

Culturomics of Anaerobic Sludge

Jack Matthew Murgatroyd

MSc by Research
University of York
Biology
January 2019

Abstract

Ever increasing concerns surrounding climate change are making alternative energy sources a focus of global discussion. Anaerobic digestion holds promise as one method of generating renewable energy with a low carbon footprint. The simple principle that underlies anaerobic digestion is the degradation of organic material by microorganisms to produce digestate and biogas. While this process occurs without human intervention in a variety of environments, there is still large scope for refinement of the process in order to increase its efficiency in industry. A key problem that limits technology advancement is the understanding of the microbiomes within anaerobic digesters. This can be attributed to the fastidious growth requirements of many of the organisms involved and as such, novel methods for culturing must be adopted if we are to advance our understanding. Work undertaken used a number of different growth backgrounds and isolation chips as a novel method for the enrichment of methanogens and anaerobic fungi from anaerobic sludge samples. 16S rRNA gene sequencing technology was used to assign taxonomy to 1,152 samples. A potential novel bacterial species was further investigated using fluorescent *in-situ* hybridisation and Sanger sequencing with a view to demonstrating the efficacy of this culturing method for the isolation and culturing of novel microbes. This alternative approach for the cultivation of multiple microbes in isolation has shown promise in terms of the number of novel species grown. Given that 16 samples were investigated at an individual level and one of these has been identified as a novel species, it is likely that novel species have also been cultured within the other 1,136 samples that have not yet been studied individually. This experiment provides a starting point from which a more streamlined method can be developed or altered to produce different microbial populations.

List of contents

Abstract	ii
List of contents	iii
List of tables	v
List of figures.....	vi
List of equations	vii
Acknowledgements	viii
Declaration.....	ix
Chapter 1: General Introduction	1
1.1 Anaerobic Digestion.....	1
1.1.1 The Industry.....	1
1.1.1.1 Industry growth and development	1
1.1.1.2 Methane as a biofuel	1
1.1.1.3 Wastewater treatment	2
1.1.2 Stages of anaerobic digestion	3
1.1.2.1 Hydrolysis	3
1.1.2.2 Acidogenesis	3
1.1.2.3 Acetogenesis	4
1.1.2.4 Methanogenesis	4
1.1.2.5 Anaerobic bacteria	4
1.2 Novel Species Discovery	7
1.2.1 16S rRNA comparative genomics	7
1.2.1.1 Overview.....	7
1.2.1.2 16S rRNA gene sequence data manipulation	8
1.2.1.3 Alternatives	8
1.2.2 Novel species definition	8
1.2.3 Rate of discovery of genomes and species	11
1.2.4 Isolation Chip.	11
1.3 Aims	13
1.4 Motivation.....	13
Chapter 2: Materials and Methods.....	14
2.1 Isolation and growth.....	14
2.1.1 Sampling	14
2.1.2 Sludge fluid preparation	14
2.1.3 Growth Media	14
2.1.4 Enrichment	15
2.1.5 Isolation Chip Loading.....	16
2.1.6 Culturing	16
2.2 Molecular work	18
2.2.1 Primers.....	18

2.2.2 PCR.....	18
2.2.3 Gel electrophoresis	18
2.2.4 Pooling and purification	19
2.2.4.1 Pooling	19
2.2.4.2 Qubit fluorometry	19
2.2.4.3 Index PCR	20
2.2.5 Library preparation and Illumina sequencing	20
2.3 Sequence data processing	21
2.3.1 Cutadapt	21
2.3.2 DADA2	21
2.4 H ₂ AD-Acetate enriched samples	21
2.4.1 Selection of samples	21
2.4.2 Sanger sequencing	21
2.4.3 Fluorescence <i>In-Situ</i> Hybridisation	22
2.4.3.1 Fixing samples	22
2.4.3.2 FISH probe binding	22
2.4.3.3 Sample Visualisation	23
Chapter 3: Results and Discussion.....	24
3.1 PCR hit rate	24
3.2 Quality assessment of Illumina raw reads	27
3.3 DADA2 Output	27
3.4 Microbial Diversity	28
3.5 Samples identified for further investigation	33
3.6 Examining metabolic capabilities of prevalent species	38
Chapter 4: General Discussion	41
4.1 Aims	41
4.2 Future work	42
4.2.1 Raman Spectroscopy	42
4.2.2 Further ichip isolation	42
4.2.3 Further data analysis	43
Appendices	45
Appendix A- PCR Gel Electrophoresis	45
Appendix B - Bioanalyser results	57
Appendix C - Sequencing data quality	59
Appendix D - Taxonomic distribution of assigned sequences (Genus level)	60-62
Appendix E - Alternative FISH probe alignment	63
Appendix F - Pure culture BLAST results	64
Appendix G - Sanger Sequence Chromatographs	65
Abbreviations.....	68
References	69

List of Tables

Table 1 - Post purification DNA concentrations	20
--	----

List of Figures

Figure 1 - Methanogenesis pathways.	5
Figure 2 - Record of Novel Archaeal Genomes and Species	11
Figure 3 - iChip illustration	12
Figure 4 - Experiment Enrichment, Isolation and Growth Stages.....	17
Figure 5 - Pooling schematic	19
Figure 6 - PCR Success Rate	24
Figure 7 - Comparison of Diversity Indices for ASVs	29
Figure 8 - Taxonomic distribution of assigned amplicon sequence variants (Order level)	31
Figure 9 - FISH probe S*-WWE1-1181-a-A-18	35
Figure 10 - Phylogenetic Tree of BLAST results for potential novel bacterium.....	37

List of equations

Equation 1 - Phred score calculation.....	27
Equation 2 - Diversity Indices	28

Acknowledgements

I would like to thank Dr James Chong for his guidance and the opportunity to undertake a masters by research in his lab. I would also like to thank Dr Anna Alessi for all her help in the laboratory and the rest of the Chong group for making the year such an enjoyable experience.

Declaration

I declare that this thesis is a presentation of original work and I, Jack Murgatroyd, 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

General Introduction

1.1 Anaerobic digestion

1.1.1 The Industry

1.1.1.1 Industry growth and development

Anaerobic digestion (AD) is a naturally occurring process that takes place in environments where there is a low abundance of oxygen such as water logged soils (Tiedje, 1984) and lake sediments (Bell, 1990). The term AD tends to describe the industrial process of a breakdown of organic material in the absence of oxygen. Energy stored in covalent bonds in polymers is ultimately transferred into bonds in methane released as an end product of this process. This is capitalised upon by industry as this energy is now more easily accessible by combustion. Although the first commercial AD plant was built over 150 years ago, and the first recorded commercial use in Europe was in 1895 in Exeter, UK, it has not been until the last fifteen years in particular that AD has grown substantially as an industry. Interest has recently increased due to its effectiveness in waste management, low carbon footprint and cheap running costs.

In the UK and more globally, government bodies have used policy to encourage growth of the AD industry (Styles, 2016; Scarlat, 2018) as it has become increasingly clear that it is a less environmentally destructive alternative energy source to fossil fuels. As biogas generation by AD makes use of relatively new technologies, by comparison to other methods of energy generation there is still a lot more to be understood. In addition to this, the ecosystem inside an anaerobic digester is immensely complex and the function and identity of many microbes that contribute to its overall productivity in terms of biogas yield is still unknown.

1.1.1.2 Methane as Biofuel

Methane has been a main source of energy for years throughout homes in many countries. Methane produces more energy per gram than any other hydrocarbon (55.5 kJg^{-1}) despite its lower heat of combustion of 891 kJmol^{-1} (Pittam, 1972). Under the current system of measurement for the contributions of different gasses to global warming, CO_2 is used as the base gas from which other pollutants are compared to in terms of Global Warming Potential (GWP). While methane remains in the atmosphere for a much shorter time than CO_2 it absorbs more energy, and this is reflected in its higher GWP value. It has been proposed recently that

there is a more informative measure termed ‘GWP*’ (Allen, 2018) that takes into account cumulative CO₂ emissions to date for comparison to the current rate of emission with short-lived climate pollutants. In terms of CO₂ production, methane is considered an environmentally friendlier fuel than other fossil fuels, producing less CO₂ per energy unit upon combustion than other hydrocarbons. Methane can be obtained in three ways, through either: geological deposits, chemical synthesis or biological synthesis.

Geological methane can be formed from the lower levels of the strata and comes about from the breakdown of organic matter under pressure and high temperatures, either organically or inorganically. Alternatively, methane can be formed inorganically in crystalline bed rock. While geological methane is important as a global source of energy it is, by its slow nature of formation, a non-renewable resource.

Alternatively, but less frequently, methane can be produced *via* a number of different industrial chemical processes. Two of these commonly used are the Sabatier process, in which CO₂ and H₂ are reacted at high pressures and temperatures in the presence of a nickel catalyst, and the Fischer-Tropsch process in which methane is produced as a by-product of the reaction of CO with H₂, the primary aim being the production of longer chain hydrocarbons.

Biological production of methane occurs naturally and can be harnessed by the industrial process of AD. The primary products of AD are the digestate and biogas, which is predominantly methane (55-70%) and CO₂ (30-45%). Before use as a fuel biogas must be upgraded to natural gas. This is done by use of a biogas upgrader which concentrates the methane by removing other components of biogas. Strict legislation is in place requiring the raw product to be upgraded as it will contain contaminants such as H₂S that damage machinery (Latosov, 2017) and have negative environmental effects.

1.1.1.3 Wastewater treatment

Aside from energy production AD is also a useful tool for environmentally friendly reduction of waste, be it on an industrial or smaller scale. Globally, AD use has increased partially due to its usefulness in waste management and particularly where government has introduced laws on disposal of organic material in an effort to decrease waste volumes. It is also important to note that AD is likely to occur in waste if it is left to decompose in landfill. Given the GWP of methane, as discussed in section 1.1.1.2, it is important that methane release into the

atmosphere is controlled. Landfill gas utilisation is the collective term for systems which control methane and other gas release from landfills. While the technology is becoming more widely used (Ahmed, 2015), it is generally more efficient and profitable for companies to use AD before waste has to be transported for disposal. As such, landfill gas utilisation is generally used as an opportunistic and damage limiting measure rather than being the primary option for methane recovery from organic waste.

1.1.2 Stages of Anaerobic digestion

1.1.2.1 Hydrolysis

The first stage of AD involves the breakdown of complex sugars and other organic polymers into shorter chain molecules by hydrolysis. At a molecular level hydrolysis involves the addition of water molecules to catalyse the breakdown of covalent bonds. This results in the addition of a hydrogen ion to one product and a hydroxide ion to the other. Hydrolytic bacteria that facilitate this process do so using a variety of different hydrolysing enzymes including proteases, carbohydrases and lipases. The majority of molecules resulting from this process are amino acids, fatty acids and monosaccharides. The breakdown of these longer chain organic molecules allows microbes in later stages of the process to metabolise the products resulting ultimately in the production of H_2 and CH_4 . Other breakdown products that result from hydrolysis, such as acetate, can be used by methanogens directly in the fourth stage of AD, methanogenesis.

Attempts to optimise the hydrolysis step have primarily involved the separation of hydrolysis and acidification from acetogenesis and methanogenesis in two stage reactor systems, as they do not all share the same optimum conditions (Bouallagui, 2005; Liu, 2006). This does, however, reduce the amount of syntrophic association possible as well as stopping the transfer of hydrogen from acidogenesis products to acetogens and methanogens. Syntrophic association here refers to a microorganism from one stage of anaerobic digestion using the products of a microorganism from another stage.

1.1.2.2 Acidogenesis

Acidogenesis is the second stage of AD. In acidogenesis fermentative bacteria further break down products of hydrolysis into long-chain fatty-acids and volatile fatty-acids. Other small molecules are produced as by-products including CO_2 , NH_3 and H_2S . While this step is generally considered to be the least rate-limiting step it is important to monitor as uncontrolled

activity of the fermentative bacteria can lead to over-acidification (Moeller, 2016). This is important to note as methanogens typically function optimally within a very specific pH range (Ye, 2012).

1.1.2.3 Acetogenesis

In acetogenesis VFAs and other molecules produced by acidogenesis are used by acetogens to primarily produce acetic acid, CO₂ and H₂. The pathway by which acetate is produced, named the ‘Wood-Ljungdahl pathway’, enables acetogens to use CO₂ as an electron acceptor and the initial molecule for synthesis and H₂ as an electron donor through three key enzymes (Ragsdale, 2008). There are two different types of acetogens: syntrophic acetogens, which require the presence of methanogens to ensure that acetate production is thermodynamically favourable; and non-syntrophic homoacetogens which are not dependent on methanogens. When considering acetogenesis for optimising AD it is important to ensure that the H₂ partial pressure remains low enough to allow reactions to proceed for the conversion of all acids to acetate.

1.1.2.4 Methanogenesis

Methanogens are a group of strictly anaerobic archaea that are characterised by their ability to produce methane through methanogenesis. They are the organisms responsible for the processes occurring in final stage of AD. In this stage several different metabolic pathways are known to be used to create ATP by reduction of different organic molecules; methane is created as a by-product in each of these pathways. To date there have been three main methanogenesis pathways identified, namely hydrogenotrophic, acetoclastic and methylotrophic methanogenesis (Figure 1).

In hydrogenotrophic methanogenesis (Figure 1.1) a CO₂ molecule is used as the starting molecule which H₂ reduces to form formylmethanofuran (CHO-MFR). From here the CHO group is transferred to tetrahydromethanopterin (H₄MTP) to produce formyl-H₄MTP. Following dehydration and reduction of the molecule to methylene-H₄MTP and then methyl-H₄MTP, the methyl group is transferred to HS-CoM producing methyl-CoM. In the final step of the process methyl-CoM is reduced to methane by Coenzyme-B (CoB). H₂ reduces the molecule that results to allow CoB to be used again (Thauer, Kaster, *et al.*, 2008). It is also possible for some methanogens to use formate as the starting molecule for hydrogenotrophic methanogenesis (Liu, 2008).

Aceticlastic methanogenesis (Figure 1.2) proceeds in a slightly different manner to hydrogenotrophic methanogenesis. The starting molecule used is acetyl-CoA, formed from an acetate molecule activated by ATP and CoA. A complex CODH/acetyl-CoA synthase splits acetyl-CoA transferring the methyl group to H₄MPT. Methanogenesis then progresses as in hydrogenotrophic methanogenesis with CoB as a reducing molecule that is recycled through reduction by H₂.

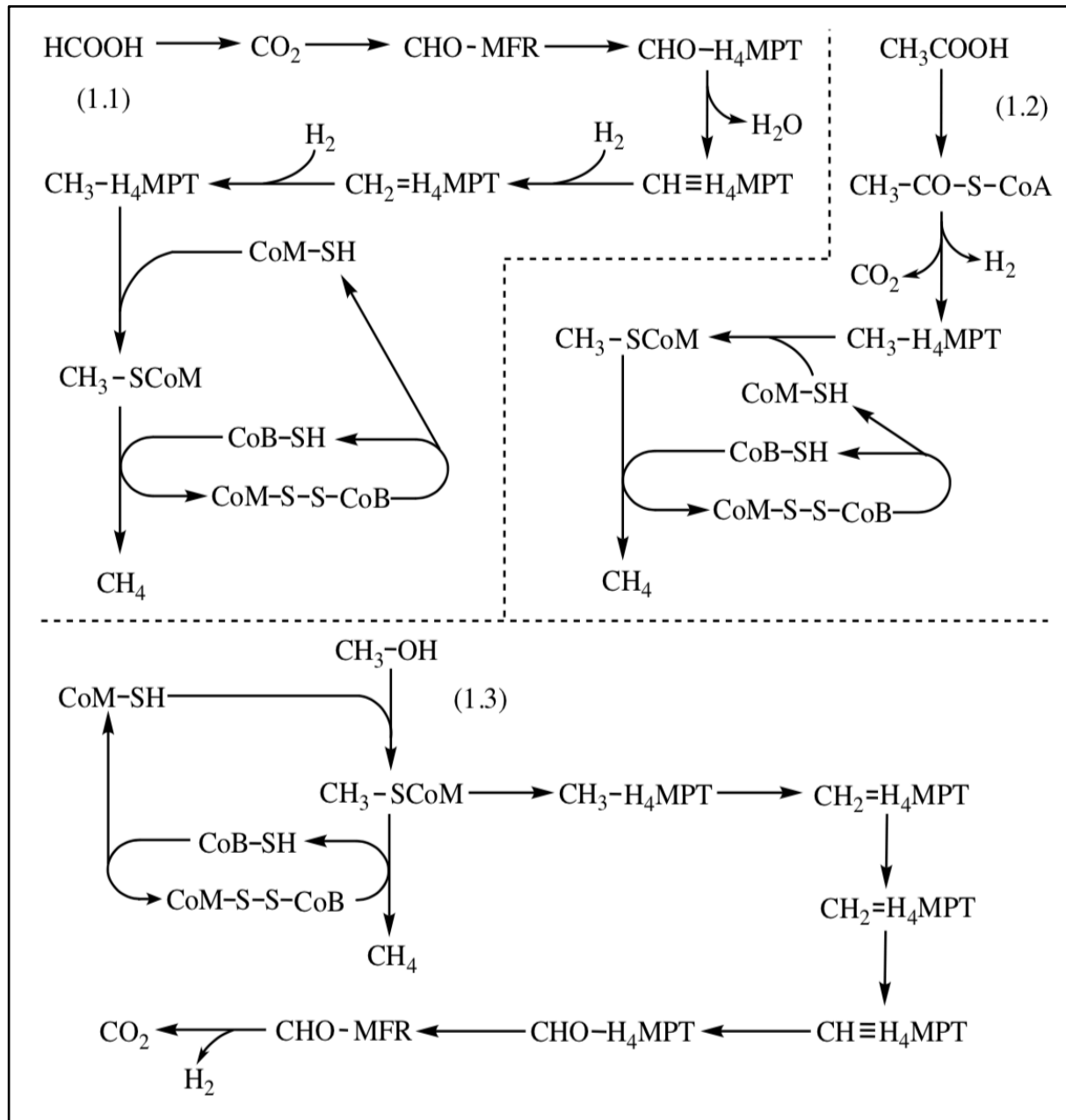


Figure 1 Methanogenesis pathways The three general methanogenesis pathways identified: (1.1) hydrogenotrophic methanogenesis, in this case formate is used as the carbon source; (1.2) aceticlastic methanogenesis and (1.3) methylotrophic methanogenesis using methanol here as a carbon source.

The final pathway is known as methylotrophic methanogenesis (Figure 1.3). In this instance methanol has been used to demonstrate the pathway, however, a variety of methylated substrates can be used including methylamines and methylated sulphur compounds. The pathway works differently to both hydrogenotrophic and acetoclastic methanogenesis in that methane is produced within the first few steps and CO₂ is released as a by-product. The first step involves two methyltransferase proteins which move the methyl group initially to a corrinoid protein and then to CoM to form methyl-CoM (Burke, 1997). Following this, one methyl-CoM molecule is oxidised to CO₂ via a pathway that functions as the hydrogenotrophic pathway in reverse. This provides a proton motive force and reducing equivalents to reduce three methyl-CoM molecules to methane. As in other pathways CoB reduces methyl-CoM and is recycled by reduction.

Methanogens are a diverse group of archaea and can be found in a variety of different environments. While the majority are mesophiles there are a number that have been found in extreme conditions (Zinder, 1984; Jones, 1989). Up until 2015 it was widely agreed that all methanogens were classified into the archaeal phylum *Euryarchaeota*, however, it has recently been suggested that there may be methane producing archaea outside of this phylum (Evans, 2015; Vanwonterghem, 2016). In terms of metabolism, hydrogenotrophic methanogenesis using H₂ and CO₂ is found across all orders of methanogens excluding *Methanomassiliicoccales*. For this reason it has been suggested that it is the original method of methanogenesis from which others have evolved (Baptiste, 2005). Our understanding of the phylogenetic linkages between the three domains of life has changed frequently (Eme, 2017). Although the origins of eukaryotic life are still debated the current most popular hypothesis is that Bacteria and Archaea evolved from the same ancestor before eukaryogenesis occurred, stemming from the archaeal branch of life (Eme, 2017).

1.1.2.5 Anaerobic bacteria

While the key aim of investigating AD is the maximisation of methane yield from reactors the importance of understanding and uncovering the roles of anaerobic bacteria within this process should not be forgotten, given their dominant presence within AD microbial communities. Three of the four stages of AD can be attributed to the metabolic activities of anaerobic bacteria: hydrolysis, acidogenesis and acetogenesis. A deeper understanding of these stages and the microbes involved will not only further AD technologies but will also work to aid development in other fields. Hydrolytic enzymes, in particular, are widely used in industry for food processing, chemical production and pharmaceuticals to name just a few areas.

1.2 Novel Species Discovery

1.2.1 16S rRNA gene comparative genomics

1.2.1.1 Overview

16S rRNA gene comparative genomic techniques are currently one of the standards for faster identification of microbes in a variety of different scenarios. As a method of microbial identification it is less time consuming and more accurate in comparison to whole genome shotgun sequencing and phenotypic methods of microbial identification respectively.

The 16S rRNA gene contains nine variable regions interspersed with conserved regions. Due to the vital role of the 16S rRNA gene in coding for ribosomal RNA, mutations to the regions of the gene coding for functional parts of 16S rRNA will likely result in reduced function or non-functioning ribosomes. This in turn is likely to result in the death of the organism containing this gene due to a reduced ability to compete. Therefore, it can be said with a degree of confidence that the regions of the 16S rRNA gene that code for functional 16S rRNA regions are conserved throughout microbes. In the same manner, mutations to the bases coding for non-functional regions of the 16S rRNA gene are much less likely to affect the competitive ability of the organism in question and so will vary in sequence between different microbes. The amount of variation seen in these nine variable regions, named V1 to V9, is taken to be representative of the amount of variation present in an organism's genome comparatively.

As the conserved regions in the 16S rRNA gene provide known sequences for primer binding, it is possible to design primers that should bind a significant percentage of the entire population of known microbes. Studies have been undertaken cataloguing primers for use in 16S rRNA gene comparative genomics and the percentage of known archaea, bacteria and eukaryota 16S rRNA genes that the primers will bind (Klindworth, 2013). This information is also available through the Silva website through the TestPrime tool (Klindworth, 2013). Forward and reverse primers are selected to amplify one or more variable regions using PCR before the amplicons are sequenced and data is compared to a database of 16S rRNA genes.

Whole 16S rRNA gene comparative genomics can be used or to save time with a large number of samples a small degree of certainty can be sacrificed and comparisons will only be made between one or two variable regions. For the purposes of identifying novel bacteria, where only previously uncultured organisms are of interest, a lower degree of certainty in the identification

of a large number of already cultured organisms is not a concern.

1.2.1.2 16S rRNA gene sequence data manipulation

Following sequencing of a region or several regions of the 16S rRNA gene it is necessary to use a pipeline to bin results into groups of varying sequence similarities. The similarity of a 16S rRNA gene to that of a known organism or group of organisms helps *in-silico* assignment of the read to different taxonomic levels. The higher the sequence similarity of a read to a known organism the more taxonomic levels of classification the two organisms will share upon assignment. This obviously requires a comprehensive database of known organisms' 16S rRNA gene sequences for comparison in order to confidently assign reads.

1.2.1.3 Alternatives

1.2.2 Novel species definition

Although the term novel is used commonly throughout the sciences, the definition of the term is widely debated. For the purposes of microbial discovery, using the 16S rRNA gene for comparison to known organisms, a 97 % similarity cut off is the widely accepted standard for the identification of a species. This 97 % identity threshold is the main topic of dispute within the debate surrounding the term novel species (Nguyen, 2016; Beye, 2018). It is generally accepted that if the 16S rRNA gene of the most similar organism is anything less than 97 % similar upon comparison of base-pair matching then this is evidence to support its definition as a novel species. In recent cases of novel species identification whole-genome comparisons have been used in conjunction with other methods to provide more sound evidence for the classification of the organism as a novel species (Diéguez, 2017; Machado, 2018). These other techniques include phenotypic comparisons, cellular fatty acid profiling and MALDI-TOF MS amongst others. There are some studies that adhere to a lower level of stringency in terms of the use of multiple methods for conformation of the novel species identity. For example, 18S rDNA sequencing data alone have been used as evidence for the identification of a novel species in recent years (Knowles, 2018). 18S rRNA gene comparative genomics is the equivalent of 16S rRNA gene comparative genomics for eukaryotes. It is important to note that the eukaryotic definition of species, while still disputed, poses fewer difficulties than defining the prokaryotic definition of species. A contributing factor to this is the alternate methods of reproduction used within each kingdom.

Of course, this threshold value is not only debated for species level classification. However,

given that there are more species classifications than any other taxonomic level it is more frequently raised in debate. The underlying issue according to Hey (Hey, J, 2001) is the two conflicting scientific aims: the categorisation of organisms and the understanding of evolutionary processes that produce different species. This is nowhere truer than in microbiology. This problem has been termed the 'species problem'. The original criteria for assignment of taxonomic names has changed as technology has developed and we are able to see differences between organisms that would not have been possible at the induction of the naming system. This is unlikely to be resolved as technology constantly develops and discoveries are made, changing opinion on phylogenetic relationships between organisms therefore making a standardised system hard to maintain. Four commonly used systems of species classification currently exist, namely: Cohan's ecological concept, the recombination species concept, phylogenetic methods of classification and pragmatic classification.

An ecological approach to classifying microbes into different species relies on the processes of natural selection and the formation of populations that have adapted to certain environments. If it is taken that in all environments there will be periodic selection resulting from the removal of organisms less adapted to the environment, then there is more potential for this method to prove useful. Examining this from a practical aspect however, shows that this method of classification doesn't hold true once the recombination of core genes is accounted for. As genes vital for survival can be transferred between organisms it is possible for two very phylogenetically distinct organisms to both survive periodic selection sweeps if one of these genes is transferred to the other organism. In terms of the ecological approach these two organisms would then be grouped together, however, in terms of a phylogenetic approach to classification they would still be classified as distinct species from one another.

This touches also on a problem that has been highlighted with the recombination species concept. This concept draws on ideas from the Biological Species Concept (BSC) (Mayr, 1999) that is commonly used for the classification of eukaryotic species. The two concepts differ as the BSC relies on the idea of the formation of closed gene pools. The formation of closed gene pools relies partially on sexual production providing a physical barrier to cross-species non-homologous gene transfer. In prokaryotes, however, non-homologous genes can be transferred laterally. The physical boundaries in place for eukaryotes are not present here and as a result it is hard to make a case for this particular definition within the prokaryotic domain.

A phylogenetic approach to classifying species poses the problem within itself of which

method to use for genetic comparisons between organisms. Different methods used to obtain data for phylogenetic comparisons include DNA:DNA hybridisation and 16S rRNA comparative genomics amongst others. In reality, it is normally a combination of these various techniques that is used and results from each technique are considered as a whole. Using the phylogenetic approach to species classification again causes problems in that it assumes microbes form clades, that is, a group of organisms that are all evolutionarily descended from one ancestor. As above, horizontal gene transfer between organisms from different genera undermines this assumption. Depending on the area of the genome that you use for comparison in order to classify a species, there is potential to classify the same organisms into different clades. This is a result of horizontal gene transfer and the evolutionary contradictions it creates between different areas of the genome within the same organism when using a phylogenetic approach to species definition (Doolittle, 2007).

The final popular approach taken for classification of species has been termed pragmatism. This concept suggests that species do not exist and are a human concept but that a system of classification is recognised as a necessity to allow reproducibility and stability (Dupré, 2001). There are a number of approaches to implementing pragmatism that use aspects of other species concepts in combination with each other. This is the approach taken in this work

1.2.3 Rate of discovery of genomes and species

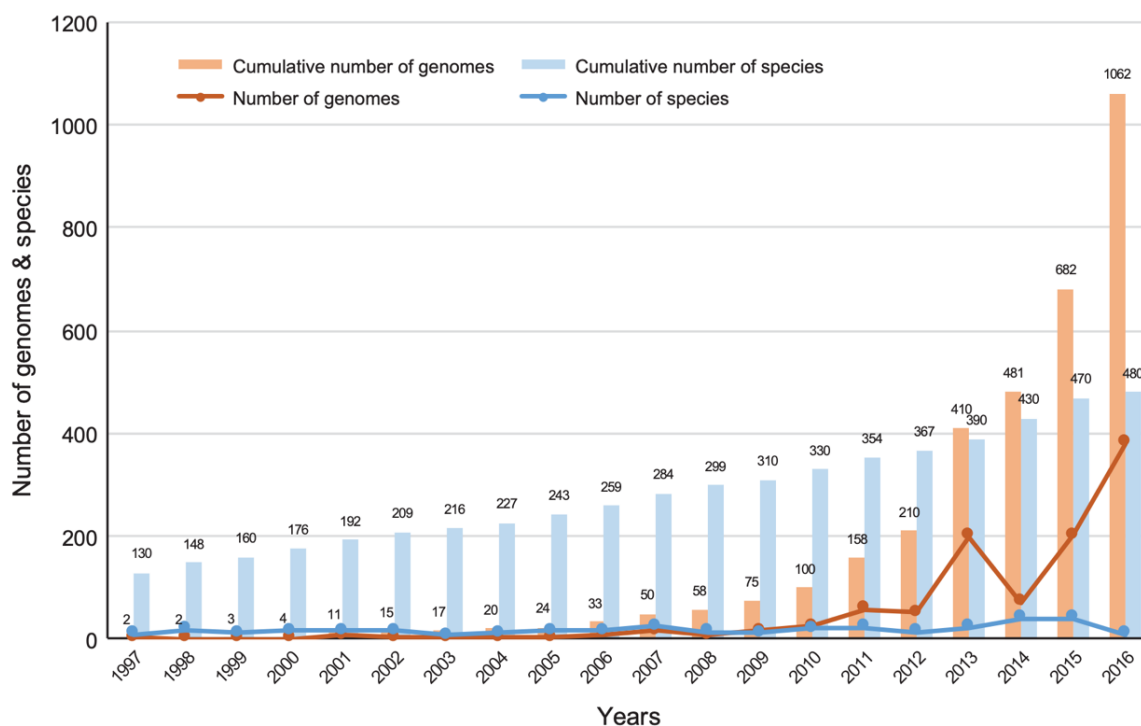


Figure 2 Record of Novel Archaeal Genomes and Species Lines represent the number of new species identified each year. Bars represent the cumulative number identified, as presented in the NCBI database (O’Leary, 2016). Figure source (Adam, 2017)

While advances in culture independent sequencing technology have allowed an increase in the number of genomes being deposited in databases each year there has been very little increase in the number of novel archaeal species identified (Figure 2). This can be attributed to the fact that culturing technology has not developed at a similar rate. Growing a pure culture of cells is often important in the identification of novel species of any microbe. This discrepancy between rates of genome and species discovery applies to bacteria also (Schloss, Girard, *et al.*, 2016).

1.2.4 Isolation Chip

The Isolation Chip (iChip) (Figure 3) was devised as a method of cultivating previously ‘uncultivable’ microbial species (Nichols, 2010). The iChip works by allowing cultures to be grown in isolation from competition from other microbes. If each culture is started from a single cell and is grown in isolation, then the resulting cultures should be pure. This method takes a step towards combating the great plate anomaly, a term that refers to the phenomenon of a large proportion of the microbes that are visible under the microscope being uncultivable in the lab. This occurs as it is hard to replicate complex *in-vivo* metabolic environments *in-vitro*, owing to the vast number of different species present and their various metabolic

interactions. After iChip loading iChips can be placed back into the original environment from where cells were taken. These cells will receive the nutrients and metabolites from other

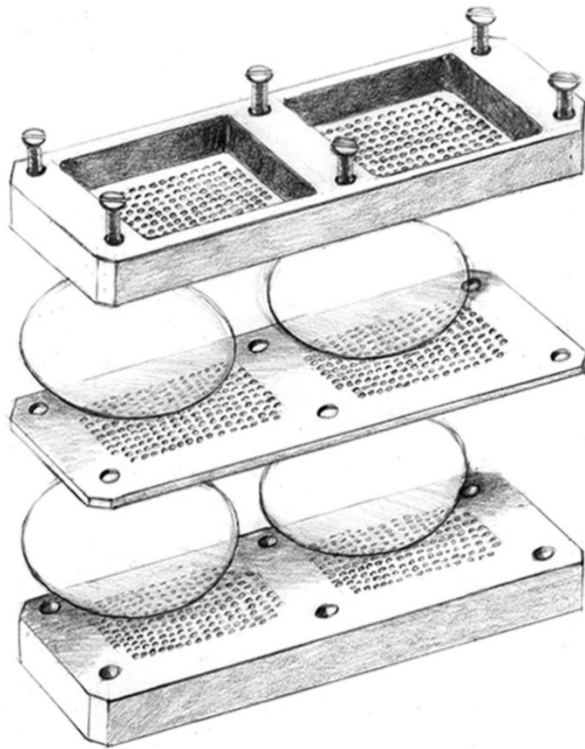


Figure 3 iChip illustration iChips are made up of three plates, commonly of plastic, containing holes that line up upon assembly. Different membranes can be selected to place in between plates before fixing the plates together using screws. Figure source (Nichols *et al.*, 2010)

microbes within the environment through semi-permeable membranes used to seal the wells. This will allow the cultures to form within the iChip which can then be studied.

A number of studies have used the iChip for the cultivation of previously uncultivable microbes since its invention in 2010. The majority of these studies have involved the placement of the iChips, after loading, back into the environment from where samples were taken. The initial experiment performed using the iChip (Nichols, 2010) found that their usage roughly yielded a 5 fold increase in the number of cultures grown using an iChip as opposed to culturing using a Petri dish. This was true for samples taken from both seawater and soil. In this study there was little overlap between the lists of species identified from

iChip and Petri dish cultivation with less than 1% of species cultivated from seawater in the iChips also being cultivated through Petri dish growth. In this instance the iChip was placed back into the sampling environment but it is also possible to simulate the sampling environment *in-vitro* (Kaeberlein, 2002). In this way organisms with shared growth requirements can be targeted by altering the media in which the iChips are submerged, to simulate appropriate conditions. iChips have been used to discover new antibiotics (Pidcock, 2015) in addition to the previously “uncultivable” microbes cultivated in 2010 their application in anaerobic digestion however is novel.

1.3 Aims

The overall aims of this work were to complete and evaluate the success of the iChip method of culturing cells from anaerobic sludge for the isolation of novel organisms. The success of the method can be evaluated by considering two different aspects. Firstly, the number of novel

organisms identified at the end of the process by sequence analysis will give an indication of the technique's efficiency. Second, by way of evaluating the efficacy of this method it is necessary to culture and isolate at least one previously uncultured organism.

1.4 Motivation

This work was undertaken with the intention of further developing our understanding of the ecosystem within anaerobic digesters. The importance of AD as a method of energy generation through methane production is an increasingly salient technology within a world where alternatives to non-renewable energy sources are of global interest. The increase in the general awareness of national, business and individuals' carbon footprints also makes AD an attractive method for energy generation.

The fastidious nature of many microbes identified within AD means there are a wealth of organisms that are as of yet uncultured from this environment. This provides ample scope for improvements in the efficiency of AD for both methane production and wastewater treatment once the system is better understood. As the iChip method for cultivation of microbes from anaerobic sludge has not yet been attempted and given the success of past iChip studies in growing novel cultures (Nichols, 2010) this approach would appear to hold promise.

Chapter 2

Materials & Methods

2.1 Isolation and Growth (Figure 4)

2.1.1 Sampling

A sludge sample was taken from anoxic post-digestion tanks from Naburn Sewage Treatment Works (53.915° N, 1.085° W) and flushed with N₂ and CO₂ (4:1) before a number of 30 mL aliquots were incubated at 20 °C and 100 rpm for two weeks in 125 mL Wheaton bottles, closed with a butyl rubber stopper and an aluminium crimp seal.

2.1.2 Sludge fluid preparation

After the bottles containing the sludge samples were degassed, they were moved to an anaerobic chamber (Coy) and yeast extract (Thermo Fisher Scientific #LP002) was added to each solution to give a final concentration of 0.4% w/v. The bottles were flushed to create a N₂/CO₂ (4:1) atmosphere and after a further 48 h incubation period, at 37 °C and 100 rpm, the sludge was centrifuged at 6,000 × g for 15 min at 4 °C. The supernatant was respun in this manner and decanted twice to produce a clarified supernatant. This was collected and autoclaved at 140 °C and 20 bar sludge fluid was stored at 4 °C in the dark.

2.1.3 Growth Media

A media with a formulation recommended by DSMZ (DSMZ, 2017) was used to enrich for methanogens using hydrogenotrophic methanogenesis. 1 L of this hydrogenotrophic media (H media) contained: 50 mL of clarified sludge fluid (2.1.2), 10 g trypticase peptone (BD Biosciences #211921), 2.5 g of yeast extract, 0.5 g of sodium formate (Sigma-Aldrich #1064430500), 0.5 g of sodium acetate (Sigma Aldrich #S7670), 10 mL of 1000× trace elements (DSMZ, 2017), 150 mL of 10× mineral solution A (3 gL⁻¹ K₂HPO₄ (Fisher Scientific #10375760)), 150 mL of 10× mineral solution B (3 gL⁻¹ KH₂PO₄ (Fisher Scientific #10429570)), 6 gL⁻¹ (NH₄)₂SO₄ (Sigma Aldrich #A4418), 6 gL⁻¹ NaCl (Fisher Scientific #10626272), 0.6 gL⁻¹ MgSO₄ × 7 H₂O (Sigma Aldrich #M5921), 0.6 gL⁻¹ CaCl₂ × 2 (H₂O Sigma #1023821000), 0.1 mL of 0.1% (w/v) sodium resazurin. Water was added to increase the volume by 10% to 1.1 L so, after boiling in order to make the solution anoxic, the final volume was not decreased. After the solution was allowed to cool at room temperature for 15

min, 4 gL⁻¹ of NaHCO₃ (Fisher Scientific #10020510), 0.5 gL⁻¹ L-cysteine (Fisher Scientific #10499740) and 5 mL of methanol (Fisher Scientific #34860) were added to the solution. This solution was aliquoted inside the anaerobic chamber into Balch-type tubes. A vacuum pump was used prior to autoclaving at 140 °C and 20 bar for 1 h to replace the atmosphere in the Balch-type tubes with a H₂/CO₂ (4:1) 40 kPa atmosphere. The pump was cycled to 40 kPa four times to ensure thorough replacement of gases.

Three different media were prepared based upon the DSMZ formulation for the hydrogenotrophic media. Each of these media contained only one of the carbon sources Na-acetate, Na-formate and methanol, all of which were included in the same concentrations in these media as in H media respectively. These media were named aceticlastic media (HA media), formate media (HF media) and methylotrophic (HM media) respectively. A fourth media was also used and was based on a modified formula designed for rumen fungal growth (Hobson, 1997). A key difference to other media being the use of wheat straw, sourced from a local farmer, as the primary carbon source within the media. 1 L of Wheat Straw media (WS media), prepared to enrich for anaerobic fungi, contained 10 g of Bacto Casamino acids (BD Biosciences #225930), 2.5 g of yeast extract, 50 mL of clarified sludge, 150 mL of 10× mineral solution A and B and 0.1% (w/v) Na-resazurin made up to 1 L with distilled H₂O. As with other media this was made anoxic by boiling for 10 min before cooling and then adding 4 gL⁻¹ of NaHCO₃ (Fisher Scientific #10020510) followed by the addition of 1 gL⁻¹ of L-cysteine. 50 mg of finely ground wheat straw was distributed in to Balch-tubes inside the anaerobic chambers along with 5 mL aliquots of WS media. Wheat straw was ground using a pestle and mortar until a fine powder resulted. Tubes were sealed with butyl rubber stoppers and a metal crimp. The atmosphere was then replaced with N₂/CO₂ (4:1) and they were autoclaved at 140 °C and 20 bar for 1 h.

2.1.4 Enrichment

The anaerobic sludge sample was mixed and diluted 10× with the H media under anaerobic conditions. 100 µL of this solution was then used to inoculate 4.9 mL aliquots of the different growth media so that it comprised 2% of the final volume. To the diluted solutions 20 µg.mL⁻¹ of neomycin and 50 µg.mL⁻¹ of chloramphenicol were added to inhibit bacterial growth. Additionally, 50 µg.mL⁻¹ of Na₂S × 9H₂O (Fisher #S426-212) was added to function as a reducing agent.

Initially cultures were grown at 37 °C and shaken at 100 rpm. After a week of growth 100 µL of each culture were used to inoculate fresh H media, HA media, HF media, HM media and WS media. A sixth set of microbes were grown using the anaerobic sludge as a growth media. These cultures were left to grow at 37 °C and 100 rpm for 10 days.

2.1.5 Isolation Chip loading

Given the volume of each well of the iChip is 1 µl, the cell density of each enrichment culture was diluted to reach an average of 1 cell µL⁻¹. The average cell density of the cultures as based on optical density measurements at 600 nm was 2.9×10^4 cell µL⁻¹ and so a dilution factor of 3×10^4 was used.

After sterilising each iChip with 70 % (v/v) ethanol they were loaded within the anaerobic chamber. 50 mL of 0.35 % (w/v) molten agar (Melford Laboratories #A20090) was used to dilute the enriched cultures and this mix was poured immediately into a Petri dish. The middle plate was submerged to fill in all the wells. One membrane (Whatman Nucleopore Track-Etched, MB PC 0.05 µm, 47 mm) on either side of the central plate was used to seal the wells and then the iChips were reassembled, placed in sterile beakers and either sludge or a specific growth medium was poured over the top of the iChip to submerge all wells. These beakers containing the submerged iChips were placed in tightly sealed containers named anaerobic bombs and pressurised with a specific gas mixture to 15-20 kPa. For bombs containing iChips submerged in H media, HF media or sludge a H₂:CO₂ (4:1) mix was used. For bombs containing iChips submerged in HA media or HM media a N₂:CO₂ (4:1) mix was used. For bombs containing iChips submerged in WS media CO₂ was used. In total eighteen iChips were loaded (3 iChips per substrate) and incubated for 10 days at 37 °C.

2.1.6 Culturing

Following the 10-day incubation period, iChips were disassembled inside the anaerobic chambers. The 1 µL agar plugs from the iChips wells were transferred into 1 mL of the respective media in which they had been grown, excluding iChips submerged in sludge which were transferred to the general H₂AD media. Agar plugs were pushed out from wells in the iChips using a sterile metal rod. From each set of iChips grown in a different media 192 plugs were transferred into fresh media resulting in 1,152 microcultures. Two 96 well plates for each enrichment background were sealed with an adhesive cover and stored in the anaerobic bombs

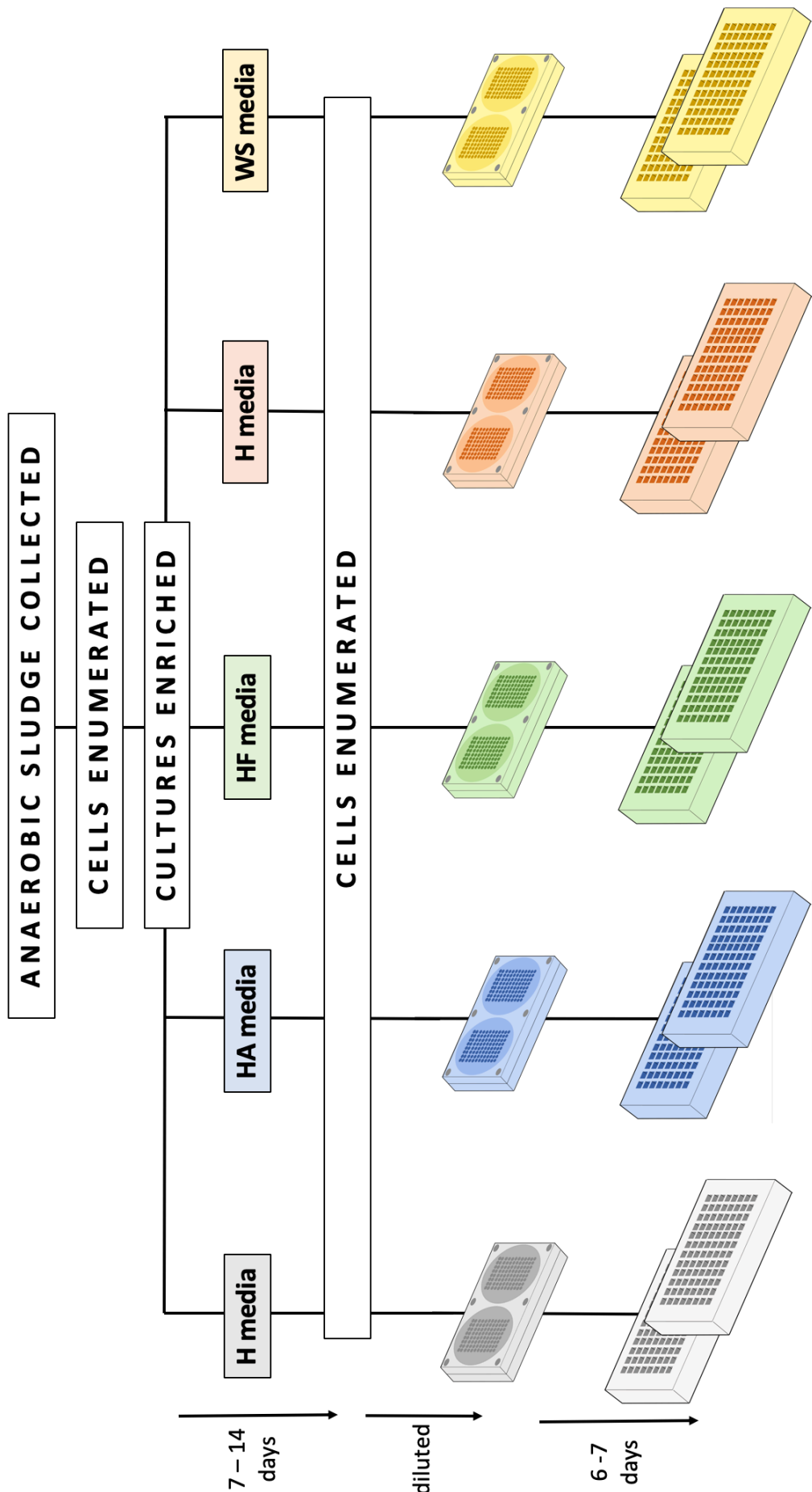


Figure 4 Experiment Enrichment, Isolation and Growth Stages Schematic showing the overview of sampling, isolation and growth stages. Periods of time indicated by arrows on the left varied depending on the growth background used. An additional set of iChips and cultures were set (not pictured) using the original sludge sample as a growth media.

and further incubated at 37 °C with frequent flushing and pressurising with appropriate combinations of fresh gas as described in section 2.1.5.

2.2 Molecular Work

2.2.1 Primers

Primers used for the amplification of the V4 region of the 16S rRNA gene were chosen based on a study by Klindworth *et al.* (2013). The primers, A519F (5'- CAG CMG CCG CGG TAA -3') and 802R (5' - TAC NVG GGT ATC TAA TCC - 3'), were selected for their high coverage of known bacterial and archaeal species (88.8% and 88.9% respectively). Primers (Integrated DNA technologies) were resuspended in nuclease-free water to give a final concentration of 100 µM before use in PCR.

2.2.2 PCR

Before PCR was performed, 50 µL of each sample was centrifuged at 1600 × g for 10 min using a Multifuge 3SR (Heraeus). The supernatant was then removed and replaced with 35 µL of nuclease-free water. A final reaction mix volume of 12.5 µL was used, halving all New England Biolabs (NEB) suggested volumes for a 25 µL reaction (New England Biolabs, 2013). One 12.5 µL reaction volume contained: 2 µL of each sample, for use as template DNA; 0.625 µL of forward primer; 0.625 µL of reverse primer; 2.5 µL of Q5 buffer (NEB #B9027S); 0.125 µL Q5 polymerase (NEB #M0491S) and 0.25 µL of 10 mM dNTPs (Promega #U1420). Reaction mixtures were pipetted into 96-well plates using an 8-tip multichannel pipette and each mixed by pipetting the volumes up and down five times. A Prime Thermal Cycler (Techne) was used for PCR. Initial denaturation was set at 98 °C for 3 min in order to ensure cell lysis, following this was 30 cycles of a 20 s, 98 °C denaturation step, 30 s, 52 °C annealing step and 45 s, 72 °C extension step followed by a single final 5 min extension step at 72 °C. Samples were then stored at 4 °C until required.

2.2.3 Gel electrophoresis

In order to visualise PCR products, they were separated using gel electrophoresis. 70 mL TAE buffer (40 mM Tris (Invitrogen #15504020), 20 mM acetic acid (Honeywell #695092), 1 mM EDTA and agarose powder (Melford Laboratories #A20090-500.0)) was used to make a 1% agarose gel containing 3.5 µL of SYBR safe DNA gel stain (Thermo Fisher Scientific #S33102). Into each well of the gel a 6 µL solution was pipetted comprising: 2 µL of PCR product, 1 µL of 6x loading dye (NEB #B7024S) and 3 µL of water. 4 µL of 1 kb GeneRuler

DNA ladder (Thermo Fisher Scientific #SM0312) was also added to each row of the gel to allow for size comparisons to PCR product bands. Gels were run for 30 min using a voltage of 110 mV.

2.2.4 Pooling and purification

2.2.4.1 Pooling

PCR products were pooled before purification as shown in Figure 5. This reduced the number of separately indexed PCR products sent for sequencing from 1,152 to 60. PCR products were purified by using 51.2 μL of AMPure XP magnetic beads (Beckman) for 64 μL PCR product volumes and 76.8 μL of beads for 96 μL PCR product volumes. Final elution volumes, following two washing steps using 80 % w/v ethanol as per the manufacturer's protocol (Illumina, 2013), were 38 μL and 48 μL for 64 μL and 96 μL of PCR product. For samples of both volumes 2 μL of

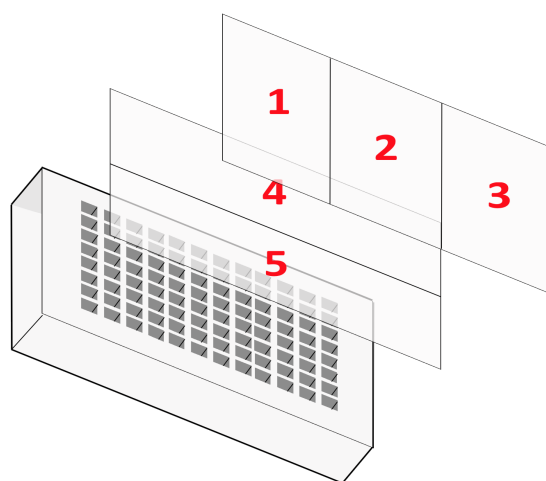


Figure 5 Pooling schematic 2 μL aliquots from 32 wells (1-3) or 48 wells (4,5) were pooled to produce pools of either 64 μL (1-3) or 96 μL (4,5). The overlapping of pools 1-3 with 4 or 5 gives a sixteen well resolution following DNA sequencing.

purified PCR product was left in the wells to avoid having beads in the purified solution after the clean-up step. As there were two plates of PCR products for each growth background, pools were labelled 1-10 for each growth background. Plates 1-5 represented pools from plate 1 and 6-10 pools from plate 2.

2.2.4.2 Qubit fluorometry

The concentration of DNA in each purified pool was measured using a Qubit fluorimeter and a Quant-iT dsDNA Assay kit, high sensitivity (Thermo Fisher Scientific #Q33120). This involved the dilution of stock reagents from the kit to give two standards which can be used to calibrate the Qubit fluorimeter before measuring the concentration of all 60 cleaned pools. Excitation and emission maxima were 502 nm and 523 nm respectively.

Pool number	DNA concentration $\mu\text{g mL}^{-1}$					
	H media	Sludge	HA media	HF media	HM media	WS media
1	3.38	23.7	1.29	9.79	4.28	0.105
2	3.96	4.04	2.69	10.9	6.03	0.839
3	0.697	24.1	2.09	14.5	5.68	0.677
4	7.2	23.8	2.9	19.6	11.3	6.1
5	3.28	17.5	5.54	16.5	7.88	2.08
6	3.09	43	2.2	8.3	8.33	6.36
7	1.73	14.1	2.48	6.96	13.2	1.24
8	4.39	5.59	2.59	7.79	11.7	5.47
9	1.74	10.3	1.77	8.71	23.4	1.01
10	9.9	42.3	1.36	14	19.7	1.85

Table 1 Post purification DNA concentration Results are a mean calculated from three repeat measurements from the same solution made of a 2 μL aliquot for each PCR product diluted in 198 μL stock solution (1:200 Quant-iT[™]dsDNA HS reagent: Quant-iT[™]dsDNA HS buffer). The three measurements were made 30 s, 2 min and 3 min 30 s after dilution of the PCR product in this stock solution.

2.2.4.3 Index PCR

Based on the concentrations of the purified DNA (Table 1), adjusted volumes of each pool were added to the indexed PCR mix. This ensured the concentration of the template DNA within the PCR reaction mixture fell between 1.5 and 7.0 $\text{ng} \cdot \mu\text{L}^{-1}$. Index PCR was performed using the Nextera XT Index Kit, following the manufacturer's recommendations for 16S PCR amplicon barcoding using a 2 \times NEBNext High-fidelity PCR master mix (NEB #M0541S). Nextera forward and reverse primers were used to attach dual indices and Illumina sequencing adapters. PCR products were visualised using gel electrophoresis as in section 2.2.3. All products were cleaned as in section 2.2.4.2 and quantified before they were diluted to give 100 μL 4 nM DNA solutions. 5 μL of each sample was taken and pooled together to give a single solution containing all indexed pools in equal concentrations.

2.2.5 Library preparation and Illumina sequencing

The pooled library and a PhiX control (Illumina #MS-102-2002) were denatured using freshly made 0.2 N NaOH, diluted to 4 pM with hybridisation buffer (Illumina #MS-102-2002) and then mixed in a 4:1 ratio. The sample was denatured at 96 $^{\circ}\text{C}$ for 2 min and then cooled at 4 $^{\circ}\text{C}$ on ice for 5 min before being immediately loaded into a MiSeq v2 cartridge (Illumina #MS-102-2002) for 2 x 200 bp sequencing.

2.3 Sequence data processing

2.3.1 Cutadapt

For every sequence in each sample the first 15 bases and the last 18 bases were removed using Cutadapt (Martin, 2011). This was to ensure that the forward and reverse primers were not included in reads for data processing further downstream.

2.3.2 DADA2

The edited reads were processed through the DADA2 (Callahan, 2016) pipeline. The pipeline filters reads and assigns taxonomy to each read based on the Silva version 132 reference database (Quast, 2012) using the Naïve Bayesian classifier method (Wang, 2007). Variables for data filtering within DADA2 were set to allow for a maximum of 1 expected error (Edgar, 2015) in each read by making a simple alteration to the script as outlined in the pipeline tutorial (Callahan, 2016).

2.4 H₂AD-acetate enriched samples

2.4.1 Subculturing

Sixteen samples selected for further investigation were subcultured under anaerobic conditions by adding 500 µL of each culture into 5 mL of freshly prepared HA media in hungate tubes. The headspace in each tube was pressurised to 15 kPa with a gas mix of N₂:CO₂ (4:1). The sixteen inoculated media were incubated at 37 °C and 100 rpm for a maximum of 48 h or until the cultures had grown to become turbid.

2.4.2 Sanger sequencing

After incubation PCR was performed using 50 µL reaction volumes, Q5 polymerase and the primers 27F (5'-AGA GTT TGA TCM TGG CTC AG-3') and 1492R (5'- TAC GGY TAC CTT GTT ACG ACT T-3'); two common primers used for amplification of all variable regions of the 16S rRNA gene (Y. Chen, Lee, *et al.*, 2015). Suggested NEB volumes for a Q5 reaction of 50 µL were used and 1 µL of each culture as template DNA. These products were sent for Sanger sequencing and results were compared to the NCBI database using BLAST (basic local alignment search tool) (Altschul, 1990).

2.4.3 Fluorescence *In-Situ* Hybridisation

2.4.3.1 Fixing Samples

FISH (Fluorescent *In-Situ* Hybridisation) was undertaken to determine the purity of the cultures, specifically investigating *Candidatus Cloacimonas*. Samples were prepared for staining using 16S rRNA gene targeted fluorescently-tagged oligos. This was done using a 4% paraformaldehyde (PFA) solution to fix the microbes, creating covalent bonds between proteins and the cytoskeleton increasing the rigidity of cell structures.

A protocol was adapted from the SILVA website (SILVA Ribosomal Database Project, 2010). Firstly, samples were spun in falcon tubes at $10\ 000 \times g$ for 8 min before the cell pellets were resuspended in 5 mL of a 1x PBS solution ($8\ \text{gL}^{-1}$ NaCl, $0.2\ \text{gL}^{-1}$ KCl (Fisher Scientific #10375810), $1.44\ \text{gL}^{-1}$ Na_2HPO_4 (Fisher Scientific #AC204851000), $0.24\ \text{gL}^{-1}$ KH_2PO_4 adjusted to pH 7.4 with HCl (Analab #30721)). Samples were spun at $10\ 000 \times g$ for a further 8 min before the resulting pellet was resuspended in 5 mL of 4 % PFA solution (1:24 PFA:PBS) and incubated for 2 h at 4 °C for fixation. Tubes were then centrifuged at $3373 \times g$ for 8 min and the pellet containing the fixed cells was resuspended in 5 mL of 4 °C distilled H_2O before centrifugation at $3373 \times g$ for 8 min. This step was repeated twice again before the cells were finally resuspended in a PBS: EtOH (1:1) solution and stored at -20 °C.

2.4.3.2 FISH probe binding

Slides were set using a combination of five different probes: three targeting the bacterial 16S rRNA gene named EUB338 I (5'-GCT GCC TCC CGT AGG AGT /Cy5/-3'), EUB338 II (5'-GCA GCC ACC CGT AGG TGT /Cy5/-3') and EUB338 III (5'-GCT GCC ACC CGT AGG TGT/Cy5/-3'); one targeting the archaeal 16S rRNA gene named ARCH915 (5'-GTG CTC CCC CGC CAA TTC CT/Cy5/-3') and one targeting the *Candidatus Cloacimonas* 16S rRNA gene (5'-CTT CCT CTG CGT TGT TAC-3') (Chouari, 2005). The consensus sequence for the bacterial probes is (5'-GCW GCC WCC CGT AGG WGT /Cy5/-3'). Both bacterial and archaeal probes contained the fluorophore Cy5 while the *Cloacimonas* probe contained Cy3. Four spots were set for each sample. One probing with the bacterial 16S probe, one with the archaeal 16S probe, one with the *Candidatus Cloacimonas* probe and one with all three probes. Each probe was diluted to $50\ \text{ng}\cdot\mu\text{L}^{-1}$ before use.

For each spot on a slide 10 μL was pipetted at the requisite location before the slide was dried at 46 °C for 30 min. The slide was then washed for 2 min periods in three ethanol solutions of increasing concentrations of 50% EtOH, 80% EtOH and 96% EtOH. The slides were dried in a fume hood for 5 min and then 8 μL of 40% formaldehyde (FA) (Sigma #F8775) hybridisation buffer was added to each spot with 1 μL of each of the appropriate probe solution. 1 mL of 40% FA hybridisation buffer contains 400 μL of FA, 399 μL of milli-Q H_2O , 180 μL of 5M NaCl, 20 μL 1M Tris/HCl (pH 7) and 1 μL of 10% SDS (Melford #L22010). Slides were placed in Falcon tubes sideways with a strip of paper towel soaked in the hybridisation buffer and then incubated for 2 h at 46 °C. After rinsing with wash buffer pre-heated to 48 °C, the slides were submerged in the buffer for 15 min and then rinsed using 4 °C Milli-Q H_2O and stored at -20 °C. 50 mL of wash buffer contained: 47.99 mL Milli-Q H_2O , 1 mL of 1M Tris/HCl (pH 8), 500 μL of 0.5M EDTA, 460 μL of 5M NaCl and 50 μL of 10% SDS . Slides were visualised using a ZEISS LSM 880 confocal microscope.

2.4.3.3 Sample visualisation

ZEN lite software was used for image capture and manipulation. Two lasers were used to excite probes and images were captured using absorption wavelengths of 538-680 nm and 638-759 nm for Cy3 and Cy5 respectively.

Chapter 3

Results and Discussion

3.1 PCR hit rate

1152 different samples were used as template DNA for PCR reactions (described in section 2.2.2). Due to a number of factors there were differing rates of DNA amplification by PCR for samples grown using different growth media (Figure 6). Gel electrophoresis used to visualise PCR results can be seen in appendix A. There are multiple potential explanations for this difference in hit rates between samples from different enrichment backgrounds. Firstly, primer bias has a large effect. Primers were designed to bind the V4 region of the 16S gene for 88.8% of bacterial species, 88.9% of archaeal species and 0.6% of eukaryotic species according to the online Silva Test Prime tool (Klindworth, 2013). The V4 region was selected as it was one of

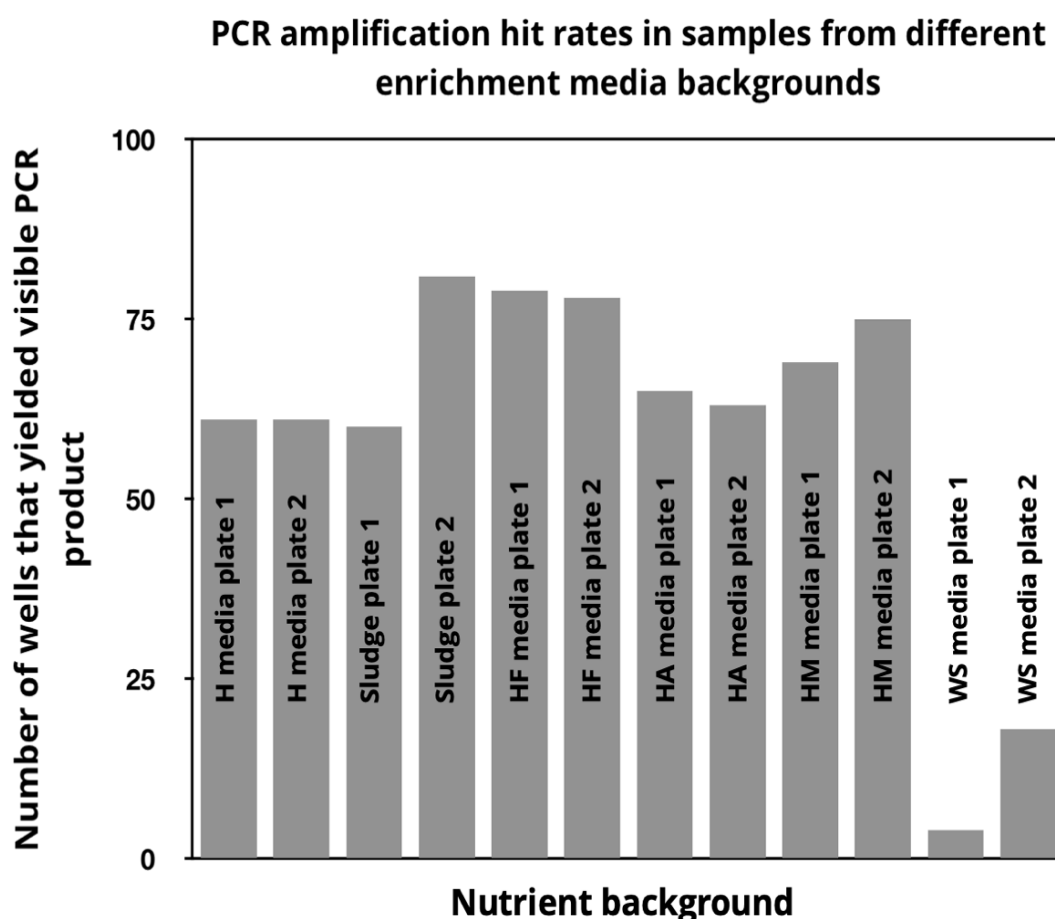


Figure 6 PCR Success Rate The average number of samples that yielded visible PCR product from each enrichment background, considering both plates, were: H media - 68.5%, sludge - 73.4%, HF media - 81.8%, HA media 66.7%, HM media - 75.0% and WS media - 11.5%.

the variable regions of the 16S gene that gave the lowest error rates when using the Naïve Bayesian method of taxonomic classification (Wang, 2007). The wheat straw medium was formulated to enrich for anaerobic fungi and so fewer PCR products were expected as the primers used in PCR targeted bacterial and archaeal 16S rRNA genes. This low level of amplification meant it likely that although PCR product was still produced from the WS media grown samples used as template DNA, it would be in lower concentrations. This may have resulted in the product not being visible following gel electrophoresis despite its presence. In order to visualise low levels of product a bioanalyser could have been used, however, given the number of samples this would not have been cost or time effective. Alternatively, more PCR product could have been loaded for gel electrophoresis, however, in order to keep gel electrophoresis results from different growth backgrounds comparable, PCR product volumes loaded were kept constant.

In terms of primer bias within kingdoms it is the case that primers will bind the target region of DNA more strongly in some species than in others, as although the primer-targeted regions in the 16S rRNA gene are highly conserved between species there will inevitably be a small amount of variability which cannot be accounted for (Polz, 1998). This effect results in the favourable amplification of certain species' DNA over others during PCR as primers are more likely to anneal successfully if there is better base-pair matching. Alternatively, there is evidence to suggest that the most abundant PCR products are more likely to self-anneal in the later stages of PCR causing the progressive inhibition of template-primer hybrids (Suzuki, 1996). As the different enrichment backgrounds should result in different microbial communities, it is therefore likely that a degree of the difference in hit rates between samples grown using different growth media can be accounted for by a combination of these two primer bias factors.

In addition to interactions between primers and amplicons it has been suggested there is scope for primer-barcode bias (Tedersoo, 2015) in second stage PCR amplification during library preparation. Although the study that has investigated this potential problem proposed that conclusions about overall drivers of community diversity were not affected by primer-barcode bias, it was highlighted that different primer-barcode combinations yielded different results in terms of relative abundances of orders in the sample. While the focus of this study was on diversity within a fungal community, it is likely that findings are comparable to diversity within bacterial and archaeal kingdoms. Sequencing results from PCR, while they will not precisely represent all organisms grown present in iChips, are likely to reflect a range of organisms.

Given the aim of the experiment, to grow a broad range of organisms in isolation, it is not so important that some organisms may not have been represented in PCR products due to either primer bias or lack of targeting with primers. The broad range of organisms grown and isolated demonstrates relative success.

Another factor that has the potential to contribute to variable amplification rates between samples is the presence of inhibitors. As cells were stored at -20 °C before PCR was performed it is probable that stress factors were released by cells. If hydrolases or inhibitors were released, then it is possible that primers or polymerases were digested or inhibited in certain samples. While attempts were made to minimise any effects of this by the resuspension of cells in fresh nuclease free water before PCR was performed, it is possible that some stress factors remained in the sample solutions used in PCR.

It is also important to note the variable thermotolerance of different bacteria and archaeal cell walls. The initial denaturation step outlined in section 2.2.2 will be more effective in breaking the cell walls of some species of microbes than others. If the cells lyse at relatively low temperature, then the higher temperatures could result in DNA denaturation reducing the efficacy of PCR. Alternatively, some cells may not lyse at all if their cell walls are thermoresistant, meaning PCR is likely to be less effective. Fortunately the microbial communities present in mesophilic and thermophilic digesters appear to differ vastly (Gagliano, 2015) and so although a degree of the community will be made up of thermophiles it is likely to be a small proportion.

Following Index PCR (section 2.2.4.3), before concentrations were obtained for indexed amplicons, eight random samples were selected and run on a bioanalyzer (Appendix B). The results were used to calculate an average product length of 443 bp which confirmed the amplification of the correct length of PCR product in addition to gel electrophoresis results. This also facilitated the calculation of more accurate DNA concentrations.

3.2 Quality assessment of Illumina raw reads

$$\text{Phred (Q)score} = -10 \log_{10}P$$

Equation 1 Phred score calculation P in this instance indicates the probability of a base-calling error.

Using the program FastQC (Andrews, 2010), the Phred (Q) scores (Ewing, 1998) of Illumina sequence data were visualised (Appendix C). Based on a random selection of samples it was decided that truncation of forward reads to 260 bp and reverse reads to 200 bp was appropriate as this was a length at which all base-pairs in reads commonly achieved a Q score of >Q28. Q scores were originally used in the human genome sequencing project for assessment of sequence quality (Ewing, 1998). Within the FastQC program reads are separated into 3 bins of <Q20, Q20-Q28 and >Q28. In order to ensure a higher confidence in the base calling accuracy of the reads used for taxonomy assignment a Q28 cut-off was used. Had a lower cut-off Q score been selected, while longer reads would have been available for taxonomic assignment, it is likely there would have been a higher frequency of miscalled bases within the reads. Whether or not this would have led to different results in terms of taxonomy assignment is debatable as a filtering step within the DADA2 program works to remove reads with a maxEE value above a certain cut-off. The maxEE (Edgar, 2015) value is an alternative value to Q score, used to assess error rates in large biological samples when coverage is low. While a Q value is arbitrary and more useful for comparison between results of different sequencing methods the maxEE value is the maximum number of expected errors allowed in a read before removal.

At this stage it is not important that sequencing results only provide semi-quantitative information as this round of sequencing is used to identify samples of interest that will then be investigated in a second round of sequencing. Given the number of samples submitted for sequencing at this stage semi-quantitative sequencing methods are sufficient.

3.3 DADA2 output

DADA2 output was organised into reads derived from samples grown in each of the six growth media from section 2.1.3. Reads were grouped into amplicon sequence variants (ASVs) following filtering and steps to remove chimeras. Each ASV is made up of reads that are identical down to a single base-pair resolution. ASVs were assigned to a taxonomic group at each level or if this was not possible, they were termed 'NA'. The number of reads that make up each ASV are taken to reflect the actual abundance of that organism in the sample

population for simplicity, although there are a number of reasons why this might not be the case. Sequences that relate to the ASVs reported here have been deposited in the ENA database and under accession numbers ERS2967428 to ERS2967487.

In total 5,781,686 reads were submitted to the DADA2 pipeline. An average of 55840 non-chimeric quality filtered ASVs per pool resulted from the DADA2 processing. The actual number of non-chimeric quality filtered ASVs produced for each pool ranged from 0 to 225,428. The lower end of this range can be attributed to the mismatch between the eukaryotic organisms which the WS media enriched for and the specificity of primers for bacterial and archaeal 16S rRNA genes used in PCR.

Something to note at this point is that taxonomy was assigned from the 283 bp long amplicon from the V4 region of the 16S gene. As the full 16S gene is around 1.5 kb long it is possible that were another variable region selected for amplification instead, the combination of organisms indicated by DADA2 output would differ. This is not a concern however as, given the number of samples being investigated, it is unlikely that large differences in taxonomy will be seen with a change in the variable region selected. At this stage in the process results are useful to highlight overall patterns and potential organisms of interest within the whole sample set, not to provide precise and accurate identification of each organisms present.

3.4 Microbial diversity

Microbial diversity was considered at an order and genus level for the entire data set. ASVs were assigned to 13 genera more commonly than others. Here, criteria for genera to be considered common across the data set were >2% of all ASVs being assigned to them. These criteria are arbitrary and used for demonstrative purposes. Of these 13 genera, 6 were from the order *Clostridiales*, 2 from *Tissierellales* and 5 were from other orders (Appendix D)

$$Shannon\ Index\ (H) = \sum_{i=1}^N p_i \ln p_i$$

$$Simpson\ Index\ (D) = \frac{1}{\sum_{i=1}^N p_i^2}$$

Equations 2 Diversity Indices In both equations ‘*p*’ represents ‘*n/N*’ i.e. the number of a particular species found as a proportion of the total number of species.

Considering ASV diversity at a genus level, as defined by the Shannon (Shannon, 1948) and Simpson (Simpson, 1949) indices (equation 2), diversity in every set of samples was higher at a genus level compared to an order (Figure 7). This is expected given the relationship of the two taxonomic classifications. The two measures of diversity, however, did not imply the same relationship between samples from different growth backgrounds at a genus level. This can be attributed to the two different methods by which these indices are calculated. The Simpson Index is a measure of evenness within a sample; evenness being the measure how equal the abundance of each taxonomic group is within that sample at any one particular level. By comparison, the Shannon index measures both evenness and richness of a sample. Richness is

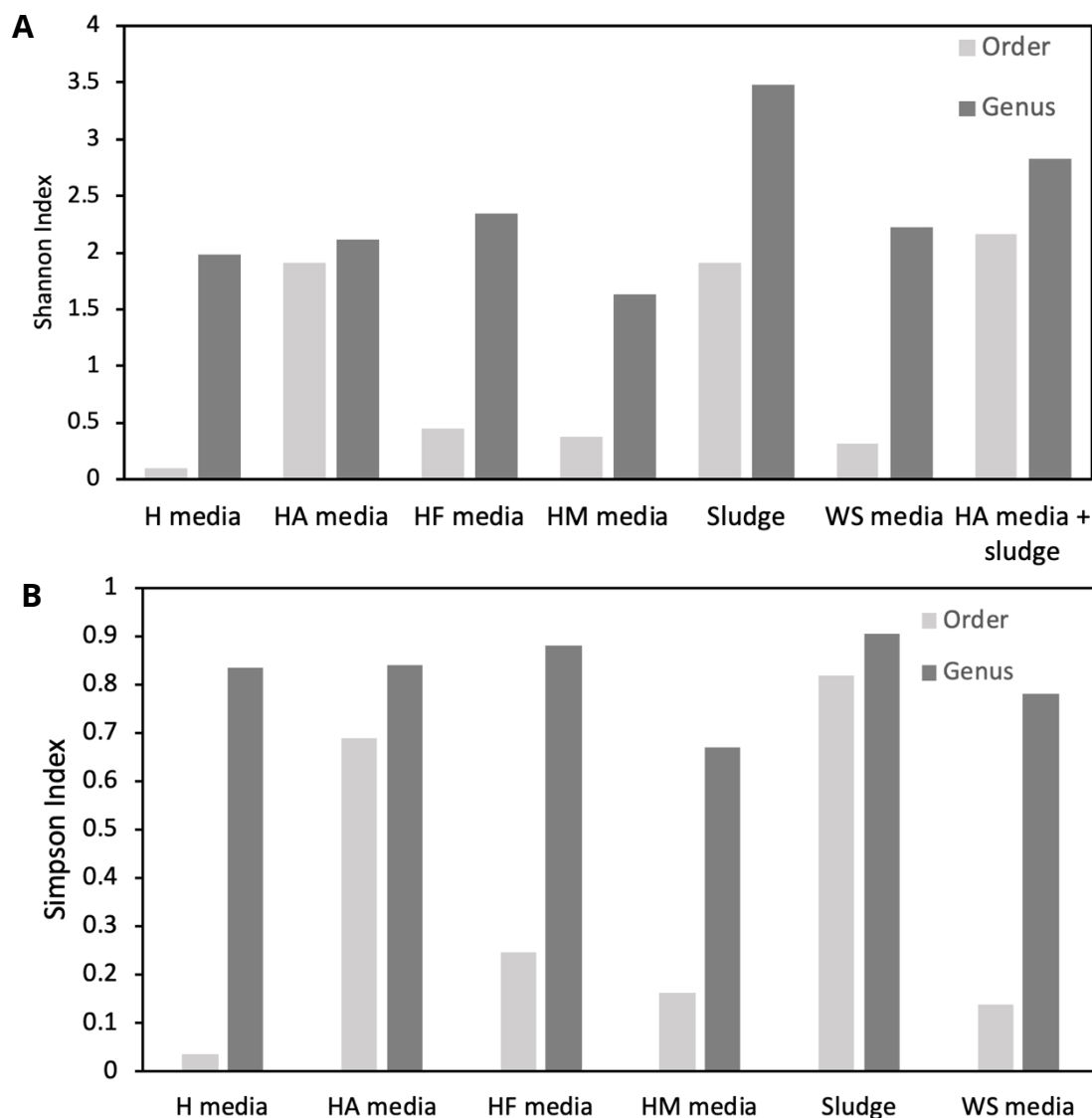


Figure 7 Comparison of Diversity Indices for ASVs Taxonomic classes were assigned as ‘other’ if the number of ASVs assigned to them contributed to less than 1% of the total ASV count for that sample set at that taxonomic level.

defined as the number of different types of a taxonomic group in a sample. Given this difference in the indices, the comparatively larger diversity of the WS sample at a genus level, calculated using the Shannon Index as opposed to the Simpson index, makes sense. While evenness of the WS media grown sample set (Simpson Index = 0.14) is less than in the samples grown using HM media (Simpson index = 0.16), HF media (Simpson Index = 0.25), HM media (Simpson index = 0.69) and H media (Simpson index = 0.82), the WS media grown sample set is richer. This makes the WS media grown samples the third most diverse of the six sample sets at a genus level when measured using the Shannon index. When considering the results of the experiment the diversity indices of both the Shannon and Simpson Index should both be considered as they provide different insights in to the data.

A useful implementation of diversity indices in this instance is to test whether the differing growth backgrounds have yielded different microbial communities. When ASVs sets are collated from samples grown using HA media and sludge, the diversity of the collated ASV set at an order level (Shannon Index = 2.17) is 13.6% higher in comparison to the two uncollated data sets (both Shannon Index = 1.91). This supports the suggestion that the enrichment media has encouraged the growth of a different set of microbes (Figure 8). If similar microbial populations were observed as a result of both nutrient backgrounds, then collation of the ASV sets from two different growth media backgrounds would not show any difference in Shannon index value.

A significant pattern seen at an order level is the prevalence of reads assigned to organisms from orders *Desulfovibrionales*, *Methanomassillicoccales*, *Selenomonadales* and *Synergistales* which are not seen in comparable amounts in samples cultured from any growth background other than sludge (Figure 8). This highlights again the effect that media composition is having on the combinations of organisms seen after assignment of reads. Considering the diversity of ASVs at an order level using the Shannon index (Figure 7), both sludge (Shannon index =1.91) and HA media (Shannon index =1.91), as growth backgrounds, yielded sequencing results suggesting significantly higher levels of diversity in comparison to the other growth backgrounds of H media (Shannon Index =0.10), HF media (Shannon Index=0.44), HM media (Shannon Index = 0.34) and WS media (Shannon Index = 0.32). This was mirrored by analysis using the Simpson index.

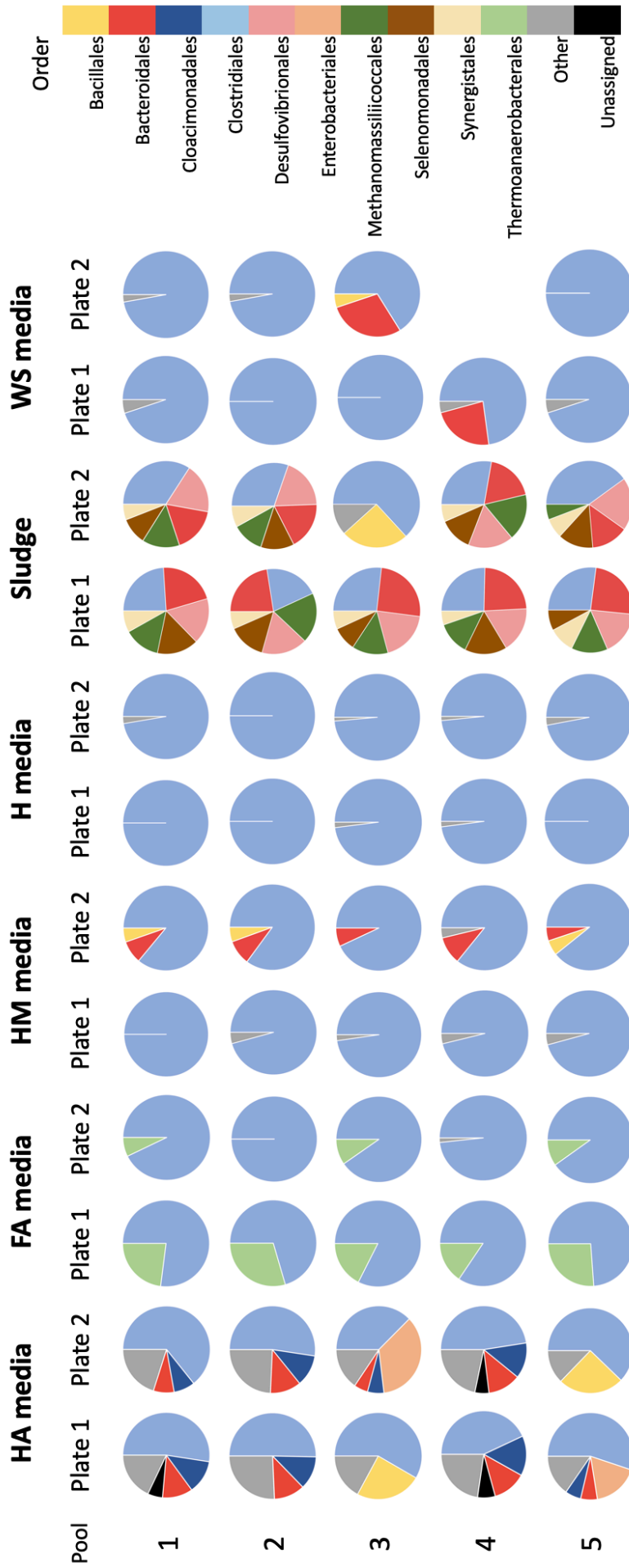


Figure 8 Taxonomic distribution of assigned amplicon sequence variants (Order level) The distribution of ASVs assigned to different orders by DADA2. Orders represented by reads that cumulatively make up less than 5% of the total number of reads for each growth background were binned as 'Other'.

It is important to note that conclusions cannot be drawn regarding the ability of the HA media to yield a more diverse population of microbes in comparison to the other enrichment media

based on these diversity indices. Instead, what should be focussed on is the microbial population that is produced using each media and the metabolic capabilities of these populations. Globally the ecosystems within anaerobic digesters will all vary due to the feedstock and environmental conditions (Vasco-Correa, 2018), this will result in a variety of different microbiomes. Bearing this in mind, the results from this experiment will have depended on the original microbiome of the anaerobic digester from Naburn Sewage Treatment Works and as such if this experiment were repeated using a sample from an anaerobic digester at a different location it is possible that different results would be obtained (Saunders, 2016). As highlighted by Saunders *et al.*, there is likely to be some similarity in a selection of organisms with the key core metabolic capabilities necessary for the AD process. Given the primary aim of this work was to gauge the successfulness of this approach for isolating and culturing novel microbes, the repeatability of this experiment in terms of specific species seen in the final sequencing data is of little concern. Given the apparent success of this method though, sampling from other anaerobic digesters could be of interest in future work.

61.7% of all ASVs were assigned to the order *Clostridiales*. As a defining characteristic of *Clostridiales* is their obligate anaerobic nature, this result is not surprising. Genera from this order are diverse, this is reflected in the data here as no more than 6.9% of ASVs were assigned to any one particular *sensu stricto* genus within the *Clostridiales* order (Figure 8). Organisms within the *Clostridium* genera can be split into several different groups at a genus level by addition of the term ‘sensu stricto’ beforehand. This implies the genus is being referred to in a ‘narrow sense’ or in the sense of the original author. In the case of the *Clostridium* genus, the term *sensu stricto* is used to separate out organisms that belong to *Clostridium* into different groups at a genus level.

Other studies into AD microbial communities have reported a high abundance of organisms from the *Clostridiales* order (Sun, 2015). Sun, Lui *et al.* (2016) have also reported a similar numerical abundance of organisms from the *Bacteroidales* order as well as *Clostridiales* which is not seen in data from this experiment. While 68.9% of ASVs were assigned to the *Clostridiales* order only 7.9% of ASVs were assigned to the *Bacteroidales* order here. This could be explained by a number of different factors. Firstly, the iChip method for culturing coupled with the enrichment media used could favour *Clostridiales* growth over *Bacteroidales*

growth through the microbial community created. This bias is likely to affect other organisms in a similar manner as well. In addition to this problem of a bias in culturing, primer bias as discussed in section 3.2 will contribute to disparity between the actual microbial population present and that presented in sequencing data. This in itself is not an issue as the aim of the experiment was to isolate and culture a wide diversity of organisms. This deviation from previously recorded ratios of organisms to one another (Sun, 2015) in fact provides evidence to suggest that there is further scope for creating different combinations of isolated organisms using other growth media. Small changes could be made to the culturing methods used such as using a phosphate containing agar that could yield different combinations of organisms (Tanaka, 2014).

The different patterns seen in Figure 8 also support the suggestion that variations in enrichment media were successful in culturing different combinations of organisms. For example, whilst organisms grown in HA media produced reads assigned to the *Cloacimonadales* order this was not seen in comparable levels in sequencing results from organisms grown in any other enrichment media. Likewise, considerable numbers of reads from samples grown in HF media were assigned to the order *Thermoanaerobacterales*, something not seen from samples in other enrichment backgrounds. While there appears to be no obvious link between formate and *Thermoanaerobacterales* metabolism the lack of acetate and methane in the media will have affected the overall microbiome composition. Interestingly all reads assigned to the *Thermoanaerobacterales*, at an order level, within reads from samples grown using the HF media have also been assigned to *Tepidanaerobacter* at a genus level and have not been assigned at a species level. Using BLAST (Altschul, 1990) it was discerned that these reads shared the highest sequence similarity with a number of uncultured bacterial clones, of particular note one within the *Thermoanaerobacterales* order. While it is not possible, given time constraints, to investigate all samples that have been assigned to uncultured species, these data provides scope for future investigations.

3.5 Samples identified for further investigation

Studying the data at a genus level (Appendix D), there were a number reads assigned to four different candidate genera: *Candidatus Cloacamonas* and *Candidatus Methanofastidiosum* from the HA media background, *Candidatus Methanoplasma* from the sludge growth background and *Candidatus Berkiella* from the HA media background. The number of ASVs assigned to these organisms suggested that only either *Candidatus Cloacamonas* or *Candidatus*

Methanoplasma would be worth pursuing in an attempt to identify a pure culture, given time constraints. Of these two *Candidatus Cloacamonas* was selected for a number of reasons.

Candidatus Cloacamonas acidaminovorans was first identified from metagenomic data of a sample from a wastewater treatment plant (Pelletier, 2008). Since then, its genome has been studied to better understand its evolutionary history, however, little progress has been made (Rinke, 2013). Based on analysis of genomic data it has been proposed that *Candidatus Cloacamonas acidaminovorans* grows optimally under anaerobic conditions but does possess some genes for the production of enzymes involved in coping with a degree of oxidative stress such as superoxide reductase and thioredoxin reductase (Pelletier, 2008). Previous studies have reported maintenance of *Cloacamonas acidaminovorans* for up to 1 year, although, have not managed to achieve any significant increase in cell count over this period. Its disappearance and reappearance in metadata from different time points within digesters has also been reported and so it appears difficult to enrich via growth experiments. It was hoped that the use of iChips may have provided a breakthrough in this respect, but it seemed as though, if *Candidatus Cloacamonas* was indeed present in any of the wells that were investigated, a further iChip experiment may be necessary (4.2.2). The wells for further investigation were selected based upon the percentage of reads from each pool assigned to *Candidatus Cloacimonas* within the HA media enriched pools. As pools 1,2 and 3 intersected with pools 4 and 5 and pools 6, 7 and 8 intersected with pools 9 and 10 the sixteen wells from the overlap of the two pools with the highest percentage of reads assigned to *Candidatus Cloacimonas* would be investigated further. For *Candidatus Cloacimonas* these were wells from the overlap of pools 2 and 4.

Visualisation of FISH stained 5A (Figure 9) and 5B samples showed the binding of the S-*-WWE1-1181-a-A-18 probe (Chouari, 2005), designed to bind *Candidatus Cloacamonas*, to all cells. This suggested that either all cells present were *Candidatus Cloacamonas* or the probe was not specific enough. After probing pure cultures of *E. coli* to test the specificity, visualisation revealed that the probe was not specific to *Candidatus Cloacamonas* as first indicated. The online SILVA TestProbe tool (Quast, 2012) suggested that the probe was specific to *Candidatus Cloacamonas*, however, upon alignment using the NCBI BLAST tool (Altschul, 1990) there were many more cultured and uncultured organisms that had a 100 % 'Query cover' and 'Identity' match to the probe sequence. This would suggest that the probe would bind the 16S gene in these other organisms. Upon further inspection with reference to a study investigating averaged variable region positions across a wide range of 16S genes (Yarza, 2014) it appeared that the probe sequence was designed to bind primarily within a conserved

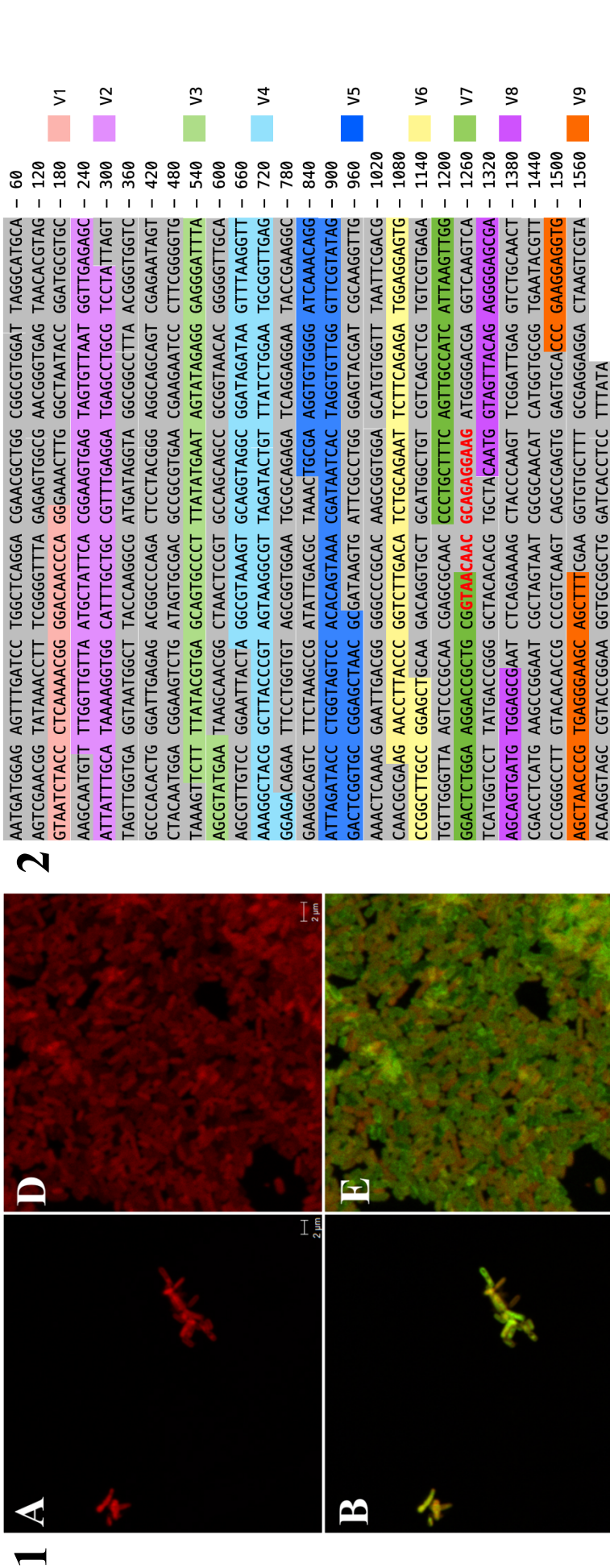


Figure 9.1 FISH probe S-*:-WWE1-1181-a-A-18 (A-C) Cells from well 5A of plate 1 of HA media grown samples. (D-F) *E. coli* cells. (A,D) probe binding indicates the presence of bacterial or archaeal cells. (C,F) probe binding indicates the presence of *Candidatus Cloacimonas*. (B,E) an overlay of A and C or D and F respectively. **Figure 9.2** The reverse complementary 16S rRNA gene from *Candidatus Cloacimonas* acidaminovorans (Pelletier, 2008) (gene position: 1799643-1801248). Grey highlighted sequences indicate conserved regions. Other colours indicate variable regions. The reverse complementary gene was used as the *Cloacimonas* 16S rRNA gene is reverse transcribed. Variable and conserved sequence positions were also shifted 53 bp downstream in relation to the positions given (Yarza, 2014), as BLAST comparison of the *E. coli* and *Candidatus C. acidaminovorans* 16S rRNA genes indicated base pair matching with a 53 bp displacement.

region of the 16S gene, for all but 4 base pairs, on the border between the V7 region and the conserved region before the V8 region. The study by Yarza, P *et al.* (2014) uses an average of all 16S genes in the SILVA database to designate the 16S variable and conserved regions, a probe more likely to have a higher specificity would be found firmly within one of the variable regions as opposed to on the border with a conserved region.

The probe S-*-WWE1-408-a-A-19 (5'- GCT CCG AAA AGC TTC ATC G -3') (Limam, 2014), could have been tried as an alternative. Images from this study, however, did not seem to provide convincing evidence of the probe functioning as a reliable indicator of the presence of *Candidatus Cloacamonas*. Upon using BLAST to compare the probe to the *Candidatus Cloacamonas acidaminovorans* 16S rRNA gene there was only limited base pairing found, suggesting probable weak probe binding (Appendix E). The SILVA TestProbe tool also gave a similar result with the highest percentage coverage match at only 20% for the order *Cloacimonadales*.

Candidatus Methanoplasma, the other candidate organism which could have been pursued, was first identified from a highly enriched sample from the intestinal tract of a termite (Lang, 2015). It has, however, been identified as one of the dominant species of methanogens in manure (Ozbayram, 2018), suggesting that its presence in anaerobic digesters is very possible (Ozbayram, 2018). Sixteen wells from the overlapping two pools containing the most reads assigned to *Candidatus Methanoplasma* can be investigated further using FISH.

Sanger sequencing of the 16S rRNA gene was performed to ascertain the purity and identity of each of the sixteen acetate cultures selected as an alternative to FISH. Of the sixteen samples, chromatograms indicated that four of these contained no microbes, four were pure cultures and eight were mixed cultures. Gel electrophoresis also suggested that PCR had been unsuccessful for the four negative samples identified through Sanger sequencing. Prior to BLAST analysis, raw reads were examined using sequence data visualisation software within Unipro UGENE (Okonechnikov, 2012), a multiplatform bioinformatics program. Based on the sequence quality as per the chromatograms (Figure S6), sequences were trimmed to ensure only low error rate reads were used for gene comparison. The results from pure cultures were identified using the NCBI BLAST tool. Two of these pure cultures were identified by BLAST tool as *Sporanaerobacter acetigenes* and another as *Tissierella praeacuta* (Appendix F). The fourth pure culture, located in well 7D of plate 1 of the HA media isolates, was identified as an uncultured bacterium (accession no. EF559175.1) by BLAST. No known cultured microbes

within the NCBI reference sequence database (O’Leary, 2016) matched to more than 97% sequence similarity; the closest cultured microbe being *Mobilitalea sibirica*. The uncultured bacterium was originally identified from a sample from an anaerobic digester run under mesophilic conditions (Li, 2009). While the BLAST tool results (Appendix F) paired with the sequencing results chromatogram (Appendix G.12) indicate a pure culture of this previously uncultured bacterium it will be necessary to perform additional experiments such as DNA:DNA hybridisation to confirm this.

Using the program MEGA7 (Kumar, 2016) evolutionary analyses were undertaken comparing nine of the closest matching sequences, as defined by the BLAST tool, from the NCBI database. The closest matching sequences were used in order to demonstrate the phylogenetic position of the previously uncultured organism at the most specific level currently possible (Figure 10). It provided evidence to support the BLAST tool results that suggested the most closely related cultured organism is *Mobilitalea sibirica*. Interestingly, analysis results suggest all of the uncultured organisms compared are more closely related than cultured organisms in the database. Analyses also highlight the fact that the cultured organisms with the highest sequence similarities, whilst they all belong to the *Firmicutes* phylum, all belong to different genera. All of these organisms have also been identified as anaerobic and as having the ability

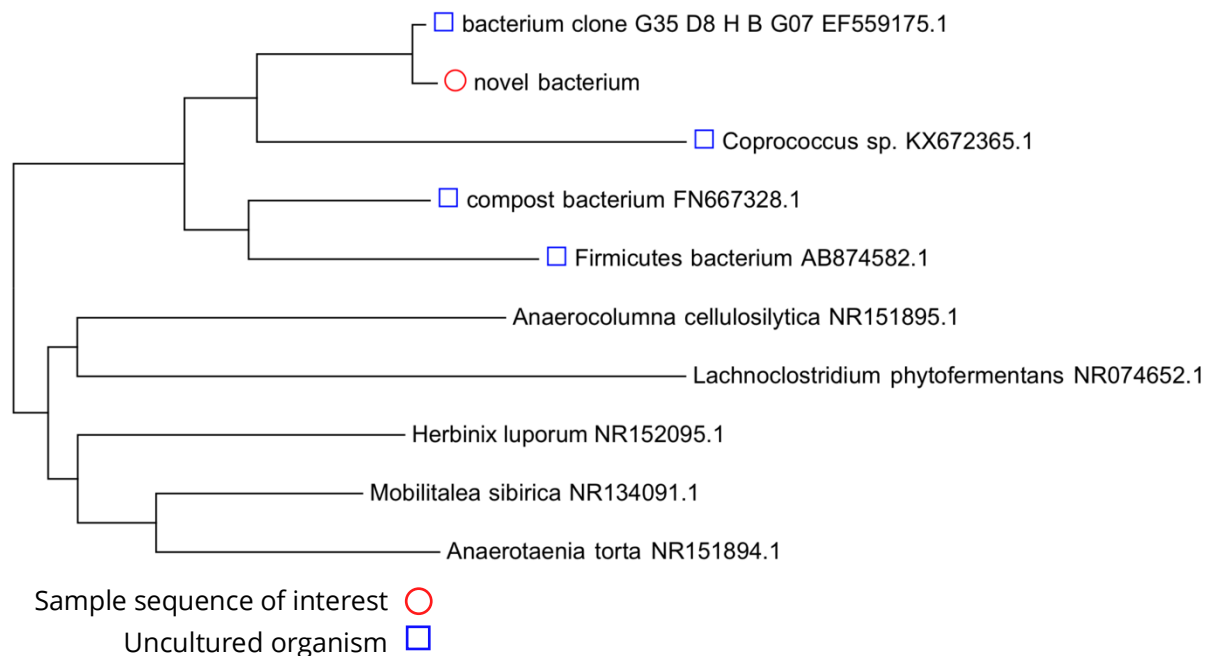


Figure 10 Phylogenetic Tree of BLAST results for potential novel bacterium The Neighbour-Joining method (Saitou, 1987) was used to calculate the evolutionary history of the shown organisms. The tree is drawn to scale, with the length of branches proportional to the evolutionary distances between organisms. The evolutionary distances were computed using the p-distance method (Thomas, 2001). 668 bp were used for gene comparison. Labels give the taxonomic name and accession number.

to produce enzymes with cellulolytic capabilities (Methé, 2002; Podosokorskaya, 2014; Koeck, 2016; Ueki, 2016). This would suggest that this potential novel organism would also share these metabolic capabilities and would contribute to the hydrolytic stages of AD. Of these five cultured organisms it is also relevant to mention that, *H. luporum*, *A. torta* and *A. cellulositytica* were isolated from anaerobic digestors or biogas plants.

3.6 Examining metabolic capabilities of prevalent species

Within samples grown from an HA media growth background there are a number of species present that are indicative of an acetate rich anaerobic environment. DADA2 did not assign 61.3% of all ASVs from samples grown in an HA media environment to a species. Of these 48.7% were not assigned at a genus level either and 7.5% are not assigned at a family level. This shows promise in future work for the isolation of novel organisms. The most prevalent species that ASVs were assigned to were *Tissierella carlieri*, *Sporanaerobacter acetigenes* and *Clostridium cochlearium* accounting for 14.8%, 10.2% and 7.2% of all HA media grown samples ASVs respectively.

Tissierella carlieri is a strict anaerobe that is known to produce acetate amongst other carboxylic acids, when grown using a Tryptose Glucose Yeast Extract broth (Alauzet, 2014). The key catabolic activities of *T. carlieri* as highlighted by Alauzet *et al.* involve the breakdown of gelatin, a material comprising the fibrous protein collagen and other peptides. Given that H media contains trypticase peptone this could partially explain the high prevalence of *T. carlieri*, however, there is little *in vivo* experimental evidence to inform us of the growth requirements of *T. carlieri*. Having been identified from a series of cultures previously classified as *T. praeacuta*, which itself has been isolated from infant and adult human faeces (Ricaboni, 2017), its presence in anaerobic sludge at a wastewater treatment facility is fitting. Interestingly, the only other growth background which yielded a significant number of *T. praeacuta* reads was sludge. This implies that the HA media and the sludge growth background were creating an environment more suited to *T. praeacuta* than either HF media or HM media.

Sporanaerobacter acetigenes, originally isolated from an upflow anaerobic sludge blanket reactor in Mexico (Garcia, 2002), is both strictly anaerobic and moderately thermophilic, although the reactor it was first isolated from operated under mesophilic conditions. Its prevalence in samples grown using different growth media is not surprising as it is a common acetogen that also produces H₂ and CO₂ in the acetogenesis process. This is therefore an

organism that is likely to exist in microbiomes that favour acetoclastic and hydrogenotrophic methanogenesis. This is reflected in the data as the only growth media background in which *S. acetigenes* reads are not seen in significant numbers is HM media, which is likely to create a microbiome in which methylotrophic methanogenesis is favoured. Samples grown using HF had a lower percentage of ASVs assigned to *S. acetigenes* in comparison to HA media (2.9% and 10.2% respectively). It could be hypothesised that the abundance of *S. acetigenes* seen in the HA media grown samples can be attributed to the fact that there may be a higher level of acetoclastic methanogenesis occurring in this microbiome creating a microbiome to favour a larger number of acetogens through co-dependance.

The third most abundant species present according to DADA2 assignment of ASVs from HA media is *Clostridium cochlearium*. *C. cochlearium* is seen in all six different growth backgrounds and is least abundant in samples grown using sludge, using the percentage of total ASVs for each growth media sample set as a measure of abundance. This, however, could be a result of a higher diversity of species in this case. First defined as a species by Nakamura, Okado, *et al.* (1979), it is an anaerobic fermentative bacteria with metabolic functions that include the catabolism of glutamate to acetate (Buckel, 1974). Of the three altered H media (HA, HF and HM media) *S. acetigenes* is least abundant in the HA background. This is the opposite pattern to that which was observed with *S. acetigenes*. Without metagenomic data it is difficult to explain this in terms of the metabolic interactions occurring within the enriched microbiome.

The original enrichment process targeted the growth of methanogens through various different methanogenesis pathways. While the only data which indicates the growth of a methanogen from the use of an enrichment media in significant amounts comes from the samples grown using WS media, this result may be useful. What has been demonstrated is the potential for the growth of different microbial populations from the same sample using these enrichment media and the isolation of a number of potential novel organisms.

Data suggest that methanogen *Candidatus Methanoplasma* was present in samples grown using WS media. It has been postulated that *Candidatus Methanoplasma* is an obligate methylotrophic methanogen (Lang, 2015). It would therefore be expected that the HM media enrichment cultures would be most likely to yield its growth. Results suggest that it is only present in samples from sludge and WS media growth backgrounds, although it is significantly more abundant in samples grown using sludge. 12.4% of ASVs from a sludge growth

background were assigned to the order *Methanomassiliicoccales* as can be seen from Figure 8. At a genus level 10.4% of all reads from the sludge growth background were assigned to *Candidatus Methanoplasma*. While this set of samples were not a result of enrichment it does demonstrate the ability of the ichip method of cultivation to potentially isolate candidate methanogens from anaerobic sludge.

Chapter 4

General Discussion

4.1 Aims

The overall aims of this study were to trial and evaluate a novel method for microbial isolation and culturing from anaerobic sludge. In doing this it was necessary to identify at least one novel organism to demonstrate the efficacy of this approach. This was approached by enriching aliquots of the sample for different methanogens as well as anaerobic fungi. Given the vast quantity of data generated from 1,152 samples there was not sufficient time to analyse all of it to the extent that would provide comprehensive information on all potential novel organisms present and potentially isolated. As the 16S rRNA genes of only sixteen of the 1,152 samples have been individually sequenced fully there is a large scope for continuation of this study. Of these sixteen samples one was a potentially previously uncultured organism. While there was not enough time to investigate any other set of samples this yield suggests that the technique has indeed worked as desired and there could be other cultures of interest within the isolate sample population.

While this success is not a direct result of the enrichment process as the sample isolated is not a methanogen, the diversity generated between enrichment backgrounds could hold promise for future studies, not necessarily limiting the enrichment process to providing growth conditions for methanogens. The complexity of the sample community within an anaerobic digester aids this in the sense that small changes in environmental conditions, including nutrients provided, can lead to the development of different microbiomes. This provides ample scope for future studies using similar methods with minor adjustments potentially yielding different results in terms of resultant microbial populations.

16S rRNA genetic data from this experiment suggests that the investigated sample cultured in an HA media growth background is a novel species based on the application of the Naïve Bayesian classifier method of taxonomic classification (Wang, 2007). In order to confirm that the culture is in fact a novel species the next step could be to use DNA:DNA hybridisation for comparisons of genetic similarity between the organism and known species. This method of identification for novel species, if it implied the genetic novelty of this organism, would fulfil the criteria for a species as defined by a pragmatic species concept as discussed in section 1.2.2.

4.2 Future Work

4.2.1 Raman Spectroscopy

The most time-consuming part of the process, with such a large sample size, was the PCR step for amplification of a region of the 16S rRNA gene. While High-Throughput PCR techniques (Schmittgen, 2008) could be used followed by Illumina dye sequencing, an alternative method of rapid microbe identification could be Raman Spectroscopy. This was not attempted in this instance due to time constraints, however, if samples were grown using appropriate ¹³C containing compounds as well as ¹²C containing compounds it is possible this technique could speed up the overall process. Raman spectroscopy works in a similar manner to other spectroscopy techniques in that a structural fingerprint is developed from which molecules and even whole microbes can be identified by comparison to a database (Strola, 2014). It differs from other spectroscopy techniques in that it relies on the inelastic scattering of monochromatic light. Raman spectroscopy is becoming an increasingly popular technology for the characterisation and identification of microbes. Tip enhanced Raman spectroscopy (TERS) enhances the signal of normal Raman spectroscopy to allow the development of a spectra from single cells (Huang, 2004). TERS is something that could vastly increase the efficiency of this whole process. By performing Raman spectroscopy on each sample, which would be relatively quick and affordable, you could eliminate any previously cultivated samples and highlight previously uncultivated or unidentified species for further investigation. This could be done at the beginning of the cultivation stage (Lorenz, 2017) and undertaken while cultures are still growing just after the ichip stage of the study.

Implementation of this approach would rely on the establishment of a comprehensive spectral database (Stöckel, 2016) for comparison of the spectral fingerprints for each microbe. While there are a number of different databases currently running, development continues.

4.2.2 Further ichip isolation

As the enumeration of cells in the primary stages of this experiment was done by estimation it is understandable that some of the resultant cultures were not pure. This experiment could be repeated using a different technique for cell enumeration in the earlier stages (section 2.1.5). For example, flow cytometry could be used. This could limit the need for a second round of isolation from iChip grown cultures required in some cases.

In addition to repeating the experiment using freshly collected anaerobic sludge, samples generated from this experiment could be used in conjunction with iChips to isolate and produce pure cultures. In the instance where it is suspected that a mixed culture comprises a microbe of interest it is possible to then repeat the process of growth and isolation using the mixed culture as a source of cells. As the samples are a product of a first round of iChip isolation and growth they should be purer than the original sludge sample that cells were isolated from. This means that, in the case of *Candidatus Cloacimonas* for example, there is an increased chance of isolating the desired cells and being able to grow them into pure cultures. There will still be the problem of obligate cogrowth in the later stages of the experiment, however, it will be clearer in this experiment whether this is the limiting factor in growing certain fastidious organisms such as *Candidatus Cloacamonas*.

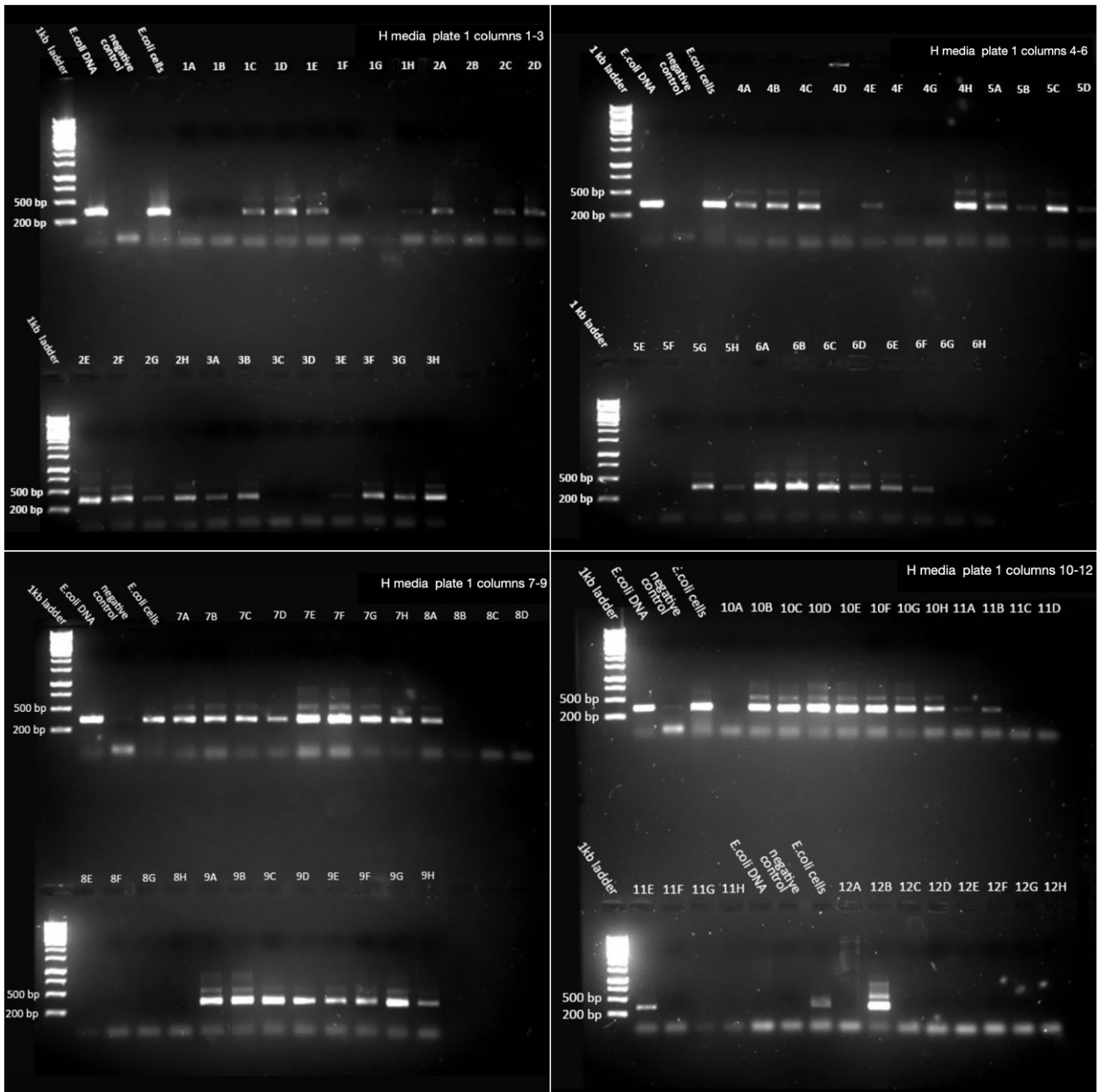
4.2.3 Further Data Analysis

Given the volume of data generated in this experiment there is still a large quantity of sequencing data to analyse. Within this, certain data of interest has already been identified such as a large volume of reads within the sludge grown samples assigned to the candidate genus *Candidatus Methanoplasma*. This is of interest firstly because 10.4% of all the sludge ASVs were assigned to this genus so the organism is relatively abundant. Additionally, a new mode of energy metabolism has recently been identified within *Candidatus Methanoplasma*. Isolating and characterising this microbe would provide further insight into this. To achieve this the sixteen wells at the overlap of pools 2 and 5 from the sludge grown samples would be further investigated using Sanger sequencing as described in section 3.3. Another candidate organism that was identified within data was *Candidatus Methanofastidiosum*. This was found to be the most abundant in pools 7 and 9 from the acetate grown samples and so would be investigated in the same manner.

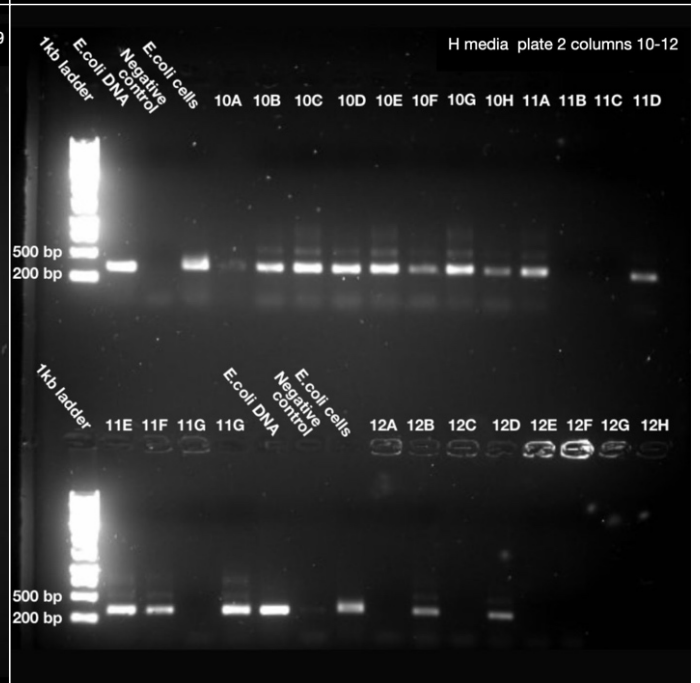
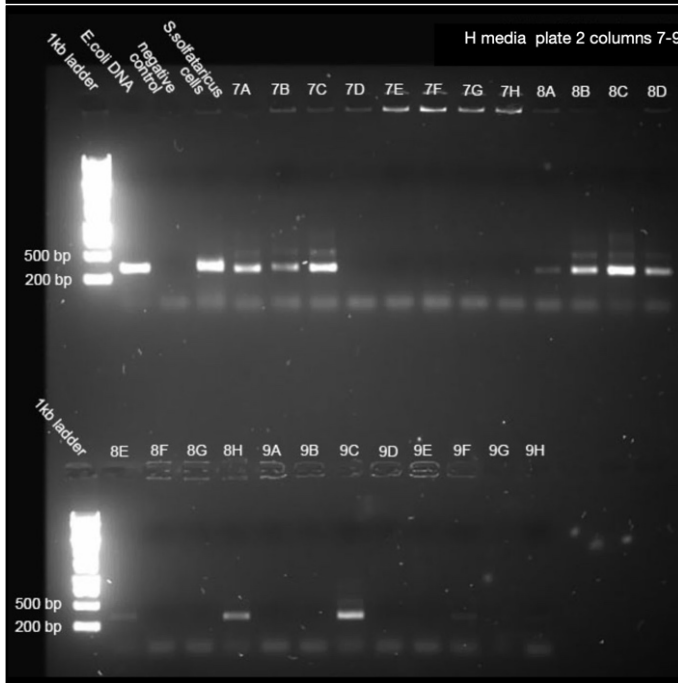
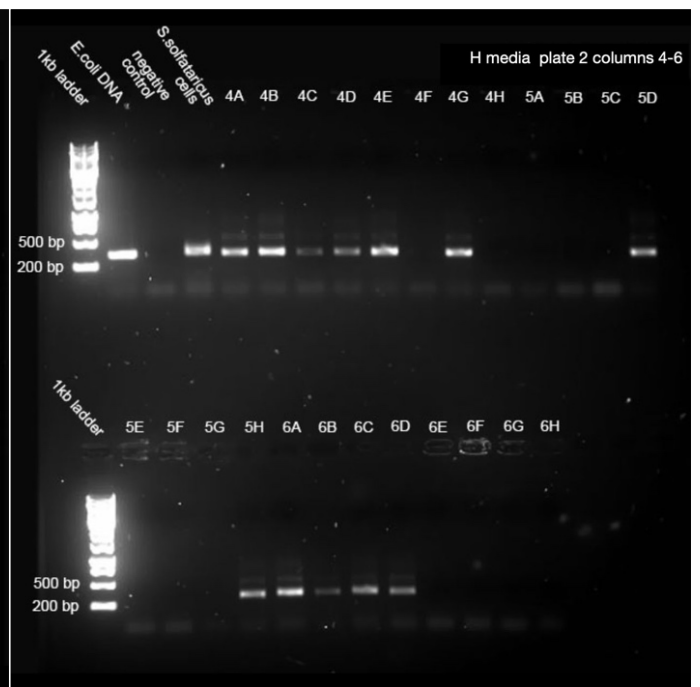
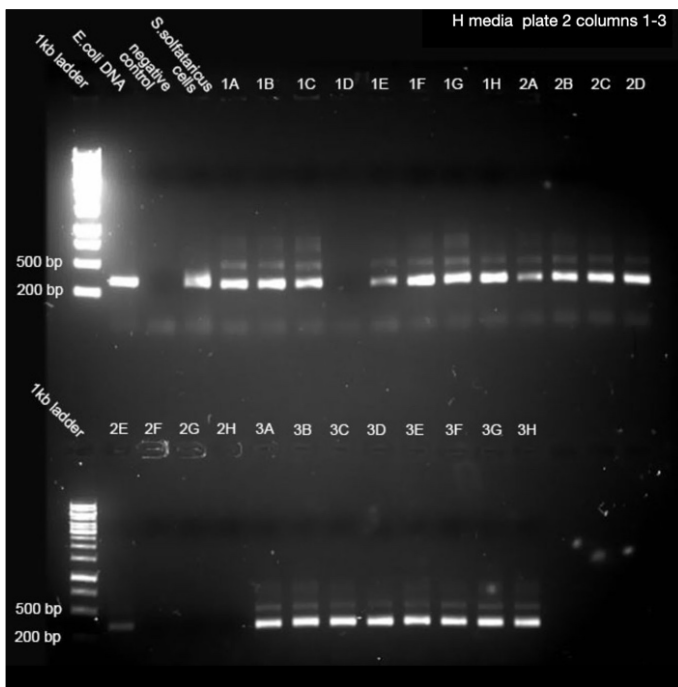
18.5 % of all reads were not assigned to a genus. All of these unassigned ASVs could be investigated further through using the BLAST tool to find the nearest known 16S rRNA gene in the NCBI database. Following this, anything with <97% sequence similarity to a known sequence would be followed up by Sanger sequencing of a section of the 16S rRNA gene comprising all variable regions. This could take a considerable amount of time given the number of unassigned reads. A suggestion would therefore be to further investigate a number of ASVs unassigned at a species level. The success of identifying novel species within these ASVs would provide further insight into the efficiency of this method for the culturing of novel species from anaerobic sludge. Based on result analysis so far this work provides support for

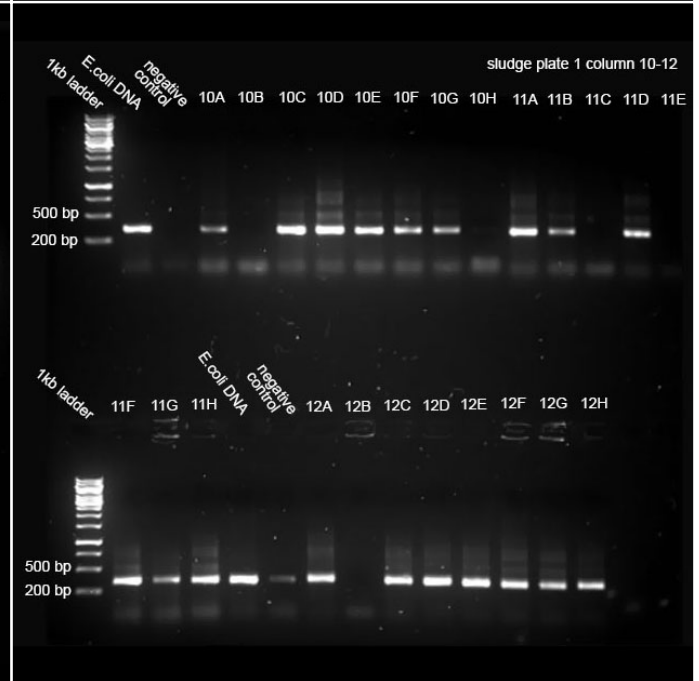
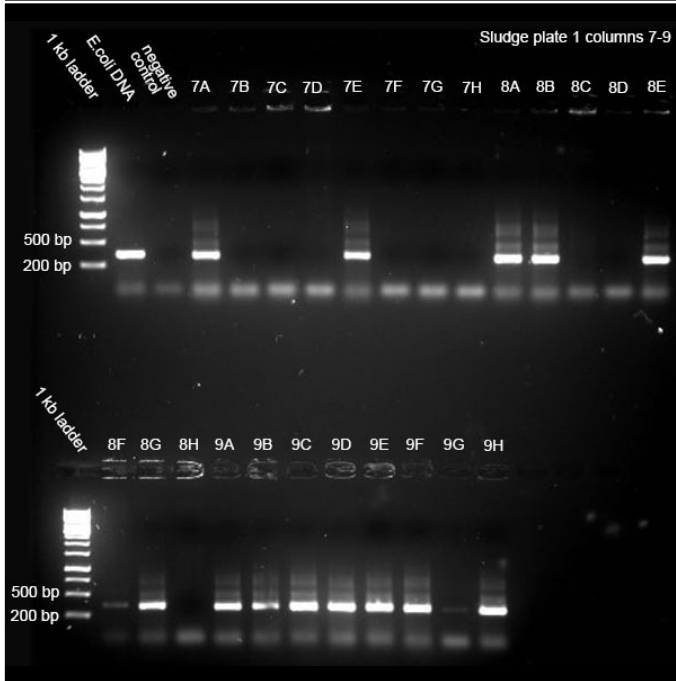
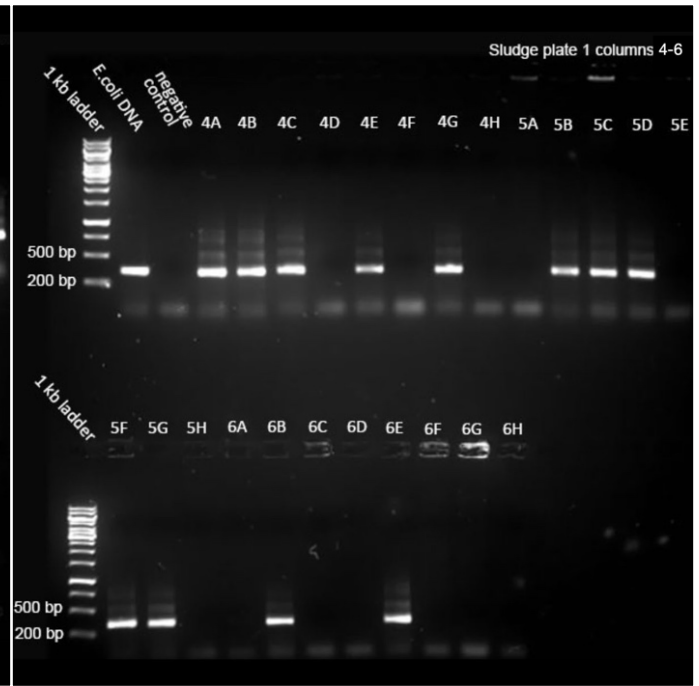
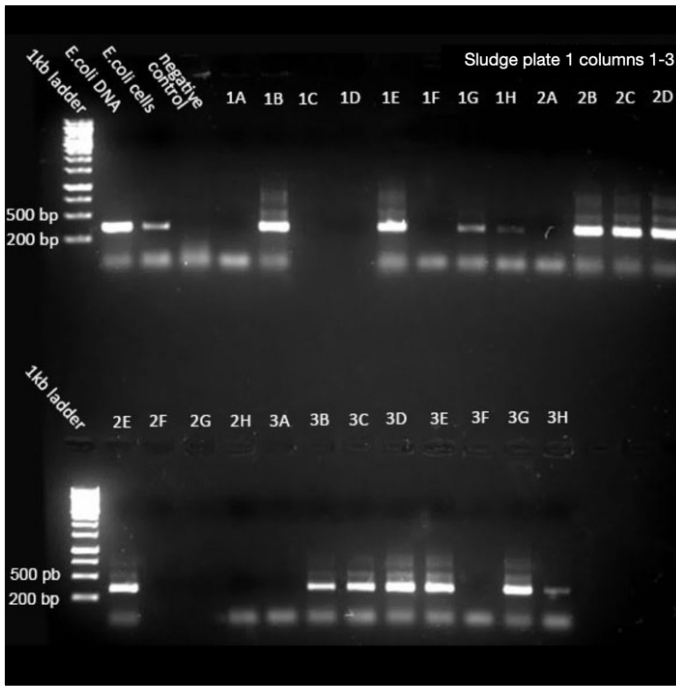
the use of iChips in the attempt to combat the problem of the great plate count anomaly in general and shows promise for future studies using the iChip method of culturing. In addition to this there is potential for the process used to be streamlined in future studies.

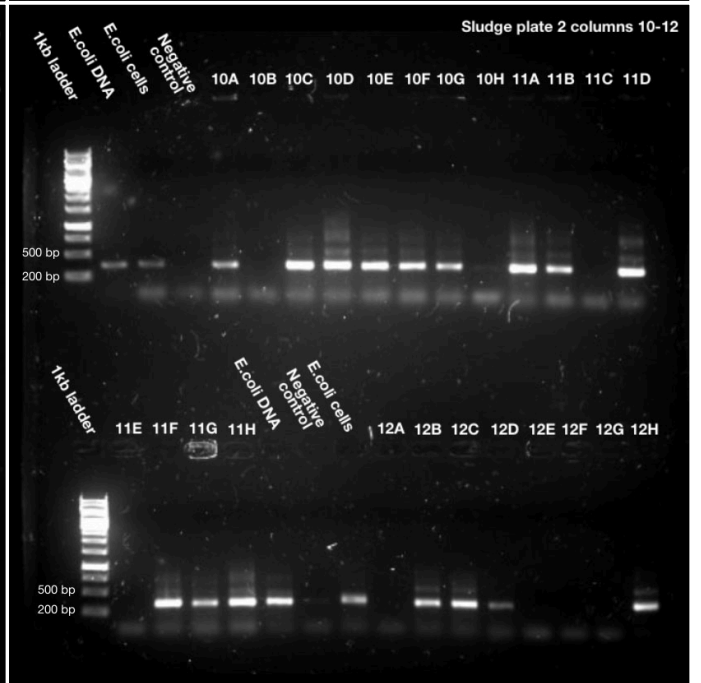
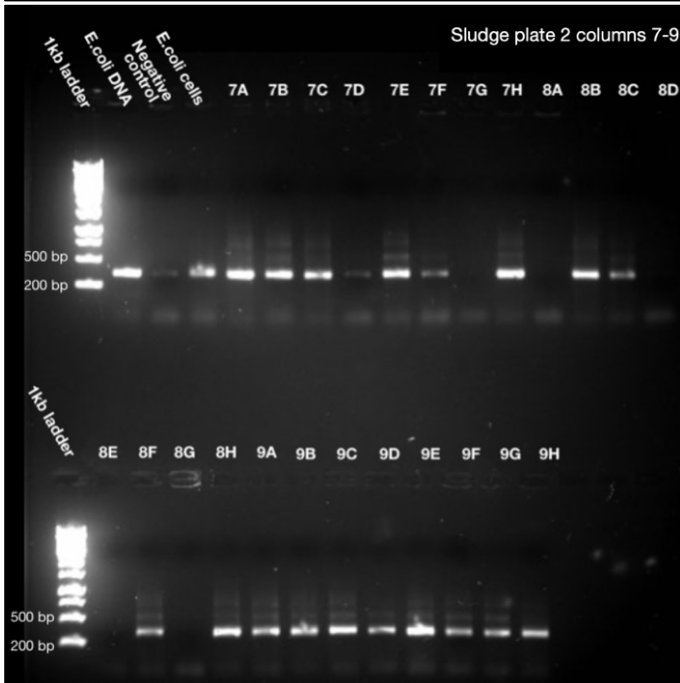
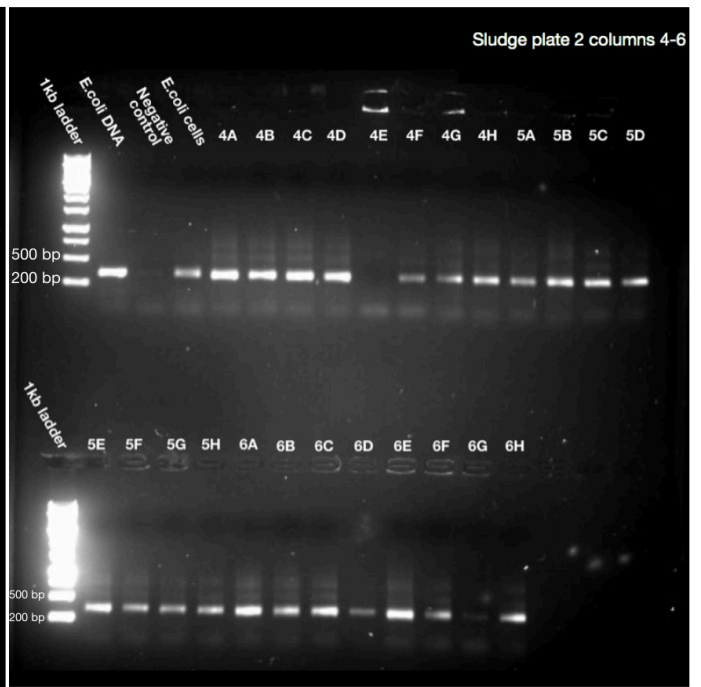
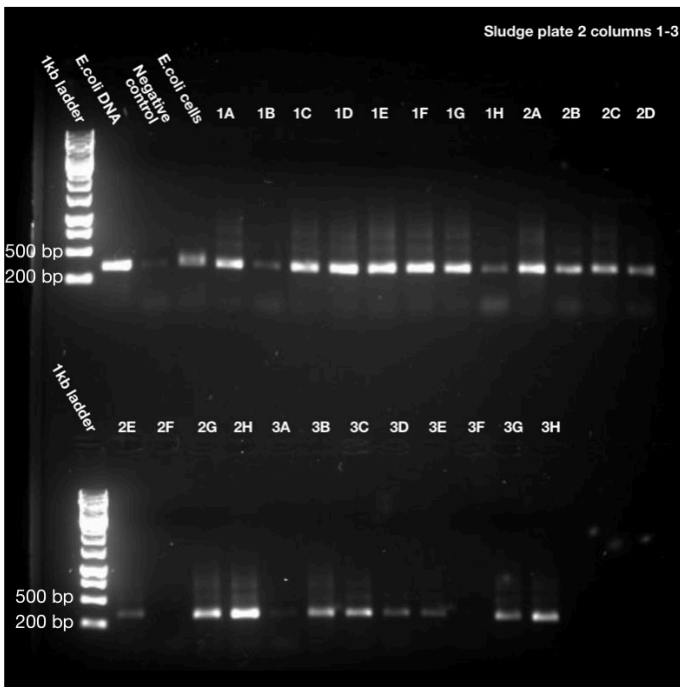
Appendices

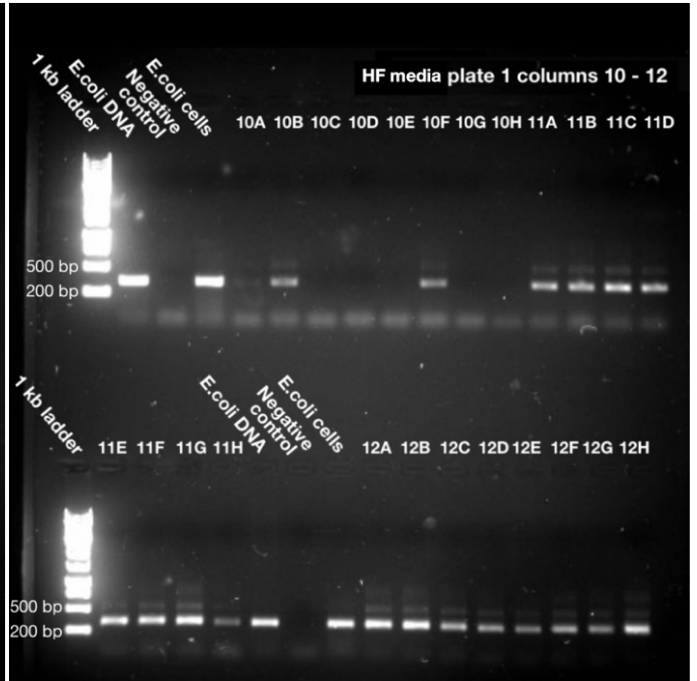
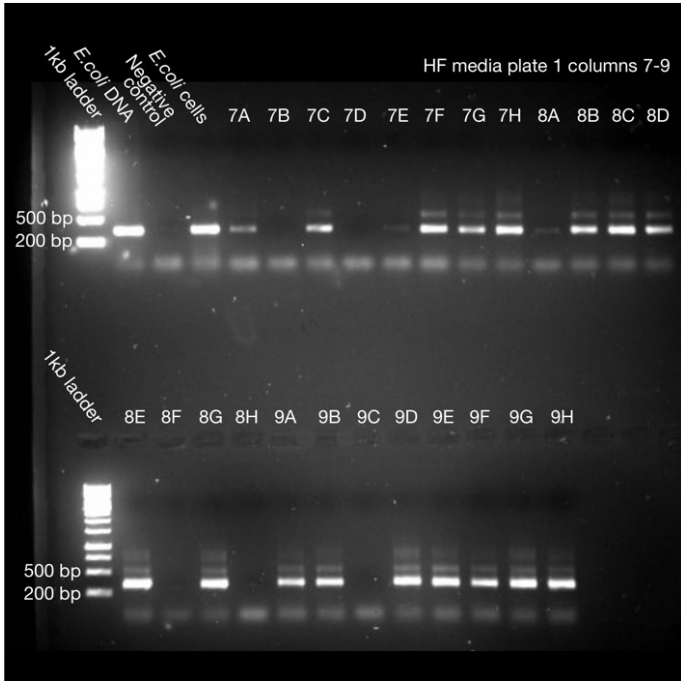
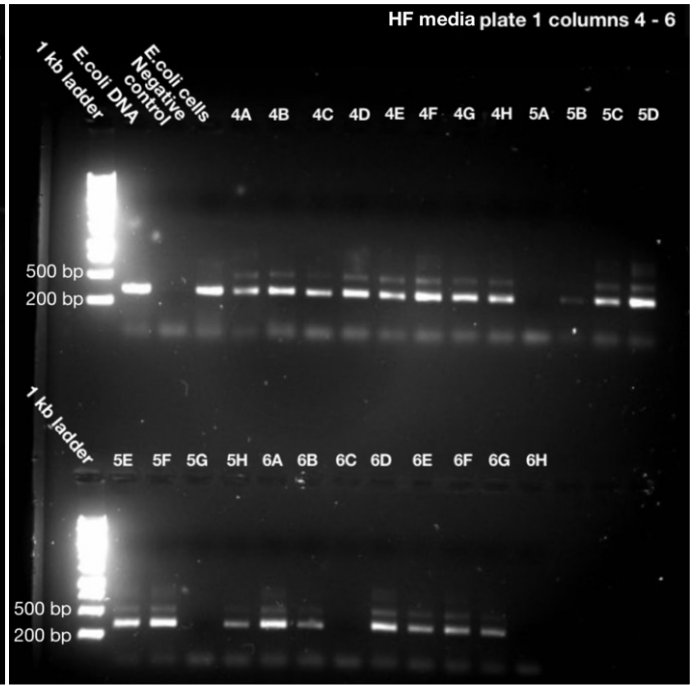
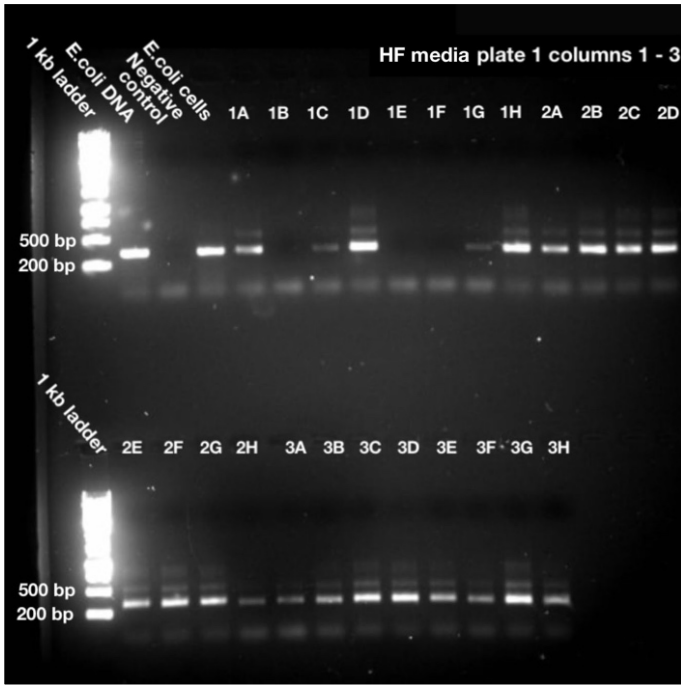


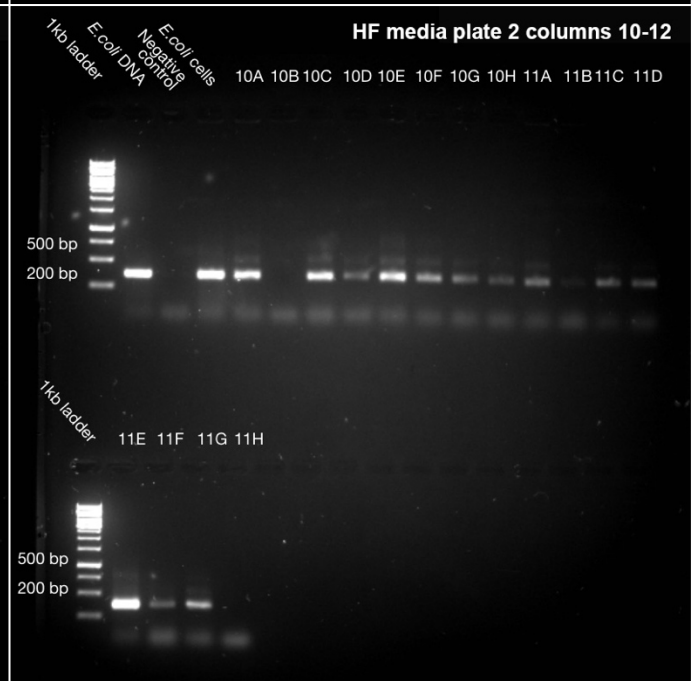
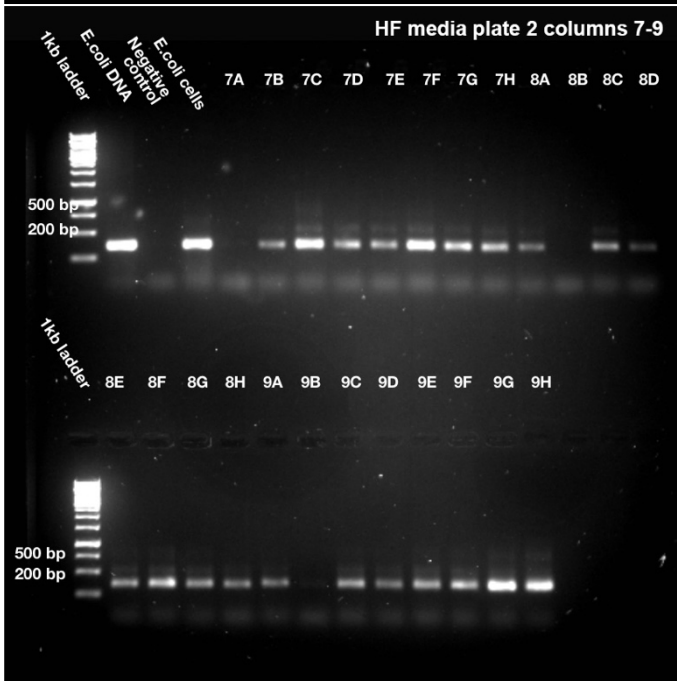
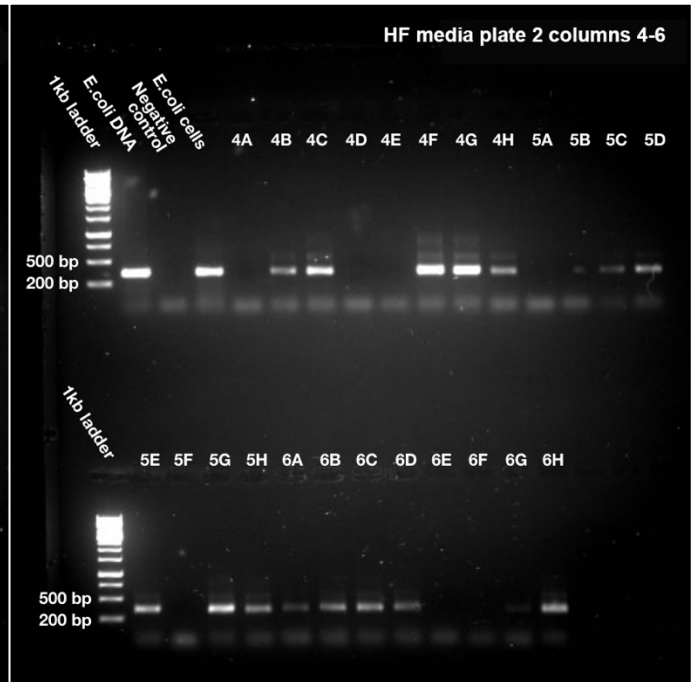
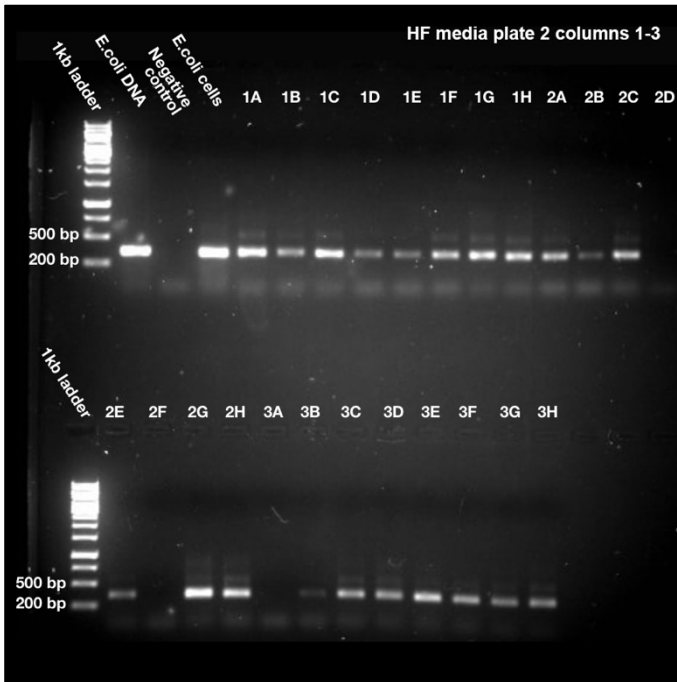
Appendix A - PCR Gel Electrophoresis Gel electrophoresis visualisation of all 1,152 PCR products. Images were visualised using a LAS-1000 Luminescent Image Analyser (Fujifilm).. The label in the top right of each image indicates the enrichment background and the columns that samples were taken from in 96 well plates. Pages 45-56

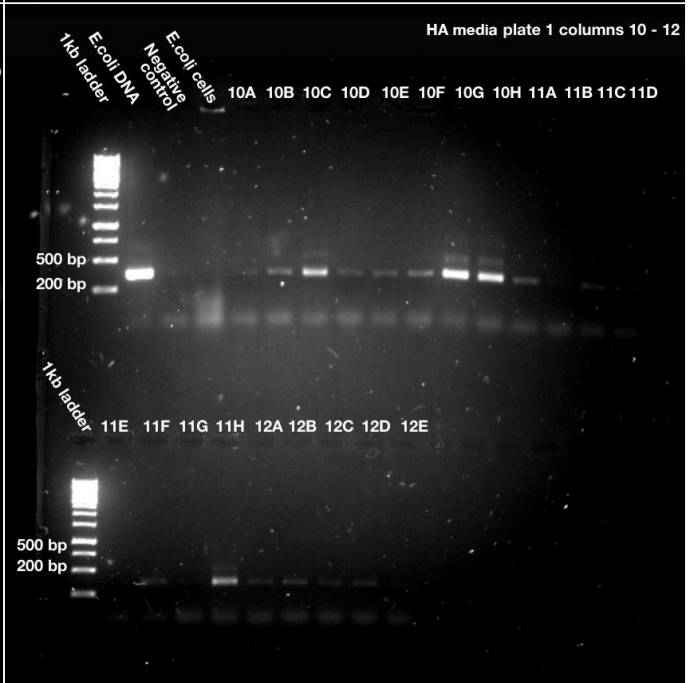
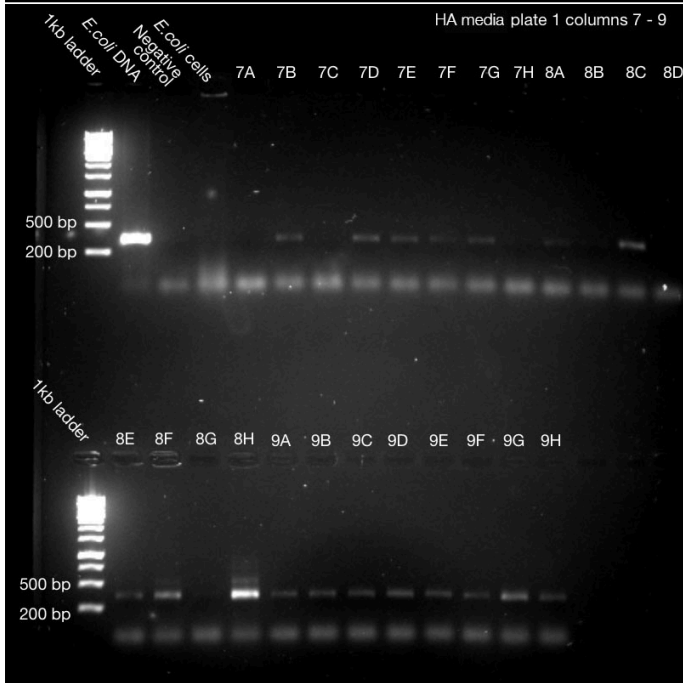
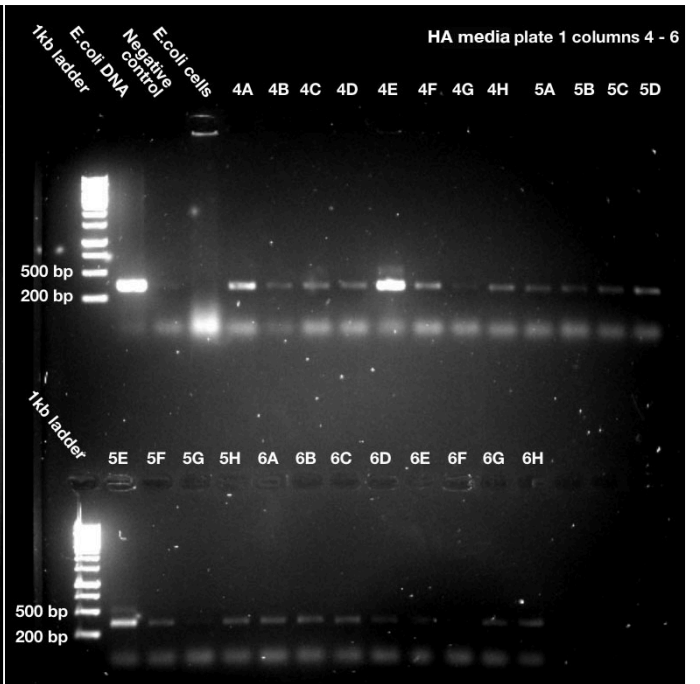
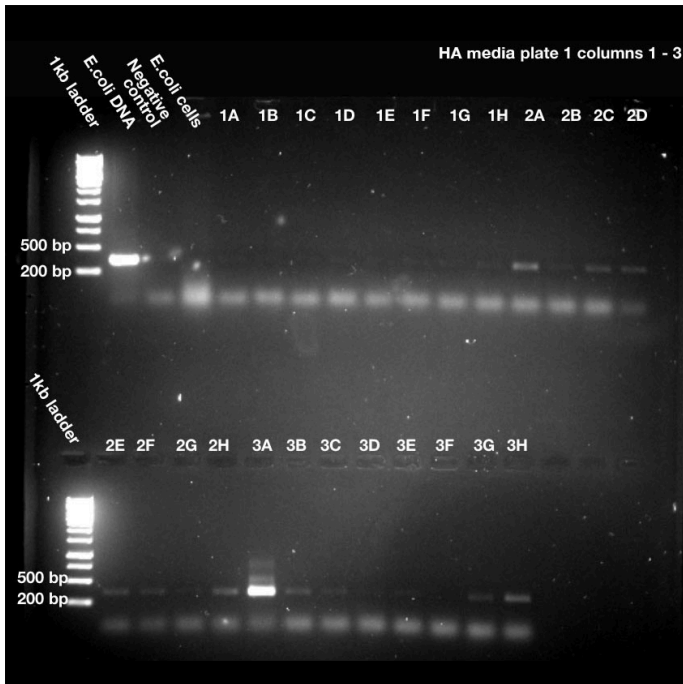


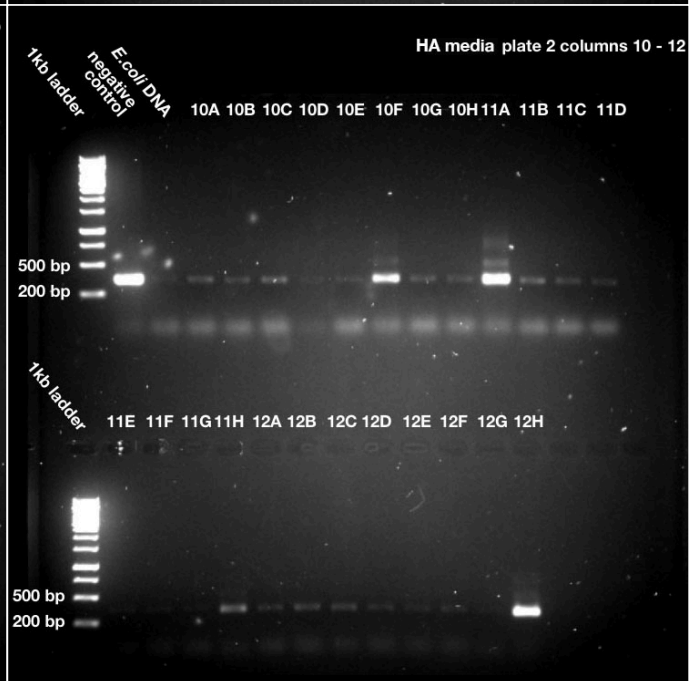
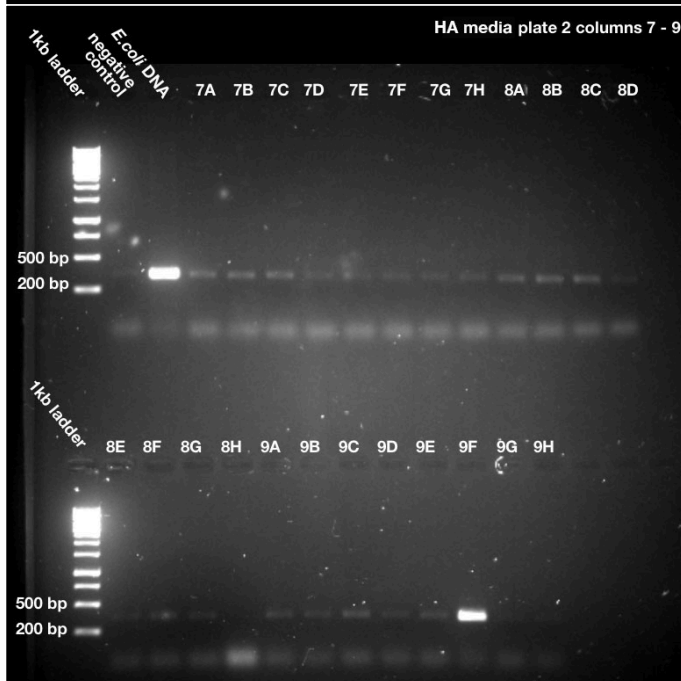
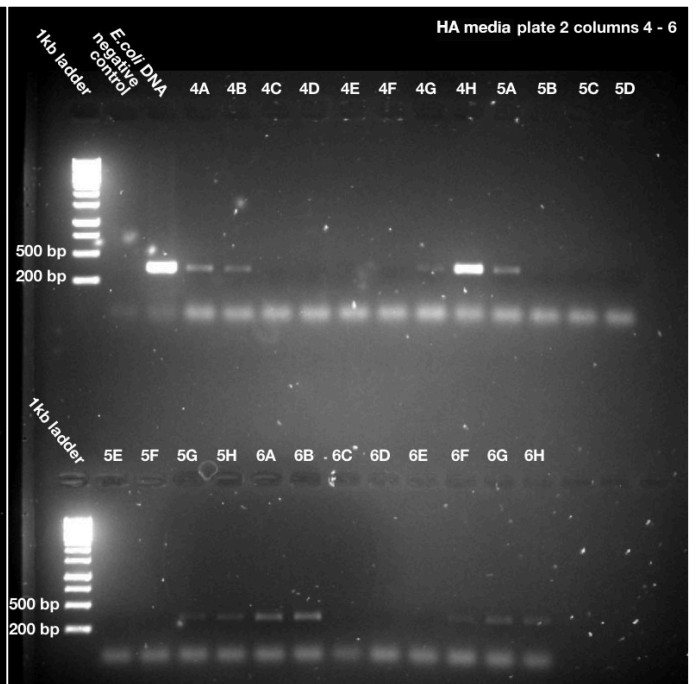
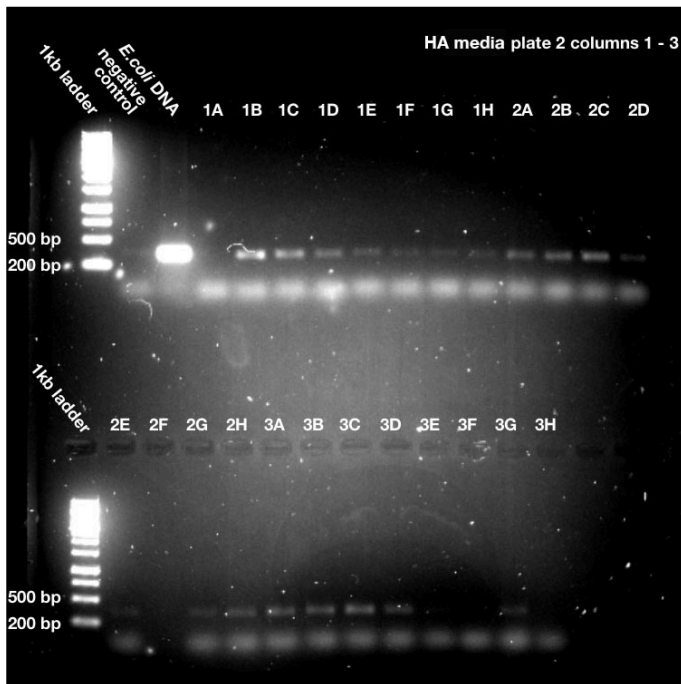


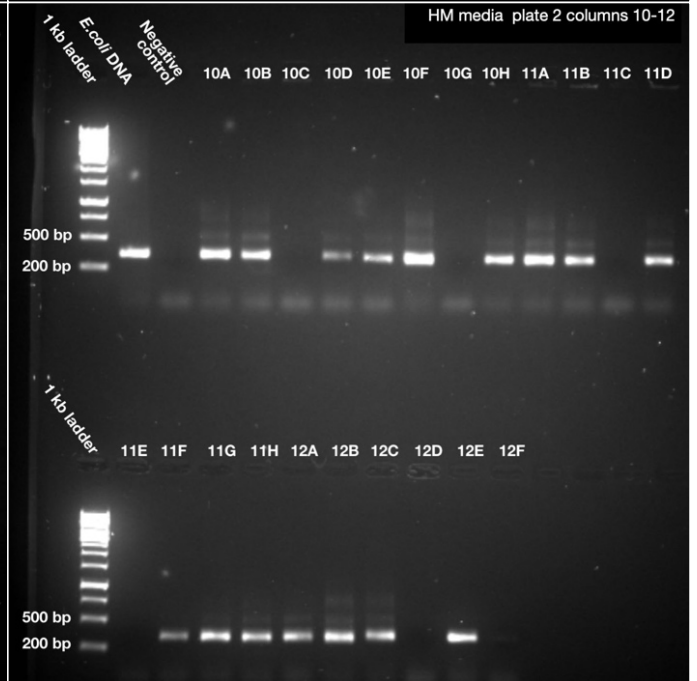
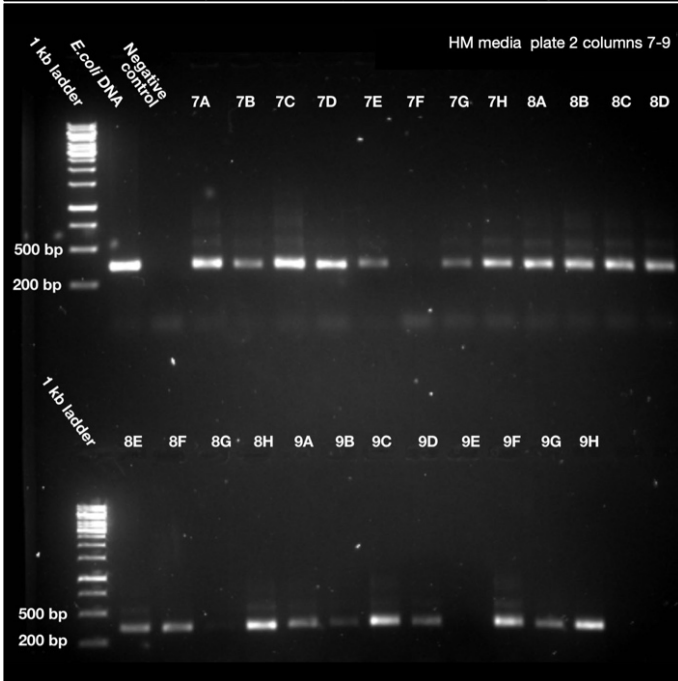
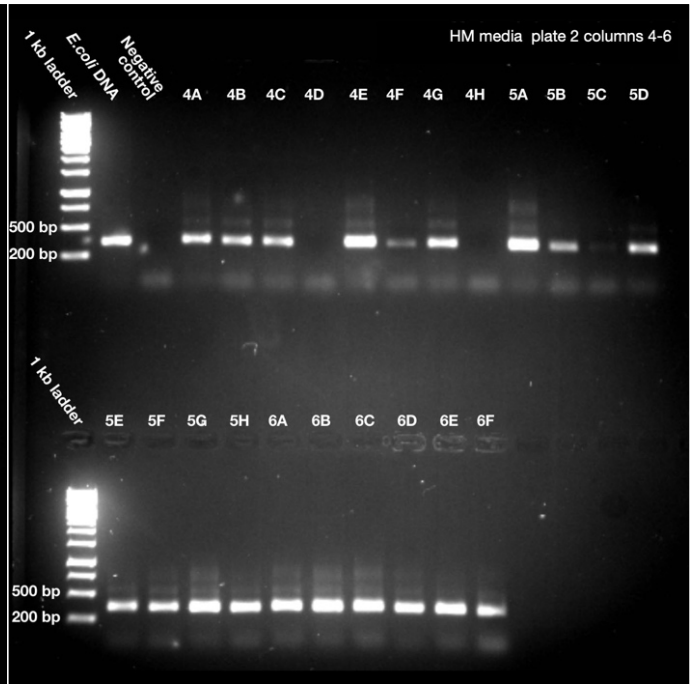
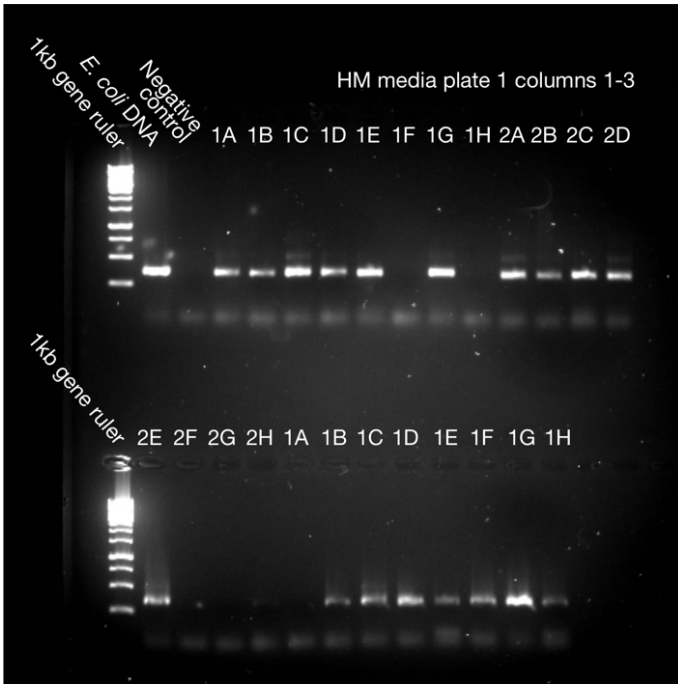


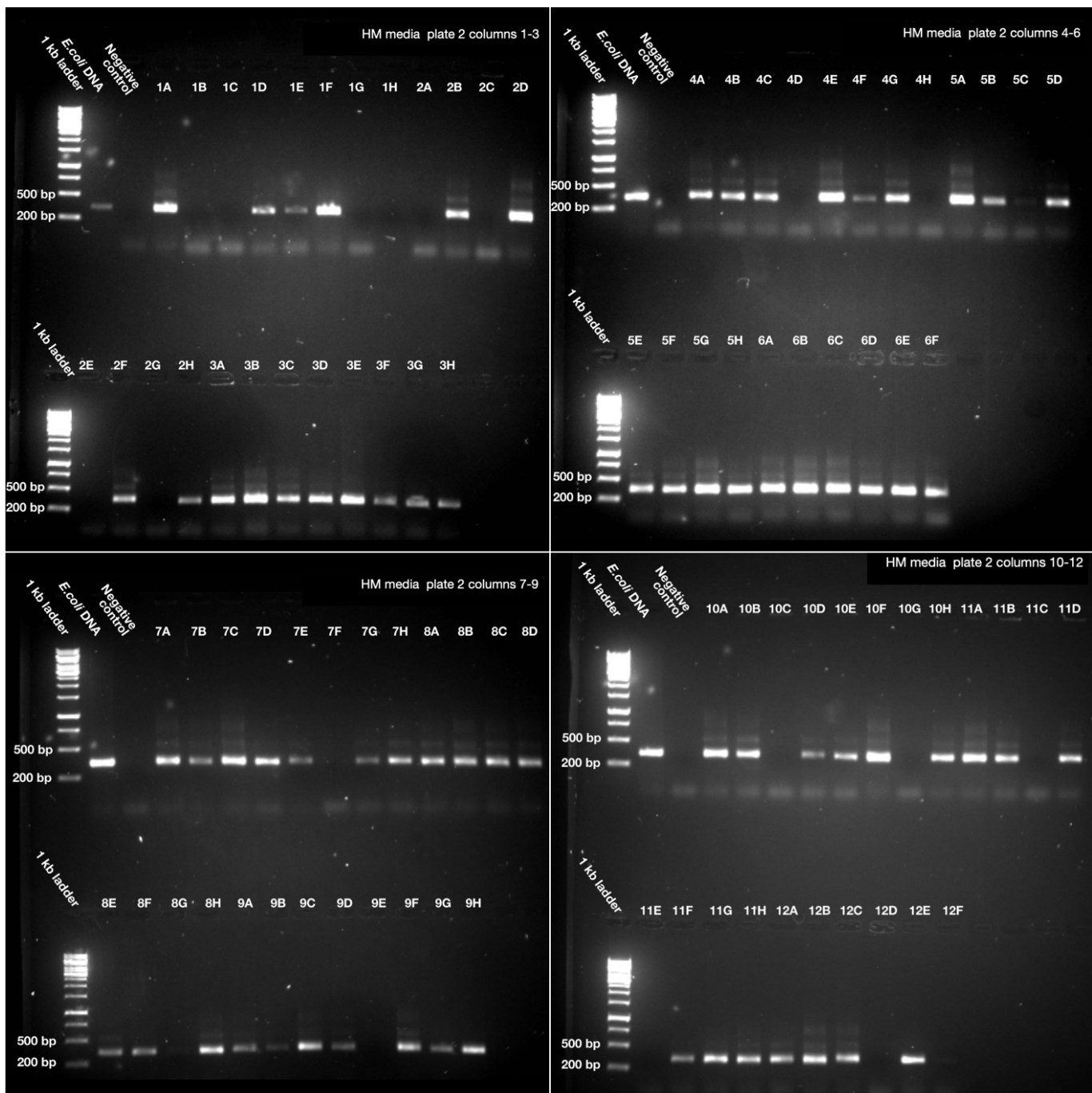


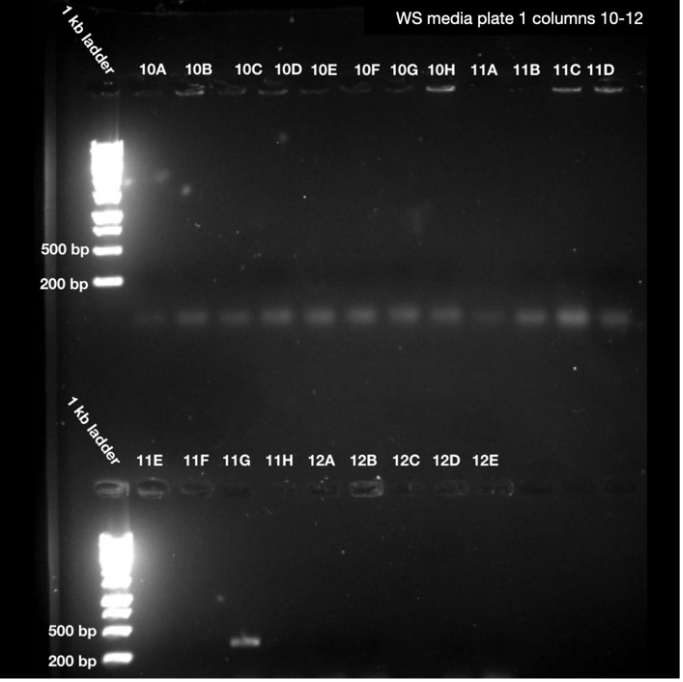
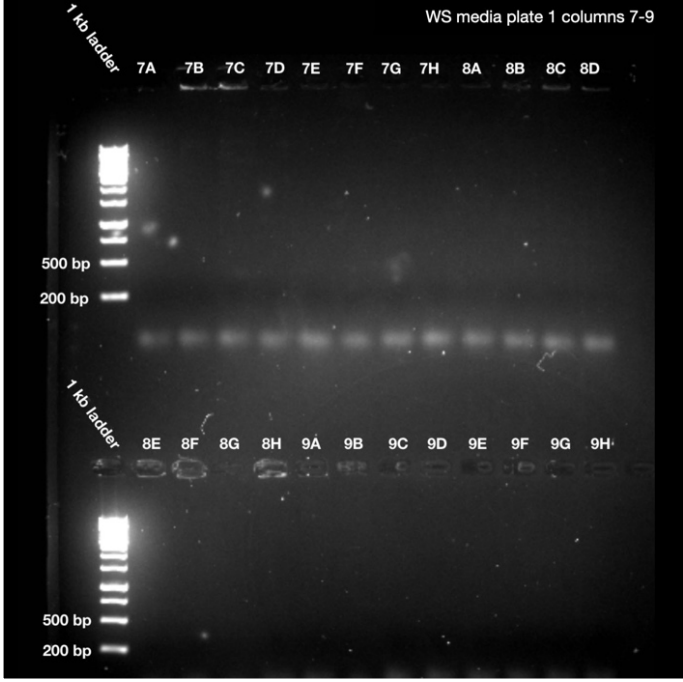
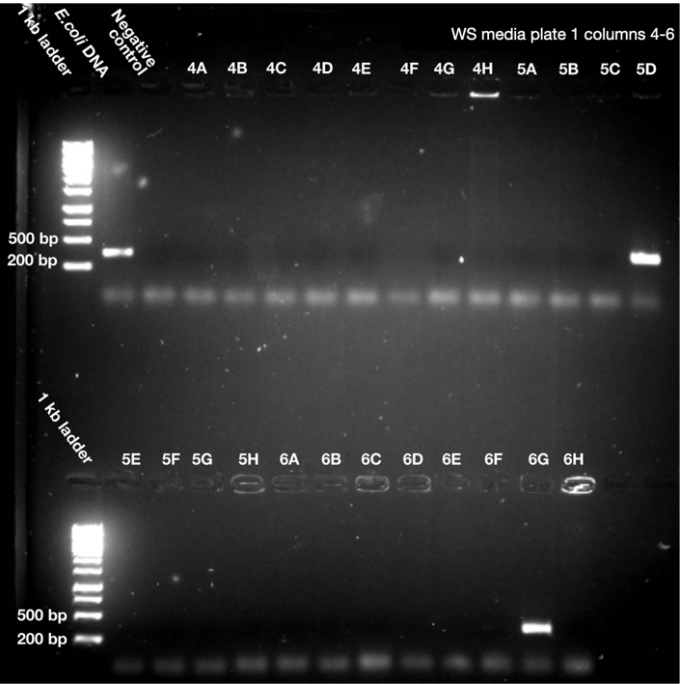
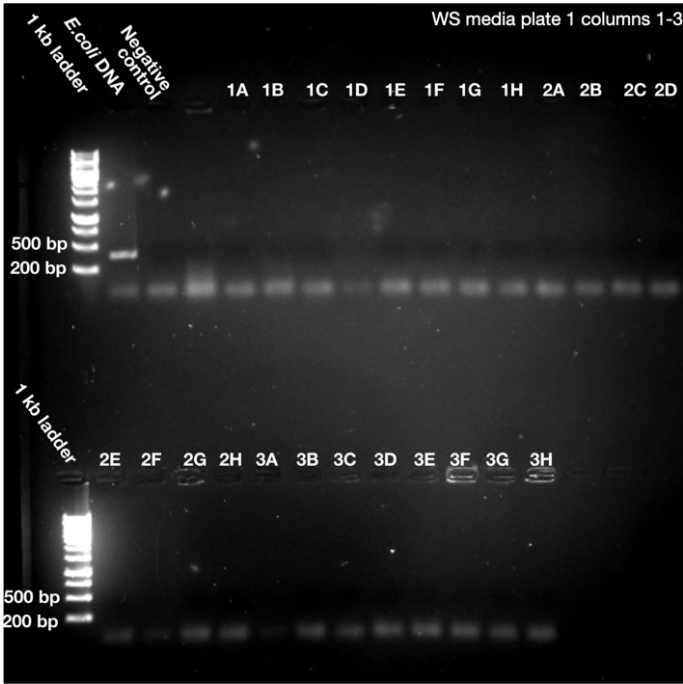


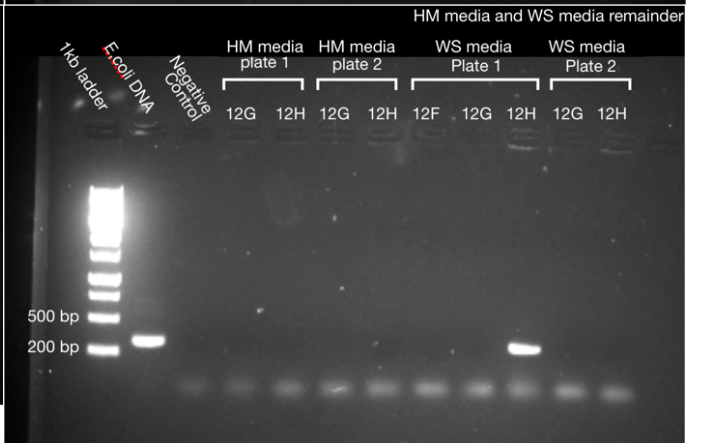
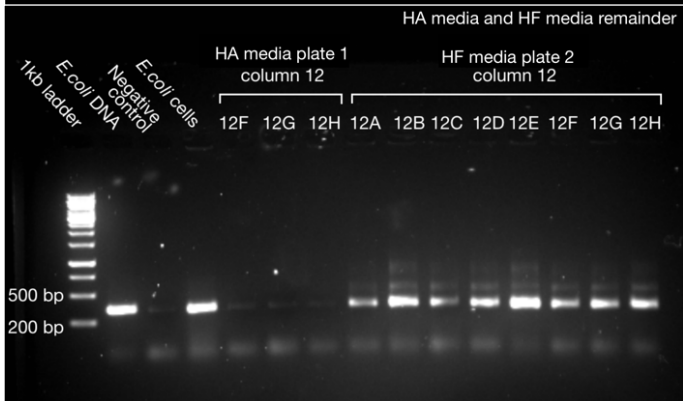
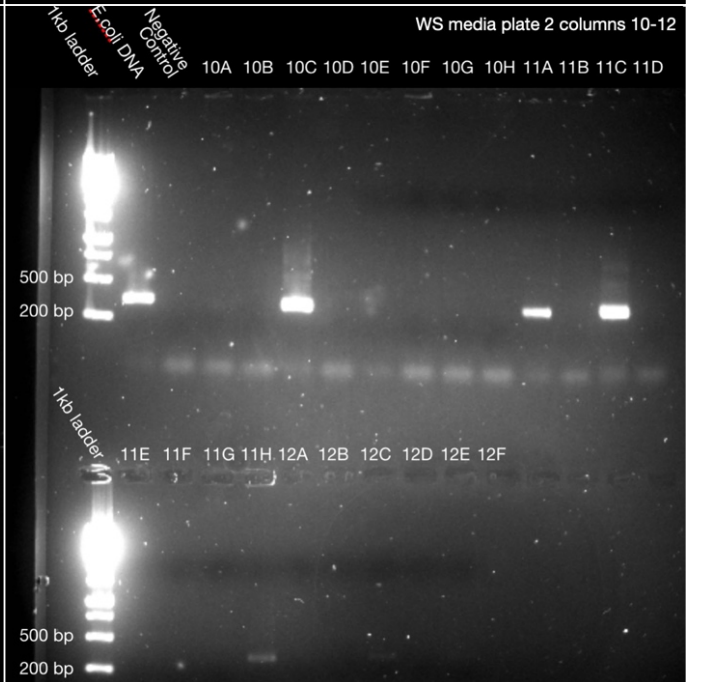
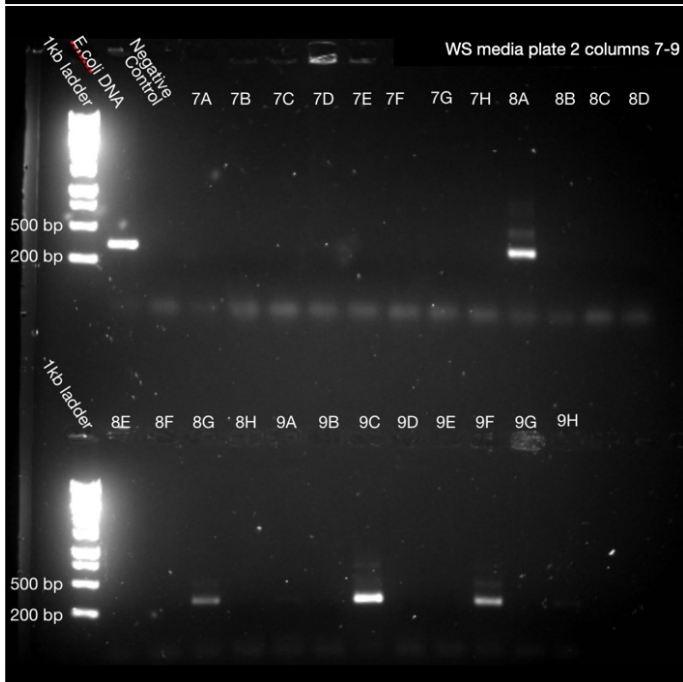
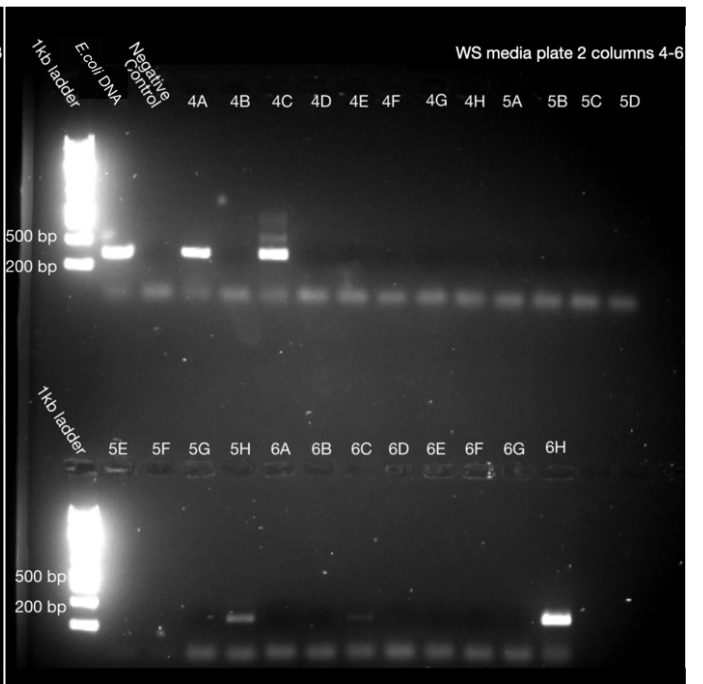
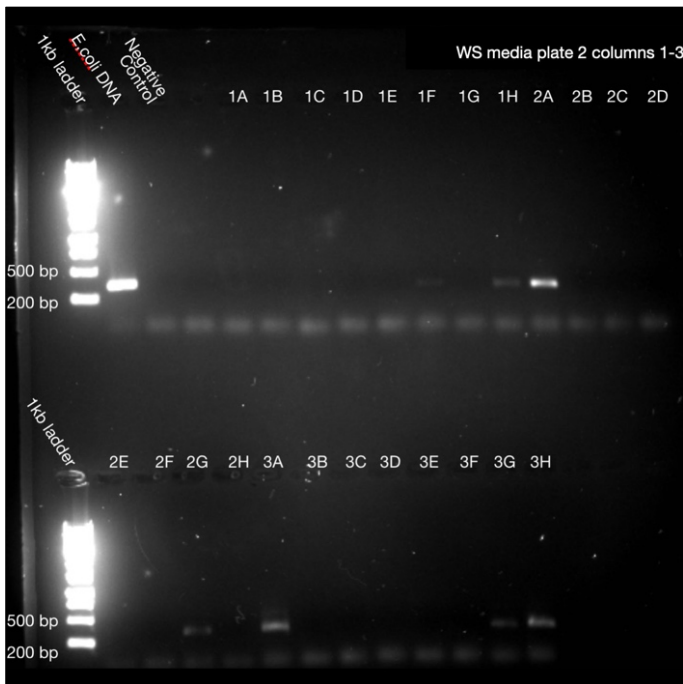


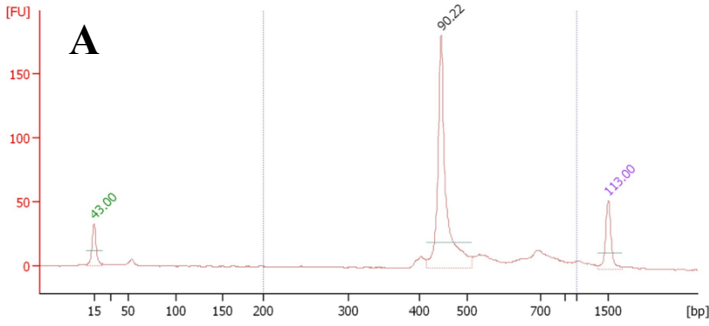












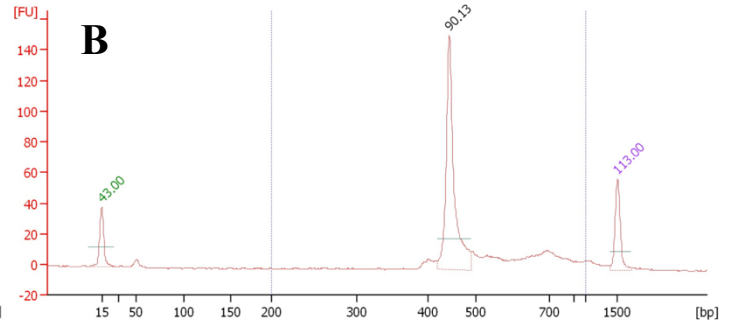
Overall Results for sample 1 : 1

Number of peaks found: 1 Area 1: 364.5

Peak	Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	445	12.90	43.9	
3	1,500	2.10	2.1	Upper Marker

Region table for sample 1 : 1

From [bp]	To [bp]	% of Total	Average Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Area	Co Size distribution in lor CV [%]
200	1,000	90	514	20.10	63.0	364.5	24.8



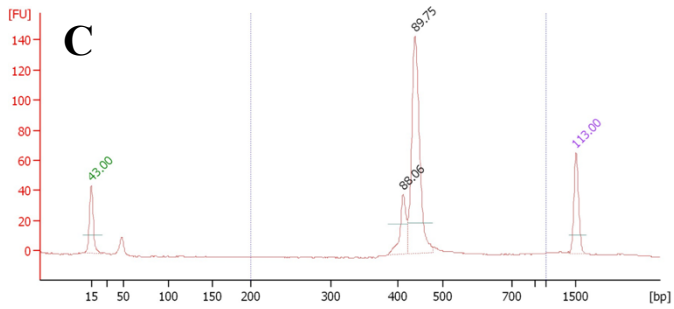
Overall Results for sample 2 : 2

Number of peaks found: 1 Area 1: 302.0

Peak	Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	443	9.84	33.6	
3	1,500	2.10	2.1	Upper Marker

Region table for sample 2 : 2

From [bp]	To [bp]	% of Total	Average Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Area	Co Size distribution in lor CV [%]
200	1,000	95	525	15.63	47.3	302.0	23.8



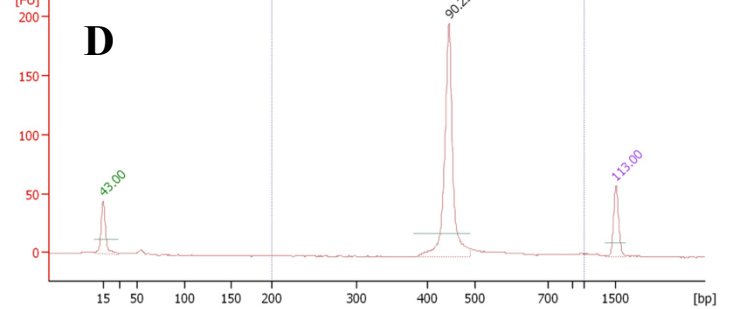
Overall Results for sample 3 : 3

Number of peaks found: 2 Area 1: 252.9

Peak	Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	411	2.11	7.8	
3	437	9.80	33.9	
4	1,500	2.10	2.1	Upper Marker

Region table for sample 3 : 3

From [bp]	To [bp]	% of Total	Average Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Area	Co Size distribution in lor CV [%]
200	1,000	91	451	13.17	44.9	252.9	15.3



Overall Results for sample 4 : 4

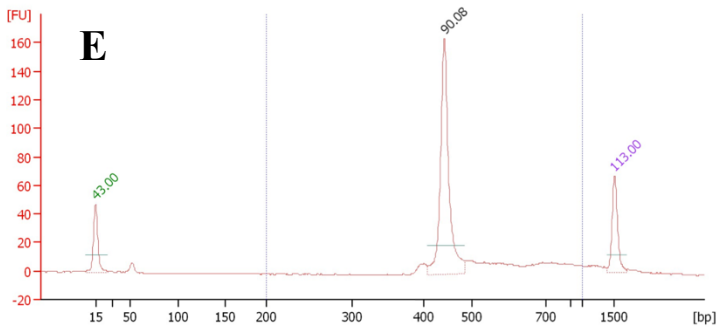
Number of peaks found: 1 Area 1: 269.5

Peak	Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	445	13.44	45.8	
3	1,500	2.10	2.1	Upper Marker

Region table for sample 4 : 4

From [bp]	To [bp]	% of Total	Average Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Area	Co Size distribution in lor CV [%]
200	1,000	97	466	14.60	48.3	269.5	16.2

Appendix B - Bioanalyser results A 2100 Bioanalyser System (Agilent) was used to ascertain the average product lengths and purity of each indexed PCR pool product. Pool selected were as follows (refer to section 2.2.4.1 for pooling schematic) (A) H media pool 4, (B) H media pool 7, (C) sludge pool 7, (D) HA media pool 6, (E) HF media pool 3, (F) HM media pool 1, (G) HM media pool 9 and (H) WS media pool 7. Pages 57-58



Overall Results for sample 5 : 5

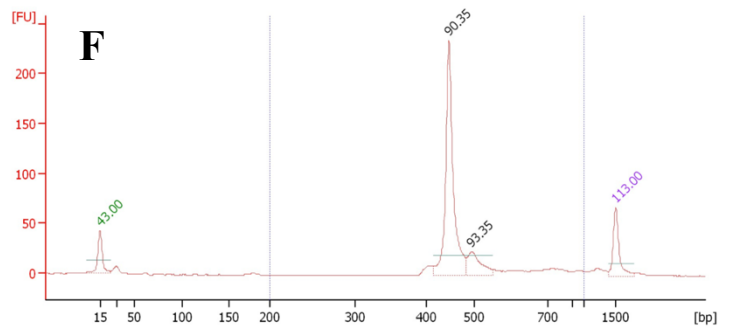
Number of peaks found: 1 Area 1: 323.0

Peak table for sample 5 : 5

Peak	Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	443	9.62	32.9	
3	1,500	2.10	2.1	Upper Marker

Region table for sample 5 : 5

From [bp]	To [bp]	% of Total	Average Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Area	Co Size distribution in lor CV [%]
200	1,000	90	515	14.56	45.0	323.0	24.6



Overall Results for sample 6 : 6

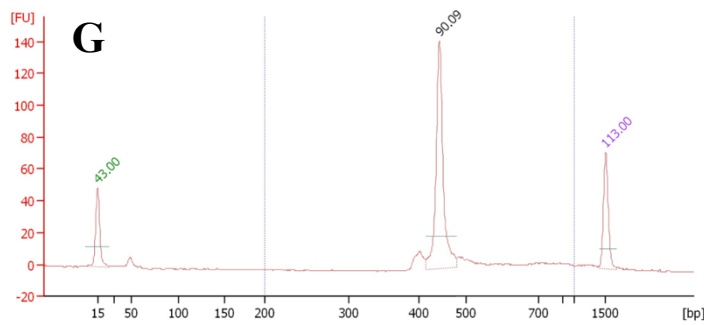
Number of peaks found: 2 Area 1: 373.6

Peak table for sample 6 : 6

Peak	Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	447	11.46	38.9	
3	493	2.01	6.2	
4	1,500	2.10	2.1	Upper Marker

Region table for sample 6 : 6

From [bp]	To [bp]	% of Total	Average Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Area	Co Size distribution in lor CV [%]
200	1,000	94	492	15.75	50.0	373.6	20.3



Overall Results for sample 7 : 7

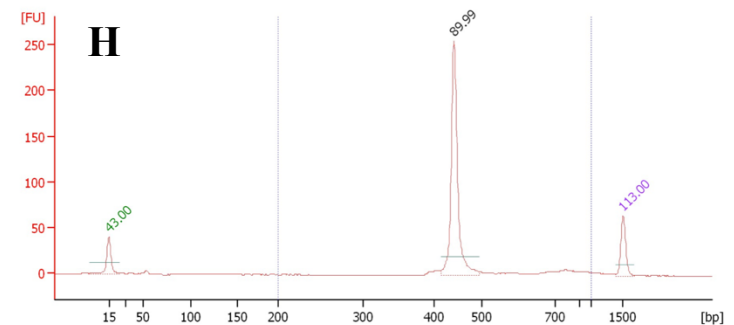
Number of peaks found: 1 Area 1: 235.3

Peak table for sample 7 : 7

Peak	Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	443	7.54	25.8	
3	1,500	2.10	2.1	Upper Marker

Region table for sample 7 : 7

From [bp]	To [bp]	% of Total	Average Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Area	Co Size distribution in lor CV [%]
200	1,000	92	497	11.31	35.9	235.3	23.4



Overall Results for sample 8 : 8

Number of peaks found: 1 Area 1: 303.2

Peak table for sample 8 : 8

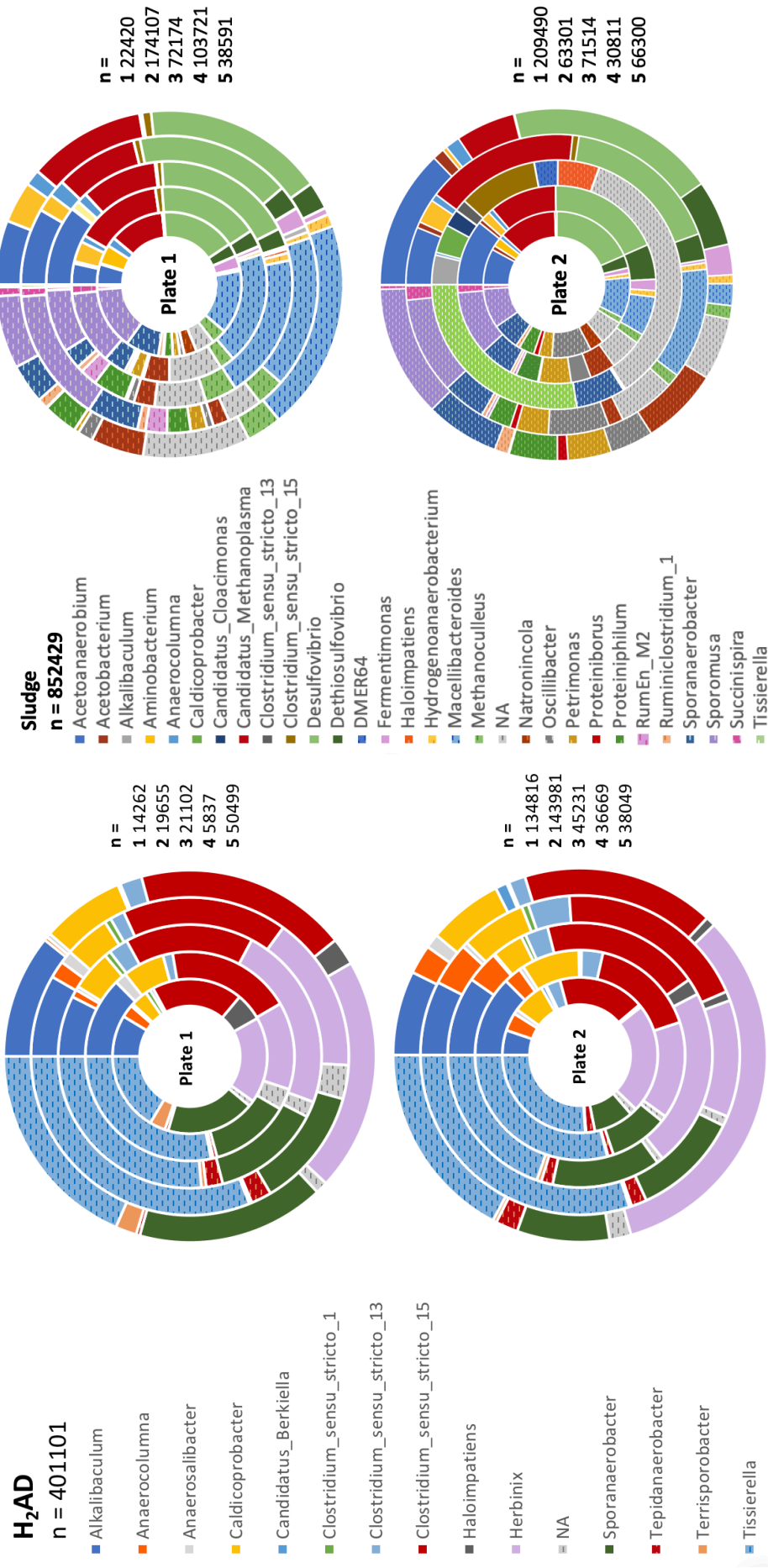
Peak	Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	441	13.42	46.1	
3	1,500	2.10	2.1	Upper Marker

Region table for sample 8 : 8

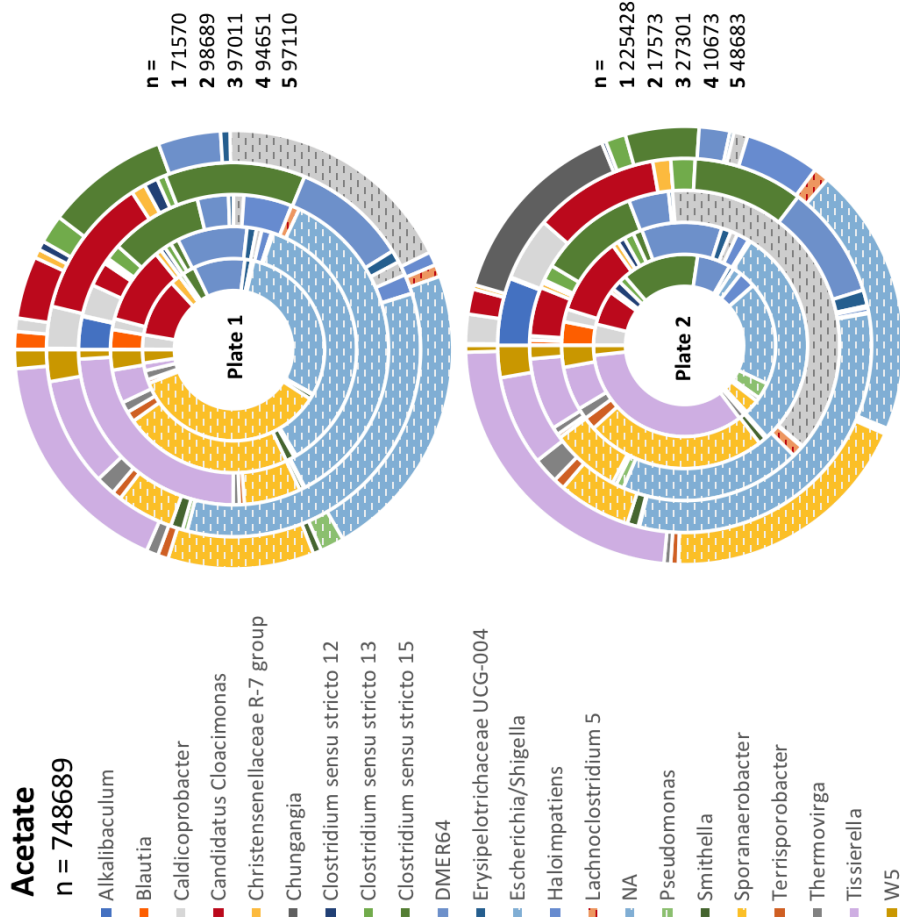
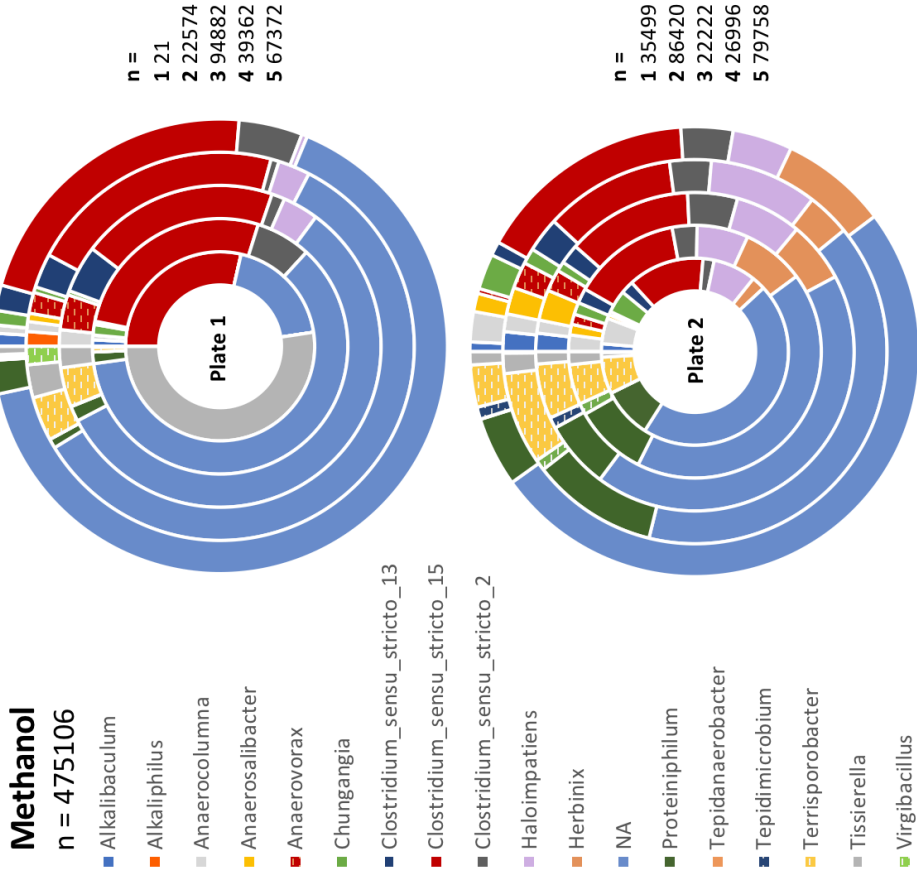
From [bp]	To [bp]	% of Total	Average Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Area	Co Size distribution in lor CV [%]
200	1,000	96	482	16.00	52.0	303.2	22.2

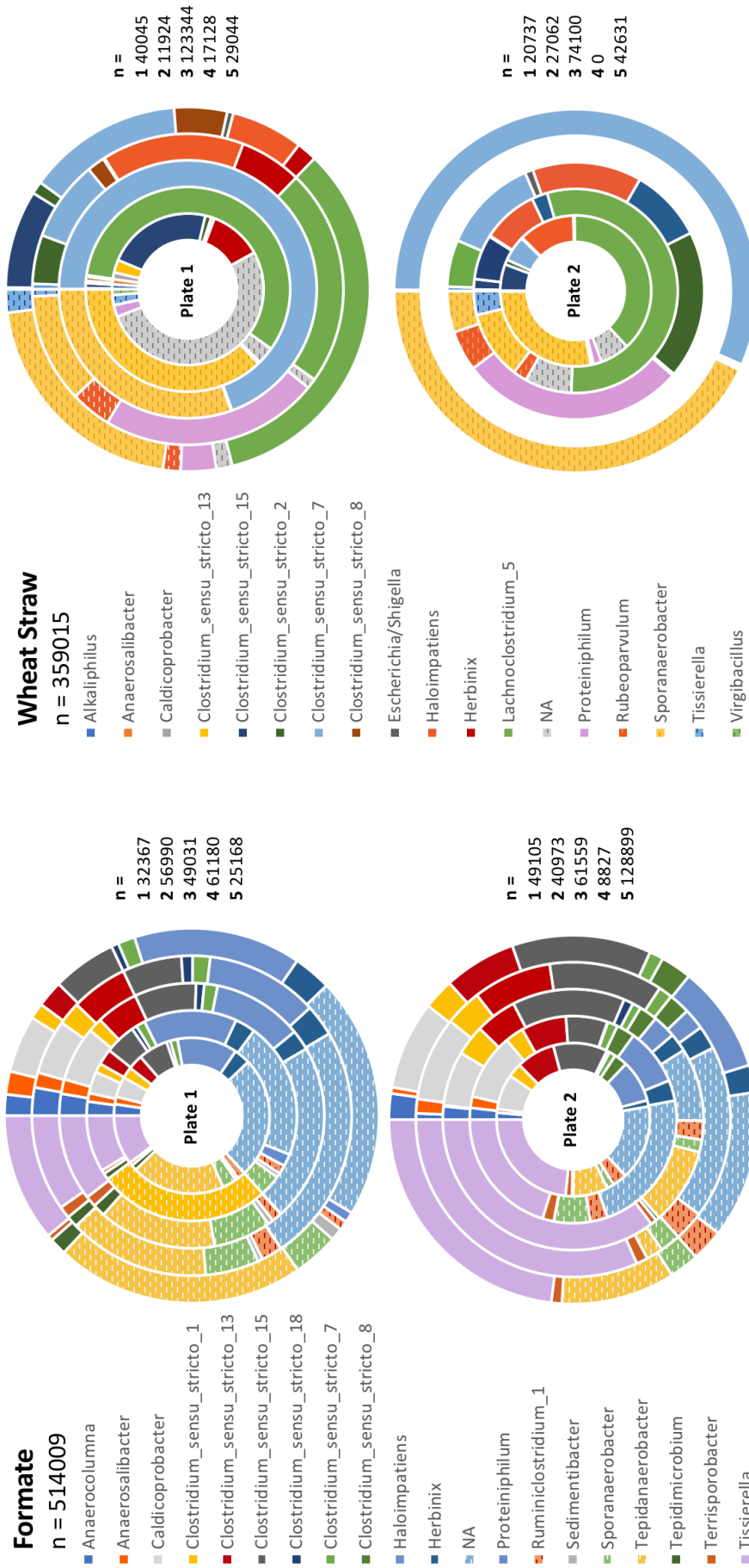


Appendix C - Sequencing data quality graphs ACE show data from forward reads and BDF show data from reverse reads. A and B show Q scores plotted against read position for sequencing results from pool 1 of the HA media grown samples. C and D show the same for samples from pool 4 of the HF media grown samples and E and F show the same for pool 9 of the Sludge grown samples. Samples were selected at random for demonstrative purposes.



Appendix D - Taxonomic distribution of assigned sequences (Genus level) DADA2 output at a genus level for all 60 pools. For each plate the most central concentric ring is pool 1 and each ring moving outwards is one pool higher so the outer pool is number 5. Reads that accounted for genera that made up >1% of the total number of reads for each growth background were binned as NA.
Pages 60-62





Download		Graphics		Sort by: E value	
Sequence ID: Query_221435 Length: 1606 Number of Matches: 2					
Range 1: 1345 to 1352 Graphics				▼ Next Match ▲ Previous Match	
Score	Expect	Identities	Gaps	Strand	
16.4 bits(8)	0.23	8/8(100%)	0/8(0%)	Plus/Plus	
Query 6	GAAAAGCT	13			
Sbjct 1345	GAAAAGCT	1352			
Range 2: 1388 to 1394 Graphics				▼ Next Match ▲ Previous Match ▲ First Match	
Score	Expect	Identities	Gaps	Strand	
14.4 bits(7)	0.91	7/7(100%)	0/7(0%)	Plus/Minus	
Query 11	GCTTCAT	17			
Sbjct 1394	GCTTCAT	1388			

Appendix E - Alternative probe alignment Alignment of the probe S*-WWE1-408-a-A-19 (Limam, *et al.*, 2014) (Query) with the *Candidatus Cloacimonas* 16S rRNA gene (Sbjct) performed using the BLAST tool.

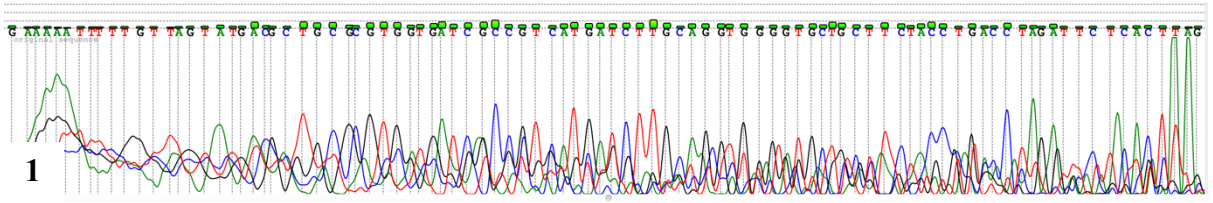
A	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	Bacterium NLAE-zl-P604 16S ribosomal RNA gene, partial sequence	1068	1068	100%	0.0	100%	JQ607196.1
<input type="checkbox"/>	Tissierella praeacuta strain LBR1 16S ribosomal RNA gene, partial sequence	1064	1064	99%	0.0	100%	GQ461808.1
<input type="checkbox"/>	Tissierella praeacuta strain VPI 10426 16S ribosomal RNA gene, partial sequence	1064	1064	99%	0.0	100%	GQ461813.1
<input type="checkbox"/>	Tissierella praeacuta strain AIP 443.98 16S ribosomal RNA gene, partial sequence	1064	1064	99%	0.0	100%	GQ461805.1
<input type="checkbox"/>	Tissierella praeacuta strain ATCC 25539 16S ribosomal RNA gene, partial sequence	1064	1064	99%	0.0	100%	NR_117376.1
<input type="checkbox"/>	Pseudomonas sp. 3e-1 16S ribosomal RNA gene, partial sequence	1064	1064	99%	0.0	100%	DQ886425.1
<input type="checkbox"/>	Tissierella praeacuta strain ATCC 25539 16S ribosomal RNA, partial sequence	1064	1064	99%	0.0	100%	NR_119111.1
<input type="checkbox"/>	Clostridium hastiforme rm gene for 16S ribosomal RNA, strain DSM 5675	1064	1064	99%	0.0	100%	X80841.1
<input type="checkbox"/>	Bacterium NLAE-zl-C280 16S ribosomal RNA gene, partial sequence	1062	1062	100%	0.0	99%	JQ607997.1

B	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	Sporanaerobacter acetigenes strain 906C 16S ribosomal RNA gene, partial sequence	976	976	100%	0.0	99%	KU950268.1
<input type="checkbox"/>	Sporanaerobacter acetigenes strain DSM 13106 16S ribosomal RNA gene, partial sequence	976	976	100%	0.0	99%	NR_117381.1
<input type="checkbox"/>	Sporanaerobacter sp. SN28 16S ribosomal RNA gene, partial sequence	976	976	100%	0.0	99%	EU887812.1
<input type="checkbox"/>	Sporanaerobacter acetigenes strain Lup33 16S ribosomal RNA, partial sequence	976	976	100%	0.0	99%	NR_025151.1
<input type="checkbox"/>	Bacterium NLAE-zl-P544 16S ribosomal RNA gene, partial sequence	972	972	100%	0.0	99%	JQ607324.1
<input type="checkbox"/>	Bacterium NLAE-zl-C426 16S ribosomal RNA gene, partial sequence	955	955	100%	0.0	99%	JQ608130.1
<input type="checkbox"/>	Bacterium NLAE-zl-C583 16S ribosomal RNA gene, partial sequence	952	952	100%	0.0	99%	JQ608283.1
<input type="checkbox"/>	Bacterium NLAE-zl-C569 16S ribosomal RNA gene, partial sequence	946	946	100%	0.0	99%	JQ608269.1
<input type="checkbox"/>	Acetanaerobacter sp. Iso-W4 16S ribosomal RNA gene, partial sequence	942	942	100%	0.0	99%	DQ677021.1

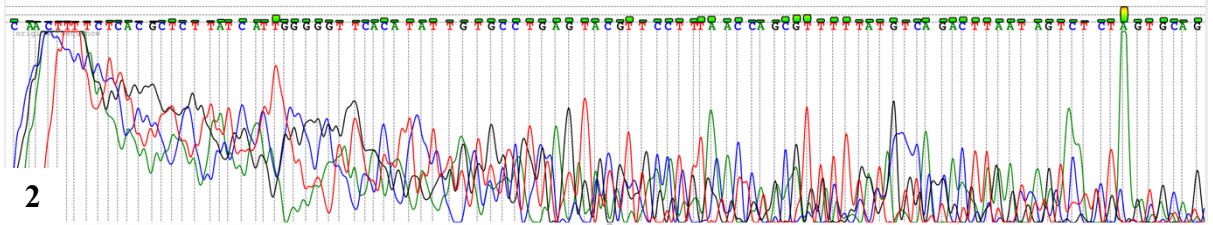
C	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	Uncultured bacterium clone G35_D8_H_B_G07 16S ribosomal RNA gene, partial sequence	1229	1229	100%	0.0	99%	EF559175.1
<input type="checkbox"/>	Uncultured compost bacterium partial 16S rRNA gene, clone BO2441	1118	1118	100%	0.0	97%	FN667328.1
<input type="checkbox"/>	Uncultured compost bacterium partial 16S rRNA gene, clone FS2202	1107	1107	100%	0.0	96%	FN667230.1
<input type="checkbox"/>	Uncultured Coprococcus sp. clone S3_E29 16S ribosomal RNA gene, partial sequence	1085	1085	100%	0.0	96%	KX672365.1
<input type="checkbox"/>	Uncultured Firmicutes bacterium gene for 16S ribosomal RNA, partial sequence, clone: B103	1081	1081	97%	0.0	97%	AB874582.1
<input type="checkbox"/>	Uncultured Coprococcus sp. clone S3_E41 16S ribosomal RNA gene, partial sequence	1064	1064	100%	0.0	95%	KX672340.1
<input type="checkbox"/>	Uncultured Coprococcus sp. clone S3_E23 16S ribosomal RNA gene, partial sequence	1059	1059	100%	0.0	95%	KX672337.1

D	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	Sporanaerobacter acetigenes strain DSM 13106 16S ribosomal RNA gene, partial sequence	1325	1325	99%	0.0	99%	NR_117381.1
<input type="checkbox"/>	Sporanaerobacter sp. SN28 16S ribosomal RNA gene, partial sequence	1325	1325	99%	0.0	99%	EU887812.1
<input type="checkbox"/>	Sporanaerobacter acetigenes strain Lup33 16S ribosomal RNA, partial sequence	1325	1325	99%	0.0	99%	NR_025151.1
<input type="checkbox"/>	Sporanaerobacter acetigenes strain 906C 16S ribosomal RNA gene, partial sequence	1321	1321	100%	0.0	99%	KU950268.1
<input type="checkbox"/>	Bacterium NLAE-zl-P544 16S ribosomal RNA gene, partial sequence	1314	1314	99%	0.0	99%	JQ607324.1
<input type="checkbox"/>	Bacterium NLAE-zl-C583 16S ribosomal RNA gene, partial sequence	1304	1304	99%	0.0	99%	JQ608283.1
<input type="checkbox"/>	Bacterium NLAE-zl-C569 16S ribosomal RNA gene, partial sequence	1293	1293	99%	0.0	99%	JQ608269.1
<input type="checkbox"/>	Bacterium NLAE-zl-C426 16S ribosomal RNA gene, partial sequence	1258	1258	100%	0.0	98%	JQ608130.1
<input type="checkbox"/>	Acetanaerobacter sp. Iso-W4 16S ribosomal RNA gene, partial sequence	1254	1254	100%	0.0	98%	DQ677021.1

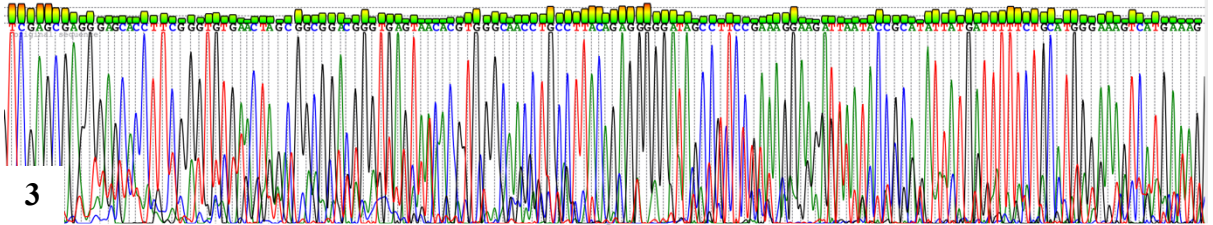
Appendix F - pure culture BLAST results Results from the BLAST tool (Sayers, 2009) on the NCBI website. Settings were set to include uncultured environmental samples but not models. The database used for comparison was the ‘nucleotide collection’ as defined by the NCBI website and the search was completed using the Megablast algorithm (Y. Chen, 2015). All sequences are derived from PCR products from plate 1 of HA media grown samples. (A) isolate 5D, (B) isolate 7A, (C) isolate 7D and (D) isolate 8A.



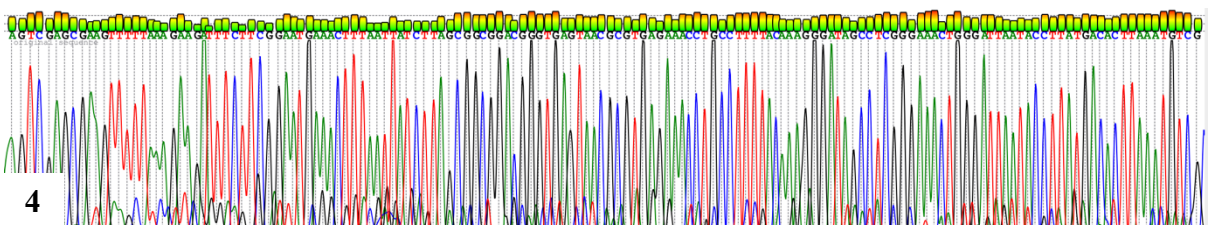
G A A A A A T T T T T G T T A G T A T G A C G C T G C G C G T G G T G A T C G C C G T C A T G A T C T T G C A G G T G G G G T G C T G C T T C T A C C T G A C C T A G A T T C T C A C T T A



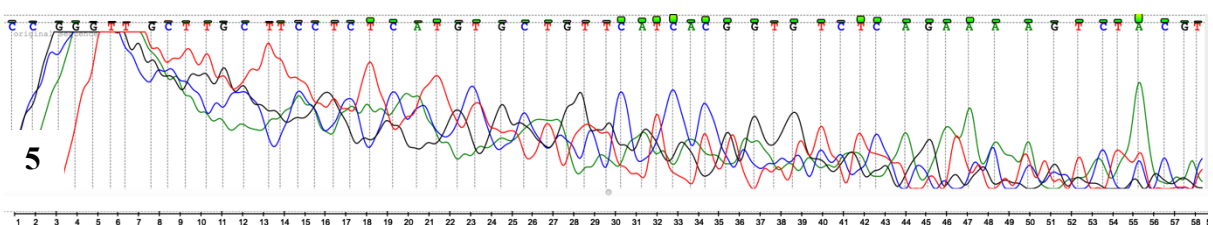
C A A C T T T C T C A C G C T C T A T C A T T G G G G G T T C A C A T A T T G T G C C T G A G T A C G T T C C T T T A A C C A G C G T T T T T A T G T C A G A C T T A A T A G T C T C T A G T G C C



T C G A C C A G G G A C C T T C G G G T G A C T A G C G G C G A G G G T G A G T A A C C T G G G A C C T G C C T A C A G A G G G G A T A G C C T T C C G A A A G G A A G A T T A A T A C C C A T A T A T G A T T T T C T G C A T G G G A A G C A T G A A G

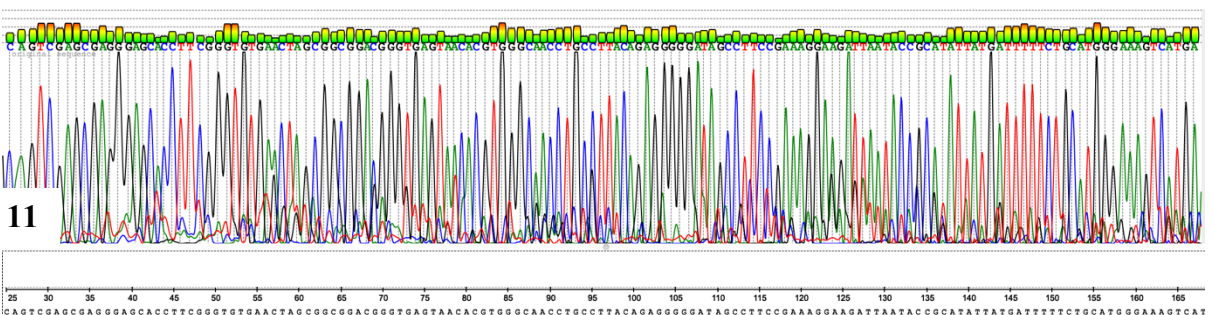
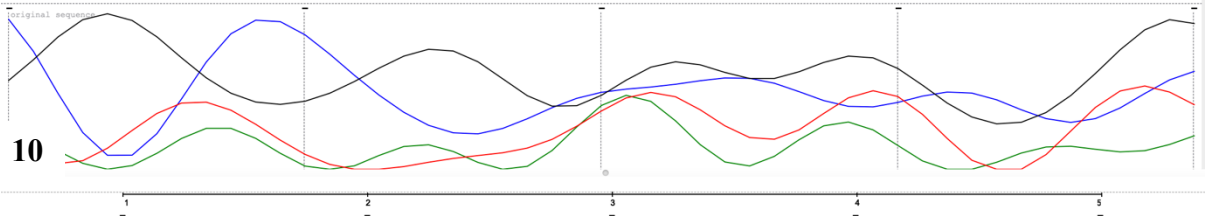
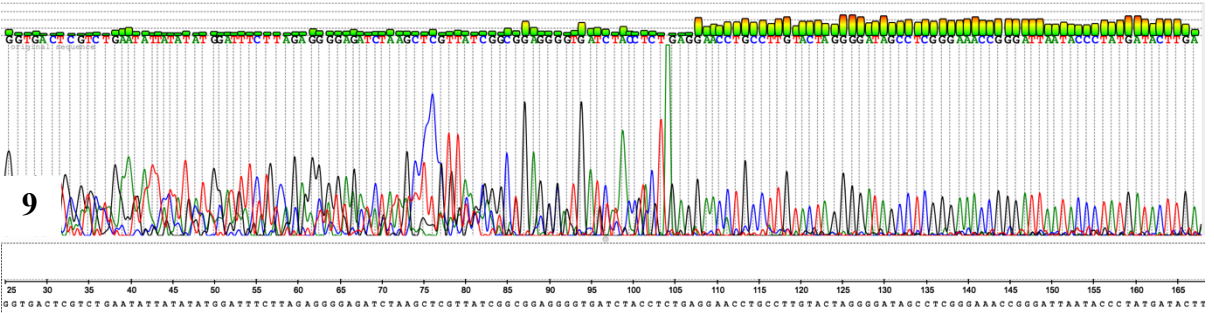
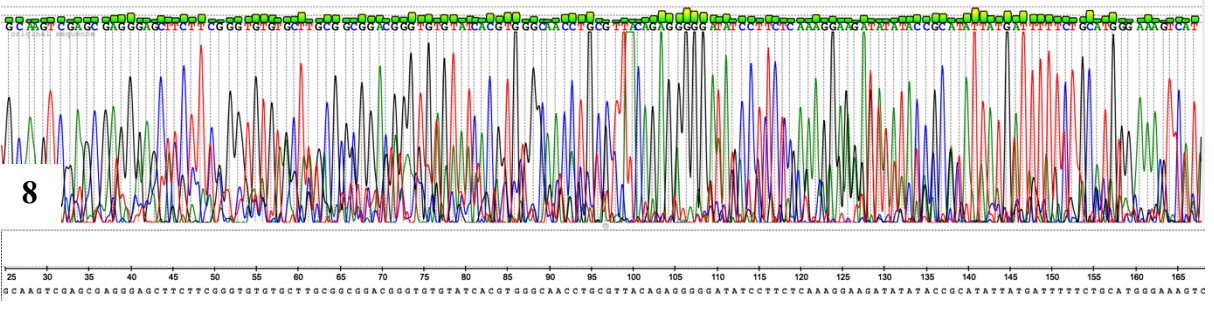
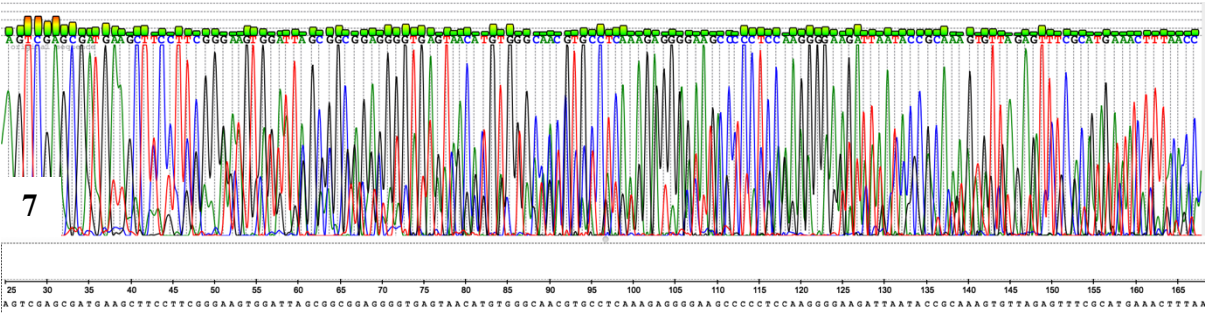
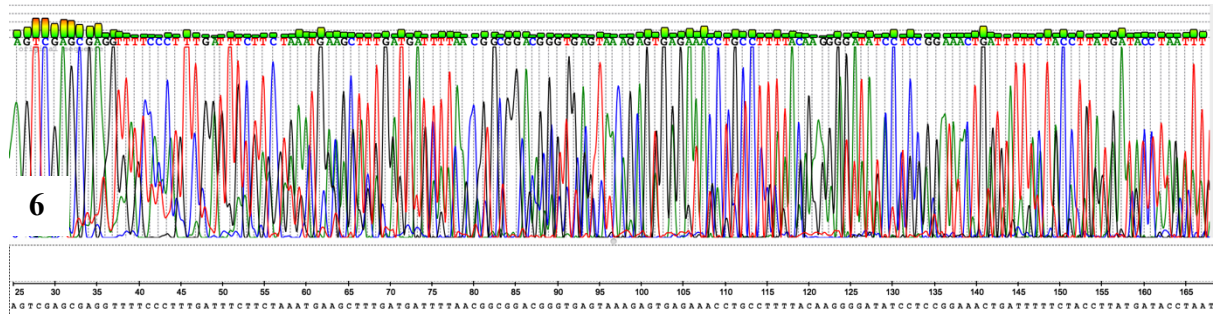


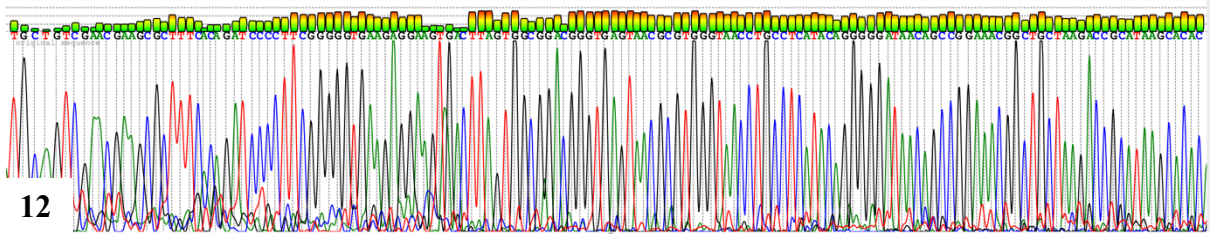
A G T C G A C C G A A G T T T T A A A G A G A T T T C C G G A T G A A A C T T A A T A T A T C T T A G C G G C G A C G G G T G A T A C C C T G A G A A A C C T C C C T T T A C A A A G G A T A G C C T C G G G A A C T G G G A T T A A C C T T A T A G A C A C T T A A A T G T C G



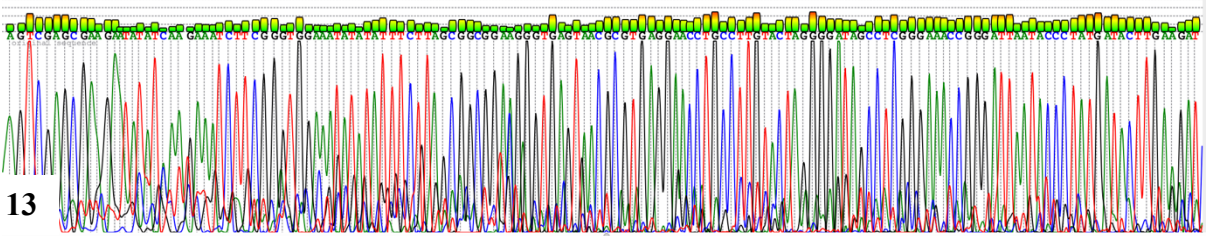
C C G G G T T G C T T G C T T C C T C T C A T G T G C T G T T C A T C A C G G T G T C T C A G A A A A G T C T A C G

Appendix G - Sanger Sequence Chromatographs Graphs were visualised using the FastQC program. Displayed are nucleotide positions 25-170. The first 25 bp are omitted as they represented the forward primer used. Pages 65 - 67

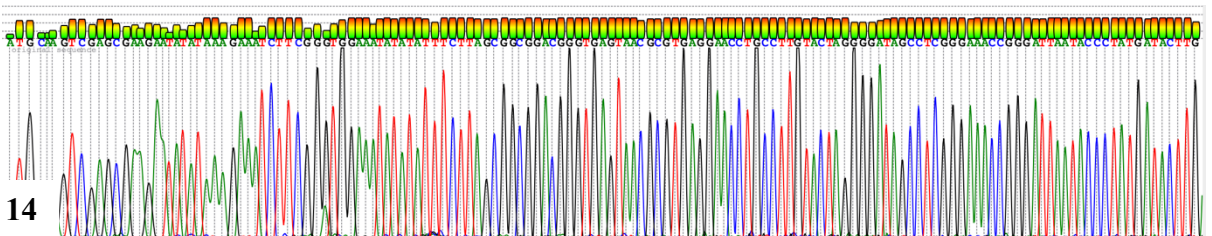




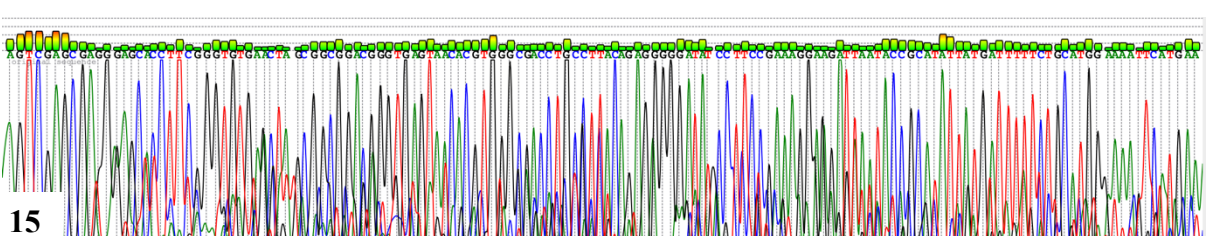
25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165
 T G C T G T C G A C G A A C C G C T T T C A C A G A T C C C C T T C G G G G T G A A G A G G A A G T G A C T T A G T G C G G G A C G G G T G A G T A A C C G C T G G G T A A C C T C C T C A T A C A G G G G A T A A C A G C C G G A A C C G G T G C T A A G A C C G C A T A A G C A C



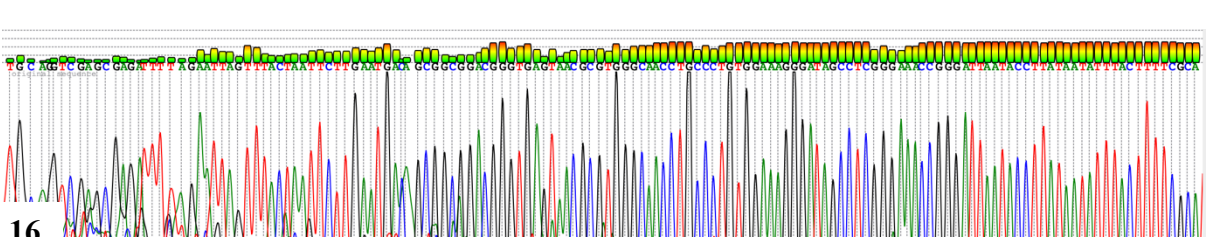
25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165
 A G C G A C C G A G G A A T A T A T A T A A G A A A T C T T C G G G T G A A A T A T A T A T T T C T T A G C G C G G A C G G G T G A G T A A C C G C T G A G G A A C C T G C C T T G T A C T A G G G A T A G C C T C G G A A A C C G G G A T T A A T A C C C T A T A T A C T T G A A G



25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165
 A T C G A C C G A G G A A T A T A T A T A A G A A A T C T T C G G G T G A A A T A T A T A T T T C T T A G C G C G G A C G G G T G A G T A A C C G C T G A G G A A C C T G C C T T G T A C T A G G G A T A G C C T C G G A A A C C G G G A T T A A T A C C C T A T A T A C T T G A A G



25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165
 A T C G A C C G A G G A A T A T A T A T A A G A A A T C T T C G G G T G A A A T A T A T A T T T C T T A G C G C G G A C G G G T G A G T A A C C G C T G A G G A A C C T G C C T T G T A C T A G G G A T A G C C T C G G A A A C C G G G A T T A A T A C C C T A T A T A C T T G A A G



25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165
 T G C T G T C G A C G A A C C G C T T T C A C A G A T C C C C T T C G G G G T G A A G A G G A A G T G A C T T A G T G C G G G A C G G G T G A G T A A C C G C T G G G T A A C C T C C T C A T A C A G G G G A T A A C A G C C G G A A C C G G T G C T A A G A C C G C A T A A G C A C

Abbreviations

AD	Anaerobic digestion
ASV	Amplicon sequence variant
BLAST	Basic Local Sequence Alignment Tool
BSC	Biological Species Concept
CoA	Coenzyme-A
CoB	Coenzyme-B
CoM	Coenzyme-M
FA	Formaldehyde
FISH	Fluorescence <i>In-Situ</i> Hybridisation
GWP	Global Warming Potential
H ₄ MPT	Tetrahydromethanopterin
iChip	isolation Chip
MFR	Methanofuran
PFA	Paraformaldehyde
TERS	Tip-Enhanced Raman Spectroscopy

References

- Adam, P. S., Borrel, G., Brochier-Armanet, C. and Gribaldo, S. 'The growing tree of Archaea: new perspectives on their diversity, evolution and ecology' *ISME J* **11**, 2407–2425 (2017).
- Ahmed, S. I., Johari, A., Hashim, H., Lim, J. S., Jusoh, M., Mat, R., *et al.* 'Economic and environmental evaluation of landfill gas utilisation: A multi-period optimisation approach for low carbon regions' *Int Biodeter Biodegr* **102**, 191–201 (2015).
- Alauzet, C., Marchandin, H., Courtin, P., Mory, F., Lemée, L., Pons, J.-L., *et al.* 'Multilocus analysis reveals diversity in the genus *Tissierella*: Description of *Tissierella carlieri* sp. nov. in the new class *Tissierellia* classis nov.' *Syst Appl Microbiol* **37**, 23–34 (2014).
- Allen, M. R., Shine, K. P., Fuglestedt, J. S., Millar, R. J., Cain, M., Frame, D. J., *et al.* 'A solution to the misrepresentations of CO₂-equivalent emissions of short-lived climate pollutants under ambitious mitigation' *npj Climate and Atmospheric Science* **1**, 16 (2018).
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. 'Basic local alignment search tool' *J Mol Biol* **215**, 403–410 (1990).
- Andrews, S. *Babraham Bioinformatics FastQC* (2010). Accessable at: www.bioinformatics.babraham.ac.uk/project/fastqc/
- Bapteste, E., Brochier, C. and Boucher, Y. 'Higher-level classification of the Archaea: evolution of methanogenesis and methanogens.' *Archaea* **1**, 353–63 (2005).
- Bell, P. E., Herlihy, A. T. and Mills, A. L. 'Establishment of anaerobic, reducing conditions in lake sediment after deposition of acidic, aerobic sediment by a major storm' *Biogeochemistry* **9**, 99–116 (1990).
- Beye, M., Fahsi, N., Raoult, D. and Fournier, P.-E. 'Careful use of 16S rRNA gene sequence similarity values for the identification of *Mycobacterium* species.' *New Microbes New Infect* **22**, 24–29 (2018).
- Bouallagui, H., Touhami, Y., Ben Cheikh, R. and Hamdi, M. 'Bioreactor performance in anaerobic digestion of fruit and vegetable wastes' *Process Biochem* **40**, 989–995 (2005).
- Buckel, W. and Barker, H. A. 'Two pathways of glutamate fermentation by anaerobic bacteria.' *J Bacteriol* **117**, 1248–60 (1974).
- Burke, S. A. and Krzycki, J. A. 'Reconstitution of Monomethylamine:Coenzyme M methyl transfer with a corrinoid protein and two methyltransferases purified from *Methanosarcina barkeri*.' *J Biol Biochem* **272**, 16570–7 (1997).

- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A. and Holmes, S. P. 'DADA2: High-resolution sample inference from Illumina amplicon data' *Nat Methods* **13**, 581–583 (2016).
- Chen, Y.-L., Lee, C.-C., Lin, Y.-L., Yin, K.-M., Ho, C.-L. and Liu, T. 'Obtaining long 16S rDNA sequences using multiple primers and its application on dioxin-containing samples' *Bioinformatics* **16**, S13 (2015).
- Chen, Y., Ye, W., Zhang, Y. and Xu, Y. 'High speed BLASTN: an accelerated MegaBLAST search tool' *Nucleic Acids Res* **43**, 7762–7768 (2015).
- Chouari, R., Le Paslier, D., Dauga, C., Daegelen, P., Weissenbach, J. and Sghir, A. 'Novel major bacterial candidate division within a municipal anaerobic sludge digester.' *Appl Environ Microbiol* **71**, 2145–53 (2005).
- Diéguez, A. L., Balboa, S., Magnesen, T. and Romalde, J. L. 'Neptuniibacter pectenicola sp. nov. and Neptuniibacter marinus sp. nov., two novel species isolated from a Great scallop (*Pecten maximus*) hatchery in Norway and emended description of the genus Neptuniibacter.' *Syst Appl Microbiol* **40**, 80–85 (2017).
- Doolittle, W. F., Bapteste, E. and Berg, O. 'Pattern pluralism and the Tree of Life hypothesis' *PNAS* **104**, 2043–2049 (2007).
- DSMZ *Methanogenium Medium* (2017). Accessable at:
www.dsmz.de/microorganisms/pdf/DSMZ_Medium141.pdf
- Dupré, J. 'In defence of classification' *Stud. Hist. Philos. Sci. C* **32**, 203–219 (2001).
- Edgar, R. C. and Flyvbjerg, H. 'Error filtering, pair assembly and error correction for next-generation sequencing reads' *Bioinformatics* **31**, 3476–3482 (2015).
- Eme, L., Spang, A., Lombard, J., Stairs, C. W. and Ettema, T. J. G. 'Archaea and the origin of eukaryotes' *Nat Rev Microbiol* **15**, 711–723 (2017).
- Evans, P. N., Parks, D. H., Chadwick, G. L., Robbins, S. J., Orphan, V. J., Golding, S. D., *et al.* 'Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics' *Science* **350**, 434–438 (2015).
- Ewing, B. and Green, P. 'Base-calling of automated sequencer traces using phred. II. Error probabilities.' *Genome Res* **8**, 186–94 (1998).
- Ewing, B., Hillier, L., Wendl, M. C. and Green, P. 'Base-calling of automated sequencer traces using phred. I. Accuracy assessment.' *Genome Res* **8**, 175–85 (1998).
- Gagliano, M. C., Braguglia, C. M., Gallipoli, A., Gianico, A. and Rossetti, S. 'Microbial diversity in innovative mesophilic/thermophilic temperature-phased anaerobic digestion of sludge' *Environ Sci Pollut Res* **22**, 7339–7348 (2015).

- Garcia, J.-L., Macarie, H., Cayol, J.-L., Fardeau, M.-L., Thomas, P., Patel, B. K. C., *et al.* ‘Sporanaerobacter acetigenes gen. nov., sp. nov., a novel acetogenic, facultatively sulfur-reducing bacterium.’ *Int J Syst Evol Microbiol* **52**, 1217–1223 (2002).
- Hey, J. ‘The mind of the species problem’ *Trends Ecol Evol* **16**, 326–329 (2001).
- Hobson, P. N. and Stewart, C. S. *The Rumen Microbial Ecosystem* (1997).
- Huang, W. E., Griffiths, R. I., Thompson, I. P., Bailey, M. J. and Whiteley, A. S. ‘Raman Microscopic Analysis of Single Microbial Cells’ *Anal Chem* (2004).
- Illumina ‘16S Metagenomic Sequencing Library Preparation’ (2013). Accessable at: http://emea.support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html
- Jones, W. J., Stugard, C. E. and Jannasch, H. W. ‘Comparison of thermophilic methanogens from submarine hydrothermal vents’ *Arch Microbiol* **151**, 314–318 (1989).
- Kaerberlein, T., Lewis, K. and Epstein, S. S. ‘Isolating "Uncultivable" Microorganisms in Pure Culture in a Simulated Natural Environment’ *Science* **296**, 1127–1129 (2002).
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., *et al.* ‘Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies’ *Nucleic Acids Res* **41**, e1–e1 (2013).
- Knowles, D. P., Kappmeyer, L. S., Haney, D., Herndon, D. R., Fry, L. M., Munro, J. B., *et al.* ‘Discovery of a novel species, *Theileria haneyi* n. sp., infective to equids, highlights exceptional genomic diversity within the genus *Theileria* : implications for apicomplexan parasite surveillance’ *Int J Parasitol* **48**, 679–690 (2018).
- Koeck, D. E., Hahnke, S. and Zverlov, V. V ‘*Herbinix luporum* sp. nov., a thermophilic cellulose-degrading bacterium isolated from a thermophilic biogas reactor’ *Int J Syst Evol Microbiol* **66**, 4132–4137 (2016).
- Kumar, S., Stecher, G. and Tamura, K. ‘MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets’ *Mol Biol Evol* **33**, 1870–1874 (2016).
- Lang, K., Schuldes, J., Klingl, A., Poehlein, A., Daniel, R. and Brunea, A. ‘New mode of energy metabolism in the seventh order of methanogens as revealed by comparative genome analysis of “*Candidatus methanoplasma termitum*”.’ *Appl Environ Microbiol* **81**, 1338–52 (2015).
- Latosov, E., Looorits, M., Maaten, B., Volkova, A. and Soosaar, S. ‘Corrosive effects of H₂S and NH₃ on natural gas piping systems manufactured of carbon steel’ *Energ Procedia* **128**, 316–323 (2017).

- Li, T., Mazéas, L., Sghir, A., Leblon, G. and Bouchez, T. ‘Insights into networks of functional microbes catalysing methanization of cellulose under mesophilic conditions’ *Environ Microbiol* **11**, 889–904 (2009).
- Limam, R. D., Chouari, R., Mazéas, L., Wu, T.-D., Li, T., Grossin-Debattista, J., *et al.* ‘Members of the uncultured bacterial candidate division WWE1 are implicated in anaerobic digestion of cellulose’ *MicrobiologyOpen* **3**, 157–167 (2014).
- Liu, D., Liu, D., Zeng, R. J. and Angelidaki, I. ‘Hydrogen and methane production from household solid waste in the two-stage fermentation process’ *Water Res* **40**, 2230–2236 (2006).
- Liu, Y. and Whitman, W. B. ‘Metabolic, Phylogenetic, and Ecological Diversity of the Methanogenic Archaea’ *Ann N Y Acad Sci* **1125**, 171–189 (2008).
- Lorenz, B., Wichmann, C., Stöckel, S., Rösch, P. and Popp, J. ‘From One to Many Cultivation-Free Raman Spectroscopic Investigations of Bacteria’ (2017).
- Machado, R. A. R., Wüthrich, D., Kuhnert, P., Arce, C. C. M., Thönen, L., Ruiz, C., *et al.* ‘Whole-genome-based revisit of Photorhabdus phylogeny: proposal for the elevation of most Photorhabdus subspecies to the species level and description of one novel species Photorhabdus bodei sp. nov., and one novel subspecies Photorhabdus laumondii subsp.’ *Int J Syst Evol Microbiol* **68**, 2664–2681 (2018).
- Martin, M. ‘Cutadapt removes adapter sequences from high-throughput sequencing reads’ *EmbnetJournal* **17**, 10 (2011).
- Mayr, E. *Systematics and the origin of species, from the viewpoint of a zoologist* (1999).
- Méthé, B. A., Leschine, S. B. and Warnick, T. A. ‘Clostridium phytofermentans sp. nov., a cellulolytic mesophile from forest soil’ *Int J Syst Evol Microbiol* **52**, 1155–1160 (2002).
- Moeller, L. and Zehnsdorf, A. ‘Process upsets in a full-scale anaerobic digestion bioreactor: over-acidification and foam formation during biogas production’ *Energy Sustainability Soc* **6**, 30 (2016).
- Nakamura, S., Okado, I., Abe, T. and Nishida, S. ‘Taxonomy of Clostridium tetani and Related Species’ *J Gen Microbiol* **113**, 29–35 (1979).
- New England Biolabs *PCR Using Q5® High-Fidelity DNA Polymerase (M0491)* (2013).
- Nguyen, N.-P., Warnow, T., Pop, M. and White, B. ‘A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity’ *NPJ Biofilms and Microbiomes* **2**, 16004 (2016).
- Nichols, D., Cahoon, N., Trakhtenberg, E. M., Pham, L., Mehta, A., Belanger, A., *et al.* ‘Use of Ichip for High-Throughput In Situ Cultivation of “Uncultivable” Microbial Species’ *Appl Environ Microbiol* **76**, 2445–2450 (2010).

- O’Leary, N. A., Wright, M. W., Brister, J. R., Ciuffo, S., Haddad, D., McVeigh, R., *et al.*
‘Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion,
and functional annotation’ *Nucleic Acids Res* **44**, D733–D745 (2016).
- Okonechnikov, K., Golosova, O., Fursov, M. and UGENE team ‘Unipro UGENE: a unified
bioinformatics toolkit’ *Bioinformatics* **28**, 1166–1167 (2012).
- Ozbayram, E., Ince, O., Ince, B., Harms, H. and Kleinstuber, S. ‘Comparison of Rumen and
Manure Microbiomes and Implications for the Inoculation of Anaerobic Digesters’
Microorganisms **6**, 15 (2018).
- Pelletier, E., Kreimeyer, A., Bocs, S., Rouy, Z., Gyapay, G., Chouari, R., *et al.* “‘Candidatus
Cloacamonas acidaminovorans’”: genome sequence reconstruction provides a first
glimpse of a new bacterial division.’ *J Bacteriol* **190**, 2572–9 (2008).
- Piddock, L. J. V. ‘Teixobactin, the first of a new class of antibiotics discovered by iChip
technology?’ *J Antimicrob Chemother* **70**, 2679–2680 (2015).
- Pittam, D. A. and Pilcher, G. ‘Measurements of heats of combustion by flame calorimetry.
Part 8.—Methane, ethane, propane, n-butane and 2-methylpropane’ *Journal of the
Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*
68, 2224 (1972).
- Podosokorskaya, O. A., Bonch-Osmolovskaya, E. A., Beskorovaynyy, A. V., Toshchakov, S.
V., Kolganova, T. V. and Kublanov, I. V. ‘Mobilitea sibirica gen. nov., sp. nov., a
halotolerant polysaccharide-degrading bacterium’ *Int J Syst Evol Microbiol* **64**, 2657–
2661 (2014).
- Polz, M. F. and Cavanaugh, C. M. ‘Bias in template-to-product ratios in multitemplate PCR.’
Appl Environ Microbiol **64**, 3724–30 (1998).
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., *et al.* ‘The SILVA
ribosomal RNA gene database project: improved data processing and web-based tools’
Nucleic Acids Res **41**, D590–D596 (2012).
- Ragsdale, S. W. ‘Enzymology of the wood-Ljungdahl pathway of acetogenesis.’ *Ann N Y
Acad Sci* **1125**, 129–36 (2008).
- Ricaboni, D., Mailhe, M., Benezech, A., Andrieu, C., Fournier, P.-E. and Raoult, D.
“‘Pseudoflavonifractor phocaeensis” gen. nov., sp. nov., isolated from human left colon’
New Microbes New Infect **17**, 15–17 (2017).
- Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., *et al.*
‘Insights into the phylogeny and coding potential of microbial dark matter’ *Nature* **499**,
431–437 (2013).

- Saitou, N. and Nei, M. 'The neighbor-joining method: a new method for reconstructing phylogenetic trees.' *Mol Biol Evol* **4**, 406–25 (1987).
- Saunders, A. M., Albertsen, M., Vollertsen, J. and Nielsen, P. H. 'The activated sludge ecosystem contains a core community of abundant organisms' *ISME J* **10**, 11–20 (2016).
- Sayers, E. W., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., *et al.* 'Database resources of the National Center for Biotechnology Information' *Nucleic Acids Res* **37**, D5–D15 (2009).
- Scarlat, N., Dallemand, J.-F. and Fahl, F. 'Biogas: Developments and perspectives in Europe' *J Renewable Energy* **129**, 457–472 (2018).
- Schloss, P. D., Girard, R. A., Martin, T., Edwards, J. and Thrash, J. C. 'Status of the Archaeal and Bacterial Census: an Update.' *MBIO* **7**, e00201-16 (2016).
- Schmittgen, T. D., Lee, E. J. and Jiang, J. 'High-Throughput Real-Time PCR' in *Methods Mol Biol* 89–98 (2008).
- Shannon, C. E. *A Mathematical Theory of Communication Bell Labs Tech J* (1948).
- SILVA Ribosomal Database Project *FISH protocols SILVA* (2010). Accessable at: www.arb-silva.de/fish-probes/fish-protocols/
- SIMPSON, E. H. 'Measurement of Diversity' *Nature* **163**, 688–688 (1949).
- Stöckel, S., Kirchhoff, J., Neugebauer, U., Rösch, P. and Popp, J. 'The application of Raman spectroscopy for the detection and identification of microorganisms' *J Raman Spectrosc* **47**, 89–109 (2016).
- Strola, S. A., Baritoux, J.-C., Schultz, E., Simon, A. C., Allier, C., Espagnon, I., *et al.* 'Single bacteria identification by Raman spectroscopy' *J Biomed OPT* **19**, 111610 (2014).
- Styles, D., Dominguez, E. M. and Chadwick, D. 'Environmental balance of the UK biogas sector: An evaluation by consequential life cycle assessment' *Sci Total Environ* **560–561**, 241–253 (2016).
- Sun, L., Liu, T., Müller, B. and Schnürer, A. 'The microbial community structure in industrial biogas plants influences the degradation rate of straw and cellulose in batch tests.' *Biotechnol Biofuels* **9**, 128 (2016).
- Sun, L., Pope, P. B., Eijsink, V. G. H. and Schnürer, A. 'Characterization of microbial community structure during continuous anaerobic digestion of straw and cow manure' *Microb Biotechnol* **8**, 815–827 (2015).
- Suzuki, M. T. and Giovannoni, S. J. 'Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR.' *Appl Environ Microbiol* **62**, 625–30 (1996).

- Tanaka, T., Kawasaki, K., Daimon, S., Kitagawa, W., Yamamoto, K., Tamaki, H., *et al.* ‘A Hidden Pitfall in the Preparation of Agar Media Undermines Microorganism Cultivability’ *Appl Environ Microbiol* **80**, 7659–7666 (2014).
- Tedersoo, L., Anslan, S., Bahram, M., Põlme, S., Riit, T., Liiv, I., *et al.* ‘Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi’ *MYCOKEYS* **10**, 1–43 (2015).
- Thauer, R. K., Kaster, A.-K., Seedorf, H., Buckel, W. and Hedderich, R. ‘Methanogenic archaea: ecologically relevant differences in energy conservation’ *Nat Rev Microbiol* **6**, 579–591 (2008).
- Thomas, R. H. ‘Molecular Evolution and Phylogenetics’ *Heredity* **86**, 385–385 (2001).
- Tiedje, J. M., Sexstone, A. J., Parkin, T. B. and Revsbech, N. P. ‘Anaerobic processes in soil’ *Plant Soil* **76**, 197–212 (1984).
- Ueki, A., Ohtaki, Y., Kaku, N. and Ueki, K. ‘Descriptions of *Anaerotaenia torta* gen. nov., sp. nov. and *Anaerocolumna cellulositytica* gen. nov., sp. nov. isolated from a methanogenic reactor of cattle waste and reclassification of *Clostridium aminovalericum*, *Clostridium jejuense* and *Clostridium xyla*’ *Int J Syst Evol Microbiol* **66**, 2936–2943 (2016).
- Vanwonterghem, I., Evans, P. N., Parks, D. H., Jensen, P. D., Woodcroft, B. J., Hugenholtz, P., *et al.* ‘Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota’ *Nat Microbiol* **1**, 16170 (2016).
- Vasco-Correa, J., Khanal, S., Manandhar, A. and Shah, A. ‘Anaerobic digestion for bioenergy production: Global status, environmental and techno-economic implications, and government policies’ *Bioresour Technol* **247**, 1015–1026 (2018).
- Wang, Q., Garrity, G. M., Tiedje, J. M. and Cole, J. R. ‘Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy.’ *Appl Environ Microbiol* **73**, 5261–7 (2007).
- Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K.-H., *et al.* ‘Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences’ *Nat Rev Microbiol* **12**, 635–645 (2014).
- Ye, R., Jin, Q., Bohannon, B., Keller, J. K., Mcallister, S. A. and Bridgham, S. D. ‘pH controls over anaerobic carbon mineralization, the efficiency of methane production, and methanogenic pathways in peatlands across an ombrotrophic-minerotrophic gradient’ *Soil Biol Biochem* **54**, 36–47 (2012).

Zinder, S. H., Cardwell, S. C., Anguish, T., Lee, M. and Koch, M. 'Methanogenesis in a Thermophilic (58 degrees C) Anaerobic Digestor: Methanotherix sp. as an Important Aceticlastic Methanogen.' *Appl Environ Microbiol* **47**, 796–807 (1984).