = 0.99). In response to increasing organic loading rate from 5 to 10 and then 15% v/v, Bacterodia levels significantly increased from 6.74 ± 0.21 to 16.36 ± 1.17 and 13.32 ± 1.97% (P value = <0.0001). On the other hand, *Anaerolina* showed no significant changes (P = 0.19). Whole community analysis reveals spatial distribution distinction between pre and post feeding samples in addition to separation of respective feeding concentrations (Figure 2.13). When examining the individual timepoints (day 7, 14 and 21) each respective feed concentration clustered closely with its partner in a 3D PCoA plot highlighting replicate similarity (Supplementary Figure 4-6). Combining samples into this 3D space also allows differentiation between pre and post feeding microbial communities as seen in previous dendrogram clustering examples (Supplementary Figure 9) although no extra information is gained.

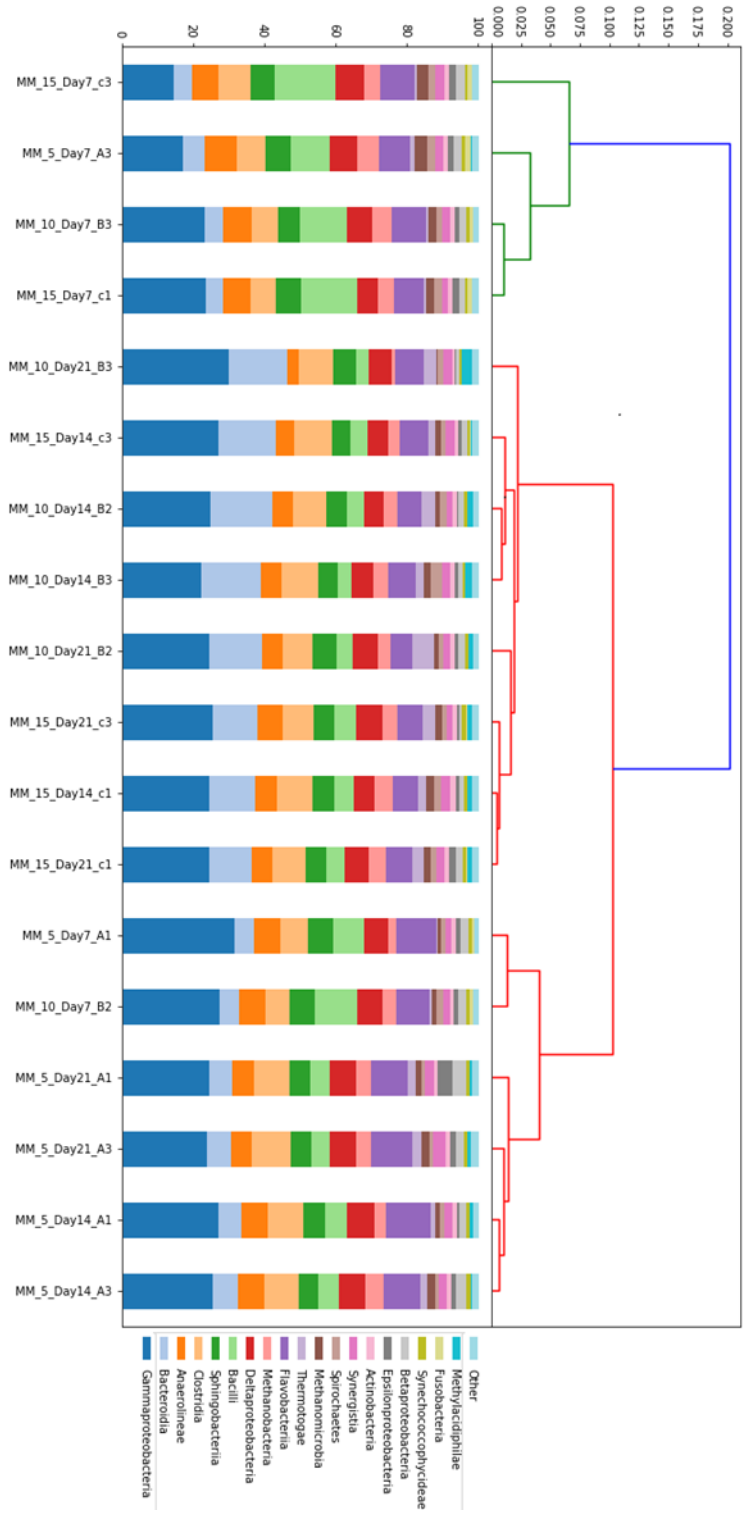


Figure 2.12 Weighted relative abundance plots of communities fed varying concentrations of pot ale

Weighted hierarchal clustering of relative abundance microbial community plots reveal main distinction due to pre (d7) and post (d14 and d21) feeding conditions. Microbial communities were inoculated from used cPAM and fed varying concentrations of pot ale (5%, 10% or 15% total volume). Increased abundance of Bacteroidia was observed in post feeding conditions.

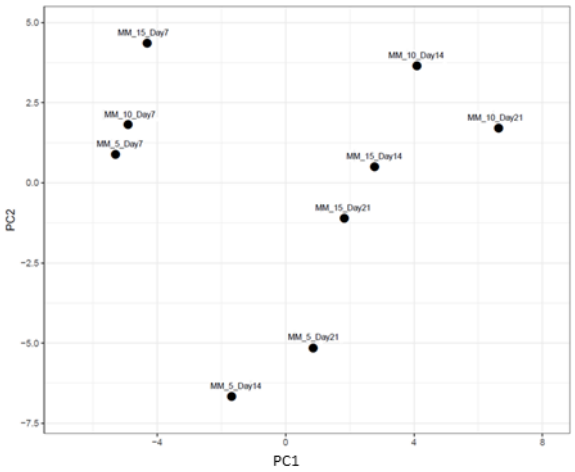


Figure 2.13 Principle component analysis of communities fed varying concentrations of pot ale

Principle component analysis of the top 50 most abundant genera present in microbial samples across timepoint that were fed either pot ale at a 5, 10 or 15% final concentration. Post feeding communities fed with 10% and 15% showed less variance than those fed with 5%.

2.6 Discussion

2.6.1 Properties of used flocculant polymer and pot ale

Composition analysis revealed the low C:N ratio present in both feedstock and inoculum (8:1 and 3:1 respectively). Optimal microbial growth has been shown to occur in ecosystems which have a carbon: nitrogen ratio of between 20-30:1 (Mao et al., 2015). Thus any long term digestion of pot ale will likely require carbon supplementation or co-digestion with substrates such as cellulose, straw or glycerol (Fountoulakis et al., 2010; Fu et al., 2016; Xia et al., 2012). An example of a high nitrogen feedstocks that showed improved digestion with carbon containing material are manures when supplemented with straw (Xu et al., 2019). In regards to the metal concentrations case, the material flocculated by cPAM could also provide trace metals in addition to inoculating microbes, providing both initial and long-term benefits for digester operation.

2.6.2 Activity and gas production kinetics of cPAM seed

The used cPAM was stored for a period of between 30 and 90 days under room temperature or chilled conditions to assess its ability to form a gas producing microbial community when fed an organic waste material. Additionally some samples were subjected to between one and three freeze thaw cycles. This was done with the goal to create season like conditions to mimic the effects of storage in either Summer (room temperature), Autumn (4 °C) or Winter (freeze thaw cycles). cPAM seed samples were

also stored in sealed containers to mimic large piles of the silage-similar material and the lack of oxygen that the majority of this material would receive in either covered or uncovered silos as according to agricultural best practice (Borreani and Tabacco, 2014; Williams, 1994). Although grass silage coverage is done to preserve dry matter contents over storage periods, the same principle can be applied to preserve microbial seed activity for use in anaerobic digestion with identical procedure. To measure cPAM seed activity following storage a biogas potential test was performed which recorded the total volume of gas produced by the seed under anaerobic conditions. Measurement of biogas production from a feedstock is a well-established method of assessing feedstock suitability and identifying inhibitors (Holliger et al., 2016).

2.6.2.1 30-day storage

Seed material that was stored for a duration of 30 days in ambient or chilled conditions produced similar gas yields of 48.67 ± 1.31 ml and 52.17 ± 1.11 respectively with no statistical differences found between the two conditions according to an unpaired t-test (P value = 0.0975). This is unsurprising as the microbial community adhered to this material would be in early dormancy due to the lack of nutrients provided by the cPAM and possibly not yet affected by the toxic and inhibitory nature of the cPAM that has been previously detected (Dai et al., 2014b; Liu et al., 2019; Wang et al., 2018). When comparing these gas yields, and thus activity, of the stored cPAM seed to a zero day reference, that produced 45.67 ± 0.24 ml of biogas, it is surprising that both the 30-day ambient and chilled samples had a 6% and 14% higher yield that were both statistically significant with P values of 0.0335 and 0.0084 respectively. However, it is unlikely that this is due to a seed maturation effect, since if this was the case the condition which should have allowed the microorganisms to mature should have been the sample stored at ambient temperature since the seed material is derived from a mesophilic AD and not a psychrophilic system. Instead, atmospheric conditions are most likely responsible as atmospheric pressure was not recorded and room temperature remained a stable 20°C.

Referencing the weather archive at www.timeanddate.com revealed a low/high pressure of 0.994 and 1.027 bar over the months of May-June 2016 (the month when the reference and 30-day storage experiment were conducted) which, when comparing gas yields, could increase the yield of the reference from 45.67 ml to 47.19. However, this does not account for the differences observe and operator error is the most likely explanation. Curiously, a related study which investigated the effects of inoculum storage at room temperature and 4°C for periods of up to 11 months also identified an increase in gas yield for only the inoculum stored for one month with a similar pattern. In the related

study a 4.5% increase in yield was observed for inoculum stored at room temperature and a 15% increase for that stored at 4°c (Hagen et al., 2015). Although Hagen *et al* do not address possible reasons for this anomalous increase, their 16S rRNA V1-V3 region sequencing data reveals a relative abundance increase in organisms such as *Bacteroidales* and *Clostridium*, two known gas producing microbes, in the sample stored under chilled condition for one month when compared to the reference. However the microbial composition will be discussed further on.

2.6.2.2 60-day storage

Gas yields from cPAM stored for 60 days at either room ambient or chilled temperatures did not significantly differ from one another with yields of 30.51 ± 1.4 (n=2) and 34.67 ± 1.93 ml of biogas (P value = 0.1175). However, when comparing the gas yields of each respective storage temperature to their 30-day counterparts significant differences were found. For instance; ambient samples stored for 30 days had a mean gas yield of 48.67 ± 1.31 ml which is on average 18.17 ml or 37% more than the 30.51 ± 1.4 ml of gas produced by the 60-day-old sample and is significantly different with a P value of 0.0010 and a 95% confidence interval between 13.68 and 22.66 ml. Chilled samples showed a similar significant difference (P value = <0.001) with cPAM inoculum stored for 30 days having on average an increased yield of 17.5ml or 33% biogas when compared to their 60-day counterpart (52.17 ± 1.11 and 34.67 ± 1.93 ml of biogas respectively with a 95% confidence interval between 12.2 and 22.80 ml). The study by Hagen *et al* also highlighted a decrease in gas yields between samples when stored for between 30 and 60 days in both ambient (427 to 321 ml g VS^{-1} or 25%) and chilled conditions (389 to 272 ml g VS^{-1} or 30%) which is similar to the decrease of 37% and 33% observed for the ambient and chilled samples in this current study.

It is worthwhile to mention that an outlier was removed from the 60-day ambiently stored samples as the gas yield was only 14.5 ml, which can be attributed to the use of older gas permeable butyl rubber stopper used to connect the anaerobic culture flask to the gas collection apparatus. After this event effort was made to ensure all tests used a more recently acquired stopper.

2.6.2.3 90-day storage

Unlike the 60-day samples, statistical differences between ambient and chilled storage conditions were found for samples stored for 90 days. Samples stored in chilled conditions produced on average 31.50 ± 0.5 ml gas which is 5.83 ml more than the 25.67 ± 0.76 ml gas produced by the inoculum stored

in ambient temperature with a P value of less than 0.001 and a 95% confidence interval between 4.37 and 7.30 ml of biogas.

Comparing each of the 90-day storage conditions to their 60 and 30-day partners reveals a significant loss of gas production. For samples stored in ambient conditions increasing the time from 60 to 90 days resulted in a loss of 4.83ml biogas with a 95% confidence interval between 1.85ml and 7.82ml (P value 0.014). Comparing seeds from 90 and 30-day samples reveals a 47% loss in activity with yields being 25.67 ± 0.76 ml compared to 48.67 ± 1.31 ml which is 23 ml biogas less with a 95% confidence interval between a loss of 20.15 and 25.85 ml gas.

These results are also partly mirrored in the chilled family of samples with 90-day gas yields on average 3.17 ml less productive than the 60-day samples, however these results are not significant (P value = 0.086) indicating that the drop in gas production between day 60 and day 90 under chilled conditions is not statistically as significant as the drop which occurs in samples stored under ambient conditions. Similar to the 47% loss of activity between the 30 and 90-day ambient samples, a decrease in biogas yield of 40%, or 20.67 ml, was observed for the day-90 chilled samples when compared to the day-30 chilled time point (P value = <0.001, 95% confidence interval between 16.88 ml and 24.45 ml.

2.6.2.4 Freeze thaw cycles

Samples subjected to cycles of freezing and thawing to mimic winter conditions performed more poorly than the unfrozen reference material with gas yields of 27.3 ± 0.12, 22.3 ± 0.1 and 20.0 ± 0.1 ml for one, two and three cycles compared to the 45.67 ± 0.24 ml biogas yield for the control indicating a significant (P value less than 0.001 when performing a one-way ANOVA) reduction in activity. Further analysis between the best performing samples subjected to freeze thawing and the control revealed that the highest performing freeze thaw group (FT-1) produced on average 18.33 ml gas significantly (P value less than 0.001) less than that of the no freeze control with a 95% confidence interval between 13.16ml and 23.51 ml. Inter-group ANOVA analysis of the three freeze thaw groups revealed a significant difference in gas production (P value = 0.024). This positive test of significant difference can be attributed to the difference in gas production between FT-1 when compared to both FT-2 and FT-3 sample groups. Comparing FT-1 to FT-2 shows that a mean decrease of 4.93 ml, or 18%, biogas occurs when an additional freeze-thaw cycle is applied, however the results are not quite significant, although close, with a t-test P value of 0.066. By subjecting the cPAM seed to two additional freeze thaw cycles a total of 7.33 ml or 27% biogas is lost which is deemed significant (P value = 0.032).

However, when comparing the inoculum subjected to two and three rounds of freeze thawing an insignificant low of 2.4 ml of biogas occurs (P value = 0.18) which indicates an unlikely further loss of activity if the cycles are increased.

Subjecting the cPAM seed to any rounds of freeze thawing seems to drastically reduce biogas producing activity and is thus definitely not recommended for any sort of anaerobic digester reactivation protocol as it presents the combined effects of both cPAM toxicity and significantly hampered gas production. The decrease in gas producing activity is in accordance with the data from Hagen et al where in each of their one, two, six and eleven month stored samples the lowest performing group was those stored under freezing conditions. However, this study focused on repeated freeze-thaw cycles, rather than the freeze-store strategy performed by Hagen et al, and reveals that no significant further loss will occur when subjecting cPAM to additional freeze thaw cycles. The likely reason for such a catastrophic loss of activity is major cell wall disruption followed by cell death as performing freeze-thaw cycles on biological samples, without the presence of cryoprotectants, is a common technique for cell lysis in DNA extraction protocols (Bey et al., 2010; Dattagupta et al., 2009). Even if cell death did not occur due to ice crystal formation, membranes may be disrupted and become more permeable to toxic cPAM in storage (Liu et al., 2019; Wang et al., 2018).

2.6.2.5 Trends across storage conditions

In concordance with the study performed by Hagen et al a general loss of activity was seen as storage duration increased apart from all samples stored for one month which showed an unexpected increase. Differing from Hagen et al's study, however, is the trend between increased gas production and chilled storage conditions. For Hagen's two, six and 11-month samples all room temperature stored inoculum produced approximately 20% more than their chilled correspondents, unlike the results presented in this study which outline that chilled samples perform better across all storage durations. A possible reason for this could be the difference in material present in the inoculum aside from differing microbial populations as the inoculum used in Hagen's study was taken directly from a biogas plant in Norway and is likely a combination of rich biodegradable material in addition to the usual microbial community. Comparatively, the inoculum used in this study is primarily comprised of cationic polymer with a microbial community attached to the surface and so presents a far less concentrated inoculum. Additionally, as we know that cPAM is toxic and inhibits methanogenesis, the effect of prolonged cPAM exposure on microbial communities is not well known although the general

consensus is that it both reduces total gas and methane yield. As the chilled samples performed better, perhaps the cold temperature reduced their metabolic activity enough to prevent them from producing cPAM degradation products in storage. Unfortunately no method of detecting acrylic acid was used in this study although acrylamide detection can be performed using HPLC with detection at 200-210 nm (X. Dai et al., 2014a; Wang et al., 2013).

Unfortunately the gas composition of the produced biogas was not measured across this study so actual methane yields were not calculated. Conventional stoichiometric and experimental data suggest that conventional methane yields of a healthy biogas producing plant are approximately 50% methane (Cavaleiro et al., 2009; Saady and Massé, 2013; Tao et al., 2019). However, as cPAM is present in the reactors and is known to specifically reduce methanogenic activity it would be irresponsible to extrapolate methane yields from biogas yields, even conservatively (Liu et al., 2019; Wang et al., 2018). As such, further experiments were planned using fresh cPAM under varying feed concentrations to collect this data.

2.6.2.6 Effects of loading rate

Both weekly gas production and weekly headspace gas composition were recorded in a series of batch tests which investigated the effect of an increased loading rate of a novel feedstock to the cPAM inoculum, which has previously been shown to foster a gas producing microbial community. Initial trials which involved increasing the pot ale concentration from 0-100% v/v (data not shown) showed a decrease in gas yield when increasing the concentration from 10% to 20% so the concentrations of 5%, 10% and 15% were chosen for further analysis, including DNA sequencing the communities. A total of 20 ± 1.41, 45.67 ± 0.24, 35 ± 2.86 ml biogas was produced over the course of the two week batch test for cultures fed with 5%, 10% and 15% pot ale respectively with specific methane yields of 1 ± 1.89 , 7.9 ± 1.13 and 4.2 ± 1.47 ml. A difference in both biogas and methane yield was deemed statistically significant using an ANOVA test (P value less than 0.001).

Headspace methane concentrations generally increased following feeding from the post starvation (and pre-feeding) level of 0.65 ± 0.52% to 6.11 (n=1 as there was issues with the instrument which lost valuable samples), 19.40 ± 0.33, 13.43 ± 1.58% for respective culture triplicates. Methane headspace concentration in both the lowest and the highest fed cultures (5 and 15%) fell the following week to 1.42 and 0.06 ± 0 whilst the 10% fed culture was slightly higher at 10.61 ± 3.84%. For these batch tests the community seemed to produce the best methane yield, both in terms of volume and richness,

when fed with 10% pot ale. Breaking down the total gas yields into weekly performance metrics reveal that the majority of both gas and methane was produced in the first week of feeding with 78.5%, 64.6% and 89.4% of biogas and 81.1%, 74.0% and 46.8% of specific methane being yielded.

The fact that the culture fed with 5% pot ale produced approximately half (53 and 50%) of both the 10% and 15% fed cultures suggests that activity of the microbial community adhered to the cPAM is working at full capacity in regards to being unable to produce gas more rapidly when increasing the feeding concentration from 10 to 15% during the first week. This is supported by the fact that no statistical significance was found between both groups when performing an unpaired *t* test (P value of 0.39). Since the community is operating at maximum capacity it is likely that the decrease in activity when fed a higher concentration (such as 20-100% in initial failed trials) is due to overloading and inhibition of microorganisms. This idea of overloading and inhibition is supported by the fact that the specific methane yield during week one of feeding decreases from 5.72 ± 0.32 to 3.97 ± 0.43 methane, a mean difference of 1.74 ml with a 95% confidence interval between day 0 (0.7 ml) and day 7 (2.7 ml) (P value = 0.013) indicating specific methanogen inhibition under the respective culture conditions. It is expected that in agreement to previous literature, the methanogenic population are the first to be inhibited in the anaerobic digestion as they are specifically sensitive to VFAs in addition to the more acidic pH conditions (Cirne et al., 2007; Palatsi et al., 2010; Rajagopal et al., 2013; Rasit et al., 2015).

The feeding strategy used in this experiment is potentially another reason for the poorer than expected gas yields when compared to similar studies which digested whisky wastes, flocculant polymers and more conventional material such as swine manure, lignocellulosic material and oil waste. An initial concentrated pulse of feed, akin to a batch feeding regime, was chosen as it was hoped it would dramatically change the microbial community and naturally select from organisms resistant to the pot ale material due to it being a more “extreme” environment compared to daily low concentrations of feeding. When comparing the organic loading rates chosen in this study in terms of total COD present at any given day to other studies of anaerobic digestion in a stirred tank system the loading rates covered conventional operational recommendations. The initial COD concentrations of 3.5, 7 and 10 g COD / L were similar to, and exceeded, previous daily COD feeding rates of malt whisky waste water, lignocellulosic biomass and oleic acid based wastewater in a stirred tank system (1-5, 1-3 and 4-8 COD / L respectively) (Cavaleiro et al., 2009; Russell, 1997; Uzal et al., 2003). These organic loading rates however are low compared to novel anaerobic digester designs such as UASB and AMBRs digesting pot ale, and other material which can process extremely high organic loading rates of 15g

COD / L as observed for an UASB or 20-40g COD/L as observed for a membrane based system (Goodwin and Stuart, 1994; Xing et al., 2010).

As the two upper organic loading rates used in this experiment exceed conventional ranges of operation for stirred tank systems, it is likely that the decrease in gas production observed in the highest fed reactor is caused by organic overloading even though the pH remained stable between 7 and 8. Additionally, typical biogas richness is expected to be approximately 50% which is far higher than even the highest headspace methane concentration of 19.4% achieved in this experiment.

The overall theoretical yields measured in this experiment however are far below the 700 ml biogas per gram of COD theoretical maximum with overall yields, assuming 100% conversion of COD to biogas although some COD will be used for microbial growth, expected to be a total of 73.5ml, 147ml and 220 ml biogas as each reactor was fed a total of 0.105 g, 0.210 g and 0.315 g COD (Dai et al., 2015; Filer et al., 2019). Actual biogas yields were far lower than this on average with 20 ± 1.41, 45.67 ± 0.24, 35 ± 2.86 ml biogas actually being produced corresponding to efficiencies of 27%, 31% and 16% respectively. This overall lower than expected yield can either be attributed to the inhibitory nature of the cPAM inoculum which has previously been discussed or that a certain degree of community acclimatisation is required for more optimum degradation of the material. It is likely that both play a significant part in the low yields recorded here since one of the major limitations of this study was the single batch feeding nature of the test which prevented a climax community forming. These types of climax community, both in microbial community composition and also microbial concentration, only occur after extended operation of bioreactors (Angelidaki and Ahring, 1992; Chen et al., 2008; Goodwin et al., 2001; Ho and Ho, 2012; Palatsi et al., 2010; Yu et al., 2001; Ziganshin et al., 2013). Following on from this lack of community adaptation and potential cPAM toxicity, future experiments involving pot ale were planned to be carried out over a period of months to allow adequate adaptation and using non-cPAM containing inoculum to avoid inhibition.

2.6.3 Ammonium is not the reason for community inhibition

The levels of potentially inhibitory ammonium were across both the stored cPAM seed and in the experiment that investigated increased organic loading rate of potale, since it was hypothesised that the nitrogen rich feedstock, and also inoculum, would cause community inhibition due to ammonium toxicity. Analysis of ammonium concentrations across experimental timepoints following the starvation phase of the experiment were comparable at 476.9 ± 49.9 mg/L indicating that cPAM alone

would not cause toxicity due to ammonium release from the anaerobic degradation of acrylamide and acrylic acid which can occur over a period of days in aquatic systems (Xiong et al., 2018).

Following feeding, the ammonium concentrations increased in the following week due to the addition of nitrogen rich pot ale. During the second week of digestion the ammonium levels remained unchanged (maximum insignificant increase of 0, 8 and 4% increase for each increasing feeding rate) indicating that the majority of yeast protein was digested during the first week following feeding (P values equal to 0.96, 0.21 and 0.69 for differences between day 14 and 21 triplicate values of each respective feeding rate).

Although there was an approximate 200 mg /L increase in ammonium when increasing the feed rate from 5% to 10%, an increase was not observed when the pot ale concentration was increased further to 15%. The reasons for this lack of change when increasing feed concentration are mirrored in the gas production kinetics and the reasons for which have been discussed in the previous section, possibly due to organic overloading and community inhibition. According to various studies, ammonium inhibition is observed only when levels increase above the 2000 mg/L mark (Gao et al., 2015; Ho and Ho, 2012; Rajagopal et al., 2013; Yenigün and Demirel, 2013). Although this cause of inhibition is still possible due to the unspecialised nature of the community, it is more likely than an alternative reason such as unknown metabolite accumulation is the explanation for the decrease in ammonia production, a by-product of protein degradation, observed in the samples fed with 15% pot ale.

2.6.4 Phylogenetic composition of stored microbial seeds

2.6.4.1 Assessment of bioinformatics pipeline

To test whether samples were sequenced to a sufficient depth to draw conclusions about beta diversity, the alpha diversity of each sample was evaluated using increasing subsets of sequence data with a variety of performance metrics. Briefly put, alpha diversity is a measure of diversity within a sample and beta diversity is a measure of diversity between samples. Typically, fewer sequences are needed to differentiate two populations based upon beta diversity as they may have distinct microbial compositions, although understanding a sample's alpha diversity indicates how complex a respective sample may be. Alpha diversity analysis was performed on all 95 samples sequenced over the course of the combined cPAM seed and pot ale study by identifying the sample with least coverage (17290 reads) and using that as the maximum number of sequences. Saturation of each respective alpha

diversity metric was visually confirmed using a moving average (period = 3) via the horizontal displacement of each subsequent moving average trend line for each 1729 subsequence interval.

Total number of species is an analysis method which calculates numerically the total number of novel species in a given sample. Between 100 and 175 species were observed at a sampling depth of 1729 which, according to previous phylogenetic studies of anaerobic digestion and microbial ecology, is more than enough species to produce a robust microbial community with adequate redundancy (Gao et al., 2015; Lennon and Jones, 2011; Zhang et al., 2017). Taking further sequence subsamples into account continues to detect new organisms up to the maximum sample size where between 250 and 400 species are present. As a stable measure of diversity was not reached using this metric, alternative methods were investigated which involved predicting the number of species based upon the discovery rate when increasing subsample (Chao 1). Total number of observed species is an inaccurate and over-rigorous metric in regards to assessing alpha diversity of complex communities because the majority of the population is typically represented by a few key players and the rest belonging to irrelevant and trace-abundant organisms (Carballa et al., 2015; Gasc et al., 2015; Vanwonterghem et al., 2014).

As previously mentioned, Chao 1 in an estimate of total number of species in a sample based upon the discovery rate of new species and rate of species re-detection. Although the predicted number of total species generally showed a positive correlation with sampling depth, some moving averages actually showed a decrease in the predicted number of species when increasing the sampling depth. For example when sampling depth was increased from 8605 to 10324 and again when (finally) increasing from 15481 to 17200 total reads. As this produced an erratic and unstable-looking trend line which showed no trend towards evenness, it was decided to avoid using this metric for further analysis as high replicate deviation has previously been detected when using the Chao 1 metric (Krakat et al., 2016). Other methods such as phylogenetic distance and Shannon diversity index show more consistency between replicates when compared to phylogenetic distance and Chao 1 (Faith and Baker, 2010; Fykse et al., 2016).

More complex than counting total number of species is using phylogenetic diversity (PD) metric which aligns OTUs to each other to create a phylogenetic tree and calculates the minimum node distance (branches) between all species present (Faith and Baker, 2010). By using this metric, the assessment of alpha diversity starts to incorporate a measure of microbial community complexity. However, like previous analysis metrics there is insufficient sequencing depth to provide a stable value.

A further measure of alpha diversity was investigated, termed the Shannon Index, which incorporates both the number of novel species present and their respective relative abundance (Fykse et al., 2016; Li et al., 2013). As previously mentioned, the anaerobic digestion microbial community is made up of a few key players which are most abundant. By taking into consideration the heavily weighted nature of the microbial communities a stable metric of alpha diversity is reached immediately with the first subsample returning a Shannon index of between 4.1 and 5.5 which shows no change when increasing the number of sequences taken into account. For complex microbial communities the Shannon Index is preferred above other methods (Freedman and Zak, 2015; Gibbons et al., 2016; Kelsic et al., 2015; L. Li et al., 2015; Unal et al., 2012; Žifčáková et al., 2016).

Analysis of the community sequencing had to be performed using both unweighted and weighted methods to identify organisms undergoing large-fold changes and major components respectively. Unweighted analysis was performed using vector scaling which provided identical minimum and maximum values across all samples. However this type of scaling, whilst applicable across all taxonomic levels, was not able to highlight the dramatic 10 and 100 fold increase observed for two specific genera; *Thermotagae* and candidatus *Methylacidiphilus* respectively (S Figures 7-9). The choice to sequence duplicates rather than triplicates was partly due to cost but also as reproducibility of sequencing results, and accuracy of microbial ecology inferred, has been shown to owe more to sequencing depth rather than PCR triplicates (Smith and Peay, 2014).

2.6.4.2 Microbial changes in comparison to previous cPAM experiments

The only report to date which investigates anaerobic microbial composition changes in response to cPAM was performed using a similar bioinformatics pipeline which involved classification of the V3-V4 16S rRNA gene region (Liu et al., 2019). Liu et al reported a decrease, in order of magnitude, of the taxonomic classes *Gammaproteobacteria* and *Bacilli,* which is in contrast to the data presented in this report as *Gammaproteobacteria* were both the most abundant class across all storage conditions and feeding rates and also only showed a minor (2.9%) decrease when comparing samples stored in ambient conditions compared to their chilled counterparts. This is presumably due to the increased metabolic activity at high temperatures, thus making them more susceptible to inhibition. Additionally Liu et al reports a decrease in organisms belonging to the class Bacilli, which is not mirrored in the samples exploring storage conditions, but is in those which investigated varying organic loading rates. However in general, the community shows no major population shifts either in this study or the study performed by Liu.

The lack of any dramatic population changes in this series of experiments can in part be attributed to the short duration and operational conditions which they were performed under (one week of starvation followed by feeding and subsequent monitoring of two weeks) although a similar result is observed in the sequencing data produced by Liu et al who fed their reactors in a continuous manner for a period of 60 days. The recalcitrant nature of the cPAM carbon backbone, and thus limitation of biologically available carbon, is likely the reason for the lack of change observed in both cases, with degradation typically performed under aerobic conditions (Gröllmann and Schnabel, 1982; Ramsden and McKay, 1986). Proliferation of certain Bacillus species using cPAM as their sole carbon source has previously been observed, and although complex microbial communities fail to thrive under conditions with cPAM, future investigations into anaerobic cPAM degradation may be more successful if they include species such as *Bacillus cereus* and *Bacillus flexu*. These two organisms have been shown to degrade 75% of cPAM over a 96 hour period under oxygenated conditions (Wen et al., 2010). If biological degradation of cPAM was a priority, this two-step anaerobic digestion process is recommended as it ensures that the majority of cPAM is degraded and is also converted into biologically available carbon, rather than just biologically available nitrogen (Kay-Shoemake et al., 1998).

2.6.4.3 Microbial changes in comparison to previous grain experiments

Unlike anaerobic degradation of cPAM, breakdown of grain and grain associated products have been more widely studied, although not specifically pot ale due to its regional production confined to Scotland. It is widely accepted that spent distillers grain is a rich source of easily accessible carbon for either sole or co-digestion (Agler et al., 2008; Ezeonu and Okaka, 1996; Fanedl et al., 2015; Wang et al., 2012; Ziganshin et al., 2011). The main difference between used grain products and pot ale is the presence of large amounts of spent yeast in the latter which recently has become a target for protein removal and subsequent re-sale (Barrena et al., 2018). In this series of experiments the major community driving force and response has been the application of a feed material rich in grain residues which promotes proliferation of Bacterodia. This group of organisms is known to be avid carbohydrate degraders, crucial in both the gut microbiome and in previous AD of spent grain products (Kettle et al., 2015; Lagier et al., 2012; Wexler, 2007; Ziganshin et al., 2013, 2011). Presence of other known hydrolytic organisms such as Clostridia have also been shown to markedly improve methane yields of similarly conducted batch tests, fed with grain residue, that the cPAM used in this study does provide microbes shown to be essential for grain residue methanisation (Fanedl et al., 2015).

2.6.4.4 Limitations of sequencing strategy used

Although the 16S rRNA gene sequencing-based strategy used in this study clearly identifies the microbial presence of certain organisms due to the cPAM inoculum in addition to some minor changes brought about by either storage conditions or feeding; the presence of an organism does not necessarily correlate with the active community population, especially in batch-test type systems like the one used in this study. This is for one very simple reason; there is insufficient time and replicative room to allow emergence, or even identification, of a climax population as a large wet volume of cPAM containing inoculum was used (approximately 25% v/v) and no further subculturing or dilution over time with additional feedstock (such as occurs in a continuous fed type system over many hydraulic retention times). A better strategy, if using 16S rRNA gene sequencing, would be to repeatedly feed the community over a long period of time. However, at the time these experiments were conducted, the researcher was still constructing the system to allow this type of longer duration experiment as the privately-owned reactors belonging to the BDC, a nearby company, were inoperable. Additionally, the restrictions and inhibition that a cPAM containing system places on the microbial community have been previously discussed at length and suggestions provided for further work, including an aerobic hydrolysis step. It was thus decided to not use cPAM as an inoculum for further experiments and, more of interest, continue the investigation into AD of pot ale.

Most importantly though, regardless of reactor operation duration and feeding schedule, DNA based analysis techniques when supplemented with RNA based study have identified an inactive majority population thus highlighting the irrelevancy of most of the DNA sequencing produced in determining key players and pathways (Ting et al., 2019). Lack of true microbial quantification, such as cell count or normalised abundance, in combination with use of toxic cPAM leaves the question if the community was so severely inhibited by the cPAM that no major growth occurred, apart from the Bacterodia, or if non-specific cell death occurred, excluding the Bacterodia. This lack of accurate microbial quantification is essential for identifying populations which possess high cell numbers with poor efficiency or low cell numbers with high efficiency especially for established microbial community-based industry such as AD. However the heterogeneous nature of AD with regards to impurities, inorganic contaminants, and large particle size makes techniques such as cell counting difficult. A solution to this problem may be the inclusion of a DNA-based spike that would be present in the sequencing reads and thus be able to normalise taxonomic counts against. This was trialled briefly by addition of *E.coli* genomic DNA into select samples (Figures 28-30) although post-experiment it was decided that addition of pure *E.coli* DNA to microbial community genomic DNA did not allow accurate

determination of extraction efficiency. Improvement of this technique would allow more accurate characterisation of microbial communities across microbial ecology and further research was performed on this idea of a microbial quantification spike in Chapter 3.

2.7 Conclusion

The biological community and inorganic material that cPAM precipitates from anaerobic digestion plant dewatering produces an active methanogenic community containing trace metals essential for methanogenesis and key genera such as *Methanobacterium*, *Methanosaeta* and the polysaccharide degrading *Methanobrevibacter* and *Parabacteroides*. Initial stages of anaerobic digestion revealed an increase of organisms known to degrade carbohydrates such as *Bacteroides* and *Parabacteroides*, most likely in response to the high levels of carbohydrate in the form of spent grain in the feedstock used for this study. The robust nature of the inoculum was also highlighted by the cultivation of phylogenetically similar communities regardless of storage temperature (ambient, chilled or cycles of freeze/thaw) or duration (between 30-90 days). Although no increase in the methanogenic population was detected across all samples when comparing start and end timepoints, their presence and activity were confirmed through the use of GC-TCD and 16S rRNA geneamplification sequencing, due to their long growth lag-times in comparison to other faster growing microbes in the community such as *Bacteroides* and *Parabacteroides*. As previous research has identified methanogen inhibition when digesting acrylamide-based polymer, we propose that the material can be used for the inoculation phase only of anaerobic digestion and is stable across a variety of environmental conditions.

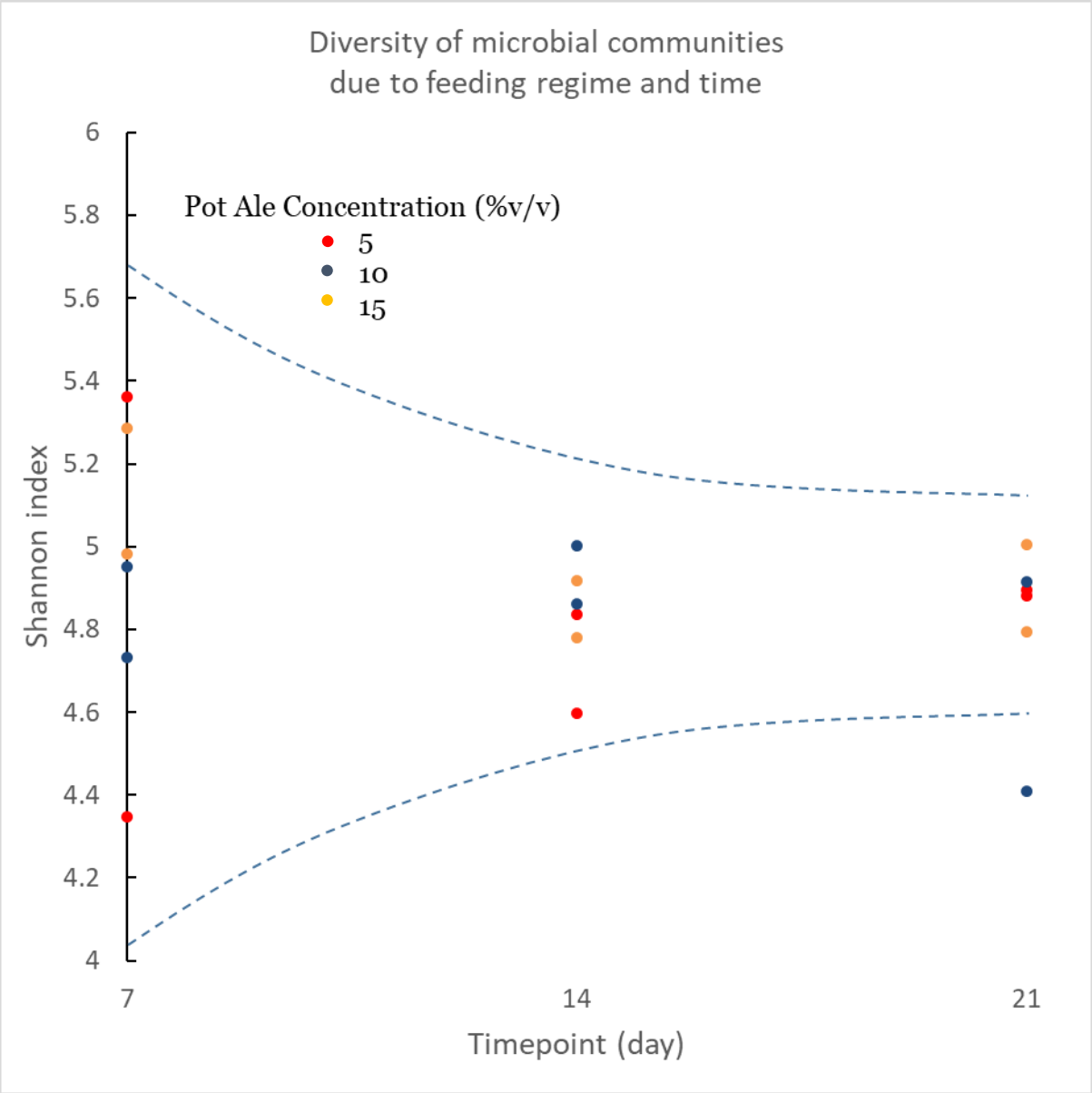
The microbial community grown from the material attached to the polymer displayed methanogenic activity throughout the experiment, confirming its suitability as an inoculum. Methane production of this material could be increased by supplementation of carbon due to its low C:N ratio of 8:1 and recalcitrant nature of the cPAM backbone. DNA analysis showed that similar core communities form when subjected to similar concentrations of pot ale. Further experiments should involve metagenomics or metatranscriptomics to identify specific genes and pathways relating to pot ale methanisation or use an improved measure of microbial concentration as population growth and death is not detected via 16S rRNA gene sequencing. These experiments highlight the plausibility of using the material precipitated alongside flocculent polymers as a methanogenic inoculum for anaerobic digestion in addition to trialling the methanisation of pot ale under varying organic loading rates. However poor methane yields, attributed to the toxic nature of cPAM, reduce the suitability of the inoculum compared to non-toxin containing alternatives such as wastewater treatment plant

sludge. Due to this poor methane yield, further experiments involving the anaerobic digestion of pot ale (already nitrogen rich) will avoid using this nitrogen rich inhibitory inoculum and opt for a more traditional reactor start up.

2.8 Acknowledgements

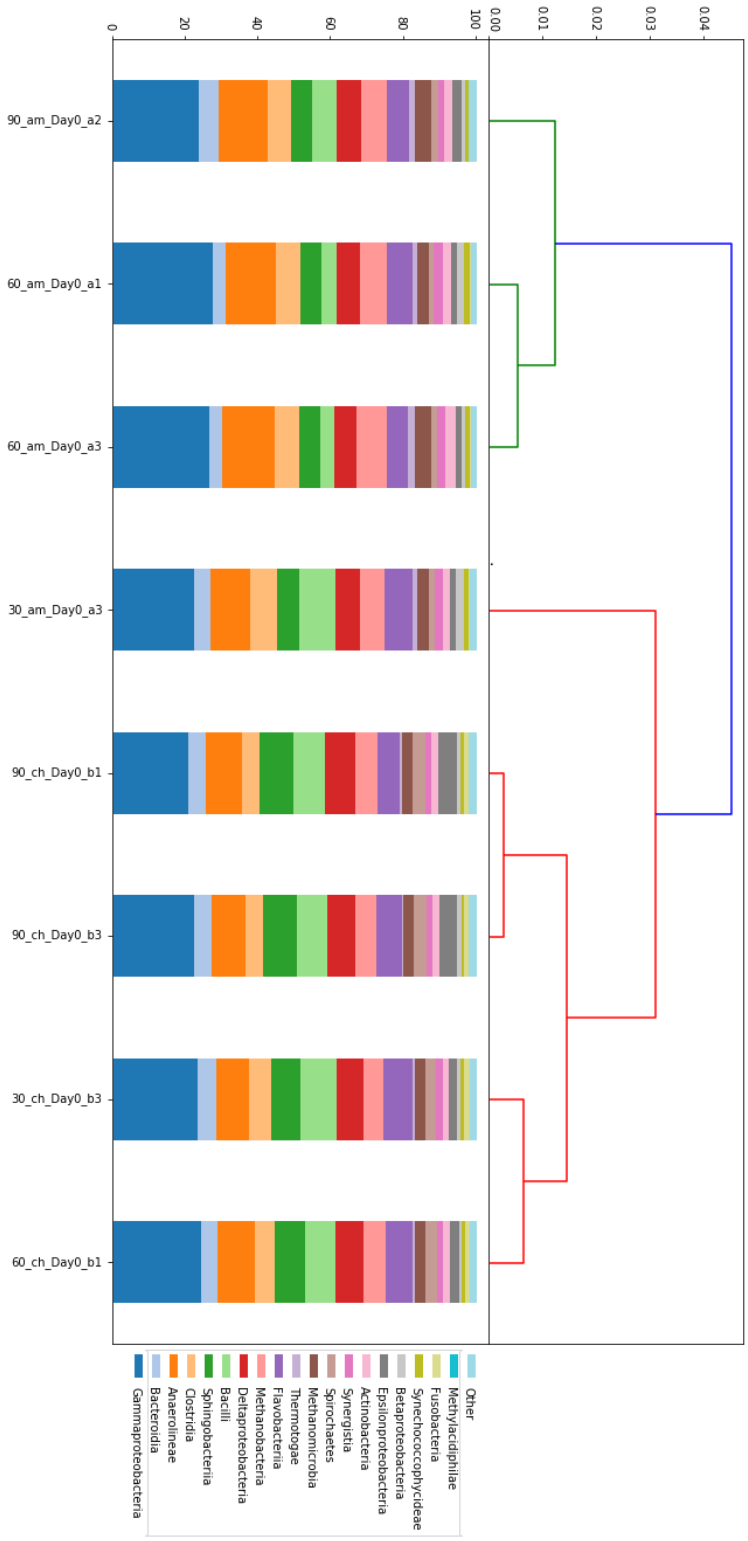
We thank Sally James (Genomics facility at the University of York) for support with library preparation and sequencing and Darren Phillips (Biorenewables Development Centre) for running the metal analysis on the ICP-MS. JPJC is a Royal Society Industry Fellow. JFR is supported by a BBSRC iCASE studentship.

2.9 Supporting Information



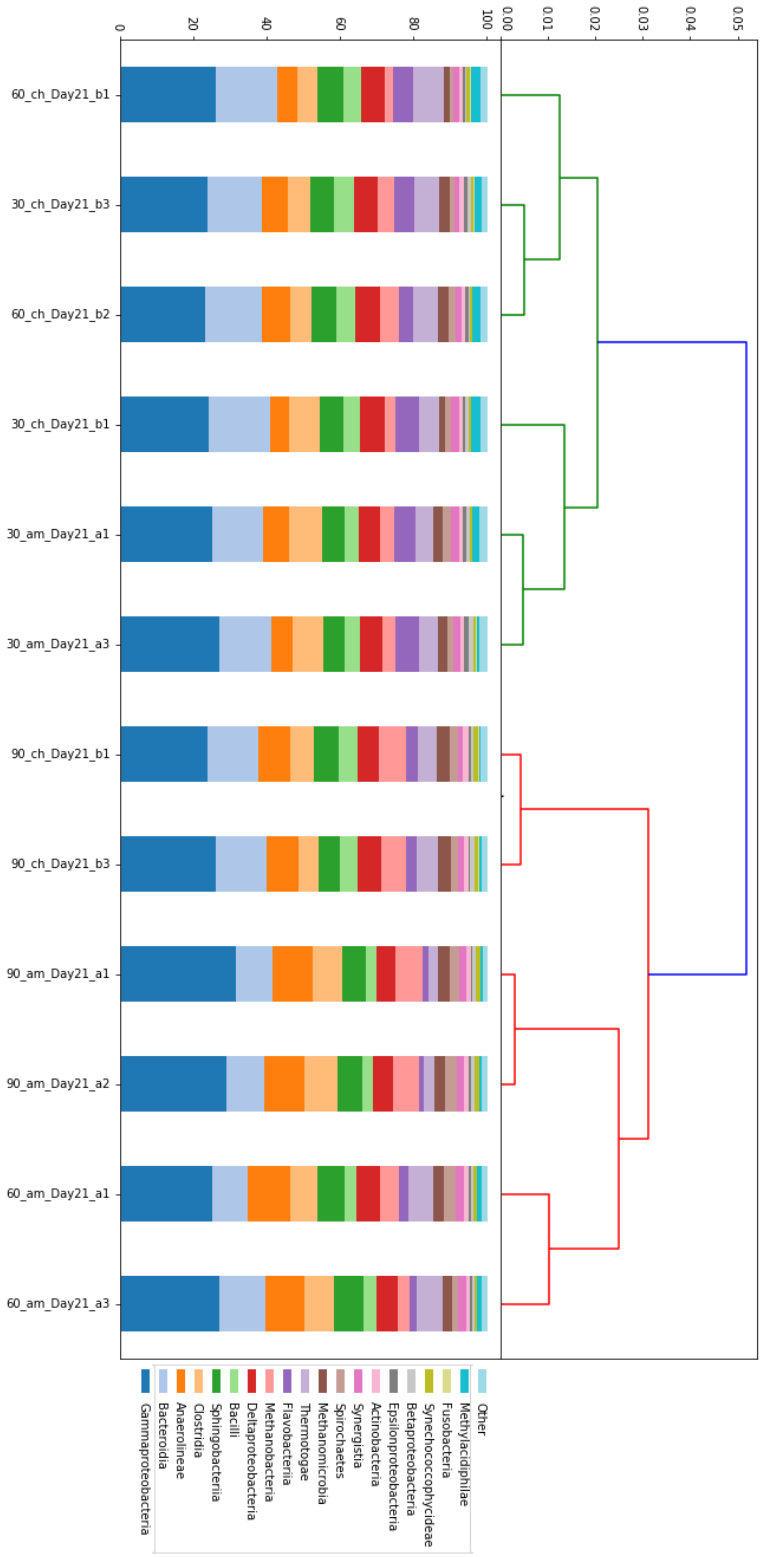
Supplementary Figure 2.1. Microbial diversity over the course of the BMP tests

Following addition of inoculum, a selective pressure, a funnelling effect (visualised by the dashed blue line) on the species diversity was imposed that reduced the Shannon index standard deviation from 0.34 to 0.13 then 0.19. To quantify the decrease in diversity of Shannon index values the difference between the 75% (Q3) and 25% (Q1) quartile ranges were calculated. Over the course of the experiment the difference between Q1 and Q3 decreased from 0.42 to 0.11 and then further to 0.09 showing the rapid community diversity decrease following feeding on day 7. Red = 5% blue = 10% and orange = 15% pot ale loading rates (v/v).



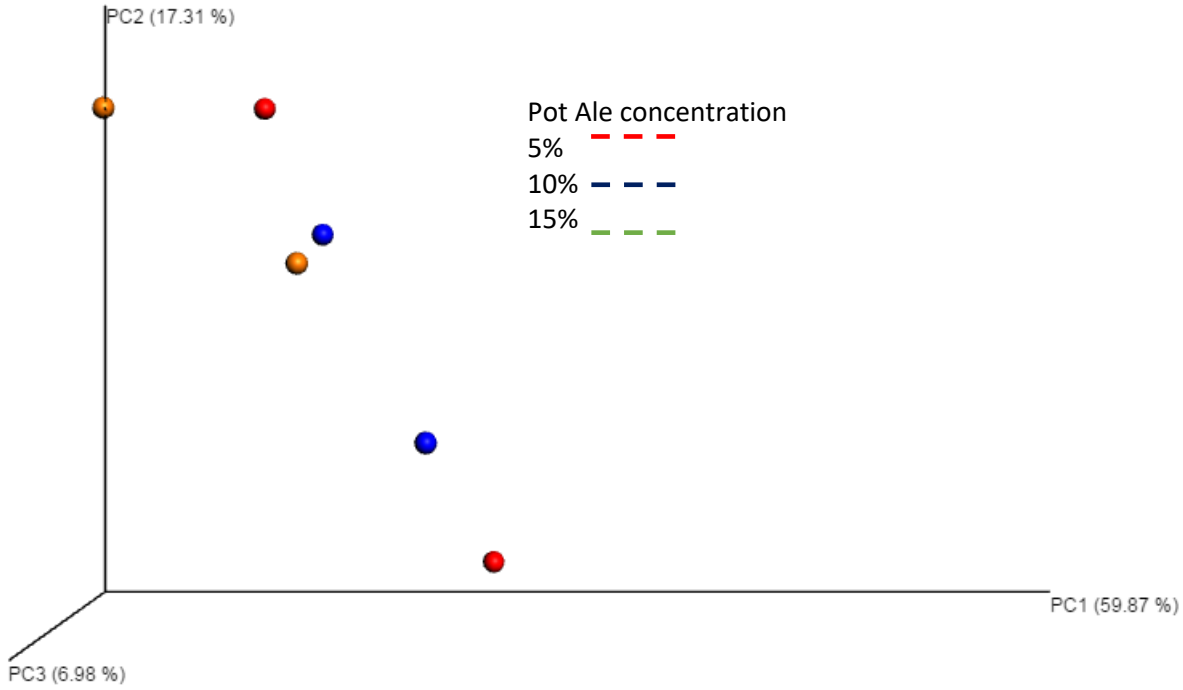
Supplementary Figure 2.2. Storage D0 – Relative abundance dendrogram

Samples clustered on the whole relating to their storage condition of either ambient (am) or chilled (ch) apart from the ambient samples stored for 30 days likely due to the minimal community change which would occur over this time.



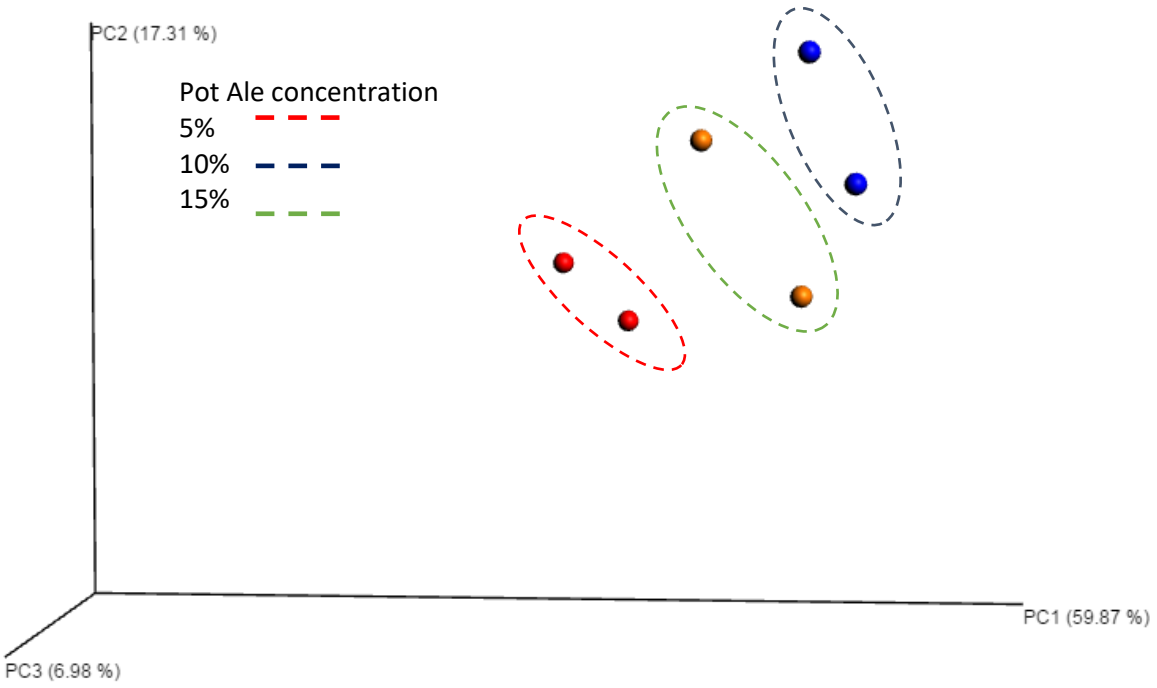
Supplementary Figure 2.3. Storage D21 – Relative abundance dendrogram

Samples clustered on the whole relating to their storage condition of either ambient (am) or chilled (ch) although more convergence (less differentiation was seen due to storage temperature compared to the initial day 0 samples).



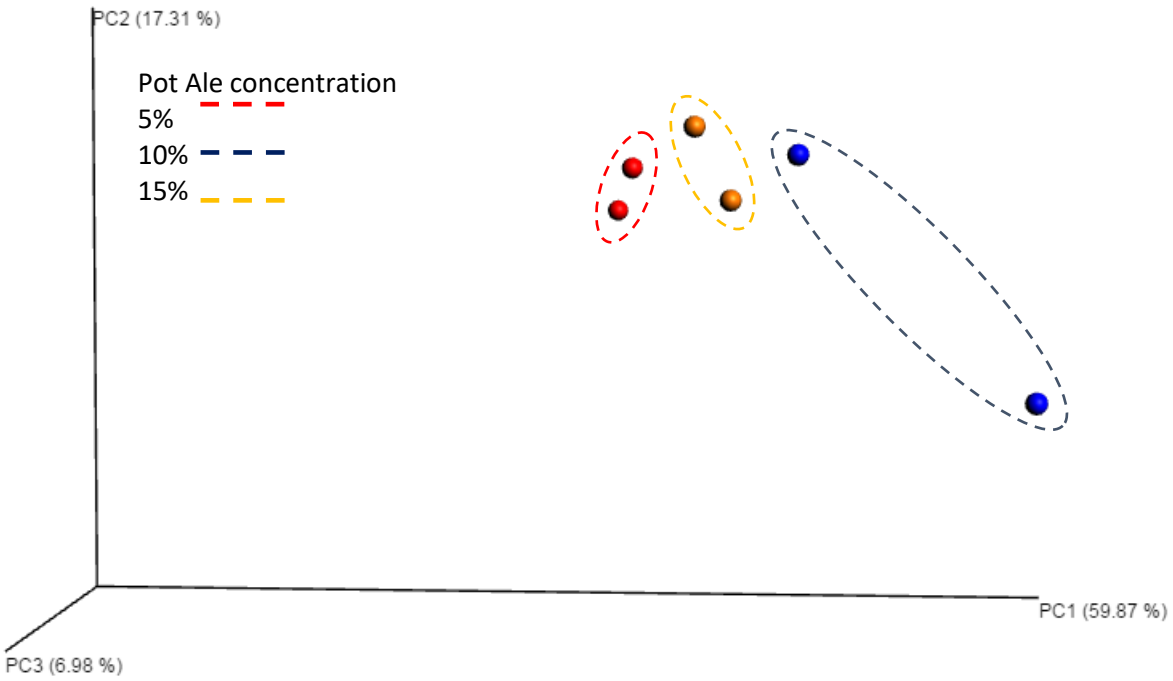
Supplementary Figure 2.4. PCoA plot on day 7

Supplementary Figure 2.4. 3D Principle coordinate analysis of communities fed varying concentrations of pot ale examines the degree of phylogenetic similarity between samples subject to varying concentrations of pot ale over the duration of the experiment using the weighted UniFrac value. Spheres present day 7 samples, community composition at the start of the experiment following starvation. The random dispersion of each sample to its respective partner indicate a high degree of community similarity when evaluating the effects of starvation prior to feeding.



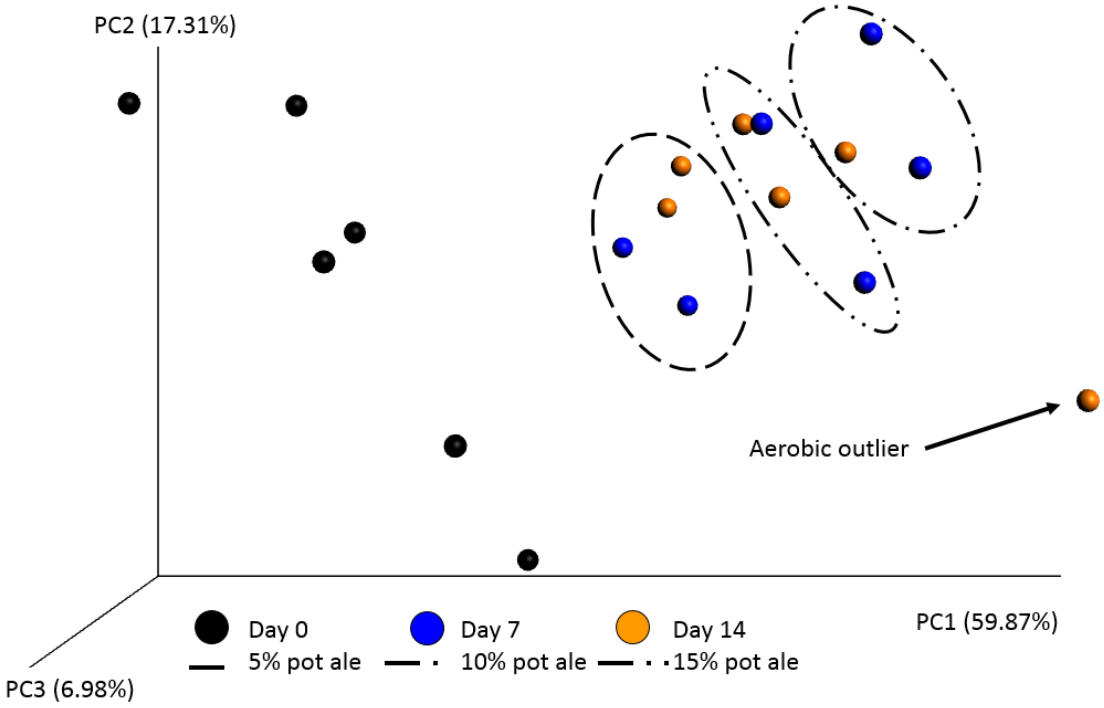
Supplementary Figure 2.5. PCoA plot on day 14

Supplementary Figure 2.5. 3D Principle coordinate analysis of communities fed varying concentrations of pot ale examines the degree of phylogenetic similarity between samples subject to varying concentrations of pot ale over the duration of the experiment using the weighted UniFrac value. Spheres present day 14 samples, community composition at the end of the experiment following feeding. The proximity of each sample to its respective partner indicate a high degree of community similarity when fed the respective pot ale concentration.

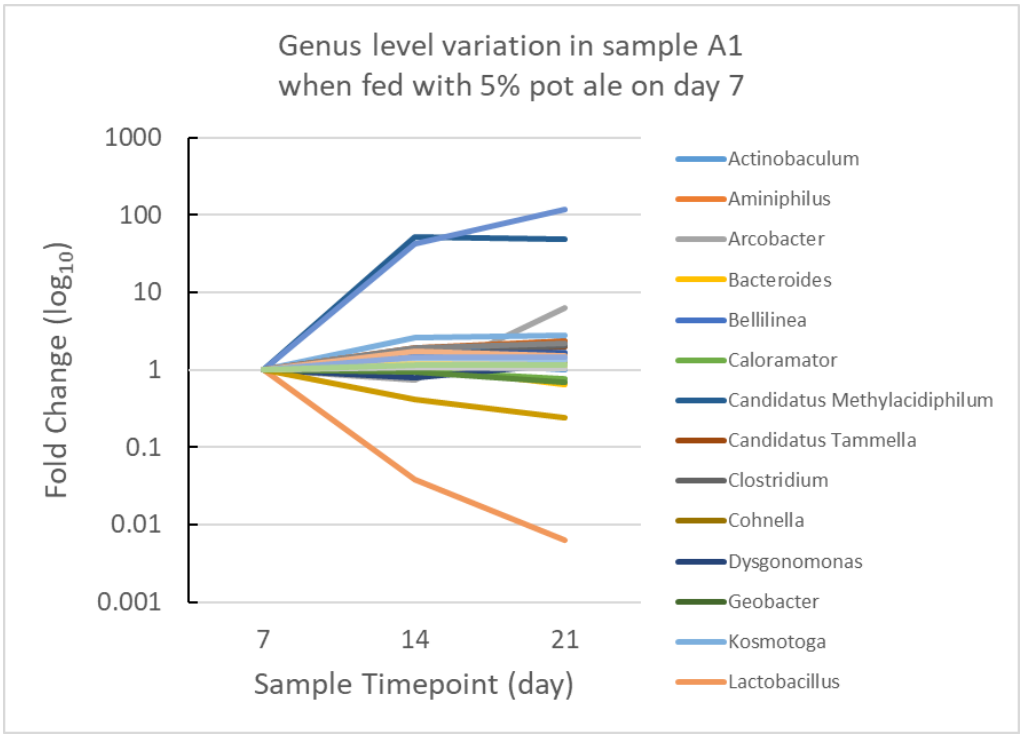


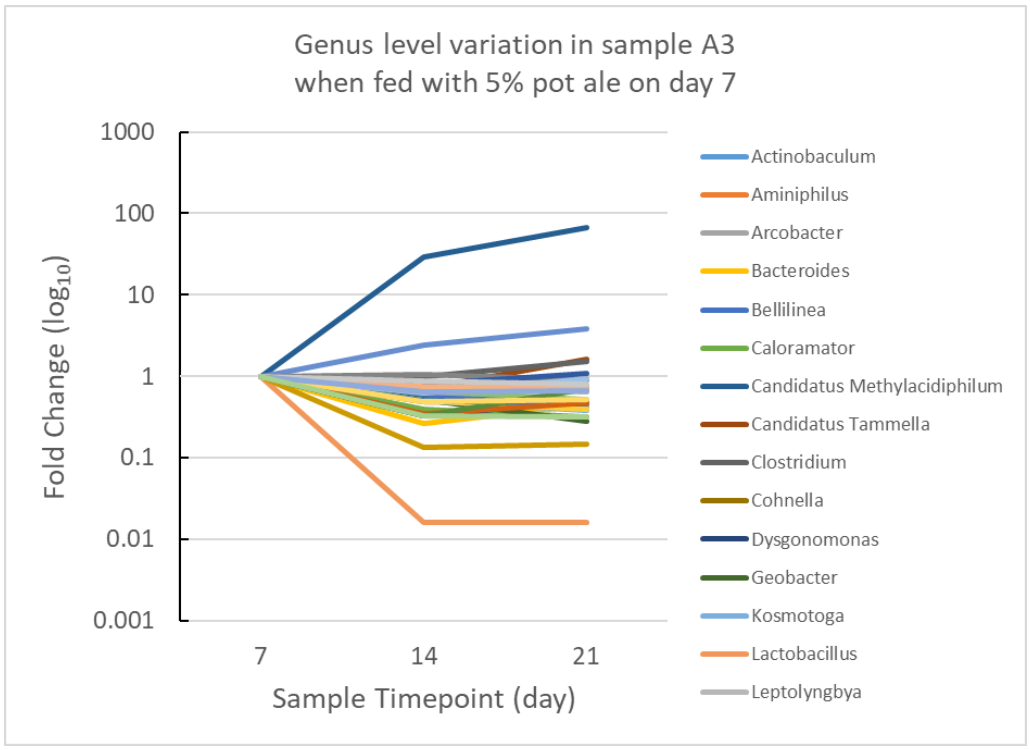
Supplementary Figure 2.6. PCoA plot on day 21

Supplementary Figure 2.6. 3D Principle coordinate analysis of communities fed varying concentrations of pot ale examines the degree of phylogenetic similarity between samples subject to varying concentrations of pot ale over the duration of the experiment using the weighted UniFrac value. Spheres present day 21 samples, community composition at the end of the experiment following feeding. The proximity of each sample to its respective partner indicate a high degree of community similarity when fed the respective pot ale concentration.



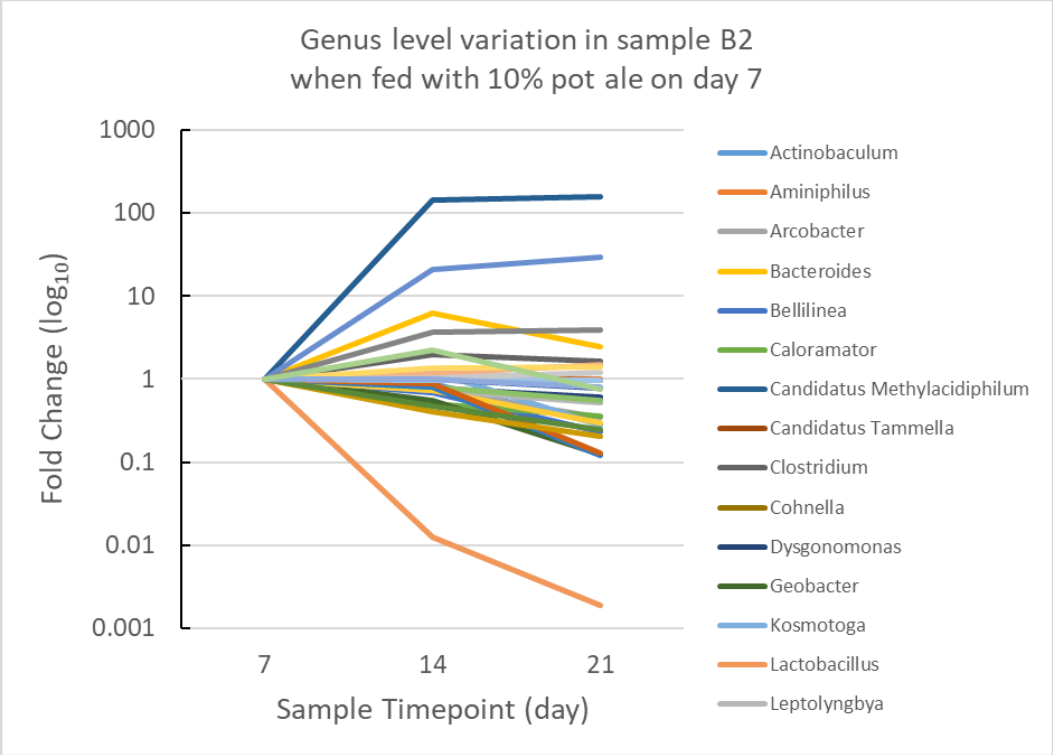
Supplementary Figure 2.7. 3D Principle coordinate analysis of communities fed varying concentrations of pot ale examines the degree of phylogenetic similarity between samples subject to varying concentrations of pot ale over the duration of the experiment using the weighted UniFrac value. Black spheres present day 0 samples, following a one week starvation period, and not subject to feeding pressure and blue and orange represent samples on day 7 and 14. A higher degree of phylogenetic similarity is observed on day 7 following pot ale feeding, regardless of concentration, with evidence to suggest presence of sub-clusters dependant on feeding concentration. By day 14 following feeding a tighter cluster is observed when compared to samples on day 7 indicating the presence of a core microbial community. Further subclusters are observed which correspond to the pot ale concentration. A 10% pot ale outlier was caused by accidental introduction of oxygen to the experimental system, with increases in aerotolerant and facultative anaerobes observed by further investigation.

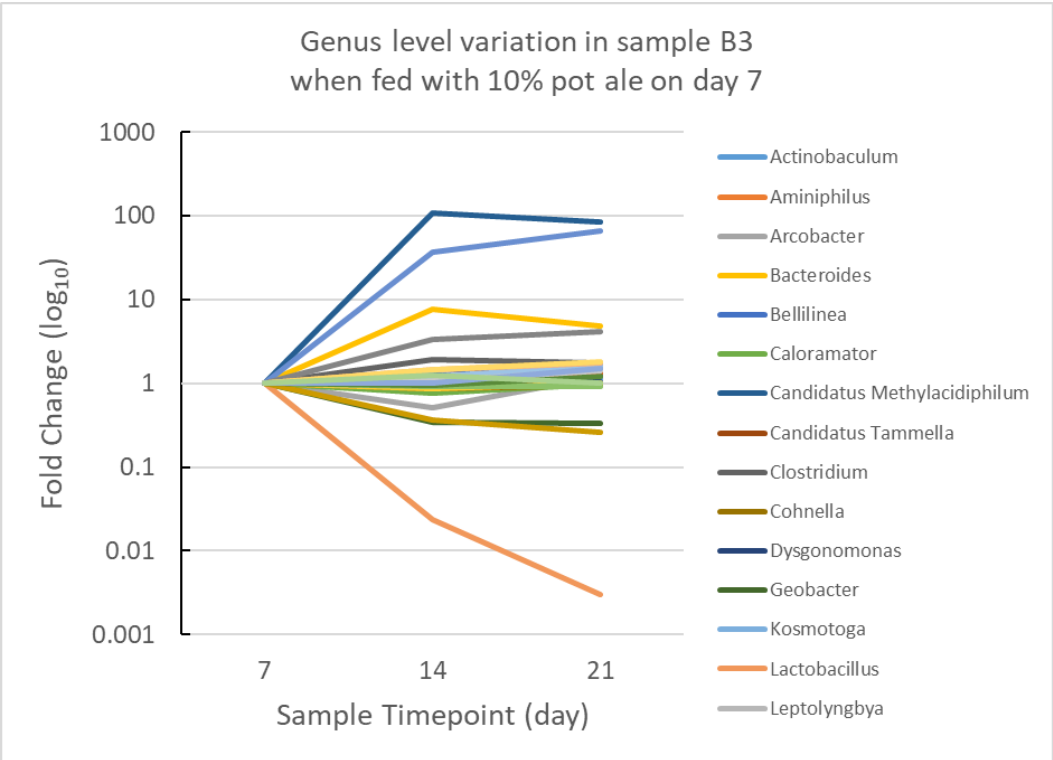




Supplementary Figure 2.8. Changes within 5% replicates

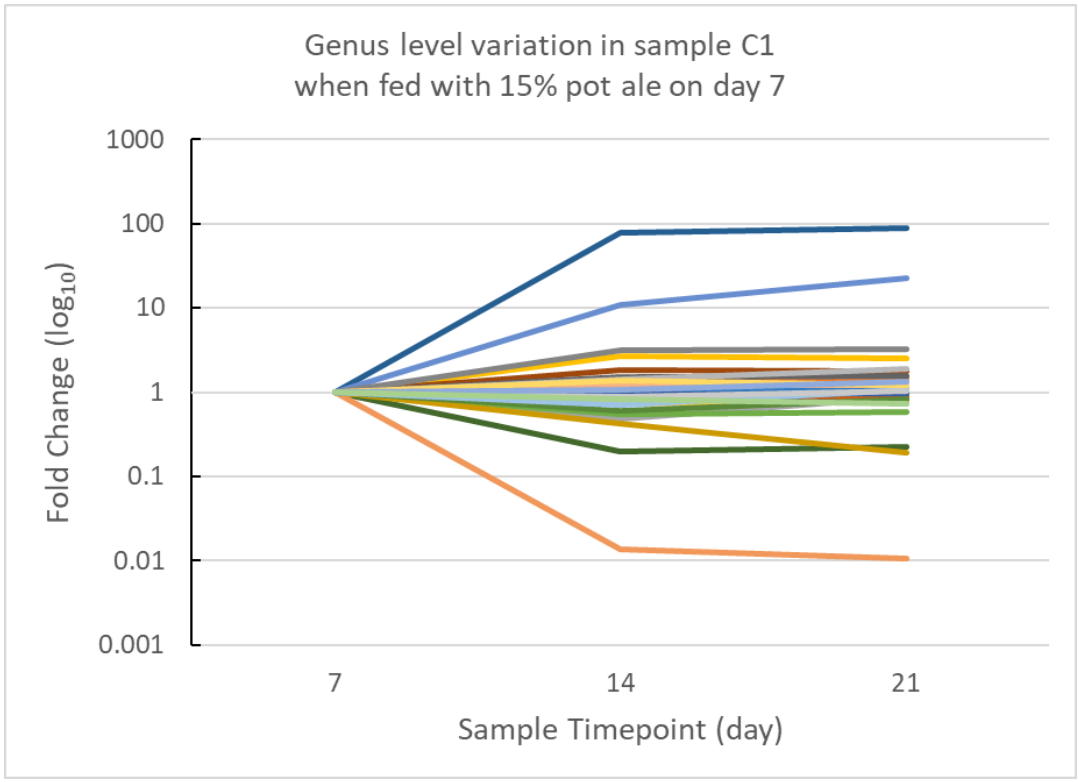
Select genera were compared to the genus *Esherecia* following sludge being spiked with a known amount of *E. coli.* Degree of change is represented as fold change (log_{10}).

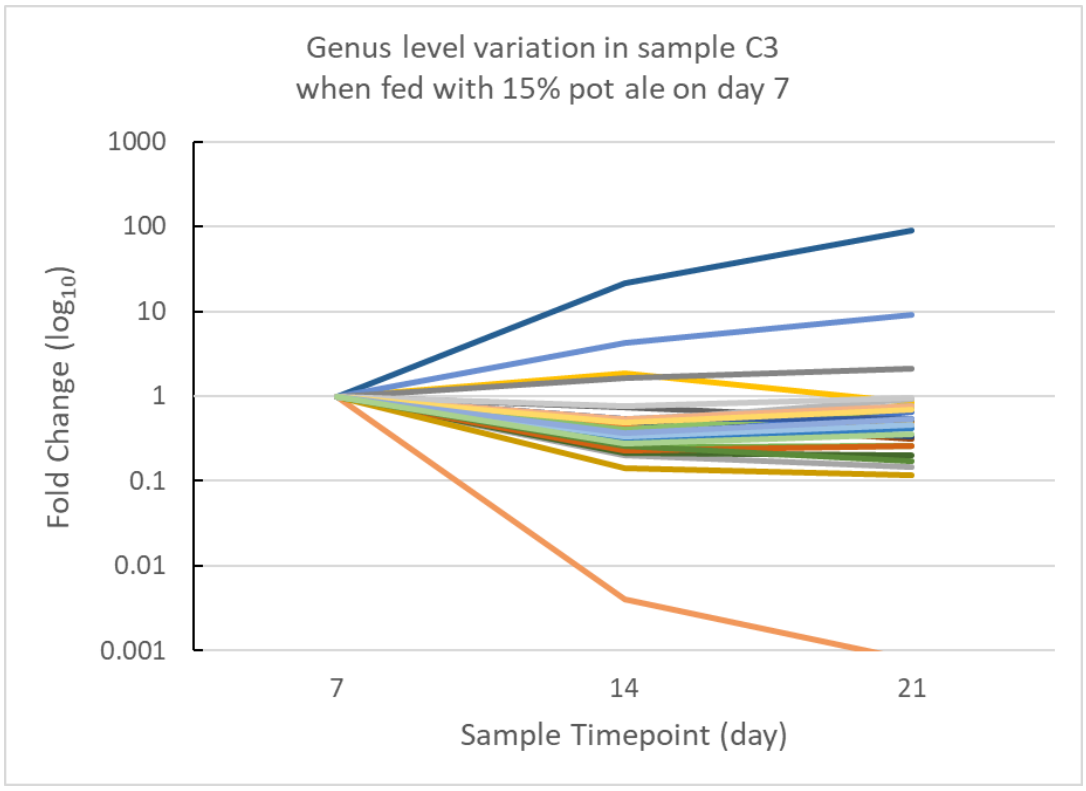




Supplementary Figure 2.9. Changes within 10% replicates

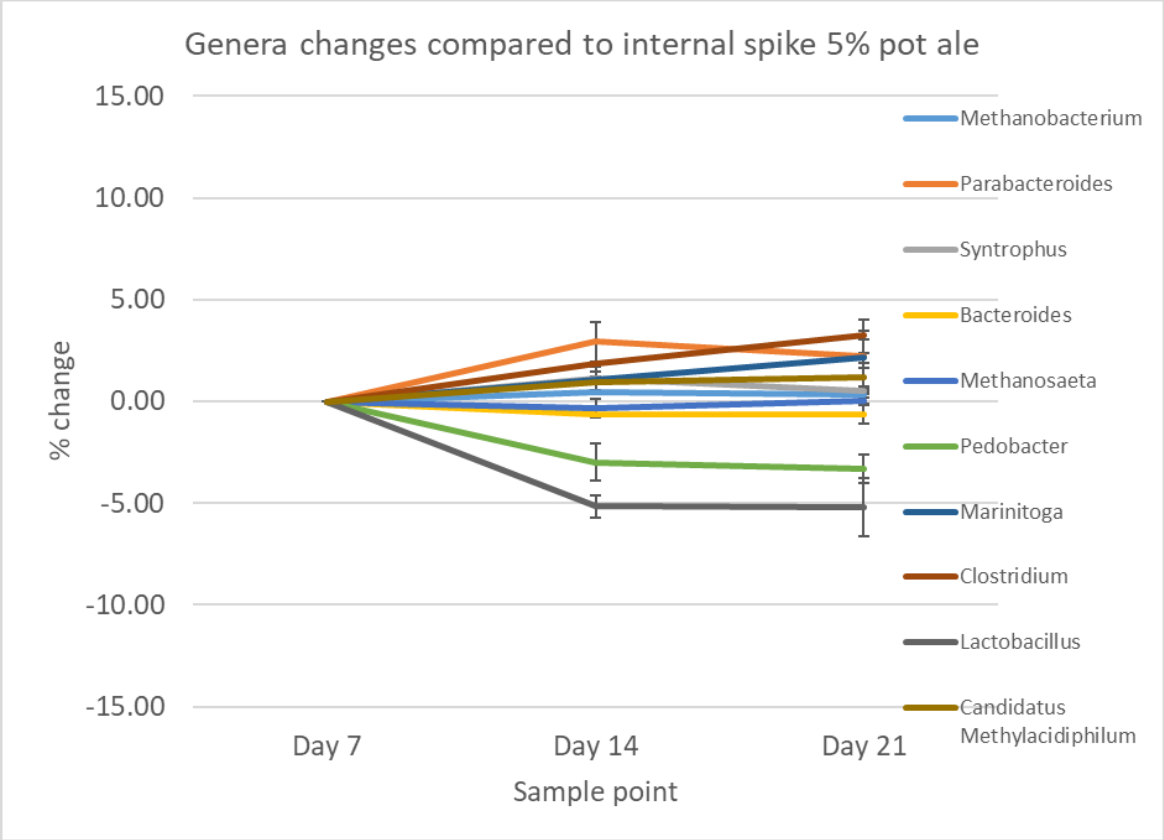
Select genera were compared to the genus *Esherecia* following sludge being spiked with a known amount of *E. coli.* Degree of change is represented as fold change (log_{10}).





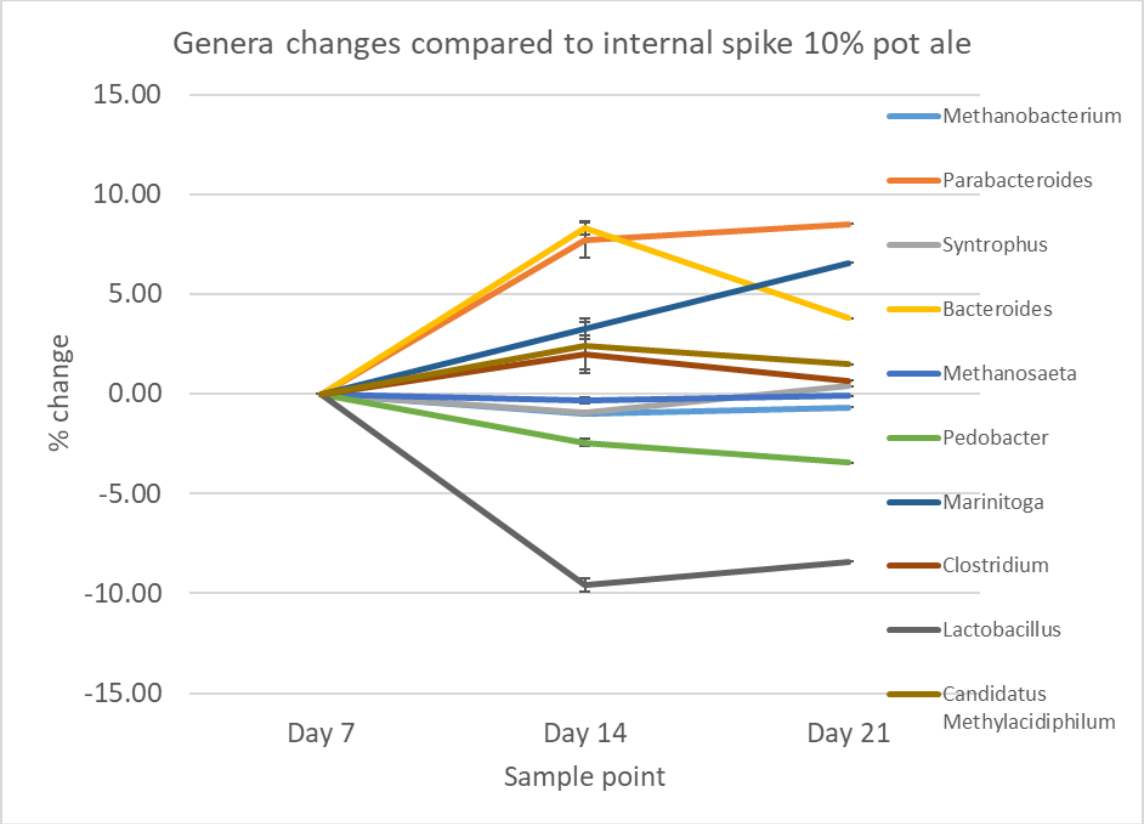
Supplementary Figure 2.10. Changes within 15% replicates

Select genera were compared to the genus *Esherecia* following sludge being spiked with a known amount of *E. coli.* Degree of change is represented as fold change (log_{10}).



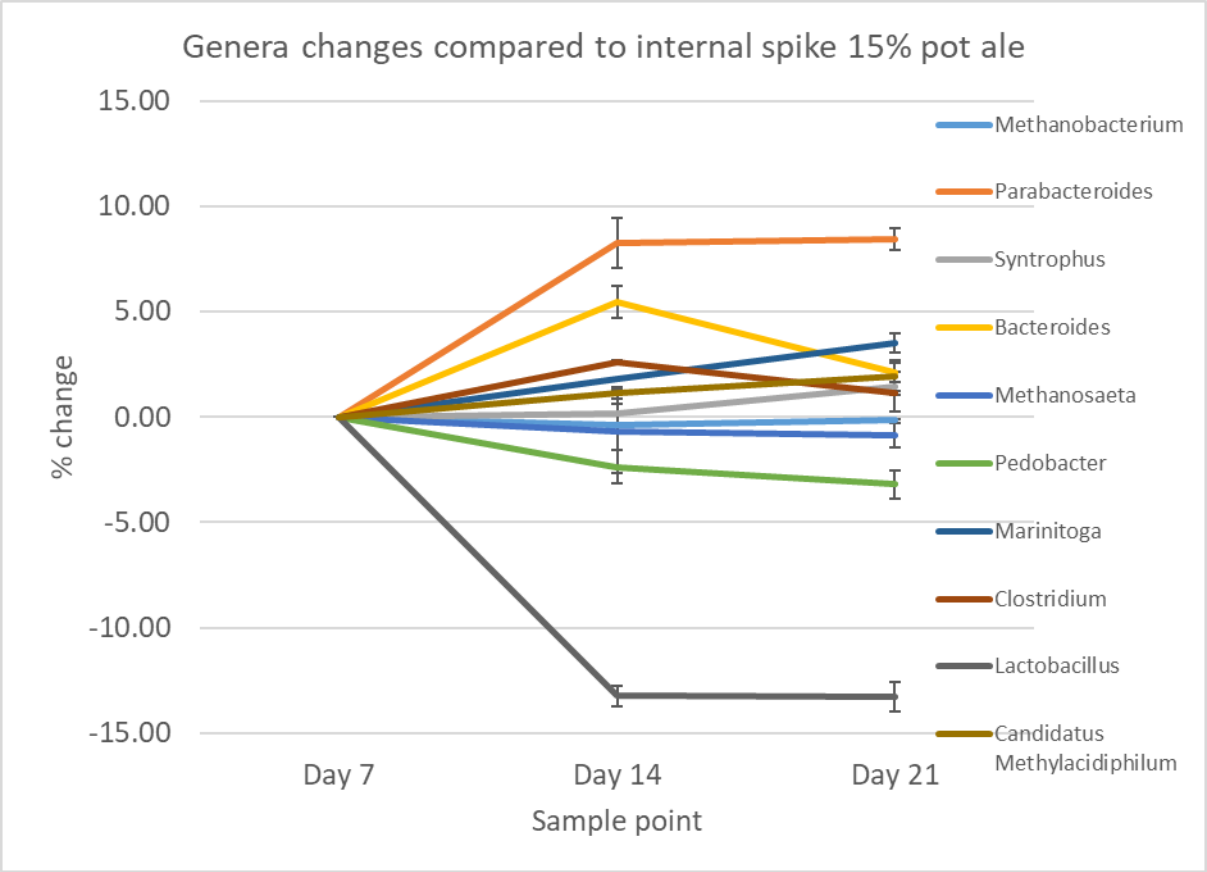
Supplementary Figure 2.11. Absolute change in select genera for 5% replicates.

By day 21 genera differences compared to the internal spike were *Methanobacterium* 0.35 ± 1.83 (SD), *Parabacteroides* 2.20 ±0.05, *Syntrophus* 0.53 ±0.46, *Bacteroides* -0.64 ±0.16, *Methanosaeta* 0.04 ±0.68, *Pedobacter* -3.32 ±0.24, *Marinitoga* 2.15 ±0.21, *Clostridium* 3.25 ±1.46, *Lactobacillus* -5.20 ±0.48, Candidatus *Methylacidiphilum* 1.19 ±0.19,



Supplementary Figure 2.12. Absolute change in select genera for 10% replicates

By day 21 genera differences compared to the internal spike were Methanobacterium -0.68, Parabacteroides 8.51, Syntrophus 0.38, Bacteroides 3.80, Methanosaeta -0.11, Pedobacter -3.45, Marinitoga 6.56, Clostridium 0.65, Lactobacillus -8.40, Candidatus Methylacidiphilum 1.48



Supplementary Figure 2.13. Absolute change in select genera for 15% replicates

Methanobacterium -0.12 ±0.52, *Parabacteroides* 8.44 ±1.20, *Syntrophus* 1.46 ±0.45, *Bacteroides* 2.13 ±0.59, *Methanosaeta* -0.87 ±0.66, *Pedobacter* -3.21 ±0.47, *Marinitoga* 3.50 ±0.08, *Clostridium* 1.14 ±0.72, *Lactobacillus* -13.27 ±0.25, Candidatus *Methylacidiphilum* 1.91 ±0.08,

2.10 File Guide

GenusComparison.xlsx

A file containing the taxonomy assigned OTUs and their respective abundance. In addition, sheets are included that present the taxa ratios within samples when normalised to the internal spike.

otu_table.biom

The file produced from QIIME preproccesing containing to OTU table

Qiime VISPP

A folder containing processed rarefaction results in spreadsheet form (CompiledMetrics.xlsx) in addition to the UniFrac visualisations (bdev_even17200>weighted_unifrac_emperor_pcoa_plot>index.html

CompiledMetrics.xlsx

A file containing rarefaction data

Weighted_unifrac_dm.txt

The weighted unifrac distance matrix

All DNA reads are submitted in the BioProject PRJNA393625

Chapter 3

Microbe-spiked anaerobic sludge samples normalise extraction conditions and introduce a degree of quantification into amplification based sequencing strategies

3.1 Abstract

Taxonomic profiling of complex microbial environments such as anaerobic digesters routinely use the amplification of highly conserved genes, such as regions of 16S rRNA subunits, followed by DNA sequencing to identify microbes, their relative abundance and infer their importance for that specific biological system to examine community structure and function. For instance, the increased relative abundance of methanogens such as *Methanosaeta*, can be used to indicate community health for anaerobic digestion with the assumption that increased abundance of methanogens results in increased biogas production. Results of this analysis are typically presented in terms of relative abundance which presents data in terms of ratios, e.g, 25% Methanogens and 75% Bacteria. However, two samples may have identical relative abundance profiles but one sample may have an overall more concentrated population. Differences in extraction methodology may also result in biased populations being extracted prior to amplification. In an attempt to both normalise the extraction and subsequent analysis between samples we have developed an enhanced method for analysis of microbial communities when performing marker amplification which involves the external spiking of a characterised foreign species. This method allows the researcher to both normalise intrastudy variables, such as differences in biomass between samples, and interstudy variables such as varying microbial richness across environments, varying extraction methodologies and sample composition, in addition to an enhanced degree of microbial quantification compared to conventional abundance based approaches. Firstly we characterised our microbial spikes ploidy using flow cytometry then identified a unique marker sequence not present in the highly complex sludge community and used the marker sequence to examine sources of variation, such as different sample weights, in community gDNA extractions through the use of qPCR. This allowed determination of sample biological and recalcitrant content in response to spike level detected to calculate level of microbial "richness". Unfortunately, although primers amplified the 16S rRNA gene of the *Sulfolobus solfataricus* microbial spike independently and qPCR detected the organism in sludge samples, no *S. solfataricus* was observed in samples following 16S rRNA sequencing. An alternative amplification strategy was adopted that aimed to amplify the entirety of the ~5 kb rRNA operon for long read sequencing on the Oxford Nanopore MinION, but reproducible amplification could not be attained using pure *S. solfataricus* gDNA.

3.2 Keywords

Microbial ecology, 16S rRNA gene PCR, microbial quantification, microbial spike

3.3 Introduction

Exploration of microbial communities in complex environments such as salt marshes, peat bogs, composting matter, geothermal hot springs, soil and anaerobic digestion present a challenging task due to differences in chemical, physical and biological composition. Amplification of the 16S ribosomal subunit (16S rRNA) gene, a highly conserved gene due to its inclusion in ribosome, allows discrimination between organisms at the species level (Weisburg et al., 1991). This gene amplification strategy has developed into one of the most extensively used techniques to analyse microbial ecology as it provides detailed information regarding what species are present and how the species rank in terms of abundance expressed as percentage of total reads.

The widespread use of this technique can be attributed to the fact that it is a relatively straight forward yet powerful technique that involves DNA purification followed by amplification of a specific region of the 16S rRNA gene using PCR then identifying the closest phylogenetic matching organism by DNA sequencing or, less commonly in recent times due to the cheap price of DNA sequencing, denaturing gradient gel electrophoresis (DGGE) where identical sized fragments are separated based on their mobility in increasing denaturing conditions. Notable successful application of this technique has resulted in; identification of 27 new species of disease causing bacteria (Drancourt et al., 2004), reclassification of whole family taxa (Everett et al., 1999) and classifying unidentified bacteria or confirming isolate identify in addition to the complex microbial ecology type analyses of environments such as peat bogs, salt marshes, soil and anaerobic digesters (Beazley et al., 2012; Sizova et al., 2003; Žifčáková et al., 2016; Ziganshin et al., 2013).

Although the 16S rRNA gene amplification technique has many broad applications there are certain caveats that the technique processes which liken it to a double edged sword. On one hand, presence of low abundance and unculturable organisms can be detected from environmental samples by amplifying small quantities of DNA which is crucial for tasks such as *Shewanella* disease prevention and detection in water bodies (Holt et al., 2005). On the other hand, actual quantification of organisms becomes ambiguous when you consider the possible variation introduced through; bias introduced through “universal” primers, DNA amplification steps performed when following this method

(between 25-30), differences in optimum annealing temperatures between organisms and polymerase fidelity (Sipos et al., 2007; Suzuki and Giovannoni, 1996). Additional difficulty arises when the investigation of an extremely diverse microbial environment is called for since 16S rRNA gene primers need to be both specific enough to not amplify off-target stretches of DNA but also broad enough to cover the majority of taxa that are most likely present. A combination of *in-silico* and experimental data has previously been generated in investigating where the fine line lies between primer pair accuracy, precision and spread for analysis of mixed communities (archaeal and bacterial or exclusively bacterial) (Klindworth et al., 2013; Marchesi et al., 1998; Watanabe et al., 2001). Accumulative error from the previously mentioned sources makes true comparison between studies and experiments extremely difficult, especially when variables such as reagents, extraction method and duration, sample volume, equipment, primer sequence and amplification profile vary from study to study (Díaz et al., 2007; Ho et al., 2013; Smith et al., 2015; Wilkins et al., 2015; Ziganshin et al., 2013). The scientific community analysing a similar microbial community to anaerobic digestion microbial community, the soil microbial community, addressed these concerns and decided the use of an exclusive protocol, to allow for important fundamentals such as interstudy comparison (Gilbert et al., 2014).

This difficulty in comparison is compounded by how the 16S rRNA gene sequencing results are presented, in multiple microbial ecology fields, by proportional abundance. For instance, comparison of sample set 1 & 2 reveals a community of identical organisms in similar proportions, e.g 25% Methanogens and 75% Bacteria. However, it does not reveal that the cell numbers in sample set 1 are far greater than that of 2 and thus the metabolic turnover capacity of sample set 1 is far greater. This is especially important for assessing industrially relevant communities whose performance is reliant on both the species present and their relative concentration. One such community is that in anaerobic digestion communities where the industrial production of biogas relies on an established methanogen community. This lack of true population distribution and density information is an issue for both the microbial ecology field as a whole but especially analysis of anaerobic digestion microbial communities in relation to their performance as data presented using relative abundance does not reveal if the community is low in concentration with high activity (quality) or vice versa (quantity).

Other techniques can be used for analysing microbial composition such as; adenosine tri-phosphate measurements, flow cytometry, quantitative real-time PCR (qPCR), phospholipid fatty acids examination and microbial biomass carbon, but none of these techniques provide the same level of understanding that 16S rRNA gene sequencing offers (Zhang et al., 2017). Concerned with the lack of

quantification Smets *et al* developed a protocol that normalised the extraction method, in addition to amplification and sequencing, with the addition of purified foreign DNA to biomass prior to extraction belonging to organisms commonly found associated with marine animals (*Aliivibrio fischeri*) and in geothermal environments (*Thermus thermophilus*) for analysis of soil communities (Smets et al., 2016). One limitation of this study was the purified DNA being added to the biomass sample prior to extraction thus exposing the marker gene to extracellular nucleases, even more so an issue with digestate sludge due to the high concentration of microbial biomass. Most importantly, this study showed that community composition profile was unchanged by the addition of a foreign spike. Addition of a foreign DNA source was also visited by Stämmler *et al* and K. Piwosz *et al* in their extensive investigation of the human and water microbiome respectively (Piwosz et al., 2018; Stämmler et al., 2016). In their study, literature values were used to calculate spike cell number and 16S rRNA gene copy number, then an aliquot of whole bacterial cells containing their 16S rRNA spike gene were added to unlysed stool samples, DNA extracted then analysed using qPCR. By placing the point of normalisation at the very beginning of sample processing intra-study sample comparison could be achieved since, although samples are normalised, 16S rRNA gene copy number could vary depending on cell cycle phase when the method is repeated. Tourlousse *et al* furthered the concept of using a marker gene not commonly found in the environmental sample by adding synthetic 16S rRNA gene sequences to the biomass sample following cell lysis and performing conventional 16S rRNA gene amplicon sequencing (Tourlousse et al., 2017). A synthetic 16S rRNA gene would reduce the chance of accidental crossover with species present in unknown environmental samples. This allows their technique to be used in environments typically home to spike-in organisms such as hydrothermal vents, hot springs and saline environments. Their method involved adding the spike to a sludge sample following cell lysis thus reducing marker gene exposure to extracellular nucleases. However, as the marker was added post cell lysis this meant that the extraction method was not normalised between samples or indeed studies which introduces variability when analysing samples of different materials e.g sand vs soil.

By identifying shared points and areas for improvement from these studies, it was clear that an ideal spike-in method should have the following characteristics; be added prior to lysis to start normalisation at the DNA extraction step, be comprised of whole cells rather than purified DNA to enable comparison of extraction techniques, be foreign to the environment being investigated or ideally be synthetic, have the marker gene copy number present in the spike quantified by measuring cell ploidy. By investigating the use of an external foreign species spike in a highly complex microbial community we aimed to build upon previous work in attempting to address the quantification issue

in microbial ecology analysis and to allow normalisation across the whole process, from DNA extraction to visualisation to provide advantages in determining true organismal abundance and microbial richness of samples.

3.4 Materials and Methods

3.4.1 Characterisation of *Sulfolobus solfataricus*' ploidy

Sulfolobus solfataricus samples for flow cytometry were prepared using an ethanol fixation and resuspension in 10mM Tris and 10mM $MgCl_2$ buffer as described previously (Bernander and Poplawski, 1997). Cells were then diluted using an appropriate volume of the same buffer. A bead standard was first run through the CytoFlex cytometer to calibrate the cell counting detection software. A blank containing dye and resuspension buffer was initially run to ensure purity of reagents and cleanliness of cytometer lines. From the cell count and relative fluorescence the spike was determined to contain 609,200 cells/ uL with the calculated number of genome copies at 1,012,800 equating to an average of 1.66 genome copies per cell. With the microbial spike fully characterised the exact number of genome copies added to samples at the stage of DNA extraction could be calculated.

To quantify the number of single stranded binding protein (SSBP) copies detected a standard curve was created from PCR amplified SSBP DNA product which was purified using AMPure beads and put through an additional PCR and purification cycle to yield pure product. Actual SSBP recovery was calculated by multiplying the number of SSBP copies / ng DNA calculated by the qPCR by the total amount of DNA present following the extraction. 25 ng of extracted DNA was used as the template for the qPCR reaction to examine copy number of SSBP and showed that extraction efficiency typically varied between 0.5 and 1.5% (Supplementary Figure 3.1).

3.4.2 Marker validation using PCR

A *Sulfolobus solfataricus* unique marker gene was identified using PCR to screen 5 primer pairs that were designed for genes flagged as unique to *Sulfolobus solfataricus* during the literature search.

Primer pairs were used to amplify DNA samples from *Sulfolobus solfataricus* spiked sludge (100 µl spike or 1,012,800 genome copies, in 200 mg sludge) DNA and sludge DNA only in triplicate. A negative control was included containing no template DNA and two positive controls were included to amplify the V4 16S rRNA gene region of both the *Sulfolobus solfataricus* and the sludge community using the A519F and 802R primer pair, all in duplicate (Klindworth et al., 2013). PCR reactions contained 50ng template gDNA, 1.25uL 10uM of each primer, 0.5uL 10 mM DNTPs, 0.25 µl (0.5 units) Q5 polymerase and rest water. Amplification was performed using a Techne Prime thermocycler with an initial denaturation for 3 minutes at 94°C followed by thirty cycles of denaturation at 94°C for 45 seconds, annealing at 50°C for 60 seconds and elongation at 72°C for 90 seconds followed by a final elongation stage for 5 mins. Primer pair selectivity was examined by running the amplification product on a 1% agarose gel stained with SYBR safe (S33102, Thermo Fisher) and visualised under UV illumination. Presence of multiple bands in the spiked lanes or any bands in the sludge only lane was interpreted as off-target amplification and non-desirable.

Once SSBP was validated for use as a unique marker, the initial amplification product was purified using AMPure beads (A63880, Beckman Coulter) using manufacturer's instructions and used as template DNA for a second round of PCR amplification using the same thermocycling settings and purification stage previously mentioned. This additional purification step was performed to ensure only SSBP amplification product was present in the solution prior to use as a standard in qPCR. The DNA concentration was quantified using Qubit HS dsDNA reagent (Q32854, Thermo Fisher) on the Qubit 4 Fluorometer.

3.4.3 Anaerobic sludge communities

Anaerobic sludge samples used for the initial spike load optimisation, DNA extraction method comparison, and biomass variation were collected from anaerobic digester 4 in the Naburn waste water treatment plant located near York, UK at coordinates 53.914156 N, -1.085661 W in October 2017. Sludge was collected in a barrel and kept anaerobic at 4°C until required when it was centrifuged at 6000 RCF for 15 minutes to pellet biomass prior to weighing and DNA extraction.

Biomass samples for the community concentration investigation were collected from the same Naburn digester in November 2018. A 50ml aliquot of sludge was pelleted as described previously and the wet remaining mass measured to determine wet solids content (as DNA extractions are typically done on wet biomass). Synthetically diluted sludge communities were prepared from the 2018 Naburn

sample pellet using 200, 150 and 100mg of wet sludge with sample weight made up to 200mg using sand to act as a dilutant to create communities of different microbial richness. Sludge samples for the environment comparison were collected from 1) in-house reactor digesting municipal waste fibre 2) in-house reactor digesting whisky waste and 3) a hydrogen fed anaerobic digestion community. All biomass samples were spiked with *Sulfolobus solfataricus* at a volume of 25uL per 100mg (characterised to 1,012,800 genome copies per µl).

3.4.4 DNA extraction, qPCR and 16S rRNA gene amplification

Anaerobic microbial community gDNA was extracted from sludge samples using the Qiagen DNeasy PowerSoil kit using manufacturer's instructions and eluted with 100uL TE buffer. Only the duration of bead beating comparison was varied with times of 12, 15 (as per instructions) and 18 minutes for the investigation into beating duration.

10-fold dilutions of spiked sample gDNA was prepared using ddH_2O and DNA concentration measured using Qubit HS dsDNA reagent on the Qubit 4 Fluorometer (Thermo Fisher). qPCR was performed on a QuantStudio 3 (Thermo Fisher) with a reaction volume of 25ul using 10ul template (25ng total gDNA) and 12.5uL 2X SYBR green master mix using 2.5uL of 10uM SSBP primers (sequence in supplementary table 3.1). The thermocycle for amplification was; initial denaturation step of 95°C for 20 seconds followed by 40 rounds of heating to 95°C at 4c/s holding for one second then cooling to 60°C at 3c/s and holding for 20 seconds. To quantify the number of SSBP copies detected a standard curve was created from PCR amplified SSBP DNA product which was purified using AMPure beads and put through an additional PCR and purification cycle to yield pure product. Actual SSBP recovery was calculated by multiplying the number of SSBP copies / ng DNA calculated by the qPCR by the total amount of DNA present following the extraction. 25 ng of extracted DNA was used as the template for the qPCR reaction to examine copy number of SSBP and showed standard curves generated for each reaction comprising of 10^7 to 10^2 total copies with R^2 values above 0.95.

3.4.5 16S rRNA gene amplification

Successful amplification of a ~460 bp product was confirmed when using *S.solfataricus* metagenomics DNA, using the primers and protocol according to the Quick-16S NGS Library Prep Kit (D6400, Zymo Research). These commercial primers were chosen since they were advertised as having increased coverage and specificity, especially for archaea, when compared to the commonly used A519F and 802R primer pair (https://www.zymoresearch.com/pages/ngs16).

The 16S rRNA genes were amplified from spiked sludge metagenomic DNA, previously confirmed to have detectable *S.solfatricus* with qPCR. Illumina library prep and indexing was kindly performed by Dr Sally James, Genomic facility, University of York. Paired end 300 bp library sequencing was performed at the University of Leeds genomics centre on a MiSeq 2000 resulting in over 4.2 million paired end reads.

3.4.6 Bioinformatic analysis

Sequencing reads were quality filtered, trimmed, denoised and assigned taxonomy using default QIIME2 parameters (Supplementary Scripts Q1 -Q4) based on both a 97 and 99% identity match to the SILVA rRNA gene database (version 138) (Bolyen et al., 2018; Quast et al., 2013). Unfortunately no *S.solfataricus* was detected in any of the samples (Supplementary figure 3.1). Absence of *Sulfolobus solfataricus* reads was also independently confirmed by a ZymoResearch bioinformatician and by Dr Katharine Newling, Bioinformatics facility, University of York.

3.4.7 rRNA gene operon amplification

The A519F and U2428R primer pair has previously been used to amplify positions between ~520bp of the 16S rRNA gene and ~2430bp of the 23S rRNA gene to produce a product approximately 4 kb in length from environmental samples (Martijn et al., 2017). As U2428R showed only 83% identity to *Sulfolobus solfataricus,* a slightly modified U2428R_2 was used which showed 99% identity as confirmed by Silva's TESTPROBE using database version 132 (Quast et al., 2013). Between 1 and 10 ng of template *Sulfolobus solfataricus* genomic DNA was used for amplification in a 25 µL reaction with either Q5 High-Fidelity DNA Polymerase (M0491S, New England BioLabs) or Phusion (M0530S, New England BioLabs) according to manufacturer's instructions. Amplification was attempted using an annealing temperature gradient between 59°C and 71°C using default conditions, default conditions with a DMSO concentration of 2.5, 5 or 7%, default conditions with the High-GC enhancer buffer for the respective polymerase and also a combination of High-GC buffer and DMSO concentrations and described previously. Unfortunately no reproducible product amplification could be achieved.

3.5 Results

3.5.1 Species and marker selection

The aerobic hyperthermophile *Sulfolobus solfataricus* was chosen for initial screening as a candidate whole-cell spike as this organism occupies environmental niches vastly dissimilar (growing optimally

at pH 3 and 78°C) to those found in commonly studies mesophilic environments such as soil, water treatment plants and marine environments. *Sulfolobus solfataricus* cells in the stationary phase were taken and characterised using flow cytometry for further use as the potential spike. Initial characterisation used qPCR to investigate the effect of the spike in the community as it was cheaper than 16S rRNA gene amplicon sequencing, provided same day results and allowed us the same level of quantification. A total of seven marker gene candidates were identified by searching literature regarding *Sulfolobus solfataricus* manually for keywords such as "novel" and "unique", as a protein unique to *Sulfolobus solfataricus* was assumed to be less likely in having shared sequence identify with other organisms. Additionally, genes were only selected if their gene occurred only once in the genome. Nucleotide sequences for these proteins were then BLASTed against the bacterial database using default parameters of blastn to optimise matches to dissimilar sequences. As all the gene sequences showed minimal occurrence in the bacterial nucleotide database (Table 3.1), sets of primers were made for each gene target to then be tested *in-vivo* (Supplementary Table 3.1).

Literature name	NCBI accession	Gene length	Closest Bacterial match	Max Score	E Value	Identity	Query coverage (bp)
Single-stranded DNA binding protein (SSBP)	AAK42515	1359	*Arcobacter nitrofigilis*	49.1	0.25	97%	28
Cellulase (CelS)	AJ296029	969	*Salegentibacter* sp. T436	46.4	0.61	82%	48
Uncharacterized ATP-dependent DNA helicase (Hel)	AAK40467.1	2628	*Prochlorococcus* sp. RS04	59.0	3e-04	74%	59
Adenylate Kinase (Kin)	AB006440	837	*Escherichia coli* strain C1	48.2	0.15	94%	30
5'-deoxy-5'-methylthioadenosine phosphorylase (SsMTAPII)	AAK42494	813	*Candidatus Pelagibacter* sp. RS40	97.8	3e-16	70%	265
Exopolyphosphatase (Set3)	AJ314593	1254	*Thermosipho melanesiensis* strain 431	50.9	0.063	88	40
α-Glucosidase (malA)	AF042494	7056	*Bacillus cereus* VPC1401 plasmid pLVP1401	48.2	0.4	94	31

Table 3.1 S.solfataricus marker gene selection

Literature data for the unique gene marker candidates are displayed here including the described name in the literature, NCBI accession value and total nucleotide length. A summary of BLAST results for the closest available bacterial match is also displayed including score metrics and % identity. The majority of candidate genes had minimal regions of nucleotide similarity varying from 30 to 60 nucleotides. Since alignment results were only obtained when using the BLASTN "somewhat similar" parameter, the first six candidates were further investigated.

3.5.2 Target verification

Initial verification of marker selectively was investigated by performing triplicate PCR on gDNA extracted from sludge taken from an anaerobic digester spiked with *Sulfolobus solfataricus*. Primers sets used included the primers designed for the target gene (Supplementary Table 3.1) and the 16S V4 rRNA gene primer pair as a positive control (in duplicate) previously used by Klindworth et al (Klindworth et al., 2013). Primer pairs which showed exclusive target amplification using the *Sulfolobus solfataricus* gDNA were SSBP and CelS (Figure 3.1a). Strong off target amplification was observed in the primer pairs for the Hel, Kin and SsMTAPII as multiple bands were visualised.

Although both the SSBP and CelS primer pairs showed potential as a target marker, the smaller product size of SSBP gene when compared with CelS (177 vs 530bp) was a more favourable sized marker for qPCR. Using qPCR, the SSBP primer pair was used to further investigate aspects of the microbial spike including; the effect of increasing the spike:biomass ratio, normalisation of initial sample biomass, effect of increasing bead beating duration, sample microbial richness, differences between sample types and extraction method (Stämmler et al., 2016).

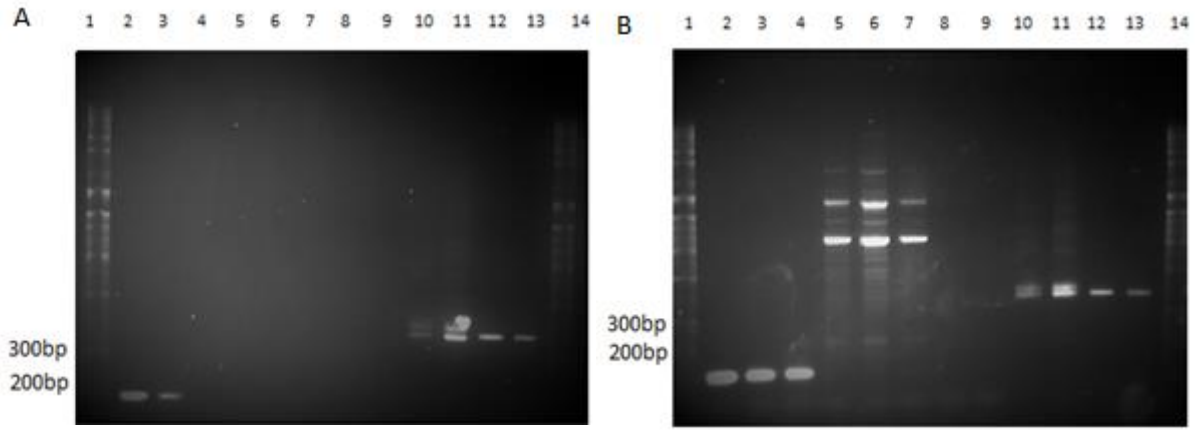


Figure 3.1 Selectivity of marker gene candidates

Selectivity of marker gene primer pairs (See Supplementary Table 1) targeting the *Sulfolobus solfataricus* single-stranded binding protein (NCBI-AAK42515, Fig A, 177bp) and Adenylate Kinase (NCBI-AB006440, Fig B, 117bp) were confirmed by using polymerase chain reaction (PCR) on both *Sulfolobus solfataricus* spiked sludge genomic DNA (Lanes 2-4) and sludge gDNA only (Lanes 5-7). A negative water control was included (Lanes 8-9) in addition to positive controls targeting the V416S rRNA gene region (Lanes 10-11 for spiked samples and lanes 12-13 for unspiked). The SSBP primer pair shows no off-target amplification for unspiked sludge DNA (Fig A lanes 5-7) whereas multiple bands were observed for the Adenylate Kinase primer pair (B lanes 5-7) indicative of off-target amplification and not desirable for use in downstream qPCR applications. Amplified samples were run on a 1% agarose gel.

3.5.3 Characterisation of *Sulfolobus solfataricus* spike using flow cytometry

Fundamental to the correct use of an internal spike is proper characterisation. To date no studies have been done which enumerated genome copy numbers added by the spike. Cell number and ploidy were identified based upon a Förster resonance energy transfer signal generated between the interaction of Mithramycin A and Ethidium Bromide, as described previously (Bernander and Poplawski, 1997). Flow cytometry was used as it allowed us to both count the number of cells used for the spike and to calculate the number of genome copies each cell possessed to generate an experimental value of actual marker gene added. Cells were identified from background noise based upon the threshold of forward and side scatter (Figure 3.2a). Both Mithramycin emission (525 nm) intensity and FRET (610 nm) signal intensity were used to further characterise the genetic content of the *Sulfolobus solfataricus* cells and to infer likely genome copy number based upon abundance of respective FRET intensities (Figure 3.2b).

From the cell count and relative fluorescence, the spike was determined to contain 609,200 cells/ µl with the calculated number of genome copies at 1,012,800 equating to an average of 1.66 genome copies per cell. With the microbial spike fully characterised, the exact number of genome copies added to samples at the stage of DNA extraction could be calculated.

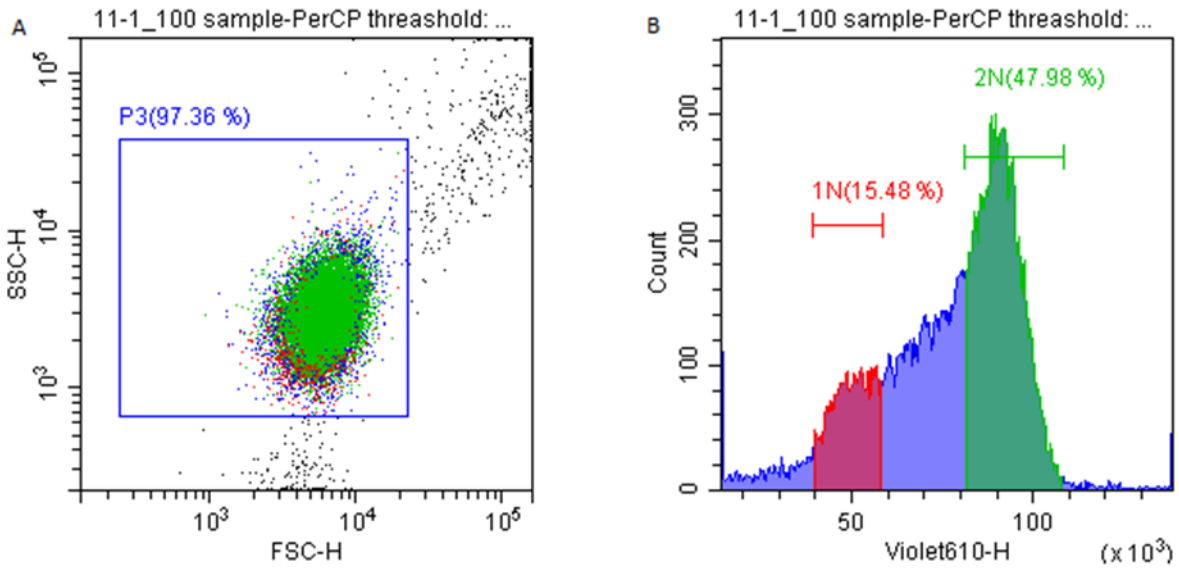


Figure 3.2 Determining Sulfolobus solfataricus *ploidy*

Sulfolobus solfataricus cells were characterised using flow cytometry coupled with a DNA dye to examine cell ploidy. Forward and side scatter at 525nm were used to count the number of cells (Figure A) and determine that 97.36% of signal was independent of background noise when compared to the control for a total of 20,000 detection events. DNA content of each event was quantified by measuring relative fluorescence at 610 ±20nm following excitation at 405 nm which takes advantage of the FRET interaction between the DNA specific Mithramycin A and more abundant Ethidium Bromide dye used. Determination of ploidy was performed manually by identifying regions of high counts with 2N cells roughly double the emission at 610nm when compared to 1N cells (100 x 10^3 vs 50 x 10^3 relative fluorescence). A total of 9342 2N events and 3014 1N events were recorded, with the rest assumed to be 1.5N, representing 47.98 % and 15.48 % of cell specific events respectively. From the cell count and relative fluorescence, the spike was determined to contain 609,200 cells/ µl with the calculated number of genome copies at 1,012,800 equating to an average of 1.66 genome copies per cell. With the microbial spike fully characterised, the exact number of genome copies added to samples at the stage of DNA extraction could be calculated.

3.5.4 Titration of *Sulfolobus solfataricus* to establish sensitivity and linearity of spike

Spike-in load was increased to investigate the relationship between *Sulfolobus solfataricus* cells added and SSBP marker gene quantification using qPCR and to identify the level at which loss of spike accuracy occurred. Whole *Sulfolobus solfataricus* cell solution volumes of 12.5, 25, 50, 75 and 100 µl were added to the bead lysis tubes with exactly 200mg sludge prior to DNA extraction.

Spike response increased linearly between dosages of 25 to 75 µl resulting in SSBP copy numbers of 1052.90 ± 316.21, 2216.89 ± 1435.02, 3913.35 ± 1602.48, 5309.74 ± 1509.99 (Figure 3.3). At a spike volume above 100 µl (4308.17 ± 1637.34 SSBP copy number), the increase spike load seemed to have a diminishing and less linear effect. For that reason, a spike volume of 50 µl ($5.1x10^7$ cells) per 200 mg sludge load was chosen for further use in this study as that lay in the linear part of the dose response curve.

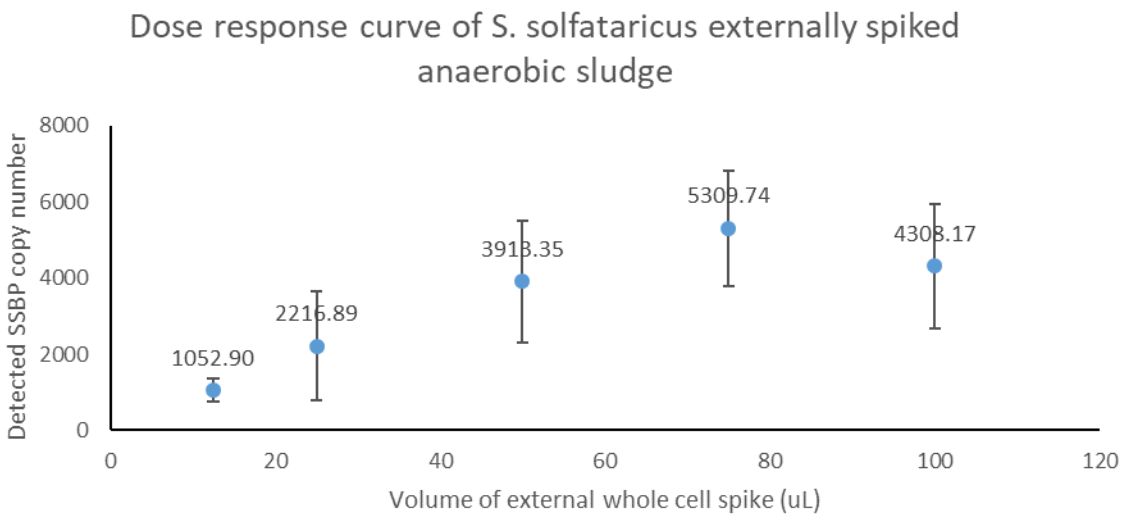


Figure 3.3 Dose response curve of spiked sludge

Volumes between 12.5 and 100 µl of *Sulfolobus solfataricus* cell culture were added to 200 mg ± 10 mg of sludge prior to genomic DNA extraction using Qiagen DNeasy PowerSoil bead beating kits (n=3, error bars = standard deviation). Using qPCR and the SSBP primer pair, which selects exclusively for a gene unique to *Sulfolobus solfataricus*, the linearity of spike response was examined as a measure of SSBP copy number detected. A linear dose:response curve was observed when increasing the spike load between 12.5 and 75 µl with a reduction in effectiveness observed when increased to a spike volume of 100 µl (error bars are standard deviation).

3.5.5 Bead beating Length

Bead beating duration in commercial kits typically varies between 15-20 minutes, depending on throughput with recommended instructions suggesting an increase in duration when more samples are being beaten. Additionally user error may cause variance of a few minutes depending on operator's dedication to protocol. Because of this, we extracted DNA from wastewater treatment plant sludge sample spiked with *Sulfolobus solfataricus* using the bead beating for a total of 12, 15 and 18 minutes beating duration. Lysing efficiency was measured by quantifying the SSBP copy number using qPCR. To ensure constant beating conditions samples were replaced with equally weighted empty tubes when removed. Overall, no trend was observed when increasing the bead beating duration from 12 to 15 and further to 18 minutes (Supplementary Figure 3.3) resulting in 2729 ±195, 2072 ±306 and 3030 ±379, respectively (n = 4).

3.5.6 Biomass normilisation

Biomass samples between 200 and 300 mg were spiked at a ratio of 50 µl per 200 mg in triplicate to investigate if spike application can act to normalise samples with different biomass and negate the need for timely accurate biomass weighing (Figure 3.4). Average SSBP copies detected for biomass sample weights of 200, 250 and 300 were 4099 ±493, 4552 ±772 and 4660 ±882 respectively. No significant difference was found to exist between each set (P value of 0.73).

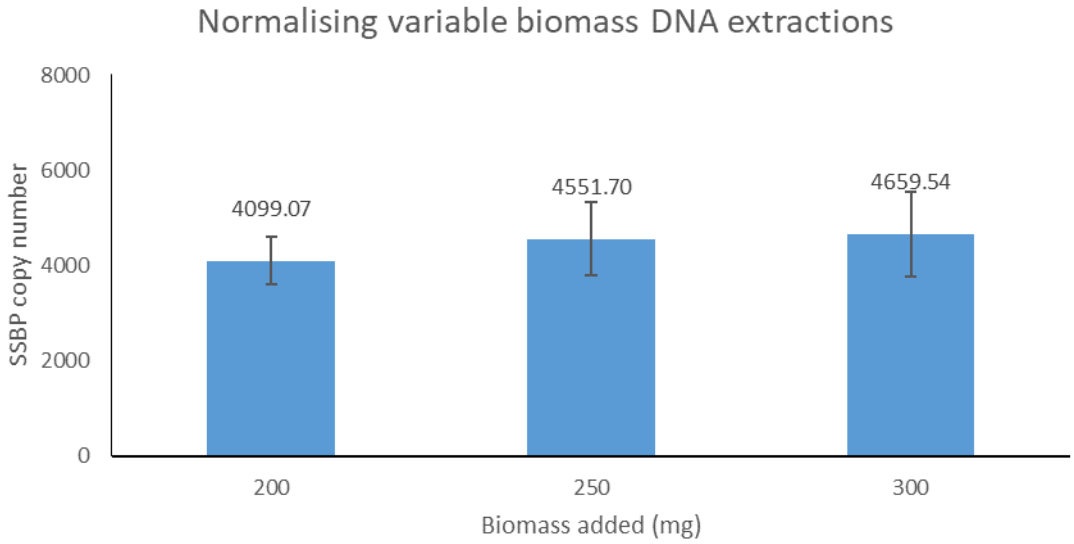


Figure 3.4 Normalising variations in sample weigh with the Sulfolobus solfataricus *spike*

Varying masses of sludge were spiked at a ratio of 25 µl per 200mg prior to genomic DNA extraction using Qiagen DNeasy PowerSoil bead beating kits (n=3, error bars = standard deviation). Using qPCR and the SSBP primer pair, which selects exclusively for a gene unique to *Sulfolobus solfataricus*, the linearity of spike response was examined as a measure of SSBP copy number detected. A near identical spike response was observed when increasing the sludge mass between 200 and 300 mg.

3.5.7 Spike response as a measure of microbial richness

Accumulation of recalcitrant material such as sand, grit and mineral precipitates is problematic for the water treatment industry as it reduces the operational volume of the respective reactors. To investigate if a linear correlation between spike response and dilution factor could be achieved, triplicate Naburn sludge samples spiked with 25 µl were diluted with grit either 2-fold or 4-fold prior to performing DNA extracting on 200 mg of the resulting mixture and qPCR analysis which revealed an extremely linear spike relationship (Figure 3.5 , $R^2 = 0.99$ and $P < 0.0001$).

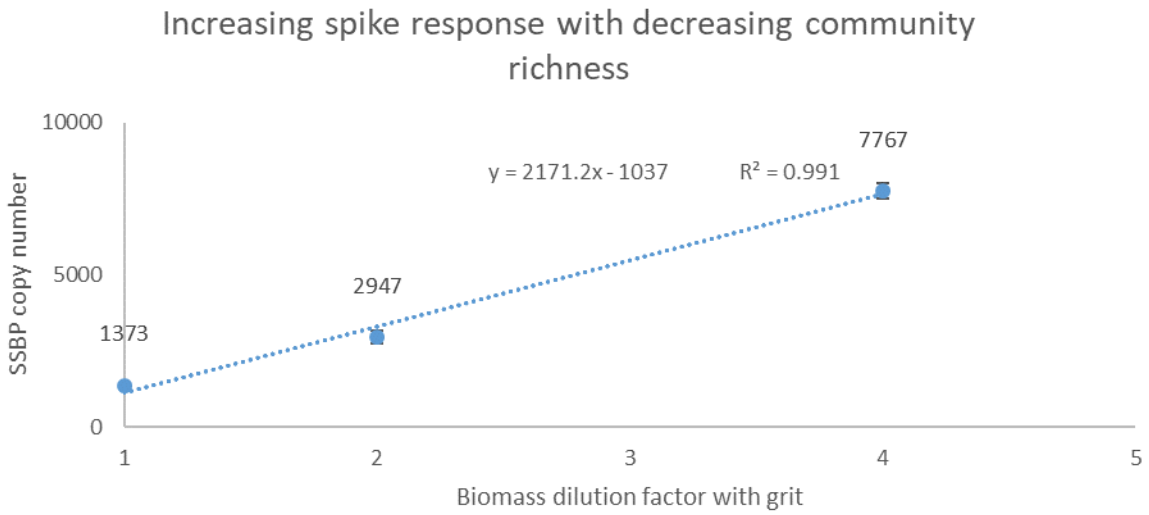


Figure 3.5 Increasing spike respoonse with decreasing microbial richness

Grit is a common recalcitrant dilutant in waste water treatment plants that results in a lower microbial biomass per gram of solids and thus decreased microbial richness or methanogenic activity per gram of solids. In this experiment sludge was taken and diluted with grit to a total weight of 200 mg. Triplicate samples with dilution factors of 1 (no dilution), 2 and 4 were comprised of only 200 mg sludge, 100 mg grit with 100 mg sludge and 150 mg grit and 50 mg sludge to investigate if a linear spike response could be produced. A strong statistically significant positive correlation (R^2 value of 0.92, P value < 0.0001, N = 4, error bars = standard deviation) was observed when comparing increasing dilution factor against an increasing spike response. This increased spike response is due to DNA from *Sulfolobus solfataricus* making up a larger fraction of total gDNA.

The spike was then applied to an additional three reactor sample types in triplicate; a beverage waste fed lab reactor (PA), A hydrogen supplemented synthetic feed lab reactor (H2), Naburn sludge (NB) and a municipal fibre waste fed reactor (MFW), resulting in a spike response of 888 ±84, 1157 ±164, 1373 ±114 and 2140 ±262 (Figure 3.6). This demonstrated the highest microbial sample richness was the lab fed reactor and the least was the municipal fibre reactor, unsurprising considering the presence of large amounts of recalcitrant material present in the form of fibre in the latter

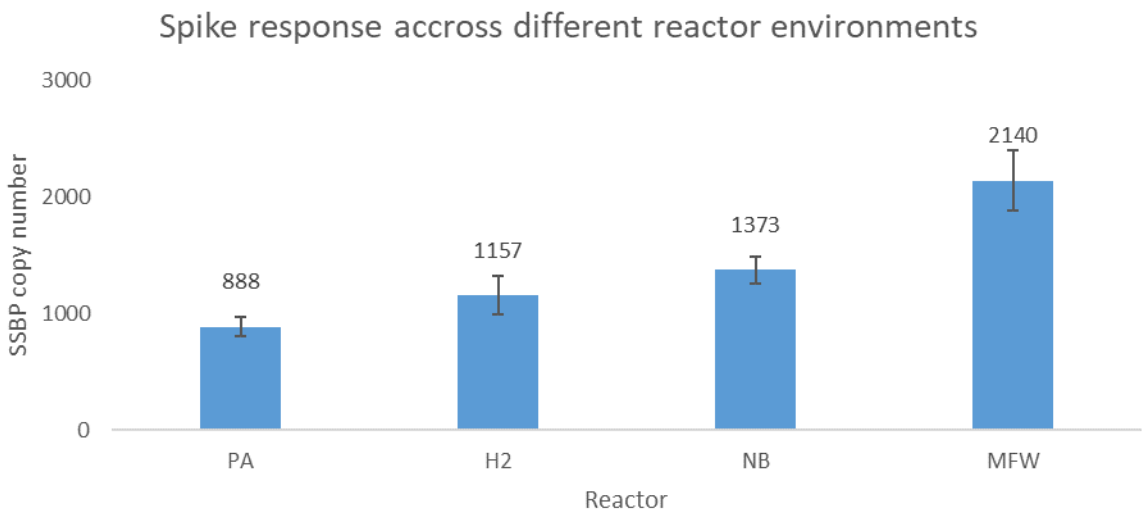


Figure 3.6 Spike response across different reactor samples

Samples from three different anaerobic digestion conditions were obtained in quadruplicate from reactors digesting either; pot ale, a beverage waste product, (PA) or a reactor fed with synthetic feed supplemented with hydrogen (H2), waste water from Naburn water treatment plant (NB) or municipal waste fibre (MWF). Samples were externally spiked prior to DNA extraction with *Sulfolobus solfataricus* whole cells to investigate if different sludge samples presented different extraction conditions and thus varying spike recovery conditions. It was observed that the reactor fed MFW has the highest spike response of 2140 ±262 SSBP copies detected and thus was the most microbially diluted (due to the presence of higher amounts of recalcitrant contamination present in the MWF) in contrast to the PA reactor which had the lowest spike response of 888 ±84 indicating highest biological content of the 200 mg sludge sample, possibly due to a lab controlled 5 month feeding schedule of rich beverage waste material, error bars = standard deviation.

Overall microbial richness of reactors were calculated by taking the inverse of spike response and multiplying by sample wet weight (mg/L). MFW and H2 reactors have similarly concentrated microbial

communities with index scores of 36.7 ±4.6 and 35.1 ±5.2, PA at 106.5 ±9.4 and NB the highest with 262.5 ±21.8 (Figure 3.7).

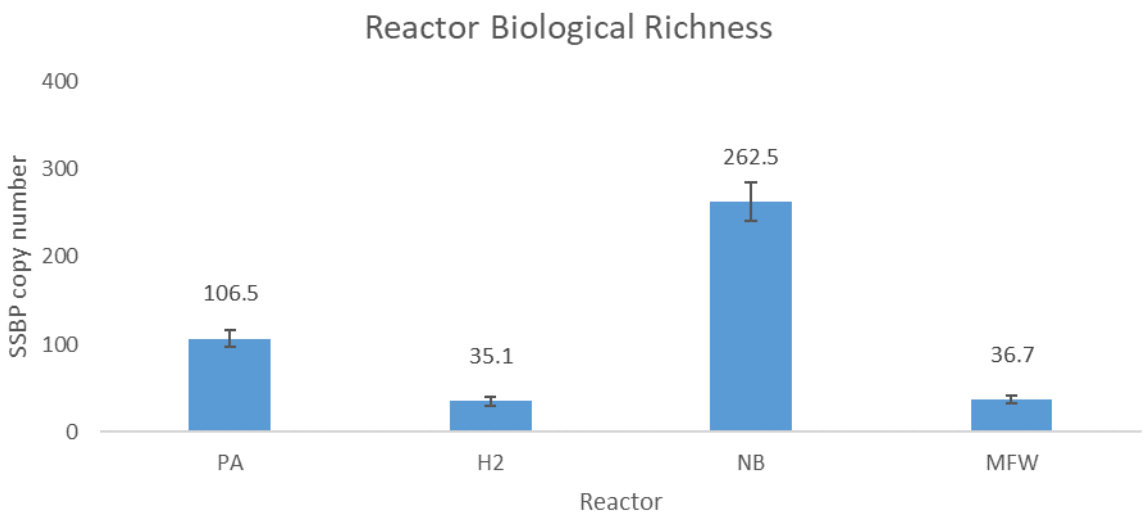


Figure 3.7 Reactor biological richness

Samples from three different anaerobic digestion conditions were obtained in quadruplicate from reactors digesting either; pot ale, a beverage waste product, (PA) or a reactor fed with synthetic feed supplemented with hydrogen (H2), waste water from Naburn water treatment plant (NB) or municipal waste fibre (MWF). Samples were externally spiked prior to DNA extraction with *Sulfolobus solfataricus* whole cells to investigate if different sludge samples presented different extraction conditions and thus varying spike recovery conditions. Spike response was then adjusted according to total sample wet weight and content to calculate degree of biological richness.

Post centrifugation wet weight was used to calculate the overall richness of sample environment as the sample spike response factor only indicates how rich the extracted sludge sample is. Wet weight was used rather than dry weight since the DNA extraction step is performed on wet sludge thus streamlining the process. A composite plot which allows for rapid characterisation of reactor microbial richness and degree of recalcitrant contamination is provided in Figure 3.8. By determining overall microbial richness of reactors it can aid identification of communities which show desirable industrial characteristics such as increased methanogenic activity per unit of microbial richness index score.

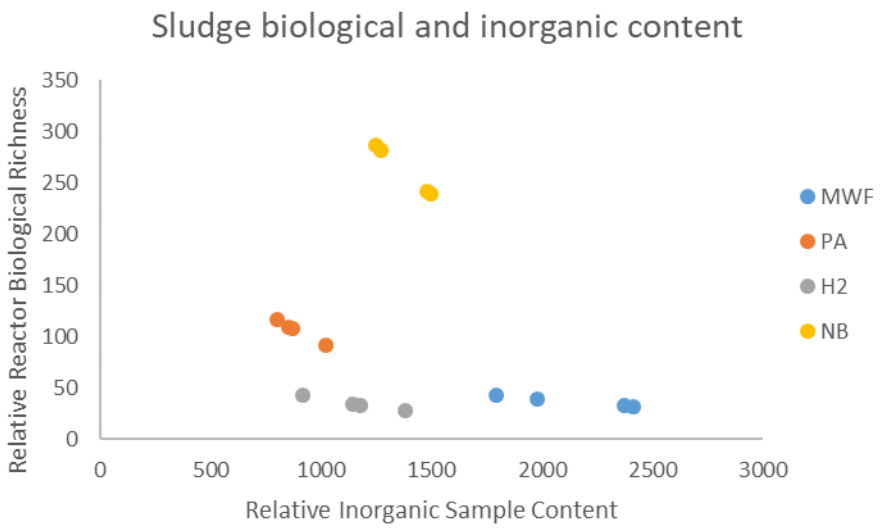


Figure 3.8 Sludge biological and inorganic content

Samples from three different anaerobic digestion conditions were obtained in quadruplicate from reactors digesting either; pot ale, a beverage waste product, (PA) or a reactor fed with synthetic feed supplemented with hydrogen (H2), waste water from Naburn water treatment plant (NB) or municipal waste fibre (MWF). Samples were externally spiked prior to DNA extraction with *Sulfolobus solfataricus* whole cells to investigate if different sludge samples presented different extraction conditions and thus varying spike recovery conditions. Spike response was then adjusted according to total sample wet weight and content to calculate degree of biological richness. Inorganic sample content index was calculated based upon the amount of DNA recovered from the wet sample in terms of ng DNA /ml sludge.

3.6 Discussion

3.6.1 Development of microbial spike

Developing a true quantitative measure of microbes in a community would allow greater correlation to environmental or industrial processes. The truly quantitative method presented in this study improved upon similar ideas, such as direct DNA spiking or optical density based spiking, by directly quantifying the number of genome copies present in the spike organism. Using *S.solfataricus* for this task was simpler than other candidate foreign organisms since it is a monoploid organism presenting a maximum of two genome copies when in exponential phase compared to Euryarchaeota species such as *Halobacterium salinarum* (25 copies), *Haloferax volcanii* (17), *Methanococcus maripaludis* (a staggering 55 copies) (Hildenbrand et al., 2011). Clearer differentiation between the spike organism's respective ploidy populations ensures a degree of reliability and ease in measurement.

Flow cytometry allows multiple properties of an individual cell to be measured based upon light scattering and light absorbance/emission. Using flow cytometry to count cells as opposed to other methods such as optical density readings or colony counting has shown to be more reliable (Biesta-Peters et al., 2010; Brown and Wittwer, 2000; Pan et al., 2014). This, coupled with the ability to accurately determine cell ploidy from using a non-specific DNA dye suggests that the spike evaluation method presented here is optimal.

3.6.2 Assessment of microbial spike

The microbial spike showed good usability in terms of qPCR compatibility (if only microbial richness metrics are required), normalising sample weight, comparing extraction method effectiveness and determining degree of recalcitrant contamination. By showing to provide normalisation across a range of sample weights it fulfils one of the key requirements in normalising extraction methodology. Additionally, it improves upon previous methods since the spike is inside an identical environment (a cell) to what is being quantified rather than direct DNA addition to sample (Smets et al., 2016). Moreover, it assigns an exact value to the strength (copy number) rather than relying on less specific methods such as growth time or optical density (Piwosz et al., 2018).

Degree of recalcitrant material contamination was simply determined as in inverse measure of spike response where a larger response indicates the spike comprises a larger fraction of the population. Determining degree of recalcitrant material, such as the higher amount found in the MFW samples, allows indirect detection of common dilutants such as sand, small plastic pieces and litter which are increasingly entering the sewage and drainage systems (Armitage and Rooseboom, 2000).

A measure of biological richness could also be inferred from the spike when considering spike response in relation to both the mostly solid sample subjected to DNA extraction and the weight of solids present in the initial homogenous samples. Although determination of solids can also be performed with dried samples, using wet solids content provided a “quick and dirty” measurement which assumed an equal level of water saturation in a given biomass pellet. Although not the most accurate solid determination, the principle of co-analysing spike response and wet solids would be identical when using dry solids.

3.6.3 16S rRNA gene amplification

Although successful amplification of the V3-V4 16S rRNA gene region with *Sulfolobus solfataricus* genomic DNA occurred, primer bias when competing against other organisms may have led to no *Sulfolobus solfataricus* being amplified from the sludge community sample (Supplementary Figure 3.4). Further discussion with ZymoResearch representatives revealed that the forward primer contained a one base pair mismatch (not shown due to confidentiality). Typically a single primer mismatch is not considered critical for gene product amplification, although a reduction in yield will occur (Kwok et al., 1990). Taq polymerase shows 1,000 fold discrimination, and subsequent underestimation, against primers paired with a single mismatch (Huang et al., 1992). Some gene amplification completely fails even with one mismatch at the 3' end of the primer (Simsek and Adnan, 2000). This is even more extreme in the case of 16S rRNA gene mismatches if occurring in the second half of the primer sequence which results in a 1,000 fold underestimation in a community equally comprised of only three species (Bru et al., 2008). It is likely that this bias against *S.solfataricus* was amplified, especially since more than 15 families of organism were detected at a relative abundance between 2 and 15%. The concept of primer bias effecting 16S rRNAgene sequencing community analysis pipelines was discussed in the previous chapter and this is an unfortunate example of it in action (Sipos et al., 2007).

3.6.4 Sequencing the entire rRNA gene operon

The recently developed Oxford Nanopore MinION is a small portable device costing approximately £1000. By barcoding samples, usage of this device is cheaper than an average 300 bp sequencing lane at Leeds genomics facility (£2,472 at the time of writing), so was a more affordable way to repeat the sequencing compared to an Illumina run. However, the lower accuracy (92% at time of study) of the Nanopore sequencing device makes sequencing short fragments of DNA, such as the V3-V4 region of the 16S rRNA gene, a difficult task (Kerkhof et al., 2017). By amplifying larger fragments, such as the whole 16S rRNA gene region or the entire rRNA operon, correct taxanomic assignment can be achieved for ~68% or ~98% of cases respectively (Cuscó et al., 2018). The U2428R_2 primer was developed from the study which showed ~98% taxanomic accuracy because sequencing the whole 4 kb rRNA operon instead of the 1.5 kb 16S rRNA gene fragment is more preferable when using more error-prone sequencing methods such as Oxford Nanopore (Cuscó et al., 2018; Ettema et al., 2006; Kerkhof et al., 2017). Unfortunately, as no reproducible amplification occurred under the examined conditions the sludge community, samples were not sequenced. If they were successfully amplified and sequenced, taxonomic assignment would be performed using a database recently curated for

whole rRNA operon taxonomy assignment supplemented with the respective *Sulfolobus solfataricus* gene (Martijn et al., 2017). Given additional time, more primer combinations and conditions would have been tried in the attempt to amplify the entire rRNA operon.

3.7 Conclusion

Use of internal spikes in work-flows offers increased quality control and allow normalisation between studies. This is especially true in the field of amplicon based microbial ecology which suffers from no universally agreed analysis pipeline, partly due to no perfect primer pair, and so efforts must be made to improve inter-study reproducibility (Krakat et al., 2016). While other studies provide a spike enabling intra-study reliability, such as a direct DNA spike or cells grown to a certain OD, this study presents the first microbial spike methodology to facilitate exact spike reproduction elsewhere (Piwosz et al., 2018; Smets et al., 2016; Stämmler et al., 2016; Tourlousse et al., 2017; Zhang et al., 2017). This external reproducibility is due to chromosome content of spike cells being quantified and characterised through flow cytometry and fluorescence-activated cell sorting. The use of *S.solfataricus* as an internal microbial spike showed highly desirable properties including easily identified ploidy and foreign nature to most environments. Additionally, characteristics such as microbial richness and degree of recalcitrant (non-biological solids) could be determined from samples using the qPCR based methodology presented in this study. Methodology induced variables such as sample weight and bead beating duration were successfully normalised and showed a high degree of reproducibility.

Unfortunately the 16S rRNA gene amplification strategy failed to amplify the *Sulfolobus solfataricus* sequence in the presence of other community organisms due to primer bias and thus could not comparatively quantify the other species present in the community. Further attempts to amplify the entire rRNA operon for long read sequencing also failed. However, recent attempts at normalising methodology and quantifying microbial populations using metagenomics approaches have been successful in synthetic and murine microbial communities (Sheth et al., 2019; Venkataraman et al., 2018).

Moving forward, the rapidly increasing accuracy of nanopore devices (from 66% to 92%) means read inaccuracy is becoming less of an issue, important since an identity of 97% is considered the cut-off threshold for clustering OTUs as identical species (Caporaso et al., 2011). As the price of metagenomics sequencing continues to fall, it is inevitable that the rRNA gene (whole or in part) amplification strategy used in this experiment becomes redundant, especially since the announced release of the

SmidgION in the coming years (Oxford Nanopore, https://nanoporetech.com/products/smidgion). However, regardless of sequencing and analysis strategy, the microbial spike characterisation methodology presented in this study is the only truly quantitative microbial spike-in method to date.

3.8 Acknowledgements

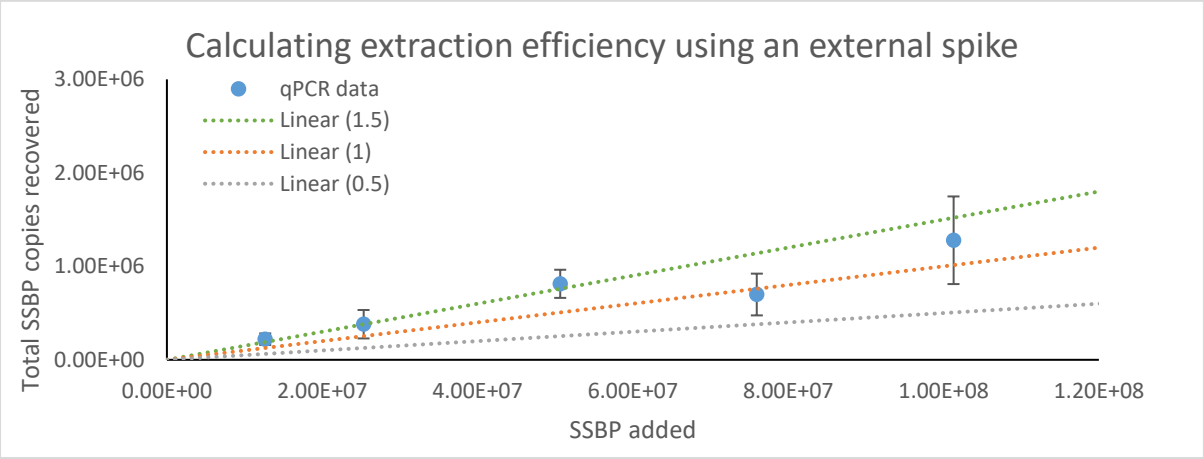
I would like to thank; Monica Patel, a hard working undergraduate summer student, for performing the screening of *S.solfataricus* marker gene targets in the summer of 2017. Luna Yuan, another fantastic undergraduate summer student, for screening 16S and whole rRNA operon primers. Aritha (recently, congratulations) Zeigler for providing samples from municipal waste fibre BMP tests and also Bing Tao for providing me with samples from a hydrogen fed AD community.

3.9 Supporting information

Gene	Sequence (5>3)	Start	Stop	Tm (°C)	GC%	Product Length
SSBP F	GGGAAAACATGCAGGTAGTATAAAAGAAGG	168	197	62.5	40	177
SSBP R	GGAGCTGTTGGTGTATTTTCTGGTATTTG	29	344	62.9	41.4	
CelS F	CTTGTGGAATATAGGATACGCTCTAGGAAA	237	266	62.2	40	530
CelS R	CTTTTAAGGGTGATAGAAAGTAAACTCCCG	766	737	62	40	
Hel F	GATGGTCTTATATATTCCCTTGTCTACCTC	1828	1857	61	40	447
Hel R	GTTAGAGTTGGAATAAGAACTAGCGATACC	2274	2245	61.5	40	
Kin F	TTCCAAGACATGTAATAGAAGTCCTTTCCC	458	487	62.9	40	117
Kin R	TATAATCAGTCCTAGCTCTACTACTATCCC	574	545	60.4	40	
SsMTAPII F	TTTCTACAAGGGCTGAGAGTAGAAC	509	533	60	44	192
SsMTAPII R	CTGCCATAACTCTAGTAACCTCCTC	700	676	60	48	
A519F	CAGCMGCCGCGGTAA	N/A	N/A	66-70		~4000
U2428R	CCRAMCTGTCTCACGACG	N/A	N/A	63-68		
U2428R_2	CCRAMCTGTCTCRCGACG	N/A	N/A	63-71		

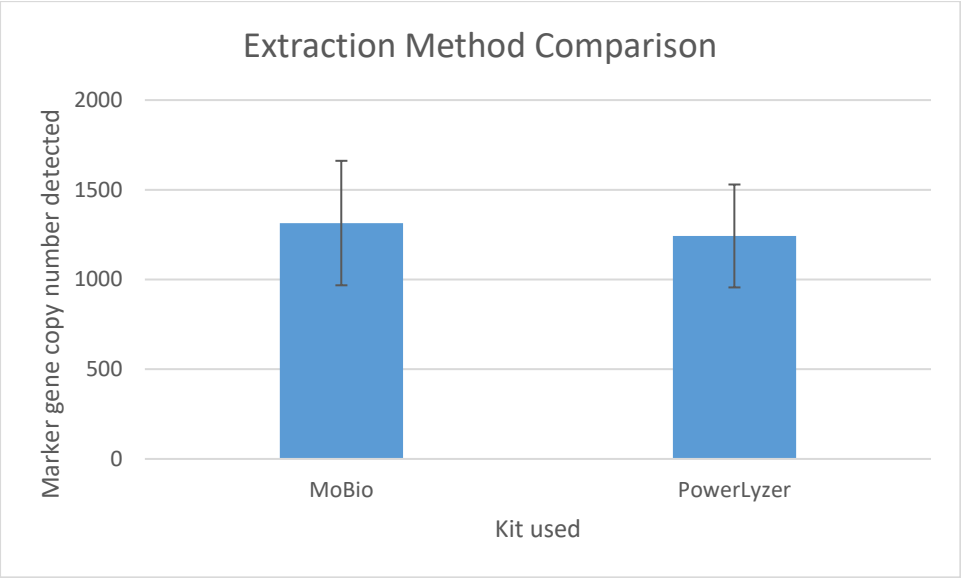
Supplementary Table 2.1. Primer pairs

Primer names, sequences and properties used over the course of this investigation are shown.



Supplementary Figure 2.1. Calculating extraction efficiencies

Extraction efficiency was calculated by measuring the number of marker gene copies present in the qPCR then multiplying by the total yield of DNA obtained following extraction. This shows that approximately 1% of the marker genes from the spike carry through into the amplification step.



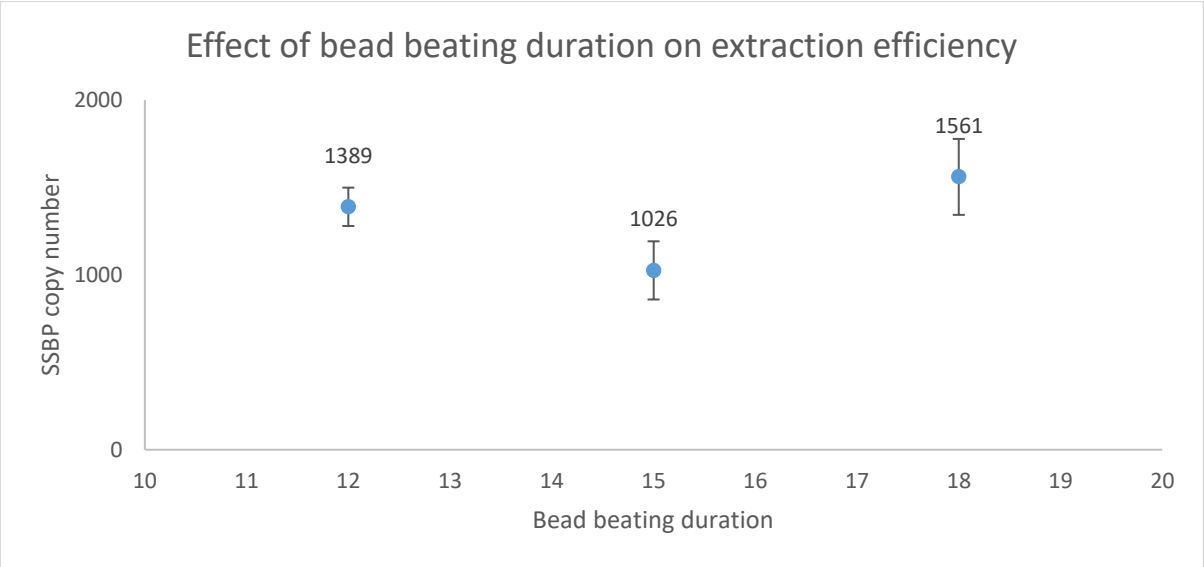
Supplementary Figure 2.2. Extraction method comparison

A comparison between commercially available bead-beating soil extraction kits was performed which revealed no significant benefit of one over the other for the lysis of the spiked *Sulfolobus* cells withing 200 mg sludge n=4.

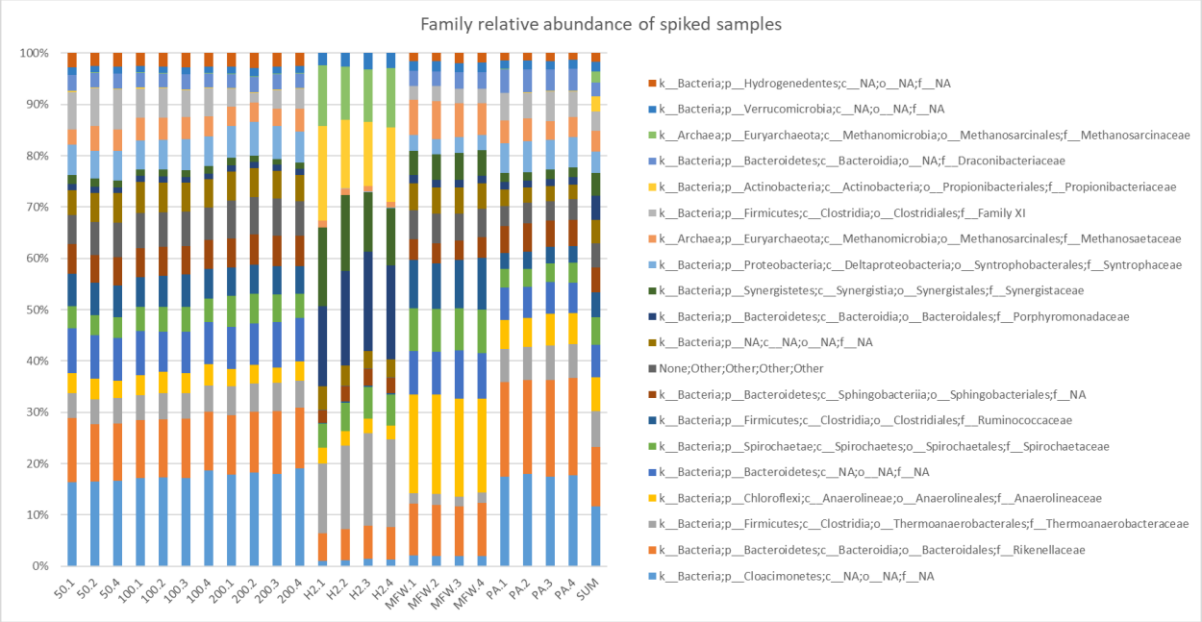
Sample	Reads
100-1	112,805
100-2	143,462
100-3	183,137
100-4	132,772
200-1	262,975
200-2	168,687
200-3	131,665
200-4	161,697
50-1	166,208
50-2	189,458
50-4	206,716
H2-1	149,008
H2-2	189,775
H2-3	243,176
H2-4	240,974
MFW-1	250,036
MFW-2	235,418
MFW-3	219,608
MFW-4	197,697
PA-1	157,414
PA-2	162,224
PA-3	129,587
PA-4	188,957
Total	4,223,456

Supplementary Table 2.2. Sequencing reads per samples

Supplementary Figure 3.3. Effect of bead beating duration on extraction efficiency



Supplementary Figure 3.2. Duration of bead beating was varied to ±3 min of the specified 15 min time according to manufacturer's instruction to investigate effects of varying beating duration on Sulfolobus solfataricus external spike recovery when measuring SSBP copy number present in gDNA from anaerobic sludge samples. No trends were observed when increasing the beating duration from 12 to 15 and then 18 minutes as SSBP copy number detected was 1389 ±110, 1026 ±166 and 1561 ±217 respectively, n = 4.



Supplementary Figure 3.4. Taxanomic assignment of OTUs at family level shows a diverse microbial community although no Sulfolobus solfataricus spiked in was detected.

Chapter 4

Metagenomics and metatranscriptomics of a pot ale fed anaerobic digestion microbiome reveal key hydrolysis and rate limiting enzymatic steps

4.1 Abstract

An anaerobic microbial community was gradually adapted to increasing concentrations of the whisky waste material pot ale. Biogas methane richness increased from 36 to 55 % and conversion efficiency from 12 to 47 % over the course of the acclimatisation period. More than 95% of chemical oxygen demand was removed after a week's digestion highlighting anaerobic digestion as an excellent treatment option for this waste material. Metagenomic community analysis of the anaerobic digestion community showed similar structure over time apart from an increasing *Bacteroidia* population of 5 to 30 % relative abundance when fed the highest pot ale concentration. Methane production repeatedly occurred in two distinct phases, one short sharp increase within the first four hours followed by a longer methanogenic phase between 12 and 108 hours. Timecourse metatranscriptomic analysis of hydrolysis enzymes revealed the first one is caused by initial phosphorylation of residual sugars in pot ale, malto-oligosaccharides, and formation of mostly beta-D-glucose-6-phosphates followed by isomerisation and further phosphorylation into D-fructose 1,6-bisphosphate and subsequent cleavage into glycerone phosphate and D-glyceraldehyde-3-phosphate. The three carbon products then are likely converted into pyruvate and then propanoic acid (as confirmed by GC-FID analysis) through an unknown means. Methane production in the second longer phase could be attributed to the increased expression of (1->3)-(1->4)-beta-D-glucan 4-glucanohydrolases involved in liberating glucose molecules from beta-glucans present in barley and further fermentation. Expression of acetoclastic methanogenesis related transcripts increased more than 100 fold over the duration of the experiment suggesting it as the dominant pathway. Although hydrolysis and methanogenesis steps could be identified, pathways associated with fatty acid production and degradation could not be identified.

4.2 Keywords

Metagenomics, Metatranscriptomics, rate limiting, enzymes, specialisation

4.3 Introduction

Complex environmental microbial communities are the underpinning driving force of many geo-chemical reactions and nutrient cycles across the planet in environments such as marine ecosystems, soils, salt marshes, caves (Beazley et al., 2012; Northup et al., 2011; Zhao et al., 2004; Žifčáková et al., 2016). Traditional culture based methods fail to accurately investigate biochemical potential, diversity and function of these communities as the majority of species cannot be grown in common laboratory conditions or require syntrophic partners (Elshahed and McInerney, 2001; Stolyar et al., 2007; Yamada et al., 2015). It is accepted that the majority of microbial species are unculturable by conventional techniques and that approximately 80% of human gut microbiome species are unculturable (Lagier et al., 2012). This can be partly attributed to some species requiring removal of by-products by a syntrophic partner but can also be associated with attempts to grow microbes in unnatural, by their standard, environments. Presence of conductive mineral pieces, redox compounds and carbon tubes more accurately reflects their highly interconnected natural environment and has allowed culture and examination of many previously unculturable microbes (Beckmann et al., 2015; Cruz Viggi et al., 2014; Sieber et al., 2012; Suanon et al., 2015; Yamada et al., 2015).

The advent of next generation sequence technologies has facilitated many culture-independent analytical frameworks for studying organisms and complex microbiomes. In the past, the most commonly used technique was amplification of a small hypervariable region of the prokaryotic 16S rRNA gene, similar in principle to a cars registration plate. The widespread use of this technique in microbial ecology is testament to its usefulness at providing meaningful data from a highly complex system such as the gut microbiome and identification of new species. However, it is entirely dependent on successful PCR amplification of a section of the 16S rRNA gene using a set of carefully selected primers. Poorly chosen primers can lead to under and over representation of select organisms and even whole taxa (Klindworth et al., 2013). To cover a wider range of organisms degenerate primers are typically used which may have been agreed upon by an international committee to ensure reproducibility and inter-study comparison, such as with the Soil Microbiome Project (Gohl et al., 2016). Primer binding bias can even be so extreme that it affects primers which show 100% specificity towards an organism (such as the unfortunate case in Chapter 2). Systematic review of steps involved in 16S rRNA gene amplification indicated that strong bias exists at all steps in the process and affects the accuracy of interpreted results (Krakat et al., 2016). Moreover, with this method, functional interpretation of microbiomes is entirely dependent on database correlation to laboratory studied

organisms and as ~80% of organisms are not laboratory studied key functional and biochemical information is lost.

A more recent culture independent method which has developed alongside the improvements and decreasing cost of DNA sequence is the field of metagenomics. Unlike 16S rRNA gene amplicon sequencing where a short region is exclusively sequenced, metagenomics involves the sequencing of all DNA present in a sample including fungi, bacteria, archaea and viral sequences as it is not constrained by selective amplification. Reconstruction of this data into complete, or near complete, genomes has allowed researchers to infer functional capabilities of these uncultured organisms and provide insight into complex environmental processes (Dutilh et al., 2014; Maus et al., 2016a). This type of reconstruction has also allowed previously unculturable organisms to be cultured, as better understanding of their metabolic pathways allows a more targeted culturing approach. One such recent event was identification of a new archael "Asgard" superphylum, from metagenomic analysis of aquatic sediments aided by the previous discovery of Lokiarchaeota (Zaremba-Niedzwiedzka et al., 2017). Members of this superhylum were enriched using a genomeiguided approach and the first species (*Candidatus Prometheoarchaeum syntrophicum* strain MK-D1) isolated in an amino acid-fed co-culture alongside *Methanogenium,* its essential hydrogen consuming partner (Imachi et al., 2020). This type of functional annotation is especially useful for understanding community structure, functional ability and effects of supplements (such as activated carbon) in anaerobic digestion microbiomes (Bertucci et al., 2019; Guo et al., 2015; Zhang et al., 2019a; Zhu et al., 2019).

Metagenomic data gathered is therefore instrumental in analysing the total composition and overall metabolic potential of any given community. However, what it fails to do is provide information regarding active populations and active pathways, which is arguably more informative for understand the anaerobic digestion of a given feedstock. A certain functional redundancy is desirable however, to preserve robust operation of anaerobic digesters, and avoid digester "crashing" when it encounters an unfamiliar feedstock, but this results in many of the organisms present not majorly contributing to the biological processes of the AD community (Cai et al., 2016; de Vladar, 2012). Most industrial scale digesters contain a highly specialised community, since they are operated for decades rather than months in most anaerobic digestion scientific studies, that preserves functional redundancy by microbial immigration due to nonsterile feedstock (Agler et al., 2008; Kirkegaard et al., 2017; Moestedt et al., 2015; Saady and Massé, 2013). This degree of specialisation may not occur in scientific studies depending on their duration, and especially not in BMP type batch tests. However, the reason behind these specialised communities is the enriched presence of microbes which can be considered

taking an active part in the hydrolysis, acidogenesis, acetogenesis and methanogenesis of material present in the feedstock.

The previous small-scale biomethane potential (BMP) batch tests of pot ale anaerobic digestion revealed the potential for a readily digestible and biomethane producing feedstock. However, from the short tests, it was not clear that stable long term digestion could be carried out as potential inhibitors may accumulate in reactors over time such as the heavy metals present in the material such as copper. In order to perform these longer duration tests a custom-designed experimental system was constructed which included; sealed vessels easily interacted with through various ports and valves, a tipping bucket gas volume analyser and an array of 0-100% methane sensors. This set up was first tested and refined with feedstocks other than pot ale such as synthetic waste water, grain residue and chemical waste to ensure robust and reliable performance. The larger reactor volume and subsequent sampling volume allowed a greater range of measurements to be taken such as chemical oxygen demand and volatile fatty acid analysis, and the less risky sampling method, in terms of oxygen introduction, complementing disturbance-free metagenomics and metatranscriptomic community analysis.

Identification of active taxa and pathways in response to feedstock addition allows key organisms and pathways to be highlighted from background noise thus paving the way for targeted bioaugmentation or supplementation. For example; acclimatising a reactor from one carbon rich feedstock to a dissimilar nitrogen rich one will take some time as the community will shift and select for protein degrading and ammonia tolerant genera such as *Sporoarcina* and *Proteiniphilum* (Xu et al., 2019). It therefore makes sense to bioaugment reactors with organisms to speed up the transition period and increase methane yield, as is the case with a spent grain fed reactor supplemented with *Pseudomonas xylanivorans* Mz5^{T} (Fanedl et al., 2015).

Metatranscriptomics provides an opportunity to identify the function and diversity of actively expressed genes, termed a transcript, and organisms. In response to certain stimuli, for instance the temperature decrease when summer turns to winter, certain organisms such as fungi become less metabolically active when compared to their bacterial counterparts, who are frantically expressing transcripts to synthesise antifreeze proteins (Lorv et al., 2014; Rampelotto, 2014; Žifčáková et al., 2016). Complex metabolic networks in microbial communities such as biofilms, soils and sludges are more clearly understood when the sequence dataset is “denoised” by removing inactive genes and genomes from inactive organisms. Doing so greatly enriches the dataset and has aided identification

of metabolically active organisms responsible for hydrogen production, hydrogenotrophic methanogenesis, acetoclastic methanogenesis and methanotrophy in the flooded soil methane cycle in addition to exclusive cellulose degraders (Jia et al., 2018; Masuda et al., 2018).

In the field of AD, metatranscriptomics has previously and successfully been applied to continuously fed AD systems to highlight active metabolic networks and identify key organisms in full scale systems since the year 2012 (Bremges et al., 2015; Hassa et al., 2018; Vanwonterghem et al., 2014; Zakrzewski et al., 2012). However, to date there has been no metatranscriptome time course investigation of an AD system to provide a temporal dimension to any of the various hydrolysis, acidogenesis, acetogenesis and methanogenesis pathways and this study represents the first temporal resolution of community wide gene transcription. This knowledge gap is compounded by the fact that the hydrolysis step is considered to be rate limiting (Appels et al., 2008; Campos et al., 2008; Cherif et al., 2014; Fanedl et al., 2015; IWA Task Group for Mathematical Modelling of Anaerobic Digestion Processes, 2002; Saady and Massé, 2013; Ward et al., 2008; Xu et al., 2015). Identification of rate limiting hydrolysis steps though metatranscriptomics would provide a target to focus pre-treatment, bioaugmentation or enzyme addition efforts to improve rates of AD.

Thus the aim of this study is to: first construct a system to, and then perform, long term digestion of the whisky waste material pot ale to identify signs of inhibition, secondly examine degradation kinetics using methane gas sensors, VFA and COD analysis, thirdly examine microbial community and functional profile dynamics over the course of the experiment using metagenomics and finally report the first AD metatranscriptome time course, identifying the respective “steps” of AD and describing specific enzymes responsible for the rate limiting hydrolysis step. This is surmised below in Figure 4.1.

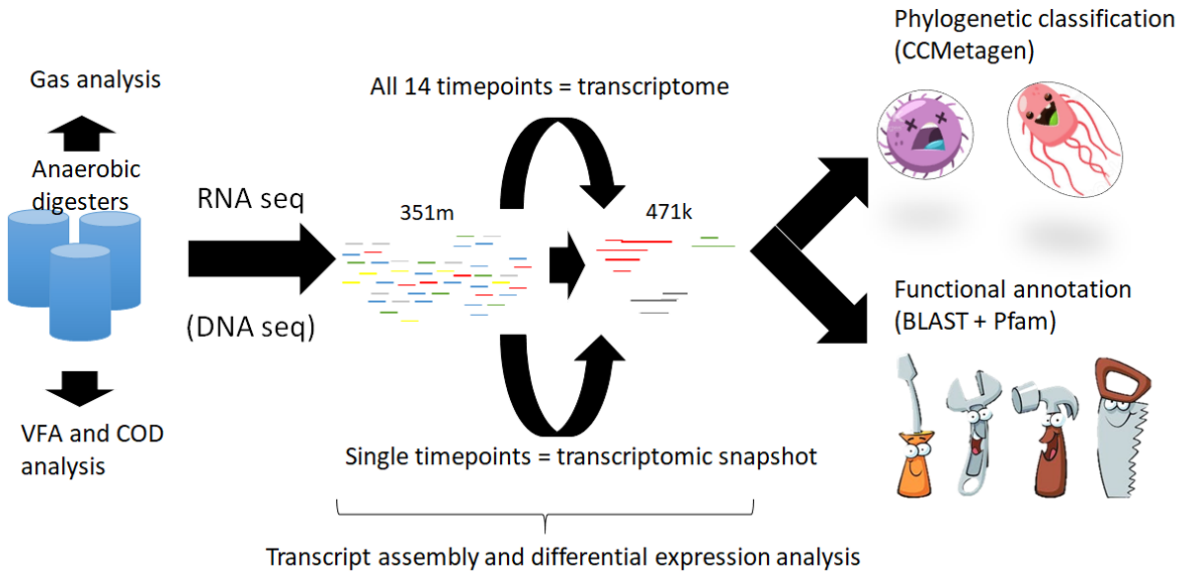


Figure 4.1 Experimental overview of a combined metagenomic and metatranscriptomic analysis of an anaerobic digestion community complemented with gas and metabolite analysis.

The experimental plan was to operate triplicate anaerobic digestion reactors and take gas composition measurements alongside COD and VFA analysis. Time point RNA sequencing reads will either be assembled together, to provide a reference transcriptome, or sample-wise to provide transcriptomic snapshots into the anaerobic digestion system. Assembled contigs were then annotated and analysed using a variety of software packages.

4.4 Materials and Methods

4.4.1 Development of lab scale AD system

4.4.1.1 Custom reactor vessels

A total of 11 reactor vessels were constructed from 1 L wide mouth bottles (Fisher Scientific, Cat no 11774329) and fitted with custom designed lids and parts produced from the biology mechanical workshop at the University of York by Mark Bentley. Figure 4.2 is an image of the system connected to an array of methane sensors and gas volume analysis.

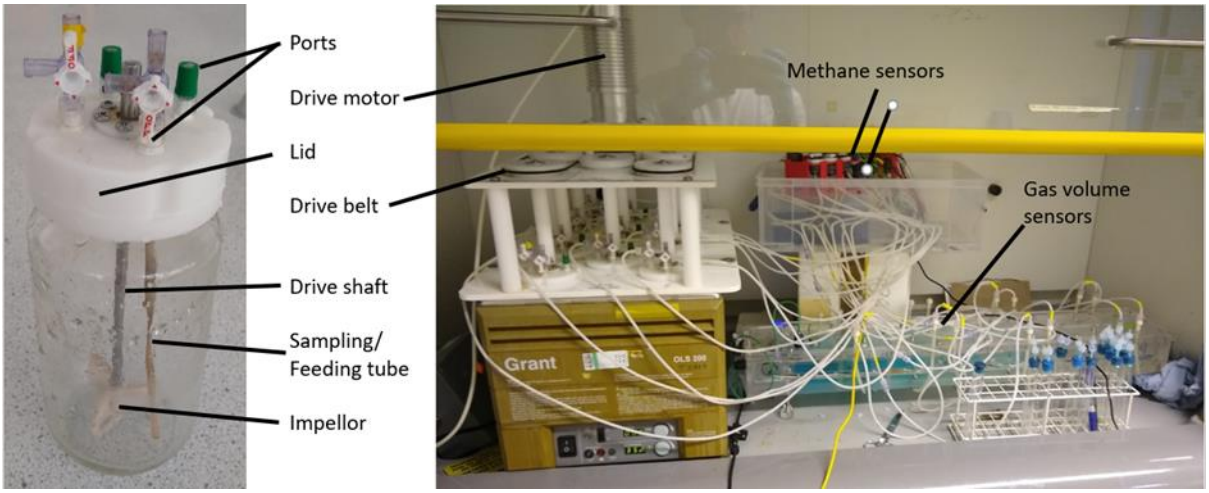


Figure 4.2 Lab scale anaerobic digestion system set up

Diagram showing the features of the reactors which contain the sludge (left) in relation to the multiple pieces of equipment used to construct the whole system (right). A water bath heats the glass reactors and gas produced travels to the methane sensors controlled by an Arduino microcontroller before passing through a tipping bubble counter to analyse gas volume.

Computer Aided Design blueprints and dimensions of all custom parts are available in Supplementary (file: Mini Digestor blueprints) and were machined out of Acrylonitrile butadiene styrene plastic unless otherwise stated. Lids were designed to allow addition or removal of liquid and gas in an airtight environment compatible with Luer type fittings and equipment and facilitate stirring. Four threaded ports were drilled into the top of the lids to allow insertion of a threaded female luer bulkhead with hose barb, a female bulkhead allowed attachment of luer fittings to the outside of the reactor, the thread providing a tight seal (which was reinforced using liquid thread sealant, Loctite 577 medium strength) and the barb on the inside of the lid to allow attachment of tubing to feed or sample directly into the liquid phase. Gas tightness of the glass:lid interface was ensured by a groove inside the lid holding an O-ring (3 mm section 60 mm bore nitrile-70 rubber O-ring, Simply Bearings).

4.4.1.1.1 Sampling interfaces

The lid port at the 6'o clock position was fitted with both a 3-way luer stopcock (WZ-30600-02, Cole-Parmer) on the female bulkhead side, and 150 mm of silicon tubing (C-Flex 3/32 ID opaque tubing, WZ-06424-63, Cole Parmer). All tubing used for construction of the lab scale system was this product and shall be referred to as tubing from this point onwards) attached to the barb to allow sampling/feeding of reactor contents without removing the lid. The lid port at 9'o clock was fitted with a 3-way luer stopcock that tubing was attached to via a male luer hose barb (WZ-45518-02, Cole Parmer) to allow downstream composition and volumetric analysis of gas produced by the reactors.

The two remaining ports were fitted with male luer stopper injectable membranes as backup (891.00, Vygon)

4.4.1.1.2 Vessel Stirring + Heating

A hole through the centre of the lid allowed a 6 mm wide metal shaft, tipped with a T-shaped cross-shaft on the outer-part, to pass through to stir the reactors with a 3D printer impellor (9 clockwise and 2 anti-clockwise). A rubber rotating gasket ensured gas tightness between the shaft and the lid (6262RS budget rubber sealed deep groove ball bearing 6x19x6 mm, 3 mm section 6 mm bore nitrile-70 rubber O-rings and 6x14x4 mm nitrile rubber rotary shaft oil seal springless design VS style, all Simply Bearings) which was held down by a cover and screws (Mini Digestor blueprints, page 6, green). The T-shaped cross-shaft protruding from the top of the vessel allowed a plastic bracket (Mini Digestor blueprints, page 3, blue) to form an interface with the vessel and the motor-powered pulley system which provided continuous stirring (Mini Digestor blueprints, page 5).

The pulley system was constructed from a baseplate, pulley-wheels and shafts which are detailed in the Mini Digester blueprints, page 2-4). Two large 277 mm bore nitrile O-rings (BS277 NBR, EAP International LTD) were placed around the central pulley wheel and fitted into the groove sections of the pulley wheels to provide drive to all vessels at an equal RPM.

Vessels were kept at 35°C in a waterbath (OLS200, Grant) filled to above the sludge level inside vessels

4.4.1.1.3 Gas composition monitoring

Gas from the vessels was carried via tubing to an array of 11 x Dynament 0-100% 1-5 V infared analogue methane sensors with PCB leads, sampling adaptors and mounting kits (P/HC/5/V/P 04ATEX1357U, Dynament LTD). Sensors were calibrated by the provider to provide a high (100% methane) and low (0% methane) signal range of 4-2V, this was done to fault indication if voltage dropped below 2 V (loss of power) or increased above 4 V (water on sensor causing a short circuit).

Hourly methane concentration was measured by connecting analogue output wires from the gas sensors to analogue input pins on an open source microcontroller (Arduino Mega 2560, Arduino) running a simple programme which sequentially;

1) turns sensors on for one minute to warm up (as recommended by the manufacturer) via a relay, 2) Records voltage input and transforms it to a methane % value where 0.5 V = 0%, 1.5 V = 50% and 2.5 V = 100% using the Arduinos analogue to digital converter which provides a 4.9 mV resolution by providing integers between 0-1023, 3) appends methane concentration values for each vessel to a text file in addition to the hour of operation and 4) pauses for 59 minutes before returning to 1).

Sensor accuracy was checked by passing 20% and 70% methane calibration gas (1 L 10 bar, CK special gases LTD) through the sensor and housing and showed deviation less than 4% i.e sensors showing 22-24% for the 20% calibration gas.

4.4.1.1.4 Gas volume monitoring

A "tipping bucket mechanism" gas flow meter was purchased from Anaerotechnology (anaerotech.com) which uses a system whereby gas is collected in an inverted bucket until is buoyant enough to flip over and trigger a magnetic switch. A thermometer and barometer inside the instrument ensure that tips are converted into STP gas measurements by the on board microcontroller (Also and Arduino).

Biogas flows from the sensor array through a silicon diaphragm one way check valve (WZ-30505-92, Cole Parmer) before reaching the tipping bucket mechanism. The check valve was installed to prevent water flowing backwards from the tipping bucket counter into the sensors in the event of a water bath failure (which did occur three times over its usage period) due to the subsequent pressure drop that would be created by the cooling vessel headspace gas.

4.4.2 Operation of lab scale AD system

4.4.2.1 Reactor vessel seeding

Sludge inoculum was sourced from the same site used to provide inoculum in Chapters 2 and 3: a full scale waste water treatment plant anaerobic digester operated by Yorkshire Water, UK at the Naburn plant site located at 53°54'52.5"N 1°05'07.9"W which digests primary and secondary sludge from the water treatment process. Sludge collected was taken from a tap which samples the centre of the reactor and transported back to the lab in a barrel kept at ambient temperature.

At the lab, sludge was stirred using a paddle prior to weighing 400 g into each glass vessel under aerobic conditions. Weight rather than volume was used as a sludge measure to ensure bubbles present in the liquid did cause uneven seeding of reactors.

4.4.2.2 Reactor sampling and feeding

If sampling and feeding were to occur they were done in that order. To sample reactors and keep avoid introduction of oxygen the syringe: vessel interface was first formed on the 12 o'clock junction of the feeding/sampling valve (see Figure 2) before moving valve position from the 12 o'clock position (closed) to the 9 o'clock position (creates an open line between the syringe and inside the vessel). The empty syringe plunger was withdrew partly before "rinsing" by withdrawing and injecting approximately 20 ml of sludge three times. Finally volumes of either 15 (only when performing RNA sampling) or 50 ml (when feeding to keep reactor volume constant) were taken.

Vessels were fed in triplicate with 50 ml pot ale aliquots at v/v concentrations of either; 25% pH 4, 25% pH 8 (adjusted using NaOH), 50% pH 4 or 100% pH 4 using a 50 ml syringe with a short section of tubing and a male luer barb for leakproof interface with the vessel feeding ports. A similar strategy was used with a designated feeding syringe (see previous sampling strategy) to feed reactors and ensure thorough mixing. Samples were placed in either 15 or 50 ml falcon tubes and spun at 4°C for 5 minutes at 10,000 g (Eppendorf 5810R with Eppendorf F-34-6-38 rotor) before separate solid and non-filtered liquid fractions were flash frozen in liquid nitrogen for five minutes. Samples fed with 25, 50 or 100% pot ale are considered to be part of phase 1,2 or 3 respectively and are abbreviated to p1d28 (phase 1, day 28) or similar.

At the end of the 6 month experimental duration reactor performance was considered to be similar to operational anaerobic digestion systems (producing ~60% v/v methane) and so the decision was made to continue onto the RNA sampling phase. As the RNA sampling phase involved extraction of reactor contents without replacement, two additional rounds of batch feeding at the highest concentration were performed to confirm reactor stability at the top end of the feeding regime covered in this experiment as gas production data could not be collected with a continually decreasing reactor volume.

4.4.3 Volatile fatty acid analysis

4.4.3.1 GC-FID set up

Volatile fatty acids (VFAs) are breakdown products of organic molecules in the anaerobic digestion process with acetic acid being dismutated to produce methane and carbon dioxide.

An Agilent 5890 GC-FID was fitted with a Nukol 30 m x 0.25 mm I.D 0.25 um (Sigma, 24107) column containing a poly (ethylene glycol) phase modified with acid functional groups. Gasses were supplied to the instrument as followed: air (purchased on-site) at 30 PSI; H_2 (supplied by a hydrogen generator, 20H, Dominik Hunter) at 20 PSI and He (Helium grade A size L, BOC) at 40 PSI. Injectors and detectors were both at 200°C. The temperature gradient for the oven was a two-step ramp followed by a hold; 75 to 150 (10°C / min), 150 to 200 (20°C / min), hold at 200°C for 10 minutes. Column head pressure was set to 20 PSI. The unmarked analogue knobs controlling the flow rate and septum purge were approximately set to minimum and halfway respectively. A 1 µL injection volume was used for all samples.

Calibration standards were prepared from serially diluted volatile fatty acid mix (CRM46975) which provides 10 mM of acetic, propanoic, isobutyric, butyric, isopentanoic, pentanoic, isohexanoic, hexanoic and heptanoic acids. The limit of detection for shorter VFAs was lower (above 0.01 mM) than that of longer VFAs (especially heptanoic acid which could only be detected above 1 mM).

4.4.3.2 Sample preparation

If liquid fraction samples were frozen, they were defrosted on ice prior to use. 2 ml of unfiltered aqueous sample was transferred to a 2 ml micro-centrifuge tube and centrifuged at max speed in a table top centrifuge for 5-30 minutes until visual inspection revealed a clear supernatant (note this may be difficult for very dark sludge samples). A sterile 1 ml syringe was attached to a 0.22 um syringe filter unit and the plunger removed. A minimum of 1.2 ml clear supernatant from the centrifuged sample was either decanted or transferred via a pipette into the barrel of the syringe and filtered into a fresh 1.5 ml micro-centrifuge tube. Exactly 1 ml of the filtered sample was transferred into a glass GC vial (29432-U, Sigma) and acidified by adding 7.5 µL orthophosphoric acid (345245, Sigma) to fully protonate organic acids in the sample ensuring their interaction with the acid-modified stationary phase.

4.4.4 COD analysis

Chemical oxygen demand (COD) is a measurement of oxidisable material, both organic and inorganic, in a given solution and is used in the water industry as a feedstock concentration measurement (Pasztor et al., 2009).

Homogenous sludge samples were first diluted ten-fold in ddH_2O then analysed using COD measurement kits (100-2000 mg/L LCK514, Hach Lange) according to manufacturer's instructions which required an oven and spectrophotometer (HT200-S and DR 3900 respectively, Hach Lange). These kits measure oxidative change by heating potassium dichromate ions to 200°C in sulphuric acid alongside an Ag_2SO_4 catalyst (Dedkov et al., 2000).

4.4.5 Metagenomic DNA sampling, sequencing and analysis

4.4.5.1 Metagenomic DNA extraction

Metagenomic DNA was extracted from solid sample fractions using the DNeasy Powerlyzer PowerSoil Kit (12855, Qiagen) according to manufacturer's protocol. Bead beating was performed using a MoBio Vortex Genie 2 with microfuge adapter. An ice/water bath was used to chill samples at 4°C when specified as it allows more rapid heat conduction compared to a fridge. In summary: Biomass is vortexed with glass beads to mechanically break open cells, proteins are precipitated out of solution using a detergent, humic substances and polyphenols (which give the characteristic brown colour) and removed using a PVP-like substance, DNA is precipitated with a guanidine salt solution and bound to the column, ethanol is used to wash the bound DNA and finally clean microbial community genomic DNA is eluted using ddH_2O. DNA purity and concentration was measured using a Nanodrop ND-1000, high concentrations over 200 ng / µL were typical with a high quality of A260/A280 values between 1.8-1.9 and A260/A230 values between 1.9-2.0. Fragment size typically varied between 10-30 kb as confirmed by gel electrophoresis.

4.4.5.2 Metagenomic short read DNA library preparation

The whole-metagenome sequencing library was prepared from samples taken across phases 1,2 and 3 using a TruSeq DNA PCR-Free High Throughput Library Prep Kit (20015963, Illumina) by the Genomics Facility at the University of York. In summary: large DNA fragments were first sheared using ultrasonication (M220, Covaris), sticky ends of DNA fragments are repaired using fill-in reactions, size selection is then performed using 350 bp magnetic purification beads, an A nucleotide is joined to the

blunt end of each strand to facilitate the adapter-index ligation containing a T nucleotide overhang. Sequencing was performed using an Illumina HiSeq 3000 at the University of Leeds with a 2 x 150 bp sequencing run. All short reads are deposited under the SUB6360036 BioProject.

4.4.5.3 Metagenomics long read DNA library preparation

Metagenomic DNA from phase 2 and phase 3 samples was pooled and sequencing libraries prepared according to manufacturer's instructions before being sequenced on the Oxford Nanopore MinION device (R9 chemistry) and base called using Albacore settings. All long reads are deposited under the SUB6360036 BioProject.

4.4.5.4 Metagenomic DNA bioinformatics analysis

4.4.5.4.1 VIKING supercomputer and analysis environment

Data from the DNA sequencing run was transferred over to the University of York VIKING supercomputing cluster. VIKING offers a high number of Intel Xeon Gold/Platinum CPU cores available (40 for a standard job and 96 for a high memory job) and large amounts of RAM (200 GB for a standard job and 1.5 TB for a high memory job) compared to a conventional office computer which may only have 4 CPU cores and 8 GB of RAM. The large amount of RAM is absolutely essential for the analysis of large sequence datasets in order to perform pairwise comparisons on the billions of sequences generated in this experiment.

VIKING was accessed via PuTTY v0.72, a free and open source SSH and telnet client for windows, developed by Simon Tatham available at (https://www.putty.org). Office computer to VIKING secure file transfers were managed using WinSCP v5.15, a free and open source SFTP and FTP client, developed by Martin Přikryl available at (https://winscp.net/eng/download.php).

A personal environment was created on VIKING using Conda, a free and open source package and environment management system available at (https://docs.conda.io/en/latest/) to allow rapid bioinformatics package installation and management. This allowed multiple environments to be created for each piece of key analysis software since each key software sometimes required different versions of a supporting packages especially for older key software which is incompatible with more recent packages.

To run a piece of analysis, a job request first had to be submitted to VIKING using the Slurm Workload Manager, a free and open source job scheduler available at (https://slurm.schedmd.com/documentation.html), which specifies requested resources such as; number of CPU cores, amount of RAM, maximum run time, script used and parameters required by the script. Additionally, it allows arrays of multiple jobs with variable permutations to be run.

4.4.5.4.2 Short read quality control and adapter trimming

FASTQC was installed via conda and ran directly in the command line of VIKING on all sets of paired end DNA reads to asses sequencing quality and GC% using the command below.

```
srun -c 40 fastqc -t 40 P*.fastq.gz -o fastqc/ --nogroup
```

Mean Phred quality score across the reads were calculated based upon chromatogram identity. The Phred score is logarithmically related to the probability of base calling error where 30 = a probability of incorrect basecalling of 1 in 1000 and 40 = a probability of incorrect basecalling of 1 in 10,000. (Supplementary Figure 1). As the first 5 bases and last of the reads has Phred scores between 31 and 35 compared to the majority of the rest with 39 and FASTQC identified that adapters were still present these were trimmed using the conda package "cutadapt" (https://cutadapt.readthedocs.io/en/stable/) with the adapter sequence "AGATCGGAAGAG" which corresponds to the Illumina universal adapter (Script 1). FASTQC was ran again which confirmed removal of adapters and suitability for sequence assembly.

4.4.5.4.3 Short read assembly

Individual samples were assembled using the short read de-novo assembler MEGAHIT already installed on VIKING (D. Li et al., 2015) using the command below as part of a Slurm job (Script 2) where N corresponds to the trimmed sequences filenames for forward (R1) or reverse reads (R2). The SPADES assembler was also trialled but repeatedly crashed due to the large dataset.

```
megahit --k-min=27 -1 NR1 -2 NR2 -o Nassembly
```

Assembly metrics such as total size, number of contigs and N50 were provided for each sample assembly by the conda available package "SeqKit" for analysis and manipulation of FASTA/Q format sequence files using the command below directly in the VIKING terminal (Shen et al., 2016).

```
seqkit stats Nassembly
```

4.4.5.4.4 Long read assembly

Long reads generated from sequencing a pooled sample on the Oxford Nanopore MinION were assembled using the CANU assembler, a branch of the Celera Assembler, which is optimised for high noise single molecule sequencing (Pinto, 2014). A metaSPAdes assembly was also performed on the dataset for comparison but yielded similar results in terms of number of contigs and contig size and was not used for subsequent analysis steps (Nurk et al., 2017). Both the passed and the failed reads were used as input as CANU performs its own correction step. As the job required significant computational power the script below was submitted via slurm with a 1 TB RAM and 80 CPU core requirement where P and D represent arbitrary directory names and N* is both the passed and failed nanopore reads.

```
canu -p P -d D minReadLength=1000 minOverlapLength=500 genomeSize=500m corOutCoverage=all
corMhapSensitivity=high corMinCoverage=0 -fast -nanopore-raw N*
```

4.4.5.4.5 Polishing the long read assembly

4.4.5.4.5.1 FAST5 signal polishing with Nanopolish

Nanopolish is a tool that was developed to improve the accuracy of assemblies by using the raw electric signal generated from Nanopore-type sequencing devices in addition to detecting modified bases such as 5-mC using a hidden Markov model (Simpson et al., 2017).

As the reads generated by the long read sequencing strategy are of high length but noisy a further round of consensus improvement was performed by indexing the basecalled sequencing reads (FASTA format) with the raw signals (FAST5 format) using slurm script 4 containing the command below where fast5 directory contains all the raw signal.

```
nanopolish index -d [fast5 directory] draftCanu.contigs.fasta
```

Indexed reads were then mapped onto the draft assembly using the minimap command below as part of script 5.

```
minimap2 -ax map-ont -t 80 JR_PA_NP_CANU_Draft
/draftCanu.contigs.fasta
/ merged.fastq.gz | samtools sort -o reads.sorted.bam -T reads.tmp
```

The draft CANU assembly was split into 16 folders, each containing just below 1,000 jobs for a 50 kb region of the genome totalling 14,883 individual segments in the script below.

```
split -l 990 NP.contigs.ranges.fasta
```

Reads were then mapped back onto 50 kb segments of the CANU assembly to generate a higher consensus segments and were submitted as a 14,000 job Slurm array using permutations of the command below as part of script 6 where $i corresponds to the ID of each 50 kb segment. This rewrote the draft CANU contigs with the improved consensus contigs using substitutions, insertions and deletions of bases.

```
nanopolish variants --consensus -o ArrayNPVCF/$i.polished.vcf -w $i -r /users/jr1214/biol-chong-
2019/2019-02-07-JR_PotAle/data/nano/merged.fastq.gz -b reads.sorted.bam -g
JR_PA_NP_CANU_Nonpolished/JR_PA_NP.contigs.fasta -t 10
```

The resulting polished .VCF filetype segments was then merged and converted into a FASTA format containing the signal polished draft CANU assembly with the command below as part of script 7.

```
nanopolish vcf2fasta --skip-checks -g JR_PA_NP_CANU_Nonpolished/JR_PA_NP.contigs.fasta
ArrayNPVCF/* > polished.fa
```

4.4.5.4.5.2 Short read polishing with Pilon

A further round of draft genome polishing was performed using the timepoint samples submitted for metagenomics short read sequencing using the software Pilon. Pilon takes advantage of the high accuracy and high coverage of short reads to correct bases in the more error prone long read assemblies (Walker et al., 2014).

Firstly forward and reverse reads from all samples were pooled in respective files using the two commands below directly In terminal.

```
cat *R1_001.fastq.trimmed.gz > R1.merged.fastq.gz
cat *R2_001.fastq.trimmed.gz > R2.merged.fastq.gz
```

The two files containing each short read direction were then input into Pilon and used to polish the FAST5 polished draft CANU assembly with the command below as part of a slurm job submission (Script 8). Between one and three rounds of polishing were performed.

```
slurm_pilon.sh Pilon5ItPolish R1.merged.fastq.gz R2.merged.fastq.gz polished.fa 40 3
```

The phylogeny of each contig in the polished metagenomics assembly (long read genome assembly base corrected using the short reads with higher accuracy) was then examined by performing a local alignment on the entire NCBI nucleotide database using BLASTN with the following command which returns the best match for each contig (Altschup et al., 1990).

```
blastn -db /mnt/lustre/groups/biol-chong-2019/databases/blast/nt/nt -query polished.fa -out
polished.fa _blastn.hm.out -outfmt 6 -max_target_seqs 1
```

Results were then sorted in order of longest alignment match to shortest to identify contigs which belong to organisms of potential interest using the command below. Both parts of the BLAST search were performed as a Slurm job (Script 9).

```
sort -t, -k1,1 -k12,12nr -k11,11n  $1_blastn.hm.out | sort -u -k1,1 --merge > bestHits.$1.hm.txt
```

4.4.6 Metatranscriptomic RNA sampling and sequencing

4.4.6.1 Metatranscriptomic RNA extraction

Metatranscriptomic RNA was extracted from solid sample fractions using the RNeasy Powerlyzer PowerSoil Kit (12866-25, Qiagen) according to manufacturer's protocol, except that 0.5 g biomass was used rather than 2 g. It was found that using more than 0.5 g biomass produced so much RNA that it prevented subsequent elution in a timely manner. An ice/water bath was used to chill samples at 4°C when specified as it allows more rapid heat conduction compared to a fridge. In summary; biomass is vortexed with glass beads alongside phenol/chloroform (pH8) and SDS to mechanically break open cells, proteins and lipids are separated from the aqueous phase by centrifigutation and further clean up with a second phenol/chloroform extraction. Humic substances and polyphenols (which give the characteristic brown colour) were removed using a PVP-like substance, a second protein precipitation step removes remaining cellular debris. Isopropanol was then used to precipitate DNA+RNA which is then pelleted and the DNA+RNA pellet resuspended in a salt solution before being passed over a column where the nucleic acids bind indiscriminately. RNA is then preferentially eluted using an acidic

solution leaving cellular debris and DNA on the column before RNA being precipitated again with isopropanol and pelleted before being resuspended in nuclease free water.

RNA purity and concentration was measured using a Nanodrop, high concentrations over 200 ng / µL were typical with a high quality of A260/A280 values between 1.9-2 and A260/A230 values between 2.0-2.1.

Quality of the RNA samples was quantified by assessing the RNA integrity number (RIN) which is the ratio of 23S:16S rRNA (Schroeder et al., 2006). Determination of this value was calculated by running 1 µL of RNA sample on a gel electrophoresis tapestation (2100 Bioanalyzer, Agilent) according to the manufacturer's instructions. Samples above RIN 7 were considered to be "intact".

4.4.6.2 Metatranscriptomic RNA library preparation

rRNA was removed from the mixed RNA sample using a Ribo-Zero Bacterial rRNA Removal Kit as per manufacturers instructions (Unfortunately a discontinued Illumina product, although their website suggests a new product is pending) to leave mRNA lncRNA and miRNA enriched samples. In summary; rRNA complementary probes hybridise rRNA which prevents binding to magnetic beads, bound mRNA is then pelleted using sodium acetate and glycogen before being washed using ethanol.

Indexed cDNA was synthesised from the purified mRNA using the NEBNext ultra II Directional RNA Library Prep Kit for Illumina as per manufacturer's instructions. In summary, mRNA is first fragmented and hybridised with random primers, the first cDNA strand is synthesised from the template mRNA, the first cDNA strand is then used as a template for synthesis of the complementary second strand cDNA, magnetic beads are used to purify the cDNA and remove unwanted mRNA, adapters and indicies are then ligated onto the doublestranded cDNA template in a similar manner as discussed previously for metagenomics DNA library prep. All RNA reads are deposited under the SUB6360036 BioProject.

4.4.6.3 Metatranscriptomic mRNA bioinformatics analysis

4.4.6.3.1 Transcriptome assembly

Trinity was used to construct a de-novo metatranscriptome from reads generated across the 14 mRNA sampling timepoints (Grabherr et al., 2013). Trinity was used as it has previously been shown to effectively assemble complex community datasets such as in soils and sludges (Bang-andreasen and Anwar, 2019; Delforno et al., 2018; Hayden et al., 2018). Unlike genome assemblers, which are optimised to overlap large sequence graphs, transcriptome assemblers account for the transcriptional complexity such as isoforms and "islands" of sequence graphs. The transcriptome assembly and adapter sequence trimming were performed using the command below as part of script 10.

```
Trinity --seqType fq --max_memory 500G\
--left [Sample1 Name]R1_001.fastq.gz, [Sample2 Name]R1_001.fastq.gz,…
--right [Sample1 Name]R2_001.fastq.gz, [Sample2 Name]R2_001.fastq.gz,…
--CPU 64 --trimmomatic
--output ../../asm/illumina/Trinity_Whole_Transcriptome/Trinity.fasta
```

4.4.6.3.2 Transcript quantification

Transcripts were quantified using RSEM software which is optimised for transcriptomic datasets without a reference transcriptome (Li and Dewey, 2011). Reads are then ambiguously mapped onto the reference genome using the command below as part of script 11.

```
align_and_estimate_abundance.pl --transcripts Trinity.fasta --seqType fq --left
JRR1_*R1_001.fastq.gz.P.qtrim.gz
--right JRR1_*R2_001.fastq.gz.P.qtrim.gz --est_method RSEM
--aln_method bowtie --trinity_mode –prep_reference
--output_dir JRR1_rsem_outdir
```

4.4.6.3.3 Gene expression matrix

The counts from the ambiguously mapped reads in the previous step are converted into normalised gene expression values TPM (transcripts per million)

```
abundance_estimates_to_matrix.pl --est_method RSEM  --gene_trans_map
Trinity.fasta.gene_trans_map --name_sample_by_basedir 1_*/RSEM.isoforms.results
2_*/RSEM.isoforms.results 3_*/RSEM.isoforms.results 4_*/RSEM.isoforms.results
5_*/RSEM.isoforms.results 6_*/RSEM.isoforms.results 7_*/RSEM.isoforms.results
8_*/RSEM.isoforms.results 9_*/RSEM.isoforms.results 10_*/RSEM.isoforms.results
11_*/RSEM.isoforms.results 12_*/RSEM.isoforms.results 13_*/RSEM.isoforms.results
14_*/RSEM.isoforms.results
```

4.4.6.3.4 Visualisation of gene counts

Counts of transcriptome features and their abundance were calculated and visualised using the command below directly in the terminal. In short it counts how many transcripts are expressed above a certain threshold as we assume that there will little experimental significance to transcripts expressed at extremely low values such as TPM = 0.001.

```
count_matrix_features_given_MIN_TPM_threshold.pl genes_matrix.TPM.not_cross_norm | tee
genes_matrix.TPM.not_cross_norm.counts_by_min_TPM
```

4.4.6.3.5 Removal of lowly expressed transcripts

Although lowly expressed transcripts may be of some biological significance, transcripts which have a TPM of zero indicate that improper assembly occurred at the first TRINITY step since paired end reads were unable to map onto the transcript and assign abundance values. This assumes that most biologically relevant transcripts were able to be quantified using RSEM and have a TPM value > 0. Removal of improperly assembled transcripts was done with the following command directly into the terminal.

```
filter_low_expr_transcripts.pl –matrix RSEM.gene.TPM.EXPR.matrix
--transcripts Trinity.fasta –min_expr_any 0 –trinity_mode
```

4.4.6.3.6 Extraction and annotation of coding regions

TransDecoder was used to identify open reading frames (ORFs) longer than 100 amino acids in the reference transcriptome using the following command directly in the terminal to generate the file "transdecoder.pep" (Haas et al., 2013).

```
TransDecoder.LongOrfs -t Trinity.fasta
```

Nucleotide and protein BLAST was run on both the nucleotide and peptide sequences detected by TransDecoder using the two following commands entered directly into the terminal.

```
blastx -query Trinity.fasta -db uniprot_sprot.pep -num_threads 80 -max_target_seqs 1 -outfmt 6 -evalue 1e-3 > blastx.outfmt6
```

```
blastp -query transdecoder.pep -db uniprot_sprot.pep -num_threads 80 -max_target_seqs 1 -outfmt 6 -evalue 1e-3 > blastp.outfmt6
```

A third round of annotation was performed using an additional protein sequence simalirty probability prediction tool HMMER as per the command below directly into the terminal (Finn et al., 2011).

```
hmmscan --cpu 12 --domtblout TrinotatePFAM.out Pfam-A.hmm transdecoder.pep > pfam.log
```

4.4.6.3.7 Adding annotations to expression matrix

BioLinux was installed on a virtual machine (VM, Oracle) for the following analysis involving a Structured Query Language (SQL) database as the VIKING file structure was incompatable with SQL databases. Trinotate package was installed using conda was then used to import the various annotation files and finally output a report file using the following commands below entered directly into the terminal.

Transcripts of peptides identified and gene:transcript relationships were populated into the SQL database.

```
Trinotate Trinotate.sqlite init --gene_trans_map Trinity.fasta.gene_trans_map --transcript_fasta Trinity.fasta --transdecoder_pep transdecoder.pep
```

Homologies for the three rounds of annotation were also loaded into the database using the three following commands.

```
Trinotate Trinotate.sqlite LOAD_swissprot_blastp blastp.outfmt6
Trinotate Trinotate.sqlite LOAD_swissprot_blastx blastx.outfmt6
Trinotate Trinotate.sqlite LOAD_pfam TrinotatePFAM.out
```

Finally an Excel compatible annotation summary report (containing both annotated and mystery transcripts) was output using the command;

```
Trinotate Trinotate.sqlite report –E 1e-3 > trinotate_annotation_report.xls
```

Note that annotated transcripts did not necessarily have information from all three rounds of annotation. In general the order of annotation counts was nucleotide blast > protein blast > Pfam annotations.

4.4.6.3.8 Taxanomic annotation of reads

Taxanomic classification of metagenomics and metatranscriptomic DNA typically aligns each read to a reference database and assigns its identity with the best match. A recently developed tool, CCMetagen avoids the pitfall of where similar sequences can appear in multiple organisms by heavily weighting highly specific alignment reads and re-mapping ambiguous reads (Marcelino et al., 2019). This strategy has shown to produce more precise annotations in both metagenomics and metatranscriptomic datasets across a variety of microbiomes when compared to alternatives such as Kraken2 and MEGAN.

The steps to map and refine read alignment using CCMetagen were performed as described for paired-end datasets as described on their GitHub (https://github.com/vrmarcelino/CCMetagen accessed on 24/09/19). Timepoint mRNA reads were first mapped against the entire NCBI nucleotide database (vJan 2018) using paired end settings with the KMA mapper (Clausen et al., 2018). CCMetagen then processed the ranked reads and assigned taxonomy at the class level.

4.4.6.3.9 Assignment of Gene Ontology terms

The Gene Ontology (GO) project is a curated database which aims to develop a unified vocabulary of genes and gene products and provide a hierarchal classification system (Ashburner et al., 2000; Carbon et al., 2019). It classifies genes and gene products into three main categories; biological process (such as methanogenesis), molecular function (such as acetate kinases) and cellular components (which identifies where in the cell a gene product is active, such as a cellulosome). A complete GO term is a series of ID numbers which indicate the complete ancestory of a gene or gene product. For example, methylselenol demethylase activity ancestry is as follows.

GO:0008150 biological_process → GO:0008152 metabolic process → GO:0003674 molecular_function →GO:0003824 catalytic activity

GO:0003824 catalytic activity → GO:0070988 demethylation → GO:0032451 demethylase activity → GO:0098608 methylselenol demethylase activity

GO annotations, including all ancestral terms, were assigned to the annotations file using the UniProt database and the following Trinotate command below (Bateman, 2019; Bateman et al., 2015).

```
extract_GO_assignments_from_Trinotate_xls.pl --Trinotate_xls trinotate.xls -G --include_ancestral_terms > go_annotations.txt
```

The files generated from the Trinotate steps inside the virtual machine were then transferred back to VIKING and also the host office computer for further GO term analysis .

4.4.6.4 Development of GOfys

Existing software packages for GO term analysis such as GOrilla, GOnet, and DAVID are built around traditional model organisms such as *Homo sapiens*, *Arabidopsis thaliana*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, Danio rerio, *Mus musculus* and *Rattus norvegicus* (Eden et al., 2009; Huang et al., 2007; Pomaznoy et al., 2018).

After a short unsuccessful period of trying to modify them for use in anaerobic digestion community analysis, an alternative more robust and universally applicable, to other studies such as microbiomes or non-model organisms, piece of software was developed termed GOfys and pronounced “go fish”.

GOfys was used to analyse relationships (fish out results with GO terms, hence GOfys) between the transcript expression matrix, transcript clusters and transcript:GO annotations using a user provided list of GO terms of interest. Alternatively, GOfys will also detect the most N abundant GO terms (although actively excludes the top three groups; biological process, molecular function and cellular components since it ran for more than one week on VIKING’s high memory node and offers little insight into transcriptional dynamics). GOfys was written in the programming language Python and extensively used the Pandas library, a free and open source library for manipulating data tables available at (https://pandas.pydata.org/).

Prior to running GOfys, a GO terms Of Interest (GOI) file was created. GO number and IDs of processes of interest were found on the curated online GO annotation and ancestry database AmiGO available at (http://amigo.geneontology.org/amigo/dd_browse) (Carbon et al., 2009). A simple Excel equation below was used to take the pasted text from the respective level, sometimes containing more than 100 GO terms, on the inferred tree view and strip unnecessary characters, break the text into the GO ID number and the text description. The final two columns containing the GOI IDs and descriptions were then pasted into a text file for input into GOfys.

Pasted AmiGO text goes into the cell A1 and irrelevant characters stripped.

Column 2 =RIGHT(A1,LEN(A1)-(FIND("]",A1)+1))

The GO ID number is extracted.

Column 3 =LEFT(B1, 10)

The text description is extracted.

Column 4 =RIGHT(B1,LEN(B1) -10)

The following steps outline what key pieces of code powering GOfys does as described in the preceding comment indicated with a “#” symbol. Note that only the peptide and nucleotide blast results were analysed since Pfam annotation was sparse.

4.4.6.4.1 Step 1 Importation of required files

#Step 1.1 Creates frame from GOI input file.

```
GOI_frame = pd.read_csv(sys.argv[1],delimiter="\t", header=None,names =
['GO_ID','GO_DESC'],encoding = "unicode_escape")
```

#Step 1.2 Creates frame from TMM.matrix input file.

```
EXPRMTRX_frame = pd.read_csv(sys.argv[2],header=None,delimiter="\t")
```

#Step 1.3 Creates frame to input desired fields from trinotate summary report. Mainly tig number and ontology

```
Tig2GO_fields = ['#gene_id','gene_ontology_blast','gene_ontology_pfam']
```

#Step 1.4 Creates frame from only the desired columns

```
Tig2GO_frame = pd.read_csv(sys.argv[3],delimiter="\t",usecols=Tig2GO_fields)
```

#Step 1.5 Counts how many nucleotide or pr

```
BlastGOCount =  Tig2GO_frame[Tig2GO_frame['gene_ontology_blast'].str.contains("GO")]
BPfamGOcount = BlastGOCount[BlastGOCount['gene_ontology_pfam'].str.contains("GO")]
PfamGOCount =  Tig2GO_frame[Tig2GO_frame['gene_ontology_pfam'].str.contains("GO")]
PBlastGOCount =  PfamGOCount[PfamGOCount['gene_ontology_blast'].str.contains("GO")]
```

#Step 1.6 Designates the folder containing transcript clusters as detected with RSEM

```
path = 'Subclusters/'
```

#Step 1.7 Making cluster files Pandas compatible (only need to run once since it permanently modifies)

```
for cluster in os.listdir(path):
    src=open(path + cluster,"r") #fixed it?
    fline="tig\t"    #Prepending string
    oline=src.readlines()
    oline.insert(0,fline)
    src.close()
    src=open(path + cluster,"w")
    src.writelines(oline)
    src.close()
```

#Step 1.8 Imports clusters into Panda dataframe

```
cluster_frames = []
for cluster in os.listdir(path):
    cluster_name = cluster[:12]
    cluster_frames.append(cluster_name)
Cluster_Frame = pd.DataFrame(columns=[cluster_frames])
```

#Step 1.9 Creates empty report files to be populated

```
Text_Summary = []
Blast_summary_frame = pd.DataFrame()
Pfam_summary_frame = pd.DataFrame()
```

4.4.6.4.2 Step 2 Iteration through GOI and detection in transcripts

#Step 2.1 Iterates through each GOI

```
for index,item in enumerate(GOI_frame["GO_ID"]):
    #create empty dataframes from each variable in GO_ID
```

```
    pfam_frame = pd.DataFrame()
    blast_frame = pd.DataFrame()
    blast_sums = pd.DataFrame()
    pfam_sums = pd.DataFrame()
    cluster_GOI_counter = [0] * cluster_lengths
    cluster_GOI_frame = []
```

#Step 2.1 Populates each frame with the rows in the tig2go frame if it contains the GOI for both

```
    enriched_blast = Tig2GO_frame[Tig2GO_frame['gene_ontology_blast'].str.contains(item)]
    enriched_pfam = Tig2GO_frame[Tig2GO_frame['gene_ontology_pfam'].str.contains(item)]
```

#Step 2.2 Creates a text report indicating how many blast and pfam matches are found for each respective GOI in the ENTRIE dataset

```
    GOI_Sum =
 ((str(item) +' term codes for ' + GOI_frame.loc[index,'GO_DESC'] +' metabolism and contains '+
          str(len(enriched_blast)) + ' blast matches and ' + str(len(enriched_pfam)) +' pfam matches'))
    Text_Summary += (GOI_Sum,)
```

#Step 2.3 Iterates through the transcripts that are annotated with the GOI, puts them in a new matrix and sums the total to condense information regarding transcripts with similar function

```
    for blast_tig in enriched_blast["#gene_id"]:
      enriched_Btig = EXPRMTRX_frame[EXPRMTRX_frame[0].str.match(blast_tig, case = False,
na=False)]
```

#creates a frame for the GO matches over timepoints

```
      blast_frame = blast_frame.append(enriched_Btig, ignore_index=True)
```

#Sums up transcripts with similar function

```
    blast_sums =
blast_frame.append(blast_frame.select_dtypes(pd.np.number).sum().rename('Total'))
```

#Step 2.4 For each transcript cluster count how many tigs are in there that match the GOI

```
    for clustidx,cluster in enumerate(os.listdir(path)):
        cluster_tig_counter = [0]
        Cluster_data = pd.read_csv(path + cluster, delimiter="\t")
        tig_counts = Cluster_data.tig.str.contains(blast_tig).sum()
        cluster_tig_counter += tig_counts
        cluster_GOI_counter[clustidx] += int(cluster_tig_counter)

  Cluster_Frame.loc[GOI_frame.loc[index,'GO_DESC']] = cluster_GOI_counter
```

#Step 2.5 Adds the condensed expression data for the GOI into the summary frame

```
  Blast_summary_frame =
Blast_summary_frame.append(blast_frame.select_dtypes(pd.np.number).sum().rename(GOI_frame.
loc[index,'GO_DESC']))
  print ('Blast Results for ' + item + ' ' + GOI_frame.loc[index,'GO_DESC'])
  print (blast_sums)
```

#Step 2.6 Repeats steps 2.3-2.5 for the protein blast annotations (code not shown)

4.4.6.4.3 Step 3 Output of GOI expression

```
Blast_summary_frame.columns = [EXPRMTRX_frame.loc[0,1:]]
Pfam_summary_frame.columns = [EXPRMTRX_frame.loc[0,1:]]

print ('Blast results summmmry')
print (Blast_summary_frame)
print ('Pfam results summmmry')
print (Pfam_summary_frame)
for i in Text_Summary:
  print (i)

Cluster_Frame.to_csv(sys.argv[1]+r'_Cluster_Summary.csv')
Blast_summary_frame.to_csv(sys.argv[1]+r'_Blast_Summary.csv')
Pfam_summary_frame.to_csv(sys.argv[1]+r'_Pfam_Summary.csv')

Text_Summary_File = open(sys.argv[1]+"_Text_Summary.txt","w+")
for i in Text_Summary:
  Text_Summary_File.write(i+ "\n")
Text_Summary_File.close()
```

4.4.6.4.4 Step 4 Automated construction of an encompassing GOI file

#Identify most common GO occurences:

```
Tig2GO_fields = ['#gene_id','gene_ontology_blast','gene_ontology_pfam']
```

#Creates frame from only the desired columns:

```
Tig2GO_frame =
pd.read_csv("trinotate_annotation_report.xls",delimiter="\t",usecols=Tig2GO_fields)
NewGOList = Tig2GO_frame[Tig2GO_frame['gene_ontology_blast'] != '.']
NewGOList = NewGOList['gene_ontology_blast'].str.replace('`','^')
NewGOList = NewGOList.str.split(pat = "^",expand=True)
CountList = NewGOList.apply(pd.Series.value_counts)
```

#Sums up all the rows for each GO term:

```
SumList = CountList.sum(axis=1)
SumList = SumList.sort_values(0, ascending=False)
SumList = SumList.reset_index()
NoGO = SumList[~SumList["index"].str.contains("GO:")]
NoGO = NoGO.iloc[3:]
YeGO = SumList[SumList["index"].str.contains("GO:")]

YeGO  = YeGO.reset_index()
NoGO  = NoGO.reset_index()
del YeGO['level_0']
del YeGO[0]
del NoGO['level_0']

AllGO = pd.concat([YeGO, NoGO], axis=1)
AllGOI = AllGO[AllGO.columns[0:1]]
```

#plots a histogram of GO feature count vs frequency:

```
AllGO.hist(column=AllGO.columns[2],label="lizard",log=True,bins=200)
```

#trims most common go terms:

```
AllGO = AllGO.iloc[6:]
```

#plots new histogram of enriched GO terms:

```
AllGO.hist(column=AllGO.columns[2],label="lizard",log=True,bins=200)
```

#exports GO term name, ID and expression over time:

```
AllGO.to_csv('AllGO_Counts_Summary.csv',sep='\t', index=False, header=False)
```

#exports GO term name, ID and total frequency:

```
AllGOI = AllGO[AllGO.columns[0:1]]
AllGOI.to_csv('AllGOI_Summary.csv',sep='\t', index=False, header=False)
```

4.5 Results

4.5.1 Acclimatisation of microbial community to pot ale

After an approximate week long starvation period, initial community acclimatisation was performed by feeding 50 ml quarter strength (diluted with ddH_2O) pot ale aliquots to the reactors each week over a course of four weeks at either pH 4 (un-buffered) or pH 8 (buffered). Performance of replicate reactors was highly reproducible (Figure 4.3) so a polynomial line of best fit was generated so that calculations regarding methane yields could be made. This conversion of data points into an equation was imperative as the time points when gas buckets were triggered were not identical. Including the starvation phase, a total of 772 and 708 ml of biogas was produced from 25% v/v pH 4 and pH 8 pot ale fed reactors. Of that gas, 214 ml and 176 ml was methane which corresponds to a methane richness of 28% and 25% (Figure 4.3a and b). Gas and methane production which can be directly attributed to pot ale is 427 ml, of which 152 ml is methane, and 376 ml, of which 124 ml was methane, corresponding to methane richness of 36 and 33%, respectively.

As each reactor was fed a total of 3.5 g COD for a theoretical maximum methane yield of 1225 ml the efficiency of conversion for reactors fed diluted pH4 pot ale is 12% and for pH8 10%. Headspace

methane concentration during the 253 hr starvation phase rose from 0 % to approximately 20-25 % in both sets of reactors. Following feeding, gas composition rose to a peak of 60-70% in the space of 50 hr before falling back to a basal level between 10-20% in the following 50 hours (Figure 4.5a and b).

Post-experiment analysis of the reactors performance during the starvation phase revealed a methane richness of 18% and 16% in total gas produced, subsequent experiments performed used the performance in the starvation phase as an indicator of inoculum health, inferring that a high methane yield during starvation indicating a higher concentration of methanogens. Reactors were then re-inoculated with fresh sludge as both the conversion efficiency and methane richness were deemed unsatisfactory. However it was clear that unbuffered pot ale performed better and no adjustment of pH was performed for subsequent experiments.

Before feeding the reactors 50% pot ale v/v aliquots reactor performance was analysed and showed an increase of methanogenic activity when compared to previous inoculum with methane representing 40% of total gas produced during starvation (350 ml gas of which 141 ml methane). In total, 7g of COD was fed to each reactor to produce a total of 1870 ml of biogas with 912 ml of that being methane (49% richness) equating to a COD conversion efficiency of 37% (Figure 4.5c). At this higher concentration of feeding, a biphasic methane production trend with an initial short increase in methane followed by an extended phase was observed in the headspace over a period of 100 hours (Figure 4.5c).

This biphasic production (Figure 4.5d) was even more pronounced when increasing to 100% pot ale v/v feeding aliquots where a total of 14 g of COD was fed to the reactor over the course of the experiment to produce a total of 4330 ml biogas including 2297 ml methane (53% richness and 47% efficiency) over the first four feedings (Figure 4.4d). Two subsequent feedings (blockages were encountered on the first sampling schedule leading to it being repeated) of this community was performed to collect mRNA and VFA samples in order to discern the biological reasons for the biphasic methane production trend and resolve whether the initial activity was caused by hydrogenotrophic or acetoclastic means.

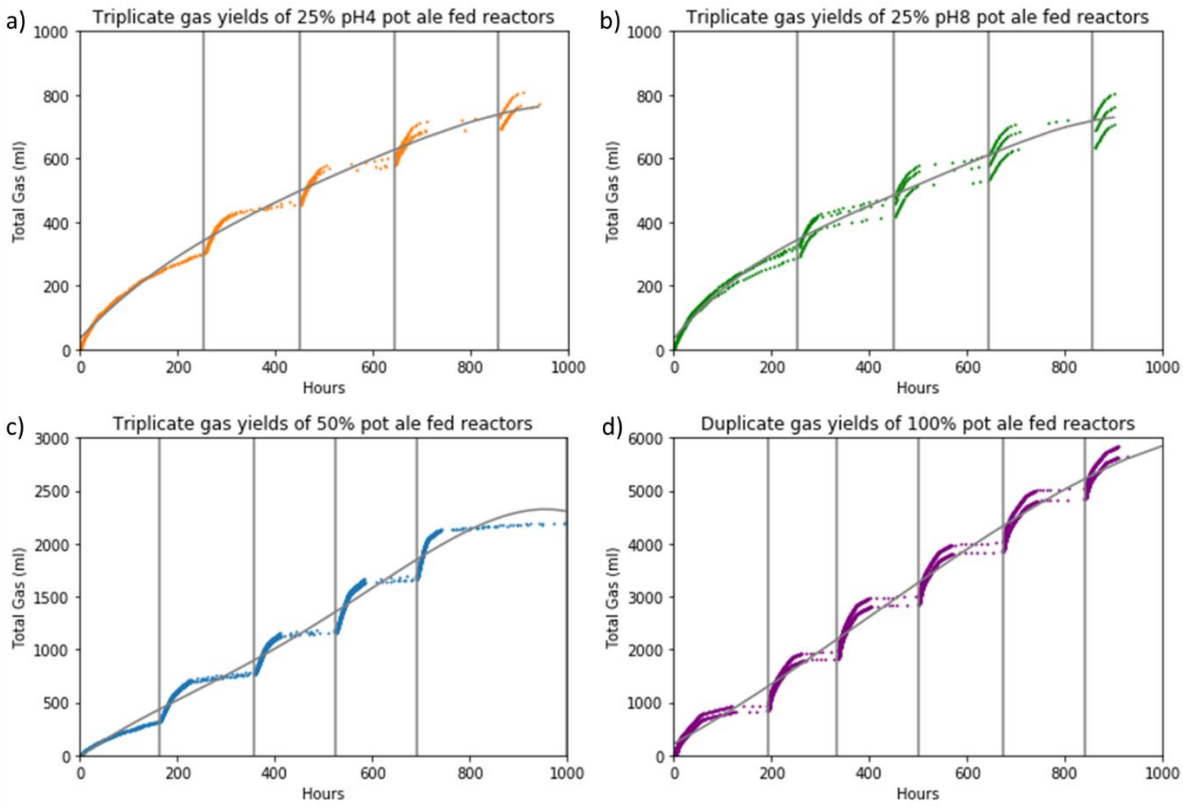


Figure 4.3 Gas yields of pot ale fed reactors at different feed strengths

Overlayed total gas yields for replicate pot ale fed reactors horizontal lines represent feeding points and the trend line is a polynomial line of best fit. Note that pH 4 pot ale was used exclusively following the initial trial at 25% pot ale strength. Each feeding point was a 50 ml aliquot at the respective concentration corresponding to a total organic loading of 7, 14 and 28 g COD / L.

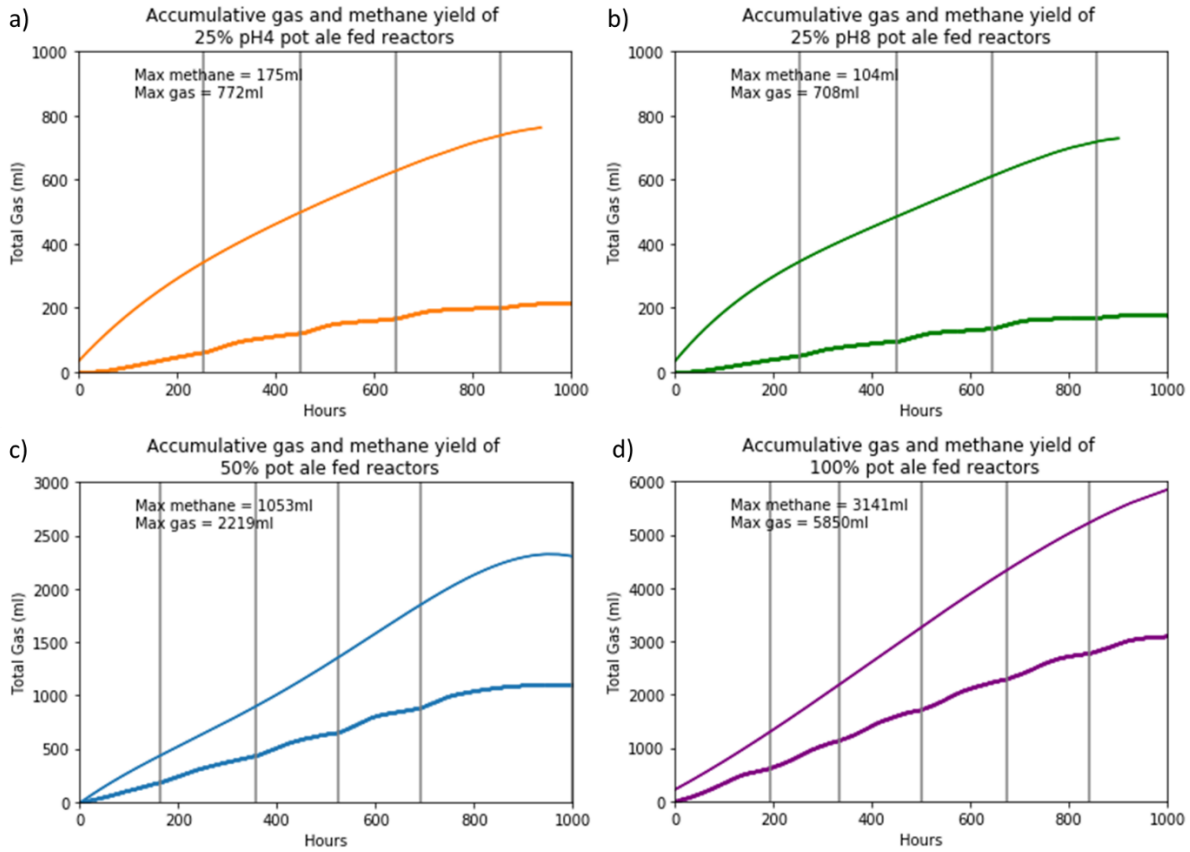


Figure 4.4 Total gas and methane yields of pot ale fed reactors at different feed strengths

Average total gas as derived from the polynomial line of best fit (upper line) and total methane as determined from a moving average across all sensors (lower line). Total gas yield increases as the feeding concentration increases until the highest concentration of feeding yields a gas containing 53% methane.

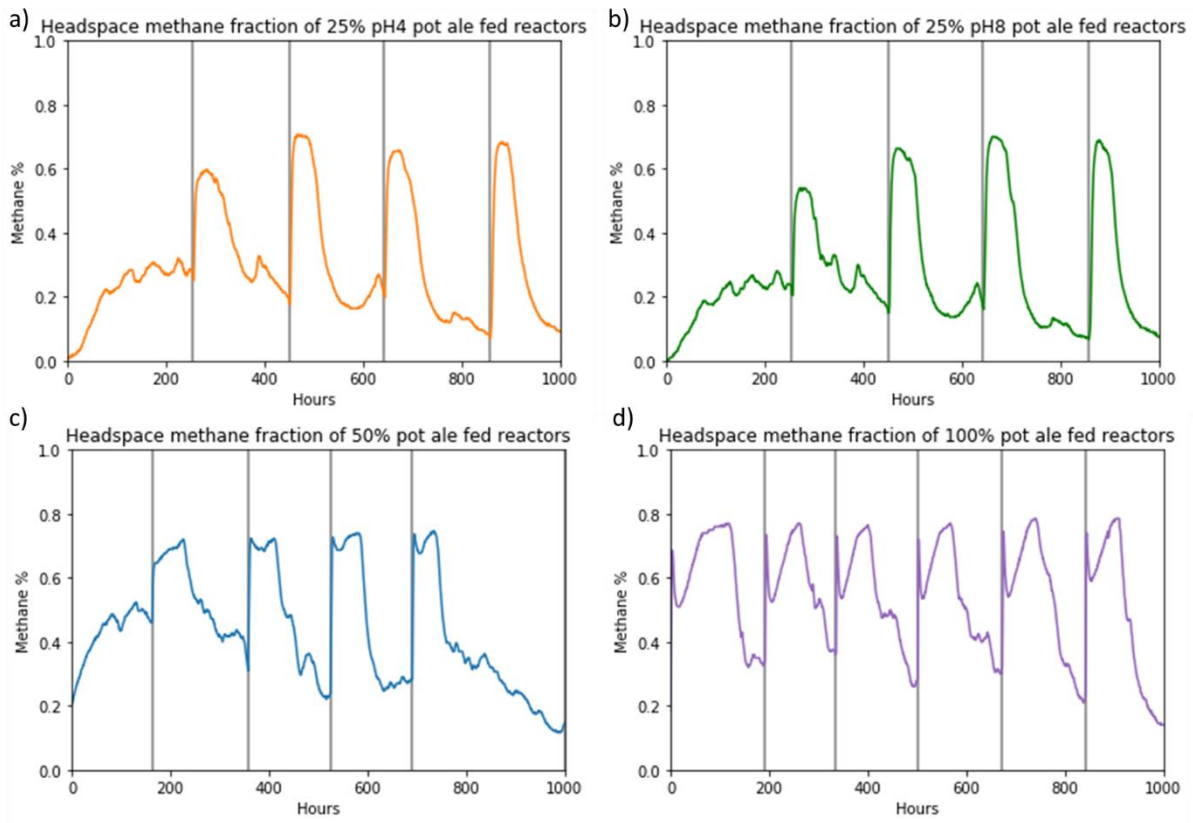


Figure 4.5 Headspace methane concentration of pot ale fed reactors at different feed strengths

Headspace methane concentration as determined from a moving average across triplicate sensors. Methane concentrations typically peaked between 60-70% following feeding and dipped to 40% whilst under starvation.

4.5.2 VFA and COD dynamics of biphasic peak features

Liquid fractions collected from reactor samples as part of the RNA sampling experiment were examined for levels of VFA metabolites, indicating organic breakdown of polymers, and COD (Sampling points are represented by red lines in Figure 4.6). Following feeding both acetic and propanoic acid levels rose sharply in the first four hours from zero to 8.6 ± 0.42 uM and 6.3 ± 0.3 uM respectively. Acetic acid further rose to a high of 11 uM at 12 hrs before gradually decreasing over the timecourse, as did propanoic acid. Butyric acid was present in the reactors between 32 and 52 hours at a concentration between 1.5 and 1.8 uM (Figure 4.7). COD removal reached 49 % by hour 32 and 88 % by hour 96 before showing complete removal by hour 115 (Figure 4.7).

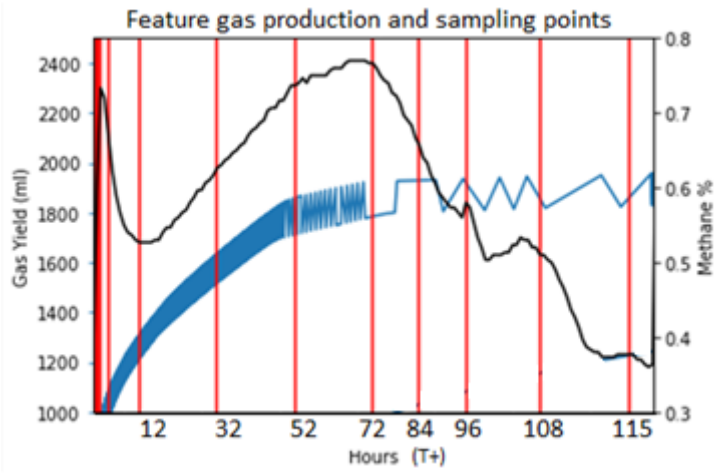


Figure 4.6 Biphasic methane production and RNA sampling points

Immediately following feeding (0 hours) a sharp rise in headspace methane concentration is observed (black line) followed by second more gradual increase. Total gas production (blue line) indicates rapid gas production activity within the first 52 hours. Points over this biphasic gas production were selected for further transcriptional, fatty acid and chemical oxygen demand analysis (vertical red lines).

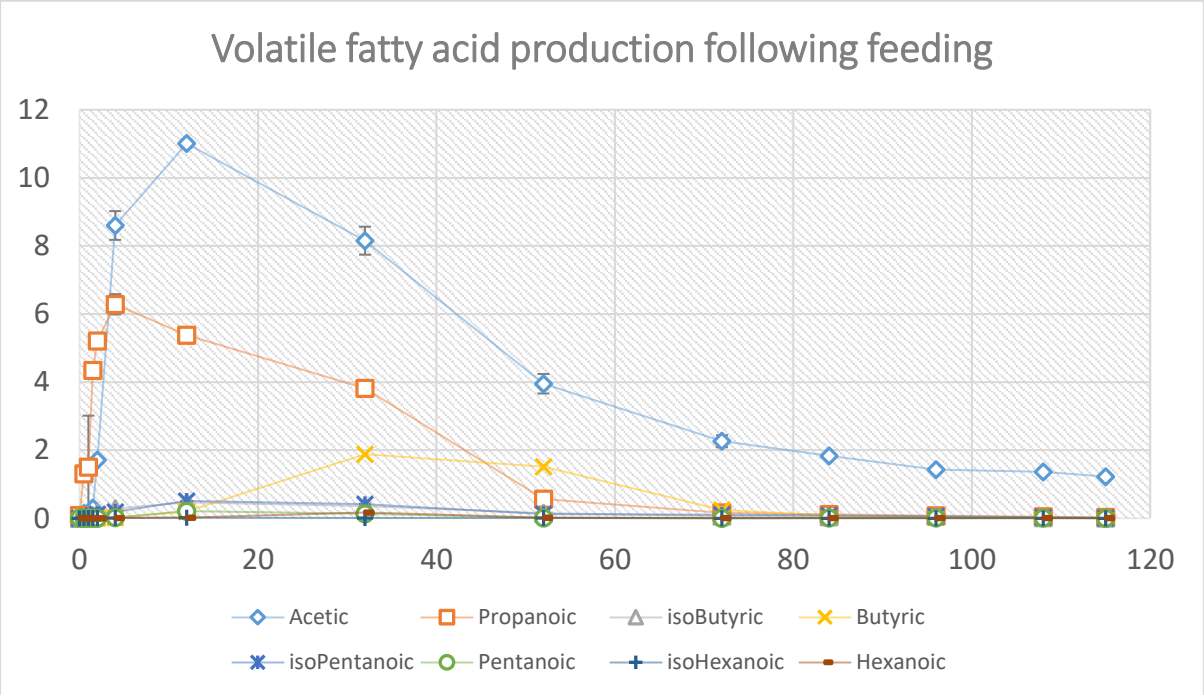


Figure 4.7 Volatile fatty acid concentrations following feeding with pot ale

Concentrations of volatile fatty acids (between 2 and 6 carbons long) in reactors immediately following feeding with a 50 ml aliquot of 100 % pot ale (3.5 g COD). Acetic acid showed the largest increase to over 10 mM 12 hours following feeding. Propanoic showed the second highest degree of change rising to 6 mM before falling back to basal levels after 80 hours.

4.5.3 Metagenomics

4.5.3.1 Short reads

Metagenomic DNA samples were collected from reactors over the course of the acclimitisation. Samples from day 6 and 28 (corresponding to the end of starvation and the middle of the 50% pot ale loading rate phase 2) were collected in addition to samples collected on day 0, 21, 35, 37, 39 and 41 when the reactor was being fed the highest concentration of pot ale (Phase 3). Each sample produced between 7.3 and 11.8 million Illumina short 200bp paired reads with a GC content between 48 and 51% (Supplementary table 4.1). Samples were initially assembled all together using the short read assembler MegaHit to produce a total of 152,869,8 contigs totalling 129,686,886,1 bp with a minimum, average and maximum size of 200 bp, 848 bp and 598867 bp respectively with an N50 of 1058 bp. Individual sample forward and reverse reads were also assembled to produce a number of contigs between 152,455 and 227,045 with a total number of assembled basepairs between 130,100,149 and 200,817,563 and a maximum length between 154,889 and 274,917 bp.

4.5.3.2 Long reads

In addition to short read sequencing, a single MinION flowcell was run using the protocol as detailed previously to collect long sequence reads from a pooled sampled containing day 6 and 28 from phase 2 (50% pot ale) in addition to day 21 and 41 of phase 3 (100% pot ale). This produced a total of 5,888,506 reads (2,834,755 pass and 3,053,751 fail) with an average and maximum length of 3,120.7 and 1,851.8 to 96,836 and 90,903 respectively. For both pass and failed reads the main distribution of GC% and read length varied between 25-75% and between 250-50,000 bases (Supplementary figure 4.2). Passed reads contained more sequences with GC content between 35-40 % and around 60% (Supplementary figure 4.3) indicating an uneven distribution of GC content between pass and failed reads. The relationship between number of reads and read length revealed an exponential decay of counts as read length increase (e.g over 900,000 1kb reads, 400,000 2kb, 250,000 3kb and so on (Supplementary figure 4.4). Nanopore reads were assembled using CANU to produce an unpolished assembly compiled of 12,214 contigs 242,805,063 bp long with an average length of 18,374.8 bases and maximum of 2,356,938 (N50 = 37,039).

4.5.3.3 Assembly Polishing

Raw Fast5 read signals were used to polish (improve areas of low/poor coverage) the long read assembly using Nanopolish and increased the total amount of sequence and longest contig to 245,714,167 bp and 2,384,488 respectively. Short reads were also used to further polish the assembly using PILON which decreased the total length and longest contig to 245,327,127 bp and 2,379,743 bp respectively.

4.5.3.4 Metagenomic Taxonomic Classification

Contigs generated from the short read assembler MegaHit were parsed into the MG-RAST metagenomics analysis server for taxonomic classification at class level. Sequences were classified using the GreenGenes database according to a 70% minimum identify of a 15 bp match and a minimum e value of $1x10^{-5}$ (Supplementary figure 4.5). Unclassified bacterial sequences represented between 30-50% of total contigs per samples. Clostridia was the most abundant class detected with samples composed of between 15-30%. Other consistent classes present included *Actinobacteria*, *Bacterodia* and *Methanobacteria*. Classes appearing intermittently between the start of phase two and end of phase three included Bacilli and *Flavobacteria*.

Comparison between the start of phase 2 (day-6) and the end of phase 3 (day-41) was performed using more encompassing settings (min abundance = 1, %-ID = 60, min length = 15 bp). This showed an increasing abundance of *Methanomicrobia* as the community matured (Supplementary figure 4.6).

Using stricter criteria (Min abundance 100, %-ID = 80, min length = 100) with an alternative reference database RefSeq was able to resolve much of the unidentified sequence (Figure 4.8). Feeding at the lower concentration of 25% v/v pot ale promoted the growth of Actinobacteria (a 10% increase from 20 to 30% relative abundance). However, this was not observed when the feeding was increased to a 50% and 100% v/v solution which reduced the Actinobacteria population to sub 1% relative abundance by P2d21. The higher feeding regime promoted growth of Bacterodia which showed an increase from 5 to 30% relative abundance.

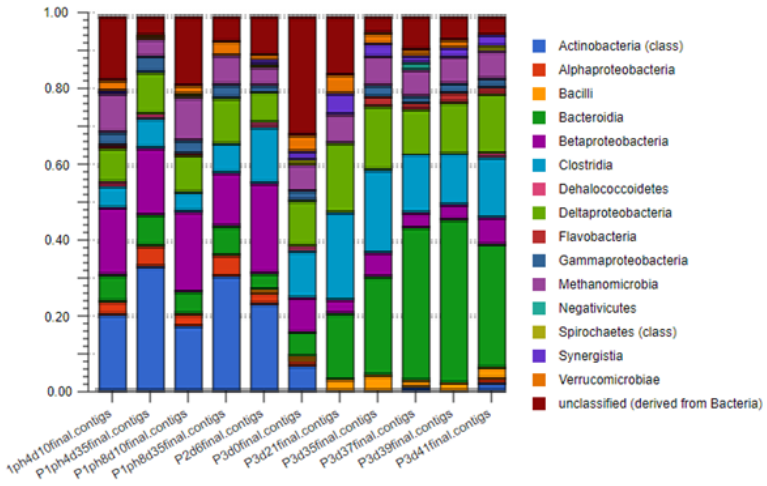


Figure 4.8 Taxonomic classification of short metagenomics read data using the RefSeq database at class level

Reads were assigned to the RefSeq database, a curated database of annotated nucleotide and protein sequences, to yield the closest phylogenetic match. This technique allowed identification of an increasing Bacterodia and Clostridia populations when feeding at the highest loading rate.

4.5.3.5 Metagenomic Gene Functions

Metabolic function was assigned to reads using both the COG database (Clusters of Orthologous Groups) and KEGG Orthology database (Kyoto Encyclopedia of Genes and Genomes) with identical settings used previously for the RefSeq database (Figure 4.9a and b). Unlike the taxonomic classification performed prior which has certain classes appearing and disappearing, assignment of metabolic function seemed to generate more similar and stable profiles between closely linked samples, e.g P3d39 and P3d41. In general, both databases produced similar metagenomics functional profiles with key processes such as carbohydrate, amino acid, transcription and translation being present. No relative abundance increase in carbohydrate associated genes was observed over the experiment, unlike the increase in Bacterodia observed when investigating taxonomy assignment previously.

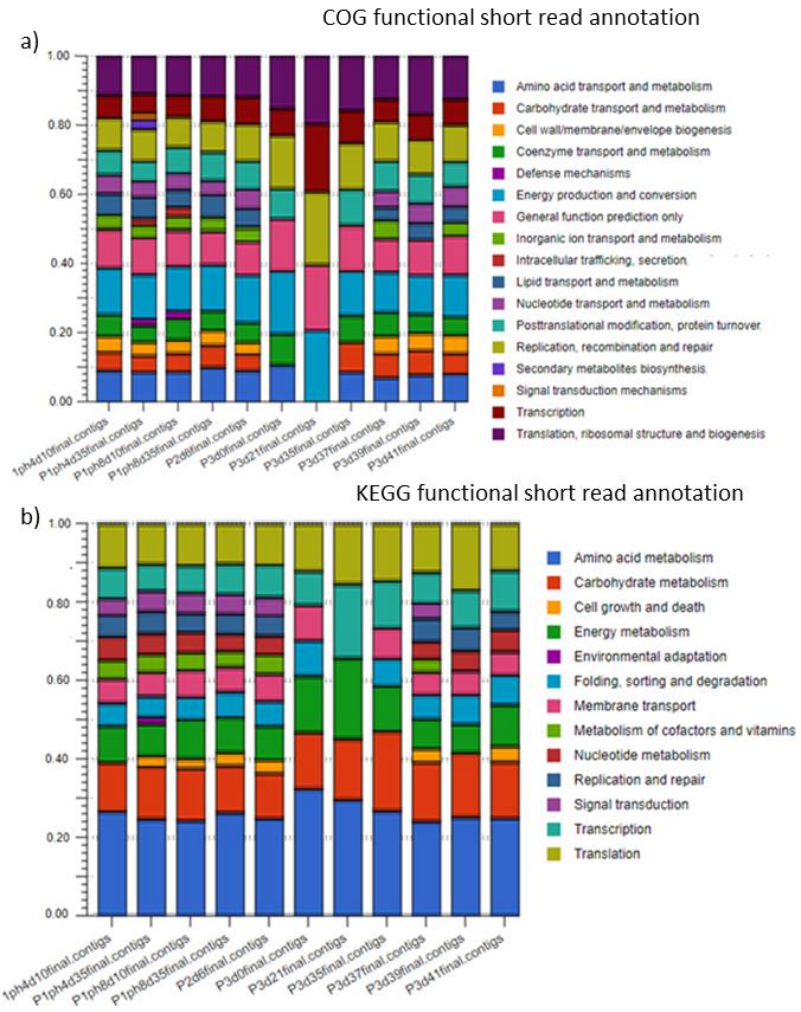


Figure 4.9 Functional annotation of short read data

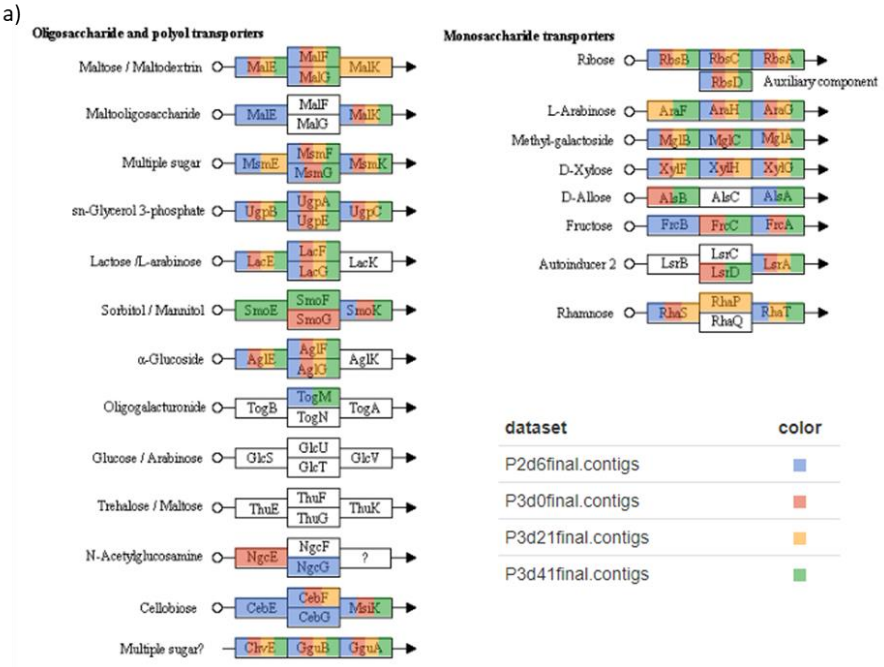
Metagenomic functional classification of short read data using the COG (a) or KEGG (b) databases highlighted an overall stable functional metabolic network. It was hoped that as the community adapted a greater degree of reads would map to carbohydrate metabolism indicating specialization towards the carbohydrate rich feedstock and methanogenesis related pathways.

4.5.3.6 Metagenomic KEGG Pathway Reconstruction

KEGG pathways corresponding to individual samples were constructed and compared to identify shifts in the community towards a certain type of metabolism with the goal that a community exposed to carbohydrates longer will have a more complete KEGG pathway for carbohydrate degradation due to functional enrichment of organisms. A number of complete or mostly complete universal KEGG pathways were constructed including; Glycolysis, TCA, Pentose Phosphate, Purine metabolism, Pyramidine metabolism, metal ion /amino acid transport and Amino acid degradation (Supplementary figures 4.7-12). More pot ale relevant pathways such as oligo/monosaccharide transportation and methane metabolism were also constructed (Figure 4.10a and b). Presence of most oligo and

monosaccharide transporters were confirmed such as fructose, rhamnose, D-xylose, ribose, L-arabinose, maltose and lactose. No temporal shifts in metabolic pathways were observed.

a)



b)

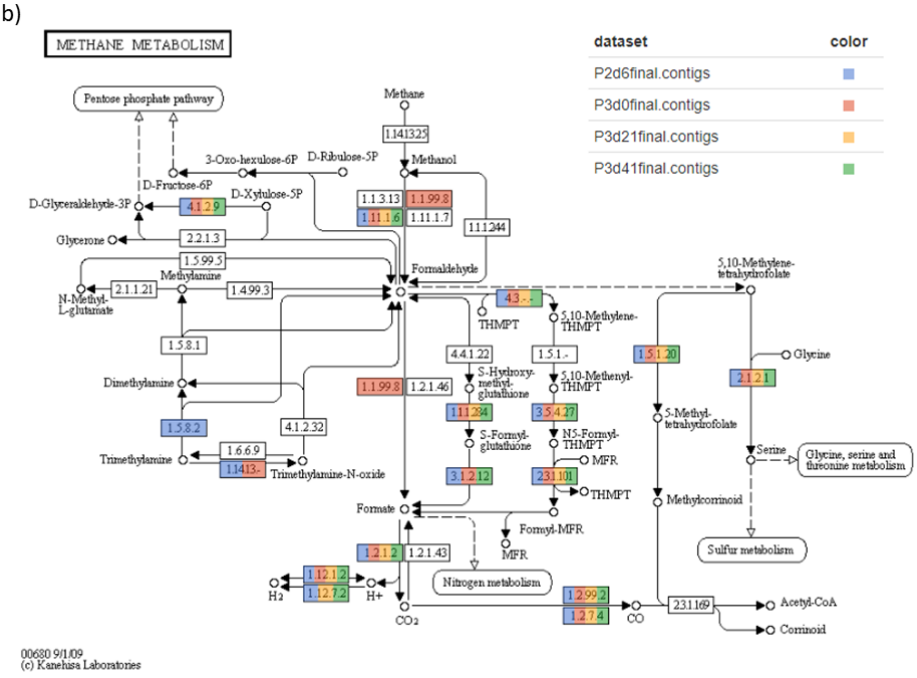


Figure 4.10 Reconstruction of sugar transport and methane metabolism KEGG pathways.

KEGG pathways were constructed using colour coded node-points in relation to timepoint in order to observe possible changes, or specialisation, towards certain pathways or functions. Over the community showed a high degree of functional redundancy towards multiple sugar metabolism.

4.5.3.7 Contig identity

Contigs from the polished long read assembly were blasted against the entire NCBI nucleotide database and the best match (based upon Bit score) was recorded. Contigs were then sorted based upon alignment length if containing 100 or more aligning basepairs. In total there were; 3716 contigs 100<1000 bp, 2207 contigs 1000<2000 bp, 745 contigs 2000<5000 bp and 23 contigs 10000<. Contigs above 10,000 belonged to species of Bacterodia, Methanomicrobia, Anaerolineae, Bacilli, Uncultured crAssphasge, Betaproteobacteria, Clostridia and Alphaproteobacteria. Alignment information, including matching species, %ID and alignment length of contigs >1000 bp is shown below in (Table 4.1).

Query seqid	Subject seqid	Class ID	Species ID	%id	alignment length	E value	Bit score
tig00016541	AP018040.2	Bacterodia	*Petrimonas* sp. IBARAKI	93.35	21307	0	30661
tig00120947	AP018040.2	Bacterodia	*Petrimonas* sp. IBARAKI	92.91	20282	0	28674
tig00006324	LR215978.1	Bacteroidia	*Parabacteroides* distasonis ATCC 8503	83.926	18216	0	17173
tig00012310	CP002565.1	Methanomicrobia	*Methanosaeta concilii*	94.202	17404	0	25957
tig00011841	LT859958.1	Anaerolineae	*Brevefilum fermentans*	96.326	16687	0	27026
tig00027754	AP018040.2	Bacterodia	*Petrimonas* sp. IBARAKI	93.956	16412	0	24157
tig00011566	JF915701.1	Bacilli	*Bhargavaea cecembensis*	97.097	15744	0	26253
tig00012878	BK010471.1	uncultured crAssphage		96.549	14721	0	24142
tig00086037	AP018040.2	Bacterodia	*Petrimonas* sp. IBARAKI	95.244	13813	0	21453
tig00023232	AP018040.2	Bacterodia	*Petrimonas* sp. IBARAKI	94.886	13335	0	20417
tig00020339	CP002565.1	Methanomicrobia	*Methanosaeta concilii*	94.405	13030	0	19632
tig00027363	CP027669.1	Betaproteobacteria	*Betaproteobacteria*	85.055	13008	0	12743
tig00032286	CU466930.1		*Candidatus Cloacimonas acidaminovorans*	93.099	12578	0	18015
tig00140877	AP018040.2	Bacterodia	*Petrimonas* sp. IBARAKI	94.682	12410	0	18849
tig00087976	CP002565.1	Methanomicrobia	*Methanosaeta concilii*	94.008	12217	0	18050
tig00078765	AP018040.2	Bacterodia	*Petrimonas* sp. IBARAKI	95.392	12088	0	18870
tig00097888	AP018040.2	Bacterodia	*Petrimonas* sp. IBARAKI	94.231	12081	0	18039
tig00044693	CP002565.1	Methanomicrobia	*Methanosaeta concilii*	94.785	11467	0	17477
tig00040665	CP003259.1	Clostridia	*Clostridium* sp. BNL1100	93.549	11378	0	16700

tig00016094	CP013244.1	Alphaproteobacteria	*Caulobacteraceae bacterium* OTSz_A_272	76.576	10596	0	5260
tig00019062	AP018040.2	Bacterodia	*Petrimonas* sp. IBARAKI	93.358	10449	0	15032
tig00016054	LT859958.1	Anaerolineae	*Brevefilum fermentans*	92.895	10148	0	14377
tig00764426	CP000562.1	Methanomicrobia	*Methanoculleus marisnigri JR1*	82.106	10132	0	8359

Table 4.1 Identity of the 23 contigs above 1000 bp in length

The identity of the 23 largest contigs were assigned to the closest match and alignment details provided such as % identity. Multiple large contigs belonging to Bacterodia were assembled although the genome was not able to be circularised due to low coverage.

4.5.4 Transcriptomics

4.5.4.1 RNA yields and quality

RNA extracted was quality and concentration checked using a Nanodrop before having the RNA integrity examined using a BioAnalyzer 2100. RNA integrity was quantified by comparing the ratio of the 16S and 23S peak areas to produce a RNA integrity number (RIN) between 0 and 10 where 10 is the highest, an example RIN plot is presented in Supplementary figure 4.13. The RIN across all 14 samples varied between 6.4 and 8.9 and was 8.05 on average with a standard deviation of 0.76. RNA sample concentrations, absorbance ratios and RINs are presented in Supplementary table 4.2 and confirm sample suitability for subsequent library preparation and sequencing.

4.5.4.2 RNA sequencing depth

Gene and isoform enrichment plots were generated to investigate the transcript abundance in identification of new transcriptome features. Both the gene and isoform plot indicate that 90% or more of most abundant transcripts code for a limited number of genes or isoforms. As the least abundant transcripts are increasingly included the number of unique features identified asymptotically approaches infinity which is undesirable for large scale data analysis (Figure 4.11).

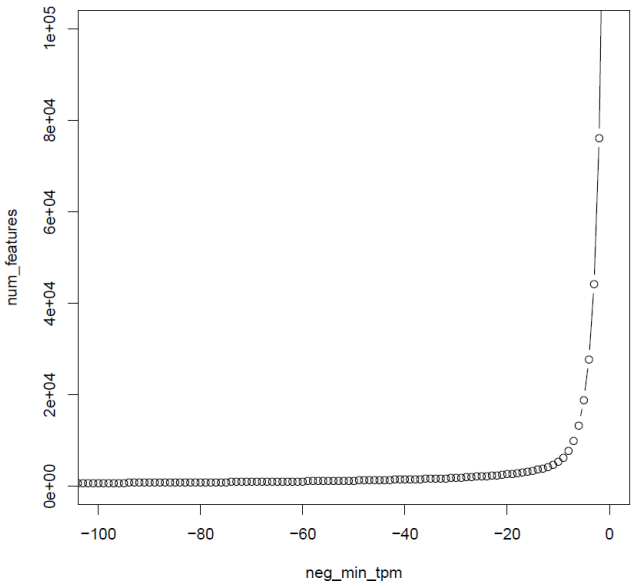


Figure 4.11 Transcript feature frequency and counts

Transcript feature frequency and expression is visualised as transcript abundance (negative transcripts per million) plotted against the number of transcripts which have the respective abundance. From the graph there is approximately 6.8 x 10^4 transcripts which occur only once across the dataset and are likely misassemblies.

4.5.4.3 Differential expression analysis

Transcripts were compared pairwise using the edgeR software package at a dispersion value between 0.1 and 0.4 to identify differentially expressed features based on log fold change and abundance (An example MA plot is shown for comparison of 0 hr and 12 hour timepoints using a dispersion setting of 0.3 below in figure 12).

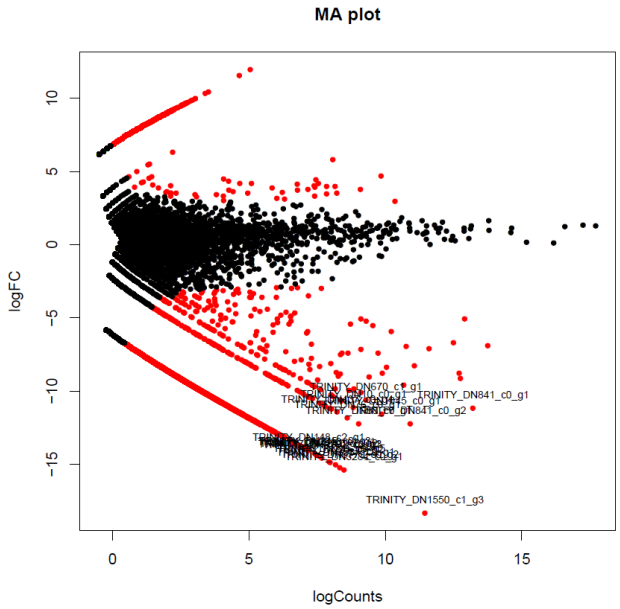


Figure 4.12 Identifying differentially expressed transcripts

The normalised mean of respective transcripts (graph objects) count is plotted against the fold expression change when comparing the 0 to the 12 hour timepoint to identify differentially expressed transcripts. Those deemed to be significant as determined by the dispersion value, or distance from main body of points, are coloured red.

4.5.4.4 Extracting and clustering differentially expressed transcripts

Transcripts were then clustered into similar expression patterns using the top X of DE features (for each respective dispersion setting) per samples as calculated previously edgeR where X varied between 100 and 10,000 (Shown for a dispersion value of 0.3 in Supplementary Figure 4.14). Pairwise sample differences is shown below in Figure 4.13a where the top 1000 DE genes at a dispersion value of 0.3 were compared across all samples. Decreasing the DE depth from 10,000 to 100 showed no significant improvement in cluster clarity (Supplementary Figure 4.14) or computation speed so the more encompassing 10,000 DE clusters were chosen for downstream analysis.

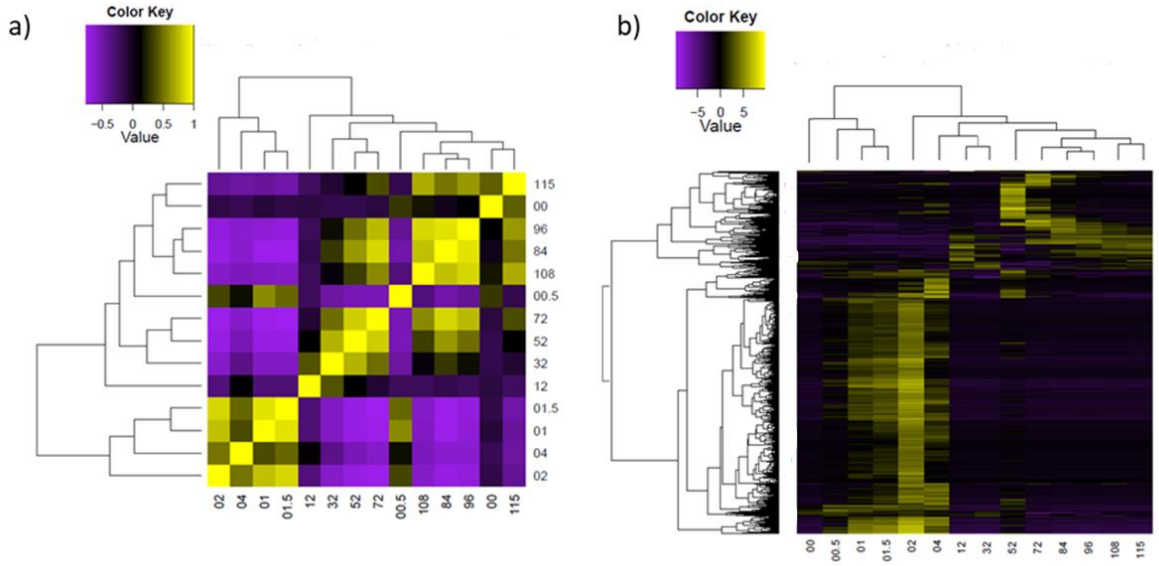


Figure 4.13 Pairwise comparison of sample transcriptomes and gene expression clustering

Pairwise comparison between sample gene expression (a) reveals tight clustering between start and endpoint expression profiles (00 and 115 hours). Visualisation and clustering of 1000 differentially expressed genes over time (b) shows groups of similarly expressed transcripts, possibly showing stages of hydrolysis, acidogenesis, acetogenesis and methanogenesis. Colours represent degree of correlation from low (purple) to high (yellow).

Sample pairwise comparison and DE gene expression profiles were visualised for the remaining datasets subject to either 0.1, 0.2 or 0.4 dispersion (Supplementary figure 4.15). The decision was made to continue with the dataset corresponding to a dispersion value of 0.3 for further analysis and functional annotation as it clustered samples into groups which likely corresponded to the sequential phases of anaerobic digestion; hydrolysis, acidogenesis and methanogenesis (Figure 4.13b). Pairwise comparison of samples at dispersion settings 0.1, 0.2 or 0.4 identified samples as either too similar (belonging to either one of two distinct groups) or too distinct (each sample a distinct entity) when

considering the likely transcriptomic activity of the complex microbiome and so were discarded. Clusters containing groups of co-expressed transcripts were segregated by cutting the tree high at 60% (Coloured bars in Figure 4.13b).

4.5.4.5 Transcript functional annotation

Coding regions in transcripts were identified using TransDecoder and respective Gene Ontology (GO) annotated using BLAST nucleotide and peptide databases in addition to the HMMER protein domain motif database using the TRINOTATE framework. Out of 422,236 transcripts; 599,601 coding regions were found and of those coding regions 154,415 were assigned function exclusively using BLAST, 154,332 were annotated using both BLAST and protein motifs, 16,041 were described exclusively using protein motifs totalling 308,747 transcript coding regions annotated. 120,481 (20% of total) coding regions were unable to be characterised using either database.

4.5.4.6 GO term cluster analysis

Custom software called “GOfys” (Pronounced “go-fish”) was used to analyse relationships (fish out results with GO terms, hence GOfys) between the transcript expression matrix, transcript clusters and transcript:GO annotations using a provided list of GO terms of interest. A list of 11,895 GO terms were identified across all transcripts.

If examining copper transport for instance, four transcripts (counted if their expression > 0) were annotated with the GO term corresponding to copper transport at hr 2 and seven transcripts were annotated with the respective GO term at hr 4 (since 3 new transcripts are being transcribed which previously had expression levels of zero) this then brings the GO term count to 10.

These terms were ranked in order of most frequent to least (Supplementary file GOIcounts.txt). Examples of terms and counts include 1,579 GO counts associated with the tricarboxylic acid cycle, 1253 for oxidative stress, 188 for hydrogenotrophic methanogenesis and 25 for acetoclastic methanogenesis. As GO terms are hierarchical counts, for common umbrella terms such as “GO:0008150 biological process” were excluded from the search as it had over 76,000 counts and slowed the software down. At the other end, 7,111 GO terms were present 10 counts or less (Figure 4.14).

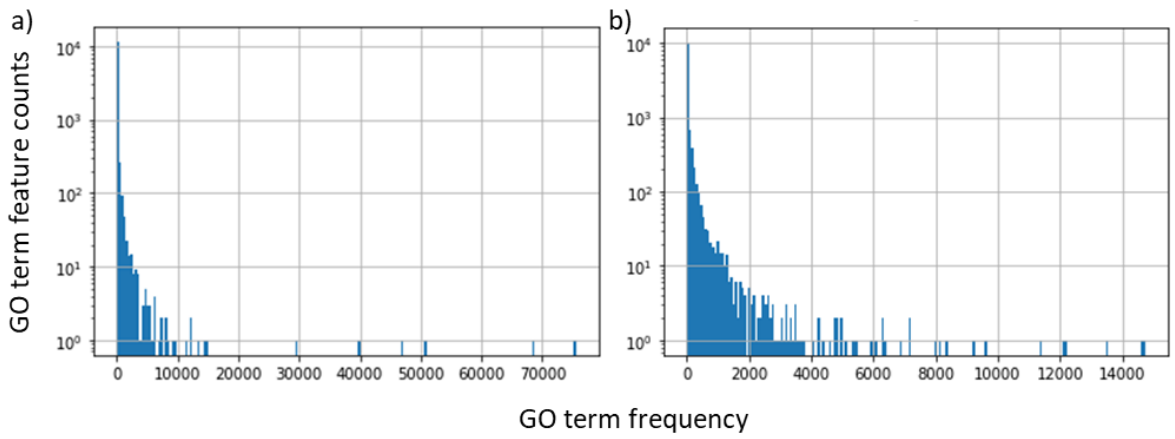


Figure 4.14 GO term feature frequency and counts

GO term feature counts and frequency before (a) and after (b) the 6 most generic terms were removed (such as "biological process").

Multiple GO term lists were provided to GO-fys containing; 1) a set GO:0016740 catalytic activity subterms including daughter terms of GO:0016787 hydrolase activity (such as GO:0016798 hydrolase activity, acting on glycosyl bonds), GO:0016853 isomerase activity, GO:0016874 ligase activity, GO:0016829 lyase activity and GO:0016740 transferase activity. 2) Fatty acid related GO terms and 3) Methanogenesis related GO terms and more general process queries such as 4) Biological processes 5) Cellular carbohydrate processes 6) Metabolic processes.

Unfortunately, no matches were detected in the transcript clusters output from identification of most differentially expressed transcripts, this was potentially due to the initial clustering of the top DE transcripts being too stringent. (The clusters of transcripts represented by coloured bars in Figure 4.13b). GOfys software was then improved to incorporate the entire expression matrix. This improved detection, or flagging, of biological processes and the 10 most abundant (in terms of TPM) were visualised using both raw transcripts per million normalised counts (TPM) and as fold change for the six GO term groups previously mentioned.

4.5.4.7 GO term expression matrix analysis

4.5.4.7.1 Hydrolases

Out of all hydrolytic related categories the most abundant by far that was those related to licheninase enzyme activity which peaked at over 3500 TPM on hour 32 closely followed by hydrolases acting on carbon-nitrogen (but not peptide) bonds (such as in amino acids) that reached approximately 3000 at 12 hrs (Figure 4.15a). Licheninase transcription was also the hydrolase category which showed the largest fold increase (over 400) followed by methyl beta-D-glucosidase 6-phosphate glucohydrolase activity levels which spiked rapidly at two hours following feeding before subsiding below 2-fold increase afterwards (Figure 4.15b).

As methanol is one of the product from methyl beta-D-glucosidase 6-phosphate glucohydrolase, the subsequent methanol dehydrogenase activity was found at extremely low expression levels between 8 and 17 TPM which provides circumstantial evidence of its presence (Supplementary file SmallMol.csv).

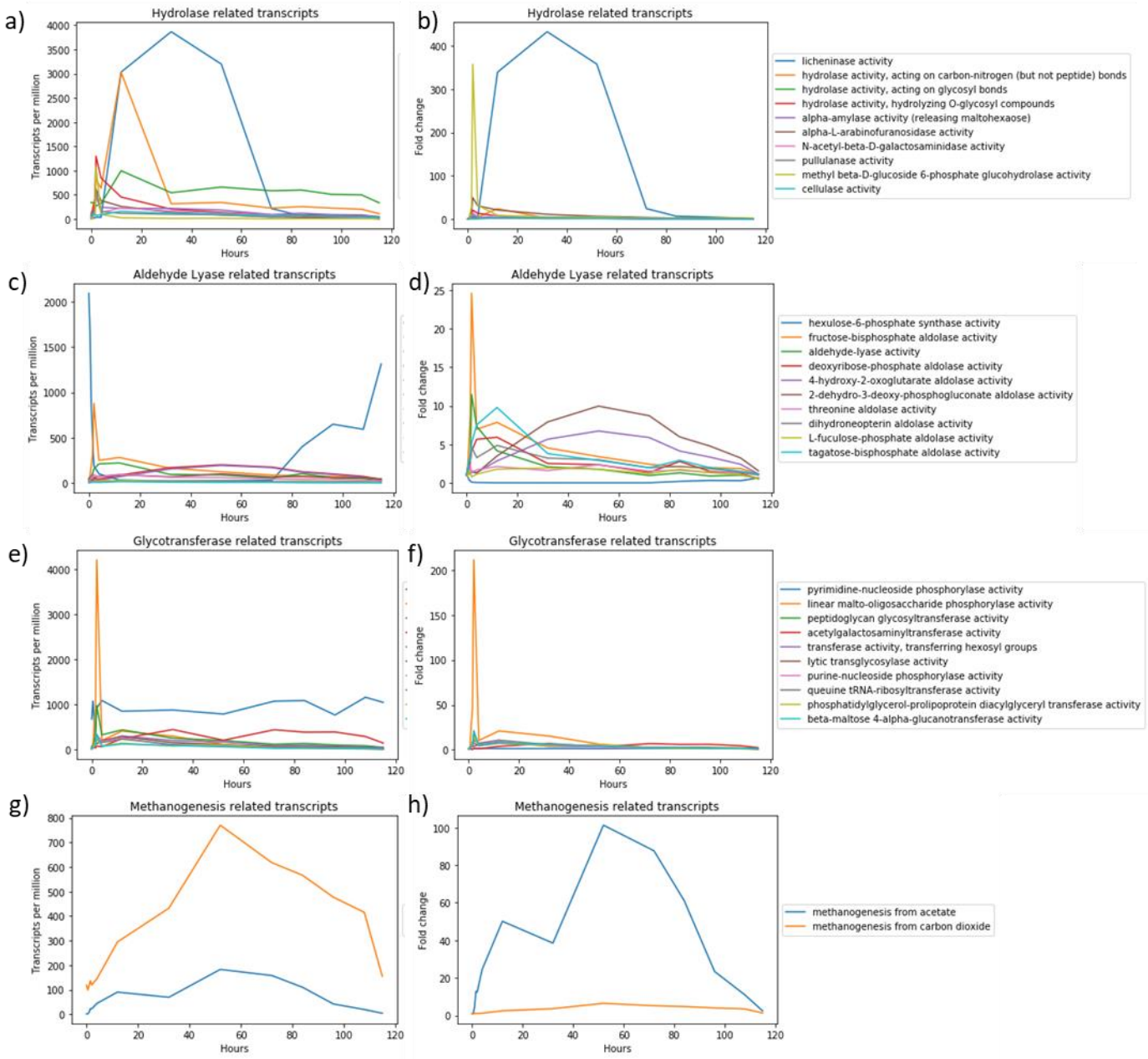


Figure 4.15 Relative and fold change in gene expression of AD hydrolysis and methanogenesis processes

Expression changes within groups of AD relevant GO categories in terms of relative abundance/transcripts per million (left) or fold change (right).

4.5.4.7.2 Lyases

Present throughout the experiment was a high level of carbon-carbon lyase activity that showed expression between 80,000 and 100,000 TPM dwarfing transcription levels of other lyase related transcripts in terms of number (Supplementary figure 4.16a). Following feeding, carbon-nitrogen lyase transcription decreased before slowly returning to basal expression levels towards the end. Fold increase in transcription of other lyases peaked around the 12 hour mark before slowly decreasing towards baseline levels (Supplementary figure 4.16b).

Further investigation into carbon-carbon lyase daughter categories revealed upregulation of aldehyde-lyase related categories. These were further investigated in Figure 4.15c and revealed upregulation of hexulose-6-phosphate synthase in conditions of community inactivity (0 and 115 hr timepoints). Within the first four hours following feeding fructose-bisphosphate aldolase increased sharply up to 878 TPM at 2 hrs (Figure 4.15c and d).

As pyruvate is one of the breakdown products from glycerone phosphate and D-glyceraldehyde-3-phosphate, subsequent pyruvate metabolism was investigated in a similar fashion and identified a isopentenyl diphosphate biosynthetic process increase of 145 to 5407 TPM between 0 and 2 hours. Furthermore, transcripts classified as having glucose-6-phosphate isomerase activity, which increased from 85 to 2028 TPM between 0 and 2 hours (Supplementary file SmallMol.csv), provide a metabolic link from fructose-bisphosphate aldolase to the methyl beta-D-glucosidase 6-phosphate glucohydrolase activity identified previously.

Two lyases, 4-hydroxy-2-oxoglutarate aldolase and 2-dehydro-3-deoxy-phosphogluconate aldolase, showed a more gradual increase in fold expression, similar to that of licheninase expression, which peaked at hour 52 resulting in a 10 and 6 fold increase in expression, respectively (Figure 4.15d). As these lyases both produce pyruvate the expression profile of pyruvate dehydrogenase was investigated and found to increase gradually from 5 to a maximum of 44 TPM at hour 12 before decreasing to a value between 22 and 28 TPM in the hours of 32 to 84 (Supplementary file pyruvate_blast_sumamry.txt).

4.5.4.7.3 Transferases

Transcription of transferases was dominated by glycosyl transferases, however unlike the consistent activity of carbon-carbon lyases, upregulation of glycosyl transferases occurred in direct response to

the feeding stimulus and rose from a basal expression level of 333 TPM to 4588 TPM by hour 2 in a linear manner (Supplementary figure 4.16c and d). Daughter GO terms relating to glycosyl transferases were further investigated which revealed two dominant processes occurring; pyrimidine-nucleoside phosphorylase activity which was stable at approximately 1000 TPM over the duration of the experiment and linear malto-oligosaccharide phosphorylase activity which sharply increased to the two hour time point mark from 20 TPM at 0 hrs to over 4200 TPM (Figure 4.15e). In terms of fold transcription change the malto-oligosaccharide phosphorylase showed over a 200 fold increase between the hours of zero and two (figure 4.15f).

4.5.4.7.4 Fatty acids

GO terms representing fatty acid biosynthetic processes and butyrate metabolic processes were upregulated during the first half of the experimental duration, especially between the zero and two hour time points after which transcripts decreased until returning to almost-baseline levels at 52 hours (Supplementary Figure 4.16e and f). Strangely, negative regulation of fatty acid biosynthetic processes increased over 70 fold between 2-12 hours (Supplementary Figure 4.16f).

4.5.4.7.5 Methanogenesis

Transcripts flagged as having either GO terms corresponding to methanogenesis from acetate or methanogenesis from carbon dioxide were the only methanogenesis pathways detected, methanogenesis from methanol was not detected. Both pathways showed a similar pattern of increasing transcription over the course of the experiment before returning close to baseline levels. Hydrogenotrophic methanogenesis was more upregulated on the whole, in terms of transcripts per million, compared to acetoclastic pathways (770 vs 183 TPM at 52 hrs as seen in Figure 14.15g). However, in terms of fold change, acetoclastic methanogensis saw a larger fold increase of transcription (6 vs 101 at 52 hrs as seen in Figure 14.15h).

4.5.4.7.6 High level processes

High level cellular processes were also investigated to provide an overview of biological processes (Supplementary Figure 4.17). In terms of cellular processes, following feeding expression of protein folding related transcripts and cytolysis response increased to over 50,000 and 10,000 TPM respectively (Supplementary Figure 4.17a and b). One carbon metabolism seemed to dominate cellular metabolic processes in terms of TPM although the process showing greatest fold change was

that of cellular aldehyde metabolic processes (Supplementary Figure 4.17c and d). Overall metabolic processes were dominated by oxidation-reduction reactions (Supplementary Figure 4.17e). Between the hours of 0 and 8 NADP metabolic processes showed the highest fold increase (40) then followed by NADH related processes which peaked at 12 hours showing a 47 fold change (Supplementary Figure 4.17f).

4.5.4.8 Transcriptional activity of classes over time

4.5.4.8.1 Unweighted heatmaps

Taxanomic identity was assigned on the class level to transcripts using CCMetagen. Respective counts of the 11 classes detected were visualised as an unweighted heatmap (Using ClustVis as described in Chapter 1) where it is clear to see when the maximum and minimum activity for each organism exists (Figure 4.16). Clostridia showed exclusive activity in the first 32 hours, which peaked between 2 and 4 hours, confirming their likely role as hydrolytic organisms in the system. Bacteroidia on the other hand were only active between 12 and 52 hours suggesting their role in the intermediary processes such as acidogenesis. Finally Methanomicrobia, later confirmed as the most transcriptionally active methanogen, showed minor activity in the hour following feeding followed by a second round of transcriptional activity between 4 and 115 hours.

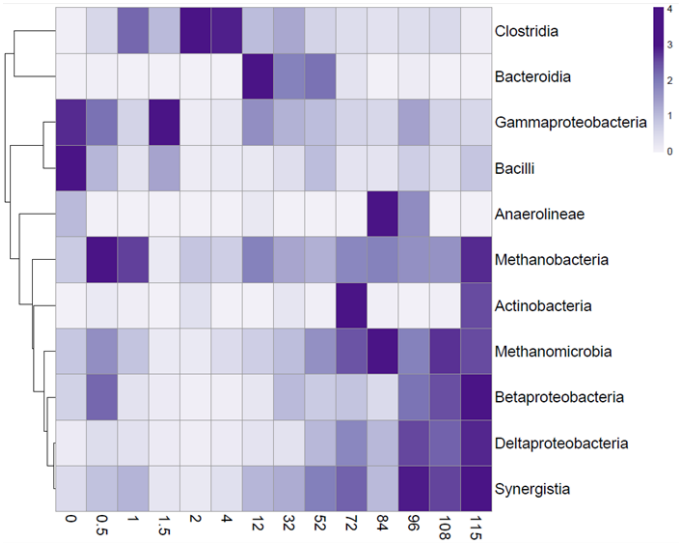


Figure 4.16 Unweighted heatmap of transcription at class level

Unweighted heatmap of classes present in transcriptomic timepoints. Hierarchal clustering shows similar trends in expression of Clostridia and Bacterodia, Gammaproteobacteria and Bacilli pairs in addition to close trends between Methanomicrobia and Synergistia.

4.5.4.8.2 Weighted line graphs

Relative class transcriptional abundance was calculated as a fraction for each timepoint to identify key players and likely phases of anaerobic digestion associated with the respective identified organism e.g hydrolysis and methanogenesis. In the first 4 hours Clostridia produce over 90% of transcripts, most likely indicating a hydrolytic phase, followed by a mixed intermediary period where, in order of abundance, Clostridia, Gammaproteobacteria and Bacterodia are active from 12 hours onwards, possibly signifying the acidogenesis phase (Figure 4.17). From 52 hours and onwards Methanomicrobia are the most active class and perform the final, and most important, stage of anaerobic digestion; methanogensis.

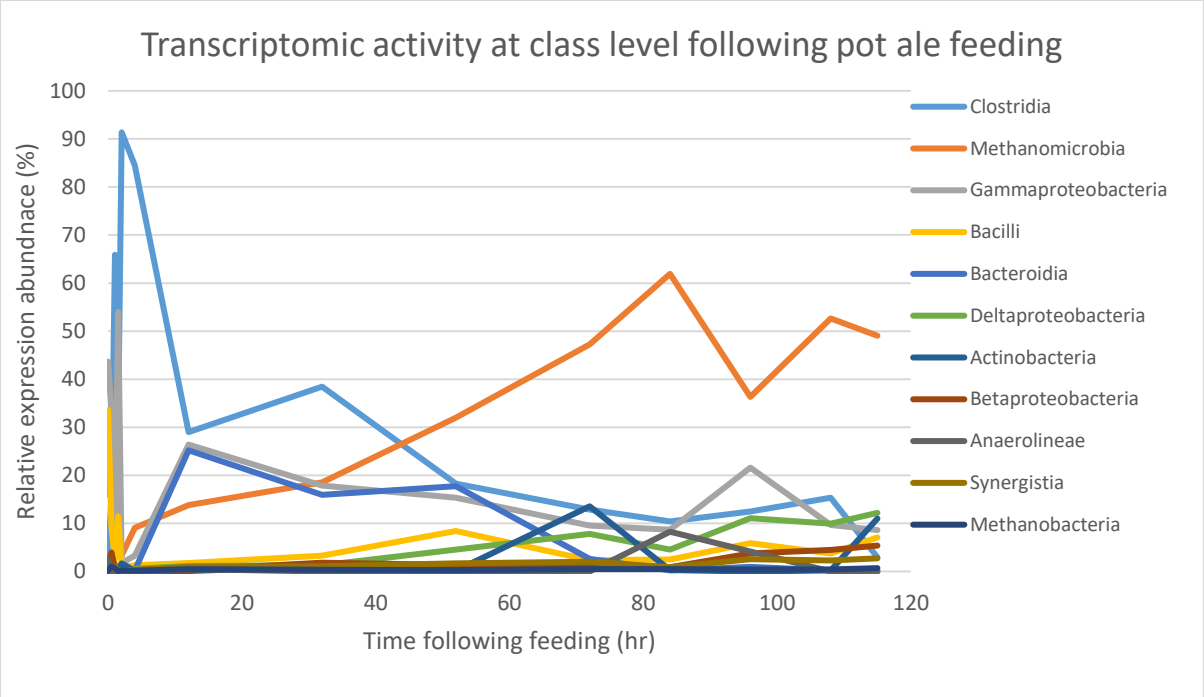


Figure 4.17 Relative abundance of transcription at class level over time

Expression relative abundance of classes following pot ale feeding. Note almost exclusive activity of Clostridia between 2 and 8 hours followed by a gradual increase of Methanomicrobia activity. The transition between these two organisms signifies the end of the hydrolytic phase and beginning of methanogenesis.

4.5.4.8.3 Identification of active vs inactive populations

Identification of active, inactive, abundant and minority populations compared the relative genomic abundance, as determined in the final metagenomics sampling point, to the overall relative

transcriptomic abundance, inclusive of all timepoints. For context, populations high in transcriptomic activity but low in genomic abundance, and vice versa, would indicate an unspecialised and inefficient community. The most genetically abundant and transcriptomically abundant were Clostridia representing 16 % and 31 % of DNA and mRNA total sequence respectively (Figure 4.18a). On the other hand, some organisms played no role in the anaerobic digestion of pot ale such as Spirochaetes, 3.3 % genomic abundance, and Sphingobacteriia, 4.5 % genomic abundance (Figure 4.18b).

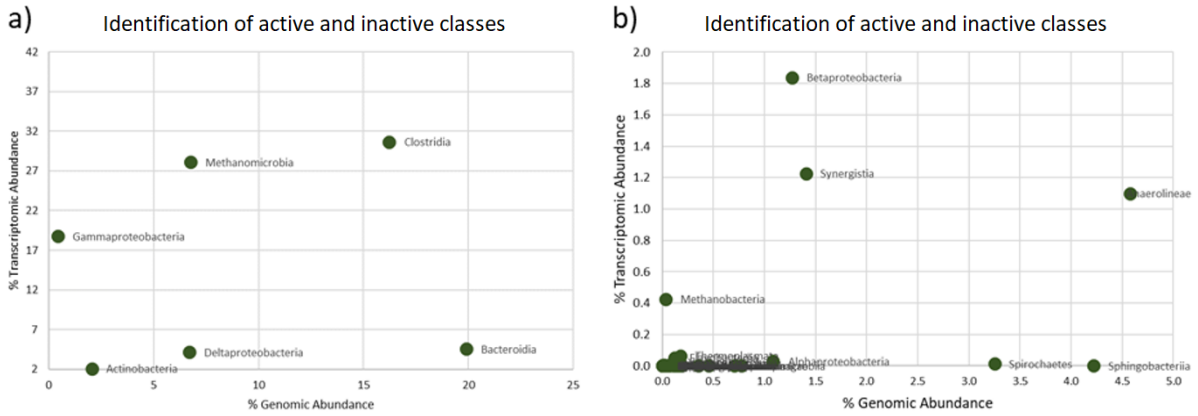


Figure 4.18 Identifying active and inactive populations using metagenomics and metatranscriptomics

Comparison between overall transcriptomic abundance and genomic abundance for transcriptionally highly active organisms (a) and less active (b). Although Gammaproteobacteria is at a very low genetic abundance over the course of the RNA-seq experiment it accounted for a total of ~18% transcripts. The identification of rare but relevant populations is a useful tool for investigation of complex microbial communities.

4.6 Discussion

4.6.1 Reactor gas production and COD removal

The microbial seed gathered from the Naburn wastewater treatment plant was eventually adapted to high efficiency pot ale degradation after a less-than-optimal initial seeding. Before starting experimentation, the inoculum sourced from Naburn on paper sounded ideal due to the common practice of sourcing anaerobic digestion inoculum from wastewater treatment plants rather than other environmental sources like ponds, soil or marshes (Appels et al., 2008; Guo et al., 2015; Jih-Gaw et al., 1997). Additionally, previous composition analysis of the pot ale feedstock revealed a high volatile solids fraction implying a low degree of inorganic recalcitrant material to suggest its suitability as a feedstock. The range of pot ale feeding concentrations chosen for this study is based upon previous successful batch studies which showed successful anaerobic treatment of pot ale between 5-15 g COD/L (Uzal et al., 2003). A higher or more frequent feeding regime could have been used as no reactor instability was observed at the highest feeding concentration and COD removal was 100%. This would have provided more information regarding the performance "ceiling" but may have placed the acclimatised community at risk and jeopardised the ability to perform a subsequent transcriptomic study.

When feeding the reactors, a low yield of between 10-12 % with gas richness between 25-28% was obtained. No major differences between the more acidic unbuffered pH 4 and buffered pH 8 pot ale were found indicating no reason to adjust native feedstock pH, which would occur additional cost on an industrial setting. The poor yields were cause for reactor reseeding, although from the same inoculum source, since typical gas richness of 50 % and above is expected for well-functioning microbial communities (Mueller, 2015; Vanwonterghem et al., 2015; Zhang et al., 2012). It is likely that the poor performance was due to factors such as sludge freshness and concentration rather than the community composition of inoculant sludge since communities in full scale plants are typically stable over a period of years (Carballa et al., 2015; Kirkegaard et al., 2017; Park, 2012; Zumstein, 2000).

Using fresh sludge improved both the conversion efficiencies and methane richness of biogas produced from 37 % efficiency and 49 % richness to 47 % efficiency and 53 % richness when fed with 50 % v/v and 100 % v/v pot ale respectively. Sludge was taken in the morning and reactor set-up complete by the evening of the same day rather than storage overnight at 4°C which likely allowed oxygen to permeate the sludge. Strictly anaerobic organisms, and especially methanogens with some having zero tolerance, are inhibited by oxygen due to their lack of superoxide dismutase genes (Rolfe

et al., 1978; Scott et al., 1983; Zitomer and Shrout, 1998). Oxygen toxicity has only been shown as less inhibitory when the microbial community are spatial organised in sludge granules which prevent oxygen interacting with the methanogen population (Kato et al., 1993).

4.6.2 Identification and importance of bi-phasic methane production

Previously undetected bi-phasic methanogenesis of pot ale was only possible due to the presence of real time methane sensors in the experimental system when feeding at pot ale concentrations of 50 % and above. This bi-phasic methane production had not been observed in previous studies where pot ale was fed either batch-wise, fed as only endpoint gas composition analysis was performed, or continuously (Barrena et al., 2018; Goodwin et al., 2001; Goodwin and Stuart, 1994; Uzal et al., 2003). The presence of bi-phasic methane production, as confirmed in this study, is evidence to support further investigation of the material using a two-stage anaerobic digestion system since it identifies two distinct biological processes. Although multiple-stage anaerobic digestion is commonly performed there is typically no biological justification, such as presented here, to support the decision (Chakrabarti et al., 1999; Kirtane et al., 2010; Mshandete et al., 2004; Yu et al., 2002).

A metatranscriptomic study was performed to determine the root cause of the bi-phasic methane production as two scenarios were possible; 1) initial exclusively acetoclastic methanogenesis from small molecules followed by acetoclastic methanogenesis from larger molecules or 2) initial hydrogenotrophic methanogenesis from small molecules followed by acetoclastic methanogenesis from larger molecules. Although the acetoclastic pathway is regarded as the dominant methanogenesis pathway, the high abundance of Clostrida in both this experiment, detected metagenomically, and in anaerobic digesters in general provided evidence to support the hydrogenotrophic methanogenesis route. Evidence in the literature states that small molecules, such as residual sugars found in pot ale, and even polymers such as cellulose are converted into hydrogen and carbon dioxide by Clostridia which could facilitate hydrogenotrophic methanogenesis pathways (Lin et al., 2007; Zhang et al., 2006; Zhang, 2019).

4.6.3 Metagenomics

4.6.3.1 Microbial community composition

Phylogenetic analysis of contigs generated via short read Illumina sequencing showed presence of microbial Classes *Actinobacteria*, *Bacteroidia*, *Clostridia* and *Methanomicrobia* across most, but not

all samples, when using the GreenGenes 16S rRNA gene database. It is unlikely that an organism, such as Methanomicrobia, is physically absent in intermediary timepoints such as P3d21 and P3d35 as the pre and proceeding samples indicate its presence. This is likely due to insufficient abundance, or misassembly, of 16S rRNA genes in the metagenomics samples resulting in a false negative when querying the GreenGenes database. Additionally, a large portion (30%) of sequence could not accurately be classified to the class level. Unclassified sequence is typically present in most complex microbial community phylogenetic studies, although usually a level between 5-15% (Guo et al., 2015; Li et al., 2013; Y. Verastegui, J. Cheng, K. Engel, D. Kolczynski, S. Mortimer, J. Lavigne, J. Montalibet, T. Romantsov, M. Hall, A B. J. McConkey, D. R. Rose, J. J. Tomashek, B. R. Scott, T. C. Charles, 2014).

Using the alternative RefSeq database, which is a curated and annotated database of genes and proteins, showed a similar but more consistent microbial community structure when compared to using the GreenGenes database. Organisms such as *Actinobacteira*, *Bacterodia*, *Clostridia* and *Methanomicrobia* were present throughout all sample points. By using the more comprehensive database, more sequence was able to be classified to leave on average 12 % of sequence unclassified. This lower level of unclassified sequence is more in line with the previously mentioned studies. Additionally by comparing the pre (P1) and post (P2) re-inoculating microbial communities little differences are seen indicating that the poor initial reactor performance was due to use of "stale" seed and not that of changing community composition. The *Bacterodia* population only proliferated, from 5 to over 40 %, once the reactors started being fed at the highest pot ale level. *Bacterodia* have shown to be positively correlated with carbohydrate, and also proteins, intake and subsequent degradation in gut microbiomes (Bermingham et al., 2017; Hall et al., 2017; Jang et al., 2017; Li et al., 2017; Wu et al., 2019). Organisms detected in the pot ale degrading microbial communities such as; Clostirida, Bacterodia, *Methanomicrobia*, *Synergistia*, *Verrumicrobiaea*, *Alpha- Beta- Delta-* and *Gamma-proteobacteria*, have previously been identified in biogas producing microbial communities degrading feedstocks such as; distillers grain with solubles, food waste, wastewater sludge, avicel (a carbohydrate) amongst others (Hagen et al., 2015; Ho et al., 2013; Manyi-Loh et al., 2013; Montalvo et al., 2012; Park, 2012; Vanwonterghem et al., 2014; Zhu et al., 2019; Ziganshin et al., 2011).

4.6.3.2 Microbial community functionality and metabolic pathways

Functional classification and quantification of metagenomic sample reads using COG and KEGG showed diverse functionality with amino acid and carbohydrate metabolism being most abundant. Relative abundance of carbohydrate metabolism associated genes did not increase over the duration

of the experiment and there was no “specialisation” of metabolic processes as expected or as suggested by the increase in Bacterodia, known carbohydrate processors. Like in this study, carbohydrate and amino acid metabolism are typically the two most abundant functional groups of genes present in environmental metagenomics in both soil, freshwater and AD microbiomes (Cai et al., 2016; Meneghine et al., 2017). The COG and KEGG databases are established resources for metagenomics functional annotation (Hassa et al., 2018; Maus et al., 2016b).

Multiple KEGG pathway were reconstructed using data from time points to identify transitions to or from certain types of metabolism. Although multiple complete or mostly complete universal KEGG pathways were constructed including; Glycolysis, TCA, Pentose Phosphate, Purine metabolism, Pyramidine metabolism and Amino acid degradation, there were no clear shifts to or away certain pathways based upon presence/absence indicators. In fact, some sugars transporters, RhaP, NgcE and MalK, showed presence only in either P3d0 or P3d21 samples and not in pre or proceeding samples. This suggests that there is insufficient sequencing depth in individual timepoints to accurately reconstruct all pathways across all organisms present. A fact not surprising when considering that most soil and sludge samples are extremely diverse and are home to over 100 species (Cai et al., 2016; Campanaro et al., 2016). One of the drawbacks that metagenomics has is that all sequence is analysed, rather than just active microbes or genes relating to feedstock degradation, possibly leading to the most relevant microbes or genes being “drowned out” by genetic noise.

4.6.4 Transcriptional analysis of the bi-phasic peak

Metatranscriptomic analysis of the microbial community in the 115 hours following pot ale feeding revealed a diverse functionality of Hydrolase, Lyase, Transferase and Methanogenesis related processes related to pot ale degradation. Unlike the metagenomics analysis, the metatranscriptomic analysis involved transcriptome contig assembly then remapping of reads back onto the assembly to calculate expression abundance to provide more accurate abundance data. As of now, no time course anaerobic digestion transcriptomics and especially no transcriptomics regarding pot ale degradation have been studied. By using transcriptomics the likely pathways for the bi-phasic methane production peak was determined in addition to identifying crucial taxa.

Two hydrolytic pathways were identified as the main methane producers in anaerobic digestion of pot ale. Firstly the rapid hydrolysis of maltose to butyric and acetic acids by Clostridia which results in a

rapid methane production spike followed by a second, slower, hydrolysis of barley beta-glucans to sugars by Clostridia, Bacteroidia and Gammaproteobacteria.

4.6.4.1 Maltose degradation is likely responsible for initial peak gas production

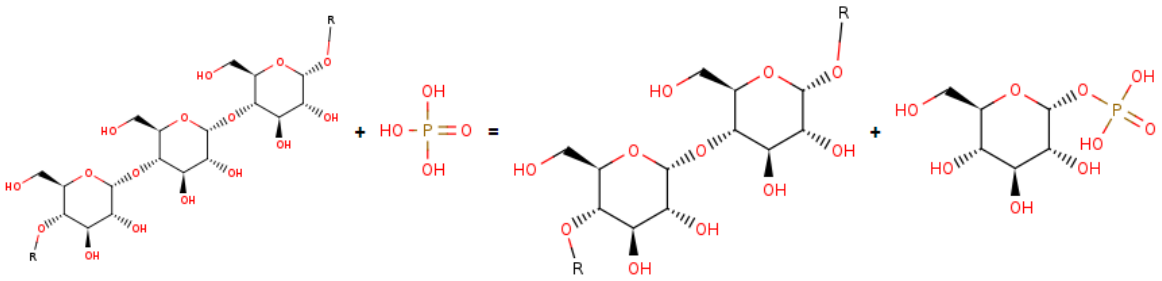
Transcripts performing a molecular function that were highly upregulated, either in terms of raw abundance or fold change, in the first four hours of digestion were considered to have an essential role in creating the first peak in the bi-phasic gas production activity of the community. Barley used for whisky production is typically high in maltose and fructose and although the majority is used in fermentation, up to 2% of residual sugars and 20% dry carbohydrates still remain (GPFeeds, 2015; "Spey Syrup," 2019). As transcripts assigned to the class Clostridia represented over 90% of the total in the first four hours following feeding, it is assumed that all the maltose degradation was performed by this group of organisms. Clostridia are well understood to be responsible for the majority of hydrolysis in anaerobic digestion systems, especially grain based feedstocks (Guo et al., 2015; Maus et al., 2016b; Vanwonterghem et al., 2014; Ziganshin et al., 2011). The enzyme catalysed reaction mechanism proposed, according to supportive transcriptomic evidence, for the degradation of maltose is as follows:

Maltose or methyl beta-D-glucoside 6-phosphate → alpha-D-glucose 1-phosphate → D-fructose 1,6-bisphosphate → glycerone phosphate + D-glyceraldehyde-3-phosphate → pyruvate →organic acids → methane

4.6.4.1.1 Transferase - Linear malto-oligosaccharide phosphorylase

Maltose in barley malts is present in either the traditional disaccharide form or as maltotriose (Yu et al., 2018). Firstly the maltose chain is phosphorylated by a linear malto-oligosaccharide phosphorylase which releases a single alpha-D-glucose 1-phosphate molecule in the reaction scheme below. As transcripts coding for linear maltooligosaccharides reached over 4000 TPM (over a 200 fold increase) within the first two hours, it is likely that this enzyme reaction represents the first step in maltose degradation within pot ale and is responsible for initiating the first methane production peak.

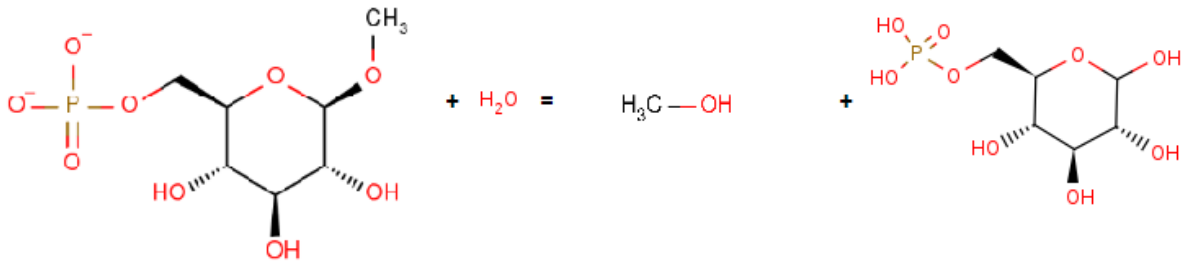
hydrogenphosphate + malto-oligosaccharide = alpha-D-glucose 1-phosphate + malto-oligosaccharide



4.6.4.1.2 Hydrolase - Methyl beta-D-glucoside 6-phosphate glucohydrolase

Dynamics in hydrolase classified transcripts revealed a short and sharp increase of Methyl beta-D-glucoside 6-phosphate glucohydrolase transcription which catalyses the reaction below which suggests that methanol is produced within the first two hours. However, analysis of methanogenesis pathways revealed that methanogenesis from methanol (and methanogenesis from other methyl substances such as methylsulfides, methylhalides and methylamides) was not present and data instead indicated hydrogenotrophic and acetoclastic methanogenesis. Conversion of methanol to acetate has been observed in anaerobic digestion systems and occurs preferentially at higher pH, such as pH 8 as used in this study (Paulo et al., 2003). Indeed this hypothesis correlated with accumulation of acetate between the hours of 0 and 4 with an increase from 0 mM to 8.6 mM. Although presence of methanol dehydrogenase was detected, newly discovered anaerobic methanol dehydrogenases are typically found in methyltrophs and require rare earth metals to function (Picone and Op den Camp, 2019; Vu et al., 2016; Wu et al., 2015). It is also possible that as pot ale was stored aerobically this facilitated aerobic oxidation of methanol when fed although no transcripts were identified relating to its subsequent oxidation product; formaldehyde. Unfortunately, the detection of alcohols was not covered in this experiment although spectroscopic, enzymatic and chromatographic methods of detection exist and could be pursued in the future (Ellis et al., 2019; Mangos and Haas, 1996; Wiśniewska et al., 2015).

methyl beta-D-glucoside 6-phosphate + H2O → beta-D-glucose 6-phosphate + methanol



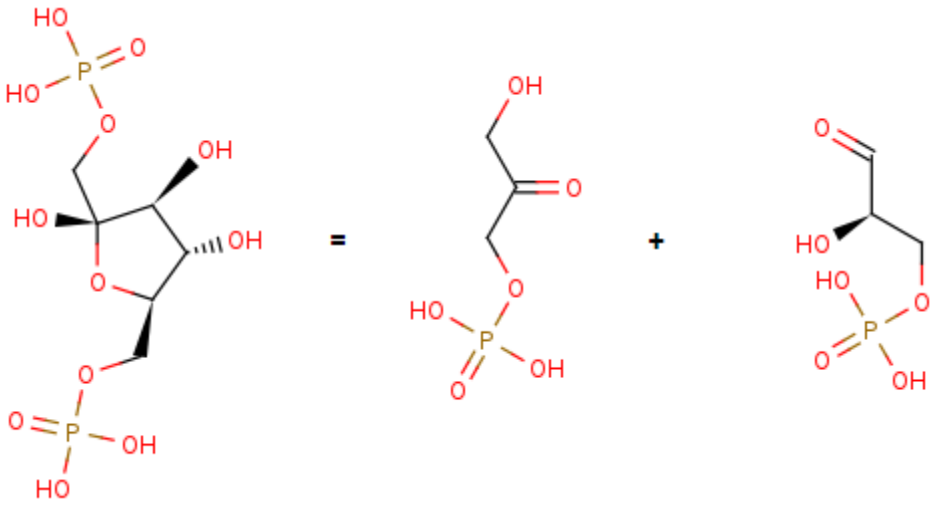
4.6.4.1.3 Lyase – fructose-bisphosphate aldolase

D-fructose 1,6-bisphosphate is first generated via isomerisation by glucose-6-phosphate isomerase activity as confirmed transcriptomically (Supplementary File SmallMol.csv). The resulting D-fructose 1,6-bisphosphate is the used by an aldehyde lyase fructose-bisphosphate aldolase which catalyses the reaction below to cleave the biphosphorylated sugar into two smaller 3 carbon chained products. This corresponded to the accumulation of propanoic acid (a three carbon VFA) between the hours of 0 and 4 increasing from 0 mM to 6.3 mM.This lyase was the most transcriptionally relevant lyase GO class in terms of both relative abundance (TPM) and fold change was transcripts that had been annotated. The phosphorylated glycerone and glyceraldehyde products can be converted into fattys acids such as propanoic and acetate (by Clostridia, which are present) or 2,3-butanediol by Enterobacteria via pyruvate anaerobic oxidation (Biebl et al., 1999; Viana et al., 2012). Thus, the sharp increase of propanoic and acetic fatty acids between the hours of 0 and 4 can partly be attributed to Clostridia, as over 90% of transcripts can be assigned taxonomically during the first 4 hours, and the increased transcription of fructose-bisphosphate aldolase.

Unfortunately, GO term analysis failed to find links between pyruvate and propanoic as described in the literature previously mentioned although a major increase in isopentenyl diphosphate biosynthetic process related transcripts, of 145 to 5407 TPM between 0 and 2 hours, suggests high activity of the isoprenoid biosynthetic methylerythritol 4-phosphate pathway (Bunney, P. E., Zink, A. N., Holm, A. A., Billington, C. J., & Kotz, 2017; Heuston et al., 2012). Recently isoprenoid pathways have been the target of metabolic engineering for the environmentally friendly production of isoprene, a rubber precursor, in addition to Clostridia mediated conversion of fructose and syngas to isoprene (Diner et al., 2018; Li et al., 2018). Further investigation into the fate of pyruvate in the system may

provide an informative basis for integrating the anaerobic digestion of pot ale into an isoprene biorefinery type process.

D-fructose 1,6-bisphosphate = glycerone phosphate + D-glyceraldehyde-3-phosphate



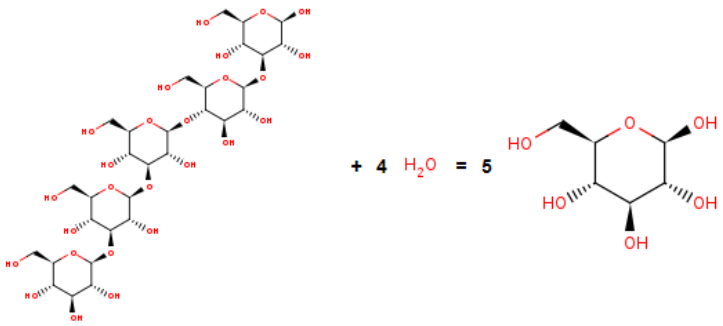
Overall the use of GO terms and reference databases failed to provide as much data regarding fatty acid related processes when compared to hydrolases, lyases and transferases. This unfortunately means the biological cause behind production and degradation of fatty acids, as confirmed via GC-FID, is unknown. This missing link in the GO based pipeline is a critical weakness since fatty acids and related metabolism is key to the anaerobic digestion process as fatty acids are a universal intermediate in the degradation of any biopolymer such as; agricultural matter, animal waste products or sewage. Thus, before further use, a standalone experiment, involving the transcriptomic analysis of a fatty acid fed reactor followed by subsequent annotation and identification of key pathways or transcript families would provide data invaluable to the field of anaerobic digestion as a whole. At the time of writing, no timecourse transcriptomic studies of VFA degradation in AD have been performed.

4.6.4.2 Likely candidates for secondary peak

4.6.4.2.1 Hydrolase- Licheninase

The only hydrolytic related GO term that showed any increase over time, aside from the initial bursts of activity in the first 4 hours, was transcripts corresponding to licheninase activity which remained over 3000 TPM (a 400 fold increase) between 12 and 52 hours. Licheninases, or the systematic (1->3)-(1->4)-beta-D-glucan 4-glucanohydrolases, cleave the β-(1,4)-D-glucosidic linkages in mixed-linkage

glucans containing both (1,3)- and (1,4)-bonds as shown below (Hahn et al., 1995). Mixed linkage beta-glucans are typically found in cereals such as barley and are part of the residual material left over following fermentation (Mangan et al., 2016; Yoo et al., 2007).



Multiple organisms could have attributed to the increased activity of licheninases as although Clostridia activity dominated the first four hours of the experiment, the organisms represent between 30 and 40 % of total transcripts between the 12 and 52 hour timepoints with the rest of the bacterial transcriptome being comprised mainly of Bacterodia, Bacilli and Gammaproteobacteria related transcripts. All of the previously mentioned groups of organisms are known to produce a variety of licheninases (Hahn et al., 1995; Schimming et al., 1991). Further fermentation of glucose molecules into butyric acid is supported by the increased concentration from 0 mM to 1.9 and 1.5 mM on hours 32 and 52 respectively, although the pathways for this are unclear since there are no transcripts were detected related to butyrate production from glucose.

4.6.4.2.2 Lyase - 2-dehydro-3-deoxy-phosphogluconate aldolase/ 4-hydroxy-2-oxoglutarate aldolase

Both 2-dehydro-3-deoxy-phosphogluconate aldolase/ 4-hydroxy-2-oxoglutarate aldolase showed similar expression increase trends to that of licheninases suggesting their linkage. Their co-expression suggests that licheninase 6-carbon (and possibly 5 carbon as suggested 4-hydroxy-2-oxoglutarate aldolase) by products are first converted into 2-dehydro-3-deoxy-6-phospho-D-gluconate (or 4-hydroxy-2-oxoglutarate) through an unknown means before being transformed from the 6 or 5 carbon intermediate into pyruvate. The fate of this pyruvate produced probably ends up in methanogenesis, as it is supported by the increase in methane concentration, however expression analysis only showed a maximum pyruvate dehydrogenase (acetyl transferring) activity less than 30 TPM.

4.7 Conclusion

Pot ale was anaerobically digested at an organic batch loading rate as high as 7 g COD / L within the reactor and showed a highly successful COD removal of over 95% and conversion into a biogas containing over 50% methane highlighting its excellent properties for long term anaerobic digestion, especially in nitrogen limited systems. Real time monitoring of biogas richness identified the presence of bi-phasic methane production which initially suggested hydrogenotrophic methane production, from small molecule hydrolysis and subsequent hydrogen production by *Clostridia*, followed by acetoclastic methanogenesis from biopolymers. Metagenomic analysis showed a relatively stable community structure over time comprised of mainly *Clostridia*, *Bacterodia*, *Deltaproteobacteria* and *Methanomicrobia*. Further transcriptomic identified that over 90% of total transcriptional activity could be attributed to four classes (*Clostridia*, *Bacterodia*, *Gammaproteobacteria* and *Methanomicrobia*) highlighting their key role in the anaerobic digestion of pot ale. Functional annotation of transcripts disproved the hydrogenotrophic hypothesis but provided evidence to support an initial burst of methane activity caused by malto-oligosaccharide hydrolysis, by Clostridia, and acetoclastic methanogenesis followed by a slower, rate limiting, liberation and subsequent degradation of glucose from mixed linkage beta-glucans by licheninase enzyme. Furthermore, the transcriptomic identification of the rate limiting enzymatic hydrolysis step, liberation of glucose from beta-glucans by licheninase, is not just useful for pot ale anaerobic digestion but the methodology used may be also be applied in the field of anaerobic digestion as a whole and opens the door towards feedstock specific enzyme addition.

The newly built lab scale AD system showed improvement to the modified BMP tests performed previously as real time methane sensors and gas volume analysis identified degradation kinetics and in this case a bi-phasic methane production process from pot ale. A larger scale system ensured that the microbial community could be fed and sampled over many months with minimal community disturbance which resulted in increased methane conversion efficiencies from 25 to 47 % as the community adapted.

Using a multi ‘omics approach on an adapting community revealed relative abundance, and increases of some, of taxa in the sludge community structure, as determined by metagenomics, and allowed identification of active and inactive populations by comparing relative genomic abundance with relative transcriptional activity abundance, over 90% of transcriptional activity could be assigned to a combination of Clostridia, Bacterodia and Methanomicrobia highlighting the possibility of a simple

synthetic community for future studies. Although some assembled contigs showed matching regions of over 20,000 bp, whole genomes of the two most abundant organisms, Clostridia and Bacterodia, were unable to be assembled due to either insufficient long read depth or assembly strategy. Additional long read sequencing runs and contig binning, such as using k-mer frequencies, would improve metagenomics assembly and provide an accurate reference metagenome which may additionally aid transcriptome assembly (Albertsen et al., 2013; Alneberg et al., 2014; Campanaro et al., 2016; Wang et al., 2015).

Correlating gas data, VFA data and transcriptomics allowed identification of key pathways and enzymatic reactions involved in anaerobic digestion of pot ale. Using a "top down" approach such as condensing transcript information into GO categories allowed rapid investigation of biological processes of interest which facilitated identification of key hydrolysis and methanogenesis steps. However, the linkage between the various hydrolytic processes and fatty acid degradation is not clear and further work needs to be done in either database curation or transcript annotation. Although community transcriptional responses to oleate (an 8 carbon fatty acid) have been investigated in an AD system, analysis was confined to pre and post feeding expression levels, rather than a timecourse investigation or testing response to specific fatty acids which could be identified by sequential fatty acid feeding of an increasing carbon length (Treu et al., 2016).

Supplementing the transcriptomics with metabolomics may have yielded valuable insights into how the fatty acids are formed. A metabolomics dataset was collected but unfortunately there was insufficient time to perform analysis on this rich and complex dataset.

Finally, the next step in the investigation of pot ale anaerobic digestion would be to purify licheninase enzyme and supplement it when feeding to examine changes in methane production kinetics and determine if the cost is economically feasible on a larger scale. However, the pipeline presented here may also be used to determine the rate limiting enzymatic hydrolysis steps in a variety of feedstocks and may be better suited to analysis of material considered difficult to degrade.

4.8 Acknowledgements

I would like to thank Dr Lesley Gilbert from the Genomics facility at the University of York for performing library preparation on short read DNA and RNA sequences and Dr Sally James for the MinION run.

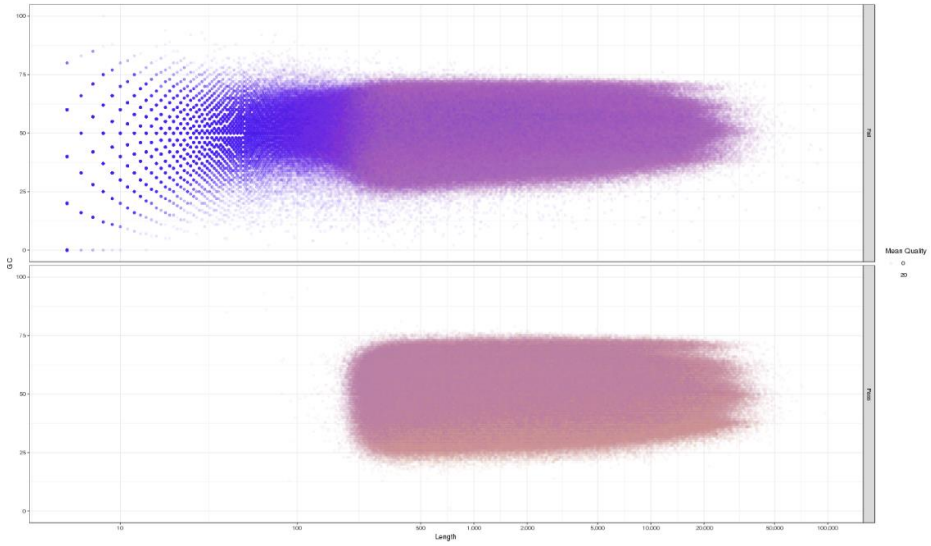
4.9 Supporting Information

Sample Name	% Dups	% GC	Sequences (Million)
P3d41_S153_L008_R2_001	10.40%	49%	11.8
P3d41_S153_L008_R1_001	11.50%	49%	11.8
P3d39_S152_L008_R2_001	9.30%	48%	9.6
P3d39_S152_L008_R1_001	10.40%	49%	9.6
P3d37_S151_L008_R2_001	10.30%	48%	11.4
P3d37_S151_L008_R1_001	11.30%	48%	11.4
P3d35_S150_L008_R2_001	9.00%	50%	8.5
P3d35_S150_L008_R1_001	9.90%	50%	8.5
P3d21_S149_L008_R2_001	8.20%	50%	7.7
P3d21_S149_L008_R1_001	9.10%	50%	7.7
P3d0_S148_L008_R2_001	8.00%	50%	9.2
P3d0_S148_L008_R1_001	8.70%	50%	9.2
P2d6_S146_L008_R2_001	7.10%	51%	8.9
P2d6_S146_L008_R1_001	7.80%	51%	8.9
P2d28_S147_L008_R2_001	8.40%	50%	10.3
P2d28_S147_L008_R1_001	9.20%	50%	10.3
P1ph8d35_S145_L008_R2_001	7.40%	51%	7.3
P1ph8d35_S145_L008_R1_001	8.10%	51%	7.3
P1ph8d10_S144_L008_R2_001	8.10%	49%	10.4
P1ph8d10_S144_L008_R1_001	8.90%	49%	10.4
P1ph4d35_S143_L008_R2_001	8.70%	51%	11
P1ph4d35_S143_L008_R1_001	9.70%	51%	11
P1ph4d10_S142_L008_R2_001	7.30%	50%	9.3
P1ph4d10_S142_L008_R1_001	8.10%	50%	9.3

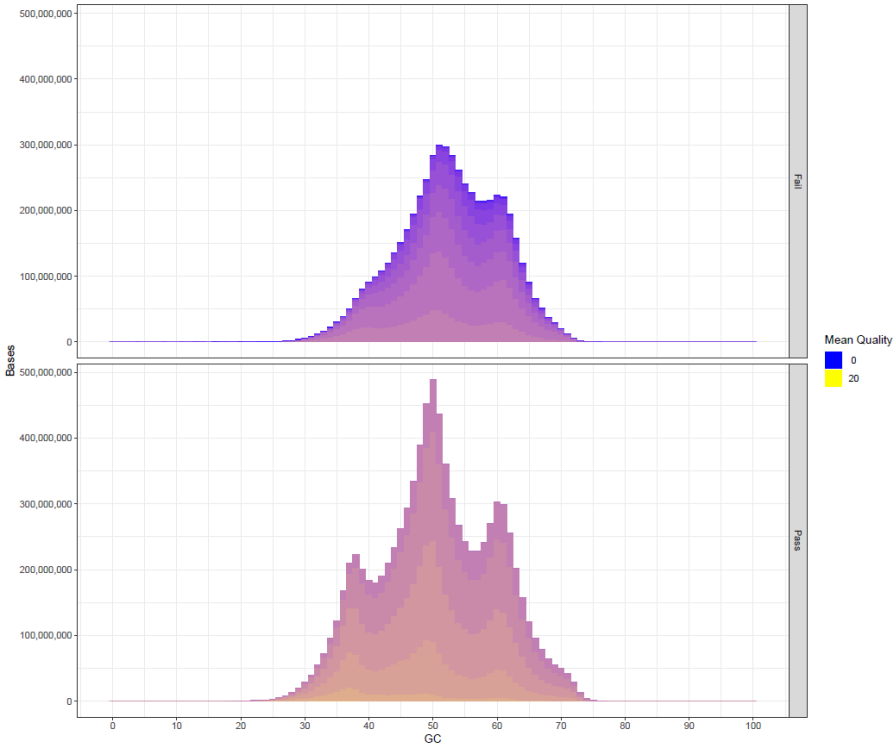
Supplementary Table 4.1. Sample sequencing read statistics.

Supplementary Figure 4.1. Positional quality scores of short reads.

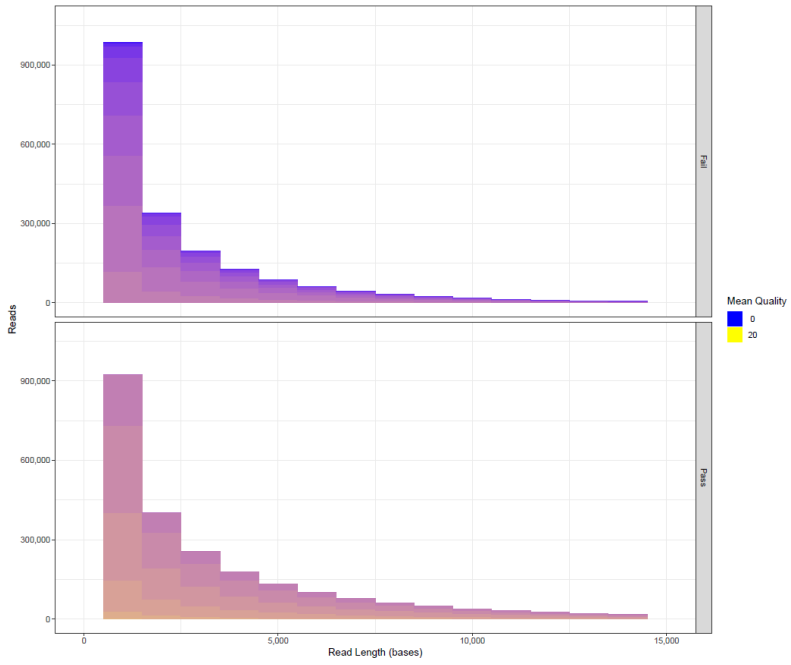
Decreased read quality was observed at the ends of each read so was trimmed accordingly.



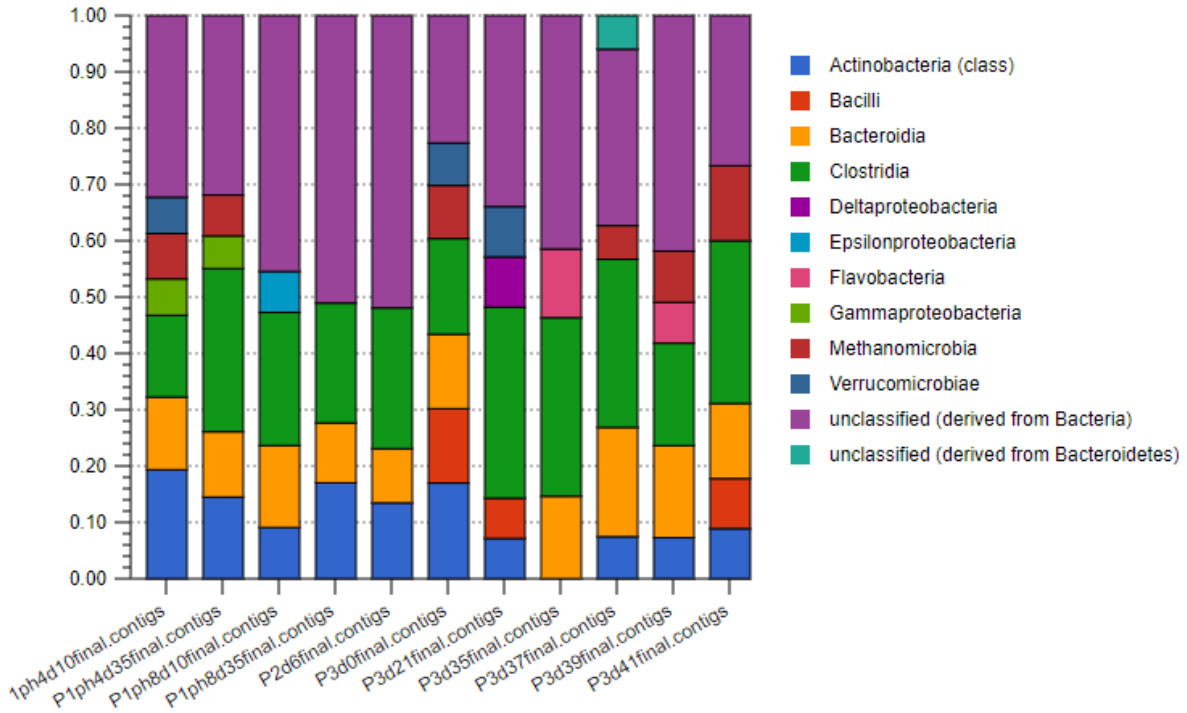
Supplementary Figure 4.2. Positional quality scores of long reads overall (top) and following low quality sequence removal (bottom).



Supplementary Figure 4.3. GC content, frequency and mean quality of pass (bottom) and failed (top) long reads.

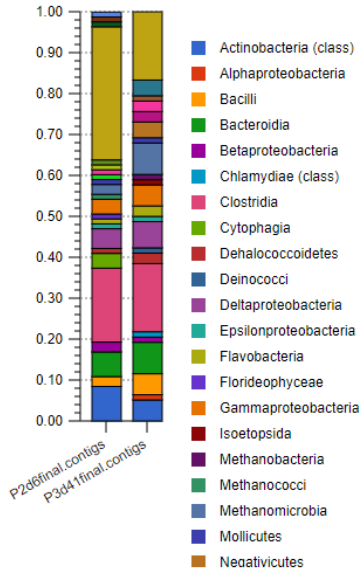


Supplementary Figure 4.4. Read length, quality, and frequency of long reads.

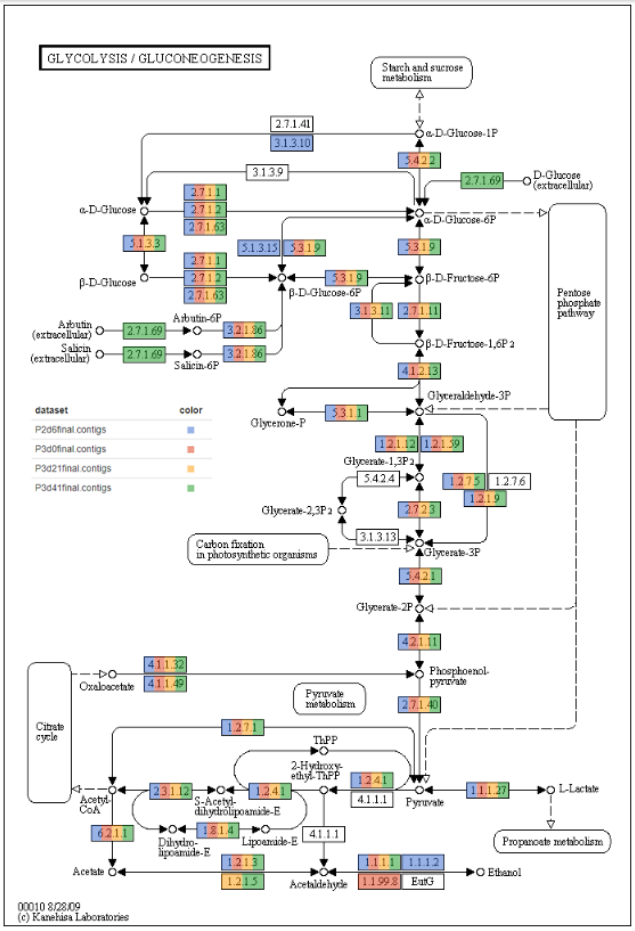


Supplementary Figure 4.5. Taxanomic assignment of metegenomic DNA samples according to the GreenGenes 16S rRNA database.

Although reads were assigned to the database the limitation of the database is that is hosts 16S rRNA gene sequences eexclusively and is not able to assign identity based upon the majority of metagenomic data.

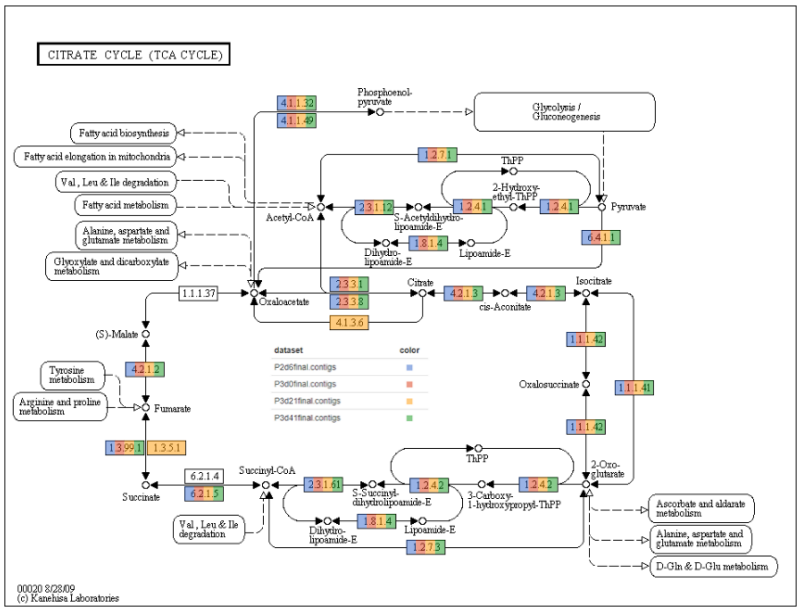


Supplementary Figure 4.6. Taxanomic comparison of the start of phase 2 and the end of phase 3 shows an increase in Methanomicrobia species according to the GreenGenes 16S rRNA database.



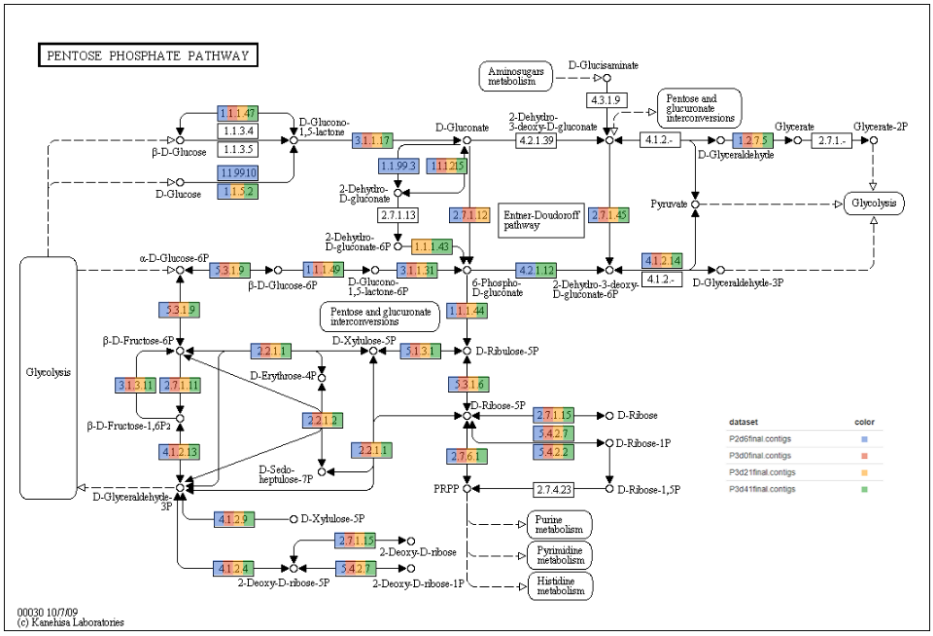
Supplementary Figure 4.7. Glycolysis KEGG pathway

Occurrence of genes related to the glycolysis pathway were mapped out which showed overall the presence of most genes at all timepoints.



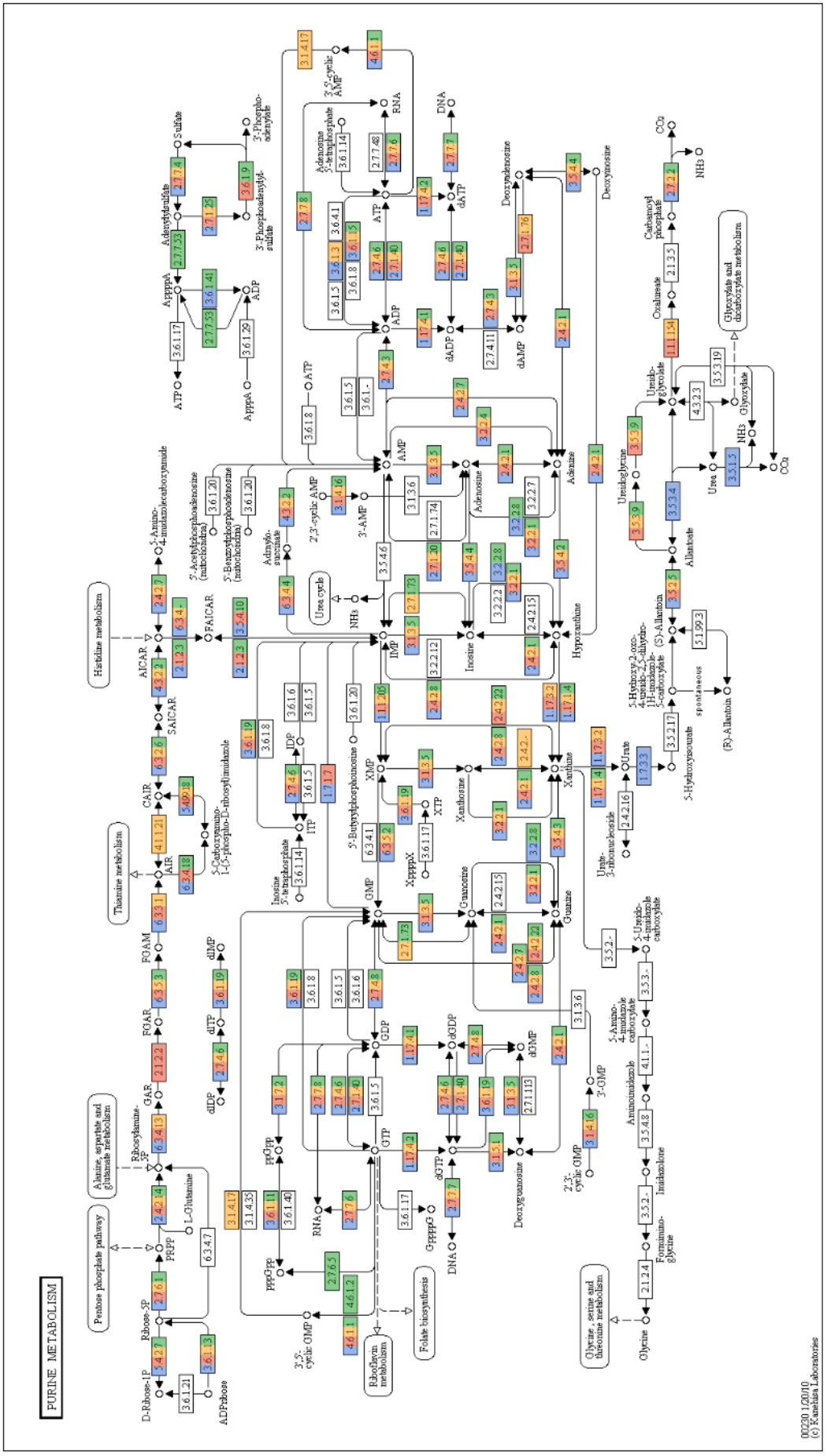
Supplementary Figure 4.8. Citric acid cycle KEGG pathway.

Over the course of the 6 month experiment no major specialisation of genes relating to citric acid cycle metabolism were detected.

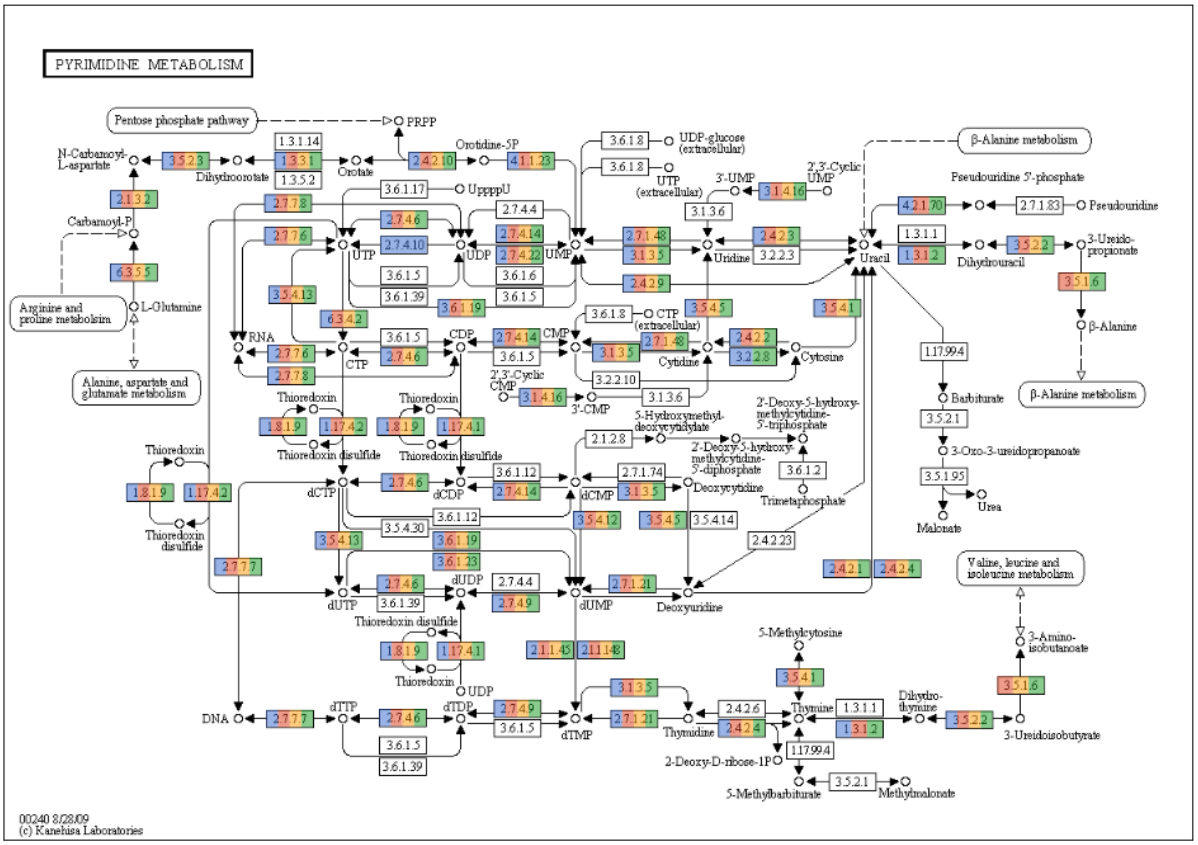


Supplementary Figure 4.9. Pentose phosphate KEGG pathway

Over the course of the 6 month experiment no major specialisation of genes relating to citric acid cycle metabolism were detected.

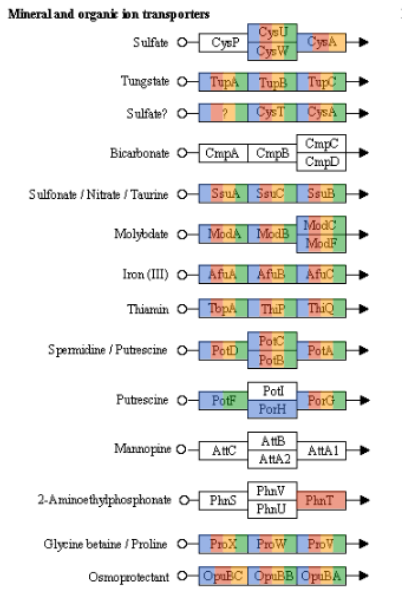


Supplementary Figure 4.10. Purine metabolism KEGG pathway



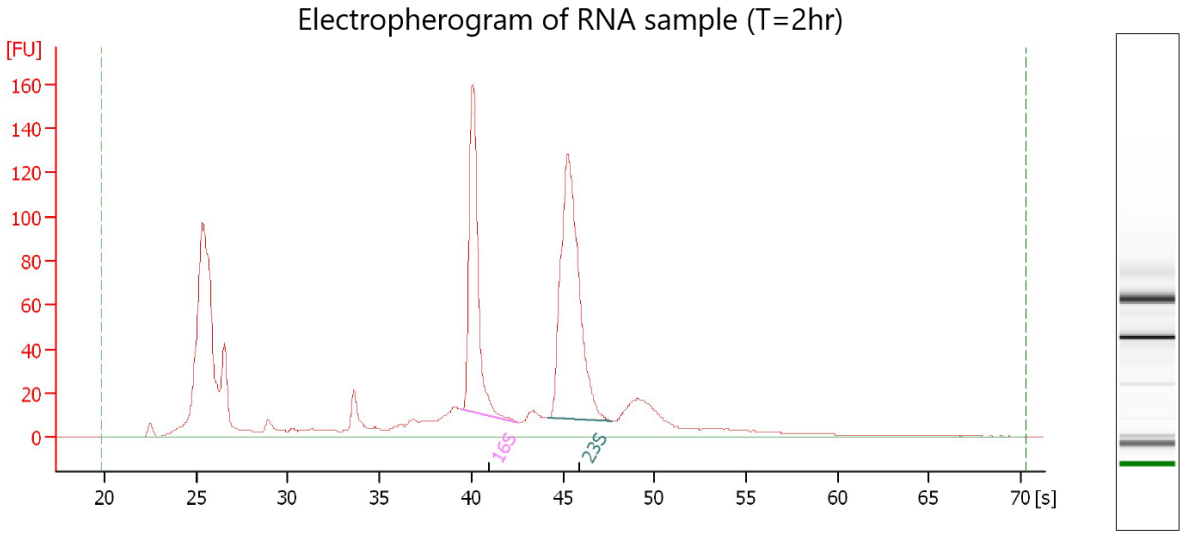
Supplementary Figure 4.11. Pyrimidine metabolism KEGG pathway

Over the course of the 6 month experiment no major specialisation of genes relating to pyrimidine metabolism were detected.



Supplementary Figure 4.12. Mineral and organic ion transport KEGG pathway

Over the course of the 6 month experiment no major specialisation of genes relating to mineral and organic ion transport were detected.



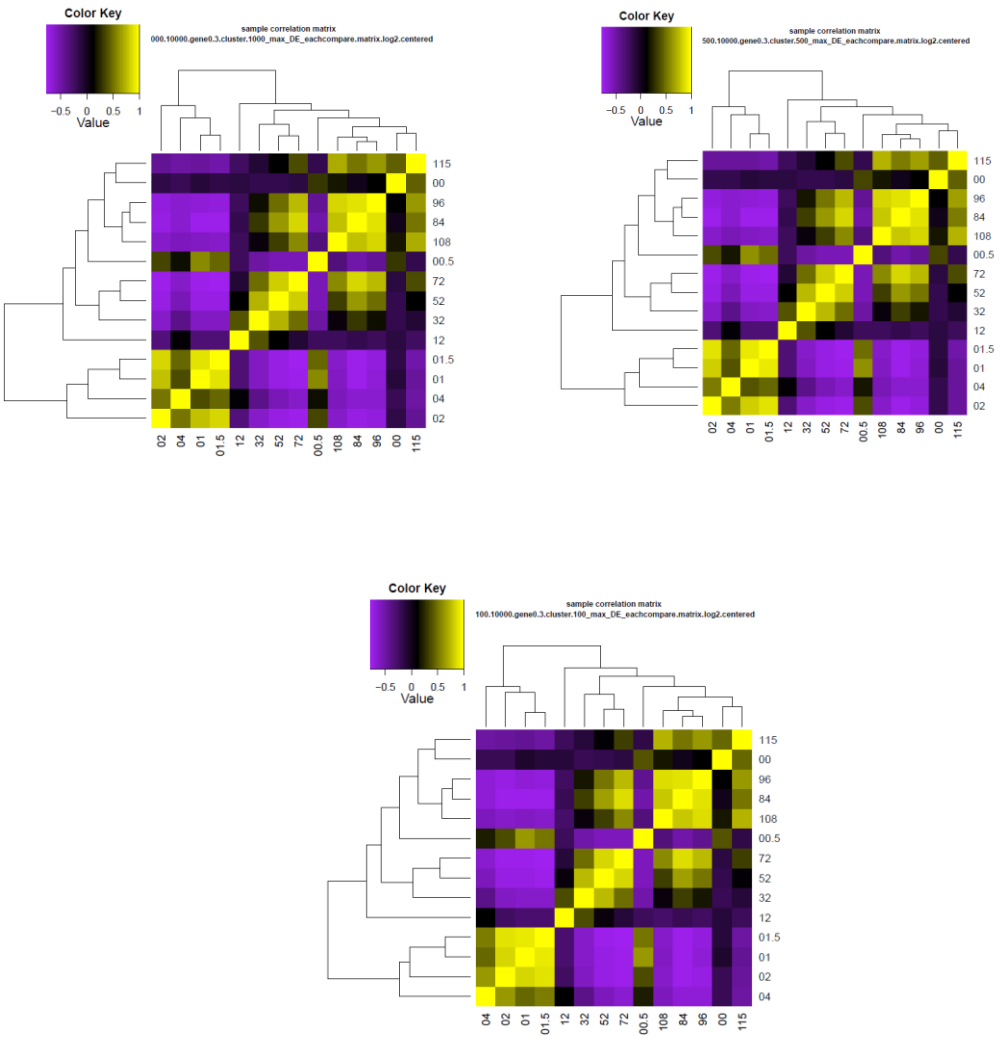
Supplementary Figure 4.13. Calculating RNA intergrity number by comparing areas of the 23 and 16S signal peaks.

Area of the peaks corresponding to 23S rRNA and 16S rRNA transcripts were compared to deduce quality of RNA.

Sample name (unique identifier)	Experiment grouping / additional details	Concentration (ng/µl)	A260:A280	A260:A230	RNA integrity number (RIN)
1	H0	222	1.91	1.97	6.8
2	H0.5	257.6	1.94	2.00	6.4
3	H1	295	1.95	2.06	7.1
4	H1.5	340.5	1.98	2.09	7.9
5	H2	639.6	2.03	2.1	8.5
6	H4	466.3	1.99	2.12	8.1
7	H12	368	2.02	2.14	8.9
8	H32	339.6	1.99	2.08	8.6
9	H52	400.5	1.98	2.13	8.5
10	H72	140.4	1.95	2.19	8.7
11	H84	356.8	1.96	2.07	8.9
12	H96	320.1	1.99	2.12	8.5
13	H108	341.5	1.96	2.05	7.7
14	H115	227.4	1.94	2.03	8.1

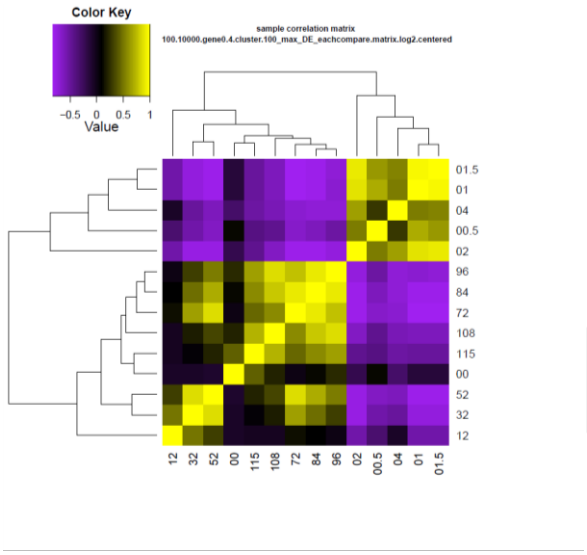
Supplementary table 4.2. Concentrations and RNA integrity values of RNA samples.

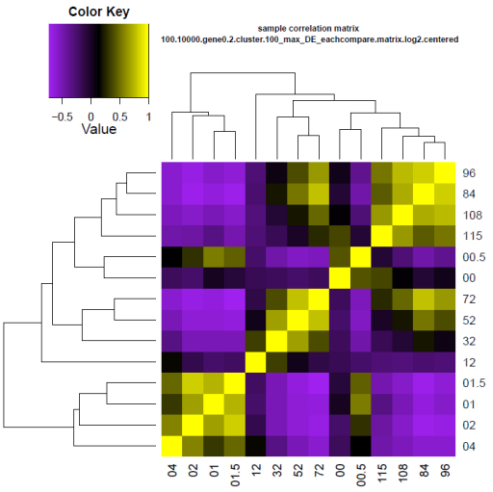
All samples showed a good degree of RNA quality and were suitable for subsequent sequencing

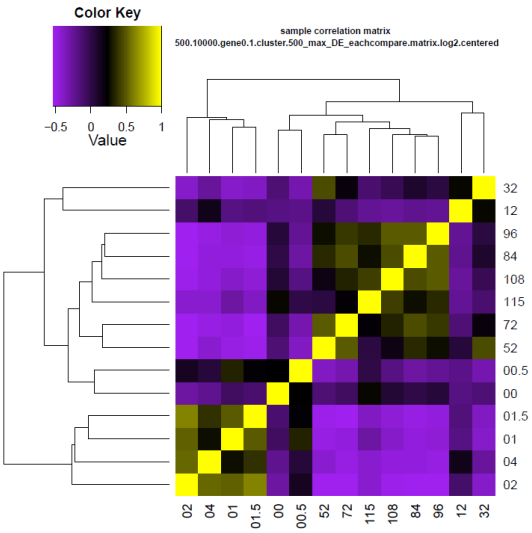


Supplementary Figure 4.14. Varying the number of differentially expressed genes included in the pairwise comparisons did not seem to affect groupings.

Increasing the number of genes considered to be part of the differentiation between samples did not show further resolution. Sufficient segregation was achieved using at least 100, although due to the power of the VIKING supercomputing cluster it was decided to use 1000 genes since the difference in computing time was negligible.

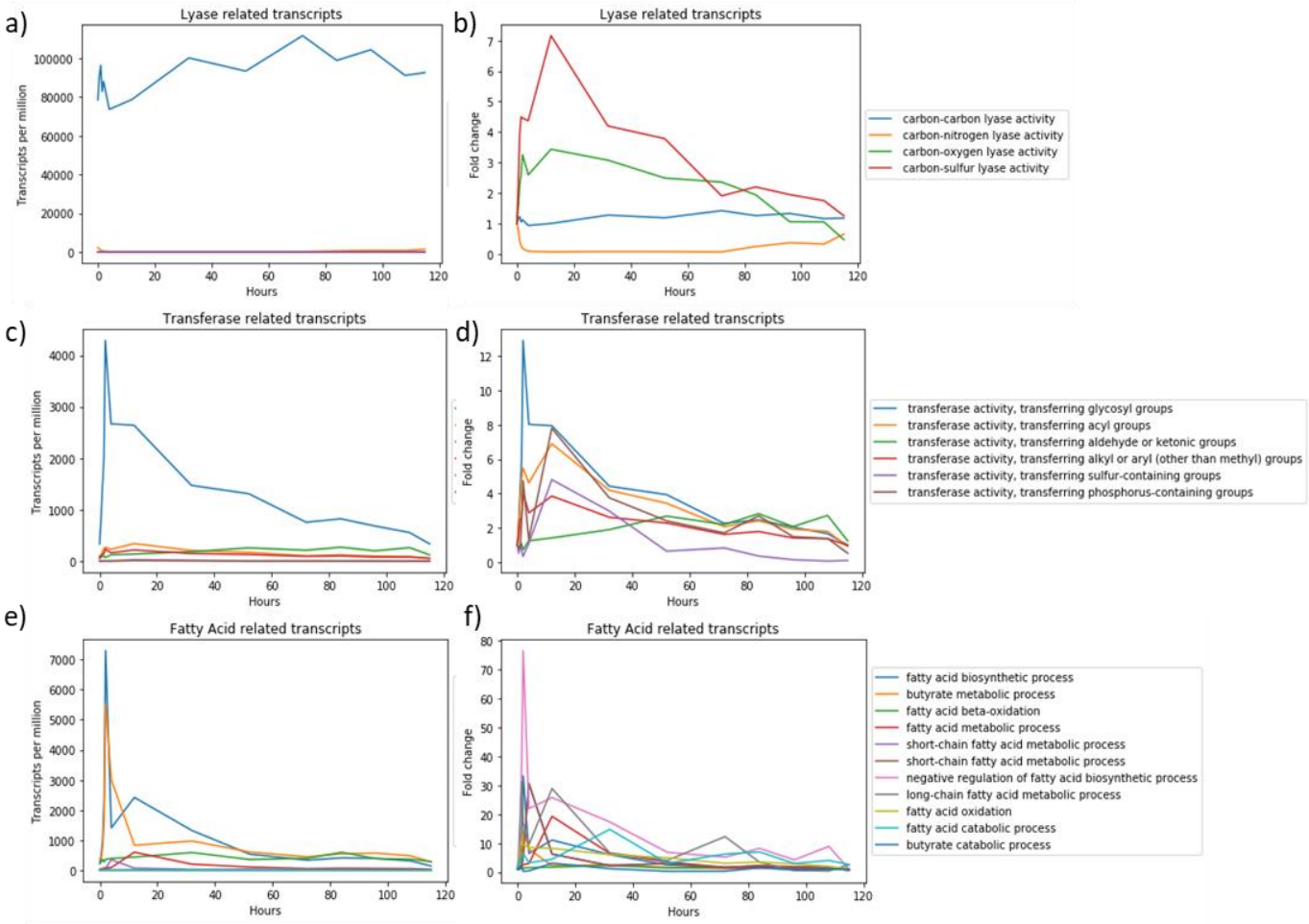




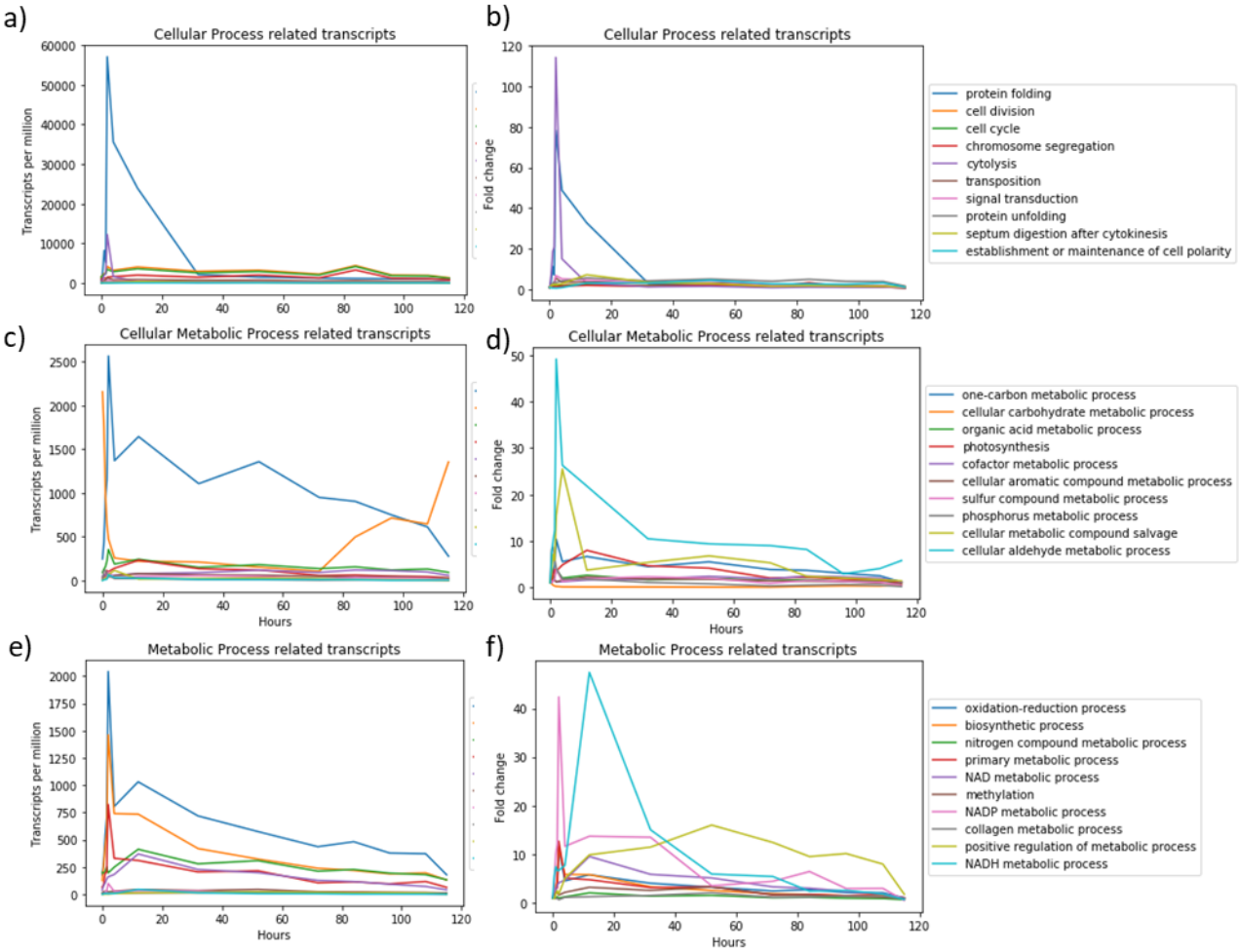


Supplementary Figure 15. A moderate dispersion value ensures differentiation between samples.

Varying the dispersion values across sample pairwise expression comparison to determine if transcripts are differentially expressed either produced clusters with not enough nuance (0.4) or too much (0.1).



Supplementary Figure 4.16. Expression profiles of GO categories and subcategories in terms of transcripts per million (left) and fold change (right)



Supplementary figure 4.17. GO term expression of high level biological processes.

4.10 File Guide

GOfys-viking.py – Expression analysis tool

GOI_terms.zip – Containing all the lists of GO categories investigated and following results

All raw short DNA reads, raw DNA long reads, assembled metagenome, raw mRNA reads and assembled metatatranscriptome are deposited under the BioProject SUB6360036

Discussion and Future work

We live in a rapidly changing world, this year (2019) saw hundreds of organised protests take place in cities across the world to voice their concerns as part of the largest act in history regarding climate change. This reflects the populations increasing economic, social and political concerns associated with ever increasing levels of pollution. The flow of money also shares these concerns and since 2004 global investment in renewable energy assets have increased from $46.6 bn to $285.9 bn in the year 2016, and is set to increase even further (FS-UNEP, 2016). Investments in technologies such as solar, wind, hydro and tidal energy have generally been well received whereas certain strategies, such primary bioenergy crops such as corn for ethanol production are questionable in a world which regularly sees protests and violence erupt over food shortages such as Pakistan, Indonesia and Yemen (Tenenbaum, 2008). As part of environmental initiatives aimed to reduce burden on both the energy and landfill sectors, the UK has reduced the amount of bio-municiple waste material sent to landfill from ~13,000 tonnes in 2010 to 7,500 tonnes in 2016. Part of this reduction can be attributed to the rapid proliferation of AD plants across the UK to treat agricultural, industrial and household wastes. In 2013 there were 100 operational sites and 648 by the summer of 2019 ("Anaerobic Digestion Plants UK (AD Plants UK or Anaerobic Digesters UK)," n.d.).

Similar to the growth of AD in the UK, the production of scotch whisky is increasing at a rate of 8% a year. The market for alternative uses for the by-products of this manufacturing process are therefore increasing. As all input material for making whisky is biological in origin, on site AD is perfectly suited to convert this waste material into renewable heat (for heating fermentation vessels), renewable energy (for powering stills) and renewable fertiliser for next years barley growing season.

AD is not a new technology, its use recorded throughout history, but this recent growth has prompted research efforts into understanding the microbial communities that perform the conversion of organic material into a methane rich biogas. Until now, biologists have not fully understood the inner workings of the community since ~80% of species living in AD are considered unculturable (Lagier et al., 2012; Nichols et al., 2010). Recent advances in DNA sequencing technologies have shed light on this "black box" of microbes resulting in new species, or even entire orders being discovered. By using modern DNA/RNA sequencing we are able to study the complex microbial methanogenic communities at an unprecedented level, reconstruct metabolic pathways, identify population changes and identify over-expressed pathways.

The initial goals set out in Chapter 1 were to anaerobically digest pot ale using a by-product from the waste water industry used in de-watering activated sludge. This material is a cationic acrylamide polymer, previously shown to be toxic to the AD community when fed in high amounts, and is covered with microorganisms that could be used to inoculate an AD system at low enough concentrations (Wang et al., 2018). Additionally, cPAM was stored between one and three months or subjected to rounds of freeze-thaw cycles to investigate if the material could be stored over long periods and what the effects of that might be in terms of seed performance. Gas and methane production from cPAM seed fed a 10% v/v solution of pot ale was recorded using a BMP system and revealed a decrease in performance after 30 days storage which was also made worse by storing the seed at room temperature rather than chilled. The pot ale feeding concentration of 10% v/v was found to be optimal in terms of both gas and methane production, although poor gas production can be attributed to the toxic effects of cPAM.

Population structure of the cPAM/pot ale communities was investigated using a V3-V4 16S rRNA amplification strategy which amplified a short fragment of hypervariable DNA. Bioinformatic analysis revealed that, although initial population differences exist, similar methanogenic populations developed after the addition of pot ale feed. Hierarchal clustering of pre and post-feeding communities revealed the main discerning feature was an increase of Bacterodia between 5 and 12% following feeding. The short 3 week duration of the tests coupled with a single feeding event limited the scale of population change within the communities and so further assessment of pot ale feedstock was performed over a longer period. It is also possible that cPAM inhibited or killed some of the community, thus preventing major changes, as gas yields were only 11% of the theoretical value. It is therefore recommended that in further studies a fully water-saturated cPAM concentration of 10% w/w be used or less and as a seed material only. The sequencing strategy used in the initial trials failed to provide a quantitative measure so overall population death or growth could not be confirmed, although biogas was produced indicating some active species.

Understanding how highly populated a reactor is in terms of microbe concentration is essential for identifying over or underperforming communities. The lack of quantitative data produced from the 16S rRNA gene sequencing strategy (used extensively across all fields of microbial ecology) inspired the development of a *Sulfolobus solfataricus* microbial spike which, when applied to sludge, would provide a way to normalise methodology in addition to a degree of quantification. Previous studies investigating microbial spikes did not quantify spike ploidy or enumerate cells as presented here in Chapter 2. Furthermore, we identified a unique marker gene (single stranded binding protein) present

in *Sulfolobus solfataricus* but absent from the sludge microbiome which allowed the use of qPCR to investigate spike response and determine sludge sample characteristics such as biological richness and inorganic contamination. The spike showed an extremely linear (R^2 = 0.97) relationship when diluting microbial biomass with sand, a common inorganic contaminant in AD. Unfortunately the 16S rRNA amplicon sequencing failed to detect the *Sulfolobus solfataricus* in all samples, although it was detecting using qPCR, and so quantitative microbial population information was produced. Further attempts to amplify the whole of the rRNA operon (~4 kb), for sequencing on the cheaper MinION platform, was inconsistent and further work is needed to optimise primers or amplification conditions. However, recent decreases of metagenomics sequencing cost mean that future attempts to investigate microbial-spiked samples may be better suited for metagenomics sequencing rather than amplicon sequencing, especially since primer bias present in amplicon sequencing can fail to accurately represent sample communities (Venkataraman et al., 2018).

Chapter 4 continued the study of pot ale anaerobic digestion over a period of 6 months to provide long term digestion data to support its suitability as a feedstock. The lab scale system constructed for this tasks provided real time gas and methane measurements and show high reproducibility amongst replicates. As part of this study a metagenomic sequencing strategy was used as it provides much more information, when compared to 16S rRNA gene profiling, regarding metabolic functionality and is not subject to amplification bias, which was experienced heavily in chapter 2. Multiple KEGG pathways including those for sugar transport and methane metabolism were constructed using this data and taxonomic annotation showed an increasing Bacterodia population (as also seen in Chapter 1). Following pot ale feeding, methane production occurred in two distinct phases, one rapid and one more gradual over a period of 115 hours. A further gene expression study, using metatranscriptomics, sought to identify the cause of this and determine what metabolic processes in the community are responsible for the two stages of methane production. This revealed initial methane production to be caused by AD of residual sugars including maltose in the pot ale and the second phase due to beta-glucan (left over from incomplete mashing of barley) hydrolysis and subsequent sugar degradation. Most interestingly, the metatranscriptomic study revealed expression of licheninases for over 72 hours indicating that beta-glucan degradation is the rate limiting hydrolysis step in the anaerobic digestion of pot ale. Taxonomic classification of transcripts revealed that Clostridia accounted for over 90% of transcription in the first 8 hours, then followed by Methanomicrobia activity which dominated the second half of the timecourse. Comparing the relative genomic abundance, as taken from the last metagenomic sampling points, to the transcriptomic abundance differentiated active and inactive populations.

To date, this is the first study in anaerobic digestion which identifies the enzyme associated with the rate limiting hydrolysis step through metatranscriptomics and opens the door towards feedstock-specific enzyme supplementation of difficult to degrade material. The pipeline and GO-fys software presented in this study is robust and can be applied to any type of AD, or environmental, system. The next stage of this research would be to purify licheninase enzyme to add alongside the pot ale feed in the hope that complete digestion of pot ale occurs in a shorter time-frame, although alternative more recalcitrant feedstocks may be a more industrially relevant target.

Over the course of this four-year project, the field of microbial ecology has transitioned away from amplicon based sequencing in favour of metagenomics to provider deeper, and less bias, understanding of unculturable organisms (Jia et al., 2018; Meneghine et al., 2017; Sheth et al., 2019; Zhang et al., 2019b; Zhu et al., 2019). This has been aided, in part, by the decreasing costs and increasing portability of devices such as the Oxford Nanopore MinION and realisation that whole phyla may be discriminated against, such as Lokiarchaeaota, when using amplification strategies (Zaremba-Niedzwiedzka et al., 2017).

In summary this research addresses many important questions in anaerobic digestion; who's there?, how many are there?, what are they doing?, why are they doing it?, how are they doing it?, when are they doing it? and why is it taking so long?

The last question, why is it taking so long?, is arguably the most import one that this research answered, using metatranscriptomics, as minor decreases in digestion time equate to major increases in methane yields on industrial scale AD plants containing huge volumes, such as the one million litre AD plant at the Glendullan distillery, Scotland ("Whisky distillery turns co-products into renewable energy," 2017).

Abbreviations

2D	-Two dimensional
3D	-Three dimensional
°C	-Degrees celcius
µl	-microlitre
µM	-micromolar
AD	-Anaerobic Digestion
bp	-base pair
COD	-Chemical oxygen demand
cPAM	-Cationic polyacrylamide
DNA	-Deoxyribonucleic acid
EDTA	-Ethylenediaminetetraacetic acid
FRET	-Fourier resonance energy transfer
GO	-Gene Ontology
kb	-kilobasepair
kg	-kilogram
mg	-milligram
ml	-milliliter
mM	-millimolar
L	-Litre
min	-M
RNA	-ribonucleic acid
SSBP	-Single stranded binding protein
TPM	-transcripts per million
VFA	-Volatile Fatty Acid

References

Agler, M.T., Garcia, M.L., Lee, E.S., Schlicher, M., Angenent, L.T., 2008. Thermophilic anaerobic digestion to increase the net energy balance of corn grain ethanol. Environ. Sci. Technol. 42, 6723–6729. https://doi.org/10.1021/es800671a

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., Nielsen, P.H., 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat. Biotechnol. 31, 533–538. https://doi.org/10.1038/nbt.2579

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., Quince, C., 2014. Binning metagenomic contigs by coverage and composition. Nat. Methods 11, 1144–1146. https://doi.org/10.1038/nmeth.3103

Altschup, S.F., Gish, W., Pennsylvania, T., Park, U., 1990. Basic Local Alignment Search Tool. Mol. Microbiol. 215, 403–410.

Anaerobic Digestion Plants UK (AD Plants UK or Anaerobic Digesters UK) [WWW Document], n.d. URL https://anaerobic-digestion.com/anaerobic-digestion-plants/anaerobic-digestion-plants-uk/ (accessed 12.12.12).

Angelidaki, I., Ahring, B.K., 1992. Effects of free long-chain fatty acids on thermophilic anaerobic digestion. Appl. Microbiol. Biotechnol. 37, 808–812. https://doi.org/10.1007/BF00174850

Appels, L., Baeyens, J., Degrève, J., Dewil, R., 2008. Principles and potential of the anaerobic digestion of waste-activated sludge. Prog. Energy Combust. Sci. 34, 755–781. https://doi.org/10.1016/j.pecs.2008.06.002

Armitage, N., Rooseboom, A., 2000. The removal of urban litter from stormwater conduits and streams: Paper 1 - The quantities involved and catchment litter management options. Water SA 26, 181–187.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese,

J.C., Richardson, J.E., Ringwald, M., Rubin, Gerald M. Sherlock, G., 2000. The Gene Ontology Consortium, Michael Ashburner1, Catherine A. Ball3, Judith A. Blake4, David Botstein3, Heather Butler1, J. Michael Cherry3, Allan P. Davis4, Kara Dolinski3, Selina S. Dwight3, Janan T. Eppig4, Midori A. Harris3, David P. Hill4, Laurie Is. Nat. Genet. 25, 25–29. https://doi.org/10.1038/75556.Gene

Awobusuyi, T.D., 2016. Concentration of Ammonium from Dilute Aqueous Solutions using Commercially Available Reverse Osmosis Membranes.

Baek, J., Kim, T.Y., Kim, W., Lee, H.J., Yi, J., 2014. Selective production of 1,3-butadiene using glucose fermentation liquor. Green Chem. 16, 3501–3507. https://doi.org/10.1039/c4gc00485j

Bang-andreasen, T., Anwar, M.Z., 2019. Total RNA-sequencing reveals multi-level microbial community changes and functional responses to wood ash application in agricultural and forest soil. bioRxiv 621557. https://doi.org/10.1101/621557

Barrena, R., Traub, J.E., Rodriguez, C., Goodwin, J.A.S., Harper, A.J., Willoughby, N.A., Sánchez, A., Aspray, T.J., 2018. Batch anaerobic digestion of deproteinated malt whisky pot ale using different source inocula. Waste Manag. 71, 675–682. https://doi.org/10.1016/j.wasman.2017.06.025

Bateman, A., 2019. UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res. 47, D506–D515. https://doi.org/10.1093/nar/gky1049

Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Apweiler, R., Alpi, E., Antunes, R., Arganiska, J., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Chavali, G., Cibrian-Uhalte, E., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., Gane, P., Castro, L.G., Garmiri, P., Hatton-Ellis, E., Hieta, R., Huntley, R., Legge, D., Liu, W., Luo, J., Macdougall, A., Mutowo, P., Nightingale, A., Orchard, S., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G., Rosanoff, S., Saidi, R., Sawford, T., Shypitsyna, A., Turner, E., Volynkin, V., Wardell, T., Watkins, X., Zellner, H., Cowley, A., Figueira, L., Li, W., McWilliam, H., Lopez, R., Xenarios, I., Bougueleret, L., Bridge, A., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., Casal-Casas, C., De Castro, E., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Jungo, F.,

Keller, G., Lara, V., Lemercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T., Nouspikel, N., Paesano, S., Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A.L., Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K., McGarvey, P., Natale, D.A., Suzek, B.E., Vinayaka, C.R., Wang, Q., Wang, Y., Yeh, L.S., Yerramalla, M.S., Zhang, J., 2015. UniProt: A hub for protein information. Nucleic Acids Res. 43, D204–D212. https://doi.org/10.1093/nar/gku989

Bauckhage, C., 2015. NumPy / SciPy Recipes for Data Science : k -Medoids Clustering 1, 1–6. https://doi.org/10.13140/2.1.4453.2009

Beazley, M.J., Martinez, R.J., Rajan, S., Powell, J., Piceno, Y.M., Tom, L.M., Andersen, G.L., Hazen, T.C., van Nostrand, J.D., Zhou, J., Mortazavi, B., Sobecky, P.A., 2012. Microbial community analysis of a coastal salt marsh affected by the Deepwater Horizon oil spill. PLoS One 7. https://doi.org/10.1371/journal.pone.0041305

Beckmann, S., Welte, C., Li, X., Oo, Y.M., Kroeninger, L., Heo, Y., Zhang, M., Ribeiro, D., Lee, M., Bhadbhade, M., Marjo, C.E., Seidel, J., Deppenmeier, U., Manefield, M., 2015. Novel phenazine crystals enable direct electron transfer to methanogens in anaerobic digestion by redox potential modulation. Energy Environ. Sci. https://doi.org/10.1039/C5EE03085D

Bermingham, E.N., Maclean, P., Thomas, D.G., Cave, N.J., Young, W., 2017. Key bacterial families (Clostridiaceae, Erysipelotrichaceae and Bacteroidaceae) are related to the digestion of protein and energy in dogs. PeerJ 2017. https://doi.org/10.7717/peerj.3019

Bernander, R., Poplawski, A., 1997. Cell cycle characteristics of thermophilic archaea. J. Bacteriol. 179, 4963–4969. https://doi.org/10.1128/jb.179.16.4963-4969.1997

Bertucci, M., Calusinska, M., Goux, X., Rouland-Lefèvre, C., Untereiner, B., Ferrer, P., Gerin, P.A., Delfosse, P., 2019. Carbohydrate Hydrolytic Potential and Redundancy of an Anaerobic Digestion Microbiome Exposed to Acidosis, as Uncovered by Metagenomics. Appl. Environ. Microbiol. 85, 1–16. https://doi.org/10.1128/aem.00895-19

Bey, B.S., Fichot, E.B., Dayama, G., Decho, A.W., Norman, R.S., 2010. Extraction of high molecular

weight DNA from microbial mats. Biotechniques 49, 631–640. https://doi.org/10.2144/000113486

Biebl, H., Menzel, K., Zeng, A.P., Deckwer, W.D., 1999. Microbial production of 1,3-propanediol. Appl. Microbiol. Biotechnol. 52, 289–297. https://doi.org/10.1007/s002530051523

Biesta-Peters, E.G., Reij, M.W., Joosten, H., Gorris, L.G.M., Zwietering, M.H., 2010. Comparison of two optical-density-based methods and a plate count method for estimation of growth parameters of bacillus cereus. Appl. Environ. Microbiol. 76, 1399–1405. https://doi.org/10.1128/AEM.02336-09

Blika, P.S., Stamatelatou, K., Kornaros, M., Lyberatos, G., 2009. Anaerobic digestion of olive mill wastewater. Glob. Nest J. 11, 364–372.

Bolland, M.D.A.I., Gilkes, R.J., 1990. Evaluation of nitrogen fertilizing value of composted household solid waste under greenhouse conditions. Agron. Sustain. Dev. 22, 79–80. https://doi.org/10.1051/agro

Bolyen, E., Dillon, M., Bokulich, N., Abnet, C., Al-Ghalith, G., Alexander, H., Alm, E., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J., Bittinger, K., Brejnrod, A., Brislawn, C., Brown, T., Callahan, B., Chase, J., Cope, E., Dorrestein, P., Douglas, G., Durall, D., Duvallet, C., Edwardson, C., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J., Gibson, D., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G., Janssen, S., Jarmusch, A., Jiang, L., Kaehler, B., Keefe, C., Keim, P., Kelley, S., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Lee, J., Ley, R., Liu, Y.-X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B., McDonald, D., McIver, L., Melnik, A., Metcalf, J., Morgan, S., Morton, J., Navas-Molina, J., Orchanian, S., Pearson, T., Peoples, S., Petras, D., Pruesse, E., Rivers, A., Robeson, M., Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Spear, J., Swafford, A., Thompson, L., Torres, P., Trinh, P., Tripathi, A., Turnbaugh, P., Ul-Hasan, S., Vargas, F., Vogtmann, E., Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K., Willis, A., Zaneveld, J., Zhang, Y., Zhu, Q., Knight, R., Caporaso, G., 2018. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. PeerJ Prepr. https://doi.org/10.7287/peerj.preprints.27295

Borreani, G., Tabacco, E., 2014. Improving corn silage quality in the top layer of farm bunker silos

through the use of a next-generation barrier film with high impermeability to oxygen. J. Dairy Sci. 97, 2415–2426. https://doi.org/10.3168/jds.2013-7632

Botheju, D., Lie, B., Bakke, R., 2010. Oxygen effects in anaerobic digestion - II. Model. Identif. Control 31, 55–65. https://doi.org/10.4173/mic.2010.2.2

Bräuer, S.L., Cadillo-Quiroz, H., Yashiro, E., Yavitt, J.B., Zinder, S.H., 2006. Isolation of a novel acidiphilic methanogen from an acidic peat bog. Nature 442, 192–194. https://doi.org/10.1038/nature04810

Bremges, A., Maus, I., Belmann, P., Eikmeyer, F., Winkler, A., Albersmeier, A., Pühler, A., Schlüter, A., Sczyrba, A., 2015. Deeply sequenced metagenome and metatranscriptome of a biogas-producing microbial community from an agricultural production-scale biogas plant. Gigascience 4, 2–7. https://doi.org/10.1186/s13742-015-0073-6

Brown, M., Wittwer, C., 2000. Flow cytometry: Principles and clinical applications in hematology. Clin. Chem. 46, 1221–1229.

Bru, D., Martin-Laurent, F., Philippot, L., 2008. Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16S rRNA gene as an example. Appl. Environ. Microbiol. 74, 1660–1663. https://doi.org/10.1128/AEM.02403-07

Brunt, J., Cross, K.L., Peck, M.W., 2015. Apertures in the Clostridium sporogenes spore coat and exosporium align to facilitate emergence of the vegetative cell. Food Microbiol. 51, 45–50. https://doi.org/10.1016/j.fm.2015.04.013

Bunney, P. E., Zink, A. N., Holm, A. A., Billington, C. J., & Kotz, C.M., 2017. Methylerythritol Phosphate Pathway of Isoprenoid Biosynthesis. Physiol. Behav. 176, 139–148. https://doi.org/10.1016/j.physbeh.2017.03.040

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, 1–12. https://doi.org/10.3389/fmicb.2016.00778

Campanaro, S., Treu, L., Kougias, P.G., De Francisci, D., Valle, G., Angelidaki, I., 2016. Metagenomic analysis and functional characterization of the biogas microbiome using high throughput shotgun sequencing and a novel binning strategy. Biotechnol. Biofuels 9, 1–17. https://doi.org/10.1186/s13068-016-0441-1

Campos, E., Almirall, M., Mtnez-Almela, J., Palatsi, J., Flotats, X., 2008. Feasibility study of the anaerobic digestion of dewatered pig slurry by means of polyacrylamide. Bioresour. Technol. 99, 387–395. https://doi.org/10.1016/j.biortech.2006.12.008

Cao, X., Liu, X., Dong, X., 2003. Alkaliphilus crotonatoxidans sp. nov., a strictly anaerobic, crotonate-dismutating bacterium isolated from a methanogenic environment. Int. J. Syst. Evol. Microbiol. 53, 971–975. https://doi.org/10.1099/ijs.0.02373-0

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G. a, Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C. a, Mcdonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W. a, Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. correspondence QIIME allows analysis of high- throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. Nat. Publ. Gr. 7, 335–336. https://doi.org/10.1038/nmeth0510-335

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., Knight, R., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. U. S. A. 108, 4516–4522. https://doi.org/10.1073/pnas.1000080107

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

Carbon, S., Douglass, E., Dunn, N., Good, B., Harris, N.L., Lewis, S.E., Mungall, C.J., Basu, S., Chisholm, R.L., Dodson, R.J., Hartline, E., Fey, P., Thomas, P.D., Albou, L.P., Ebert, D., Kesling, M.J., Mi, H., Muruganujan, A., Huang, X., Poudel, S., Mushayahama, T., Hu, J.C., LaBonte, S.A., Siegele, D.A., Antonazzo, G., Attrill, H., Brown, N.H., Fexova, S., Garapati, P., Jones, T.E.M., Marygold, S.J.,

Millburn, G.H., Rey, A.J., Trovisco, V., Dos Santos, G., Emmert, D.B., Falls, K., Zhou, P., Goodman, J.L., Strelets, V.B., Thurmond, J., Courtot, M., Osumi, D.S., Parkinson, H., Roncaglia, P., Acencio, M.L., Kuiper, M., Lreid, A., Logie, C., Lovering, R.C., Huntley, R.P., Denny, P., Campbell, N.H., Kramarz, B., Acquaah, V., Ahmad, S.H., Chen, H., Rawson, J.H., Chibucos, M.C., Giglio, M., Nadendla, S., Tauber, R., Duesbury, M.J., Del, N.T., Meldal, B.H.M., Perfetto, L., Porras, P., Orchard, S., Shrivastava, A., Xie, Z., Chang, H.Y., Finn, R.D., Mitchell, A.L., Rawlings, N.D., Richardson, L., Sangrador-Vegas, A., Blake, J.A., Christie, K.R., Dolan, M.E., Drabkin, H.J., Hill, D.P., Ni, L., Sitnikov, D., Harris, M.A., Oliver, S.G., Rutherford, K., Wood, V., Hayles, J., Bahler, J., Lock, A., Bolton, E.R., De Pons, J., Dwinell, M., Hayman, G.T., Laulederkind, S.J.F., Shimoyama, M., Tutaj, M., Wang, S.J., D'Eustachio, P., Matthews, L., Balhoff, J.P., Aleksander, S.A., Binkley, G., Dunn, B.L., Cherry, J.M., Engel, S.R., Gondwe, F., Karra, K., MacPherson, K.A., Miyasato, S.R., Nash, R.S., Ng, P.C., Sheppard, T.K., Shrivatsav Vp, A., Simison, M., Skrzypek, M.S., Weng, S., Wong, E.D., Feuermann, M., Gaudet, P., Bakker, E., Berardini, T.Z., Reiser, L., Subramaniam, S., Huala, E., Arighi, C., Auchincloss, A., Axelsen, K., Argoud, G.P., Bateman, A., Bely, B., Blatter, M.C., Boutet, E., Breuza, L., Bridge, A., Britto, R., Bye-A-Jee, H., Casals-Casas, C., Coudert, E., Estreicher, A., Famiglietti, L., Garmiri, P., Georghiou, G., Gos, A., Gruaz-Gumowski, N., Hatton-Ellis, E., Hinz, U., Hulo, C., Ignatchenko, A., Jungo, F., Keller, G., Laiho, K., Lemercier, P., Lieberherr, D., Lussi, Y., Mac-Dougall, A., Magrane, M., Martin, M.J., Masson, P., Natale, D.A., Hyka, N.N., Pedruzzi, I., Pichler, K., Poux, S., Rivoire, C., Rodriguez-Lopez, M., Sawford, T., Speretta, E., Shypitsyna, A., Stutz, A., Sundaram, S., Tognolli, M., Tyagi, N., Warner, K., Zaru, R., Wu, C., Chan, J., Cho, J., Gao, S., Grove, C., Harrison, M.C., Howe, K., Lee, R., Mendel, J., Muller, H.M., Raciti, D., Van Auken, K., Berriman, M., Stein, L., Sternberg, P.W., Howe, D., Toro, S., Westerfield, M., 2019. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 47, D330–D338. https://doi.org/10.1093/nar/gky1055

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S., Hub, A., Presence, W., 2009. AmiGO : online access to ontology and annotation data 25, 288–289. https://doi.org/10.1093/bioinformatics/btn615

Cavaleiro, A.J., Salvador, A.F., Alves, J.I., Alves, M., 2009. Continuous high rate anaerobic treatment of oleic acid based wastewater is possible after a step feeding start-up. Environ. Sci. Technol. 43, 2931–2936. https://doi.org/10.1021/es8031264

Chakrabarti, S.K., Roychoudhury, P.K., Bajpai, P.K., 1999. Biphasic biomethanation of wood-

hydrolysate effluent. Artif. Cells. Blood Substit. Immobil. Biotechnol. 27, 461–467. https://doi.org/10.3109/10731199909117720

Chen, Y., Cheng, J.J., Creamer, K.S., 2008. Inhibition of anaerobic digestion process: A review. Bioresour. Technol. 99, 4044–4064. https://doi.org/10.1016/j.biortech.2007.01.057

Cherif, S., Aloui, F., Carrière, F., Sayadi, S., 2014. Lipase Pre-Hydrolysis Enhance Anaerobic Biodigestion of Soap Stock from an Oil Refining Industry. J. Oleo Sci. J. Oleo Sci 63, 109–114. https://doi.org/10.5650/jos.ess13150

Cirne, D.G., Paloumet, X., Björnsson, L., Alves, M.M., Mattiasson, B., 2007. Anaerobic digestion of lipid-rich waste-Effects of lipid concentration. Renew. Energy 32, 965–975. https://doi.org/10.1016/j.renene.2006.04.003

Clausen, P.T.L.C., Aarestrup, F.M., Lund, O., 2018. Rapid and precise alignment of raw reads against redundant databases with KMA. BMC Bioinformatics 19, 1–8. https://doi.org/10.1186/s12859-018-2336-6

Colón, J., Cadena, E., Pognani, M., Barrena, R., Sánchez, A., Font, X., Artola, A., 2012. Determination of the energy and environmental burdens associated with the biological treatment of source-separated municipal solid wastes. Energy Environ. Sci. 5, 5731. https://doi.org/10.1039/c2ee01085b

Comittee, J.N.C., 2013. The global biodiversity footprint of UK biofuel consumption. Jt. Nat. Conserv. Com. 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Cruz Viggi, C., Rossetti, S., Fazi, S., Paiano, P., Majone, M., Aulenta, F., 2014. Magnetite particles triggering a faster and more robust syntrophic pathway of methanogenic propionate degradation. Environ. Sci. Technol. 48, 7536–7543. https://doi.org/10.1021/es5016789

Cuscó, A., Catozzi, C., Viñes, J., Sanchez, A., Francino, O., 2018. Microbiota profiling with long amplicons using nanopore sequencing: Full-length 16s rRNA gene and whole rrn operon [version 1; referees: 2 approved, 3 approved with reservations]. F1000Research 7, 1–29. https://doi.org/10.12688/f1000research.16817.1

Cysneiros, D., Banks, C.J., Heaven, S., Karatzas, K.A.G., 2012. The effect of pH control and "hydraulic flush" on hydrolysis and Volatile Fatty Acids (VFA) production and profile in anaerobic leach bed reactors digesting a high solids content substrate. Bioresour. Technol. 123, 263–271. https://doi.org/10.1016/j.biortech.2012.06.060

Dai, W., Xu, X., Liu, B., Yang, F., 2015. Toward energy-neutral wastewater treatment: A membrane combined process of anaerobic digestion and nitritation–anammox for biogas recovery and nitrogen removal. Chem. Eng. J. 279, 725–734. https://doi.org/10.1016/j.cej.2015.05.036

Dai, X., Luo, F., Yi, J., He, Q., Dong, B., 2014a. Biodegradation of polyacrylamide by anaerobic digestion under mesophilic condition and its performance in actual dewatered sludge system. Bioresour. Technol. 153, 55–61. https://doi.org/10.1016/j.biortech.2013.11.007

Dai, X., Luo, F., Yi, J., He, Q., Dong, B., 2014b. Biodegradation of polyacrylamide by anaerobic digestion under mesophilic condition and its performance in actual dewatered sludge system. Bioresour. Technol. 153, 55–61. https://doi.org/10.1016/j.biortech.2013.11.007

Dai, Z., Wu, Z., Hang, S., Zhu, W., Wu, G., 2014. Amino acid metabolism in intestinal bacteria and its potential implications for mammalian reproduction. Mol. Hum. Reprod. 21, 389–409. https://doi.org/10.1093/molehr/gav003

Dattagupta, S., Schaperdoth, I., Montanari, A., Mariani, S., Kita, N., Valley, J.W., Macalady, J.L., 2009. A novel symbiosis between chemoautotrophic bacteria and a freshwater cave amphipod. ISME J. 3, 935–43. https://doi.org/10.1038/ismej.2009.34

de Vladar, H.P., 2012. Amino acid fermentation at the origin of the genetic code. Biol. Direct 7, 6. https://doi.org/10.1186/1745-6150-7-6

DECC, 2009. National Renewable Energy Action Plan for the United Kingdom Article 4 of the Renewable Energy Directive. Dep. Energy Clim. Chang. 1–160. https://doi.org/1st July 2010

DECC (Department of Energy & Climate Change), 2016. UK Energy Statistics. Stat. Press release 1–16.

Dedkov, Y.M., Elizarova, O. V, Kei, S.Y., 2000. Dichromate Method for the Determination of Chemical

Oxygen Demand. J. Anal. Chem. 55, 777–781.

Defra, 2016. UK Statistics on Waste 1–24.

Delforno, T.P., Macedo, T.Z., Midoux, C., Lacerda, G. V, Rué, O., Mariadassou, M., Loux, V., Varesche, M.B.A., Bouchez, T., Bize, A., Oliveira, V.M., 2018. Comparative metatranscriptomic analysis of anaerobic digesters treating anionic surfactant contaminated wastewater. Sci. Total Environ. 649, 482–494. https://doi.org/10.1016/j.scitotenv.2018.08.328

Department for Bussiness Energy & Industrial Stratergy, 2016. Energy Consumption. Crown Copyr.

Department for Environment Food and Rural Affairs, 2015. 2010 to 2015 government policy: waste and recycling.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72, 5069–5072. https://doi.org/10.1128/AEM.03006-05

Díaz-gonzález, F., Milano, M., Olguin-araneda, V., Pizarro-cerda, J., Castro-córdova, P., Tzeng, S., Maier, C.S., Sarker, M.R., Paredes-sabja, D., 2015. Protein composition of the outermost exosporium-like layer of Clostridium difficile. J. Proteomics 123, 1–13. https://doi.org/10.1016/j.jprot.2015.03.035

Díaz, C., Baena, S., Fardeau, M.L., Patel, B.K.C., 2007. Aminiphilus circumscriptus gen. nov., sp. nov., an anaerobic amino-acid-degrading bacterium from an upflow anaerobic sludge reactor. Int. J. Syst. Evol. Microbiol. 57, 1914–1918. https://doi.org/10.1099/ijs.0.63614-0

Diner, B.A., Fan, J., Scotcher, M.C., Wells, D.H., Whited, G.M., 2018. Synthesis of heterologous mevalonic acid pathway enzymes in Clostridium ljungdahlii for the conversion of fructose and of syngas to mevalonate and isoprene. Appl. Environ. Microbiol. 84, 1–16. https://doi.org/10.1128/AEM.01723-17

DiPippo, J.L., Nesbø, C.L., Dahle, H., Doolittle, W.F., Birkland, N.K., Noll, K.M., 2009. Kosmotoga olearia

gen. nov., sp. nov., a thermophilic, anaerobic heterotroph isolated from an oil production fluid. Int. J. Syst. Evol. Microbiol. 59, 2991–3000. https://doi.org/10.1099/ijs.0.008045-0

Drancourt, M., Berger, P., Raoult, D., 2004. Systematic 16S rRNA Gene Sequencing of Atypical Clinical Isolates Identified 27 New Bacterial Species Associated with Humans. Society 42, 2197–2202. https://doi.org/10.1128/JCM.42.5.2197

Driks, A., 2002. Maximum shields: The assembly and function of the bacterial spore coat. Trends Microbiol. 10, 251–254. https://doi.org/10.1016/S0966-842X(02)02373-9

Dutilh, B.E., Cassman, N., Mcnair, K., Sanchez, S.E., Silva, G.G.Z., Boling, L., Barr, J.J., Speth, D.R., Seguritan, V., Aziz, R.K., Felts, B., Dinsdale, E.A., Mokili, J.L., Edwards, R.A., 2014. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat. Commun. 5, 1–11. https://doi.org/10.1038/ncomms5498

Eastman, J. a, Ferguson, J.F., 1981. Solubilization organic phase of of carbon anaerobic particulate during the digestion acid. J. (Water Pollut. Control Fed. 53, 352–366.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., Yakhini, Z., 2009. GOrilla : a tool for discovery and visualization of enriched GO terms in ranked gene lists 7, 1–7. https://doi.org/10.1186/1471-2105-10-48

Ellis, D.I., Muhamadali, H., Xu, Y., Eccles, R., Goodall, I., Goodacre, R., 2019. Rapid through-container detection of fake spirits and methanol quantification with handheld Raman spectroscopy. Analyst 144, 324–330. https://doi.org/10.1039/c8an01702f

Elshahed, M.S., McInerney, M.J., 2001. Benzoate Fermentation by the Anaerobic Bacterium Syntrophus aciditrophicus in the Absence of Hydrogen-Using Microorganisms. Appl. Environ. Microbiol. 67, 5520–5525. https://doi.org/10.1128/AEM.67.12.5520-5525.2001

Ettema, T.J.G., Brinkman, A.B., Lamers, P.P., Kornet, N.G., de Vos, W.M., van der Oost, J., 2006. Molecular characterization of a conserved archaeal copper resistance (cop) gene cluster and its copper-responsive regulator in Sulfolobus solfataricus P2. Microbiology 152, 1969–1979. https://doi.org/10.1099/mic.0.28724-0

Everett, K.D.E., Bush, R.M., Andersen, A.A., 1999. Emended description of the order Chlamydiales, proposal of Parachlamydiaceae fam. nov. and Simkaniaceae fam. nov., each containing one monotypic genus, revised taxonomy of the family Chlamydiaceae, including a new genus and five new species, and standards. Int. J. Syst. Bacteriol. 49, 415–440. https://doi.org/10.1099/00207713-49-2-415

Ezeonu, F.C., Okaka, A.N.C., 1996. Process kinetics and digestion efficiency of anaerobic batch fermentation of brewer's spent grains (BSG). Process Biochem. 31, 7–12.

Faith, D.P., Baker, A.M., 2010. Phylogenetic diversity (PD) and biodiversity conservation : some bioinformatics challenges 121–128.

Fanedl, L., Malovrh, Š., Marinšek, R., 2015. Biogas production from brewery spent grain enhanced by bioaugmentation with hydrolytic anaerobic bacteria. Bioresour. Technol. 186, 261–269. https://doi.org/10.1016/j.biortech.2015.03.029

Filer, J., Ding, H.H., Chang, S., 2019. Biochemical methane potential (BMP) assay method for anaerobic digestion research. Water (Switzerland) 11. https://doi.org/10.3390/w11050921

Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: Interactive sequence similarity searching. Nucleic Acids Res. 39, 29–37. https://doi.org/10.1093/nar/gkr367

Fountoulakis, M.S., Petousi, I., Manios, T., 2010. Co-digestion of sewage sludge with glycerol to boost biogas production. Waste Manag. 30, 1849–1853. https://doi.org/10.1016/j.wasman.2010.04.011

Freedman, Z., Zak, D.R., 2015. Soil bacterial communities are shaped by temporal and environmental filtering: evidence from a long-term chronosequence. Environ. Microbiol. 17, n/a-n/a. https://doi.org/10.1111/1462-2920.12762

FS-UNEP, 2016. Global Trends in Renewable Energy 84.

Fu, S.-F., Wang, F., Shi, X.-S., Guo, R.-B., 2016. Impacts of microaeration on the anaerobic digestion of corn straw and the microbial community structure. Chem. Eng. J. 287, 523–528.

https://doi.org/10.1016/j.cej.2015.11.070

Fykse, E.M., Aarskaug, T., Madslien, E.H., Dybwad, M., 2016. Microbial community structure in a full-scale anaerobic treatment plant during start-up and first year of operation revealed by high-throughput 16S rRNA gene amplicon sequencing. Bioresour. Technol. 222, 380–387. https://doi.org/10.1016/j.biortech.2016.09.118

Gao, S., Zhao, M., Chen, Y., Yu, M., Ruan, W., 2015. Tolerance response to in situ ammonia stress in a pilot-scale anaerobic digestion reactor for alleviating ammonia inhibition. Bioresour. Technol. 198, 372–379. https://doi.org/10.1016/j.biortech.2015.09.044

Garcia, H., Rico, C., Garcia, P.A., Rico, J.L., 2008. Flocculants effect in biomass retention in a UASB reactor treating dairy manure. Bioresour. Technol. 99, 6028–6036. https://doi.org/10.1016/j.biortech.2007.11.037

Gasc, C., Ribière, C., Parisot, N., Beugnot, R., Defois, C., Petit-Biderre, C., Boucher, D., Peyretaillade, E., Peyret, P., 2015. Capturing prokaryotic dark matter genomes. Res. Microbiol. 166, 814–830. https://doi.org/10.1016/j.resmic.2015.06.001

Gibbons, S.M., Scholz, M., Hutchison, A.L., Dinner, A.R., Gilbert, J.A., Colemana, M.L., 2016. Disturbance regimes predictably alter diversity in an ecologically complex bacterial system. MBio 7, 1–10. https://doi.org/10.1128/mBio.01372-16

Gilbert, J.A., Jansson, J.K., Knight, R., 2014. The Earth Microbiome project: Successes and aspirations. BMC Biol. 12, 1–4. https://doi.org/10.1186/s12915-014-0069-1

Gohl, D.M., Vangay, P., Garbe, J., MacLean, A., Hauge, A., Becker, A., Gould, T.J., Clayton, J.B., Johnson, T.J., Hunter, R., Knights, D., Beckman, K.B., 2016. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. Nat. Biotechnol. 34, 942–949. https://doi.org/10.1038/nbt.3601

Goodwin, J., Finlayson, J.M., Low, E.W., 2001. A further study of the anaerobic biotreatment of malt whisky distillery pot ale using an UASB system. Bioresour. Technol. 78, 155–160. https://doi.org/10.1016/S0960-8524(01)00008-6

Goodwin, J., Stuart, J., 1994. Anaerobic digestion of malt whisky distillery pot ale using upflow anaerobic sludge blanket reactors. Bioresour. Technol. 49, 75–81. https://doi.org/10.1016/0960-8524(94)90175-9

Gowda, M.C., Raghavan, G.S. V, Ranganna, B., Barrington, S., 1995. Rural Waste Management in a South Indian Village a Case Study 53, 157–164.

GPFeeds, 2015. Pot Ale Syrup Composition [WWW Document]. URL http://www.gpfeeds.co.uk/analysis/pot_ale_syrup.htm

Grabherr, M.G., Haas, B.J., Moran, Y., Joshua Z. Levin, Dawn A Thompson, I.A., Xian Adiconis, L.F., Raktima, Raychowdhury Qiandong, Z., Zehua, C., Evan, M., Hacohen, N., Andreas Gnirke, Nicholas Rhind, F. di P., Aviv, R., 2013. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883.Trinity

Gröllmann, U., Schnabel, W., 1982. Free radical-induced oxidative degradation of polyacrylamide in aqueous solution. Polym. Degrad. Stab. 4, 203–212. https://doi.org/10.1016/0141-3910(82)90027-1

Guo, J., Peng, Y., Ni, B.-J., Han, X., Fan, L., Yuan, Z., 2015. Dissecting microbial community structure and methane-producing pathways of a full-scale anaerobic reactor digesting activated sludge from wastewater treatment by metagenomic sequencing. Microb. Cell Fact. 14, 33. https://doi.org/10.1186/s12934-015-0218-4

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512. https://doi.org/10.1038/nprot.2013.084

Hagen, L.H., Vivekanand, V., Pope, P.B., Eijsink, V.G.H., Horn, S.J., 2015. The effect of storage conditions on microbial community composition and biomethane potential in a biogas starter

culture. Appl. Microbiol. Biotechnol. 99, 5749–5761. https://doi.org/10.1007/s00253-015-6623-0

Hahn, M., Pons, J., Planas, A., Querol, E., Heinemann, U., 1995. Crystal structure of Bacillus licheniformis 1,3-1,4-β-d-glucan 4-glucanohydrolase at 1.8 Å resolution. FEBS Lett. 374, 221–224. https://doi.org/10.1016/0014-5793(95)01111-Q

Hall, A.B., Tolonen, A.C., Xavier, R.J., 2017. Human genetic variation and the gut microbiome in disease. Nat. Publ. Gr. 18, 690–699. https://doi.org/10.1038/nrg.2017.63

Hansen, K.H., Angelidaki, I., Ahring, B.K., 1998. Anaerobic Digestion of Swine Manure: Inhibition By Ammonia. Water Res. 32, 5–12. https://doi.org/10.1016/S0043-1354(97)00201-7

Hassa, J., Maus, I., Off, S., Pühler, A., Scherer, P., Klocke, M., Schlüter, A., 2018. Metagenome, metatranscriptome, and metaproteome approaches unraveled compositions and functional relationships of microbial communities residing in biogas plants. Appl. Microbiol. Biotechnol. 102, 5045–5063. https://doi.org/10.1007/s00253-018-8976-7

Hayden, H.L., Savin, K.W., Wadeson, J., Gupta, V.V.S.R., Mele, P.M., 2018. Comparative metatranscriptomics of wheat Rhizosphere microbiomes in disease suppressive and non-suppressive soils for Rhizoctonia solani AG8. Front. Microbiol. 9, 1–19. https://doi.org/10.3389/fmicb.2018.00859

Henard, C.A., Smith, H., Dowe, N., Kalyuzhnaya, M.G., Pienkos, P.T., Guarnieri, M.T., 2016. Bioconversion of methane to lactate by an obligate methanotrophic bacterium. Sci. Rep. 6, 21585. https://doi.org/10.1038/srep21585

Heuston, S., Begley, M., Gahan, C.G.M., Hill, C., 2012. Isoprenoid biosynthesis in bacterial pathogens. Microbiol. (United Kingdom) 158, 1389–1401. https://doi.org/10.1099/mic.0.051599-0

Highfield, A.C., El Nagar, A., Mackinder, L.C.M., Noël, L.M.L.J., Hall, M.J., Martin, S.J., Schroeder, D.C., 2009. Deformed wing virus implicated in overwintering honeybee colony losses. Appl. Environ. Microbiol. 75, 7212–7220. https://doi.org/10.1128/AEM.02227-09

Hildenbrand, C., Stock, T., Lange, C., Rother, M., Soppa, J., 2011. Genome copy numbers and gene conversion in methanogenic archaea. J. Bacteriol. 193, 734–743. https://doi.org/10.1128/JB.01016-10

HMRC data shows Scotch exports hit record high in 2018 [WWW Document], 2019. URL https://www.bbc.co.uk/news/uk-scotland-scotland-business-47211794 (accessed 9.17.19).

Ho, D.P., Jensen, P.D., Batstone, D.J., 2013. Methanosarcinaceae and acetate-oxidizing pathways dominate in high-rate thermophilic anaerobic digestion of waste-activated sludge. Appl. Environ. Microbiol. 79, 6491–6500. https://doi.org/10.1128/AEM.01730-13

Ho, L., Ho, G., 2012. Mitigating ammonia inhibition of thermophilic anaerobic treatment of digested piggery wastewater : Use of pH reduction , zeolite , biomass and humic acid 6. https://doi.org/10.1016/j.watres.2012.05.016

Holliger, C., Alves, M., Andrade, D., Angelidaki, I., Astals, S., Baier, U., Bougrier, C., Buffiere, P., Carballa, M., de Wilde, V., Ebertseder, F., Fernandez, B., Ficara, E., Fotidis, I., Frigon, J.-C., de Laclos, H.F., Ghasimi, D.S.M., Hack, G., Hartel, M., Heerenklage, J., Horvath, I.S., Jenicek, P., Koch, K., Krautwald, J., Lizasoain, J., Liu, J., Mosberger, L., Nistor, M., Oechsner, H., Oliveira, J. V., Paterson, M., Pauss, A., Pommier, S., Porqueddu, I., Raposo, F., Ribeiro, T., Rusch Pfund, F., Stromberg, S., Torrijos, M., van Eekert, M., van Lier, J., Wedwitschka, H., Wierinck, I., 2016. Towards a standardization of biomethane potential tests. Water Sci. Technol. 1–9. https://doi.org/10.2166/wst.2016.336

Holt, H.M., Gahrn-Hansen, B., Bruun, B., 2005. Shewanella algae and Shewanella putrefaciens: Clinical and microbiological characteristics. Clin. Microbiol. Infect. 11, 347–352. https://doi.org/10.1111/j.1469-0691.2005.01108.x

Hou, Y., Peng, D., Wang, B., Zhang, X., Xue, X., 2013. Effects of stirring strategies on the sludge granulation in anaerobic CSTR reactor. Desalin. Water Treat. 52, 1–8. https://doi.org/10.1080/19443994.2013.841102

Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C., Lempicki, R.A., 2007. The DAVID Gene Functional Classification Tool: A novel

biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. 8. https://doi.org/10.1186/gb-2007-8-9-r183

Huang, M. mei, Arnheim, N., Goodman, M.F., 1992. Extension of base mispairs by Taq DNA polymerase: Implications for single nucleotide discrimination in PCR. Nucleic Acids Res. 20, 4567–4573. https://doi.org/10.1093/nar/20.17.4567

Hulshoff Pol, L.., de Castro Lopes, S.., Lettinga, G., Lens, P.N.., 2004. Anaerobic sludge granulation. Water Res. 38, 1376–1389. https://doi.org/10.1016/j.watres.2003.12.002

Hwang, K., Song, M., Kim, W., Kim, N., Hwang, S., 2010. Effects of prolonged starvation on methanogenic population dynamics in anaerobic digestion of swine wastewater. Bioresour. Technol. 101, S2–S6. https://doi.org/10.1016/j.biortech.2009.03.070

Illumina, 2013. 16S Metagenomic Sequencing Library Preparation. Illumina.com 1–28.

Imachi, H., Nobu, M.K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano, Y., Uematsu, K., Ikuta, T., Ito, M., Matsui, Y., Miyazaki, M., Murata, K., Saito, Y., Sakai, S., Song, C., Tasumi, E., Yamanaka, Y., Yamaguchi, T., Kamagata, Y., Tamaki, H., Takai, K., 2020. Isolation of an archaeon at the prokaryote–eukaryote interface. Nature 577, 519–525. https://doi.org/10.1038/s41586-019-1916-6

IWA Task Group for Mathematical Modelling of Anaerobic Digestion Processes, 2002. Anaerobic Digestion Model No.1 (ADM1). IWA Publishing.

Jamail, D., 2012. Gulf seafood deformities alarm scientists. Al-Jazeera.

Jang, H.B., Choi, M., Kang, J.H., Park, S.I., Lee, H., 2017. Association of dietary patterns with the fecal microbiota in Korean adolescents. BMC Nutr. 3, 1–11. https://doi.org/10.1186/s40795-016-0125-z

Jáuregui-Jáuregui, J.A., Méndez-Acosta, H.O., González-Álvarez, V., Snell-Castro, R., Alcaraz-González, V., Godon, J.J., 2014. Anaerobic treatment of tequila vinasses under seasonal operating conditions: Start-up, normal operation and restart-up after a long stop and starvation period.

Bioresour. Technol. 168, 33–40. https://doi.org/10.1016/j.biortech.2014.04.006

Jia, Y., Ng, S.K., Lu, H., Cai, M., Lee, P.K.H., 2018. Genome-centric metatranscriptomes and ecological roles of the active microbial populations during cellulosic biomass anaerobic digestion. Biotechnol. Biofuels 11, 1–15. https://doi.org/10.1186/s13068-018-1121-0

Jih-Gaw, L., Chang, C.N., Chang, S.C., 1997. Enhancement of anaerobic digestion of waste activated sludge by alkaline solubilization. Bioresour. Technol. 62, 85–90. https://doi.org/10.1016/S0960-8524(97)00121-1

Kato, M.T., Field, J.A., Lettinga, G., 1993. High Tolerance of Methanogens in Granular Sludge to Oxygen. Biotechnol. Bioeng. 42, 1360–1366.

Kavanagh, M., 2013. United Kingdom: Boilers on the backburner as consumers rein in waste. Financ. Times.

Kay-Shoemake, J.L., Watwood, M.E., Sojka, R.E., Lentz, R.D., 1998. Polyacrylamide as a substrate for microbial amidase in culture and soil. Soil Biol. Biochem. 30, 1647–1654. https://doi.org/10.1016/S0038-0717(97)00251-4

Kelsic, E.D., Zhao, J., Vetsigian, K., Kishony, R., 2015. Counteraction of antibiotic production and degradation stabilizes microbial communities. Nature 521, 516–519. https://doi.org/10.1038/nature14485

Kerkhof, L.J., Dillon, K.P., Häggblom, M.M., McGuinness, L.R., 2017. Profiling bacterial communities by MinION sequencing of ribosomal operons. Microbiome 5, 116. https://doi.org/10.1186/s40168-017-0336-9

Kettle, H., Louis, P., Holtrop, G., Duncan, S.H., Flint, H.J., 2015. Modelling the emergent dynamics and major metabolites of the human colonic microbiota. Environ. Microbiol. 17, 1615–1630. https://doi.org/10.1111/1462-2920.12599

Kieu, H.T.Q., Müller, E., Horn, H., 2011. Heavy metal removal in anaerobic semi-continuous stirred tank reactors by a consortium of sulfate-reducing bacteria. Water Res. 45, 3863–3870.

https://doi.org/10.1016/j.watres.2011.04.043

Kirkegaard, R.H., McIlroy, S.J., Kristensen, J.M., Nierychlo, M., Karst, S.M., Dueholm, M.S., Albertsen, M., Nielsen, P.H., 2017. The impact of immigration on microbial community composition in full-scale anaerobic digesters. Sci. Rep. 7, 1–11. https://doi.org/10.1038/s41598-017-09303-0

Kirtane, R.D., Suryawanshi, P.C., Chaudhari, A.B., Kothari, R.M., 2010. Studies on biphasic biomethanation of spoiled mango puree. Indian J. Biotechnol. 9, 308–312.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., Glöckner, F.O., 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 41, 1–11. https://doi.org/10.1093/nar/gks808

Knab, N.J., Dale, A.W., Lettmann, K., Fossing, H., Jørgensen, B.B., 2008. Thermodynamic and kinetic control on anaerobic oxidation of methane in marine sediments. Geochim. Cosmochim. Acta 72, 3746–3757. https://doi.org/10.1016/j.gca.2008.05.039

Kodama, H., Takeshita, K., Araki, T., Tanaka, H., 2004. Fluid particle dynamics simulation of charged colloidal suspensions. J. Phys. Condens. Matter 16, L115–L123. https://doi.org/10.1088/0953-8984/16/10/L01

Kotsyurbenko, O.R., Friedrich, M.W., Simankova, M. V., Nozhevnikova, A.N., Golyshin, P.N., Timmis, K.N., Conrad, R., 2007. Shift from acetoclastic to H2-dependent methanogenesis in a West Siberian peat bog at low pH values and isolation of an acidophilic Methanobacterium strain. Appl. Environ. Microbiol. 73, 2344–2348. https://doi.org/10.1128/AEM.02413-06

Krakat, N., Anjum, R., Demirel, B., Schröder, P., 2016. Methodological flaws introduce strong bias into molecular analysis of microbial populations. J. Appl. Microbiol. 49. https://doi.org/10.1111/jam.13365

Kwok, S., Kellogg, D.E., Mckinney, N., Spasic, D., Goda, L., Levenson, C., Sninsky, J.J., 1990. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. Nucleic Acids Res. 18, 999–1005. https://doi.org/10.1093/nar/18.4.999

Ladd, A.J.C., Verberg, R., 2001. Lattice-Boltzmann simulations of particle-fluid suspensions. J. Stat. Phys. 104, 1191–1251. https://doi.org/10.1023/A:1010414013942

Lagier, J.-C., Armougom, F., Million, M., Hugon, P., Pagnier, I., Robert, C., Bittar, F., Fournous, G., Gimenez, G., Maraninchi, M., Trape, J.-F., Koonin, E.V., La Scola, B., Raoult, D., 2012. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin. Microbiol. Infect. 18, 1185–1193. https://doi.org/10.1111/1469-0691.12023

Lee, J., Hong, W.-Y., Jang, E., Kim, J., 2015. FCMM: A comparative metagenomic approach for functional characterization of multiple metagenome samples. J. Microbiol. Methods 115, 121–128. https://doi.org/10.1016/j.mimet.2015.05.023

Leeds Council, n.d. Composting [WWW Document]. URL https://www.leeds.gov.uk/residents/bins-and-recycling/composting (accessed 8.29.19).

Lennon, J.T., Jones, S.E., 2011. Microbial seed banks: the ecological and evolutionary implications of dormancy. Nat. Rev. Microbiol. 9, 119–130. https://doi.org/10.1038/nrmicro2504

Li, A., Chu, Y., Wang, X., Ren, L., Yu, J., Liu, X., Yan, J., Zhang, L., Wu, S., Li, S., 2013. A pyrosequencing-based metagenomic study of methane-producing microbial community in solid-state biogas reactor. Biotechnol. Biofuels 6, 3. https://doi.org/10.1186/1754-6834-6-3

Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323.

Li, D., Liu, C.M., Luo, R., Sadakane, K., Lam, T.W., 2015. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31, 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, L., Sun, Y., Yuan, Z., Kong, X., Wao, Y., Yang, L., Zhang, Y., Li, D., 2015. Effect of microalgae supplementation on the silage quality and anaerobic digestion performance of Manyflower silvergrass. Bioresour. Technol. 189, 334–340. https://doi.org/http://dx.doi.org/10.1016/j.biortech.2015.04.029

Li, M., Nian, R., Xian, M., Zhang, H., 2018. Metabolic engineering for the production of isoprene and isopentenol by Escherichia coli. Appl. Microbiol. Biotechnol. 102, 7725–7738. https://doi.org/10.1007/s00253-018-9200-5

Li, Q., Lauber, C.L., Czarnecki-maulden, G., Pan, Y., 2017. Effects of the Dietary Protein and Carbohydrate Ratio on Gut Microbiomes in Dogs of Different Body Conditions 8, 1–14.

Liang, Y., 2013. SIMULATION OF FLOW, MASS TRANSFER AND BIO-CHEMICAL REACTIONS IN ANAEROBIC DIGESTION. J. Chem. Inf. Model. 53, 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Lichtfouse, E., Schwarzbauer, J., Robert, D., 2015. CO2 Sequestration, Biofuels and Depollution. Environ. Chem. a Sustain. World 5, 233–274. https://doi.org/10.1007/978-94-007-2442-6

Lin, P.Y., Whang, L.M., Wu, Y.R., Ren, W.J., Hsiao, C.J., Li, S.L., Chang, J.S., 2007. Biological hydrogen production of the genus Clostridium: Metabolic study and mathematical model simulation. Int. J. Hydrogen Energy 32, 1728–1735. https://doi.org/10.1016/j.ijhydene.2006.12.009

Liu, X., Xu, Q., Wang, D., Wu, Y., Yang, Q., Liu, Y., Wang, Q., Li, X., Li, H., Zeng, G., Yang, G., 2019. Unveiling the mechanisms of how cationic polyacrylamide affects short-chain fatty acids accumulation during long-term anaerobic fermentation of waste activated sludge. Water Res. 155, 142–151. https://doi.org/10.1016/j.watres.2019.02.036

Lorv, J.S.H., Rose, D.R., Glick, B.R., 2014. Bacterial ice crystal controlling proteins. Scientifica (Cairo). 2014, 976895. https://doi.org/10.1155/2014/976895

Louis, R., 2016. The role of additives on anaerobic digestion: A review. Renew. Sustain. Energy Rev. 58, 1486–1499. https://doi.org/10.1016/j.str.2014.12.012

Lozupone, C., Knight, R., 2005. UniFrac : a New Phylogenetic Method for Comparing Microbial Communities. Appl. Environ. Microbiol. 71, 8228–8235. https://doi.org/10.1128/AEM.71.12.8228

Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J., Knight, R., 2011. UniFrac: An effective distance

metric for microbial community comparison. ISME J. 5, 169–172. https://doi.org/10.1038/ismej.2010.133

Mangan, D., Liadova, A., Ivory, R., McCleary, B. V., 2016. Novel approaches to the automated assay of β-glucanase and lichenase activity. Carbohydr. Res. 435, 162–172. https://doi.org/10.1016/j.carres.2016.10.006

Mangos, T.J., Haas, M.J., 1996. Enzymatic Determination of Methanol with Alcohol Oxidase, Peroxidase, and the Chromogen 2,2'-Azinobis(3-ethylbenzthiazoline-6-sulfonic acid) and Its Application to the Determination of the Methyl Ester Content of Pectins. J. Agric. Food Chem. 44, 2977–2981. https://doi.org/10.1021/jf960274z

Manyi-Loh, C.E., Mamphweli, S.N., Meyer, E.L., Okoh, A.I., Makaka, G., Simon, M., 2013. Microbial anaerobic digestion (bio-digesters) as an approach to the decontamination of animal wastes in pollution control and the generation of renewable energy. Int. J. Environ. Res. Public Health 10, 4390–4417. https://doi.org/10.3390/ijerph10094390

Mao, C., Feng, Y., Wang, X., Ren, G., 2015. Review on research achievements of biogas from anaerobic digestion. Renew. Sustain. Energy Rev. 45, 540–555. https://doi.org/10.1016/j.rser.2015.02.032

Marcelino, V.R., Clausen, P.T.L.C., Buchmann, J.P., Wille, M., Iredell, J.R., Meyer, W., Lund, O., Sorrell, T.C., Holmes, E.C., 2019. CCMetagen: comprehensive and accurate identification of eukaryotes and prokaryotes in metagenomic data. bioRxiv 641332. https://doi.org/10.1101/641332

Marchesi, J.R., Sato, T., Weightman, A.J., Martin, A., Fry, J.C., Hiom, S.J., Wade, W.G., Martin, T. a, 1998. Design and Evaluation of Useful Bacterium-Specific PCR Primers That Amplify Genes Coding for Bacterial 16S rRNA Design and Evaluation of Useful Bacterium-Specific PCR Primers That Amplify Genes Coding for Bacterial 16S rRNA 64, 795–799.

Martijn, J., Lind, A.E., Spiers, I., Juzokaite, L., Bunikis, I., Pettersson, O.V., Ettema, T.J.G., 2017. Amplicon sequencing of the 16S-ITS-23S rRNA operon with long-read technology for improved phylogenetic classification of uncultured prokaryotes. bioRxiv 234690. https://doi.org/10.1101/234690

Masuda, Y., Itoh, H., Shiratori, Y., Senoo, K., 2018. Metatranscriptomic insights into microbial consortia driving methane metabolism in paddy soils. Soil Sci. Plant Nutr. 64, 455–464. https://doi.org/10.1080/00380768.2018.1457409

Maus, I., Cibis, K.G., Bremges, A., Stolze, Y., Wibberg, D., Tomazetto, G., Blom, J., Sczyrba, A., König, H., Pühler, A., Schlüter, A., 2016a. Genomic characterization of Defluviitoga tunisiensis L3, a key hydrolytic bacterium in a thermophilic biogas plant and its abundance as determined by metagenome fragment recruitment. J. Biotechnol. 1–11. https://doi.org/10.1016/j.jbiotec.2016.05.001

Maus, I., Koeck, D.E., Cibis, K.G., Hahnke, S., Kim, Y.S., Langer, T., Kreubel, J., Erhard, M., Bremges, A., Off, S., Stolze, Y., Jaenicke, S., Goesmann, A., Sczyrba, A., Scherer, P., König, H., Schwarz, W.H., Zverlov, V. V., Liebl, W., Pühler, A., Schlüter, A., Klocke, M., 2016b. Unraveling the microbiome of a thermophilic biogas plant by metagenome and metatranscriptome analysis complemented by characterization of bacterial and archaeal isolates. Biotechnol. Biofuels 9, 1–28. https://doi.org/10.1186/s13068-016-0581-3

Meneghine, A.K., Nielsen, S., Varani, A.M., Thomas, T., Alves, L.M.C., 2017. Metagenomic analysis of soil and freshwater from zoo agricultural area with organic fertilization. PLoS One 12, 1–20. https://doi.org/10.1371/journal.pone.0190178

metoffice.gov.uk, 2017. The Great Smog of 1952 [WWW Document]. URL http://www.metoffice.gov.uk/learning/learn-about-the-weather/weather-phenomena/case-studies/great-smog (accessed 1.27.17).

Metsalu, T., Vilo, J., 2015. ClustVis: A web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. Nucleic Acids Res. 43, W566–W570. https://doi.org/10.1093/nar/gkv468

Moestedt, J., Nordell, E., Shakeri Yekta, S., Lundgren, J., Mart??, M., Sundberg, C., Ejlertsson, J., Svensson, B.H., Bj??rn, A., 2015. Effects of trace element addition on process stability during anaerobic co-digestion of OFMSW and slaughterhouse waste. Waste Manag. 47, 11–20. https://doi.org/10.1016/j.wasman.2015.03.007

Möller, K., Müller, T., 2012. Effects of anaerobic digestion on digestate nutrient availability and crop growth: A review. Eng. Life Sci. 12, 242–257. https://doi.org/10.1002/elsc.201100085

Montalbo-Lomboy, M., Khanal, S.K., van Leeuwen, J. (Hans), Raj Raman, D., Dunn, L., Grewell, D., 2010. Ultrasonic pretreatment of corn slurry for saccharification: A comparison of batch and continuous systems. Ultrason. Sonochem. 17, 939–946. https://doi.org/10.1016/j.ultsonch.2010.01.013

Montalvo, S., Guerrero, L., Borja, R., S??nchez, E., Mil??n, Z., Cort??s, I., Angeles de la la Rubia, M., 2012. Application of natural zeolites in anaerobic digestion processes: A review. Appl. Clay Sci. 58, 125–133. https://doi.org/10.1016/j.clay.2012.01.013

Montgomery, L., Bochmann, G., 2014. Pretreatment of feedstock for enhanced biogas production. IEA Bioenergy 24.

Mshandete, A., Kivaisi, A., Rubindamayugi, M., Mattiasson, B., 2004. Anaerobic batch co-digestion of sisal pulp and fish wastes. Bioresour. Technol. 95, 19–24. https://doi.org/10.1016/j.biortech.2004.01.011

Mueller, M., 2015. Final Project Report Air Quality Issues Related To Using Biogas From Anaerobic.

Nichols, D., Cahoon, N., Trakhtenberg, E.M., Pham, L., Mehta, a., Belanger, a., Kanigan, T., Lewis, K., Epstein, S.S., 2010. Use of Ichip for High-Throughput In Situ Cultivation of "Uncultivable" Microbial Species. Appl. Environ. Microbiol. 76, 2445–2450. https://doi.org/10.1128/AEM.01754-09

North Yorkshire Council, n.d. Composting [WWW Document]. URL https://www.northyorks.gov.uk/composting (accessed 8.29.19).

Northern Ireland Environment Agency, 2012. Quality Protocol: Compost - End of waste criteria for the production and use of quality compost from source-segregated biodegradable waste. Quality.

Northup, D.E., Melim, L. a, Spilde, M.N., Hathaway, J.J.M., Garcia, M.G., Moya, M., Stone, F.D., Boston, P.J., Dapkevicius, M.L.N.E., Riquelme, C., 2011. Lava cave microbial communities within mats and

secondary mineral deposits: implications for life detection on other planets. Astrobiology 11, 601–18. https://doi.org/10.1089/ast.2010.0562

Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., 2017. MetaSPAdes: A new versatile metagenomic assembler. Genome Res. 27, 824–834. https://doi.org/10.1101/gr.213959.116

O'Mahony, L., Mccarthy, J., Kelly, P., Hurley, G., Luo, F., Chen, K., O'Sullivan, G.C., Kiely, B., Collins, J.K., Shanahan, F., Quigley, E.M.M., 2005. Lactobacillus and Bifidobacterium in irritable bowel syndrome: Symptom responses and relationship to cytokine profiles. Gastroenterology 128, 541–551. https://doi.org/10.1053/j.gastro.2004.11.050

Orzi, V., Scaglia, B., Lonati, S., Riva, C., Boccasile, G., Alborali, G.L., Adani, F., 2015. The role of biological processes in reducing both odor impact and pathogen content during mesophilic anaerobic digestion. Sci. Total Environ. 526, 116–126. https://doi.org/10.1016/j.scitotenv.2015.04.038

Ostrem, K., 2004. Greening waste: anaerobic digestion for treating the organic fraction of municipal solid waste.

Pakarinen, A., Zhang, J., Brock, T., Maijala, P., Viikari, L., 2012. Enzymatic accessibility of fiber hemp is enhanced by enzymatic or chemical removal of pectin. Bioresour. Technol. 107, 275–281. https://doi.org/10.1016/j.biortech.2011.12.101

Palatsi, J., Illa, J., Prenafeta-Boldú, F.X., Laureni, M., Fernandez, B., Angelidaki, I., Flotats, X., 2010. Long-chain fatty acids inhibition and adaptation process in anaerobic thermophilic digestion: Batch tests, microbial community structure and mathematical modelling. Bioresour. Technol. 101, 2243–2251. https://doi.org/10.1016/j.biortech.2009.11.069

Pan, H., Zhang, Y., He, G.X., Katagori, N., Chen, H., 2014. A comparison of conventional methods for the quantification of bacterial cells after exposure to metal oxide nanoparticles. BMC Microbiol. 14. https://doi.org/10.1186/s12866-014-0222-6

Park, H., 2012. Monitoring bacterial community structure and variability in time scale in full-scale anaerobic digesters 1893–1905. https://doi.org/10.1039/c2em10958a

Pasztor, I., Thury, P., Pulai, J., 2009. Chemical oxygen demand fractions of municipal wastewater for modeling of wastewater treatment. Int. J. Environ. Sci. Technol. 6, 51–56. https://doi.org/10.1007/BF03326059

Paulo, L.M., Stams, A.J.M., Sousa, D.Z., 2015. Methanogens, sulphate and heavy metals: a complex system. Rev. Environ. Sci. Biotechnol. 14, 537–553. https://doi.org/10.1007/s11157-015-9387-1

Paulo, P.L., Villa, G., Van Lier, J.B., Lettinga, G., 2003. The Anaerobic Conversion of Methanol under Thermophilic Conditions: pH and Bicarbonate Dependence. J. Biosci. Bioeng. 96, 213–218. https://doi.org/10.1016/S1389-1723(03)80184-6

PFPI Partnership for Policy Integrety, 2011. Air pollution from biomass energy.

Pianta, C., Passos, D.T., Hepp, D., Oliveira, S.J. De, 2007. Isolation of Arcobacter spp from the milk of dairy cows in Brazil. Ciência Rural 37, 171–174. https://doi.org/10.1590/S0103-84782007000100027

Picone, N., Op den Camp, H.J., 2019. Role of rare earth elements in methanol oxidation. Curr. Opin. Chem. Biol. 49, 39–44. https://doi.org/10.1016/j.cbpa.2018.09.019

Pinto, A., 2014. Secure because Math: A deep-dive on Machine Learning- based Monitoring. Black Hat Briefings 25, 1–11. https://doi.org/10.1101/gr.215087.116.Freely

Piwosz, K., Shabarova, T., Tomasch, J., Šimek, K., Kopejtka, K., Kahl, S., Pieper, D.H., Koblížek, M., 2018. Determining lineage-specific bacterial growth curves with a novel approach based on amplicon reads normalization using internal standard (ARNIS). ISME J. 1–15. https://doi.org/10.1038/s41396-018-0213-y

Pomaznoy, M., Ha, B., Peters, B., 2018. GOnet: A tool for interactive Gene Ontology analysis. BMC Bioinformatics 19, 1–8. https://doi.org/10.1186/s12859-018-2533-3

Puyuelo, B., Ponsá, S., Gea, T., Sánchez, A., 2011. Determining C/N ratios for typical organic wastes using biodegradable fractions. Chemosphere 85, 653–659. https://doi.org/10.1016/j.chemosphere.2011.07.014

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. 41, 590–596. https://doi.org/10.1093/nar/gks1219

Rajagopal, R., Massé, D.I., Singh, G., 2013. A critical review on inhibition of anaerobic digestion process by excess ammonia. Bioresour. Technol. 143, 632–641. https://doi.org/10.1016/j.biortech.2013.06.030

Rampelotto, P., 2014. Polar Microbiology: Recent Advances and Future Perspectives. Biology (Basel). 3, 81–84. https://doi.org/10.3390/biology3010081

Ramsden, D.K., McKay, K., 1986. Degradation of polyacrylamide in aqueous solution induced by chemically generated hydroxyl radicals: Part I-Fenton's reagent. Polym. Degrad. Stab. 14, 217–229. https://doi.org/10.1016/0141-3910(86)90045-5

Rasit, N., Idris, A., Harun, R., Wan Ab Karim Ghani, W.A., 2015. Effects of lipid inhibition on biogas production of anaerobic digestion from oily effluents and sludges: An overview. Renew. Sustain. Energy Rev. 45, 351–358. https://doi.org/10.1016/j.rser.2015.01.066

Rasouli, Z., Valverde-Pérez, B., D'Este, M., De Francisci, D., Angelidaki, I., 2018. Nutrient recovery from industrial wastewater as single cell protein by a co-culture of green microalgae and methanotrophs. Biochem. Eng. J. 134, 129–135. https://doi.org/10.1016/j.bej.2018.03.010

Rico-Martínez, R., Snell, T.W., Shearer, T.L., 2013. Synergistic toxicity of Macondo crude oil and dispersant Corexit 9500A ?? to the Brachionus plicatilis species complex (Rotifera). Environ. Pollut. 173, 5–10. https://doi.org/10.1016/j.envpol.2012.09.024

Ritala, A., Häkkinen, S.T., Toivari, M., Wiebe, M.G., 2017. Single cell protein-state-of-the-art, industrial landscape and patents 2001-2016. Front. Microbiol. 8. https://doi.org/10.3389/fmicb.2017.02009

Robson, J., Alessi, A., Bochiwal, C., Malley, C.O., Chong, J.P.J., 2016. Biomethane as an Energy Source, in: Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids: Production of Fuels and Chemicals. pp. 1–12. https://doi.org/10.1007/978-3-319-31421-1

Rolfe, R.D., Hentges, D.J., Campbell, B.J., Barrett, J.T., 1978. Factors Related to the Oxygen Tolerance of Anaerobic Bacteria 36, 306–313.

Russell, N.J., 1997. Psychrophilic bacteria—Molecular adaptations of membrane lipids. Comp. Biochem. Physiol. Part A Physiol. 118, 489–493. https://doi.org/10.1016/S0300-9629(97)87354-9

Saady, N.M.C., Massé, D.I., 2013. Psychrophilic anaerobic digestion of lignocellulosic biomass: a characterization study. Bioresour. Technol. 142, 663–71. https://doi.org/10.1016/j.biortech.2013.05.089

Sakamoto, M., Kitahara, M., Benno, Y., 2007. Parabacteroides johnsonii sp. nov., isolated from human faeces. Int. J. Syst. Evol. Microbiol. 57, 293–296. https://doi.org/10.1099/ijs.0.64588-0

Schimming, S., Schwarz, W.H., Staudenbauer, W.L., 1991. Properties of a thermoactive β-1,3-1,4-glucanase (lichenase) from Clostridium thermocellum expressed in Escherichia coli. Biochem. Biophys. Res. Commun. 177, 447–452. https://doi.org/10.1016/0006-291X(91)92004-4

Schröder, C., Selig, M., Schönheit, P., 1994. Glucose fermentation to acetate, CO2 and H2 in the anaerobic hyperthermophilic eubacterium Thermotoga maritima: involvement of the Embden-Meyerhof pathway. Arch. Microbiol. 161, 460–470. https://doi.org/10.1007/BF00307766

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., Ragg, T., 2006. The RIN: An RNA integrity number for assigning integrity values to RNA measurements. BMC Mol. Biol. 7, 1–14. https://doi.org/10.1186/1471-2199-7-3

Scotch whisky exports hit record high [WWW Document], 2018. URL https://www.bbc.co.uk/news/uk-scotland-scotland-business-43009106 (accessed 6.7.18).

Scott, R.I., Williams, T.N., Lloyd, D., 1983. OXYGEN SENSITIVITY OF METHANOGENESIS IN RUMEN AND ANAEROBIC DIGESTER POPULATIONS USING MASS SPECTROMETRY. Biotechnol. Lett. 5, 375–380.

Shen, W., Le, S., Li, Y., Hu, F., 2016. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file

manipulation. PLoS One 11, 1–10. https://doi.org/10.1371/journal.pone.0163962

Sheth, R.U., Li, M., Jiang, W., Sims, P.A., Leong, K.W., Wang, H.H., 2019. Spatial metagenomic characterization of microbial biogeography in the gut. Nat. Biotechnol. 37, 877–883. https://doi.org/10.1038/s41587-019-0183-2

Show, K.Y., Wang, Y., Foong, S.F., Tay, J.H., 2004. Accelerated start-up and enhanced granulation in upflow anaerobic sludge blanket reactors. Water Res. 38, 2292–2303. https://doi.org/10.1016/j.watres.2004.01.039

Sieber, J.R., McInerney, M.J., Gunsalus, R.P., 2012. Genomic Insights into Syntrophy: The Paradigm for Anaerobic Metabolic Cooperation. Annu. Rev. Microbiol. 66, 429–452. https://doi.org/10.1146/annurev-micro-090110-102844

Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J., Timp, W., 2017. Detecting DNA cytosine methylation using nanopore sequencing. Nat. Publ. Gr. 14, 407–410. https://doi.org/10.1038/nmeth.4184

Simsek, M., Adnan, H., 2000. Effect of single mismatches at 3'-end of primers on polymerase chain reaction. Sultan Qaboos Univ. Med. J. 2, 11–14.

Sipos, R., Székely, A.J., Palatinszky, M., Révész, S., Márialigeti, K., Nikolausz, M., 2007. Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. FEMS Microbiol. Ecol. 60, 341–350. https://doi.org/10.1111/j.1574-6941.2007.00283.x

Sizova, M. V, Panikov, N.S., Tourova, T.P., Flanagan, P.W., 2003. Isolation and characterization of oligotrophic acido-tolerant methanogenic consortia from a Sphagnum peat bog. FEMS Microbiol. Ecol. 45, 301–315. https://doi.org/10.1016/S0168-6496(03)00165-X

Smets, W., Leff, J.W., Bradford, M.A., McCulley, R.L., Lebeer, S., Fierer, N., 2016. A method for simultaneous measurement of soil bacterial abundances and community composition via 16S rRNA gene sequencing. Soil Biol. Biochem. 96, 145–151. https://doi.org/10.1016/j.soilbio.2016.02.003

Smith, A.L., Skerlos, S.J., Raskin, L., 2015. Anaerobic membrane bioreactor treatment of domestic wastewater at psychrophilic temperatures ranging from 15 °C to 3 °C. Environ. Sci. Water Res. Technol. 1, 56–64. https://doi.org/10.1039/C4EW00070F

Smith, D.P., Peay, K.G., 2014. Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. PLoS One 9. https://doi.org/10.1371/journal.pone.0090234

Spey Syrup [WWW Document], 2019. URL https://www.kwalternativefeeds.co.uk/products/view-products/spey-syrup/ (accessed 9.19.19).

Stämmler, F., Gläsner, J., Hiergeist, A., Holler, E., Weber, D., Oefner, P.J., Gessner, A., Spang, R., 2016. Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. Microbiome 4, 1–13. https://doi.org/10.1186/s40168-016-0175-0

Stitt, E.H., 2002. Alternative multiphase reactors for fine chemicals: A world beyond stirred tanks? Chem. Eng. J. 90, 47–60. https://doi.org/10.1016/S1385-8947(02)00067-0

Stolyar, S., Van Dien, S., Hillesland, K.L., Pinel, N., Lie, T.J., Leigh, J. a, Stahl, D. a, 2007. Metabolic modeling of a mutualistic microbial community. Mol. Syst. Biol. 3, 92. https://doi.org/10.1038/msb4100131

Street, B., 2006. The Power of Scotland : Cutting Carbon with Scotland ' s Renewable Energy.

Suanon, F., Sun, Q., Mama, D., Li, J., Dimon, B., Yu, C.-P., 2015. Effect of nanoscale zero-valent iron and magnetite (Fe3O4) on the fate of metals during anaerobic digestion of sludge. Water Res. 88, 897–903. https://doi.org/10.1016/j.watres.2015.11.014

Suzuki, M.T., Giovannoni, S.J., 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by B ia s C a u s e d b y T e m p la te A n n e a lin g in th e A m p lifi c a tio n o f M ix tu r e s o f 1 6 S r R N A G e n e s b y P C R † 62, 2–8.

Szabó, I., Szoboszlay, S., Kriszt, B., Háhn, J., Harkai, P., Baka, E., Táncsics, A., Kaszab, E., Privler, Z., Kukolya, J., 2011. Olivibacter oleidegradans sp. nov., a hydrocarbondegrading bacterium isolated from a biofilter cleanup facility on a hydrocarbon-contaminated site. Int. J. Syst. Evol. Microbiol.

61, 2861–2865. https://doi.org/10.1099/ijs.0.026641-0

Takai, K., Kobayashi, H., Nealson, K.H., Horikoshi, K., 2003. Deferribacter desulfuricans sp. nov., a novel sulfur-, nitrate- and arsenate-reducing thermophile isolated from a deep-sea hydrothermal vent. Int. J. Syst. Evol. Microbiol. 53, 839–846. https://doi.org/10.1099/ijs.0.02479-0

Tao, B., Alessi, A.M., Zhang, Y., Chong, J.P.J., Heaven, S., Banks, C.J., 2019. Simultaneous biomethanisation of endogenous and imported CO 2 in organically loaded anaerobic digesters. Appl. Energy 247, 670–681. https://doi.org/10.1016/j.apenergy.2019.04.058

Tenenbaum, D.J., 2008. Food vs. fuel diversion of crops could cause more hunger. Environ. Health Perspect. 116, 254–257. https://doi.org/10.1289/ehp.116-a254

Thanh, P.M., Ketheesan, B., Yan, Z., Stuckey, D., 2015. Trace metal speciation and bioavailability in anaerobic digestion: A review. Biotechnol. Adv. https://doi.org/10.1016/j.biotechadv.2015.12.006

The Energy and Climate Change Committee, 2016. 2020 renewable heat and transport targets.

Themelis, N.J., 2002. Anaerobic Digestion of Biodegradable Organics in Municipal Solid Wastes. Found. Sch. Eng. Appl. Sci. Columbia Univ. 1–56. https://doi.org/10.1016/j.biotechadv.2010.10.005

Ting, T., Zhao, W., Sun, Y., Peng, S., Yue, W., Tang, Q., Kida, K., 2019. Succession of Total and Active Microbial Community During the Composting of Anaerobic Digested Residue. Waste and Biomass Valorization. https://doi.org/10.1007/s12649-019-00779-7

Tourlousse, D.M., Yoshiike, S., Ohashi, A., Matsukura, S., Noda, N., Sekiguchi, Y., 2017. Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. Nucleic Acids Res. 45, e23. https://doi.org/10.1093/nar/gkw984

Treu, L., Campanaro, S., Kougias, P.G., Zhu, X., Angelidaki, I., 2016. Untangling the Effect of Fatty Acid Addition at Species Level Revealed Different Transcriptional Responses of the Biogas Microbial Community Members. Environ. Sci. Technol. 50, 6079–6090. https://doi.org/10.1021/acs.est.6b00296

Ulger-Toprak, N., Summanen, P.H., Liu, C., Rowlinson, M.-C., Finegold, S.M., 2010. Gemella asaccharolytica sp. nov., isolated from human clinical specimens. Int. J. Syst. Evol. Microbiol. 60, 1023–6. https://doi.org/10.1099/ijs.0.001966-0

Unal, B., Perry, V.R., Sheth, M., Gomez-Alvarez, V., Chin, K.-J., Nüsslein, K., 2012. Trace elements affect methanogenic activity and diversity in enrichments from subsurface coal bed produced water. Front. Microbiol. 3, 175. https://doi.org/10.3389/fmicb.2012.00175

Uzal, N., Gökçay, C.F., Demirer, G.N., 2003. Sequential (anaerobic/aerobic) biological treatment of malt whisky wastewater. Process Biochem. 39, 279–286. https://doi.org/10.1016/S0032-9592(03)00071-2

Vanwonterghem, I., Jensen, P.D., Ho, D.P., Batstone, D.J., Tyson, G.W., 2014. Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. Curr. Opin. Biotechnol. 27, 55–64. https://doi.org/10.1016/j.copbio.2013.11.004

Vanwonterghem, I., Jensen, P.D., Rabaey, K., Tyson, G.W., 2015. Temperature and solids retention time control microbial population dynamics and volatile fatty acid production in replicated anaerobic digesters. Sci. Rep. 5, 8496. https://doi.org/10.1038/srep08496

Vavilin, V.A., Fernandez, B., Palatsi, J., Flotats, X., 2008. Hydrolysis kinetics in anaerobic degradation of particulate organic material: An overview. Waste Manag. 28, 939–951. https://doi.org/10.1016/j.wasman.2007.03.028

Venkataraman, A., Parlov, M., Hu, P., Schnell, D., Wei, X., Tiesman, J.P., 2018. Spike-in genomic DNA for validating performance of metagenomics workflows. Biotechniques 65, 315–321.

Viana, M.B., Freitas, A. V., Leitão, R.C., Pinto, G.A.S., Santaella, S.T., 2012. Anaerobic digestion of crude glycerol: a review. Environ. Technol. Rev. 1, 81–92. https://doi.org/10.1080/09593330.2012.692723

Vu, H.N., Subuyuj, G.A., Vijayakumar, S., Good, N.M., Martinez-Gomez, N.C., Skovran, E., 2016. Lanthanide-dependent regulation of methanol oxidation systems in Methylobacterium extorquens AM1 and their contribution to methanol growth. J. Bacteriol. 198, 1250–1259.

https://doi.org/10.1128/JB.00937-15

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9. https://doi.org/10.1371/journal.pone.0112963

Wang, D., Liu, X., Zeng, G., Zhao, J., Liu, Y., Wang, Q., Chen, F., Li, X., Yang, Q., 2018. Understanding the impact of cationic polyacrylamide on anaerobic digestion of waste activated sludge. Water Res. 130, 281–290. https://doi.org/10.1016/j.watres.2017.12.007

Wang, L., Wang, Q., Cai, W., Sun, X., 2012. Influence of mixing proportion on the solid-state anaerobic co-digestion of distiller ' s grains and food waste. Biosyst. Eng. 112, 130–137. https://doi.org/10.1016/j.biosystemseng.2012.03.006

Wang, Q., Kuninobu, M., Ogawa, H.I., Kato, Y., 1999. Degradation of volatile fatty acids in highly efficient anaerobic digestion. Biomass and Bioenergy 16, 407–416. https://doi.org/10.1016/S0961-9534(99)00016-1

Wang, Y., Du, R., Yu, T., 2013. Systematical method for polyacrylamide and residual acrylamide detection cosmetic surgery products and example application. Sci. Justice 53, 350–357. https://doi.org/10.1016/j.scijus.2013.04.007

Wang, Y., Hu, H., Li, X., 2015. MBBC: an efficient approach for metagenomic binning based on clustering. BMC Bioinformatics 16, 36. https://doi.org/10.1186/s12859-015-0473-8

Ward, A.J., Hobbs, P.J., Holliman, P.J., Jones, D.L., 2008. Optimisation of the anaerobic digestion of agricultural resources. Bioresour. Technol. 99, 7928–7940. https://doi.org/10.1016/j.biortech.2008.02.044

Ward, A.J., Lewis, D.M., Green, F.B., 2014. Anaerobic digestion of algae biomass: A review. Algal Res. 5, 204–214. https://doi.org/10.1016/j.algal.2014.02.001

Watanabe, K., Kodama, Y., Harayama, S., 2001. Design and evaluation of PCR primers to amplify

bacterial 16S ribosomal DNA fragments used for community fingerprinting. J. Microbiol. Methods 44, 253–262. https://doi.org/10.1016/S0167-7012(01)00220-2

Wei, J., Wang, J., 2013. Enhanced hydrolysis and methane yield by applying microaeration pretreatment to the anaerobic co-digestion of brown water and food waste. Waste Manag. 33, 813–819. https://doi.org/10.1016/j.wasman.2012.11.013

Weisburg, W.G., Barns, S.M., Pelletier, D.A., Lane, D.J., 1991. 16S ribosomal DNA amplification for phylogenetic study. J. Bacteriol. 173, 697–703. https://doi.org/n.a.

Wen, Q., Chen, Z., Zhao, Y., Zhang, H., Feng, Y., 2010. Biodegradation of polyacrylamide by bacteria isolated from activated sludge and oil-contaminated soil. J. Hazard. Mater. 175, 955–959. https://doi.org/10.1016/j.jhazmat.2009.10.102

Wexler, H.M., 2007. Bacteroides: The good, the bad, and the nitty-gritty. Clin. Microbiol. Rev. 20, 593–621. https://doi.org/10.1128/CMR.00008-07

Whisky distillery turns co-products into renewable energy [WWW Document], 2017. URL https://www.thesourcemagazine.org/whisky-distillery-turns-co-products-renewable-energy/

Wielgosiński, G., Łechtańska, P., Namiecińska, O., 2015. Emission of some pollutants from biomass combustion in comparison to hard coal combustion. J. Energy Inst. 1–10. https://doi.org/10.1016/j.joei.2016.06.005

Wilkins, D., Lu, X.Y., Shen, Z., Chen, J., Lee, P.K.H., 2015. Pyrosequencing of mcrA and archaeal 16s rRNA genes reveals diversity and substrate preferences of methanogen communities in anaerobic digesters. Appl. Environ. Microbiol. 81, 604–613. https://doi.org/10.1128/AEM.02566-14

Williams, A.., 1994. The permeability and porosity of grass silage as affected by dry matter. J. Agric. Eng. Res. 59, 133–140.

Wiśniewska, P., Śliwińska, M., Dymerski, T., Wardencki, W., Namieśnik, J., 2015. The Analysis of Vodka: A Review Paper. Food Anal. Methods. https://doi.org/10.1007/s12161-015-0089-7

Wongnate, T., Ragsdale, S.W., 2015. The reaction mechanism of methyl-coenzyme M reductase. J. Biol. Chem. 290, 9322–9334. https://doi.org/10.1074/jbc.M115.636761

Wu, M.L., Wessels, H.J.C.T., Pol, A., Op den Camp, H.J.M., Jetten, M.S.M., van Niftrik, L., Keltjens, J.T., 2015. XoxF-type methanol dehydrogenase from the anaerobic methanotroph "Candidatus Methylomirabilis oxyfera." Appl. Environ. Microbiol. 81, 1442–1451. https://doi.org/10.1128/AEM.03292-14

Wu, T., Lin, C., Chang, C., Lin, T., Martel, J., Ko, Y., Ojcius, D.M., Lu, C., 2019. Gut commensal Parabacteroides goldsteinii plays a predominant role in the anti-obesity effects of polysaccharides isolated from Hirsutella sinensis. Gut microbiota 68, 248–262. https://doi.org/10.1136/gutjnl-2017-315458

Xia, Y., Cai, L., Zhang, T., Fang, H.H.P., 2012. Effects of substrate loading and co-substrates on thermophilic anaerobic conversion of microcrystalline cellulose and microbial communities revealed using high-throughput sequencing. Int. J. Hydrogen Energy 37, 13652–13659. https://doi.org/10.1016/j.ijhydene.2012.02.079

Xia, Y., Wang, Y., Fang, H.H.P., Jin, T., Zhong, H., Zhang, T., 2014. Thermophilic microbial cellulose decomposition and methanogenesis pathways recharacterized by metatranscriptomic and metagenomic analysis. Sci. Rep. 4, 1–9. https://doi.org/10.1038/srep06708

Xiao, Y., De Araujo, C., Sze, C.C., Stuckey, D.C., 2015. Controlling a toxic shock of pentachlorophenol (PCP) to anaerobic digestion using activated carbon addition. Bioresour. Technol. 181, 303–311. https://doi.org/10.1016/j.biortech.2015.01.080

Xing, B.-S., Guo, Q., Jiang, X.-Y., Chen, Q.-Q., He, M.-M., Wu, L.-M., Jin, R.-C., 2016. Long-term starvation and subsequent reactivation of anaerobic ammonium oxidation (anammox) granules. Chem. Eng. J. 287, 575–584. https://doi.org/10.1016/j.cej.2015.11.090

Xing, W., Ngo, H.H., Guo, W., Wu, Z., Nguyen, T.T., Cullum, P., Listowski, A., Yang, N., 2010. Enhancement of the performance of anaerobic fluidized bed bioreactors (AFBBRs) by a new starch based flocculant. Sep. Purif. Technol. 72, 140–146. https://doi.org/10.1016/j.seppur.2010.01.015

Xiong, B., Loss, R.D., Shields, D., Pawlik, T., Hochreiter, R., Zydney, A.L., 2018. Polyacrylamide degradation and its implications in environmental systems Poly-acrylamido-2-. npj Clean Water. https://doi.org/10.1038/s41545-018-0016-8

Xu, F., Li, Y., Wang, Z.W., 2015. Mathematical modeling of solid-state anaerobic digestion. Prog. Energy Combust. Sci. https://doi.org/10.1016/j.pecs.2015.09.001

Xu, L., Peng, S., Dong, D., Wang, C., Fan, W., Cao, Y., Huang, F., Wang, J., Yue, Z., 2019. Performance and microbial community analysis of dry anaerobic co-digestion of rice straw and cow manure with added limonite. Biomass and Bioenergy 126, 41–46. https://doi.org/10.1016/j.biombioe.2019.04.026

Y. Verastegui, J. Cheng, K. Engel, D. Kolczynski, S. Mortimer, J. Lavigne, J. Montalibet, T. Romantsov, M. Hall, A B. J. McConkey, D. R. Rose, J. J. Tomashek, B. R. Scott, T. C. Charles, J.D.N., 2014. Multisubstrate Isotope Labeling and Metagenomic Analysis of Active. MBio 5, 1–12. https://doi.org/10.1128/mBio.01157-14.Editor

Yamada, C., Kato, S., Ueno, Y., Ishii, M., Igarashi, Y., 2015. Conductive iron oxides accelerate thermophilic methanogenesis from acetate and propionate. J. Biosci. Bioeng. 119, 678–682. https://doi.org/10.1016/j.jbiosc.2014.11.001

Yenigün, O., Demirel, B., 2013. Ammonia inhibition in anaerobic digestion: A review. Process Biochem. 48, 901–911. https://doi.org/10.1016/j.procbio.2013.04.012

Yoo, D.H., Lee, B.H., Chang, P.S., Hyeon, G.L., Yoo, S.H., 2007. Improved quantitative analysis of oligosaccharides from lichenase-hydrolyzed water-soluble barley β-glucans by high-performance anion-exchange chromatography. J. Agric. Food Chem. 55, 1656–1662. https://doi.org/10.1021/jf062603l

Yu, H.Q., Tay, J.H., Fang, H.H.P., 2001. The roles of calcium in sludge granulation during UASB reactor start-up. Water Res. 35, 1052–1060. https://doi.org/10.1016/S0043-1354(00)00345-6

Yu, H.W., Samani, Z., Hanson, A., Smith, G., 2002. Energy recovery from grass using two-phase anaerobic digestion. Waste Manag. 22, 1–5. https://doi.org/10.1016/S0956-053X(00)00121-5

Yu, W., Quek, W.P., Li, C., Gilbert, R.G., Fox, G.P., 2018. Effects of the starch molecular structures in barley malts and rice adjuncts on brewing performance. Fermentation 4. https://doi.org/10.3390/fermentation4040103

Zakrzewski, M., Goesmann, A., Jaenicke, S., Jünemann, S., Eikmeyer, F., Szczepanowski, R., Al-Soud, W.A., Sørensen, S., Pühler, A., Schlüter, A., 2012. Profiling of the metabolically active community from a production-scale biogas plant by means of high-throughput metatranscriptome sequencing. J. Biotechnol. 158, 248–258. https://doi.org/10.1016/j.jbiotec.2012.01.020

Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, Di., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., Stott, M.B., Nunoura, T., Banfield, J.F., Schramm, A., Baker, B.J., Spang, A., Ettema, T.J.G., 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature 541, 353–358. https://doi.org/10.1038/nature21031

Zhang, D., Zhu, W., Tang, C., Suo, Y., Gao, L., Yuan, X., Wang, X., Cui, Z., 2012. Bioreactor performance and methanogenic population dynamics in a low-temperature (5-18 °C) anaerobic fixed-bed reactor. Bioresour. Technol. 104, 136–43. https://doi.org/10.1016/j.biortech.2011.10.086

Zhang, H., Bruns, M.A., Logan, B.E., 2006. Biological hydrogen production by Clostridium acetobutylicum in an unsaturated flow reactor. Water Res. 40, 728–734. https://doi.org/10.1016/j.watres.2005.11.041

Zhang, L., 2019. Lignocellulosic hydrogen production using dark fermentation by Clostridium lentocellum strain Cel10 newly isolated from Ailuropoda melanoleuca 11179–11185. https://doi.org/10.1039/c9ra01158g

Zhang, L., Loh, K., Wei, J., Zhang, J., 2019a. Bioinformatics analysis of metagenomics data of biogas-producing microbial communities in anaerobic digesters : A review. Renew. Sustain. Energy Rev. 100, 110–126. https://doi.org/10.1016/j.rser.2018.10.021

Zhang, L., Loh, K.C., Lim, J.W., Zhang, J., 2019b. Bioinformatics analysis of metagenomics data of biogas-producing microbial communities in anaerobic digesters: A review. Renew. Sustain. Energy Rev. 100, 110–126. https://doi.org/10.1016/j.rser.2018.10.021

Zhang, Y., Schauer, J.J., Zhang, Y., Zeng, L., Wei, Y., Liu, Y., Shao, M., 2008. Characteristics of particulate carbon emissions from real-world Chinese coal combustion. Environ. Sci. Technol. 42, 5068–5073. https://doi.org/10.1021/es7022576

Zhang, Z., Qu, Y., Li, S., Feng, K., Wang, S., Cai, W., Liang, Y., Li, H., Xu, M., Yin, H., Deng, Y., 2017. Soil bacterial quantification approaches coupling with relative abundances reflecting the changes of taxa. Sci. Rep. 7, 1–11. https://doi.org/10.1038/s41598-017-05260-w

Zhao, J.-S., Spain, J., Thiboutot, S., Ampleman, G., Greer, C., Hawari, J., 2004. Phylogeny of cyclic nitramine-degrading psychrophilic bacteria in marine sediment and their potential role in the natural attenuation of explosives. FEMS Microbiol. Ecol. 49, 349–57. https://doi.org/10.1016/j.femsec.2004.04.008

Zheng, H., Li, D., Stanislaus, M.S., Zhang, N., Zhu, Q., Hu, X., Yang, Y., 2015. Development of a bio-zeolite fixed-bed bioreactor for mitigating ammonia inhibition of anaerobic digestion with extremely high ammonium concentration livestock waste. Chem. Eng. J. 280, 106–114. https://doi.org/10.1016/j.cej.2015.06.024

Zhu, X., Campanaro, S., Treu, L., Kougias, P.G., 2019. Novel ecological insights and functional roles during anaerobic digestion of saccharides unveiled by genome-centric metagenomics. Water Res. 151, 271–279. https://doi.org/10.1016/j.watres.2018.12.041

Žifčáková, L., Větrovský, T., Howe, A., Baldrian, P., 2016. Microbial activity in forest soil reflects the changes in ecosystem properties between summer and winter. Environ. Microbiol. 18, 288–301. https://doi.org/10.1111/1462-2920.13026

Ziganshin, A.M., Liebetrau, J., Pröter, J., Kleinsteuber, S., 2013. Microbial community structure and dynamics during anaerobic digestion of various agricultural waste materials. Appl. Microbiol. Biotechnol. 97, 5161–5174. https://doi.org/10.1007/s00253-013-4867-0

Ziganshin, A.M., Schmidt, T., Scholwin, F., Il'Inskaya, O.N., Harms, H., Kleinsteuber, S., 2011. Bacteria and archaea involved in anaerobic digestion of distillers grains with solubles. Appl. Microbiol. Biotechnol. 89, 2039–2052. https://doi.org/10.1007/s00253-010-2981-9

Zitomer, D.H., Shrout, J.D., 1998. Feasibility and benefits of methanogenesis under oxygen-limited conditions. Waste Manag. 18, 107–116.

Zumstein, E., 2000. Examination of two years of community dynamics in an anaerobic bioreactor using Fluorescence polymerase chain reaction (PCR) single-strand conformation polymorphism analysis 2.