THE UNIVERSITY
of EDINBURGH

Metataxonomic analysis of a novel anaerobic digester and engineering of isolated bacteria

Lukas Frederik Mühlbauer

THE UNIVERSITY *of* EDINBURGH

Thesis submitted for the degree of
Doctor of Philosophy

The University of Edinburgh
School of Biological Sciences

2022

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where it states otherwise by reference or acknowledgment, the work presented is entirely my own.

(Digital signature redacted for electronic thesis submission)

_____

Lukas Frederik Mühlbauer, 7th January 2022

Acknowledgments

First and foremost, I would like to express my sincere thanks and gratefulness to my supervisors, Prof Louise Horsfall and Dr Andrew Free, for their invaluable support throughout the last 4.5 years and for making this project possible in the first place.

Huge thank you also to the postdocs in the Horsfall Lab, Dr Annegret Honsbein, Dr Matthew Edmundson, Dr Michael Capeness and Dr Virginia Echavarri-Bravo, for their helpful comments on my thesis drafts and/or the advice they have given me throughout this project. I would also like to thank Mrs Fiona Strathdee from the Free Lab, who was of great help during the first few months of my project. Thanks also to Dr Jimmy Roussel and Ms Louise Austin from the industrial partner, Blue Sky Bio Ltd., for providing the digestate samples and information about their bioreactor.

I would also like to acknowledge all friends and family who have been with me the longest throughout my studies, with special thanks to Sharif, Will, Gwen, Omar, Mairi, George and Sonya. I want to express my special and most genuine gratitude to my parents Martina and Günther, for their moral, emotional, and financial support.

Finally, I want to thank Xiawen, for her love, support, encouragement, and bioinformatics advice over the past few years.

Lay summary

Anaerobic digestion (AD) is the biological process, in which microorganisms break down organic matter and produce biogas, which can be used as a renewable energy source. The process is divided into four stages and different groups of microorganisms carry out each of them. When used industrially, AD is performed in large vessels, so-called bioreactors. Here, different groups of microorganisms share the same space, despite having different preferences for their ideal growth conditions. This project's industrial partner, Blue Sky Bio Ltd. (Chester, UK) has developed a bioreactor that, unlike traditional systems, is divided into seven chambers.

In this project, the microorganisms involved in its operation have been characterised for the first time. The idea behind this design is to physically divide the AD process into several compartments, allowing the different groups of microorganisms to grow better and produce more biogas. The organic matter that was digested in this reactor consisted of either grass or seaweed. When the industrial partner was operating their reactor, samples from each of the chambers were taken. These contained the microorganisms that were present at the time, and appropriate methods were used to identify what types of organisms were involved in digesting grass and seaweed in the bioreactor.

From the results, it was possible to see that the design of the reactor physically divided the four groups of microorganisms that are responsible for the performance of each of the AD process stages, confirming initial expectations. Furthermore, it was seen that, in the first AD stages, different microorganisms were involved in the breakdown of either grass or seaweed. Then, some of the microorganisms were taken from the samples and propagated in the laboratory to assess their capabilities to produce butyric acid and hydrogen gas, two products that are of wide industrial interest as potential sources of fuel.

For example, butyric acid can be converted into butanol, which is used in the perfume industry but can also be used as a replacement for gasoline, while hydrogen can be a zero-carbon fuel for use in internal combustion engines or to produce electricity in fuel cells, both methods which only result

in water as the end product. Two organisms with desirable features were identified and methods were developed to modify their behaviour with the aim to enhance their productivity. The behaviour of modified organisms was altered, albeit not in a way that was initially expected. Still, the work carried out as part of this project has generated insights into the AD process and methods to modify organisms for industrial purposes.

Abstract

Biomolecular tools such as metataxonomic and metagenomic DNA sequencing approaches have facilitated our detailed understanding of the microbiology of anaerobic digestion (AD) systems in recent years. However, most studies have focussed on single-stage or two-stage AD systems.

In this project, the microbiology of a novel multi-stage AD reactor developed by Blue Sky Bio Ltd. (Chester, UK) was characterised for the first time. The results of this analysis suggest that changes in key species abundance follow the longitudinal separation of AD stages (i.e. hydrolysis, acidogenesis, acetogenesis, methanogenesis).

Furthermore, the impact of a change in feedstock on the reactor's microbiology was observed, indicating that microbial communities become less feedstock-specific as biological matter is digested from complex molecules to more simple compounds.

After characterisation of the community, the isolation and identification of several species involved in the production of volatile fatty acids (VFAs) was carried out. This was finally narrowed down to two species of the genus *Clostridium*, one of which was shown to be dissimilar enough to other members of the genus to be declared a newly identified species by established delimitation thresholds.

The isolates were recognised for their capabilities to produce high amounts of butyric acid as well as hydrogen. Protocols to genetically engineer these bacteria have been established and the GUS reporter from *E. coli* has been expressed in both bacteria to assess the strength of several promoters. After working native promoters were identified, several hydrogenases and butyrate kinase genes were overexpressed in both isolates using the pMTL80000 shuttle plasmids. Effects on gas production and VFA yield was measured and significant changes were observed, however contrary to the initial hypothesis based on comparative genomics and previous studies in related *Clostridium* species, observed changes in the isolates' metabolism were not characterised by a substantial increase in the production of butyric acid and hydrogen.

Both end products are desirable for their industrially useful properties, mainly as they can be converted or used directly as transport fuels. In particular, butyric acid can be converted into butanol, which is used in the perfume industry but can also be used as a replacement for gasoline, while hydrogen can be a zero-carbon fuel for use in internal combustion engines or to produce electricity in fuel cells, both methods which only result in water as the end product. Still, the work carried out as part of this project has generated insights into multi-stage AD systems and the potential to engineer these newly identified non-model organisms for industrial applications.

Table of contents

List of abbreviations

Abbreviation	Meaning
16S	16 Svedberg (Non-SI metric unit for sedimentation coefficients)
AA	Amino acid
ABE	Acetone-Butanol-Ethanol
ABR	Anaerobic baffled reactor
ABS	Absorbance
AD	Anaerobic Digestion
adk	Adenylate kinase
ANI	Average nucleotide identity
ASV	Amplicon sequence variant
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BPROM	Bacterial Promoter Prediction Software
buk	Butyrate kinase
cDNA	Complementary DNA
COD	Chemical oxygen demand
Cq	Quantification cycle
D-GENIES	Dot plot large Genomes in an Interactive, Efficient and Simple way
DGGE	Denaturing gradient gel electrophoresis
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
dsDNA	Double-stranded DNA
dT	Deoxythymidylic acid
EC	Enzyme commission number
FC	Relative expression fold change (e.g. logFC)
Fd	Ferredoxin

FDR	False Discovery Rate
Fig.	Figure
GC	guanine-cytosine (content)
GC-FID	Gas chromatography with flame-ionisation detection
gDNA	Genomic DNA
gen.	Genus
GUS	Beta-glucuronidase
GW	Gateway (vector)
HRT	Hydraulic retention time
hydA	[FeFe] hydrogenase
IEA	International Energy Agency
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
IUPAC	International Union of Pure and Applied Chemistry
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Lysogeny broth
Mb	Mega base pairs
MCS	Multiple cloning site
MRC PPU	Medical Research Council Protein Phosphorylation and Ubiquitylation Unit
MRCA	Most recent common ancestor
mRNA	Messenger RNA
NAD(H)	Nicotinamide adenine dinucleotide (+ hydrogen)
NCBI	National Center for Biotechnology Information
NEB	New England BioLabs
OD_{600}	Optical density at 600 nm
ORF	Open reading frame
OTU	Operational taxonomic unit
PAGE	Polyacrylamide gel electrophoresis
PATRIC	Pathosystems Resource Integration Center
PBS-T	Phosphate buffered saline containing 0.1% Tween 20
PCoA	Principal coordinates analysis

PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PERMADISP	Permutational multivariate analysis of dispersion
PERMANOVA	Permutational multivariate analysis of variance
PES	Polyethersulfone
PFV	Polyvinyl fluoride
PHA	Polyhydroxyalkanoates
PNPG	P-nitrophenyl-β-glucuronide
POCP	Percentage of conserved proteins
PVC	Polyvinyl chloride
PYG	Peptone-yeast-glucose medium
QIIME	Quantitative Insights into Microbial Ecology
qPCR	Quantitative PCR
RNA	Ribonucleic acid
ROX	Carboxyrhodamine
rRNA	Ribosomal RNA
rrs	16S rRNA
RT-qPCR	Reverse transcriptase qPCR
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SMRT	single-molecule real-time sequencing
sp.	Species
T	Transmittance
TSS	Transformation and Storage Solution
UK	United Kingdom
USA	United States of America
V(1-9)	Variable region 1-9 of 16S rRNA
VFA	Volatile fatty acid
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

List of figures

List of tables

1 Introduction

1.1 Bioenergy

In the latest report on Key World Energy Statistics by the International Energy Agency (IEA, 2021b), the world total energy generation has more than quadrupled from 1973 to 2019. This has been accompanied by a 117% increase in CO_2 emissions, as coal, oil and natural gas remain the key providers of energy. In an effort to reduce dependence on fossil fuels, the relative amount of non-hydro renewables (i.e. solar, wind, tidal, geothermal, biofuels, waste and heat) used for electricity production has increased from 0.6% in 1973 to 10.8% in 2019. Biofuels, which include liquid fuels such as bioethanol and biodiesel as well as gaseous fuels such as biogas, play a large role in the aim to reduce CO_2 emissions and reach the IEA's Net Zero Scenario. A main benefit of biofuels is that they are one of the few non-hydro renewable energy options that can be stored easily for energy production and as transport fuels. Over the last decade, their consumption has increased by 5% per year (IEA, 2021a) as liquid biofuels such as bioethanol and biodiesel have been adopted as a replacement for fossil fuel-derived petrol on a large scale. Most transport biofuels are so-called first generation biofuels that are made from high energy crops such as sugar cane and corn (Mumm *et al.*, 2014). Due to the competition for agricultural land by using traditional food crops as an energy source, these have been identified as a driver for food price spikes and volatility, which has caused a controversy about their use (Tadasse *et al.*, 2016). Among the biogas produced in Europe, 40% is still made from high energy crops, however a majority of biogas today already comes from valorised waste products such as animal manure, municipal solid waste and municipal wastewater. It is estimated that biogas from these wastestreams could cover 20% of today's worldwide gas demand if their full potential was utilised (IEA, 2020), which means that additional feedstocks for their production will have to be explored if biofuels are to completely replace fossil fuels.

Biofuels produced from sources that do not compete with crops are called second-generation biofuels and are generally lignocellulosic feedstocks grown on non-arable land. For example, shrubs and grasses have been proposed as a feedstock for biogas production, as they can easily grow on agriculturally degraded lands without added fertiliser and irrigation (Naik *et al.*, 2010). Tilman, Hill and Lehman (2006) showed that the use of low-input high-diversity grassland perennials is carbon negative due to net CO_2 sequestration in soil and roots and could become a sustainable method to fulfil global energy demands. Nevertheless, Russelle *et al.* (2007) disagreed, arguing that management inputs were understated and that findings from one localised site were wrongly extrapolated to apply globally. Also, due to the many different types of feedstock, production routes (e.g. biogas, biodiesel) and the fact that second-generation biofuels are yet to be fully commercialised, the consensus on their sustainability is mixed (Jeswani, Chilvers and Azapagic, 2020).

Another alternative is found in the so-called third-generation biofuels, which are made from algal biomass. Compared to shrubs and grasses, the use of algae benefits from their low lignocellulose contents, making them easier to degrade biologically (Montingelli, Tedesco and Olabi, 2015). The largest proportion of third-generation biofuels is represented by biodiesel made from microalgae, which can produce higher yields of processable oil per area compared to second-generation crops such as *Jatropha* or sunflower (Chowdhury and Loganathan, 2019). Still, according to Jeswani *et al.* (2020), a majority of studies concluded that the production of algal biodiesel currently has higher green house gas emissions than that of fossil diesel, due to it currently being at an early stange of development. Conversely, as fewer necessary processing steps are involved, the production of biogas from microalgae has shown to be potentially carbon negative, which means that more carbon dioxide is captured from the air and turned into biodiesel than is being emitted in the processing of the microalgae (Chowdhury and Freire, 2015). Compared to land-based second-generation feedstocks, marine macroalgae (i.e. seaweed) have the advantage of higher growth rates, while studies comparing several feedstocks found that the anaerobic digestion of

selected seaweed species can produce more biogas than the digestion of grass and rice straw (Vanegas and Bartlett, 2013)

1.2 Anaerobic digestion

The process to produce biogas from a wide range of digestible feedstocks (e.g. manure, wastewater, high energy crops, grasses, seaweed) using microbial communities is called anaerobic digestion (AD). AD is a multi-step process in which organic matter is broken down from complex molecules (i.e. proteins and carbohydrates) into simple chemicals such as volatile fatty acids (VFAs, e.g. acetic acid, butyric acid, propionic acid), CO_2, H_2 and methane, in the absence of oxygen (Adekunle and Okolie, 2015). The biogas produced by this process primarily consists of methane but also contains carbon dioxide and other trace gases, which can be removed through water scrubbing and membrane separation to produce pure biomethane (Ofori-Boateng and Kwofie, 2009; Vrbová and Ciahotný, 2017). The breakdown of complex substrates in AD occurs in several stages (Fig. 1), which are carried out by distinct groups of microorganisms within the domains Bacteria and Archaea. The first stage is hydrolysis, in which hydrolytic-fermentative bacteria cover the surface of non-soluble substrates such as complex sugars, lipids or proteins and secrete hydrolysing enzymes (e.g. cellulase, amylase, protease, lipase). These enzymes catalyse reactions to break down complex substrates into simple substrates such as monosaccharides and amino acids. (Gunaseelan, 1997). In the second stage (acidogenesis), different facultative and obligatory anaerobic bacteria take up the produced monomers and ferment them into organic acids (e.g. VFAs), alcohols (e.g. ethanol, methanol) and carbon dioxide. Simple molecules that are produced in this step are then consumed by acetogenic bacteria, which produce acetic acid, hydrogen gas and carbon dioxide in a process called acetogenesis (Adekunle and Okolie, 2015). Finally, methane is produced by methanogenic archaeal species in the last stage of the process, methanogenesis. Methanogenic archaea fall into different groups, depending on the compounds they use in the production of methane. Hydrogenotrophic species produce methane from hydrogen and

carbon dioxide, acetoclastic methanogens utilise acetic acid to produce methane and carbon dioxide, and methyltrophic species produce methane from methanol (Demirel and Scherer, 2008)

The review by Thauer (1998) covers the biochemistry of methanogenesis in great detail, explaining how the production of methane occurs in step-wise processes from different substrates, such as glucose, formate, acetate, methanol, H_2 and CO_2. All methanogens ultimately facilitate the production of methane through methyl-coenzyme M reductase, which combines coenzyme B as the hydrogen donor and coenzyme M as the methyl donor. Within the enzyme, this reaction is catalysed by the nickel-containing coenzyme F_{430}, which is highly unstable in contact with oxygen, explaining the strict anaerobic nature of methanogens that produce ATP through anaerobic methane oxidation (Kaneko *et al.*, 2014).

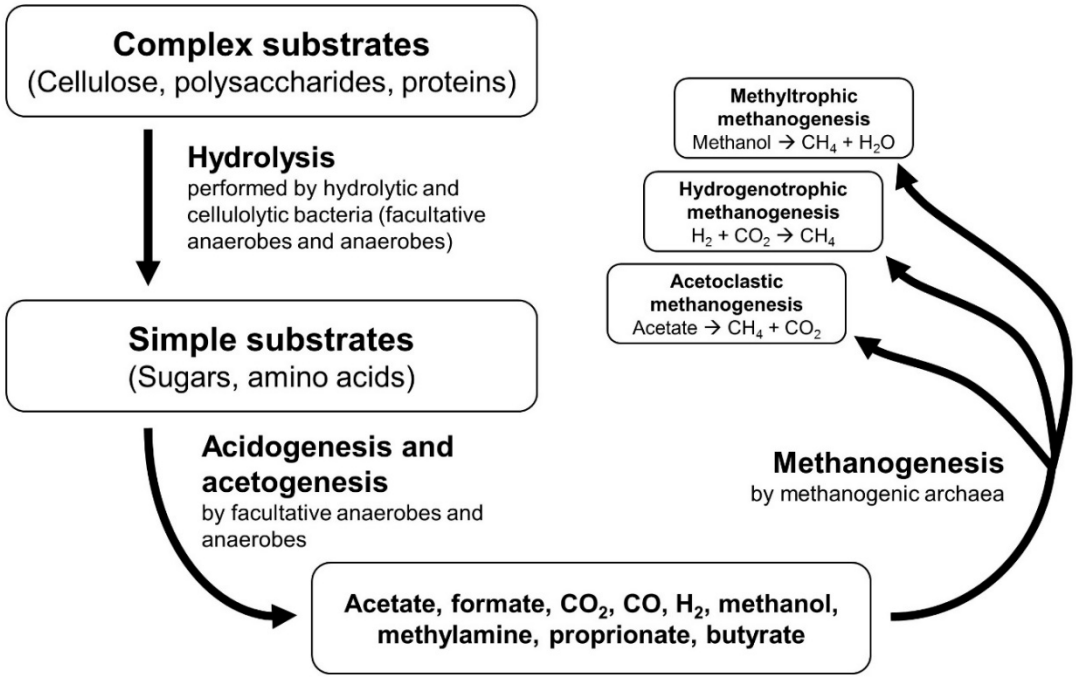


Figure 1. Stages of anaerobic digestion, in which microbial communities break down organic matter into its substrates and metabolise them into fatty acids and biogas. (Gunaseelan, 1997; Demirel and Scherer, 2008)

1.2.1 Types of anaerobic digesters

Anaerobic reactor types with varying levels of complexity are used to produce biogas from organic material. Overall, they differ in the number of compartments, moving parts (e.g. stirred tank reactors), whether they are fed continuously or in batches and their operating temperature (i.e. mesophilic or thermophilic). In Europe, the most common reactors are continuous mesophilic one-stage systems as their main advantage is their simple design and low investment costs (Bouallagui *et al.*, 2005). Nevertheless, due to multiple AD processes occurring in the same space, bacterial communities with distinct optimal growing conditions are negatively influencing each other. For example, the accumulation of VFAs produced by acidogenic bacteria leads to a drop in pH that can inhibit methanogenic archaea, causing an imbalance that reduces energy recovery from the digested biomass (Griffin *et al.*, 1998). As a result, the use of two-stage AD has been found to increase methane production performance by up to 43%, as methanogenesis is taking place in a separate compartment which can be controlled independently (Schievano *et al.*, 2014).

1.2.2 Anaerobic Baffled Reactors

Another type of AD reactor design is the anaerobic baffled reactor (ABR), which is not commonly found on a large scale. Like single stage reactors, these reactors also consist of a single vessel, which is however separated into interconnected chambers by a series of baffles, forcing digestate to flow under and over them. The group that first published on the concept of the ABR (Bachmann, Beard and McCarty, 1985) described its benefits in the fact that bacteria within the reactor move at a relatively slow rate, allowing digestate to come into contact with a large active biological mass. Similar to two-stage systems, the design of ABRs allows for the separation of acidogenesis and methanogenesis longitudinally down the reactor, which can increase both acidogenic and methanogenic activity as different microbial communities can grow under more favourable conditions (Barber and Stuckey, 1999). While behaving very similarly to two-stage reactors, ABRs have the

benefit of costing less and requiring less control (Weiland and Rozzi, 1991). Besides the high efficiency, further advantages include higher resistance to shock loadings and higher biomass retention capacity. Nevertheless, the design of ABR reactors also has drawbacks, which are found in a higher building complexity compared to single stage reactors, resulting in difficulties to distribute the influent evenly. Also, since ABRs have to operate within the limits of liquid and gas up/downflow velocities, they tend to be shallow, which causes a potential problem when it comes to process upscaling (Tilche and Vieira, 1991). Requiring a steady flow from inlet to outlet, ABRs have therefore mainly been used to digest high-liquid feedstocks such as wastewater (Barber and Stuckey, 1999).

1.2.3 Microbial communities involved in AD

Despite the widespread use of anaerobic digesters in the production of renewable energy, the structures of underlying microbial communities are not very well understood, so much so that the microbiome of bioreactors is often referred to as a "black box" (Koch et al., 2014). Studies on the characterisation of bacterial diversity within anaerobic digesters have yet to establish a common core of microbial key players, meaning a consortium of species that is shared by a majority of AD systems. One likely reason for this is that the microbial communities are very much dependent on the type of biomass that they digest, as well as process parameters such as temperature. For example, Klocke et al. (2007) reported that digesters fed with beet silage are rich in members of the Clostridiales, Deltaproteobacteria, Bacilli and Bacteroidetes. Likewise, Kampmann et al. (2012) described reactors fed with casein, starch and cream to be abundant in Firmicutes and Bacteroidetes. (Cirne et al. (2007) found that the microbial communities in the digestion of grass and sugar beets differed substantially, with either Alphaproteobacteria or Firmicutes being the dominant bacterial phyla for the two respective feedstocks. Overall, this is likely due to different specialisations within these taxa. For example, as later discussed in chapter 3.3.4., the Clostridiales and Bacteroidetes contain an abundance of cellulose-digesting species that would be well-suited for the

digestion of cellulose-rich feedstocks such as the aforementioned beet silage or grass.

Differences in AD communities are also dependent on factors such as operating temperatures. While thermophilic reactors (around 45–60°C) are faster at converting feedstocks to biogas and achieve higher methane yields, their microbial communities are less diverse (Moset *et al.*, 2015).

As single-stage reactor designs are most commonly used in AD, the majority of studies on the involved communities have focused on this widespread type of anaerobic digester. In comparison, studies on anaerobic baffled reactors are relatively scarce and there are wide variations in reactor designs (e.g. number of chambers and size) and feedstocks. Most publications have studied the communities of four- and five-chamber reactors digesting high-liquid feedstocks such as wastewater in bench-scale reactors (Table 1). As a result of the diversity of setups, common features are seldomly found in ABR communities of these studies. Nevertheless, along their ABRs, Gulhane *et al.* (2017), Tran *et al.* (2017), Zhang and Shi (2021) and Ziganshin *et al.* (2019) have reported that the abundance of species associated with the phylum Bacteroidetes was highest towards the feedstock inflow. This phylum is associated with the hydrolysis step in anaerobic digestion and metagenomics studies have argued that the phylum's hydrolytic potential and dominance in some anaerobic digesters is due to it comprising a high number of diverse hydrolytic enzymes (Bertucci *et al.*, 2019).

Table 1. Selected studies on microbial communities in anaerobic baffled reactors (ABR)

Author	Feedstock	Chambers	Total vol.	HRT	Organic load rate
Zhang *et al.*, 2011	Acetone-Butanol-Ethanol (ABE) fermentation wastewater	4	8 L	40 h	0.96 kg COD m^{-3} d^{-1}
Ban *et al.*, 2013	Sugar refinery wastewater	4	27.8 L	36 h	5.33 kg COD m^{-3} d^{-1}
Gulhane *et al.*, 2017	Vegetable waste	4	39 L	30 d	0.5 g VS/L/d
Su *et al.*, 2019	Pharmaceutical factory wastewater	4	2.5 L	12 – 24 h	Not mentioned (variable)
Ziganshin *et al.*, 2019	Artificial substrate	4	5 L	7 – 9 d	Not mentioned (see paper for substrate composition)
Zhang and Shi, 2021	Brown sugar	4	28.4 L	0.7 – 1.7 d	2.12 – 5.87 kg COD m^{-3} d^{-1}
Lin *et al.*, 2012	Nitrobenzene wastewater	5	15 L	24 h	0.34 – 3.32 mg nitrobenzene L^{-1} h^{-1}
Peng *et al.*, 2013	Glucose	5	15 L	24 h	1.2 kg COD m^{-3} d^{-1}
Chen *et al.*, 2016	Synthetic wastewater	5	6.43 L	N/A	71 – 284 mg COD m^{-3} d^{-1}
Giordani *et al.*, 2021	Artificial dairy waste	5	25 L	24 – 72 h	1 – 2.96 kg COD m^{-3} d^{-1}
Jiang *et al.*, 2018	Synthetic wastewater	7	20.5 L	30 h	2 g COD m^{-3} d^{-1}
Tran *et al.*, 2017	Factory wastewater	10	68 L	3.5 d	1.4 kg COD m^{-3} d^{-1}

1.3 Biohydrogen and volatile fatty acids

While the process of anaerobic digestion has mainly been used for the production of methane-containing biogas, the process can be optimised to produce other products such as hydrogen gas or volatile fatty acids (Khan *et al.*, 2016).

Hydrogen has been proposed as a clean transport fuel for almost 50 years (Dickson, Ryan and Smulyan, 1976), however historic forecasts about its future adoption have generally been too optimistic as predictions on “peak

oil" were defied (McDowall, 2016). As passenger transport is now increasingly converting to electric vehicles, Moriarty and Honnery (2019) predict that hydrogen will still play an important role in the future of freight transport.

The production of hydrogen from anaerobic digestion is usually referred to as "dark fermentation", which is a term that derives from its differences compared to photofermentation, employing photosynthetic bacteria such as *Rhodobacter* to produce hydrogen (Tao *et al.*, 2007). In order to shift the anaerobic digestion process towards the production of hydrogen, methanogenic activity has to be inhibited. This is generally done by lowering the pH to around 6 (Łukajtis *et al.*, 2018) and decreasing the hydraulic retention time so that the process does not advance to the methanogenesis stage (Romero Aguilar *et al.*, 2013). Furthermore, two-stage digesters can be optimised to produce both hydrogen and methane (Intanoo, Chaimongkol and Chavadej, 2016), while Jürgensen *et al.* (2015) found that gas produced in the first compartment of an anaerobic baffled reactor primarily consisted of hydrogen and contained almost no methane. Lee *et al.* (2011) reviewed studies on hydrogen production in dark fermentation reactors using pure cultures instead of microbial consortia and found that the most productive species are within the genera *Clostridium*, *Enterobacter, Thermotoga* and *Caldicellulosiruptor*.

Volatile fatty acids (VFAs) such as acetic acid and butyric acid are intermediate metabolites in the process of anaerobic digestion, which have increasingly come into focus as value-added end products of the process, as their usual chemical synthesis from fossil fuel sources is increasingly being scrutinised (Wainaina *et al.*, 2019). VFAs have a variety of applications in the chemical industry. For example, they can be used in the synthesis of polyhydroxyalkanoates (PHA), which are polyesters that are used in the production of biodegradable plastics (Cai *et al.*, 2009). Zhang *et al.* (2009) demonstrated that the conversion of PHA to R-3-hydroxybutyrate methyl ester and hydroxyalkanoate methyl ester could yield biofuels with combustion heats similar to ethanol. Kleerebezem *et al.* (2015) argued that the valorisation of VFAs into PHA could yield significantly higher revenue compared to methane

production, as the price per kilo of PHA is 5 times higher. Other possible uses of VFAs include the production of butanol from butyric acid, which is re-emerging as a potential replacement for petrol due to its desirable characteristics such as being compatible with modern internal combustion engines without modification (Kumar *et al.*, 2012). Butyric acid can also be used to produce other potential fuels such as ethyl butyrate and butyl butyrate, which can be used as gasoline additive and aviation fuel, respectively (Olson *et al.*, 2003; Jenkins *et al.*, 2013). Furthermore, butyric acid has uses in the pharmaceutical industry. Anti-cancer effects of its derivative pivalyoxmethyl butyrate have been shown to prolong survival in a mouse leukaemia model (Kasukabe, Rephaeli and Honma, 1997), and it also serves as precursor for prodrugs used in the treatment of alopecia and sickle cell anemia (Rephaeli, Zhuk and Nudelman, 2000). In the chemical industry, it can be used to manufacture cellulose acetate butyrate plastics and the food and cosmetic industry uses butyric acid esters as fragrance and flavouring agents (Dwidar *et al.*, 2012).

Overall, these examples show how anaerobic digestion is a very versatile process to produce value-added goods and fits into the International Energy Agency's definition of "biorefinery", which is the sustainable processing of biomass into a range of marketable products and energy (IEA, 2012).

1.4 Blue Sky Bio Ltd.

This project's industrial partner is a start-up company called Blue Sky Bio Ltd., which was founded in 2013 and is based at Thornton Science Park at the University of Chester, United Kingdom. As part of an InnovateUK Knowledge Transfer Partnership with the University of Birmingham, the company developed a concept to process sustainable second-generation and third-generation feedstocks (e.g. grass, seaweed) as well as food industry and agricultural waste products (e.g. orange juice waste, olive oil mill effluent, oil palm husk) using anaerobic digestion. This includes an undisclosed pre-hydrolysis and solubilisation step before feedstocks are digested using an

anaerobic baffled reactor. By the end of 2016, the company completed trials using a four-chamber ABR at bench-scale (i.e. 6 litre volume) and commissioned the construction of a seven-stage ABR holding 500 litres (Fig. 2). Operation of this reactor began in January 2017 after wastewater treatment plant sludge was introduced to the reactor as a seed inoculum. From January to July 2017, the reactor digested bladderwrack seaweed (*Fucus vesiculosus*). After this, the feedstock was seamlessly switched to grass in the form of lawn clippings that were sourced on site. During the reactor's operation, solubilised feedstock with volatile solid concentrations between 10–20 g/L was constantly introduced to the reactor using a peristaltic pump, with a hydraulic retention time of 7 days.

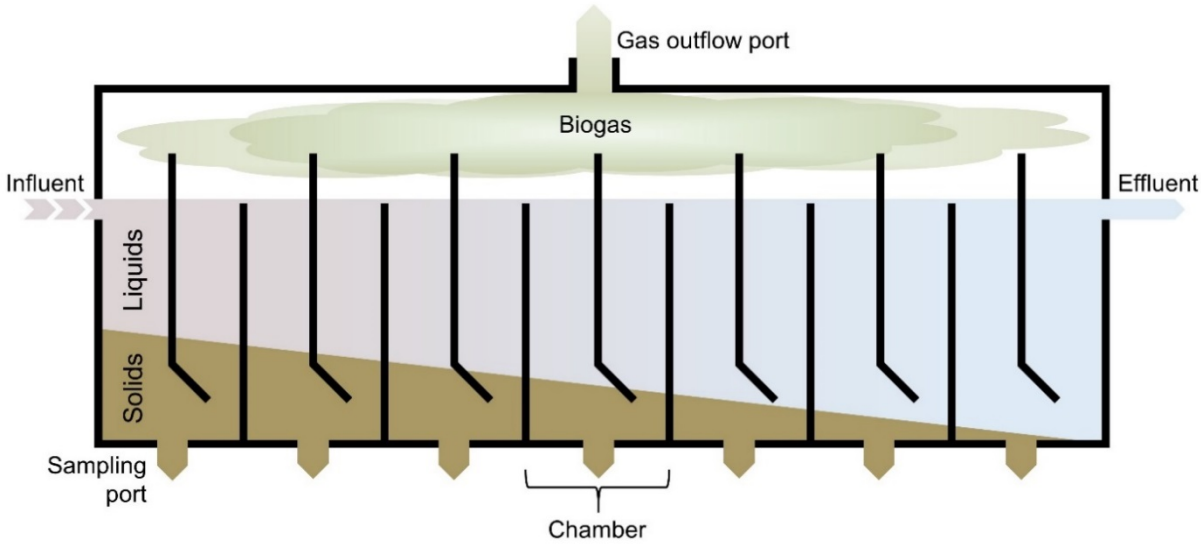


Figure 2: Simplified schematic of the Blue Sky Bio 500 litre Anaerobic Baffled Reactor. As feedstock passes through the reactor, its solid content is reduced as it is digested to biogas and other fermentation products (e.g. VFAs, alcohols, H_2, CO_2).

1.5 Aims and objectives

The primary aim of this project was to characterise the microbial community in the industrial partner's ABR using a metataxonomic approach. The hypothesis for this analysis was that the baffled design of the reactor would be reflected through changes in the microbial community along the seven different chambers. For this, the company provided several sets of digestate samples taken throughout 2017 (see chapter 2.1).

Since much of the available literature on ABRs has focused on small-scale reactors digesting waste products, characterising the microbial community of this seven-chamber 500 litre reactor has the potential to offer new insights into the anaerobic digestion of lignocellulosic second and third-generation feedstocks using anaerobic baffled reactors.

The secondary aim of this project was to find novel species with potentially useful characteristics for industrial applications, such as the production of volatile fatty acids, which fits with the industrial partner's ambitions to develop a "biorefinery" to produce value-added chemicals. Here, the objective was to first identify non-model organisms through the metataxonomics analysis before isolating and characterising them.

2 Materials and Methods

2.1 Digestate samples

Digestate was sampled by Blue Sky Bio Ltd. (Chester, UK) from the company's 500 litre 7-stage ABR. Sampling took place between February to June 2017 (seaweed digestate) and November to December 2017 (grass digestate) (Table 1). During this time, the bioreactor was operated with a hydraulic retention time of 7 days. For each date, one sample per chamber was taken (77 samples in total). The collected digestate was stored at −20°C at the company site until it was shipped to Edinburgh for DNA extractions in January 2018.

Table 2. Summary of digestate samples taken by Blue Sky Bio Ltd. (Chester, UK). On each date, samples were taken from each of the seven chambers of the company's Anaerobic Baffled Reactor (ABR).

Sampling date	Feedstock	Number of samples
16th February 2017	Seaweed	7
15th March 2017	Seaweed	7
7th April 2017	Seaweed	7
1st June 2017	Seaweed	7
1st November 2017	Grass	7
8th November 2017	Grass	7
15th November 2017	Grass	7
22th November 2017	Grass	7
29th November 2017	Grass	7
1st December 2017	Grass	7
7th December 2017	Grass	7
Total number of samples		**77**

2.2 Extraction, gel purification and quantification of DNA

DNA was extracted from digestate using the PowerSoil kit (Qiagen, UK). DNA extracts that displayed a faint green tint, as well as reduced or lacking amplification when used as template in PCR reactions, underwent the inhibitor removal steps of the PowerSoil kit for a second time. The amount of double stranded DNA in samples was quantified in Corning Falcon Black 96 Well Tissue Culture Imaging Plates, using the Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher, UK) on a SpectraMax Gemini XS plate reader (Molecular Devices, UK). Extraction and purification of DNA fragments from agarose gels was carried out using the Wizard SV Gel and PCR Clean-Up System (Promega, UK).

2.3 Amplification of 16S ribosomal RNA (rRNA) genes for metataxonomic analysis and bacterial identification

To perform the metataxonomic analysis of digestate samples, 5 ng of extracted genomic DNA from each digestate sample was used in a PCR reaction (Table 4) to amplify the hypervariable region V4 of the 16S rRNA gene. The a515F-806R primer pair (Table 3) was adapted from Caporaso *et al.* (2011), including degeneracies to remove known biases against Crenarchaeota/Thaumarchaeota (Parada, Needham and Fuhrman, 2016) and the marine and freshwater Alphaproteobacterial clade SAR11 (Apprill *et al.*, 2015). Primers also contained appropriate adapters as well as barcoded sequences for Illumina sequencing, as described in Caporaso *et al.* (2012) (Table 3). The PCR products with an average size of around 400 bp were excised from agarose gels, purified and pooled. The resultant sample pool with a DNA concentration of 6.83 ng/μL was submitted to Edinburgh Genomics for quality control, library preparation and sequencing using an Illumina MiSeq v2 with 2 × 250 bp paired-end reads.

For the identification of isolated bacteria using Sanger sequencing, the BactF-BactR primer pair (Fernández *et al.*, 1999) was used to first amplify the complete 16S rRNA gene in a colony PCR, before the resulting product (1532

bp) was purified from a gel and sent for sequencing with the a515F-806R primer pair as sequencing primers to sequence the 16S rRNA V4 region. Alternatively, if colony PCR with BactF-BactR primers did not result in a product (i.e. primers not binding due to their lack of degeneracy and sequence variability), the a515F-806R primer pair was used to generate a product (254 bp) for sequencing instead. For colony PCR, a loopful of bacterial colony was resuspended in 10 µL water and heated to 95°C for 10 minutes before 1 µL of the mixture was used as template in the PCR reaction. Sanger sequencing was performed by MRC PPU DNA Sequencing and Services (Dundee, UK). Two reactions per primer were carried out and the resulting sequences were aligned using MEGA X Version 10.0.2 (Kumar *et al.*, 2018). Concatenated sequences were queried against the NCBI BLAST nucleotide collection as well as the SILVA database version 132 (Quast *et al.*, 2012).

Table 3. Sequences of primers used in qPCR, PCR and Sanger sequencing.

Name	**Sequence**
a515F	GTGYCAGCMGCCGCGGTAA
806R	GGACTACHVGGGTWTCTAAT
IL-a515F	**AATGATACGGCGACCACCGAGATCTACAC**TATGGTAATTGTGTGYCAGCMGCCGCGGTAA
IL-806R	**CAAGCAGAAGACGGCATACGAGAT**TCCCTTGTCTCCAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
BactF	AGAGTTTGATCCTGGCTCAG
BactR	AAGGAGGTGATCCAGCCGCA

Nucleotide redundancies are described following the IUPAC nucleotide code.

Features of the Illumina primers are shown in different colours:
Illumina adapter (**red**), Barcode (**purple**), Pad (**blue**), Linker (**green**)

Table 4. PCR protocol to amplify the 16S rRNA V4 region from digestate for Illumina sequencing or Sanger sequencing using primers without the Illumina primer features.

Master Mix (25 µL reaction)		**Cycling conditions**		
Reagent	**µL**	**Temperature**		**Time (mm:ss)**
NEB *Taq* DNA polymerase (5 U/µL)	0.25	94°C		3:00
NEB Standard *Taq* Buffer (10x)	2.5	32 cycles	94°C	0:45
Nuclease-free Water	18.5		50°C	1:00
$MgCl_2$ (25 mM)	1.0		72°C	1:30
dNTP mix (10 mM)	0.5	72°C		10:00
(IL-)a515F forward primer (10 µM)	0.625	4°C		∞
(IL-)806R reverse primer (10 µM)	0.625			
DNA template	1			

Table 5. PCR protocol to amplify the 16S rRNA gene from cultured bacteria for Sanger sequencing.

Master Mix (50 µL reaction)		**Cycling conditions**		
Reagent	**µL**	**Temperature**		**Time (mm:ss)**
NEB *Taq* DNA polymerase (5 U/µL)	0.25	94°C		5:00
NEB Standard *Taq* Buffer (10x)	5	32 cycles	94°C	1:00
Nuclease-free Water	34.75		55°C	1:00
$MgCl_2$ (25 mM)	5		72°C	3:00
dNTP mix (10 mM)	1	72°C		7:00
BactF forward primer (20 µM)	1	4°C		∞
BactR reverse primer (20 µM)	1			
DNA template	2			

2.4 Metataxonomic analysis

The metataxonomic analysis of the sequencing data was carried out using the bioinformatics pipeline QIIME 2 Version 2018.4 (Bolyen *et al.*, 2019). The DADA2 algorithm (Callahan *et al.,* 2016) was used to correct Illumina-sequenced amplicon errors and remove chimaeras. On average, DADA2 retained 75% of all sequences (±7%). Due to the good quality of the raw sequence data throughout the reads, the data was neither trimmed nor truncated in further quality control steps.

For the Alpha and Beta diversity analyses, QIIME2 uses a rarefaction approach as diversity metrics are sensitive to different sampling depths across samples. In order to avoid samples being discarded from the diversity analysis, the number of sequences to be randomly subsampled was chosen to match the sample with the lowest number of counts (2,023 counts in digestate sample from 1st June 2017, Chamber 4). Alpha diversity as a function of sampling depth was explored using the QIIME diversity alpha-rarefaction visualiser. The value provided for the parameter max-depth was chosen based on the median frequency per sequence feature (i.e. 81,990). For the taxonomy bar plots, a classifier was trained using the QIIME-compatible 99% OTU reference database of SILVA Release 132 (April 2018). Reference reads were extracted by supplying the a515F forward and 806R reverse primers.

2.5 Culturing

Peptone-Yeast-Glucose medium (PYG) adjusted to pH 7 (Table 6) was filter sterilised (CytoOne 0.2 µm PES vacuum filter; Starlab, UK) and used to isolate organisms from digestate and grow all subsequent *Clostridium* cultures, unless otherwise mentioned. Medium pH was measured using a Mettler Toledo FiveEasy pH meter.

To isolate organisms from digestate, 1 mL of semi-solid digestate (from chamber 5, sampled on 15th March 2017) was briefly vortexed and allowed to settle for 30 seconds. Then, 50 µL of supernatant were transferred to a PYG agar plate and spread using a disposable spreader. Cultures were grown at

30°C in a Whitley A95 Anaerobic Workstation, under an atmosphere containing 10% hydrogen, 10% carbon dioxide and 80% nitrogen at 75% humidity. After colonies of a single type (Gram-positive, facultative anaerobe cocci) were seen to grow on the plates initially, the experiment was repeated using PYG with nutrients reduced by 75%. Low-nutrient media was used in an attempt to culture a more diverse consortium of microorganisms, as in other studies such as Connon and Giovannoni (2002).

Table 6. Ingredients for Peptone-Yeast-Glucose (PYG) medium.

Ingredient	**Amount added**	
	Normal medium	**Low-nutrient medium**
Peptone (Bacto)	10 g	2.5 g
Yeast Extract (Oxoid)	10 g	2.5 g
Glucose	10 g	1.25 g
Dipotassium phosphate	2 g	2 g
Water	975 mL	975 mL
Hydrochloric acid	variable (to pH 7)	variable (to pH 7)
Agar (for solid medium only)	10 g	10 g

For the assessment of organism morphology, colonies were picked and suspended in 50 µL phosphate buffered saline, of which 5 µL were placed on a microscope slide and covered with a cover slip. Microscopy was carried out using an Olympus CX41 Microscope and 100X objective. Images were taken through the 10X eyepiece using a LG G6 Smartphone. The software ImageJ was used to apply scale bars to the images and measure the size of the microorganisms.

For the generation of frozen stocks, colonies were picked into 25 mL PYG medium in 50 mL centrifuge tubes. Cultures were grown for 1–3 days, until cells had settled into visible pellets at the bottom of the tubes. Cultures were centrifuged in a Sigma 3-16KL centrifuge at 3850 × *g* for 10 minutes. Pellets were re-suspended in 10 mL 50% glycerol, distributed as 1 mL aliquots and snap frozen in liquid nitrogen before storage at −70°C.

2.6 Antibiotics testing

To choose a suitable antibiotic to select for *Clostridium* mutants and to select against *E. coli* S17-1 conjugation donors, antibiotic resistance was assessed by spreading 50 µL of frozen stocks on PYG agar containing a range of antibiotics at various concentrations (Table 7).

Table 7. Antibiotics used to test antibiotic resistance of *Clostridium* and *E. coli* S17-1*

Ingredient	**Tested concentrations in PYG agar**			
Ampicillin	1 µg/mL	10 µg/mL	100 µg/mL	
Chloramphenicol	1 µg/mL	10 µg/mL	100 µg/mL	
Kanamycin	1 µg/mL	10 µg/mL	100 µg/mL	
Spectinomycin	1 µg/mL	10 µg/mL	100 µg/mL	
Tetracycline	0.1 µg/mL	1 µg/mL	10 µg/mL	
Thiamphenicol	1 µg/mL	15 µg/mL		
Erythromycin	10 µg/mL	50 µg/mL		
Cycloserine	25 µg/mL	50 µg/mL	100 µg/mL	300 µg/mL

* *E. coli* S17-1 was only tested for cycloserine resistance

2.7 Bacterial genome sequencing

Genomic DNA from the two isolated *Clostridium* species was extracted with the GenElute Bacterial Genomic DNA Kit (Sigma Aldrich, UK) using the manufacturer's protocol with the added lysozyme treatment for Gram-positive bacteria. Samples were submitted to MicrobesNG (University of Birmingham, UK) for Illumina sequencing (2 × 250 bp paired-end reads with 30x minimum coverage). The bioinformatical analysis including assembly, annotation and variant calling was also carried out by the facility.

2.8 Carbon source utilisation assays

This assay was designed to test whether a range of carbon sources can be utilised by the isolated *Clostridium* species to grow and produce acids. For this, a synthetic complete medium (Table 8) was supplemented with a range of carbon sources (Table 9) at 5 g/L. The medium contained bromocresol purple (260 mg/L), which is a pH indicator that is purple above pH 6.8 and yellow below pH 5.2.

The supplemented medium was distributed on a clear 96-well plate in 200 µL aliquots, which were inoculated with the *Clostridium* isolates. For this, frozen glycerol stocks were centrifuged at 16,000 × *g* and washed twice with water to remove any residual glycerol and medium. The cell suspension with an OD_{600} of around 1 was then used to aliquot 20 µL into the wells containing synthetic complete medium with selected carbon sources. All cultures were set up in triplicates and plates were incubated in the anaerobic workstation at 30°C for 4 days. Plates were analysed using a Multiskan Go plate reader (Thermo Scientific, UK) to measure transmittance between 500 – 650 nm. The transmittance in this spectrum differs depending on whether the indicator is purple or yellow. Therefore, these measurements served as a quantitative way to indicate carbon source utilisation, instead of relying on subjectively assessing the visible colour change in the wells, which was recorded using a flatbed scanner.

To test a wider range of carbon sources, two Phenotype MicroArrays PM1 MicroPlates (Biolog, USA) were used to test the ability of both *Clostridium* isolates to utilise a total of 95 compounds. After these were inoculated the same way as the pH indicator plates, growth was confirmed visually and by measuring OD_{600} using the plate reader and comparing to the control well.

Table 8. Ingredients for 100 mL synthetic complete medium (10x) without carbon source.

Ingredient	**Amount added**
Yeast Nitrogen Base Without Amino Acids (Sigma-Aldrich, UK)	6.7 g
Complete Supplement Mixture (Formedium, UK)	790 mg
Peptone (Oxoid, USA)	5 g
Bromocresol purple (Sigma-Aldrich, UK), 0.04% solution in water	35 ml
Water	65 ml

Table 9. Ingredients for 10 mL synthetic complete medium (1x) with carbon source.

Ingredient	**Amount added**
Synthetic complete medium (10x) without carbon source	1000 µL
Water	8950 µL
1M NaOH (to adjust to pH 7)	50 µL
Carbon source	50 mg
Arabinose *Galactose* *Glucose* *Glycerol* *Lactate* *Lactose* *Maltose* *Manitol* *Sorbitol* *Sucrose* *Trehalose* *Xylose*	

2.9 Gas volume

To assess the amount of gas produced by *Clostridium* cultures, frozen glycerol stocks were washed and resuspended in 25 mL PYG medium (containing 15 µg/mL chloramphenicol if mutants were grown) to an OD_{600} of 0.1. From these suspensions, six replicate 3 mL aliquots (unless otherwise mentioned) were taken into 10 mL syringes (Becton Dickinson, UK) and any gas bubbles left in the syringe were pushed out. Then, the syringes were attached to non-autoclavable Luer-Lok™ stopcocks (Promega, UK) via Millex 0.22 µm (Millipore, UK) filters to ensure sterile conditions. The syringes were suspended from custom-made 3D printed hooks in the anaerobic cabinet (Fig. 3) and photos were taken automatically in 1-hour intervals using a Canon PowerShot S95 camera. As the syringe plunger was pushed upwards due to increased gas pressure, the amount of gas produced was recorded from the photos. After gas production ceased, the optical density (OD_{600}) of syringe cultures was determined using a WPA Biowave CO8000 Cell Density Meter (Biochrom, UK). Negative controls in this experiment were non-inoculated samples of PYG medium as well as wild-type *Clostridium* cultures in selective medium.

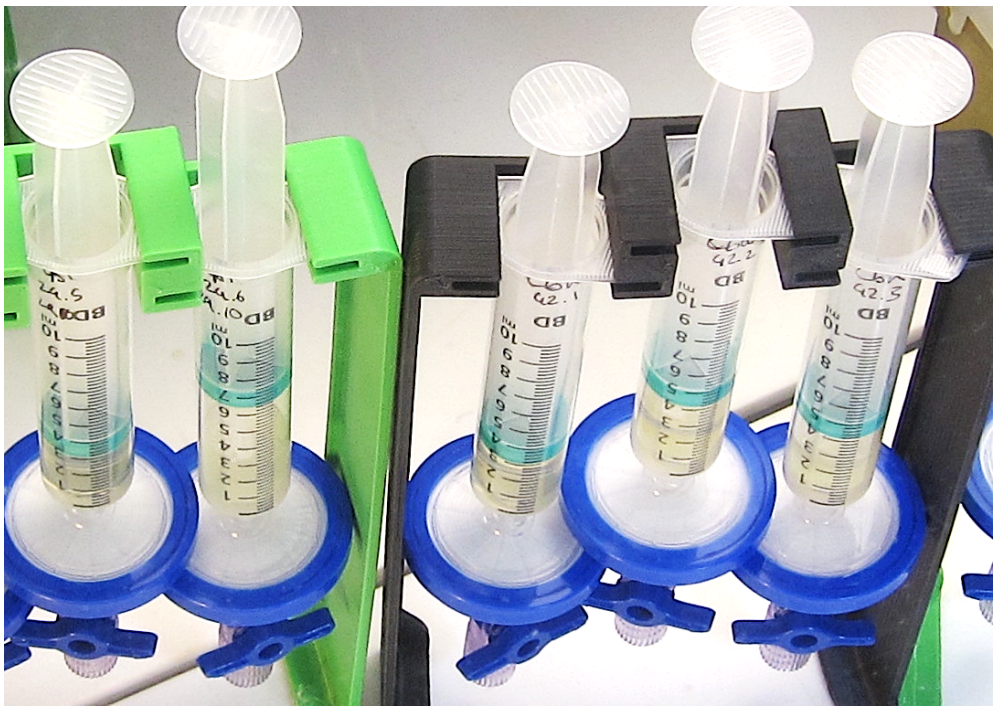


Figure 3. Syringe reactors containing gas-producing *Clostridium* cultures

2.10 Gas composition

The composition of gas produced by both *Clostridium* isolates was measured using a Rapidox 5100 Portable Multigas Analyser (Cambridge Sensotec, UK). As this device needs at least about 250–500 mL of gas to produce a reading, 500 mL cultures were grown in 500 mL Duran bottles, with 1 L Tedlar PFV gas bags (Environmental Sampling Supply, USA) connected through the Duran GL 45 Connection system and PVC tubing. To avoid introduction of hydrogen and carbon dioxide-containing anaerobic gas mix, the cultures were grown in a pure nitrogen atmosphere. The gas produced was analysed for relative concentrations of methane, carbon dioxide, carbon monoxide and hydrogen. Prior to taking measurements, the gas analyser was calibrated using a gas mixture containing 25% carbon dioxide, 20% carbon monoxide, 7% methane, 5% hydrogen and 3% ethane. Subsequently, the analyser was flushed with nitrogen gas before determining gas composition.

Only hydrogen and carbon dioxide could be detected in gas samples, however their concentration never added up to 100%. As hydrogen and carbon dioxide are the only gases known to be produced by *Clostridium* species (Lin *et al.*, 2007), it was therefore assumed that the remainder of the gas was the non-measurable nitrogen from the cultures' headspaces. The measured values of hydrogen and carbon dioxide concentration were adjusted accordingly to add up to 100%.

2.11 Volatile fatty acid analysis

To determine the quantity of VFAs produced by the *Clostridium* isolates, supernatants of 25 mL stationary cultures were analysed by gas chromatography with flame-ionisation detection (GC-FID) on a HP 5890 II (Hewlett Packard, UK) fitted with a PERMABOND FFAP 30 m × 0.32 mm column with 0.5 µm film thickness (Macherey-Nagel, UK). Data was collected and processed by Mr Stephen Mowbray, Lab Manager at the Organic Geochemistry Facility, School of Geosciences, University of Edinburgh. The VFA concentration in the samples was determined by comparing acid peak heights against those generated by a standard solution (Table 10) with known VFA concentrations (Table 11). The OD_{600} and pH of cultures was measured to determine a possible correlation with the measured VFA concentrations.

Table 10. Ingredients for VFA analysis standard solution (1 mL).

Ingredient	Amount (µL)
Volatile Free Acid Mix (Supelco, UK)	100
2-ethylbutyric acid, diluted in water (1:1000)	250
10% (w/v) phosphoric acid	100
Deionised water	550

Table 11. Final acid concentrations in the standard solution used in GC-FID analysis.

Acid	Concentration (mM)	Concentration (µg/mL)
Formic acid	10.0	46.03
Acetic acid	9.9	59.45
Proprionic acid	10.0	74.08
Isobutyric acid	10.1	88.99
Butyric acid	10.1	88.99
Isovaleric acid	10.1	103.15
Valeric acid	10.0	102.13
Isocaproic acid	10.2	118.48
Hexanoic acid	10.3	119.65
n-Heptanoic acid	10.2	132.78
2-ethylbutyric acid	-	227.7

2.12 pMTL80000 shuttle plasmids

The modular pMTL80000 *E. coli* – *Clostridium* shuttle plasmids designed by Heap *et al.* (2009) and purchased as a kit from CHAIN Biotechnology (Nottingham, UK) were used to genetically modify the two *Clostridium* species that were isolated from digestate samples. The nomenclature of the pMTL80000 plasmids is based on the prefix "pMTL8" and the combination of 18 exchangeable modules (Table 12). For example, the supplied pMTL83151 plasmid contains the pCB102 Gram-positive replicon from *C. butyricum*, the chloramphenicol marker gene *catP*, the Gram-negative ColE1 replicon with *tra* genes necessary for bacterial conjugation as well as a multiple cloning site (MCS) containing *lacZ* without a promoter.

Table 12. pMTL80000 plasmid kit standard modules and numbering.

Gram+ replicon	Marker	Gram− replicon	Application-specific module
0. Spacer	0. Spacer	0. Spacer	0. Spacer
2. pBP1	1. *catP* (Chloramphenicol)	1. p15a	1. MCS
3. pCB102	2. *ermB* (Erythromycin)	2. p15a + *tra*	2. P_{thl} + MCS
4. pCD6	3. *aad9* (Spectinomycin)	4. ColE1	3. P_{fdx} + MCS
5. pIM13	4. *tetA(P)* (Tetracycline)	5. ColE1 + *tra*	4. catP reporter

In order to generate plasmids containing the MCS as well as the thiolase promoter from *C. acetobutylicum* (P_{thl} / *Cac thl*) or ferredoxin promoter from *C. sporogenes* (P_{fdx} / *Csp fdx*), plasmids pMTL83151, pMTL83353 and pMTL84422 were digested with restriction enzymes *SbfI* and *AscI* (see Fig. 4). The backbone of pMTL83151 was then ligated with the application-specific modules from pMTL84422 and pMTL83353 to generate plasmids pMTL83152 and pMTL83153, respectively. The promoter regions in the pMTL80000 plasmids are flanked by restriction sites for *NotI* and *NdeI*, which were used to construct plasmids containing promoters from the two *Clostridium* isolates.

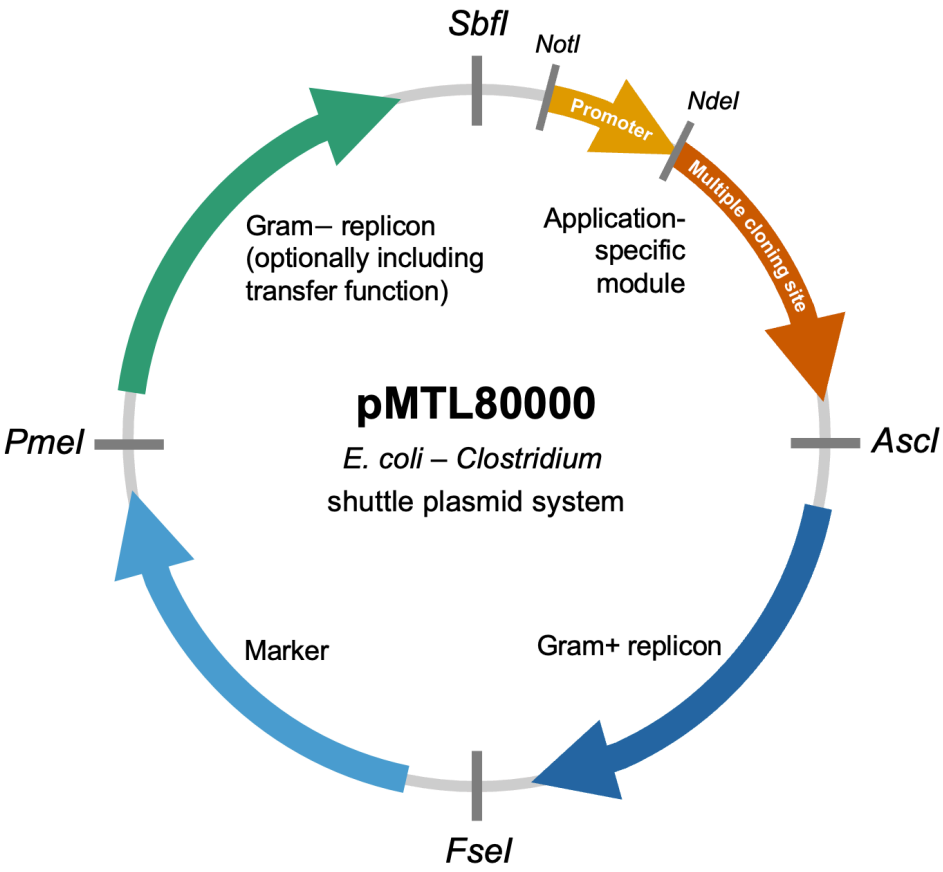


Figure 4. Modular *E. coli – Clostridium* shuttle plasmid system pMTL80000. Interchangable modules are flanked by four rare (8 bp) type II restriction enzyme recognition sites, while the promoter region in the Application-specific module is flanked by recognition sites for *NotI* and *NdeI*.

2.13 Construction of plasmids

For double digestions, 1 µg of either plasmid or insert DNA was incubated with the appropriate restriction enzymes (New England Biolabs, UK) and buffers according to the manufacturer's protocol for 1 hour at 37°C. Following plasmid digestion and separation by gel electrophoresis (1% agarose gel, 100 V for 1 hour), gel fragments were excised and the desired plasmid DNA was purified using the Wizard SV Gel and PCR Clean-Up System (Promega, UK) according to the manufacturer's protocol. Digested PCR products were cleaned up using the QIAquick PCR Purification Kit (QIAGEN, UK) according to the manufacturer's protocol. After purification, DNA fragments were ligated using T4 DNA Ligase (New England Biolabs, UK).

Ligation reactions were carried out overnight at 16°C in a Bio-Rad T100 Thermal Cycler. Afterwards, 5 µL of the ligation mixes were used to transform NEB 5-alpha Competent *E. coli* (High Efficiency) (New England Biolabs, UK) according to the NEB High Efficiency Transformation Protocol. Transformants were selected on LB agar containing 15 µg/mL chloramphenicol. For blue-white screening, 100 µM IPTG and 50 µg/mL X-gal were added to the LB agar as well. Plasmids were purified from transformants using the Monarch Plasmid Miniprep Kit (New England Biolabs, UK). In addition to blue-white screening, where applicable, diagnostic digests and PCR was used to screen for the presence of the inserts. For this, M13 UP Fwd/Rev sequencing primers designed by Dr Annegret Honsbein (Horsfall Group, University of Edinburgh) as well as Pro Fwd/Rev (designed to amplify the region between the *NotI* and *NdeI* sites, Table 13) were used for PCR. Furthermore, these primers as well as the M13 (GW) primers for Gateway vectors (supplied by the sequencing facility) were used to validate the sequence of all constructed plasmids using Sanger sequencing provided by MRC PPU DNA Sequencing and Services at the University of Dundee, UK.

Table 13. Primers used to validate plasmid sequences via Sanger sequencing.

Primer name	Sequence
M13 UP Fwd	CTCACATGTTCTTTCCTGCG
M13 UP Rev	ACTGTTGGGAAGGGCGATC
M13 Fwd (GW)	TGTAAAACGACGGCCAGT
M13 Rev (GW)	CAGGAAACAGCTATGACC
Pro Fwd	GAGCAAGGCAAGACCGATC
Pro Rev	AGCTCGAATTCGTAATCATGG

2.14 Transformation of *E. coli* S17-1

After constructing the plasmids in *E. coli* NEB 5-alpha, they were transformed into *E. coli* S17-1, which is a strain capable of bacterial conjugation. In order to make the *E. coli* S17-1 cells chemically competent, 5 mL overnight cultures of cells in LB media were diluted back into 50 mL media in a 200 mL conical flask. The diluted culture was then grown to OD_{600} 0.5 before it was split into two 50 mL Falcon tubes and incubated on ice for 10 minutes. Subsequently, the chilled culture was centrifuged for 10 minutes at 3850 × *g* and 4°C. Afterwards, the supernatant of the two 25 mL cultures was removed and replaced with 2.5 mL pre-chilled, filter-sterilised TSS buffer (5 g PEG 8000, 1.5 mL 1M $MgCL_2$, 2.5 mL DMSO, LB medium to 50 mL). The cultures were vortexed to fully resuspend the pellet and remove small cell aggregates. Then, 100 µL aliquots were distributed into pre-chilled microcentrifuge tubes and snap frozen in liquid nitrogen before storage at −70°C. The transformation protocol was identical to the commercial protocol by New England Biolabs for use with their *E. coli* NEB 5-alpha chemically competent cells. Single colonies of successful transformants were grown in LB medium with 15 µg/mL chloramphenicol overnight to generate frozen stocks in 25% glycerol, stored at −70°C.

2.15 Transconjugation of *Clostridium* isolates

For the transformation of *Clostridium* isolates, *E. coli* S17-1 mutants harbouring the pMTL80000 plasmids were used as conjugation donors. For this, cultures of the *E. coli* S17-1 donor as well as the *Clostridium* recipient were grown to stationary phase in LB (with 15 µg/mL chloramphenicol) and PYG medium, respectively. Initially, 1 mL of *E. coli* culture was centrifuged at 4000 × *g* for 2 minutes. To wash off the antibiotic, the supernatant was discarded and cells were resuspended very gently with 1 mL of PBS before the centrifugation step was repeated. The final pellet was resuspended and mixed gently with 200 µL of *Clostridium* recipient culture. The mixture of donor and recipient cells was pipetted drop-by-drop on PYG agar without antibiotics and incubated for 8 hours at 30°C under anaerobic conditions. To recover the transconjugants, all cells from the plate were then collected using an inoculation loop and re-suspended in 1 mL PBS, partially-open bottles of which were placed in the anaerobic hood 24 hours prior to use, in order to reduce the amount of dissolved oxygen in it. From this mixture, 50 µL were then spread on PYG plates containing 100 µg/mL cycloserine and 15 µg/mL chloramphenicol to select against the donor and non-recipients, respectively. The colonies that were grown on these plates were picked and spread again on fresh PYG plates with cycloserine and chloramphenicol to rule out any carry-over of donor *E. coli* cells. Finally, single colonies were used to grow liquid cultures in PYG with the two antibiotics. After these cultures were grown to an OD_{600} larger than 1, they were aliquoted and mixed in equal amounts with 50% glycerol to set up frozen stocks in 25% glycerol for storage at −80°C.

2.16 Selecting promoters to be tested using the beta-glucuronidase (GUS) assay

As the pMTL80000 kit contains the *Cac thl* and *Csp fdx* promoters, the aim was to construct plasmids that contain the endogenous thiolase and ferredoxin promoters from the two *Clostridium* isolates (*C. butyricum* and *C. pluributyricum*). For this, a BLAST search (Madden, 2013) of the *Cac thl* and *Csp fdx* promoters was carried out against the genomes of *C. acetobutylicum* and *C. sporogenes,* respectively. The *Cac thl* promoter was found in the genome of *C. acetobutylicum* strain LJ4, upstream of a gene described as acetyl-CoA C-acetyltransferase (GenBank CP030018.1, Location 263,285 to 264,463), which is also known as thiolase (Stim-Herndon *et al.*, 1995). Likewise, the *Csp fdx* promoter was found in the genome of *C. sporogenes* strain DSM 795, upstream of the ferredoxin gene (GenBank CP011663.1, Location 86,476 to 86,646). In order to find homologues of both genes in *C. butyricum* and *C. pluributyricum*, their sequences were compared to the genomes of both isolates. For *C. acetobutylicum* thiolase, matching homologues were found in *C. butyricum* (96% query cover, 77.09% identity) and *C. pluributyricum* (79% query cover, 77.85% identity). Likewise, matches for *C. sporogenes* ferredoxin were also found in *C. butyricum* (92% query cover, 85.09% identity) and *C. pluributyricum* (92% query cover, 83.02% identity). The BPROM software (Solovyev and Salamov, 2011) was used to ensure that the upstream regions of these genes, between their start codon and the stop codon of the next gene, contained putative promoters (Table 14). Primers with restriction modifications for *NotI* and *NdeI* were designed (Table 15) to amplify their upstream regions from *C. butyricum* and *C. pluributyricum* genomic DNA using PCR (Table 17). Furthermore, the inducible XylS-PM promoter from *Pseudomonas putida* (Gawin, Valla and Brautaset, 2017) was cloned from pSEVA238 (Martínez-Garćía *et al.*, 2015) for insertion into the pMTL8315x backbone.

Using the sequence of *E. coli* DH5α (Chen *et al.*, 2018), primers were designed to clone the beta-glucuronidase (*gusA*) from *E. coli* NEB 5-alpha competent cells (New England Biolabs, UK) for insertion into the pMTL80000

plasmids. For the construction of plasmids with the endogenous *Clostridium* promoters, the backbone of the pMTL83153 plasmid was used after excising the *Cac fdx* promoter.

Table 14. *Clostridium* spp. ferredoxin and thiolase upstream regions with *NotI* and *NdeI* restriction modifications for insertion into pMTL80000 plasmids.

Promoter	Sequence
Cac thl	gcggccgcTTTTTAACAAAATATA**TTGATA**AAAATAATAATAGT**GGGTATA** **AT**TAAGTTGTTAGAGAAAACGTATAAATTAGGGATAAACTATGGAACTTAT GAAATAGATTGAAATGGTTTATCTGTTACCCCGTATCAAAATTTAGGAGGT TAGTTcatatg
Csp fdx	gcggccgcGTGTAGTAGCCTGTGAAATAAGTAAGGAAAAAAAAGAAGTAAG TGTTATATATGATGATTATTTTGTAGATGTAGATAGGATAATAGAATCCAT AGAAAATATAGGTTATACAGTTATATAAAAATTAC**TTTAAA**AATTAATAAA AACA**TGGTAAAAT**ATAAATCGTATAAAGTTGTGTAATTTTTAAGGAGGTGT GTTAcatatg
Cbu thl	gcggccgcAGATGAATTTAAAAGATAATTAATTTTATAAATATAAAGCAGA AATTTACGCATAATGTAATTTCTGCTTTATTTATTTTCTAAAATTATATAC TTTTGATAAAAAATAAAACTTTGTCTAAAATTTAACAATAAACTATTGATA AAAAAGTTAAAAATGGTATAATTATCTTATGTTAATAAT**TTAACG**AAAAGT AAATAAAAAACAT**AATTATAAT**TTTAAAATAATTTATGATATTTTTAGGAG GTAAATTcatatg
Cbu fdx	gcggccgcATAATAATTATAAATA**TTGATT**AAATTTTGGAAATCATA**ATAT** **AAAAT**ACTTTAAAAACATTATGATACATGATAAAATGTAAGAGGTTAATTA CTACAATTTATTAGGAGGTGTTCTGAAcatatg
Cpl thl	gcggccgcTGAGCAGTCCTCCTCTTGAAATAATCTATATTAAAATTCTATA AAGACATTGTCACAATTTGGTTACAAAAGTTAGGTAAATATGTAAGAATTA AATTTCGGAAAATTAAGCGTC**TTAAAG**AAAATAGCAAATTCAAAA**TGTTAT** **AAT**TATTTTTTATTAAGTCAGGAAACAGATTTATCTTATAGTTAAGTCTAT TGGTAAGGTATTCTATGATAAGCGTGCCATTTTTTTATTGTGTCTTAATTT AGATAATAGTTAGACATGGTAATAAGAATGAAAAATGTATTATTGATACAG TCATGAATTATTATAATATAAAAATACATTAAAGCATAATTAAAAGTAATA GGTTATCATACACAATTCATTTTATATATTATTTTTTTATGTGTCTAAATC GATAAAGTACGAAGAAAGGTTATTTATAATACATCTTTTGGTCTATGAAAT ATGTGCTTAATACAATTTTGTAGTGAAAATGAA**TTTCAA**GATAACGTTTTT **TATTAAAAT**AAAGAAAAAATAATTATTGAAATTAAAATAATTAGATAAAAC TTACTAGAGAAAACGAGCATTTTCAGGATTTTAAATAATTGTATAAGAAGA AATAGCATGTTATTAAAATAATTGGCATATTAATTGCTTTATATAATTCTA AAAGGATATAAATAAGATCAAATTATAATTAAAAATACCTTTATAATGTGT CTCAAATAAATATATGAAAGAGGGAGAAAATTTTTcatatg
Cpl fdx	gcggccgcCAATACTCTGTAATGGCATTATTTCTTTGCATTTCAAATAAAA TGAAATAATAAATATTATTAAAAAATAACTCCTATAGAATATAAAACAC**TT** **TAAA**AACATTAGAATACA**TGTTAAAAT**GTAAGTGGTTAATTACCGCAATTT ATTAGGAGGTGTTCTATcatatg

Putative -10 (**BLUE**) and -35 (**RED**) boxes as predicted by BPROM.
Restriction sites for *NotI* and *NdeI* shown in lower case.

Table 15. Primers with appropriate restriction modifications used to clone *Clostridium* endogenous promoters, the XylS-PM inducible promoter and the GUS reporter for plasmid construction.

Primer name	Sequence
Cbu thl F (NotI)	CATGCGGCCGCAGATGAATTTAAAAGATAATTAATTTTATAAATATAAAGC
Cbu thl R (NdeI)	GGGAATTCCATATGAATTTACCTCCTAAAAATATCATAAATTATT
Cbu fdx F (NotI)	CATGCGGCCGCATAATAATTATAAATATTGATTAAATTTTGG
Cbu fdx R (NdeI)	CATCATATGTTCAGAACACCTCCTAATAAATTG
Cpl thl F (NotI)	CATGCGGCCGCTGAGCAGTCCTCCTCTTG
Cpl thl R (NdeI)	CATCATATGAAAAATTTTCTCCCTCTTTCATATATTTATTTGAGACAC
Cpl fdx F (NotI)	CATGCGGCCGCCAATACTCTGTAATG
Cpl fdx R (NdeI)	CATCATATGATAGAACACCTCCTAATAAATTGCGG
XylS-PM F (NotI)	CATGCGGCCGCTCAAGCCACTTCCTTTTTG
XylS-PM R (NdeI)	CTTCACCTTTAGAAACCATATGCTTTTTCTCCTCTTTCCCG
GusA F (NdeI)	GGGAATTCCATATGTTACGTCCTGTAGAAACCC
GusA R (HindIII)	CCAAGCTTCAGGAGAGTTGTTGATTCATTGTTTG

Table 16. Plasmids constructed to test promoter strength with the GUS reporter system.

Plasmid name	Size	Description
pMTL83152	4620 bp	pMTL80000 *E. coli – Clostridium* shuttle plasmid with *C. butyricum* replicon (pCB102), chloramphenicol resistance gene, *E. coli* replicon ColE1, conjugation gene *tra* and *Cac thl* promoter.
pMTL83153	4670 bp	As above but with *Csp fdx* promoter.
pMTL8315A	4724 bp	As above but with *Cbu thl* promoter.
pMTL8315B	4591 bp	As above but with *Cbu fdx* promoter.
pMTL8315C	5211 bp	As above but with *Cpl thl* promoter.
pMTL8315D	4632 bp	As above but with *Cpl fdx* promoter.
pMTL8315X	6450 bp	As above but with *XylS-PM* promoter.
pMTL83152-gusA	6356 bp	As pMTL83152 but with GUS reporter in the MCS.
pMTL83153-gusA	6406 bp	As pMTL83153 but with GUS reporter in the MCS.
pMTL8315A-gusA	6460 bp	As pMTL8315A but with GUS reporter in the MCS.
pMTL8315B-gusA	6327 bp	As pMTL8315B but with GUS reporter in the MCS.
pMTL8315C-gusA	6947 bp	As pMTL8315C but with GUS reporter in the MCS.
pMTL8315D-gusA	6368 bp	As pMTL8315D but with GUS reporter in the MCS.
pMTL8315X-gusA	8168 bp	As pMTL8315X but with GUS reporter in the MCS.

Table 17. PCR protocol to amplify genes from genomic DNA for cloning into pMTL80000.

Master Mix (50 µL reaction)		Cycling conditions		
Reagent	**µL**	**Temperature**		**Time (mm:ss)**
NEB *Q5* High-Fidelity DNA polymerase	0.5	98°C		0:30
NEB Q5 Reaction Buffer (5x)	10	32 cycles	98°C	0:10
Nuclease-free Water	31.5		60°C	0:30
dNTP mix (10 mM)	1		72°C	0:30 / kb
Forward primer (10 µM)	2.5	72°C		2:00
Reverse primer (10 µM)	2.5	4°C		∞
DNA template	2			

2.17 GUS reporter assay

Strains of *Clostridium* mutants transformed with pMTL80000 plasmids (Table 16) were grown in triplicate cultures of 10 mL in 50 mL falcon tubes. After the cultures reached OD_{600} of 1, aliquots (1.5 mL) were taken into microcentrifuge tubes and pelleted using centrifugation. After the supernatant was removed, pellets were frozen at −70°C and thawed before carrying out the assay. Pellets were resuspended in 800 µL of buffer Z (60 mM $Na_2HPO_4 \cdot 7 H_2O$, 40 mM $NaH_2PO_4 \cdot H_2O$, 10mM KCl, 1 mM $MgSO_4 \cdot 7 H_2O$, pH adjusted to 7.0, and 50 mM freshly added 2-mercaptoethanol). In order to normalise GUS absorbance values against OD_{600}, 200 µL aliquots of resuspended pellet were taken aside into the well of a Falcon Black 96 Well Tissue Culture Imaging Plate (Corning, UK). To the remaining 600 µL samples, 6 µL toluene was added before the samples were vortexed for 1 minute, followed by a 10 min incubation on ice before a 30 min pre-incubation at 37°C with the lids open. After that, 120 µL buffer Z containing 6 mM p-nitrophenyl-β-glucuronide (PNPG) was added to each sample, which started the enzymatic reaction. Samples were incubated for 5 minutes at 37°C before the reactions were stopped via the addition of 300 µL 1M Na_2CO_3. Subsequently, cells were removed by centrifugation at 10000 × *g* for 10 min. To measure the absorbance of samples at 405 nm, 200 µL from each sample was transferred

to the 96-well plates mentioned above. Absorbance (A_{405}) and optical density (OD_{600}) was measured using a Multiskan GO microplate reader (Thermo Scientific, UK). Relative β-glucuronidase activity was calculated by dividing the A_{405} value by the sample OD_{600}.

For strains that were transformed with the pMTL8315X(-gusA) plasmids in order to test the inducible *Pseudomonas putida* XylS promoter (Zwick, Lale and Valla, 2013), cultures were grown in medium containing the m-tuloic acid inducer at different concentrations (0 M, 1 µM, 10 µM, 100 µM, 1 mM).

2.18 Cloning of *Clostridium* genes into pMTL80000

Genomes of both *Clostridium* species were uploaded to the Pathosystems Resource Integration Center (PATRIC, Davis *et al.*, 2020). The Genome View function was used to find genes associated with the butanoate metabolism pathway (KEGG: map00650, Kanehisa and Goto, 2000), specifically hydrogenase (*hydA*) and butyrate kinase (*buk*) genes. Primers with appropriate restriction modification sites and an optional His-tag sequence in the reverse primer were generated for four hydrogenases and three butyrate kinases using Snapgene (Table 18). The nomenclature of the primers and resulting plasmids is based on the species the gene originated from (i.e. Cbu for *C. butyricum,* Cpl for *C. pluributyricum*) and the contig number within the full genome sequence. The format used is species-contig-gene (e.g. Cpl-c67-hydA) followed by the modifications for the optional His-tag and restriction site (e.g. His-BamHI). For the forward primers, restriction sites for *NdeI* were implemented, as the enzyme cuts right at the start codon and therefore allows the gene to be inserted as closely to the promoter as possible. Genes Cbu-c16-hydA and Cpl-c42-buk however both have *NdeI* recognition sites within their coding sequence. To circumvent this incompatibility, recognition sites for the Type IIS restriction enzyme *FauI* were designed to generate an overhang compatible with plasmids cut by *NdeI*. The PCR protocol used to amplify genes from *Clostridium* genomic DNA is described in Table 17. Constructed plasmids are listed in Table 19.

Table 18. Primers with appropriate restriction modifications used to clone *Clostridium* hydrogenase (*hydA*) and butyrate kinase (*buk*) genes into plasmid pMTL8315B.

Primer name	Sequence
Cbu-c2-hydA-F-NdeI	GGGAATTCCATATGAATAGTAAATATACTGATTTATTTCATGAAC
Cbu-c2-hydA-R-BamHI	CGGGATCCGGTTGAAGTTGTTTTTAGTTTAAC
Cbu-c2-hydA-R-His-BamHI	CGGGATCCTCAGTGGTGATGGTGATGATGTTTTTTAACATTATTTGAATTAAAATC
Cbu-c16-hydA-F-FauI	CCTATATACCCGCTTTTTATGTTTCAATTTGAAAATC
Cbu-c16-hydA-R-XbaI	GACTCTAGACTTTTATTTTGCTAAATCAC
Cbu-c16-hydA-R-His-XbaI	GACTCTAGATCAGTGGTGATGGTGATGATGCTTCTTTAAGCTTATATTGTC
Cbu-c24-hydA-F-NdeI	GGGAATTCCATATGATAAATTTAATTATTGATGAAAAAAATGTTC
Cbu-c24-hydA-R-BamHI	CGCGGATCCTAATTATTTAGTATATTTTAAGTGTAATAATTCATGG
Cbu-c24-hydA-R-His-XbaI	GACTCTAGATTAGTGGTGATGGTGATGATGTTTAGTATATTTTAAGTGTAATAATTC
Cpl-c67-hydA-F-NdeI	GGGAATTCCATATGTTTCAATTTGAAAATCAACTTTTAAAG
Cpl-c67-hydA-R-BamHI	CGCGGATCCTAAGTCCTATACTTATCTTTC
Cpl-c67-hydA-R-His-BamHI	CGGGATCCTTAGTGGTGATGGTGATGATGTCTTTCTAAATTTATATTGTGG
Cpl-c3-buk-F-NdeI	GGTAAATTCATATGAATTACAAAGTGTTAGCTATAAATC
Cpl-c3-buk-R-XbaI	CGTCTAGAGCTTTTTAAACTTTATATACAAACAAATTTATC
Cpl-c3-buk-R-His-XbaI	GACTCTAGATTAGTGGTGATGGTGATGATGTCTTTCTAAATATTCTTTAGC
Cbu-c15-buk-F-NdeI	GGAGGTAAATTCATATGGCTTACAAATTATTAATAATTAATCCAGG
Cbu-c15-buk-R-XbaI	CGTCTAGAGTATGTATATAAAATTAAAATTAATAAACTTTAGCC
Cbu-c15-buk-R-His-XbaI	GACTCTAGATTAGTGGTGATGGTGATGATGATAAACTTTAGCCTTTTCTTC
Cpl-c42-buk-F-FauI	CCTATATACCCGCCATCTATGTCATATAAGCTATTAATAATTAATCCAG
Cpl-c42-buk-R-XbaI	CGTCTAGAGTTCCTAATTGCTTTAATAAATTCTAG
Cpl-c42-buk-R-His-XbaI	GACTCTAGATTAGTGGTGATGGTGATGATGATAAATTCTAGCTTGTTCTTC

Table 19. Plasmids constructed to express butanoate pathway genes in *Clostridium*.

Plasmid name	Size	Description
pMTL8315B Cbu-c2-hydA	5951 bp	pMTL8315B (see Table X) with [FeFe] Hydrogenase (EC 1.12.7.2) from *C. butyricum* (1371 bp, 456 aa).
pMTL8315B Cbu-c2-hydA-His	5945 bp	As above, with added 6-His-tag at 5' end.
pMTL8315B Cbu-c16-hydA	6054 bp	pMTL8315B with [FeFe] hydrogenase (EC 1.12.7.2) from *C. butyricum* (1488 bp, 495 aa).
pMTL8315B Cbu-c16-hydA-His	6056 bp	As above, with added 6-His-tag at 5' end.
pMTL8315B Cbu-c24-hydA	6279 bp	pMTL8315B with [FeFe] hydrogenase, cytoplasmic, one subunit form (EC 1.12.7.2) from *C. butyricum* (1725 bp, 574 aa).
pMTL8315B Cbu-c24-hydA-His	6293 bp	As above, with added 6-His-tag at 5' end.
pMTL8315B Cpl-c67-hydA	6059 bp	pMTL8315B with [FeFe] hydrogenase (EC 1.12.7.2) from *C. pluributyricum* (1491 bp, 496 aa).
pMTL8315B Cpl-c67-hydA-His	6065 bp	As above, with added 6-His-tag at 5' end.
pMTL8315B Cpl-c3-buk	5649 bp	pMTL8315B with butyrate kinase (EC 2.7.2.7) from *C. pluributyricum* (1071 bp, 356 aa).
pMTL8315B Cpl-c3-buk-His	5639 bp	As above, with added 6-His-tag at 5' end.
pMTL8315B Cbu-c15-buk	5638 bp	pMTL8315B with butyrate kinase (EC 2.7.2.7) from *C. butyricum* (1068 bp, 355 aa).
pMTL8315B Cbu-c15-buk-His	5636 bp	As above, with added 6-His-tag at 5' end.
pMTL8315B Cpl-c42-buk	5631 bp	pMTL8315B with butyrate kinase (EC 2.7.2.7) from *C. pluributyricum* (1068 bp, 355 aa).
pMTL8315B Cpl-c42-buk-His	5636 bp	As above, with added 6-His-tag at 5' end.

2.19 RT-qPCR for recombinant gene expression

In order to confirm that the inserts introduced into *Clostridium* isolates on pMTL80000 plasmids were transcribed into mRNA, cDNA libraries were created from extracted RNA to perform RT-qPCR. For the extraction of RNA, cell pellets from triplicate stationary *Clostridium* cultures (25 mL) were lysed using the Monarch Total RNA Miniprep Kit (New England BioLabs, USA) following the manufacturer's protocol for tough-to-lyse samples. Mechanical disruption of cells was carried out by adding an equal amount (by weight) of acid-washed glass beads (≤ 106 µm, Sigma-Aldrich, UK) to the microcentrifuge tubes containing cell pellets in RNA protection reagent. The tubes were then vortexed in a bead mill adapter (Vortex-Genie 2; Mo Bio Laboratories, UK) for 5 minutes.

Extracted RNA samples were treated with DNA-free DNA Removal Kit (Invitrogen, UK) to remove genomic DNA (gDNA) contamination. Treated RNA samples were included in a standard PCR reaction (Table 17) with 33 cycles to confirm the absence of gDNA. RNA samples without detectable gDNA contamination were used to generate cDNA libraries using GoScript Reverse Transcription Mix with Oligo(dT) (Promega, UK), following the manufacturer's protocol.

Short-length primers for qPCR were designed using Primer-BLAST (Ye *et al.*, 2012) (Table 20). For gene expression normalisation, primers for reference genes 16S rRNA (*rrs*) and adenylate kinase (*adk*) were designed. These genes were chosen based on a study by Metcalf, Sharif and Weese (2010), who identified them as the two most stable housekeeping genes in *C. difficile*.

RT-qPCR reactions were carried out on a Roche Lightcycler 96, using Takyon No ROX SYBR 2X MasterMix (Eurogentec, Belgium) following the manufacturer's protocol. All qPCR reactions for each of the triplicate RNA samples were carried out in duplicate.

The expression fold change of genes on pMTL80000 plasmids was calculated using the method described by Vandesompele *et al.* (2002), which accounts for the use of multiple reference genes. For this, primer efficiencies

were first calculated by generating a 5-point standard curve, based on multiple serial dilutions. Plasmids containing the relevant target gene were diluted from 1:10 to $1:10^5$ to be used in qPCR reactions in order to determine the Cq values for the standard curve. For housekeeping genes *rrs* and *adk*, purified PCR product was used instead of plasmids. The slope value of the resultant regression equation was used in the following formula to calculate primer efficiency and converted primer efficiency to be used in downstream calculations.

$$Primer\ efficiency\ (\%) = \left(10^{\frac{-1}{Slope}} - 1\right) \times\ 100$$

$$Converted\ primer\ efficiency = \left(\frac{Primer\ efficiency\ (\%)}{100}\right) + 1$$

Next, the Cq values generated from control cDNA libraries (e.g. *C. butyricum* pMTL8315B without insert) were selected as a calibrator to determine gene expression values relative to the empty plasmid control. Selecting a calibrator is necessary to calculate ΔCq values using the following formula:

$$\Delta Ct = Calibrator\ Cq - Sample\ Cq$$

The relative quantity (RQ) value is then calculated by taking into account the previously determined converted primer efficiency (E):

$$RQ = E^{\Delta Ct}$$

Finally, the relative gene expression value, which describes the log10 expression fold change compared to the empty plasmid control, is determined by dividing the relative quantity of the gene of interest by the geometric mean of the relative quantity values of both reference genes:

$$Relative\ gene\ expression = \frac{RQ_{GOI}}{Geomean[RQ_{REFs}]}$$

Table 20. Primers used in RT-qPCR reactions.

Primer name	Sequence
Cbu-c2-hydA-F qPCR	TTGAAGTTGCCTTTGCAGCAG
Cbu-c2-hydA-R qPCR	ACATGCCAACCCACATAGGG
Cbu-c16-hydA-F qPCR	ACCTTGTAAGAGAGCGTGCC
Cbu-c16-hydA-R qPCR	ATGATGCCATACATGCTCCACA
Cbu-c24-hydA-F qPCR	GTTGACCACCGCCATTTACAC
Cbu-c24-hydA-R qPCR	TGAGGGCGGACAAATGGATG
Cpl-c67-hydA-F qPCR	TATGCTGTAGTGGCACCGGC
Cpl-c67-hydA-R qPCR	TAACCGCATCTGCTCCACAG
Cpl-c3-buk-F qPCR	ACATACCAAGGGCAAGTACGAG
Cpl-c3-buk-R qPCR	TTCCGCCACCTAAATGAGCTAC
Cbu-c15-buk-F qPCR	GATCAATAGCTGATGAAATAGG
Cbu-c15-buk-R qPCR	CTTTTTCTTGGAATATCTGGTG
Cpl-c42-buk-F qPCR	GTAGTGGTACATATGGGCGGAG
Cpl-c42-buk-R qPCR	CCGACTGGAACTGAACCTGC
Cbu-c68-rrs F qPCR	CCGCGGTAATACGTAGGTGG
Cbu-c68-rrs R qPCR	CAGTTTGGAATGCAGCACCC
Cpl-c114-rrs F qPCR	CATGCAAGTCGAGCGAGAGA
Cpl-c114-rrs R qPCR	ATTCCCCTCTATGAGGCAGGT
Cbu-c15-adk F qPCR	GGCTCAGGCTGAAGCTCTAAA
Cbu-c15-adk R qPCR	TTGCTCCACATGATGGACATACTC
Cpl-c46-adk F qPCR	CCACATGATGGACACACTCTTCT
Cpl-c46-adk R qPCR	CTAGGACTGTGGCTCAAGCTG

2.20 Protein extraction and Western blots

To confirm the expression of His-tagged recombinant proteins by *Clostridium* species, protein extracts were generated from *Clostridium* mutant cultures. For this, pellets from stationary 25 mL cultures were lysed using BugBuster Protein Extraction Reagent (Sigma-Aldrich, UK), following the manufacturer's protocol with the added lysozyme (5 mg/mL) lysis step for 45 minutes. In addition to the fraction of soluble proteins obtained from cell cultures, non-soluble proteins within pellets of cell debris were solubilised by resuspension in 1 mL of 8 M urea and incubated at room temperature for 30 minutes with gentle mixing at the beginning and the end of the incubation time. According to the Bio-Rad protocol for protein solubilisation, urea is used for this purpose as it disrupts hydrogen bonds, destroys secondary protein structures and therefore brings otherwise-insoluble proteins into solution.

Protein preparations were loaded onto precast Mini-PROTEAN TGX 4–12% polyacrylamide gels (Bio-Rad, USA) and ran in Tris-Glycine-SDS buffer at 150 V for 70 minutes using the Mini-PROTEAN Tetra Cell (Bio-Rad, USA). After gel electrophoresis was completed, gels were rinsed in trays of distilled water for 2 minutes. Subsequently, gels were transferred to trays of ice-cold transfer buffer (25 mM Tris, 192 mM glycine, 10% v/v methanol, pH 8.3) and washed on an orbital shaker for 10 minutes. Pre-cut sheets of gel blotting paper (Whatman, UK) were soaked in ice-cold transfer buffer. Nitrocellulose membranes (Bio-Rad, USA) were first wetted in distilled water for one minute and then washed in ice-cold transfer buffer on an orbital shaker for 5 minutes. Using the gel holder cassette of the Mini Trans-Blot cell (Bio-Rad, USA), gel sandwiches were assembled for blotting. An ice block was suspended in the cell to maintain cold buffer temperatures during the transfer, which was carried out at room temperature for 1 hour at 100 V.

After completion of the transfer, the nitrocellulose membranes were stained with Ponceau S (Sigma-Aldrich, UK) for 10 minutes and rinsed in distilled water to confirm the presence of protein bands on the membrane. Membranes were then washed overnight on an orbital shaker in blocking buffer (PBS containing 0.1% Tween 20 (PBS-T), 3% bovine serum albumin and 0.01%

sodium azide) at 4°C. The next morning, membranes were transferred into primary antibody solution (6*His, His-Tag Antibody 66005-1-Ig (Proteintech, UK) at 1:1000 dilution in PBS-T) and incubated for one hour at room temperature at the slowest speed of the orbital shaker. Next, the membrane was washed in PBS-T for 15 minutes on an orbital shaker at high speed. This was repeated for a total of 5 washes before the membrane was incubated on a slow orbital shaker in secondary antibody solution (IRDye 800CW Donkey anti-Mouse IgG (LI-COR, USA) at 1:20000 dilution in PBS-T) for 1 hour at room temperature, covered in aluminium foil to protect the infrared dye from light to avoid bleaching. As before, the membrane was then washed in PBS-T for five times and shielded from light prior to image development using the Odyssey DLx Imaging System (LI-COR, USA).

2.21 Statistical analyses

Statistical comparisons between groups, such as in the gas volume and composition analyses, were made using the Student's *t*-test function in Microsoft Excel (T.TEST). The tests were two-tailed, assuming a heteroscedastic distribution of values (i.e. larger averages will produce larger errors than smaller averages). Standard deviations were calculated using the Microsoft Excel STDEV function.

Regression analyses were done using the Microsoft Excel Regression Data Analysis function. For gas volume and volatile fatty acid analyses, the constant was set to zero as controls for both experiments remained at a OD_{600} of 0 and neither produced gas or VFAs.

2.22 Other software

Figures were created using the ggplot2 package for R (Wickham, 2016), Microsoft Excel and Paint.NET. Data from the QIIME2 files was extracted using the R package QIIME2R (Bisanz, 2018). Phylogenetic trees were drawn using FigTree 1.14 (Rambaut, 2012).

3 Metataxonomic analysis of an anaerobic baffled reactor

3.1 Introduction

Historically, the classification of species into various taxa has been based on phenotypic characteristics. With the advent of sequencing technologies, the focus has been shifted to molecular characteristics, which reflect evolutionary relationships better than classical phenotypic criteria. On the basis of differences found in ribosomal RNA genes, Woese *et al.* (1990) proposed the classification of life into the three domains Archaea, Bacteria and Eukarya. Since then, the design of universal primers has facilitated the study of microbial taxonomy and allowed the classification of organisms without culturing (Weisburg *et al.*, 1991). Nowadays, newly identified species and genera are mainly classified via the similarity of their 16S rRNA genes to members of other taxa (Yarza *et al.*, 2014).

The ability to identify uncultured organisms has greatly advanced the field of microbial ecology, as the inability to cultivate the majority of microscopically visible organisms had long been recognised as a considerable handicap (Head, Saunders and Pickup, 1998). Initially, the commonly used approaches were based on methods such as the amplification of rRNA genes by PCR and Sanger sequencing of individual clones, or molecular fingerprinting techniques such as denaturing gradient gel electrophoresis (DGGE), with the option to sequence individual bands. However, these approaches are relatively low-throughput and were superseded by new high-throughput sequencing approaches within the last 15 years. Initially, the first metataxonomics studies to employ second-generation sequencing used 454 pyrosequencing to generate large datasets from 16S rRNA PCR amplicons (Sogin *et al.*, 2006). As the first available pyrosequencers only allowed read lengths of 100 bp, early studies, such as Sogin *et al.* (2006), focused on the short V6 variable region of the bacterial 16S rRNA gene. Within the 16S rRNA gene in bacteria and archaea, nine hypervariable regions (V1–V9) have been identified. Overall, the gene is approximately 1600 bp long, however the short read length of second-generation technologies has been a limitation and

requires researchers to focus on the variable regions instead of the whole gene. Most recent metataxonomics studies have been based on the Illumina platform, allowing paired-end reads up to 2 × 300–350 bp (Bukin *et al.*, 2019).

In this project, an amplicon sequencing approach based on protocols by Caporaso *et al.* (2012) was followed. First, genomic DNA was extracted from digestate samples, from which the 16S rRNA V4 region was amplified using a degenerate primer pair. Due to this degeneracy, sequence variation between bacteria and archaea are accounted for so that known biases against certain taxa are removed (Apprill *et al.*, 2015; Parada, Needham and Fuhrman, 2016). This approach uses primers with different barcode regions, so that obtained sequences can later be associated with the digestate samples they were obtained from. The length of the 16S rRNA V4 region is about 254 bp, which means that the Illumina MiSeq v2 sequencer with overlapping 2 × 250 bp paired-end reads can resolve the variable region in its entirety. Bukin *et al.* (2019) discussed that increased paired read lengths of up to 350 bp allow to use additional regions of the 16S rRNA for metataxonomic studies. In their study, they found that sequencing the combined V2-V3 region resulted in a higher resolution than the commonly sequenced V4 region, especially for lower-rank taxa such as genera and species.

Another advancement that has been made in the last few years is the use of amplicon sequence variants (ASVs) over operational taxonomic units (OTUs; Callahan, McMurdie and Holmes, 2017). Previously, clusters of sequencing reads were grouped together in OTUs by a fixed similarity threshold, most commonly 97%, which is below the species delimitation threshold of 98.7% 16S rRNA similarity (Stackebrandt and Ebers, 2006) and therefore primarily resolved sequence taxonomy to genus level and higher. These clusters were then compared to databases containing closed-reference OTUs (e.g. Greengenes), which would discard non-matching OTUs into *de novo* OTUs. The downside of using *de novo* OTUs is that they are unique to the study they were generated from, and therefore cannot be compared to other studies (Edgar, 2017). The use of close-reference OTUs has the downside that the databases are favoured towards well-characterised

communities and that reads from more unusual environments are often discarded. Due to new methods to control sequencing errors, ASVs, which represent single sequence variants, have mostly replaced OTUs. According to Callahan *et al.* (2017), they allow for the better discrimination of ecological patterns, while higher resolutions can help to distinguish species in genera with high 16S rRNA similarity (e.g. *Neisseria*). Furthermore, they are also comparable across studies and not limited by incomplete reference databases. Still, 16S rRNA reference databases are required to infer sequence relatedness to known species and the sequences within the analysis presented here were matched against the SILVA database at 99% similarity.

With the development of third-generation sequencing technologies such as single-molecule real-time sequencing (SMRT) by Pacific Biosciences (PacBio) and nanopore sequencing by Oxford Nanopore Technologies, long reads (> 100 kb) with no theoretical upper limit can be generated (Leggett and Clark, 2017). While these new technologies are relatively error prone (97% accuracy) compared to second-generation technologies, the ongoing development of new bioinformatics tools for error correction has helped to alleviate this issue (Ciuffreda, Rodríguez-Pérez and Flores, 2021). Using these third-generation technologies, the limitations of short-read technologies (e.g. Illumina) can be overcome, which allows the full sequencing of the 16S rRNA gene for metataxonomics studies, which can dramatically increase the taxonomic resolution at species and strain level (Johnson *et al.*, 2019).

Further advances in sequencing technology are continuing to bring together the fields of Metataxonomics and Metagenomics, which allows deeper investigations of microbial communities, sequencing the entire genomic content of microbial samples. However, Ciuffreda *et al.* (2021) argue that numerous challenges such as standardised protocols for microbial characterisation and improvements in taxonomic analysis and sequence comparison software are yet to be solved.

3.1.1 Aims and objectives

The objective of this metataxonomic analysis was to provide insights into the microbial community of a novel 500-litre anaerobic baffled reactor digesting pre-treated seaweed and grass. As previous studies on the communities of anaerobic baffled reactors used older metataxonomics methods such as low-throughput clonal libraries (Lin *et al.*, 2012; Jiang *et al.*, 2018), DGGE fingerprinting (Zhang *et al.*, 2011; Ban *et al.*, 2013; Peng *et al.*, 2013), pyrosequencing (Ziganshin *et al.*, 2013) and/or 97% OTU-based classifications (Gulhane *et al.*, 2017; Tran *et al.*, 2017; Su *et al.*, 2019; Giordani *et al.*, 2021; Zhang and Shi, 2021), the work presented as part of this project stands out as it uses more modern metataxonomics techniques to characterise the microbial community of a large pilot-scale ABR. The aim of this analysis was to show how the reactor's microbial community is shaped by the baffled design of the reactor as well as the switching of feedstocks from seaweed to grass.

3.2 Results

3.2.1 Abundance at the domain level

Across the sequences of all 77 digestate samples around 98% of 16S rRNA reads came from bacteria, while the other identified sequences were associated with members of the domain Archaea. The relative abundance of archaea varied greatly by the feedstock used in the anaerobic baffled reactor, with about 3% and 0.5% of sequences belonging to this domain in grass and seaweed digestate, respectively. Across the reactor, a steady increase of archaeal abundance was seen in seaweed digestate from chamber 1 to chamber 7, while the abundance of archaea in grass digestate was more variable, with peaks in chambers 3 and 7 (Fig. 5).

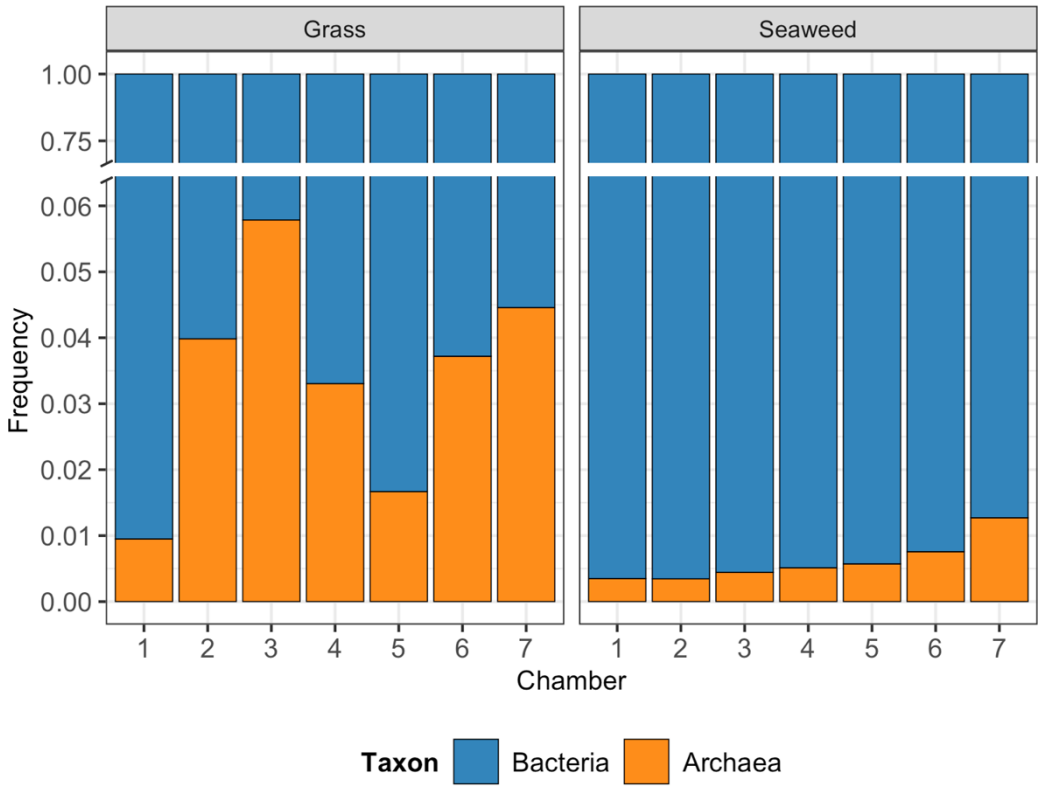


Figure 5. Taxonomic bar plots showing the relative average abundance of sequences and their association by kingdom in grass and seaweed digestate samples, grouped by ABR chamber 1 to 7.

3.2.2 Abundance at the class level

The largest proportion of sequences in both grass and seaweed digestate across all reactor chambers was associated with bacteria from the class Bacteroidia (35% in grass, 42% in seaweed) (Fig. 6). In samples from both feedstock digestates, the relative abundance of sequences associated with Bacteroidia decreased as feedstock was moved through the reactor.

Clostridia was the second and third most abundant class in seaweed and grass digestate samples, respectively (18% in both). Along the chambers of the anaerobic baffled reactor, the abundance of sequences associated with this class was relatively stable in grass digestate but increased steadily in seaweed digestate samples from 15% (chamber 1) to 22% (chamber 7).

Thermotogae was another common class with increasing abundance along the reactor, with low abundance in chambers 1–3 (0.3% in grass, 0.2% in seaweed) and increasing abundance towards chambers 4–7 (5% in grass, 3% in seaweed).

A larger difference between grass and seaweed digestate was seen in the abundance of Spirochaetia, which was the second most common class in grass digestate (19%) but only the fourth most common class in seaweed digestate samples (6%), with stable relative abundance across the reactor.

Conversely, the class of Gammaproteobacteria was of higher abundance in seaweed samples (9%, third most common) than in grass digestate. In grass, the average abundance of Gammaproteobacteria was 1.6% overall, with a steady decrease from chamber 1 (6%) to chambers 5–7 (0.7%).

Other noticeable differences were seen in the class of Oxyphotobacteria, which accounted for 2.6% of sequences in seaweed digestate but only 0.6% in Grass digestate. In seaweed digestate, it was found that these sequences were almost exclusively associated with bladderwrack (*Fucus vesiculosus)* chloroplasts, while the sequences in grass digestate were mostly related to other unidentified chloroplasts. Therefore, these sequences most likely come from the plant material used as feedstock, rather than the digesting microbial communities.

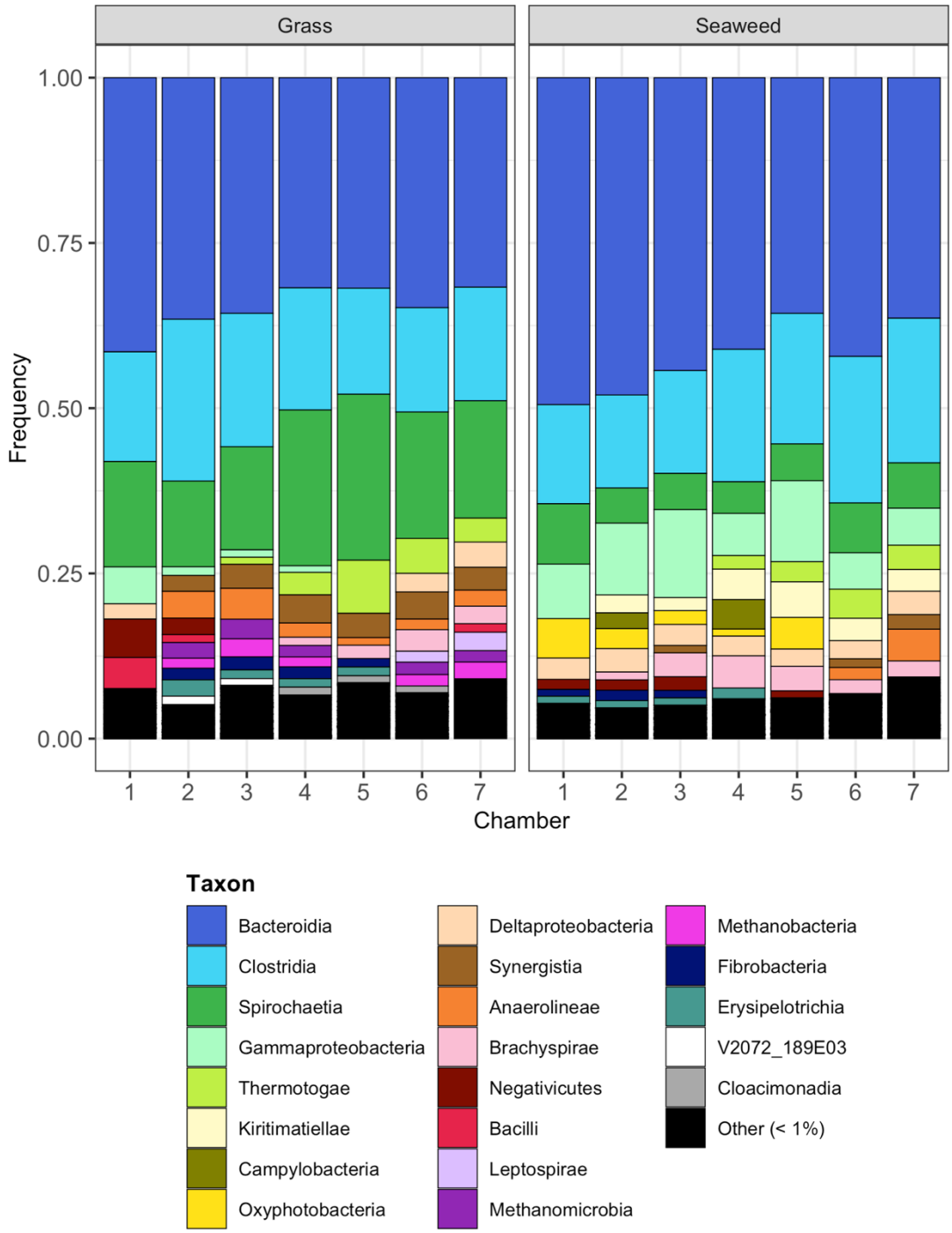


Figure 6. Taxonomic bar plots showing the relative average abundance of sequences and their association by class in grass and seaweed digestate samples, grouped by reactor chamber 1 to 7. The plot shows the 22 classes with an abundance higher than 1% in any of the chambers, while the remaining 79 classes were grouped as “Other”.

3.2.3 Abundance of methanogenic archaea

The archaeal classes of Methanomicrobia and Methanobacteria each accounted for over 1% of sequences in grass digestate from chambers 2–4, 6 and 7. In seaweed digestate, both classes only reached a maximum abundance of 0.8% and 0.2%, respectively (both chamber 7).

Trends in the relative abundance within the domain Archaea were explored further when only sequences associated with archaeal species were included in taxonomic bar plots (Fig. 7 and 8).

Most sequences associated with class Methanomicrobia were from the genus *Methanosarcina*, which made up 48% of all archaeal sequences in both digestate samples.

From the same class, the overall second most abundant archaeal species across all samples was *Methanosaeta*, which was however only found in seaweed digestate. Here, its relative abundance considerably increased from chamber 1 (2%) to chamber 7 (28%), where it was the most abundant methanogen species.

The class of Methanobacteria was mostly represented by the species *Methanobacterium*, which was more abundant in grass (38%) than seaweed digestate (16% of all archaeal sequences), and *Methanobrevibacter*, which was only seen in grass digestate, where it was the third most common archaeal species in chamber 1 (10%) but rapidly decreased in relative abundance to 4% in chamber 2 and an average of 1% in chambers 3 to 7.

The third most abundant taxon was Odinarchaeia, even though the sequences associated with this class could not be associated with any known species. This class was primarily found in seaweed digestate, with increasing abundance towards the outflow of the reactor. For both types of feedstock, the highest average abundance of Odinarchaeia was in chamber 7 (15% of archaea in seaweed, 2% in grass), while very low abundance was seen in chambers 1–4 (0.1% in seaweed, none in grass).

Other species exclusive to either type of feedstock were *Methanolinea* and *Methanofastidiosum* in seaweed digestate (both about 5% in chambers 6 and 7) and *Methanosphaera* in grass digestate (4% in chamber 1).

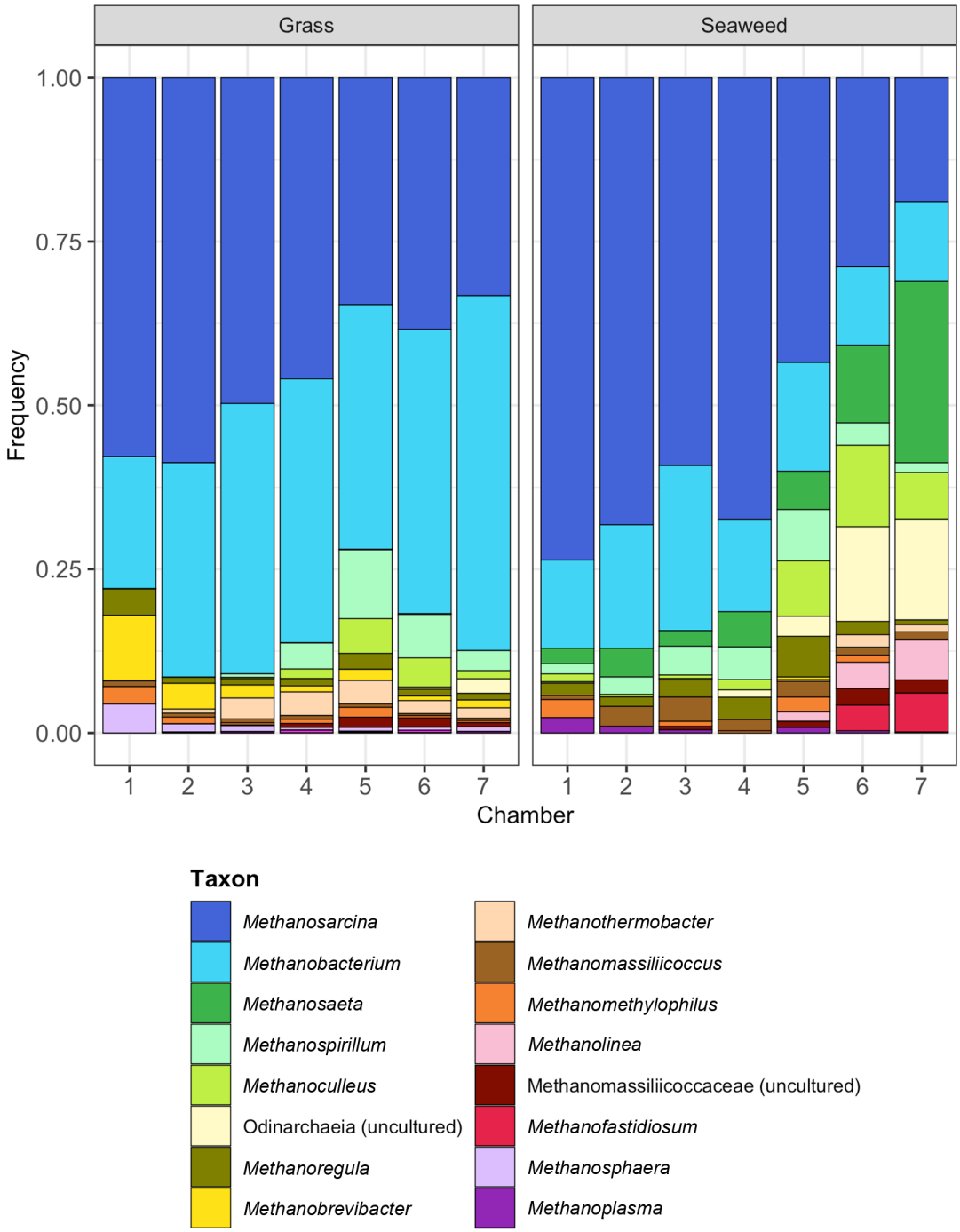


Figure 7. Taxonomic bar plots showing the relative average abundance of archaeal sequences (excluding bacterial sequences) and their association by genus (or next highest known taxon) in grass and seaweed digestate samples, grouped by reactor chamber 1 to 7.

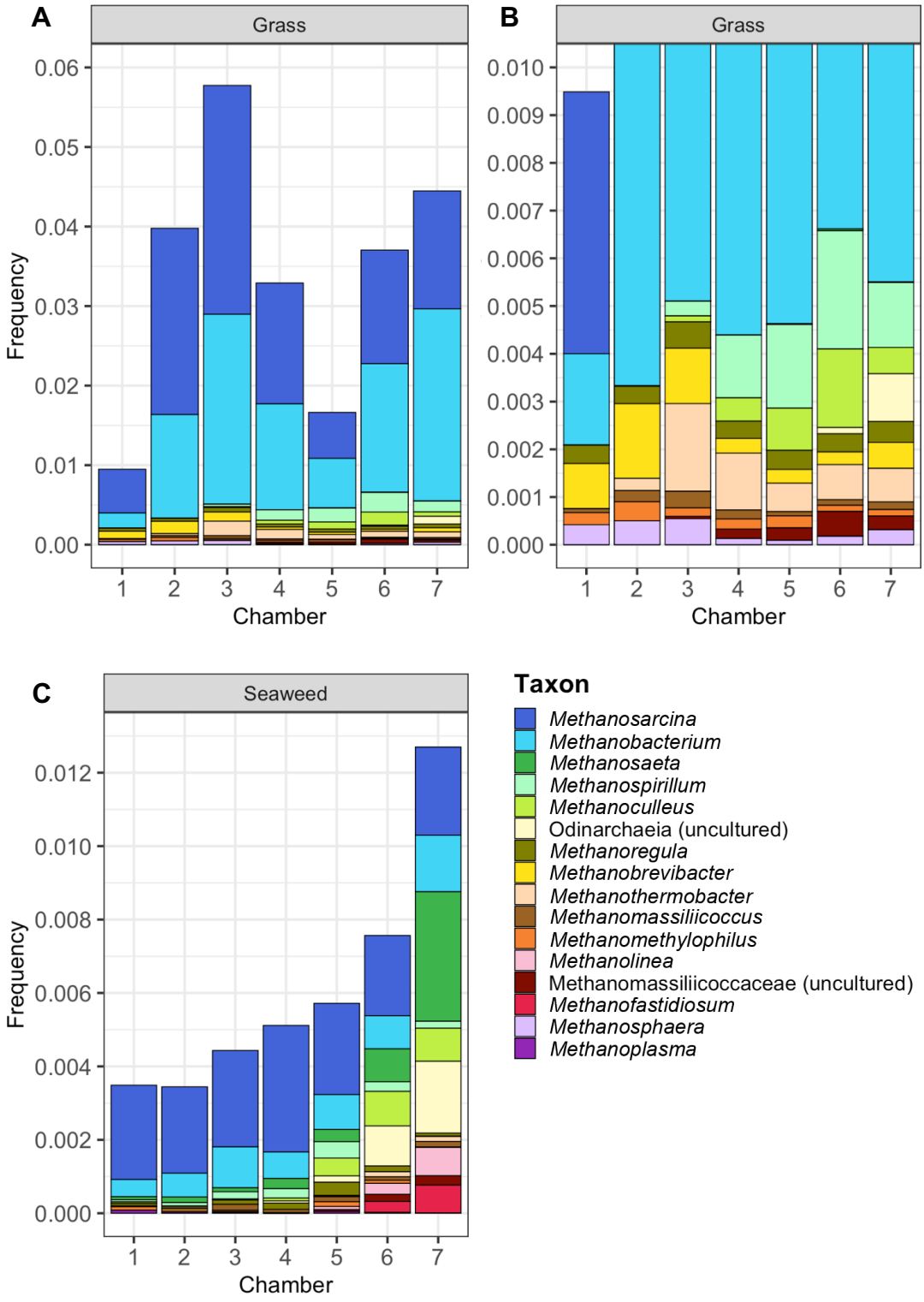


Figure 8. Taxonomic bar plots showing the relative average abundance of archaeal sequences (as a fraction of all sequences from digestate samples) and their association by genus (or next highest known taxon) in grass (A and magnified in B) and seaweed (C) digestate samples, grouped by reactor chamber 1 to 7.

In chamber 5 of grass digestate, it was especially obvious that the relative abundance of archaea experienced a severe drop (Fig. 8). This drop was primarily characterised by a decrease in the abundance of *Methanosarcina* and *Methanobacterium*, which together made up 2.8% of all microbial sequences in grass digestate from chamber 4 (86% of all archaeal sequences), but only 1.2% of all sequences in chamber 5 (72% of all archaea). Conversely, the genera *Methanoculleus, Methanospirillum, Methanoregula* together experienced a rise in their relative abundance from chamber 4 to chamber 5. Their abundance increased from 0.2% to 0.32% of all microbes and from 7% to 18% of all archaeal sequences. By chamber 7 however, the relative abundance of *Methanosarcina* and *Methanobacterium* had recovered to 87% of all archaea and 4% of all microorganisms.

3.2.4 Abundance at the genus level

Shifting the focus back to sequences of both archaea and bacteria, their taxonomy was identified up to the genus level. This split them into a total of 827 taxa, of which 66 reached an average relative abundance of at least 1% in any of the 7 reactor chambers. Across the whole reactor, 24 taxa reached an overall average relative abundance of over 1%, representing 64% of all sequences (Fig. 9).

At this level, the taxon with the highest overall relative abundance was an uncultured genus associated with the family Paludibacteraceae, which reached an average of 7% in both grass and seaweed digestate across all chambers, peaking in the mid to rear-section of the reactor (12% in grass digestate, chamber 6; 11% in seaweed digestate, chamber 4).

The second most abundant taxon was the species with the designation BSV13 from the family Prolixibacteraceae, which also accounted for around 7% of all sequences. However, this species was almost exclusively found in seaweed digestate, where it was the largest taxon at the species level (13%; 0.6% in grass digestate).

The third most abundant species was *Hydrogenispora*, which was common in both types of digestate (7% in seaweed, 5% in grass), peaking in abundance towards chambers 5 and 6.

Other abundant taxa with large discrepancies between the two feedstocks included the species *Treponema* (*Treponema and Treponema 2*; 7% in grass, 2% in seaweed), *Alkaliflexus* (1% in grass, 6% in seaweed), an uncultured species from the family Dysgonomonadaceae (5% in grass, 0.5% in seaweed), group MSBL3 from the family Kiritimatiellaceae (0.03% in grass, 3% in seaweed) and the family Enterobacteriaceae (0.4% in grass, 3% in seaweed), among others.

Many of the most abundant species in the ABR were primarily found either near the inflow or the outflow port, showing either an increase or decrease in abundance from chamber 1 to chamber 7. Among the 30 most abundant genera across all samples (reflecting 70% of all sequences), the highest average decrease in abundance from chamber 1 to chamber 7 was seen in the species *Bacteroides*, while the highest increase from chamber 1 to chamber 7 was in *Fervidobacterium* (Fig. 10).

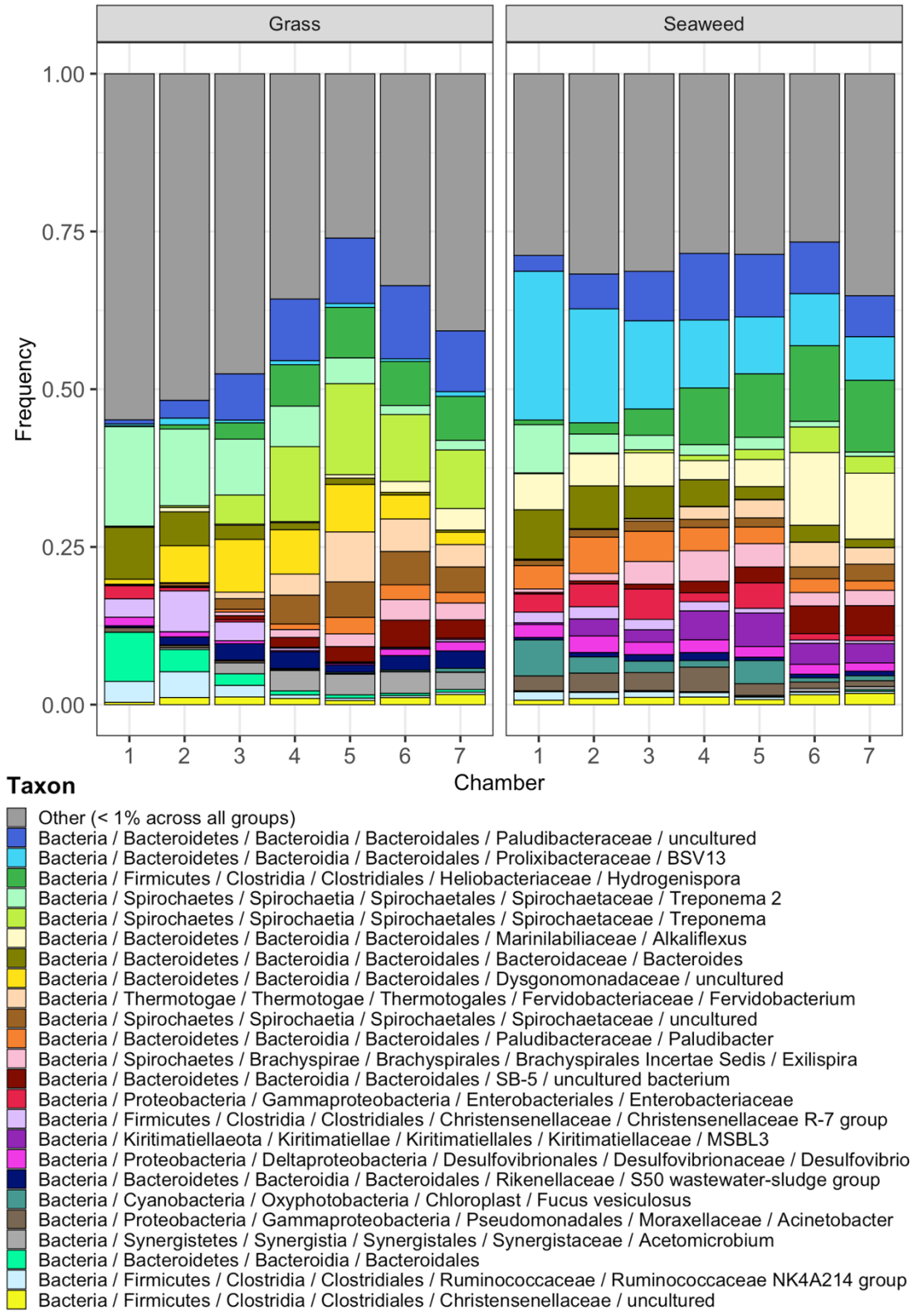


Figure 9. Taxonomic bar plots showing the relative average abundance of sequences and their association by genus (or next highest known taxon, grouped by feedstock and reactor chamber. 803 out of 827 taxa did not reach an average abundance of over 1% across all groups and are classed as “Other”. Taxa are sorted by average relative abundance.

Figure 10. Top 30 taxa (by average abundance across all samples) identified to genus level or higher. Average relative abundance (log10) by feedstock and ABR chamber is shown in a heatmap. Rows are labelled by the SILVA-assigned taxonomy at genus level or higher. Taxa are sorted by average difference in abundance between chamber 1 and chamber 7.

3.2.5 Diversity measures

3.2.5.1 Alpha diversity measures

Alpha diversity (i.e. the diversity of each sample) was determined in QIIME2 by Faith phylogenetic diversity and Pielou's evenness indeces. Faith's index describes taxon richness, expressed as the number of tree units found in a sample (Faith, 1992). Pielou's evenness index measures if different species are equally abundant. Comparing the microbial make-up of digestate samples, it was found that the communities in grass digestate were significantly more diverse (q = 0.004, Kruskal-Wallis) and more even (q = 0.0001) than the communities of seaweed digestate (Fig. 11). Comparing the different reactor chambers running on either grass or seaweed, there was no significant change in community evenness, while a significant increase of Faith diversity from chamber 1 to 7 (q = 0.009) could be observed in both.

3.2.5.2 Beta diversity measures

Secondly, beta diversity (i.e. the comparison of diversity between samples) was determined. This was done using the Unifrac metric, which measures the distance between samples based on phylogenetic information. This metric can either be weighted, which takes into account the relative abundance of taxa between samples, or unweighted, which only considers the presence or absence of taxa. (Lozupone and Knight, 2005). For both the weighted and unweighted Unifrac metric, Principal Coordinates Analysis plots showed that samples clustered most visibly by feedstock (grass vs. seaweed, Fig. 11 A and 11 B), with a statistically significant difference between the two groups as a whole, i.e. when separation by chambers is not taken into account (q = 0.001, determined via permutational multivariate analysis of variance, PERMANOVA). Dispersion was measured using PERMDISP and no significant differences were observed for any of the groups that were compared here, which implies that all significant differences observed through PERMANOVA represent a change in community composition (Anderson, 2001).

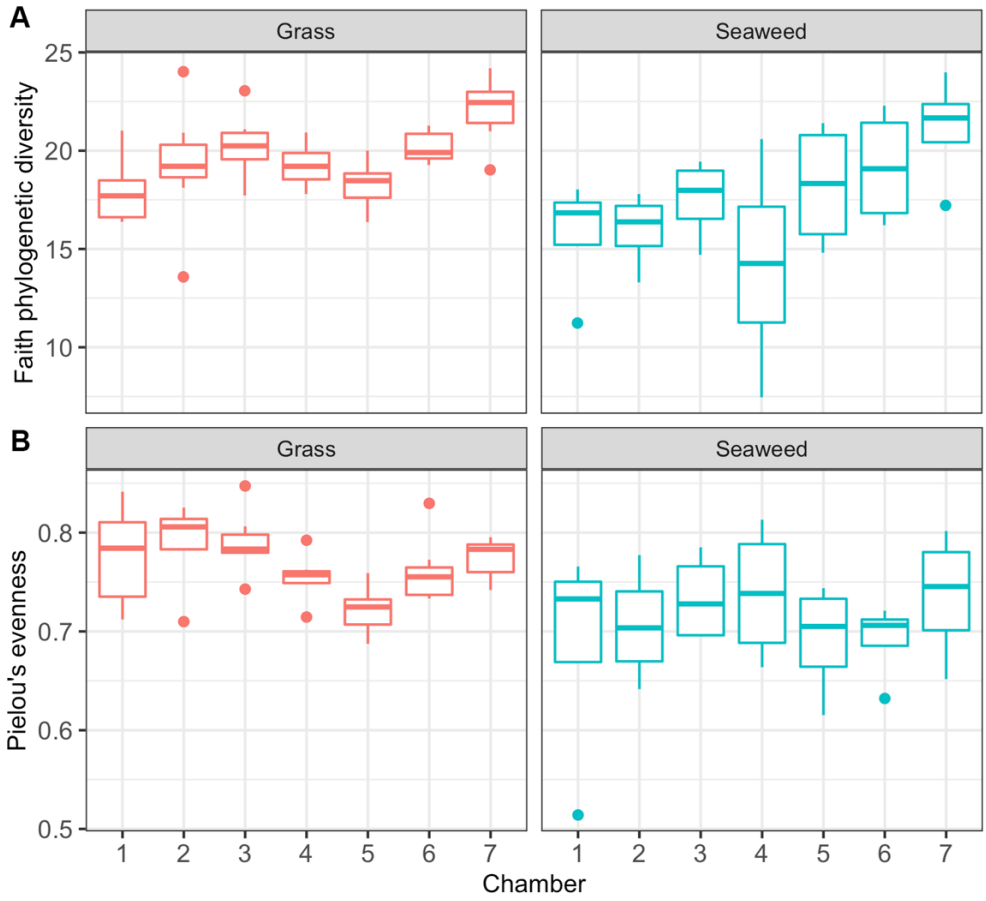


Figure 11. Alpha diversity indices describing Faith phylogenetic diversity (A) and Pielou's evenness (B) of samples (n = 7 in grass, n = 4 in seaweed) grouped by feedstock and reactor chamber. Upper and lower hinges of boxes represent 25th and 75th percentiles, respectively, middle line denotes median. Whiskers extend to the largest/smallest values up to 1.5 times the inter-quartile range. Outlying points are plotted individually.

3.2.5.3 Grouping samples by chamber and feedstock

Comparing sets of samples that were grouped by chamber and feedstock (Table 21 and 22), the grass and seaweed digestate samples from each of the seven chambers were statistically significantly different to one another (e.g. grass chamber 1 compared to seaweed chamber 1). Nevertheless, when considering the average Unifrac distances between sets of samples from both feedstocks, the communities of grass and seaweed digestate were further apart in chamber 1 than they were in chamber 7, indicating that the microbial communities of both feedstocks converged and became increasingly similar towards the rear of the reactor. This decrease in distance was highly significant using both weighted and unweighted Unifrac metrics ($p < 0.0001$) but was more pronounced when the relative abundance of taxa was taken into account (i.e. weighted Unifrac).

Between groups of samples of the same feedstock, it was seen that grass digestate samples of directly adjacent chambers (e.g. samples from chamber 1 and chamber 2; chamber 2 and chamber 3 etc.) were not statistically significantly different, when measured using the unweighted Unifrac metric (Table 21). This was also generally true when using the weighted Unifrac metric (Table 22), with the exception of communities from chamber 4, which were not significantly different to communities in chamber 5 and 6; as well as communities from chamber 5, which were significantly different to the communities in the adjacent chamber 6.

For samples of seaweed digestate, the differences between chambers only reached statistical significance (i.e. $q < 0.05$) when comparing the samples from chambers 1–3 with chambers 6 and 7, as well as chamber 4 with chamber 7 without considering the abundance of present taxa. When this was taken into account (i.e. weighted Unifrac), the only significant difference was seen when comparing the communities from chamber 1 and chamber 7.

Still, the Unifrac distances correlated with the increasing physical distance between chambers in the ABR, showing the dynamic changes of the reactor's microbial community as feedstock is passed from inflow to outflow.

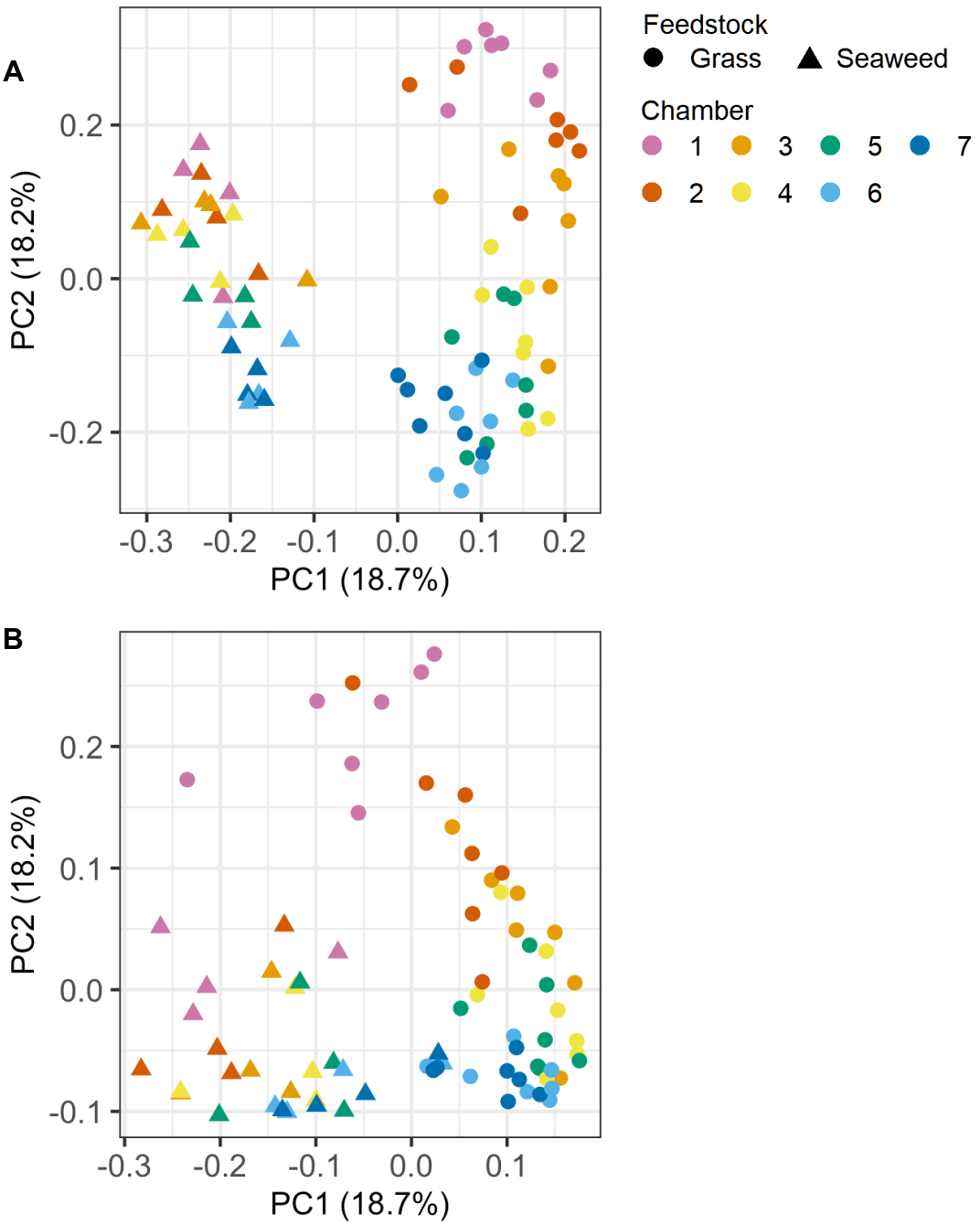


Figure 12. Principal coordinates analysis (PCoA) plot of all 77 analysed digestate samples, showing the unweighted (A) and weighted (B) Unifrac distances on the two axis (PC1 and PC2) with the largest eigenvalues, capturing 36.9% of the variation in the input data. Points with closer ordination represent samples that are more similar to each other in community composition. Samples are grouped by feedstock and ABR chamber.

Table 21. Statistical significance (FDR-adjusted *p*-value i.e. *q*-value) of beta diversity distances between sample groups by feedstock and reactor chamber 1–7, determined by PERMANOVA using the **unweighted Unifrac** metric. Comparisons with a *q*-value < 0.05 are marked in green, others in red.

A		Grass						
		1	2	3	4	5	6	7
Seaweed	1	0.008	0.008	0.008	0.009	0.008	0.008	0.008
	2	0.009	0.008	0.012	0.008	0.008	0.011	0.008
	3	0.012	0.009	0.008	0.008	0.009	0.008	0.009
	4	0.008	0.009	0.009	0.008	0.008	0.008	0.009
	5	0.009	0.008	0.008	0.009	0.008	0.009	0.008
	6	0.009	0.009	0.008	0.008	0.008	0.009	0.008
	7	0.008	0.008	0.008	0.012	0.008	0.009	0.016

B		Grass					
		1	2	3	4	5	6
Grass	2	0.058					
	3	0.008	0.210				
	4	0.008	0.008	0.080			
	5	0.008	0.008	0.008	0.613		
	6	0.008	0.008	0.008	0.020	0.085	
	7	0.008	0.008	0.008	0.008	0.008	0.150

C		Seaweed					
		1	2	3	4	5	6
Seaweed	2	0.946					
	3	0.951	0.885				
	4	0.976	0.964	0.961			
	5	0.657	0.735	0.735	0.976		
	6	0.044	0.048	0.037	0.090	0.666	
	7	0.033	0.041	0.032	0.075	0.168	0.655

Table 22. Statistical significance (FDR-adjusted *p*-value i.e. *q*-value) of beta diversity distances between sample groups by feedstock and reactor chamber 1–7, determined by PERMANOVA using the **weighted Unifrac** metric. Comparisons with a *q*-value < 0.05 are marked in green, others in red.

A		Grass						
		1	2	3	4	5	6	7
Seaweed	1	0.009	0.008	0.008	0.010	0.008	0.009	0.008
	2	0.011	0.009	0.009	0.008	0.009	0.008	0.008
	3	0.010	0.008	0.008	0.011	0.008	0.008	0.008
	4	0.012	0.009	0.008	0.009	0.010	0.013	0.009
	5	0.011	0.008	0.009	0.009	0.008	0.009	0.009
	6	0.008	0.008	0.008	0.008	0.009	0.008	0.008
	7	0.008	0.008	0.008	0.008	0.009	0.009	0.009

B		Grass					
		1	2	3	4	5	6
Grass	2	0.041					
	3	0.008	0.263				
	4	0.008	0.008	0.239			
	5	0.008	0.008	0.011	0.375		
	6	0.008	0.008	0.012	0.131	0.028	
	7	0.008	0.008	0.008	0.027	0.008	0.239

C		Seaweed					
		1	2	3	4	5	6
Seaweed	2	0.788					
	3	0.456	0.796				
	4	0.362	0.690	0.868			
	5	0.181	0.517	0.690	0.806		
	6	0.079	0.255	0.471	0.690	0.972	
	7	0.043	0.128	0.139	0.425	0.786	0.795

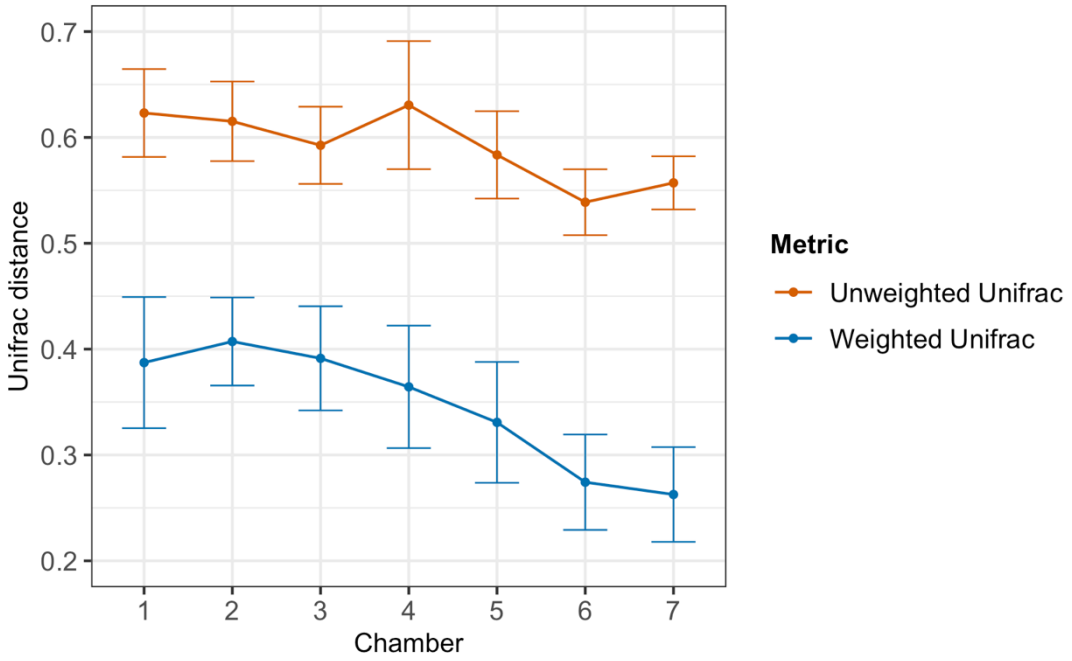


Figure 13. Unifrac distance between sets of seaweed and grass digestate by chamber. Error bars show standard deviation (n = 7).

3.2.5.4 Grouping samples by sampling date

As sampling of the reactor occurred over several weeks, it was investigated whether the sampling date influenced the communities in digestate samples (Fig. 14). Overall, samples from grass and seaweed digestate both showed statistically significant differences by sampling date, but this was much more pronounced in seaweed digestate samples, which had longer intervals between sampling compared to grass samples. Apart from the seaweed digestate sampled in week 11 and week 14, all sets were significantly different here, using both unweighted and weighted Unifrac metrics.

In grass samples, which were taken in seven consecutive weeks from early November to mid-December 2017, there were generally very few statistically significant changes seen in the communities grouped by sampling week. Using the unweighted Unifrac metric, samples collected in weeks 44 to 46 were significantly different to the final set of samples from week 50, while the weighted Unifrac distances between week 44, 45 and 48 as well as 44 and 50 were significantly different.

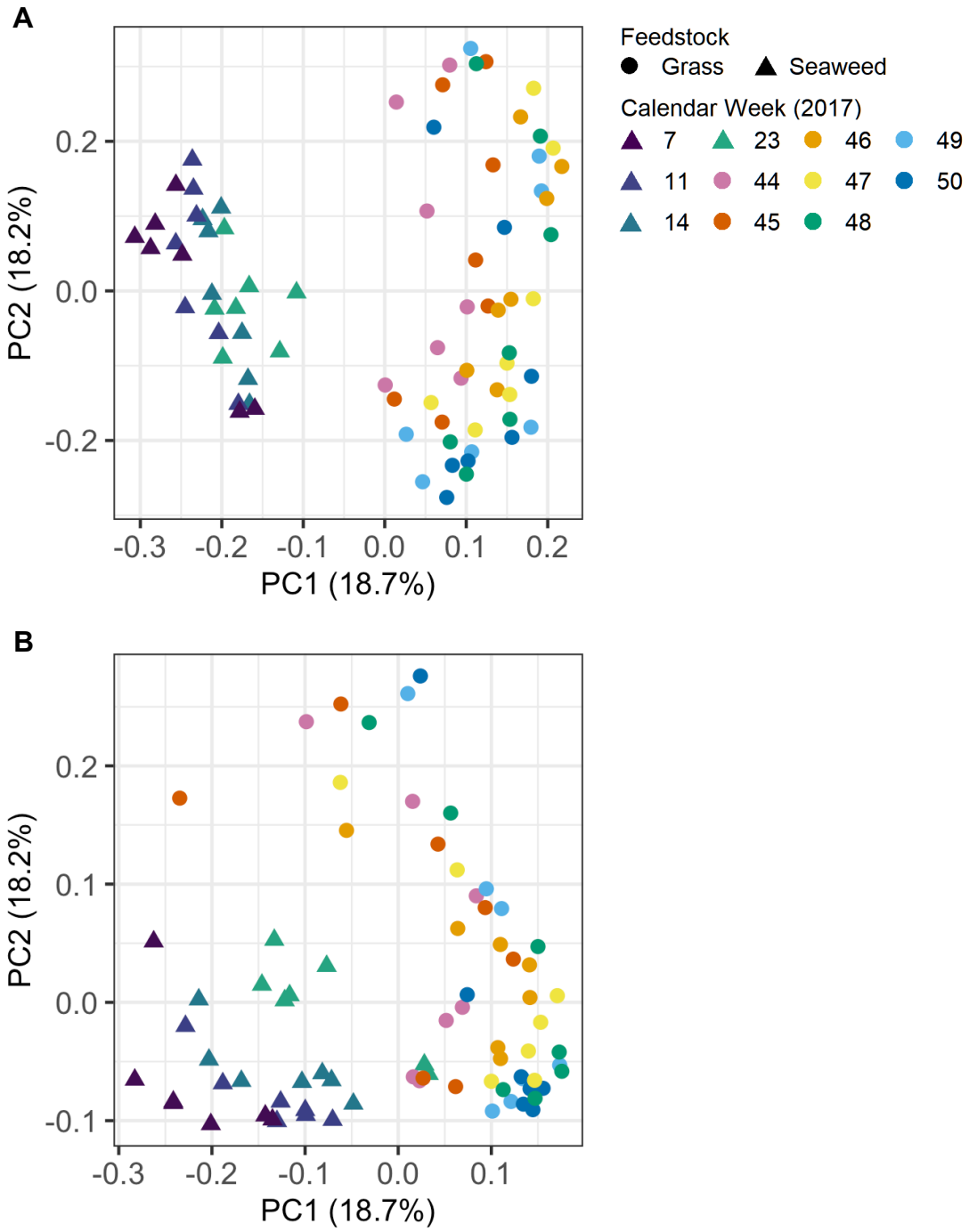


Figure 14. Principal coordinates analysis (PCoA) plot of all 77 analysed digestate samples, showing the unweighted (A) and weighted (B) Unifrac distances on the two axis (PC1 and PC2) with the largest eigenvalues, capturing 36.9% of the variation in the input data. Points with closer ordination represent samples that are more similar to each other in community composition. Samples are grouped by feedstock and sampling week.

Table 23. Statistical significance (FDR-adjusted *p*-value i.e. *q*-value) of beta diversity distances between sample groups by feedstock and sampling week, determined by PERMANOVA using the **unweighted Unifrac** metric. Comparisons with a *q*-value < 0.05 are marked in green, others in red.

		Seaweed				Grass					
	Week	7	11	14	23	44	45	46	47	48	49
Seaweed	11	0.014									
	14	0.007	0.078								
	23	0.003	0.003	0.003							
Grass	44	0.003	0.003	0.004	0.003						
	45	0.003	0.003	0.004	0.005	0.655					
	46	0.003	0.004	0.003	0.004	0.067	0.167				
	47	0.004	0.005	0.004	0.004	0.062	0.097	0.394			
	48	0.003	0.003	0.003	0.003	0.062	0.079	0.165	0.716		
	49	0.003	0.004	0.003	0.004	0.079	0.099	0.055	0.269	0.875	
	50	0.003	0.005	0.003	0.003	0.012	0.040	0.014	0.116	0.432	0.781

Table 24. Statistical significance (FDR-adjusted *p*-value i.e. *q*-value) of beta diversity distances between sample groups by feedstock and sampling week, determined by PERMANOVA using the **weighted Unifrac** metric. Comparisons with a *q*-value < 0.05 are marked in green, others in red.

		Seaweed				Grass					
	Week	7	11	14	23	44	45	46	47	48	49
Seaweed	11	0.009									
	14	0.013	0.147								
	23	0.003	0.003	0.004							
Grass	44	0.003	0.004	0.004	0.006						
	45	0.003	0.004	0.004	0.006	0.784					
	46	0.003	0.003	0.003	0.006	0.072	0.081				
	47	0.003	0.003	0.003	0.004	0.061	0.058	0.746			
	48	0.006	0.004	0.003	0.009	0.047	0.046	0.344	0.784		
	49	0.003	0.003	0.004	0.004	0.067	0.073	0.352	0.724	0.873	
	50	0.003	0.003	0.003	0.006	0.024	0.051	0.054	0.243	0.595	0.784

3.3 Discussion

One of the main questions at the beginning of the project was whether the separation of the reactor into seven chambers has an impact on the resident microbial community and whether it reflects the separate stages of anaerobic digestion.

The microbial changes along the reactor chambers are evident, and the most significant impact on the reactor communities was due to the company's decision to change feedstock from seaweed to grass. Still, while the abundance of certain classes and genera along the reactor can be associated with the two feedstocks, the different stages of anaerobic digestion are also clearly reflected in the metataxonomics data.

3.3.1 Comparison with previous studies on anaerobic baffled reactors (ABRs)

Several studies have previously investigated the microbial communities in ABRs, digesting a wide range of feedstocks (see Chapter 1, Table 1)

While all studies recorded changing trends in the relative abundance of certain taxonomic groups of bacteria and archaea along their reactors, there were not many obvious commonalities to be found when comparing the studies with the results presented here. Differences in the microbial communities are to be expected with the different feedstocks as well as sources of seed inoculum used in these studies, ranging from wastewater treatment plant sludge (e.g. Jiang *et al.*, 2018; Ziganshin *et al.*, 2019; Giordani *et al.*, 2021; Zhang and Shi, 2021) to sludge from digesters treating pig manure (Zhang *et al.*, 2011) or cow dung (Gulhane *et al.*, 2017).

One trend that was found in Blue Sky Bio's ABR as well as the studies by Gulhane *et al.* (2017), Zhang and Shi (2021) and Ziganshin *et al.* (2019) was a drop in the relative abundance of the class Bacteroidia along the reactor. In these studies, however, the abundance of Bacteroidia was lower compared to the reactor studied here and dropped to a greater extent (i.e. under 5–10%).

Overall, none of the cited studies on microbial communities in ABRs found a similarly strong dominance of Bacteroidia as the most abundant class throughout the reactor. Together with the class Clostridia, which was not a dominant feature in most ABR studies, they accounted for over 50% of all sequences in the results presented here.

The dominance of these two classes is likely due to the type of feedstock used in the reactor. Microbial communities that are dominated by Bacteroidia and Clostridia have previously been identified in traditional single-stage AD reactors that have digested lignocellulosic biomass such as grass (Joyce *et al.*, 2018), switchgrass (Deangelis *et al.*, 2012), corn straw (Qiao *et al.*, 2013) and microcrystalline cellulose (Xia *et al.*, 2014), as well as in the rumen of grass-digesting animals such as cows, sheep and deer (Glendinning *et al.*, 2021).

3.3.2 Abundance of hydrolytic taxa

Focusing on the abundance of bacterial genera in Blue Sky Bio's ABR, an increased abundance of *Bacteroides* (class Bacteroidia), *Ruminococcus* (class Clostridia) and *Treponema 2* (*Spirochaeta*[1], class Spirochaetia) was observed towards the inflow port in both grass and seaweed digestate samples.

There are several studies showing that species within these classes and genera have cellulolytic activity. For example, *Ruminococcus albus* and *Treponema bryantii* have been associated with the enhanced degradation of cellulose in co-cultures with *Bacteroides succinogenes* (Kudo, Cheng and Costerton, 1987), while *Spirochaeta caldaria* has been found to enhance

[1] With release 138 of the SILVA RNA database, the genus *Treponema 2* was renamed *Spirochaeta*. The metaxonomic analysis was carried out using the previous version, release 132. Differences between the two taxa are not immediately obvious, as the database entries of both genera contain RNA sequences associated with *Treponema* and *Spirochaeta* species. Conversely, species formerly associated with the genus *Spirochaeta* have been reclassified into the genus *Treponema* and the physiological and functional genomic characteristics of both genera have been described to be "intermixed and of little taxonomic value" (Abt *et al.*, 2013).

cellulose degradation in cultures with *Clostridium thermocellum* (Pohlschroeder, Leschine and Canale-Parola, 1994), and Pagaling *et al.* (2017) described the enrichment of *Treponema* and Bacteroidetes in cellulose-enriched soil microcosms. Chatellard *et al.* (2016) studied the effect of different carbohydrates on microbial community structures and found that the digestion of lignocellulosic feedstocks such as wheat straw was facilitated by the high abundance of hydrolytically active species within the family Ruminococcaceae.

Furthermore, the 16S rRNA gene of one of the organisms that was later isolated from seaweed digestate (Chapter 4, Table 32) was a 100% match with that of *Lacrimispora saccharolyticum* (class Clostridia), which has been associated with the enhanced hydrolysis of cellulose in co-cultures with *Bacteroides cellulosolvens* (Murray, 1986).

Even though the listed species serve as an example how members of the identified genera are known to digest lignocellulosic biomass together, the majority of 16S rRNA V4 sequences obtained from the reactor were only discernible at the genus level or higher. Therefore, sequences in the dataset could not be associated with these particular species, which however does not rule out their presence in the microbial community.

Also, while comparing the abundance of the aforementioned genera in grass and seaweed digestate revealed similar trends along the reactor, the relative abundance of *Treponema* and the class Spirochaetia was noticeably higher in grass digestate compared to seaweed digestate. One explanation might be the relatively high content of lignocellulose in grass species (60–90%, about 30% cellulose; Greenhalf *et al.*, 2012) compared to *Fucus vesiculosus*, where cellulose accounts for about 3% of the algal biomass, while hemicellulose and lignin are only found in trace amounts (Bogolitsyn *et al.*, 2017). Sequences associated with the genus *Treponema 2* were assessed using a BLAST search. Matches with 100% similarity were all from uncultured organisms, some of which were previously found in anaerobic digesters treating lignocellulosic feedstocks such as rice straw and wheat straw.

In the gut microbiome of termites and wood-eating cockroaches, *Treponema* and *Spirochaeta* species have been found to facilitate the

digestion of lignocellulosic biomass (Dröge *et al.*, 2008). Their apparent specialisation in digesting plant material with higher lignin and hemicellulose contents may explain why grass seems to be a better substrate for them to proliferate in, compared to seaweed.

In grass samples, Gammaproteobacteria were highly abundant in the first chamber of the reactor, which coincides with their suggested role in grass hydrolysis (Joyce *et al.*, 2018). Conversely, the high abundance of Gammaproteobacteria throughout seaweed digestate samples might be an example of how feedstock-associated consortia shape the microbial communities in anaerobic digesters, as Quigley *et al.*, (2020) have shown that *Fucus vesiculosus* is generally colonised by a biofilm that predominately consists of bacteria from this class.

Unfortunately, as no aliquots of the initial inoculum (sludge from a wastewater treatment plant) or samples of undigested feedstock were kept, the microbial community at the start-up of the reactor and initial effects of adding feedstock could not be assessed.

Overall, the characteristics of the microbial communities indicate that the hydrolysis step of the anaerobic digestion process is primarily localised within the first chambers of the Blue Sky Bio ABR.

3.3.3 Abundance of acidogenic and acetogenic taxa

Even though the metataxonomics data indicated that the reactor chambers towards the inflow were dominated by species associated with hydrolysis, some of these organisms are also likely playing a role in the production of volatile fatty acids.

Data recorded by the industrial partner has shown that the concentration of acetate was highest in chamber 1. These sporadic measurements were only taken while the reactor was running on seaweed digestate, and just one of their 9 measurements was taken on a day when digestate was sampled for the subsequent metataxonomic analysis (16th February 2017).

On that day, measurements showed that the concentration of acetate was highest in chamber 1 (1.23 g/L) before it dropped towards chamber 4 (0.29 g/L). Then, chamber 5 saw an increase (0.55 g/L) before the concentration of acetate fell drastically in chamber 7 (0.01 g/L). The average of all measurements followed the same decrease-increase-decrease pattern.

This would suggest that the abundance of species associated with the production of acetate was high towards the inlet as well as the mid-section of the reactor, while the consumers of acetate were present in the mid-section as well as towards the outlet of the reactor.

In fact, some of the genera that are associated with the hydrolysis of lignocellulosic biomass are also known to produce volatile fatty acids. For example, species of *Bacteroides* produce VFAs such as acetate, propionate, butyrate, succinate (Macy, Ljungdahl and Gottschalk, 1978; Miller, 1978; Touw, van Steenbergen and de Graaff, 1982), *Ruminococcus* produces acetate, formate, succinate, lactate and ethanol (Ezaki, 2015) and the genus *Treponema* includes acidogenic species such as *T. saccharophilum* (Paster and Canale-Parola, 1985), *T. succinifaciens* (Cwyk and Canale-Parola, 1979) and *T. bryantii* (Stanton and Canale-Parola, 1980) that produce acetate, formate, succinate, lactate and ethanol.

As the abundance of the acidogenesis-associated genera fell, other taxa increased in abundance towards chambers 5–7. These mainly included species that are known producers of acetate, hydrogen and carbon dioxide, which are the compounds converted to methane in acetoclastic and hydrogenotrophic methanogenesis (Tomei *et al.*, 2009).

The taxon with the highest average abundance across samples of both feedstock digestates was associated with the family Paludibacteraceae, which so far includes the acetate and propionate-producing species *Paludibacter jiangxiensis* (Qiu *et al.*, 2014) and *Paludibacter propionicigenes* (Ueki *et al.*, 2006). Representative sequences of this taxon have mainly been found in other anaerobic digesters and wastewater treatment plants, but the associated species have not been isolated.

Another taxon that was highly abundant towards the outflow port in both types of feedstock was the genus *Hydrogenispora*. This genus includes the species *Hydrogenispora ethanolica,* which produces acetate, ethanol and hydrogen (Liu *et al.*, 2014) as well as the thermophilic *Hydrogenispora sp.* strain UUS1-1, producing lactate, acetate, hydrogen and carbon dioxide (Ungkulpasvich *et al.*, 2021).

Other genera with increased abundance in the end section of the reactor included *Fervidobacterium* and *Treponema*. While this trend was seen in both grass and seaweed digestate, their relative abundance was noticeably increased in grass digestate.

While the genus *Fervidobacterium* is recognised to be acetogenic due to their production of acetate, CO_2 and H_2, it is also thermophilic with optimum growth temperatures around 65–70°C (Huber and Stetter, 2015). The apparent presence of this genus is therefore at odds with the reactor's reported mesophilic operation temperature of 37°C, even though the industrial partner did report issues with overheating towards the outflow port of the reactor. As no temperature measurements were recorded, the extent of these fluctuations remains unknown.

Both in grass and seaweed digestate from chambers 4–7, sequences were increasingly associated with the class Thermotogae, of which *Fervidobacterium* is a member. Although this class primarily includes organisms that are extreme thermophiles (Pollo, Zhaxybayeva and Nesbø, 2015), one mesophilic genus (*Mesotoga*) with optimal growth temperatures between 37–40°C has been found in oil reservoirs, harbours and wastewater treatment plants (Nesbø *et al.*, 2019).

Within the reactor, the genus *Mesotoga* was however only the third most abundant genus of this class, with an average abundance of 0.05% across all samples. The second most abundant genus was *Pseudothermotoga*, with an average abundance of 0.08% and the vast majority of sequences associated with the class Thermotogae were linked with the genus *Fervidobacterium* (2.4% across all samples). When performing a BLAST search on the sequences associated with *Fervidobacterium*, matches with 100%

similarity were mostly from uncultured organisms that were sequenced from wastewater treatment plants, microbial fuel cells as well as mesophilic anaerobic digesters. The highest match to a cultured organism (*Fervidobacterium riparium*) was only at 92% identity, which is below the recognised 16S rRNA similarity threshold for genus delimitation (94.5%, Yarza et al., 2014).

Pseudothermotoga-associated sequences also matched highly with uncultured organisms, which were previously found in high temperature settings such as thermophilic anaerobic digesters and geothermal vents.

Therefore, while some Thermotogae sequences obtained from the anaerobic baffled reactor were matched with sequences isolated from high temperature environments, others were previously found in mesophilic settings. Overall, this means that the higher abundance of the class Thermotogae does not necessarily reflect on issues related to temperature control but demonstrates that there are likely other mesophilic species within the class Thermotogae, which have however not yet been cultured.

The genus *Treponema*, which was primarily seen in chambers 4–7 of the ABR is distinct from the genus *Treponema 2* / *Spirochaeta* associated with hydrolytic/acidogenic processes in the first stages of the reactor. Therefore, due to the diversity of species in the genus *Treponema*, it is difficult to predict which kind of organisms could be responsible for the genus's large increase in abundance in chambers 4–7 of the ABR. Considering the overall increase of acetogenic genera, the proliferating members of the genus *Treponema* may also have been producers of acetate, H_2 and CO_2, such as *T. isoptericolens* (Dröge *et al.*, 2008), *T. primitia* or *T. azotonutricium* (Graber, Leadbetter and Breznak, 2004).

3.3.3.1 Evidence of process disruption

Although the relative abundance of the most common genera in the metataxonomic dataset generally peaked in chambers 1–2 or 6–7, only an uncultured genus from the family Dysgonomonadaceae stood out for its high abundance in the mid-section of the reactor. This genus was very abundant in

chambers 3–5 when the reactor was digesting grass, but virtually absent when it was digesting seaweed.

Members of this family (e.g. *Petrimonas*, *Proteiniphilum*) have previously been associated with increased organic loading rates (Maus *et al.*, 2020) and can act as indicators for process failure due to organic overloading (He, Li and Peng, 2017). In fact, besides the aforementioned problems with overheating, for which experimental details are not available due no data having been recorded by the industrial partner, Blue Sky Bio also reported issues with clogging and extensive foaming after they had switched feedstocks from seaweed to grass. He *et al.* (2017) investigated foaming in mesophilic food waste digesters and found that it correlated with an increased abundance of *Petrimonas* and *Proteiniphilum*, which are both acetogens (Zhang *et al.*, 2017).

3.3.3.2 Summary

In summary, it appears that the anaerobic digestion stages of acidogenesis and acetogenesis are spatially separated along the reactor, with a higher prevalence of acidogens in chambers 1–4 and an increased abundance of acetogens in chambers 4–7.

When the reactor was loaded with grass, the increased abundance of species in the family Dysgonomonadaceae may have shifted the process of acetogenesis towards the mid-section of the reactor.

As with the samples from seaweed digestate, it would have been worthwhile to compare the abundance of genera in grass digestate with the concentration of VFAs in the reactor, however such measurements were not taken by the industrial partner.

It was very interesting to see how the reported overloading and temperature issues during the operation of the reactor may have led to plausible changes in the microbial community. Still, despite indicators that point to the overloading of grass feedstock, the reactor finally shared common highly abundant taxa in chamber 7 when compared to seaweed digestate.

3.3.4 Abundance of methanogenic taxa

The most noticeable difference between the different feedstocks was the low abundance of methanogenic archaea in seaweed samples when compared to grass digestate. Unfortunately, the industrial partner did not record any data on the methane output of their digester, so that no direct connections could be made between gas production and archaeal abundance.

Furthermore, a second obvious difference between the feedstocks was that the abundance of methanogenic species in seaweed digestate experienced a steady increase from chamber 1 to chamber 7, while the abundance of archaea in grass digestate samples peaked in chamber 3 and in chamber 7.

The high abundance of archaea in grass digestate from chamber 3 coincides with the increased number of bacterial sequences associated with the family Dysgonomonadaceae. Even though genera within this family are associated with process disruption, they are primarily acetogenic, and it has been suggested that their production of acetate and CO_2 promotes methanogenic activity (Zhang *et al.*, 2017).

3.3.4.1 Impact of feedstock

The increased abundance of archaea in grass digestate compared to seaweed digestate could also be an indication that the use of grass as a feedstock enhanced the production of methane when compared to the type of seaweed sourced by the industrial partner (*Fucus vesiculosus*).

Vanegas and Bartlett (2013) studied the potential of marine algae for methane production and compared seaweed with grass. They described that the produced amount of methane from seaweed was largely species dependent. The use of sugar kelp (*Saccharina latissima*) and furbellow (*Saccorhiza polyschides*) as feedstock resulted in higher methane yields compared to grass, while the methane production from oarweed (*Laminaria digitata*) was comparable. When sea lettuce (*Ulva sp.*) and toothed wrack (*Fucus serratus*) were used, much less methane was produced. In their study,

the use of *F. serratus*, which is a close relative to *F. vesiculosus*, resulted in the lowest amount of methane produced. The authors suggested that this might be due to varying concentrations of inhibitory compounds present in the seaweed species.

Brown algae such as *F. vesiculosus* accumulate iodine and produce a range of halogenated secondary metabolites such as brominated and chlorinated hydrocarbons, which include strong inhibitors of methane production (Nielsen *et al.*, 2020). In a fuel analysis comparing five common British seaweed species as well as terrestrial plant feedstocks, *F. vesiculosus* was found to contain the highest total halogen content, about 5 times higher than *L. digitata* and 36 times higher than the grass species *Miscanthus* (Ross *et al.*, 2008). A good example of the inhibitory effects of seaweed on methanogenic archaea has recently been shown in cattle feed studies, where this characteristic was proven to be exploitable in reducing beef cattle methane emissions by over 80% (Roque *et al.*, 2021).

Methanogenesis can be inhibited by a wide range of compounds such as nitroethane, 2-nitroethanol, 2-nitro-1-propanol (Zhang *et al.*, 2020), ethyl-2-butynote, trans-2-butenoate, propynoic acid, 2-bromoethanesulphonate (Zhou, Meng and Yu, 2011), chloroform, chlorophenols and bromophenols (Nielsen *et al.*, 2020). Due to the complexity of compounds found in seaweed species, it is not possible to pinpoint the compounds that may be responsible to the effects seen in this study, however many known mechanisms of inhibition directly involve the action of the vital enzyme methyl-coenzyme M reductase. For example, the inhibitory effects of 2-bromoethanesulphonate are mediated through its structural analogy with methyl-coenzyme M and result in a reversible competitive inhibition (Smith, 1983), while other compounds such as 3-nitrooxypropanol and 3-bromopropanesulfonate inhibit the enzyme irreversibly by interacting with the nickel-containing F_{430} cofactor within its active sites. For other methanogenesis-inhibiting nitrocompounds such as nitroethane and 2-nitroethanol, it is not known whether they attack the enzyme (Thauer, 2019).

3.3.4.2 Impact of high salinity

Another inhibitor of methanogenesis is high salinity, which has been reported as an issue when treating food industry wastewater or macroalgae (Hierholtzer and Akunna, 2012). The salinity of Blue Sky Bio's reactor wasn't reported, but it is a possibility that the sodium introduced by the seaweed used in this anaerobic digester could have had an inhibitory effect on methanogenesis as well. The sodium chloride content of algae tends to be much higher than that of terrestrial plants, accounting for up to 15% of their dry weight in some species before washing (Milledge and Harvey, 2018).

Feijoo *et al.* (1995) found that sodium concentrations as low as 3 g/L inhibited the production of methane from a VFA mix by 50%. For perspective, the sodium concentration of the Irish Sea, from where the feedstock was sourced, is about 11 g/L (Nygård and Dring, 2008), while the sodium content of washed and dried *F. vesiculosus* from the south coast of England was found to be 29 g/kg (Ross *et al.*, 2008). According to the industrial partner, sourced seaweed was not washed before it was macerated and used as feedstock, which therefore very likely would have led to conditions of high salinity in the reactor. Another indicator for this was the high relative abundance of the halophilic genus *Alkaliflexus* in seaweed digestate samples, which grows at an optimum sodium concentration of 20 g/L (Whitman, 2015) and was of much lower abundance in grass digestate samples. Furthermore, one of the most abundant genera in seaweed samples was BSV13 from the family Prolixibacteraceae, which comprises numerous species isolated from seawater and marine sediments (e.g. Iino *et al.*, 2014; Gwak *et al.*, 2015; Zhou *et al.*, 2019).

3.3.4.3 Genus abundance as indicator for process parameters

The relative abundance of methanogenic genera in the digestate samples can also serve as an indicator for the conditions along the reactor. *Methanosarcina*, which is primarily an acetoclastic methanogen (Kurade *et al.*, 2019), was dominant in the first chambers of the reactor regardless of

feedstock. In grass digestate, the gradual increase in the abundance of the hydrogenotrophic *Methanobacterium* (Siegert *et al.*, 2014) indicates the increased production of H_2 and CO_2 along the reactor. In grass and seaweed samples from chambers 4–7, this is also evident due to the increased abundance of *Methanospirillum* and *Methanoculleus*, which are hydrogenotrophic methanogens commonly seen in anaerobic digesters (Cai *et al.*, 2016).

When the reactor digested seaweed, *Methanosaeta* replaced *Methanosarcina* as the most abundant methanogen in chamber 7. Even though *Methanosaeta* is also an acetoclastic methanogen, it thrives in lower acetate concentrations and can outcompete *Methanosarcina* due to a lower half-saturation coefficient, meaning that it requires less acetate for growth (Conklin, Stensel and Ferguson, 2006). In grass digestate, *Methanosaeta* was virtually absent, which might indicate higher concentrations of acetate when the reactor was digesting grass, favouring *Methanosarcina*. Unfortunately, unlike with seaweed samples, measurements of the acetate concentration weren't made while the reactor was digesting grass.

As mentioned, the industrial partner acknowledged issues with temperature control in chambers 5–7, which likely had an effect of the abundance of bacterial species.

Most archaeal genera have large ranges in their optimal growth temperature, so it is not known whether the changes in abundance might have been an effect of increased temperatures selecting for thermophilic species. While many species in the reactor's most abundant genera are mesophilic (Boone, 2015; Chong and Boone, 2015; Patel, 2015; Wagner, 2020), they also include some thermophilic species such as *Methanoculleus thermophilus, Methanosarcina thermophila*, *Methanosaeta thermophila* or *Methanobacterium thermaggregance* (Suryawanshi, Chaudhari and Kothari, 2010).

Nevertheless, a large increase in the abundance of sequences associated with the phylum Odinarchaeia was observed in seaweed digestate samples. As sequences belonging to this proposed and uncultured phylum

have been found in several metagenomes of hot springs (Zaremba-Niedzwiedzka *et al.*, 2017), this might be another indicator for high temperatures in the reactor.

Other changes in the abundance of archaeal genera may be the effect of increased organic load rates. In grass digestate samples from chamber 5, the decrease in the abundance of *Methanosarcina* and *Methanobacterium* and increase in *Methanospirillum* and *Methanoculleus* mirrored findings by Lerm *et al.*, (2012), who investigated the effects of organic overloads on the composition of archaeal communities in anaerobic digesters. Here, they attributed the changes in community composition with the increased abundance of hydrogen from VFA degradation.

3.3.4.4 Summary

Overall, the distribution of methanogens within the reactor suggests that the process of methanogenesis increasingly occurs towards the outflow port of the reactor. Furthermore, the abundance of the various methanogen genera implies that the process is primarily acetoclastic at first and becomes increasingly hydrogenotrophic towards chamber 7.

Due to the lack of data on methane output as the industrial partner did not attempt to measure methane production, it is unknown if the increase in methanogen abundance caused an increase in methane production. Otherwise, it might have been possible to recommend whether a revised reactor design would benefit from additional chambers, allowing to further exploit the steady increase in methanogen abundance towards the reactor's outlet.

3.3.5 Alpha- and Beta diversity measures

Measures such as Faith diversity or Pielou's evenness are not always predictive of process performance, as studies that attempted to associate these characteristics have often shown very different outcomes (De Vrieze *et al.*, 2018).

Nevertheless, communities in grass digestate samples that showed higher diversity and evenness than those of seaweed digestate, experienced a notable decrease of both measures in chamber 5. Again, like changes in the abundance of certain bacterial and archaeal genera, the decrease of evenness and diversity is yet another indicator that has been associated with process disruption due to organic load shocks (Braz *et al.*, 2019).

Still, it seems like the reactor recovered in subsequent chambers as both diversity and evenness saw a steady increase in chambers 6 and 7. Overall, even though there are many indicators in the dataset of grass digestate samples that correlate to issues of process disruption in the mid-stage of the reactor, it also looks like all of them return to their prior state in chamber 7. It is difficult to infer process performance characteristics without measurements such as VFA concentrations or methane output, however this may support previous observations, which established the idea that ABRs are more resilient to process disruptions as the result of organic shock loadings when compared to single-stage AD systems (Barber and Stuckey, 1999).

The increase in Faith's phylogenetic diversity observed along Blue Sky Bio's ABR is similar to observations within the aforementioned literature. Here, six studies that investigated alpha diversity in ABRs reported an increase in microbial diversity when comparing the first to the last chamber. While this was also true for Jiang *et al.*, (2018) and Ziganshin *et al.* (2019), they found the highest diversity in the mid-sections of their reactors.

In the analysis of beta diversity measures, it was seen that the seamless change in feedstock from seaweed to grass had the biggest impact on the microbial communities in the reactor, and that communities in digestate samples from both feedstocks showed more similarities along the reactor.

As the communities in grass and seaweed digestate were most distant in samples from chamber 1, it appears that the changing of feedstock had its biggest effect on communities near the inlet.

Considering that anaerobic digestion involves the breakdown of complex molecules into simple sugars, amino acids and VFAs (Adekunle and Okolie, 2015), this could mean that the communities found closer to the inlet were more specialised and therefore less diverse at first, followed by the proliferation of generalist species in later chambers.

When feedstocks were changed, specialists in the degradation of seaweed would have been replaced with those better suited to digest grass. Downstream of these specialised hydrolytically active communities, already established communities in chambers 6–7 were less affected by the change of feedstock, as they likely still would have received similar quantities of the simpler compounds required in acetogenesis and methanogenesis.

3.3.5.1 Community specialisation

As the most abundant genera followed similar patterns in abundance, regardless of feedstock, it may be that the specialised organisms were within genera that were of low abundance individually.

Unifrac distances between communities of the two feedstocks were larger when they were unweighted, compared to when community differences were weighted for taxa of higher abundance (Fig. 13 and 14 A vs. 14 B). This suggests that taxa of low abundance were a defining feature that differentiated communities in both feedstocks.

In grass digestate samples from chamber 1 and chamber 2, the number of genera with an abundance of less than 1% was especially high and accounted for over 50% of all sequences. Ziganshin *et al.* (2013) studied the effect of various feedstock types on microbial communities. In their study, they described that specialist genera such as *Fibrobacter* thrived on high-fibre feedstock, but still only contributed less than 2% of all sequences, which showed that specialised consortia do not necessarily dominate the microbial communities of anaerobic digesters.

3.3.6 Limitations of 16S rRNA metataxonomics

A limitation of using 16S rRNA to identify species in a sample may be found when trying to identify species with multiple gene copies. Pei *et al.*, (2010) found that about 75% of bacterial species in the GenBank database contained more than one copy of the 16S rRNA gene. Among these species, 55% showed polymorphisms within the gene copies.

Also, as the number of 16S rRNA copies varies between taxa, the abundance of certain phyla, such as Gammaproteobacteria and Firmicutes, may often be overestimated as they tend to have higher 16S rRNA copy numbers. Conversely, the abundance of Acidobacteria are more likely to be underestimated due to lower 16S rRNA copy numbers (Větrovský and Baldrian, 2013). Single copy housekeeping genes such as the chaperone *hsp60* have been proposed as alternative phylogenetic markers, with comparable resolution to 16S rRNA and may be used, as universal degenerate primers are available (Schellenberg *et al.*, 2011), however the availability of large sequence databases for 16S rRNA sequences still makes this gene the gold standard for metataxonomic studies.

3.3.7 Outlook

It would have been noteworthy to investigate how quickly the overall community of the reactor changed when the feedstock was replaced, however this wasn't possible as the grass digestate samples for the metataxonomic analysis were only taken about 3 months after this change was made. Nevertheless, it is fair to assume that a large proportion of the organisms that were initially active in the digestion of seaweed were carried over to digest grass. Communities in digestate that would have been sampled immediately after the feedstock was switched to grass possibly would have initially clustered closer to the original seaweed communities seen on the PCoA plots.

Therefore, it would have been interesting to see the effects on the community composition if the change from seaweed to grass digestate would have included a complete evacuation of the reactor.

3.3.8 Conclusion

The metataxonomic data gathered from the industrial partner's ABR has shown many indicators that the design with seven chambers has created niches for microbial consortia that are specialised in the anaerobic digestion processes of hydrolysis, acidogenesis, acetogenesis and methanogenesis.

As the changes along the reactor were gradual, the relatively high number of chambers allowed the observation of these trends more clearly than in systems that are less complex.

Nevertheless, a lack of measurements on VFA and methane production means that process performance could not be associated with the metataxonomic data. This has made it difficult to reach conclusions about whether the performance of this pilot-stage reactor has been adequate.

Still, several independent indicators within the metataxonomic dataset support reports that the operation of this anaerobic baffled reactor has not been without difficulties. Reported issues with overloading, as well as possible effects of high salinity were undoubtedly reflected in the abundance of certain genera and changes in community diversity.

Despite these complications, the obtained results have clearly demonstrated that a lot of information about the processes within anaerobic baffled reactors can be deducted from metataxonomic data alone.

However, changes in feedstock combined with possible differences in organic load rate and temperature are all additional variables that made it more difficult to pin down the effects of the reactor's baffled design.

If this reactor was to be set up again, its operation would benefit from more regular sampling, with measurements of parameters such as temperature, pH, chemical oxygen demand (COD), methane and VFA concentration. Ideally, several replicates of the reactor would be run in parallel, which however would be difficult at this scale. Overall, the assessment of this reactor would have benefited from more stable experimental conditions.

4 Isolating VFA-producing bacteria from digestate

4.1 Introduction

Microorganisms with useful properties have been used since prehistoric times and strains of industrial importance continue to be identified for their potential to produce compounds or enzymes with applications in a wide variety of industries such as food, cosmetic, environmental and pharmaceutical (Pessôa *et al.*, 2019). In the chemical and energy industry, there has been a recent focus on the biological production of VFAs and hydrogen. In order to explore the cost-effective production of these compounds from biomass, Baeyens *et al.* (2020) suggests the isolation of new bacterial species or modifying the pathways of present microorganisms to enhance productivity.

4.2 Aims and objectives

In the techniques for the selection of industrially important microorganisms proposed by Steele and Stowers (1991), the first step is to identify the desired product, followed by the identification of known microbes with the desired activity. After identifying high abundances of genera associated with VFA production in the metataxonomics data, the aim was to isolate novel species from digestate and study their potential industrial use. The focus was on the uncultured genus of Paludibacteraceae, which was the most abundant genus in the reactor overall. The only genus in this family contains the species *Paludibacter propionicigenes* (Ueki *et al.*, 2006) and *Paludibacter jiangxiensis* (Qiu *et al.*, 2014). These species have been found to produce propionate and the uncultured genus was therefore presumed to be of possible industrial use. Since *P. propionicigenes* was first isolated in Peptone-Yeast-Glucose (PYG) medium at 30°C, this medium was also used at the same temperature to attempt an isolation from digestate. Based on sequencing data, seaweed digestate sampled from chamber 5 on 15th March 2017 was identified to contain the highest abundance of the unidentified member of the Paludibacteraceae (28% of all sequences), and this sample was spread on PYG agar.

4.3 Results

After digestate was spread on agar plates, they were incubated in the anaerobic cabinet for 5 days. Subsequently, twelve colonies were picked based on differences in morphology to be re-spread on PYG agar. Colonies from axenic cultures were used in colony PCR reactions for the 16S rRNA gene and nine PCR reactions resulted in product that was sent for Sanger sequencing. Based on differences found in the 16S rRNA V4 gene sequences obtained here, morphology and characteristics such as observed gas production, 7 unique organisms were identified (Table 25 and Fig. 15). Using the SILVA database, one organism was identified to species level (*Clostridium butyricum*), two to genus level and four to family level. Species 6 and 8 showed practically identical 16S rRNA V4 sequences, that were only slightly different in the length of the sequence obtained by Sanger sequencing (256 bp vs 253 bp, respectively). Despite this, there were clear differences in their morphology (rods vs. cocci), indicating that the sequence of the V4 region alone is not sufficient to differentiate the two organisms.

Table 25. Summary of organisms isolated and identified from digestate.

ID	Taxon (16S rRNA)	Shape	Size (approx.)	Spore forming	Gas production
1	*Clostridium butyricum*	Rods	8 µm × 2 µm	Yes	Yes
3	*Enterococcus sp.*	Cocci	0.8 - 1 µm diameter	Not observed	Not observed
4	*Clostridium sp.*	Rods	1.2–1.5 µm × 10–25 µm 1.2 × 2.5 µm (spores)	Yes	Yes
6	Ruminococcaceae *gen. sp.*	Rods	2.5 µm × 0.6 µm 0.8 µm diameter (spores)	Yes	Not observed
8	Ruminococcaceae *gen. sp.*	Cocci	1.2 µm diameter	Not observed	Not observed
9	Lachnospiraceae *gen. sp.*	Rods	3–4 µm × 0.6 µm	Yes	Not observed
11	Lachnospiraceae *gen. sp.*	Rods	3–4 µm × 0.6 µm	Yes	Yes

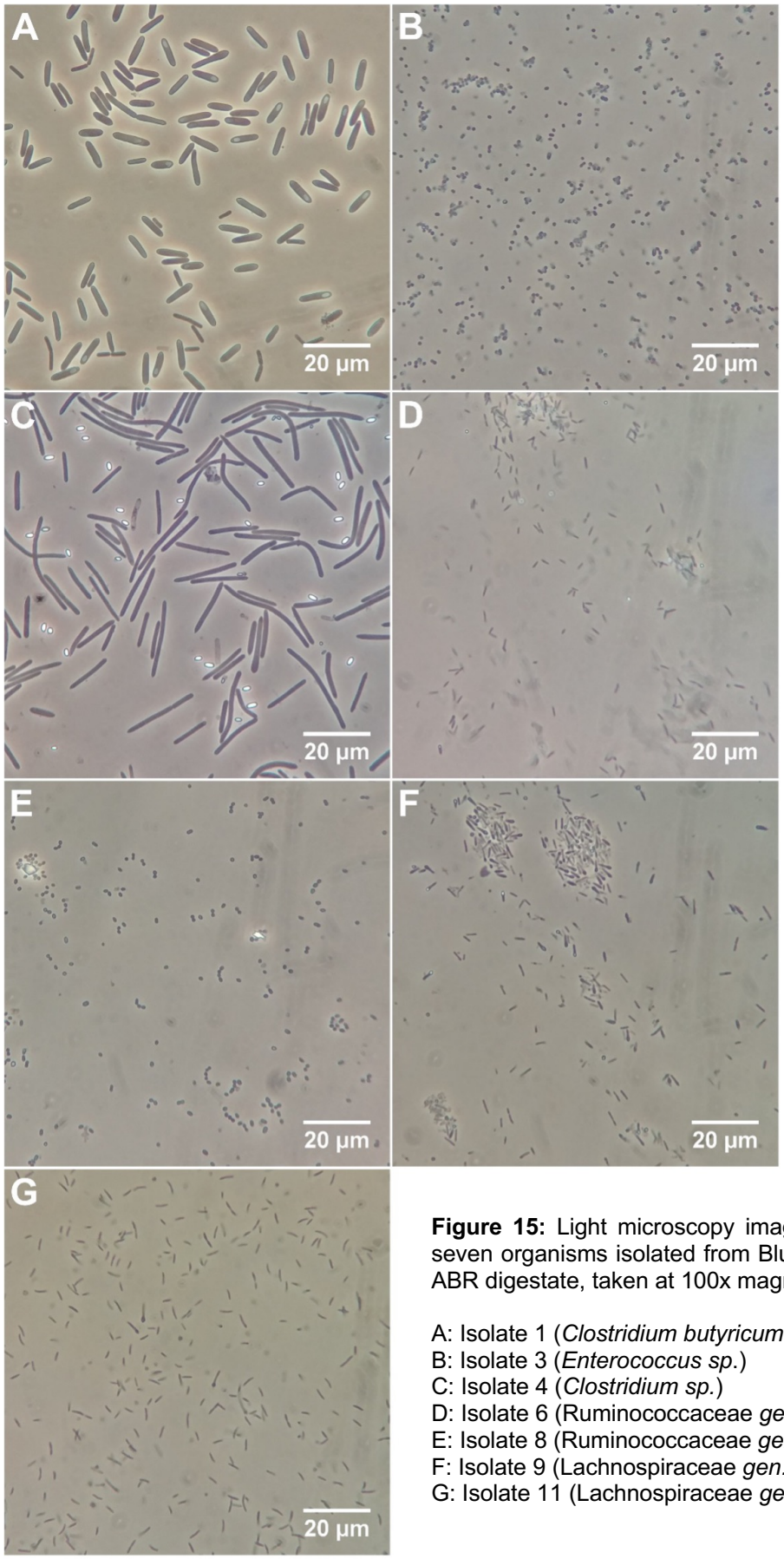


Figure 15: Light microscopy images of the seven organisms isolated from Blue Sky Bio ABR digestate, taken at 100x magnification.

A: Isolate 1 (*Clostridium butyricum*)
B: Isolate 3 (*Enterococcus sp.*)
C: Isolate 4 (*Clostridium sp.*)
D: Isolate 6 (Ruminococcaceae *gen. sp.*)
E: Isolate 8 (Ruminococcaceae *gen. sp.*)
F: Isolate 9 (Lachnospiraceae *gen. sp.*)
G: Isolate 11 (Lachnospiraceae *gen. sp.*)

4.3.1 Abundance of isolated organisms in metataxonomics data

Despite the high abundance of Paludibacteraceae sequences in the digestate sample, none of the isolated organisms were identified as a member of the family.

Apart from isolates 3 (*Enterococcus sp.*) and 9 (Lachnospiraceae *gen. sp.*), representative 16S rRNA sequences of all isolates were found in the data associated with the digestate samples they were isolated from, albeit at very low abundance (Table 26). Nevertheless, the 16S rRNA sequences of the other two isolates 3 and 9 were present within the overall dataset, representing 1552 and 957 sequences counted from a total of 7.5 million, respectively.

Sequences identical to the 16S rRNA sequence of isolate 3 were found in 34 of the 77 digestate samples, while those matching to isolate 9 were found in just 6 digestate samples. Compared to all other digestate samples, the number of sequences obtained from the digestate sample the organisms were isolated from (81,680) was 20% lower than the median, which would make it more likely that sequences from organisms with very low abundance are not represented in the dataset. Considering that the DADA2 denoising algorithm removes singletons (Callahan *et al.*, 2016), it can be concluded that the highest possible abundance of the two undetected isolates would likely be less than 0.002%, which would be the relative abundance of two sequences in the dataset of 81,680 sequences.

Table 26. Abundance of identified isolate sequences in the Illumina sequencing data of the digestate sample the bacteria were isolated from (Chamber 5, 15th March 2017). Isolates 6 and 8 had identical 16S rRNA V4 sequences despite clear morphological differences.

ID	Taxon (16S rRNA)	Representative sequence abundance
1	*Clostridium butyricum*	0.002%
3	*Enterococcus sp.*	not detected
4	*Clostridium sp.*	0.006%
6, 8	Ruminococcaceae *gen. sp.*	0.017%
9	Lachnospiraceae *gen. sp.*	not detected
11	Lachnospiraceae *gen. sp.*	0.038%

4.3.2 Isolate taxonomy

4.3.2.1 BLAST searches

In addition to the taxonomy assigned by the SILVA database, 16S rRNA V4 sequences from the seven isolates were used in a BLAST search to identify their closest known relatives (Tables 27–32).

Table 27. Isolate 1 (*Clostridium butyricum)* 16S rRNA V4 region BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium butyricum strain JCM 1391	99%	100%	NR_113244.1
Clostridium butyricum strain VPI3266	99%	100%	NR_042144.1
Clostridium butyricum strain ATCC 19398	99%	100%	NR_112170.1
Clostridium puniceum strain DSM 2619	99%	99%	NR_119031.1
Clostridium saccharoperbutylacetonicum strain N1-4(HMT)	99%	99%	NR_102516.1

Table 28. Isolate 3 (*Enterococcus sp.)* 16S rRNA V4 region BLAST results (Top 5).

Description	Query cover	Ident	Accession
Enterococcus gallinarum strain LMG 13129	99%	99%	NR_104559.2
Enterococcus saccharolyticus strain LMG 11427	99%	99%	NR_114786.2
Enterococcus gallinarum strain NBRC 100675	99%	99%	NR_113924.1
Enterococcus saccharolyticus strain NBRC 100493	99%	99%	NR_113909.1
Enterococcus casseliflavus strain NBRC 100478	99%	99%	NR_104560.1

Table 29. Isolate 4 (*Clostridium sp.)* 16S rRNA V4 region BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium chromiireducens strain GCAF-1	99%	99%	NR_122090.1
Clostridium saccharobutylicum strain NCP 262	99%	99%	NR_122051.1
Clostridium puniceum strain DSM 2619	99%	99%	NR_119031.1
Clostridium saccharoperbutylacetonicum strain N1-4(HMT)	99%	99%	NR_102516.1
Clostridium beijerinckii strain JCM 1390	99%	99%	NR_113388.1

Table 30. Isolate 6 and 8 (Ruminococcaceae *gen. sp.)* 16S rRNA V4 region BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium jeddahense strain JCD	100%	99%	NR_144697.1
Clostridium sporosphaeroides strain DSM 1294	100%	98%	NR_044835.2
Clostridium merdae strain Marseille-P2953	100%	98%	NR_147400.1
Caproiciproducens galactitolivorans strain BS-1	100%	93%	NR_145929.1
Neglecta timonensis strain SN17	100%	92%	NR_144736.1

Table 31. Isolate 9 (Lachnospiraceae *gen. sp.)* 16S rRNA V4 region BLAST results (Top 5).

Description	Query cover	Ident	Accession
Faecalicatena contorta strain DSM 3982	99%	97%	NR_117147.1
Faecalicatena fissicatena strain DSM 3598	99%	97%	NR_104800.1
Faecalicatena orotica strain DSM 1287	99%	97%	NR_117130.1
Faecalicatena orotica strain JCM 1429	99%	97%	NR_114392.1
Faecalicatena orotica strain ATCC 13619	99%	96%	NR_118729.2

Table 32. Isolate 11 (Lachnospiraceae *gen. sp.)* 16S rRNA V4 region BLAST results (Top 5).

Description	Query cover	Ident	Accession
Lacrimispora saccharolyticum strain WM1	99%	100%	NR_102852.1
Lacrimispora celerecrescens strain 18A	99%	99%	NR_026100.1
Lacrimispora sphenoides strain ATCC 19403	99%	99%	NR_112174.1
Lacrimispora sphenoides strain DSM 4024	99%	99%	NR_026409.1
Lacrimispora sphenoides strain DSM 632	99%	99%	NR_119035.1

4.3.2.2 16S rRNA phylogenetic analysis

To explore the relatedness between isolates and species identified through BLAST searches, a molecular phylogenetic analysis was carried out using the 16S rRNA sequences of the isolates and closely related species identified by BLAST search (Fig. 16). Overall, it was seen that isolates, which have been initially identified as members of the same family or genus using SILVA, were recognised as being more closely related by this analysis as well.

An exception to this were top matches to isolates 1, 4, 8 and 6. The closest relatives to the two isolates identified as Ruminococcaceae *gen. sp.* belong to the same genus as isolates 1 and 4 (*Clostridium*), however the phylogenetic analysis separated them on distant branches. This highlights that the taxonomy of the genus *Clostridium* is still largely based on historical classifications by phenotype, which does not necessarily reflect 16S rRNA similarity (Cruz-Morales *et al.*, 2019).

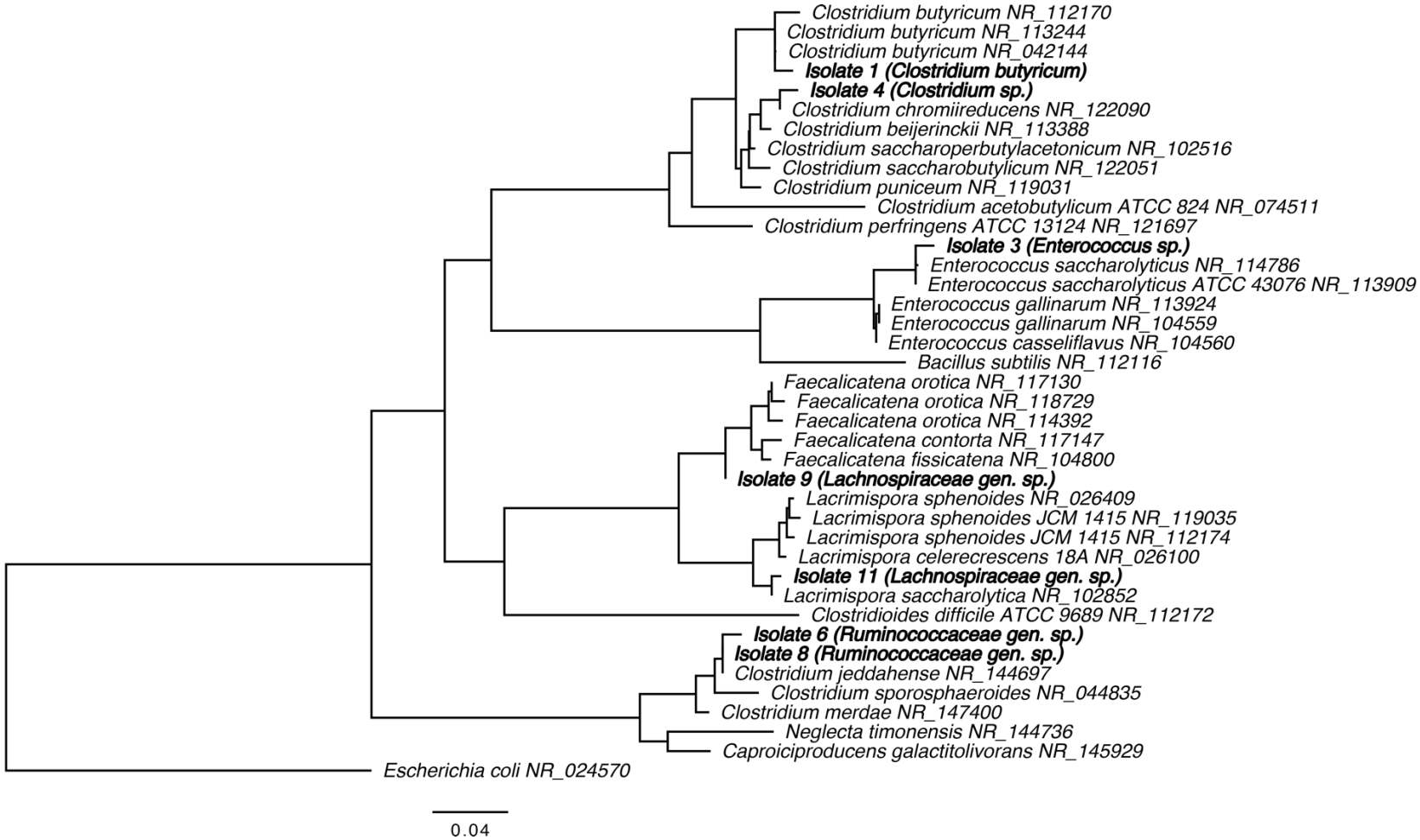


Figure 16. Maximum likelihood tree inferred under the GTR+GAMMA model and rooted to the *E. coli* outgroup. The branches are scaled in terms of the expected number of substitutions per site. The compared sequences are 16S rRNA V4 sequences of the isolates (showing SILVA-assigned taxonomy), their top 5 BLAST matches and selected other species. Tree was drawn using the DSMZ single-gene phylogeny server (Meier-Kolthoff *et al.*, 2021).

4.3.3 Acid production by bacterial isolates

As the aim was to find isolates producing VFAs, the pH of isolate cultures was measured to detect whether they produce acids. After the isolates had been grown for five days, all cultures showed visible signs of growth (i.e. increase in OD_{600}). The use of pH indicator strips showed that the medium of all cultures except the non-inoculated control and isolate 6 (Ruminococcaceae *gen. sp.*) had been acidified.

Following this observation, the culture supernatants of the six acid-producing organisms were further analysed using GC-FID to determine the concentration of VFAs (Fig. 17). Compared to the control (non-inoculated PYG medium with 5 g/L glucose) the most significant increases were found for acetic acid and butyric acid among the tested compounds. Acetic acid was the only VFA detected in the control (0.11 g/L) and the final measurements were adjusted to reflect the amount produced by the organisms. Other acids such as lactic acid or formic acid were not measurable with the column used on the GC-FID machine.

Production of more than 0.5 g/L acetic acid was observed in cultures of *Clostridium butyricum* and the two cultures containing species of the family *Lachnospiraceae*. Butyric acid was not detected in the control and 1.5–2 g/L were produced by the two *Clostridium* isolates with lower amounts (0.5 g/L) produced by *Lachnospiraceae* isolate 9.

As the two isolates of *Clostridium* were identified to produce high amounts of butyric acid, they were selected to be studied further.

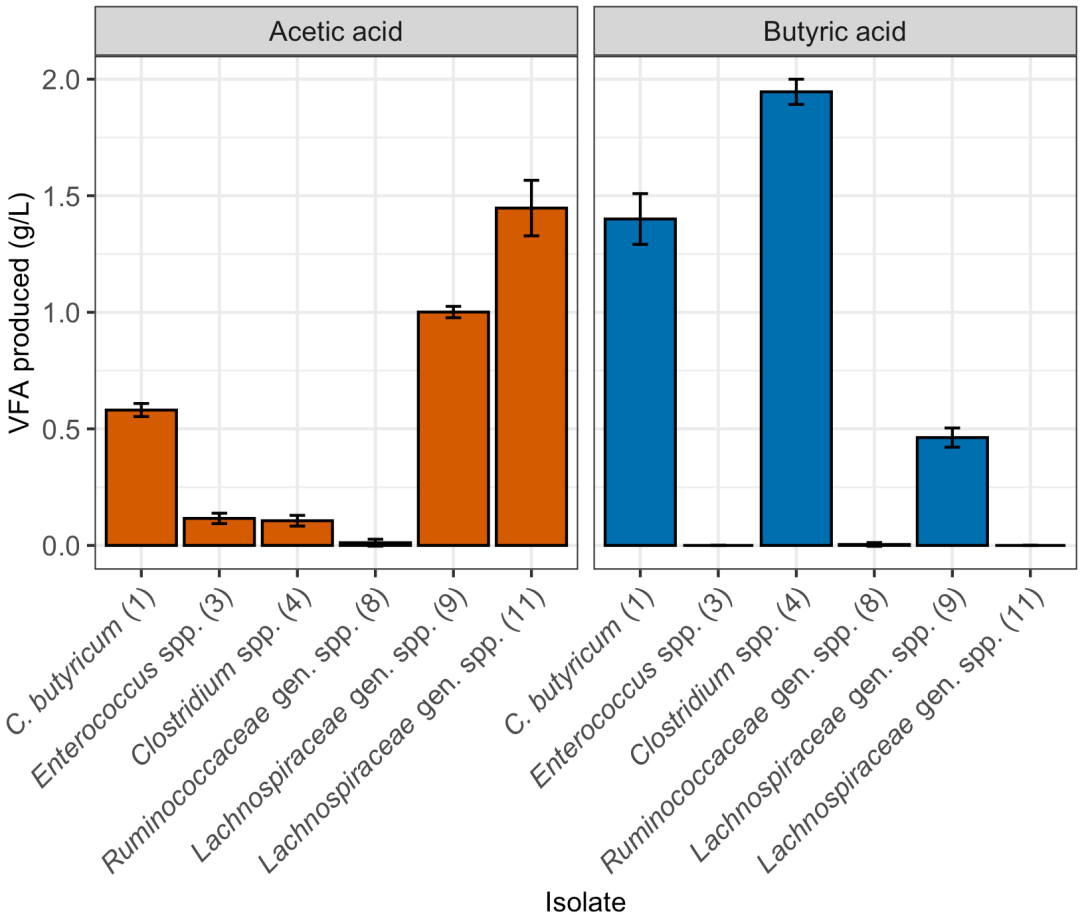


Figure 17. Concentration of acetic acid and butyric acid in supernatant of stationary phase cultures of acid-producing bacteria isolated from digestate samples. Error bars show standard deviation (n = 3).

4.3.4 Culturing *Clostridium* isolates

4.3.4.1 Antibiotic resistance

The resistance of the two *Clostridium* isolates to antibiotics was determined, which allowed choosing the suitable resistance genes that could be used in transformation experiments to select for genetically engineered strains. Pure cultures from frozen stocks were streaked on PYG agar containing various concentrations of antibiotics (Table 33). It was found that both species are highly susceptible to low concentrations of ampicillin, chloramphenicol, erythromycin and tetracycline. Therefore, these antibiotics would be most suitable to select for mutants containing the appropriate antibiotic resistance genes. For use in bacterial conjugation reactions to transform plasmids from *E. coli* to *Clostridium*, it was found that cycloserine can be used at concentrations of 100 µg/mL to select against *E. coli* conjugation donors while allowing growth of *Clostridium* recipients.

Table 33. Growth of *Clostridium* isolates tested on PYG agar containing various antibiotics and the tested concentrations that allowed/inhibited growth.

	C. butyricum		***Clostridium sp.***	
Antibiotic	Resistant	Susceptible	Resistant	Susceptible
Ampicillin	N/A	≥ 1 µg/mL	N/A	≥ 1 µg/mL
Chloramphenicol	N/A	≥ 1 µg/mL	N/A	≥ 1 µg/mL
Cycloserine	≤ 100 µg/mL	≥ 300 µg/mL	≤ 300 µg/mL	N/A
Erythromycin	N/A	≥ 10 µg/mL	N/A	≥ 10 µg/mL
Kanamycin	≤ 10 µg/mL	≥ 100 µg/mL	N/A	≥ 1 µg/mL
Spectinomycin	≤ 100 µg/mL	≥ 600 µg/mL	≤ 1 µg/mL	≥ 10 µg/mL
Tetracycline	N/A	≥ 0.1 µg/mL	N/A	≥ 0.1 µg/mL
Thiamphenicol	≤ 1 µg/mL	≥ 15 µg/mL	N/A	≥ 1 µg/mL

4.3.4.2 Colony morphology

In the experiments to determine antibiotic resistance, it was noted that *C. butyricum* colonies displayed two morphologies. Primarily, they appeared as opaque and cream-coloured, with round and well-defined margins. In this shape, the colonies required some force when scraping them off an agar plate with an inoculation loop. Colonies that had been scraped off plates tended to stick together and required vigorous shaking to break them up in liquid. Under the microscope, the bacteria in these colonies were seen to form spores and looked like cells of *C. butyricum* shown in Fig. 15, Panel 1.

The second observed colony shape on *C. butyricum* agar plates was pale and translucent, with faint and irregular margins. Colonies of this shape were easier to lift off agar plates. When observed under a light microscope, bacteria in these colonies were of similar length to the bacteria found in opaque colonies, however they were thinner and very motile. Spores were not observed in these colonies.

It was initially assumed that one of the two colony shapes was the result of contamination, however they were both confirmed to be *C. butyricum* through colony PCR and 16S rRNA sequencing.

No such differences were seen for the other *Clostridium sp.* isolate, which presented as opaque colonies that looked similar to those seen in *C. butyricum*. In contrast, the colonies from this isolate were not as adherent to agar and readily dissipated in liquid, as they did not clump together.

4.4 Discussion

4.4.1 Inability to culture species with high sequence abundance

Despite the high abundance of sequences in the reactor associated with the family Paludibacteraceae, and the use of similar culturing methods that were initially used in the isolation of related bacteria (Ueki et al., 2006), none of the isolates obtained from seaweed digestate were associated with this family. The isolation of organisms that were primarily of very low abundance raised the question why it was not possible to isolate bacteria from any of the genera that made up the majority of sequences in the dataset.

It was noteworthy that all isolates apart from the one identified as *Enterococcus*, were associated with the class Clostridia and that, out of the isolated bacteria, 5 out of 6 had formed visible spores. This suggests that spore-forming isolates may have been selected for, likely as a result of the sampling and storage procedures at the industrial partner's laboratories. Here, digestate samples were not processed under strict anaerobic conditions, and when the bacteria were isolated, digestate aliquots had been stored for over 12 months at around –20°C without the use of any protective agents such as glycerol.

Sporulating bacteria like *Clostridium* are among the life forms that are most resistant to extreme temperatures and can withstand multiple cycles of freeze-thawing (Setlow, 2014). In fact, *Bacillus* and *Clostridium* species have been cultured from viable spores found in Antarctic human faecal waste that had been frozen for 40 years (Hughes and Nobbs, 2004).

In comparison, several species of *Bacteroides*, which like *Paludibacter* is a member of the class Bacteroidia, survive freezing in medium for up to 6 weeks at –20°C and only 2 weeks at –10°C (Egwari and Rotimi, 1991). Species of the highly abundant genus *Treponema* are also very susceptible to cold temperatures and do not survive freezing for longer than one month unless they are stored in glycerol at –70°C (Hollander and Nell, 1954). Therefore, it is expected that the number of viable non-spore forming bacteria would have been very low in the digestate samples.

4.4.2 Low abundance of isolated species in metataxonomic dataset

Within the sequencing data associated with the digestate sample that several organisms were isolated from (Chamber 5 from 15th March 2017), the 16S rRNA sequences from five isolates were of very low abundance, while the sequences of the other two organisms were not detected. Still, the sequences that were not detected in this particular sample were found in sequences associated with several other digestate samples. This indicates that the organisms were present in the reactor, however their abundance in the digestate they were isolated from was likely so low their DNA was not included in the sequencing reaction by chance.

Furthermore, the possibility that the isolates have multiple 16S rRNA copies was not investigated, as the 16S rRNA genes were amplified using PCR and then sequenced, meaning that variable domains would have likely been obscured (Cilia, Lafay and Christen, 1996). As species with multiple 16S rRNA genes only show an average divergence of about 1% between copies (Acinas *et al.*, 2004), it is unlikely that such polymorphisms would cause significant misinterpretations about their taxonomy. Still, if multiple divergent 16S rRNA copies were present in the isolates, their reported abundance in the digestate sample they were isolated from (Table 26) may have been underestimated.

4.4.3 Isolate taxonomy

When the taxonomy of the isolated organisms was identified by sequencing their 16S rRNA, it was seen that two phenotypically distinct isolates shared identical sequences. As this comparison for the isolates was only done on the hypervariable region V4, differences might have been found if the whole 16S rRNA gene had been sequenced. For example, the lack of information by only using the 16S rRNA V4 region was also the reason why *Clostridium chromiireducens* had initially been identified as the closest relative to the *Clostridium sp.* isolate, while a later comparison of full 16S rRNA sequences (Chapter 5.2.1.2) placed these organisms further apart.

Regarding the isolates with shared 16S rRNA V4 sequences, distinct bacterial species with 16S rRNA sequence similarities of more than 99.5% have been found and such high commonalities may indicate that the two isolated species diverged relatively recently (Fox, Wisotzkey and Jurtshuk, 1992).

Recommended thresholds to classify bacteria and archaea by 16S rRNA similarity have been proposed to be 98.7% to differentiate species and 94.5% to identify different genera (Yarza et al., 2014). By these measures, isolate 9 (Lachnospiraceae gen. sp.) could likely be classified as a new species as its 16S rRNA V4 region only showed a 97% similarity to the next known species (*Faecalicatena contorta*).

When carrying out the phylogenetic tree analysis of the isolated species, it became apparent that the closest relatives of isolate 6 and 8 were within the genus *Clostridium*, despite being located on very distant branches from the *Clostridium* isolates 1 (*C. butyricum*) and 4 (*Clostridium sp.*). Comparing the 16S rRNA V4 sequence between isolates 6/8 and 1/4, they only shared 81% similarity, which would place them in separate genera by established delimitation thresholds. This observation highlighted issues with the taxonomy of a lot of *Clostridium* species, which was discussed by Rossi-Tamisier *et al.* (2015). In their study, they found that the correlation between previously assigned classifications and 16S rRNA similarity thresholds varied significantly between genera. For example, only 6.4% of species associated with the genus *Clostridium* shared 16S rRNA gene similarities above the threshold, while the worst adherence to the threshold was noted for the species *Spirochaeta* (1.5%) and *Treponema* (2.1%).

Recent advances in molecular analysis have increasingly led to reclassifications to reflect the established identity thresholds for 16S rRNA sequences. This has been especially relevant for the genus *Clostridium,* which was proposed to be restricted to those species that are closely related to *C. butyricum* (Lawson and Rainey, 2016). Reclassifications included species like the important human pathogen *C. difficile* into the genus *Clostridioides* (Lawson *et al.*, 2016), but also the identified close relatives of isolates 9 and

11 within the newly named genera *Lacrimispora* and *Faecalicatena*, which have historically also been associated with the genus *Clostridium* (Cruz-Morales *et al.*, 2019).

4.4.4 Volatile fatty acid production

After the taxonomy of the isolates was resolved, it was investigated whether they acidified their growth medium and produced volatile fatty acids such as acetic acid and butyric acid.

Production of acid was initially observed in cultures of isolate 8 (Ruminococcaceae *gen. sp*.), while the culture of the closely related isolate 6 did not produce a drop in pH, further highlighting their phenotypical differences despite identical 16S rRNA sequences. Interestingly, *Clostridium jeddahense*, the closest known related species to both isolates, has also been reported not to produce any acid from glucose (Lagier *et al.*, 2015). Nevertheless, none of the tested volatile fatty acids were detected in the culture supernatants of isolate 8, suggesting that this isolate may produce other acids such as lactic acid or formic acid, which was not measurable with the column used on the GC-FID machine. The detection threshold for acetic acid and butyric acid using the GC-FID machine was not determined, however the lowest detected amounts were detected within non-inoculated PYG medium, which contained 0.04 g/L of each of the two VFAs.

Cultures of isolate 3 (*Enterococcus sp.*) produced small amounts of acetic acid and studies on bacteria associated with its relative *E. gallinarum* have reported the production of lactic acid, formic acid, acetic acid and CO_2 (Kim *et al.*, 2005).

The highest amount of acetic acid was produced by the isolate with close similarity to *Lacrimispora saccharolytica*. This species is a known producer of acetic acid, ethanol, H_2 and CO_2, that can hydrolyse cellulose and hemicellulose (Murray and Khan, 1983). The isolation of an organism closely related to this species might mean that it also played an important role in the digestion of seaweed, as the described characteristics would certainly be relevant to the processes of hydrolysis and acetogenesis.

The other isolate (9) in the family Lachnospiraceae had 16S rRNA similarities of 96–97% to members of the genus *Faecalicatena*. The closest identified relative (*F. contorta*) is described as non-spore forming and produces acetic acid, formic acid, propionic acid, traces of succinic acid and lactic acid, ethanol and propanol (Holdeman, Cato and Moore, 1971). Furthermore, *F. fissicatena* is also non-sporeforming and produces acetic acid, H_2 and CO_2 (Taylor, 1972) and *F. orotica* has also been described as non-spore forming and mainly produces ethanol and CO_2 with small amounts of acetic acid, lactic acid and formic acid from glucose (Wachsman and Barker, 1954).

As none of the closely related species of this isolate share the characteristics of being spore-forming or producing butyric acid, further analysis may be needed to indicate whether this organism is a new species within the genus *Faecalicatena* (as 16S rRNA V4 similarity may suggest) despite these phenotypical discrepancies or perhaps an entirely novel genus.

Finally, the two isolates that were directly identified as *Clostridium* species and grouped with so-called “true” members of genus *Clostridium “sensu stricto”*, of which *C. butyricum* is the type species (Gupta and Gao, 2009). Therefore, their high production of butyric acid and spore-forming characteristics are typical. Compared to the butyric acid production of other *Clostridium* species grown in batch-fermentation experiments, their production of butyric acid was average (Table 34), however most of the other studies were using pH-controlled fermenters. At glucose concentrations of 5 g/L, the relative amount of butyric acid produced by isolate 4 (*Clostridium sp.*) was higher than in pH-uncontrolled cultures of *C. tyrobutyricum* (Wu and Yang, 2003), however medium composition and culture volumes differed.

Controlling the culture pH would therefore likely increased productivity of the isolated *Clostridium* species, as pH-uncontrolled batch fermentations undergo what is called “acid crash”, which stops the bacteria’s metabolic activity when the pH gets too low (Maddox *et al.*, 2000).

Table 34. Comparing butyric acid yields from various *Clostridium* batch-fermentation experiments reported in the literature.

Author	Organism	Glucose (g/L)	Butyric acid (g/L)	Ratio
He *et al.* (2005)	*C. butyricum* ZJUCB	20	12.25	0.61
Zigová *et al.* (1999)	*C. butyricum* S21	30	7.3	0.24
Fayolle *et al.* (1990)	*C. tyrobutyricum* CIP 1-776	130	45	0.35
Michel-Savin *et al.* (1990)	*C. tyrobutyricum* CNRZ 596	80	44	0.55
Jiang *et al.* (2010)	*C. tyrobutyricum* ATCC 25755	30	13.7	0.46
Luo *et al.* (2017)	*C. tyrobutyricum* ATCC 25755	5	1.57	0.31
	C. tyrobutyricum ATCC 25755	15	13.6	0.35
Wu *et al.* (2003)	*C. tyrobutyricum* ATCC 25755	50	20.4	0.38
Patel *et al.* (1988)	*C. populeti*	57	6.3	0.11
This study	Isolate 1 (*C. butyricum*)	5	1.40	0.28
	Isolate 4 (*Clostridium sp.*)	5	1.94	0.39
	Isolate 9 (Lachnospiraceae *gen. sp.*)	5	0.46	0.09

The production of acetic acid and butyric acid is also common in other species of bacteria, some of which can display much higher yields than observed here. For example, within the family Acetobacteraceae, species of *Acetobacter* and *Komagataeibacter* release acetic acid into the growth medium, where it accumulates to a maximum of 5–10% and 10–20% (v/v), respectively (Gomes *et al.*, 2018). While the production of butyric acid is not a feature of wild-type *E. coli*, engineered strains expressing butyrate pathway genes from *C. acetobutylicum* and a transenoyl-CoA reductase from *T. denticola* have produced yields as high as 28 g/L (Wang *et al.*, 2019) and other studies have extended similar work by engineering *E. coli* for the production of butanol, producing yields of 20 g/L butyric acid (Jawed *et al.*, 2020).

4.4.5 Antibiotic resistance

After the two *Clostridium* isolates were selected to be studied further due to their relatively high production of butyric acid, their resistance to commonly used antibiotics was tested to select suitable resistance genes to select for genetically engineered strains. Noticeably, the isolate identified as *C. butyricum* did not show resistance to chloramphenicol, in contrast to the type strains *C. butyricum* ATCC 19398 and NCIB 7423 (Sebald, 1994). Still, the chloramphenicol resistance of these type strains is uncommon for *C. butyricum*, as confirmed susceptibilities of the *C. butyricum* isolate matched results from a study of 39 *C. butyricum* isolates, where all strains were inhibited by chloramphenicol concentrations between 0.5 and 2 µg/mL and only two strains were resistant to erythromycin (Ferraris, Butel and Aires, 2010).

Previously, Popoff and Truffaut (1985) studied the antibiotic resistance of 50 strains of *C. butyricum* and 14 strains of *C. beijerinckii* (close relative to isolate 4 by 16S rRNA similarity). Unlike the strain of *C. butyricum* that was isolated from digestate, all tested *C. butyricum* isolates were susceptible to kanamycin and a large majority (72%) were resistant to ampicillin. Here, all strains were also susceptible to chloramphenicol.

For *C. beijerinckii*, none of the investigated isolates displayed any antibiotic resistance, which coincides with the comparison of the two *Clostridium* isolates, where isolate 4 (*Clostridium sp.*) was only resistant to cycloserine and low concentrations of spectinomycin (not tested in the two studies), while *C. butyricum* was resistant to moderate concentrations of cycloserine, kanamycin, spectinomycin and thiamphenicol.

4.4.6 Variations in *C. butyricum* colony morphology

When *C. butyricum* was grown on agar plates, it was observed that the isolate had two different morphologies, pale with smaller, motile cells and opaque with large spore-forming cells.

This observation of several different colony morphologies mirrors findings for other related species, such as in *C. perfringens*, where either translucent or opaque spore-forming colonies were observed (Fisher Costerton, 1956), or *C. difficile*, where different colony morphotypes are associated with genetic switches that control the expression of flagella and toxins (Anjuwon-Foster and Tamayo, 2017).

In liquid cultures, Zhou *et al.* (2020) described that *C. butyricum* goes through spontaneous oscillatory cycles, that are characterised by changes in cell metabolism as well as morphology. The different morphologies were described as either slimmer and elongated or inhomogeneous in length, aggregated and spore-forming. These observations mirror the two distinct states that were described here, so that it is possible that the differences in colony morphologies represent these oscillation cycles.

5 Further study of the two *Clostridium* isolates

5.1 Introduction

The genus *Clostridium* contains around 200 species, which tend to show as Gram-positive rods. However, the genus is phylogenetically and phenotypically heterogeneous and includes anaerobic and non-anaerobic species as well as spore formers and non-spore formers (Public Health England, 2016). Several species within the genus are of medical importance (e.g. *C. botulinum*, *C. novyi*, *C. haemolyticum*) however most species are non-pathogenic. In recent years, the phylogenetic heterogenicity of this genus has been recognised through differences in 16S rRNA sequences and the genus is now defined by its type species, *Clostridium butyricum* (Lawson and Rainey, 2016). This has led to an extensive re-classification of species that were historically associated with the genus due to phenotypical similarities. This effort to define the genus in its newly described narrow sense is still ongoing, which is why its members with close relationship to *C. butyricum* are also referred to as *Clostridium sensu stricto*. Members of the genus are recognised for their high production of hydrogen and VFAs, specifically butyric acid and acetic acid (e.g. *C. butyricum*, *C. tyrobutyricum*) but also include species that produce acetone and butanol, such as *C. acetobutylicum and* C. beijerinckii (Udaondo *et al.*, 2017).

C. acetobutylicum has been used industrially since it was first isolated by Chaim Weizmann in 1915. The species was recognised for its potential to ferment a wide variety of starchy feedstocks to produce acetone and butanol (Gabriel, 1928). This process, called acetone-butanol-ethanol (ABE) fermentation was initially used to produce acetone required to make cordite, an explosive heavily used in the First World War, while butanol was considered a by-product, for which there only was demand a few years later when the developing car industry increasingly required it as a lacquer solvent (Jones and Woods, 1986). The use of *C. acetobutylicum* for ABE fermentation remained the method of choice for acetone and butanol production until the

1960s, when it was replaced by chemical processes as ABE fermentation became less cost-effective. This was due to the rise of the petrochemical industry as well as the competitive buying of feedstocks that were now more commonly used as cattle feed (Hastings, 1978). Nevertheless, the increasing scarcity of crude oil and the search for green alternatives to fossil fuels have shifted ABE fermentation back into the focus of researchers and investors alike.

While *C. acetobutylicum* and *C. beijerinckii* produce butanol directly, higher production yields are limited by butanol toxicity (Bowles and Ellefson, 1985). Therefore, there is also an increased interest in bacteria that metabolise feedstocks into butyric acid, which can be used as a precursor chemical for butanol and a wide range of other compounds (Dwidar *et al.*, 2012). The production of butyric acid, acetic acid, butanol, acetone and other compounds occurs through the butanoate pathway (Fig. 18). In this fermentative process, 1 mol of glucose is converted to 4 moles of ATP when acetic acid is the end product, while butyric acid production only yields 3 moles of ATP. Furthermore, less NADH + H^+ is reduced to NAD^+ during the biosynthesis of acetic acid, compared to the production of more reduced end products (e.g. lactic acid, butyric acid, ethanol and butanol). Therefore, the production of acetic acid in *Clostridium* is associated with a higher production of hydrogen gas through ferredoxin-NAD^+ reductase and hydrogenase (Levin, Pitt and Love, 2004).

Other examples of the industrial use of *Clostridium* species include the newly designed gas fermentation process that is being commercialised by the American company LanzaTech, in which *Clostridium autoethanogenum* has been shown to grow and produce ethanol from steal mill off-gas containing carbon monoxide as well as CO_2 and H_2 (Heffernan *et al.*, 2020)

5.1.1 Aims and objectives

After the two *Clostridium* isolates were identified for their high butyric acid yields, the objective was to further investigate their potential as VFA-producing species and study other characteristics such as their ability to produce hydrogen gas. Also, the full genome of both species was sequenced in order to find out more about their taxonomy within the genus *Clostridium sensu stricto* and their metabolism, with a focus on usable carbon sources and their ability to produce other fermentation products such as acetone and butanol. Furthermore, the aim was to identify potential genes in the butanoate pathway that may be a target to engineer these organisms for the enhanced production of VFAs, which would be facilitated by having full genome sequences.

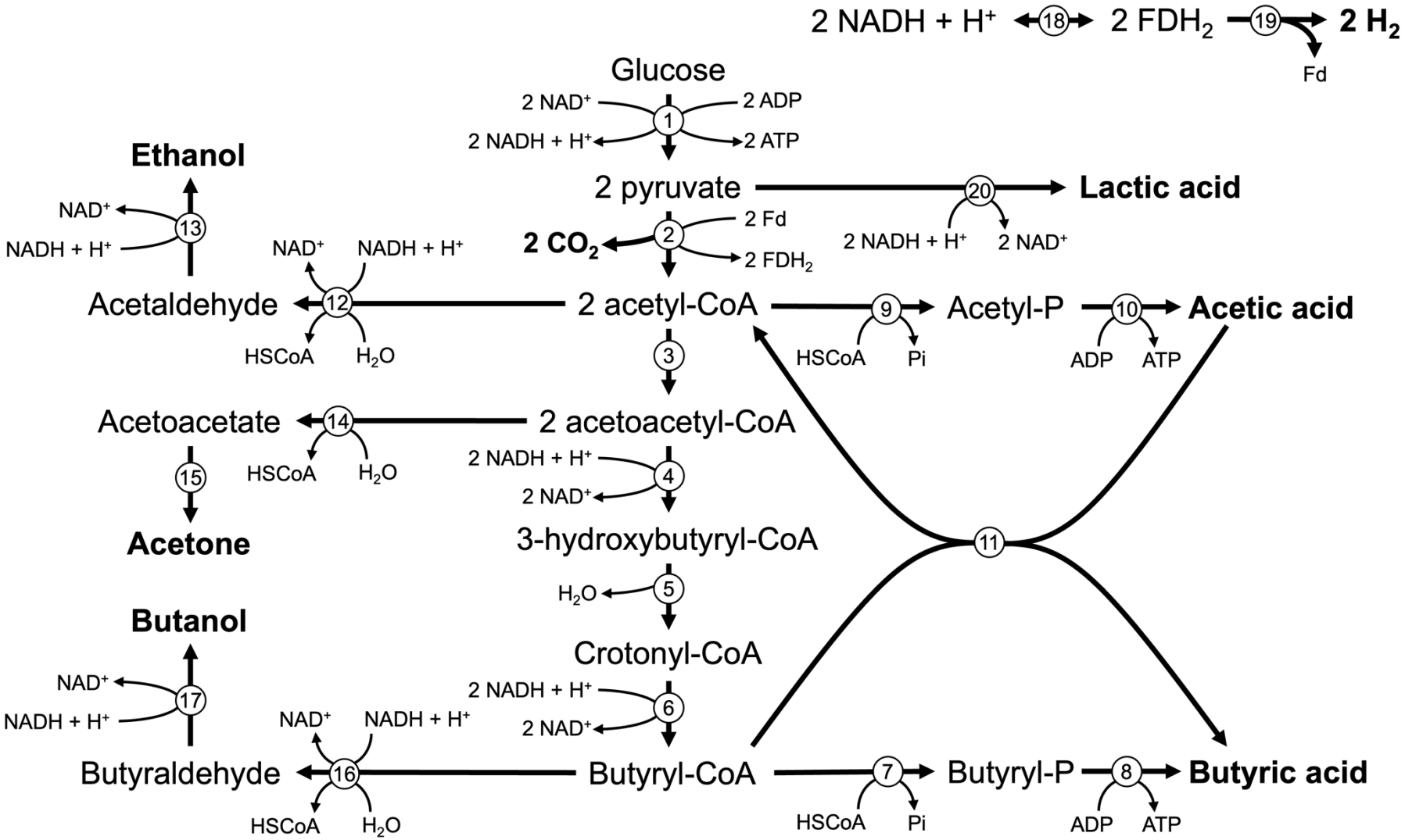


Figure 18. Butanoate pathway in Clostridia. (1) Embden-Meyerhof-Parnas pathway, (2) pyruvate-ferredoxin oxidoreductase, (3) acetyl CoA-acetyl transferase (thiolase), (4) 3-hydroxybutyryl CoA dehydrogenase, (5) 3-hydroxybutyryl-CoA dehydratase, (6) butyryl-CoA dehydrogenase, (7) phosphotransbutyrylase, (8) butyrate kinase, (9) phosphotransacetylase, (10) acetate kinase, (11) butyryl-CoA-acetate transferase, (12) acetaldehyde dehydrogenase, (13) ethanol dehydrogenase, (14) acetoacetyl-CoA hydrolase, (15) acetoacetate decarboxylase, (16) butyraldehyde dehydrogenase, (17) butanol dehydrogenase, (18) ferredoxin-NAD^+ reductase, (19) hydrogenase, (20) lactate dehydrogenase. The figure was modified from Dwidar *et al.* (2012) and further information about the involved enzymes was retrieved from the KEGG butanoate metabolism pathway (https://www.genome.jp/pathway/ec00650). Note that Acetyl/Butyryl-P stands for Acetyl/Butyryl phosphate..

5.2 Results

5.2.1 Full genome sequencing of *Clostridium* isolates

The genomes of both *Clostridium* isolates were fully sequenced to find out more about their lineage and subsequently facilitate their engineering.

The two genome assemblies for *C. butyricum* and the unidentified *Clostridium* species were made up of 285 (4.6 Mb) and 191 (7.4 Mb) contigs, respectively (Table 35). The PATRIC genome browser identified 4234 and 6811 coding regions for *C. butyricum* and *Clostridium sp.*, respectively.

Table 35. Genome assembly statistics.

	C. butyricum	***Clostridium sp.***
Contigs ≥ 0 bp	285	191
Total length ≥ 0 bp (bp)	4,573,330	7,385,832
Contigs ≥ 1 kb	72	129
Total length ≥ 1 kb (bp)	4,514,816	7,358,485
Largest contig (bp)	420,944	376,872
GC (%)	28.70	29.73
N50	126,823	108,242
N75	72,242	78,032
L50	10	21
L75	22	41

The NCBI database was then used to compare the average size and GC content of published *Clostridium* genomes to the genomes of the two isolates.

The genome of the *C. butyricum* isolate was 0.08 Mb larger (4.57 Mb) than the average *C. butyricum* genome (4.49 Mb ± 0.43 SD, n = 59), while its GC content of 28.7% was 0.17% less compared to the average (28.9% ± 1.9 SD). The genome of the isolated *C. butyricum* species is therefore of average size and GC content.

With a size of 7.39 Mb, the genome of the second isolated *Clostridium* species is atypically large. Compared to the published reference genomes of other related *Clostridium sensu stricto* species, it is 3.38 Mb larger than the genus average (4.01 Mb) and 724 kb larger than the genome of its close relative *C. saccharoperbutylacetonicum*, which is the NCBI database's largest reference genome within this genus.

Its GC content (29.73%) is very similar to the published genomes of the closely related species *C. saccharoperbutylacetonicum* (29.5%) and *C. beijerinckii* (5.95 Mb, 29.7%), which is the 4th largest genome in the genus.

5.2.1.1 Codon usage

Codon usage between the two *Clostridium* isolates was investigated to determine whether genes that were to be expressed in either species would potentially require codon optimisation. Relative use of codons in all open reading frames (ORFs) was calculated using the Sequence Manipulation Suite (Stothard, 2000) and while some differences were seen between the species, most of the relative codon frequencies were within 5 percentage points and no species-exclusive codon bias was observed (Table 36). It was therefore decided that genes from *C. butyricum* would not be codon optimised for transformation into the other isolate and vice-versa, as the species are very closely related.

Table 36. Codon usage in ORFs of *C. butyricum (C. bu)* and *Clostridium sp. (C. sp).*

		C. bu.	*C. sp.*			*C. bu.*	*C. sp.*
AA	**Codon**	**%**	**%**	**AA**	**Codon**	**%**	**%**
Ala	**GCG**	5	5	**Pro**	**CCG**	5	6
	GCA	48	46		**CCA**	48	48
	GCT	39	40		**CCT**	37	38
	GCC	9	9		**CCC**	9	8
Cys	**TGT**	59	65	**Gln**	**CAG**	32	29
	TGC	41	35		**CAA**	68	71
Asp	**GAT**	85	86	**Arg**	**AGG**	23	18
	GAC	15	14		**AGA**	63	62
Glu	**GAG**	18	21		**CGG**	3	3
	GAA	82	79		**CGA**	5	5
Phe	**TTT**	71	79		**CGT**	5	9
	TTC	29	21		**CGC**	2	3
Gly	**GGG**	11	9	**Ser**	**AGT**	24	22
	GGA	49	50		**AGC**	15	11
	GGT	30	30		**TCG**	2	3
	GGC	10	11		**TCA**	25	28
His	**CAT**	78	79		**TCT**	22	27
	CAC	22	21		**TCC**	11	9
Ile	**ATA**	41	47	**Thr**	**ACG**	5	5
	ATT	45	44		**ACA**	43	42
	ATC	14	9		**ACT**	38	41
Lys	**AAG**	29	28		**ACC**	14	12
	AAA	71	72	**Val**	**GTG**	10	11
Leu	**TTG**	13	13		**GTA**	37	41
	TTA	41	45		**GTT**	45	42
	CTG	6	5		**GTC**	8	6
	CTA	10	12	**Trp**	**TGG**	100	100
	CTT	25	21	**Tyr**	**TAT**	77	82
	CTC	5	4		**TAC**	23	18
Met	**ATG**	100	100	**Stop**	**TGA**	39	21
Asn	**AAT**	78	82		**TAG**	13	22
	AAC	22	18		**TAA**	48	57

5.2.1.2 BLAST search of 16S rRNA and housekeeping genes

To further assess the relatedness of the isolated *Clostridium* species to known organisms, their full 16S rRNA sequences as well as the sequences of five other genes were used in a BLAST search to identify closely related species. Furthermore, the species boundary thresholds for 16S rRNA (98.7%; Stackebrandt and Ebers, 2006) and other genes (95% average nucleotide identity (ANI); Richter and Rosselló-Móra, 2009) were considered to determine whether the two isolates are novel species within the genus *Clostridium.*

In the analysis, it was found that the 16S rRNA sequence of the isolate previously identified as *C. butyricum* using the SILVA database was a 100% match to various *C. butyricum* strains (Table 37). Comparing the gene sequences of various housekeeping genes (Tables 38–42) also produced close matches in the region of 99–100%, mostly with the strain *C. butyricum* TK520. The average nucleotide identity (ANI) with *C. butyricum* TK520 was calculated from the full genome sequences of both organisms and found to be 99%. The isolate was therefore identified to be a strain of *C. butyricum*.

The closest 16S rRNA match for the other *Clostridium* isolate was *C. saccharoperbutylacetonicum* with 97.75% identity (Table 43). The tested housekeeping genes generally matched most closely to *C. beijerinckii* with 80–92% identity (Tables 44–48). The ANI shared with these two species was determined to be 82% for *C. beijerinckii* and 80% compared to *C. saccharoperbutylacetonicum*.

Furthermore, it was confirmed that neither of the two isolates possess genes to produce *C. difficile* toxin or botulinum toxin, nor restriction modification systems found in species such as *C. acetobutylicum* and *C. beijerinckii*, which would hinder their possible use in industrial settings as well as the introduction of unmethylated plasmids without prior generation of knock-out mutants (Dong *et al.*, 2010).

All in all, it can therefore be argued that this isolate represents a novel species within the genus *Clostridium.* As this isolate was initially recognised for its increased production of butyric acid, the name ***Clostridium pluributyricum*** (“pluri-“ meaning “more”) was chosen to refer to the isolate.

Table 37. *C. butyricum* 16S rRNA (full gene), BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium butyricum strain MSK	100%	100%	NR_102852.1
Clostridium butyricum strain LQ25	100%	100%	NR_026100.1
Clostridium butyricum strain X-44	100%	100%	NR_112174.1
Clostridium butyricum strain CDC_51208	100%	100%	NR_026409.1
Clostridium butyricum strain TK520	100%	100%	NR_119035.1

Table 38. *C. butyricum* DNA polymerase I, BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium butyricum strain TK520	100%	99.10%	CP016332.1
Clostridium butyricum strain TOA	100%	99.10%	CP014704.1
Clostridium butyricum strain JKY6D	100%	99.10%	CP013352.1
Clostridium butyricum strain KNU-L09	100%	99.10%	CP013252.1
Clostridium butyricum strain CDC_51208	100%	98.72%	CP013239.1

Table 39. *C. butyricum* DnaK (hsp70), BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium butyricum strain TK520	100%	99.84%	CP016332.1
Clostridium butyricum strain TOA	100%	99.84%	CP014704.1
Clostridium butyricum strain JKY6D1	100%	99.84%	CP013352.1
Clostridium butyricum strain KNU-L09	100%	99.84%	CP013252.1
Clostridium butyricum strain CDC_51208	100%	99.78%	CP013239.1

Table 40. *C. butyricum* dihydrofolate reductase, BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium butyricum strain TK520	100%	99.19%	CP016332.1
Clostridium butyricum strain TOA	100%	99.19%	CP014704.1
Clostridium butyricum strain JKY6D1	100%	99.19%	CP013352.1
Clostridium butyricum strain KNU-L09	100%	99.19%	CP013252.1
Clostridium butyricum strain CDC_51208	100%	98.78%	CP013239.1

Table 41. *C. butyricum* GroEL (hsp60), BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium butyricum strain TK520	100%	100.00%	CP016332.1
Clostridium butyricum strain TOA	100%	100.00%	CP014704.1
Clostridium butyricum strain JKY6D1	100%	100.00%	CP013352.1
Clostridium butyricum strain KNU-L09	100%	100.00%	CP013252.1
Clostridium butyricum strain CDC_51208	100%	99.75%	CP013239.1

Table 42. *C. butyricum* guanylate kinase, BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium butyricum strain TK520	100%	100.00%	CP016332.1
Clostridium butyricum strain TOA	100%	100.00%	CP014704.1
Clostridium butyricum strain JKY6D1	100%	100.00%	CP013352.1
Clostridium butyricum strain KNU-L09	100%	100.00%	CP013252.1
Clostridium butyricum strain CDC_51208	100%	99.84%	CP013239.1

Table 43. *C. pluributyricum* 16S rRNA (full gene), BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium saccharoperbutylacetonicum strain N1-504	99%	97.75%	CP016087.1
Clostridium saccharobutylicum strain NCP 195	99%	97.69%	CP016092.1
Clostridium saccharobutylicum strain NCP 258	99%	97.69%	CP016091.1
Clostridium saccharobutylicum strain BAS/B3/SW/136	99%	97.69%	CP016089.1
Clostridium saccharobutylicum strain NCP 200	99%	97.69%	CP016086.1

Table 44. *C. pluributyricum* DNA polymerase I, BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium beijerinckii ATCC 35702	100%	84.50%	NR_102852.1
Clostridium beijerinckii NCIMB 8052	100%	84.50%	NR_026100.1
Clostridium beijerinckii isolate C. beijerinckii DSM 6423	100%	84.42%	NR_112174.1
Clostridium beijerinckii strain NCIMB 14988	100%	84.15%	NR_026409.1
Clostridium beijerinckii NRRL B-598	100%	84.08%	NR_119035.1

Table 45. *C. pluributyricum* DnaK (hsp70), BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium beijerinckii isolate C. beijerinckii DSM 6423	100%	92.81%	LN908213.1
Clostridium sp. MF28	100%	92.76%	CP014331.1
Clostridium beijerinckii NRRL B-598	100%	92.71%	CP011966.3
Clostridium beijerinckii ATCC 35702	100%	92.71%	CP006777.1
Clostridium beijerinckii NCIMB 8052	100%	92.71%	CP000721.1

Table 46. *C. pluributyricum* dihydrofolate reductase, BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium beijerinckii isolate WB53	98%	80.29%	CP029329.1
Clostridium saccharobutylicum strain NCP 195	98%	79.06%	CP016092.1
Clostridium saccharobutylicum strain BAS/B3/SW/136	98%	79.06%	CP016089.1
Clostridium saccharobutylicum strain NCP 200	98%	79.06%	CP016086.1
Clostridium saccharobutylicum DSM 13864	98%	79.06%	CP006721.1

Table 47. *C. pluributyricum* GroEL (hsp60), BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium beijerinckii strain NCIMB 14988	100%	92.19%	CP010086.2
Clostridium beijerinckii ATCC 35702	100%	92.13%	CP006777.1
Clostridium beijerinckii NCIMB 8052	100%	92.13%	CP000721.1
Clostridium beijerinckii strain BAS/B3/I/124	100%	92.07%	CP016090.1
Clostridium beijerinckii NRRL B-598	100%	92.00%	CP011966.3

Table 48. *C. pluributyricum* guanylate kinase, BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium beijerinckii isolate C. beijerinckii DSM 6423	100%	85.96%	LN908213.1
Clostridium beijerinckii NRRL B-598	100%	85.65%	CP011966.3
Clostridium sp. MF28	100%	85.34%	CP014331.1
Clostridium beijerinckii strain NCIMB 14988	100%	85.34%	CP010086.2
Clostridium beijerinckii strain BAS/B3/I/124	100%	85.18%	CP016090.1

The five housekeeping genes were selected based on their mention in Cai *et al.* (2007), Cleenwerck *et al.* (2010) and Vandecasteele *et al.* (2001).

5.2.1.3 Genome comparison dot plots

The D-GENIES tool was used to align the isolates' genome contigs to the genomes of their closest relatives identified by BLAST search. The resultant dot plots are a visual aid to compare these sequence and see features such as insertions, deletions, repeats or inversions (Cabanettes and Klopp, 2018).

Comparing the genome of the isolate identified as *C. butyricum* to the genome of the closely related strain TK520 (Fig. 19) revealed that their genomes are essentially co-linear, highly similar and that the second *C. butyricum* chromosome (820 kb) is included in the isolate's sequencing data. One noticeable mismatch between the isolate's genome and TK520 was an inverted 44 kb segment, which matched with the second chromosome of TK520, while the surrounding sequence matched its chromosome 1. Considering that this mismatch was not visible in comparison with the botulinum neurotoxin-producing strain CDC 51208 (Fig. 20), which is the isolate's more distant relative compared to strain TK520 (Tables 38–42), it indicates that a recombination event has taken place in TK520 rather than in the isolated species.

Larger differences were seen when the genome data from *C. pluributyricum* was aligned against the genomes of the closest relatives *C. saccharoperbutylacetonicum* (Fig. 21), *C. saccharobutylicum* (Fig. 22) and *C. beijerinckii* (Fig. 23). All comparisons showed large mismatches, especially when comparing to the two species that were the closest matches by 16S rRNA sequence similarity (*C. saccharoperbutylacetonicum* and *C. saccharobutylicum*). The alignment to the *C. beijerinckii* assembly showed the highest degree of co-linearity out of the three comparisons, which is noteworthy as *C. beijerinckii* was identified as *C. pluributyricum*'s closest relative when BLAST searches were carried out on selected housekeeping genes. Noticeably, while mismatches were overall smaller than in the comparison to the other two species, the *C. pluributyricum* genome did not match to a large semi-continuous section of the *C. beijerinckii* genome,

spanning about 750 kb in length. Still, out of the three examined species, it seems like *C. beijerinckii* is the closest known relative to *C. pluributyricum.*

Nevertheless, as the short-read Illumina sequencing splits the genomes into several contigs, this may make the alignment performed by D-GENIES less accurate in predicting rearrangements or translocations that span several contigs. This could be alleviated by using long read sequencing technologies (e.g. Nanopore) and aligning the contigs to a scaffold.

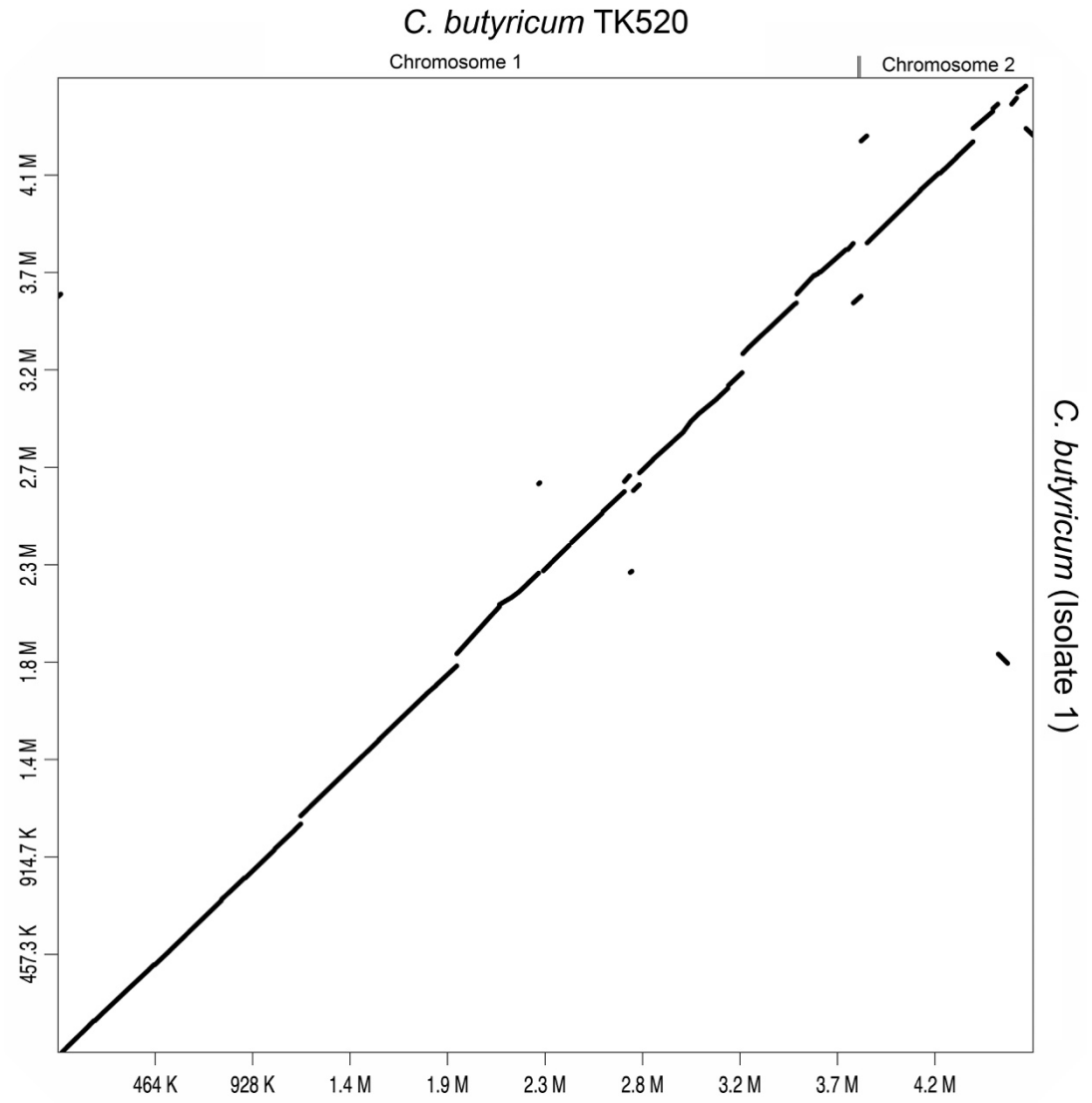


Figure 19. D-GENIES comparative genome plot, showing the alignment of genome contigs from the isolated *C. butyricum* (query) to the full genome of *C. butyricum* TK520 (reference).

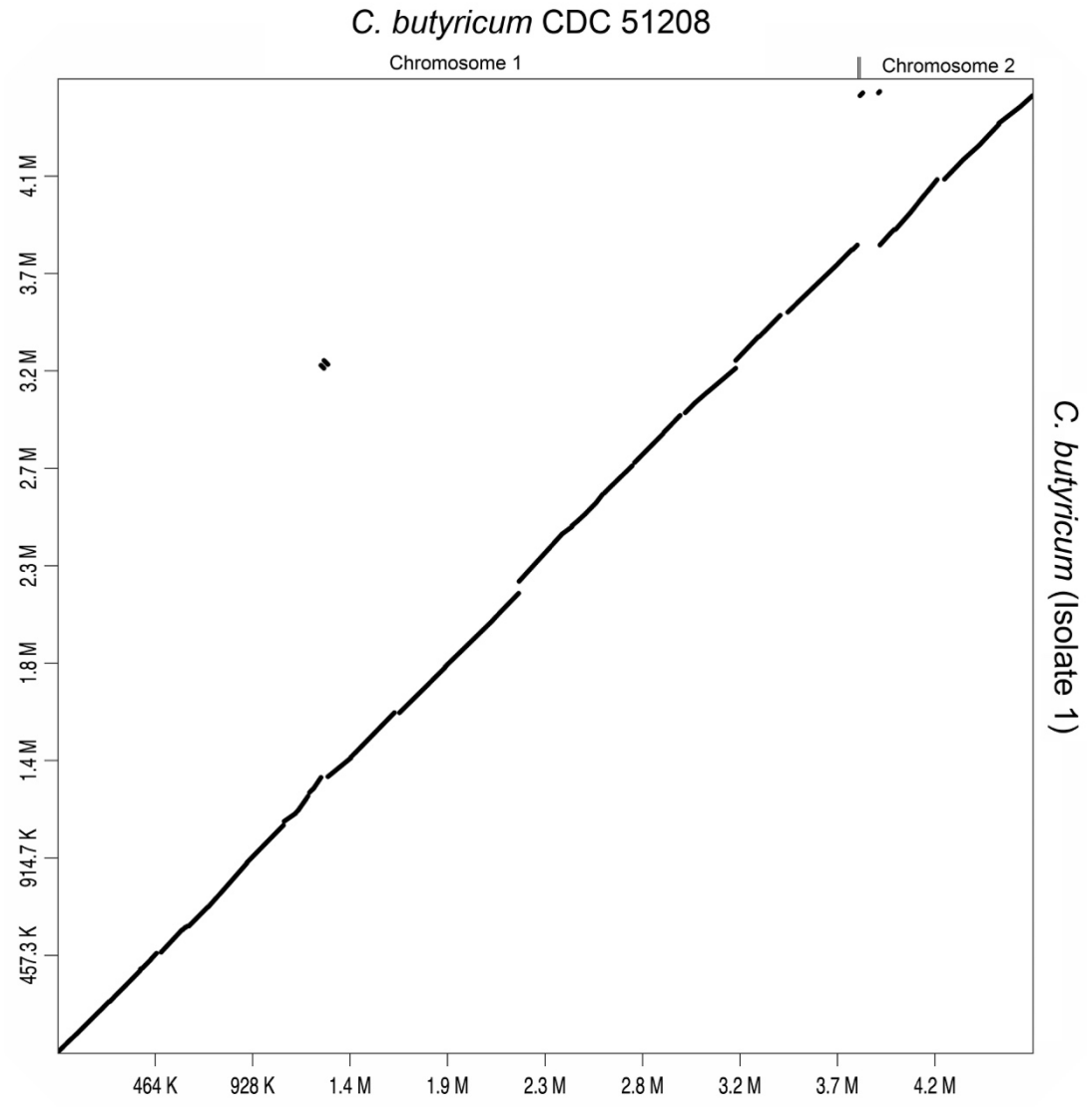


Figure 20. D-GENIES comparative genome plot, showing the alignment of genome contigs from the isolated *C. butyricum* (query) to the full genome of *C. butyricum* CDC 51208 (reference).

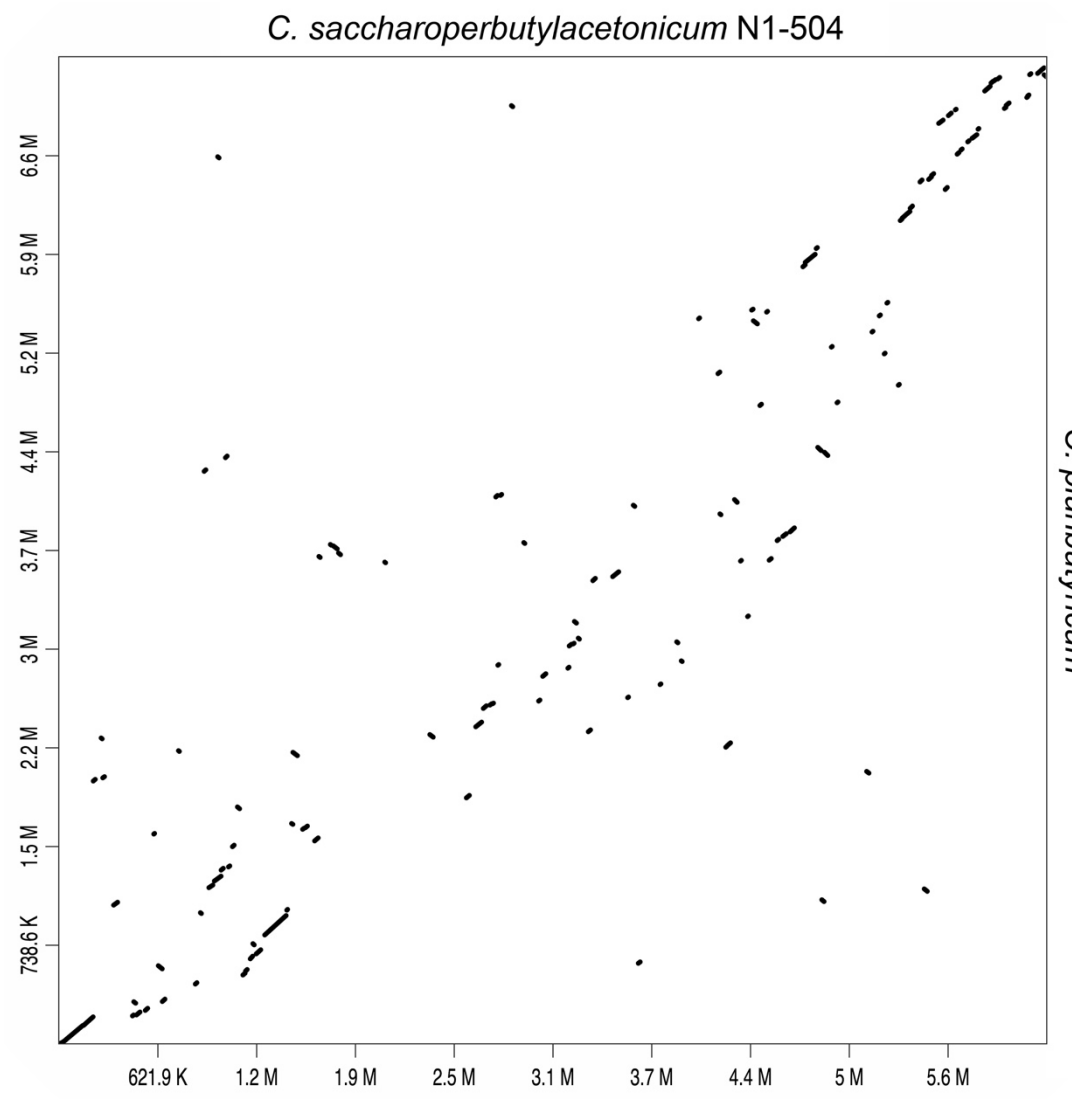


Figure 21. D-GENIES comparative genome plot, showing the alignment of genome contigs from the isolated *C. pluributyricum* (query) to the full genome of *C. saccharoperbutylacetonicum* N1-504 (reference).

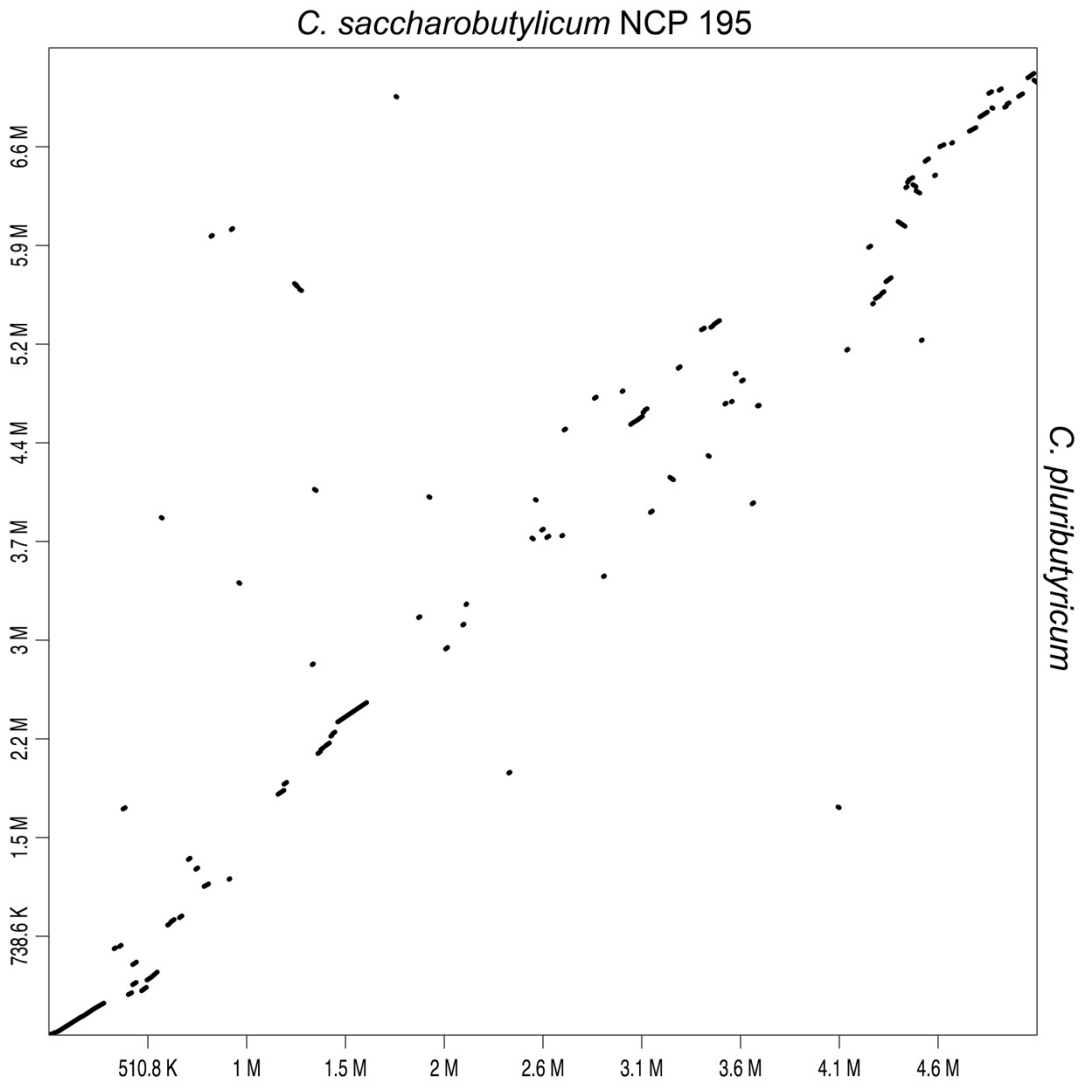


Figure 22. D-GENIES comparative genome plot, showing the alignment of genome contigs from the isolated *C. pluributyricum* (query) to the full genome of *C. saccharobutylicum* NCP 195 (reference).

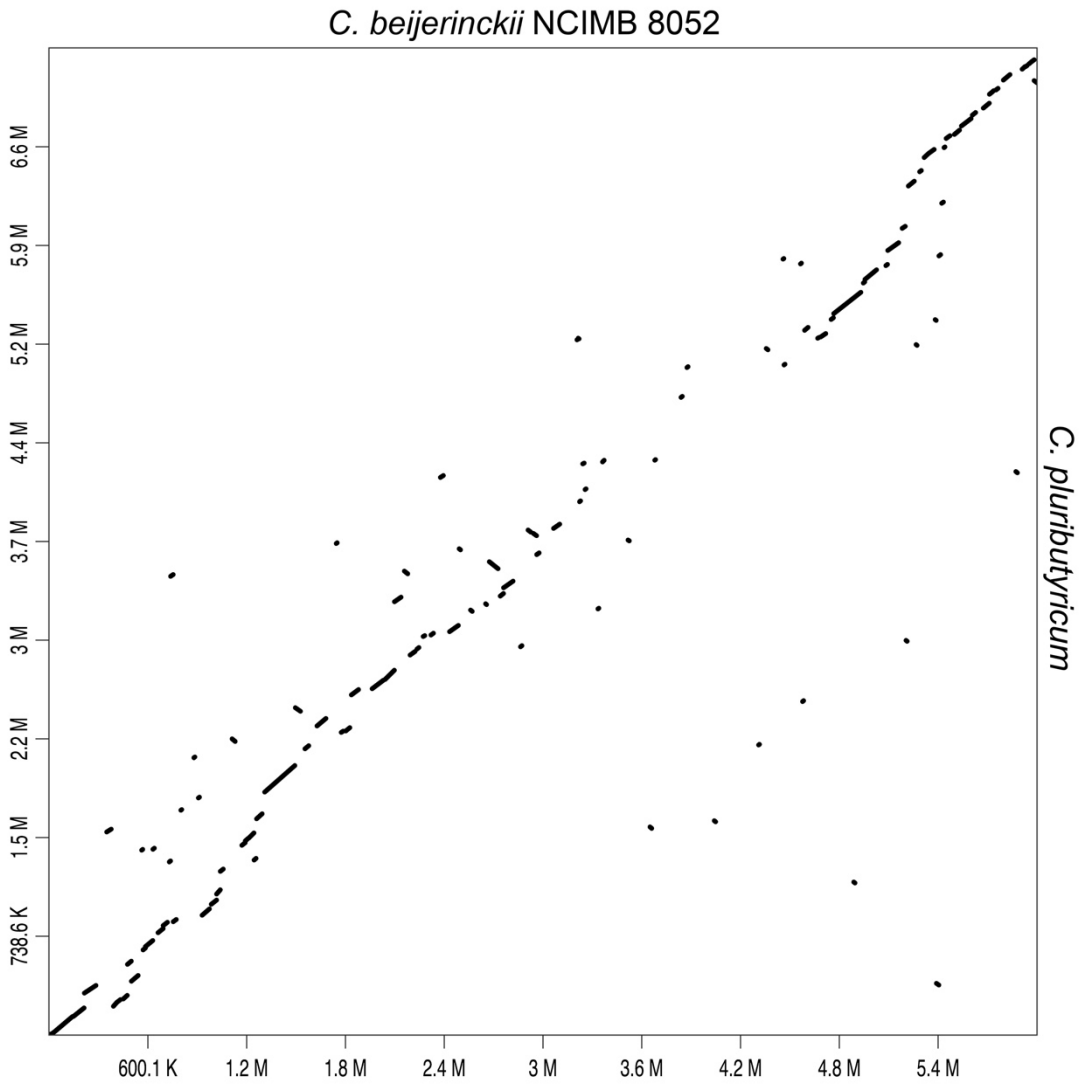


Figure 23. D-GENIES comparative genome plot, showing the alignment of genome contigs from the isolated *C. pluributyricum* (query) to the full genome of *C. beijerinckii* NCIMB 8052 (reference).

5.2.1.4 Butanoate pathway genes

The close relatedness of the isolated *Clostridium* species to known industrial producers of acetone and butanol (i.e. *C. beijerinckii* and *C. saccharobutylicum*) led to the question whether they are also capable of producing these fermentation products. BLAST-searching for orthologues of acetoacetate decarboxylase and butyraldehyde dehydrogenase, key genes in the production of acetone and butanol, respectively, returned no results in both *Clostridium* isolates, indicating the absence of these genes. Furthermore, both isolates were found to possess genes to produce formic acid (formate c-acetyltransferase, EC 2.3.1.54) and lactic acid (lactate dehydrogenase, EC 1.1.1.27) from pyruvate. Unlike *C. butyricum*, C*. pluributyricum* was also found to possess an acetate CoA-transferase gene (EC 2.8.3.8) as well as a butyrate-acetoacetate CoA-transferase (EC 2.8.3.9), which may enable it to produce butyrate from acetate and acetoacetate.

Between the two isolates, differences were found in the number of genes identified as butyrate kinase (*buk*, EC 2.7.2.7) and [FeFe] hydrogenase (*hydA,* EC 1.12.7.2), which play a role in the production of butyric acid and hydrogen, respectively. Genes were identified through the PATRIC genome feature browser as well as BLAST search against *buk* and *hydA* orthologues and were given designations according to the contigs they were found in.

C. butyricum was found to possess at least three single-subunit [FeFe] hydrogenase genes, while *C. pluributyricum* was found to only have one (Table 49). Furthermore, the *C. butyricum* isolate's genome was also found to contain a three-subunit bifurcating [FeFe] hydrogenase, which was however not studied further as part of this project.

Comparing the butyrate kinases of *C. butyricum* and *C. pluributyricum,* the single butyrate kinase identified in *C. butyricum* (*Cbu-c15-buk*) was found to have two orthologs in *C. pluributyricum* with *Cpl-c42-buk* being a closer match (100% query cover, 82.58% identity) than *Cpl-c3-buk* (87% query cover, 65.71% identity). The *buk* and *hydA* genes from both species were compared to the BLAST database to identify closely related gene orthologs in other species (Tables 50–56).

Table 49 Butyrate kinase and single-subunit [FeFe] hydrogenase genes identified in genome sequences of both *Clostridium* isolates and selected for further study in this project.

Gene	Name	Species	Length (nt)	Length (aa)
Butyrate kinase (buk)	*Cbu-c15-buk*	*C. butyricum*	1068	355
Butyrate kinase (buk)	*Cpl-c3-buk*	*C. pluributyricum*	1071	356
Butyrate kinase (buk)	*Cpl-c42-buk*	*C. pluributyricum*	1068	355
[FeFe] hydrogenase	*Cbu-c2-hydA*	*C. butyricum*	1371	456
[FeFe] hydrogenase	*Cbu-c16-hydA*	*C. butyricum*	1488	495
[FeFe] hydrogenase	*Cbu-c24-hydA*	*C. butyricum*	1725	574
[FeFe] hydrogenase	*Cpl-c67-hydA*	*C. pluributyricum*	1491	496

Table 50. *C. butyricum* butyrate kinase *Cbu-c15-buk*, BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium butyricum strain CBUT	100%	99.91%	CP073277.1
Clostridium butyricum strain 16-3	100%	99.72%	CP053292.1
Clostridium butyricum strain CFSA3989	100%	99.72%	CP033249.1
Clostridium butyricum strain CFSA3987	100%	99.72%	CP033247.1
Clostridium butyricum NBRC 13949	100%	99.72%	AP019716.1

Table 51. *C. pluributyricum* butyrate kinase *Cpl-c3-buk*, BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium drakei strain SL1	99%	83.19%	CP020953.1
Clostridium scatologenes strain ATCC 25775	99%	82.72%	CP009933.1
Clostridium fermenticellae strain JN500901	100%	82.54%	CP032416.1
Clostridium carboxidivorans P7	99%	82.35%	CP011803.1
Clostridium sp. AWRP	97%	79.89%	CP029758.2

Table 52. *C. pluributyricum* butyrate kinase *Cpl-c42-buk*, BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium beijerinckii isolate WB53	100%	89.61%	CP029329.1
Clostridium sp. C5S11	100%	89.51%	AP024849.1
Clostridium diolis strain DSM 15410	98%	89.69%	CP043998.1
Clostridium beijerinckii strain DSM 791	98%	89.69%	CP073653.1
Clostridium sp. MF28	98%	89.59%	CP014331.1

Table 53. *C. butyricum* [FeFe] hydrogenase *Cbu-c2-hydA*, BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium butyricum NBRC 13949	100%	100%	AP019716.1
Clostridium butyricum strain DSM 10702	100%	100%	CP040626.1
Clostridium butyricum strain 16-3	100%	99.85%	CP053292.1
Clostridium butyricum strain 4-1	100%	99.85%	CP039705.1
Clostridium butyricum strain S-45-5	100%	99.85%	CP030775.1

Table 54. *C. butyricum* [FeFe] hydrogenase *Cbu-c16-hydA*, BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium butyricum strain 29-1	100%	100%	CP039702.1
Clostridium butyricum strain TK520	100%	100%	CP016332.1
Clostridium butyricum strain TOA	100%	100%	CP014704.1
Clostridium butyricum strain JKY6D1	100%	100%	CP013352.1
Clostridium butyricum strain KNU-L09	100%	100%	CP013252.1

Table 55. *C. butyricum* [FeFe] hydrogenase *Cbu-c24-hydA*, BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium butyricum strain 29-1	100%	100%	CP039702.1
Clostridium butyricum strain TK520	100%	100%	CP016332.1
Clostridium butyricum strain TOA	100%	100%	CP014704.1
Clostridium butyricum strain KNU-L09	100%	100%	CP013252.1
Clostridium butyricum strain CBUT	100%	100%	CP073277.1

Table 56. *C. pluributyricum* [FeFe] hydrogenase *Cpl-c67-hydA*, BLAST results (Top 5).

Description	Query cover	Ident	Accession
Clostridium beijerinckii isolate DSM 6423	100%	85.92%	LN908213.1
Clostridium beijerinckii ATCC 35702	100%	84.17%	CP006777.1
Clostridium beijerinckii strain CBEI	100%	84.17%	CP073279.1
Clostridium beijerinckii NCIMB 8052	100%	84.17%	CP000721.1
Clostridium beijerinckii strain ASCUSBR67	100%	84.04%	CP070895.1

As with other previously examined genes, all tested *C. butyricum* isolate genes were practically identical (99.9–100% similarity) to genes found in previously isolated strains of *C. butyricum*. Like the tested housekeeping genes, *C. pluributyricum* genes *Cpl-c42-buk* and *Cpl-c67-hydA* also associated most closely with *C. beijerinckii*, however the butyrate kinase *Cpl-c3-buk* did not. Compared to other previously tested genes, the gene was a noticeably lower match to the *C. beijerinckii* genome, with 87% query cover and only 66.8% identity, suggesting that this gene is a distantly related butyrate kinase ortholog.

Instead, the butyrate kinase *Cpl-c3-buk* was found to share its highest similarity (82–83%) to *buk* orthologs from a group of species closely related to *C. drakei*, which only share 91% 16S rRNA sequence similarity and about 72% ANI with *C. pluributyricum.* The relatedness was assessed using whole genome sequence phylogenetic analysis, which included C*. drakei, C. carboxidivorans*, *C.* scatologenes and *C. fermenticellae* and other butyric acid producing species of *Clostridium* (e.g. *C. acetobutylicum, C. tyrobutyricum*) (Fig. 24).

Furthermore, the BLAST search revealed that the three highest-matching species to *Cpl-c3-buk* possess three butyrate kinase orthologs. To further evaluate the relatedness between the *buk* orthologs of *C. pluributyricum* and those found in other species, a phylogenetic comparison of 24 *buk* genes from 11 *Clostridium* species was carried out (Fig. 25). Here, it was found that the *C. pluributyricum* butyrate kinase *Cpl-c3-buk* is closely related to two out of three butyrate kinases in *C. scatologenes*, *C. carboxidivorans* and *C. drakei* and both *buk* genes in *C. fermenticellae*.

Noticeably, the other *buk* genes of *C. scatologenes*, *C. carboxidivorans* and *C. drakei* were recognised to be more closely related to the butyrate kinases *Cpl-c15-buk* from *C. pluributyricum* and *Cbu-c42-buk* from *C. butyricum*.

Clustal Omega Multiple Sequence Alignment (Madeira *et al.*, 2019) was used (see appendix) to identified conserved regions among the 24 orthologues and found that all genes were between 353 and 356 amino acids in length, out

of which 207 residues (58%) shared either weakly similar properties (37/207), strongly similar properties (78/207) or were fully conserved (92/207).

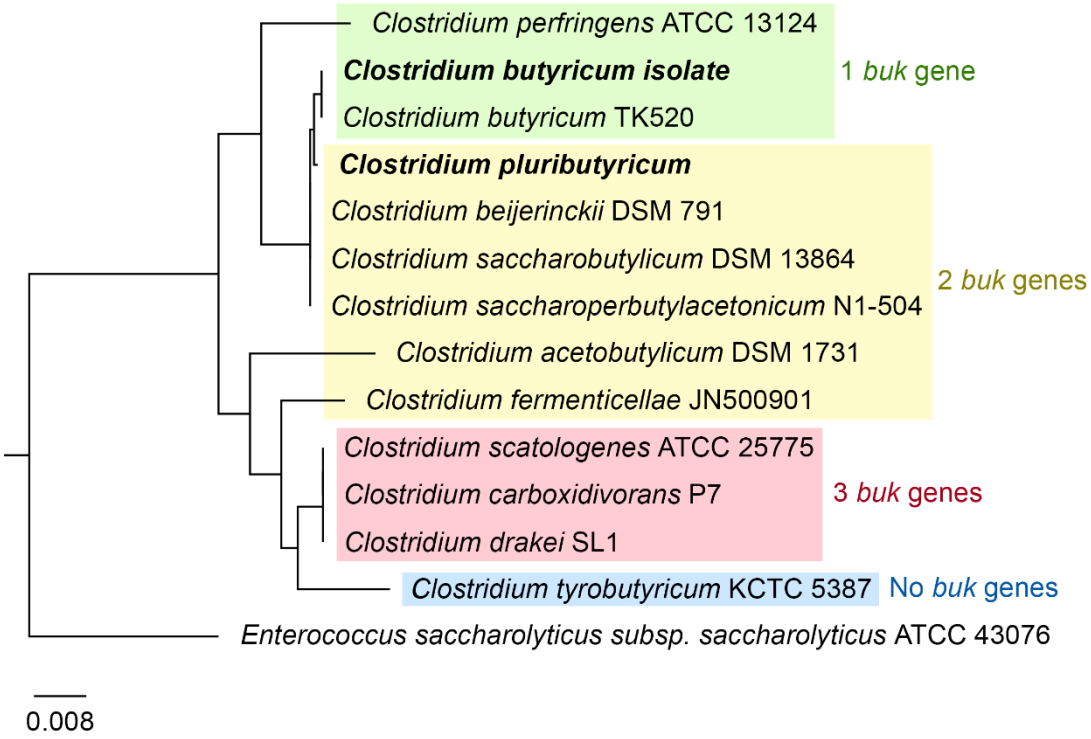


Figure 24. Phylogenetic tree inferred from whole genome sequence data of 13 *Clostridium* species and *Enterococcus saccharolyticus* as the outgroup. Reference genomes were obtained from the NCBI Genome database, the tree was built using the Reference Sequence Alignment Based Phylogeny Builder (REALPHY) version 1.13 (realphy.unibas.ch/realphy/, Bertels *et al.*, 2014). Species are grouped by the number of unique orthologs to *C. butyricum Cbu-c15-buk* and *C. pluributyricum Cpl-c42-buk* butyrate kinase genes within their genomes.

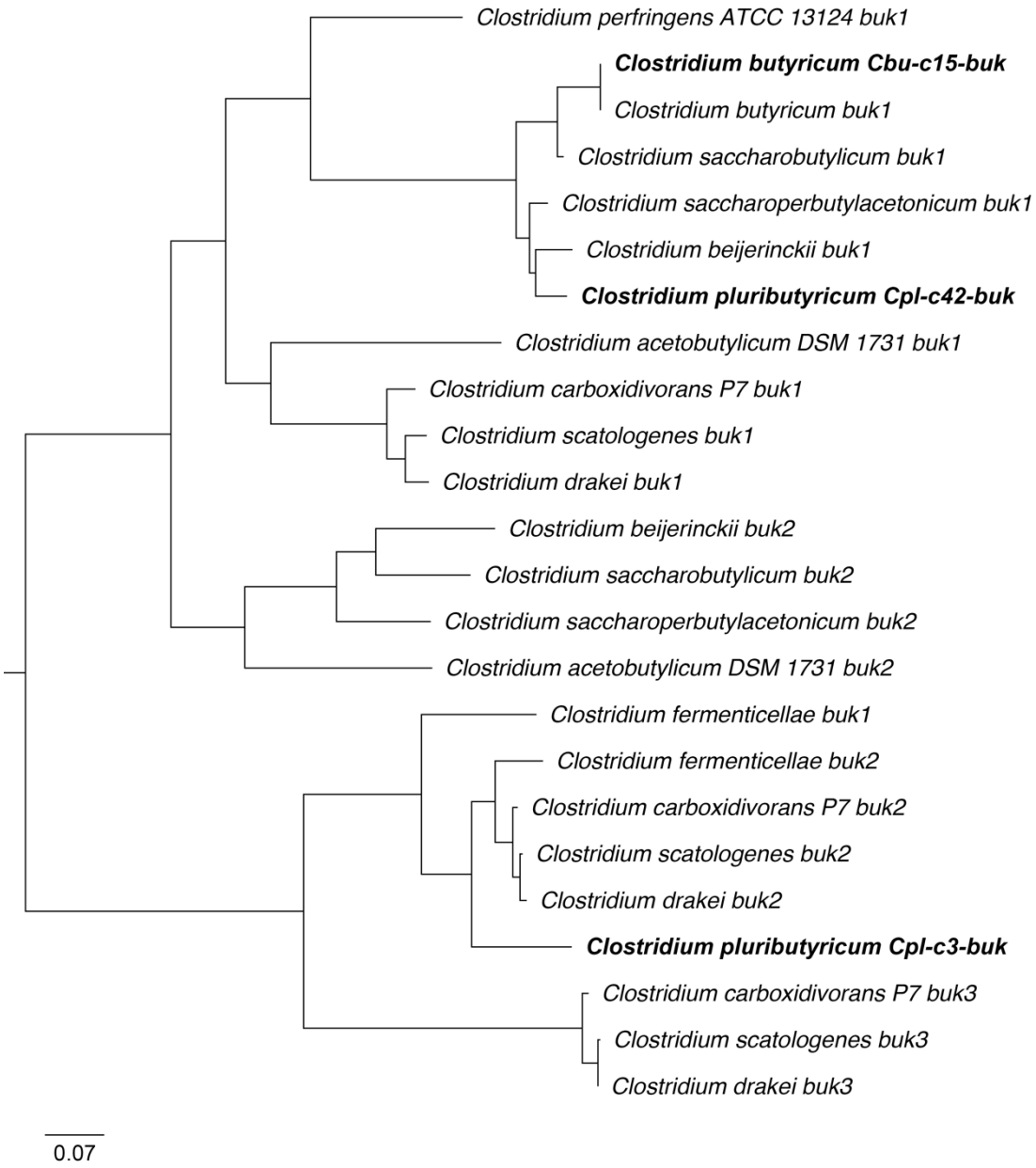


Figure 25. Maximum likelihood tree of 24 butyrate kinase amino acid sequences from 11 *Clostridium* species. The genes were identified as orthologs of the butyrate kinase genes in the *C. butyricum* and *C. pluributyricum* isolates (*Cbu-c15-buk*, *Cpl-c3-buk* and *Cpl-c42-buk*, in bold). The tree was inferred under the GTR+GAMMA model. The branches are scaled in terms of the expected number of substitutions per site. Tree was drawn using the DSMZ single-gene phylogeny server (Meier-Kolthoff *et al.*, 2021).

5.2.1.5 Percentage of conserved proteins

It is recognised that species that belong to the same genus generally share higher similarities in their 16S rRNA genes than in other genes, however Qin *et al.* (2014) argues that the measure of average nucleotide identity (ANI), while suitable to distinguish species, is not a suitable standard of prokaryotic genus delineation. They therefore propose to use both 16S rRNA similarity as well as the percentage of conserved proteins (POCP) as a measure to establish whether two species belong to the same genus or not, with the proposed threshold being 50%. According to their definition, proteins between two species are considered conserved when they have a BLAST match with an E-value of less than 1e-5, a sequence identity of more than 40%, and an alignable region of the query protein sequence of more than 50%.

Whole genome sequences of the species that were included in the phylogenetic analysis (Fig. 24) were converted to their protein-coding sequences using Prokka (Seemann, 2014) and POCP was calculated using the software by Hölzer (2020).

The results show that the *C. butyricum* isolate shares 90% of conserved proteins with *C. butyricum* TK520, which matched the isolate most closely in the BLAST analysis of several housekeeping genes (Fig. 26).

For *C. pluributyricum*, the highest POCP was found with *C. beijerinckii* (66%) followed by *C. saccharoperbutylacetonicum* (62%). This reflected the BLAST comparisons of several housekeeping gene sequences but not the 16S rRNA sequence comparison, as *C. saccharoperbutylacetonicum* was the closest match to *C. pluributyricum*.

Furthermore, the resulting POCP values largely mirror the analysis of the *Clostridium* species' phylogeny by whole genome sequence (Fig. 24), with more closely related species reaching higher values, even though some species appear most distantly related by POCP than by the phylogenetic tree (e.g. *C. butyricum* and *C. pluributyricum*).

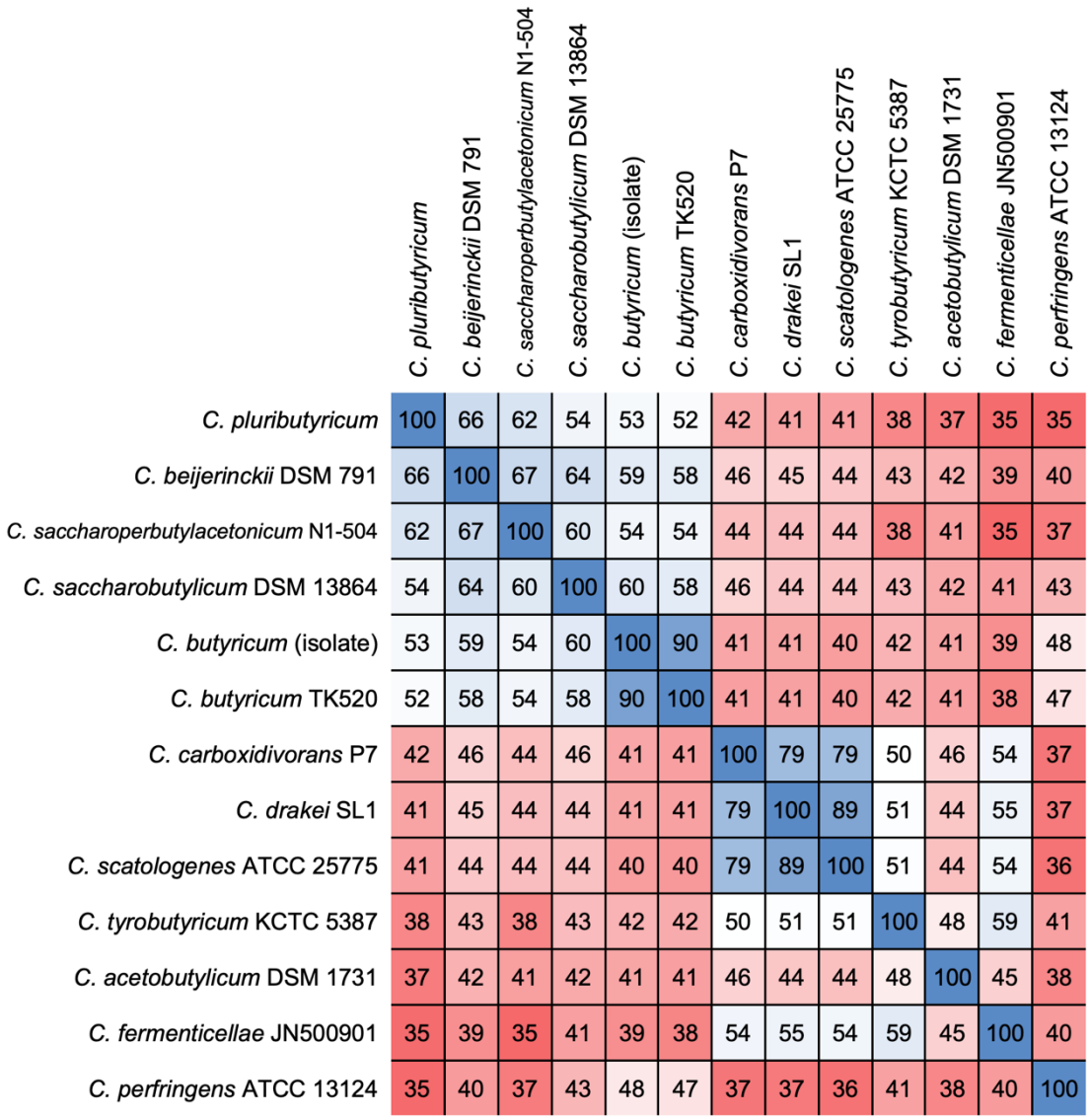


Figure 26. Percentage of conserved proteins (POCP) calculated for 13 *Clostridium* species. Reference genomes for the species were retrieved from NCBI, their protein-coding sequences were identified using Prokka (Seemann, 2014) and POCP was calculated using the software by Hölzer (2020). Rows and columns are sorted by descending POCP with *C. pluributyricum.* Similarities under 50% are shown in red, which according to Qin *et al.* (2014) should be regarded as a threshold for genus delimitation.

5.2.2 Carbon source utilisation

Using the two *Clostridium* isolates to inoculate synthetic complete medium with 12 different carbon sources revealed that there are differences in carbon source utilisation between them. Observing the colour change of the bromocresol purple indicator (Fig. 27) showed that both isolates grew and produced acid in the presence of arabinose, galactose, glucose, sucrose, and trehalose. *C. butyricum* also grew well in the presence of maltose, while medium inoculated with *C. pluributyricum* showed clear signs of growth when lactose, mannitol and sorbitol were present.

An aqueous bromocresol purple solution exhibits changes in its absorption/transmittance spectrum, depending on solution pH. The higher the medium pH, the lower the transmittance of the solution between 500 and 650 nm, with a trough at 590 nm (El-Ashgar *et al.*, 2012). Transmittance of the cultures was measured at 500–650 nm and the observed colour in Fig. 27 closely followed the average transmittance within this spectrum (Fig. 28), with more yellow wells showing higher transmittance. Some cultures that had very low transmittance at 590 nm and were visually indistinguishable from the non-inoculated control in Fig. 27 (i.e. *C. butyricum* in lactose, trehalose and glycerol and *C. pluributyricum* in xylose) showed a more noticeable difference in their transmittance around 550 nm (Fig. 29) when compared to both the non-inoculated control and the control without added carbon sources, suggesting weak acid production.

Figure 27. Growth of *Clostridium* isolates in synthetic complete medium with added carbon sources. Control wells did not contain any added carbon source. The colour change of the bromocresol purple indicator shows change in pH due to acid production (pH > 6.8 = purple; pH < 5.2 = yellow).

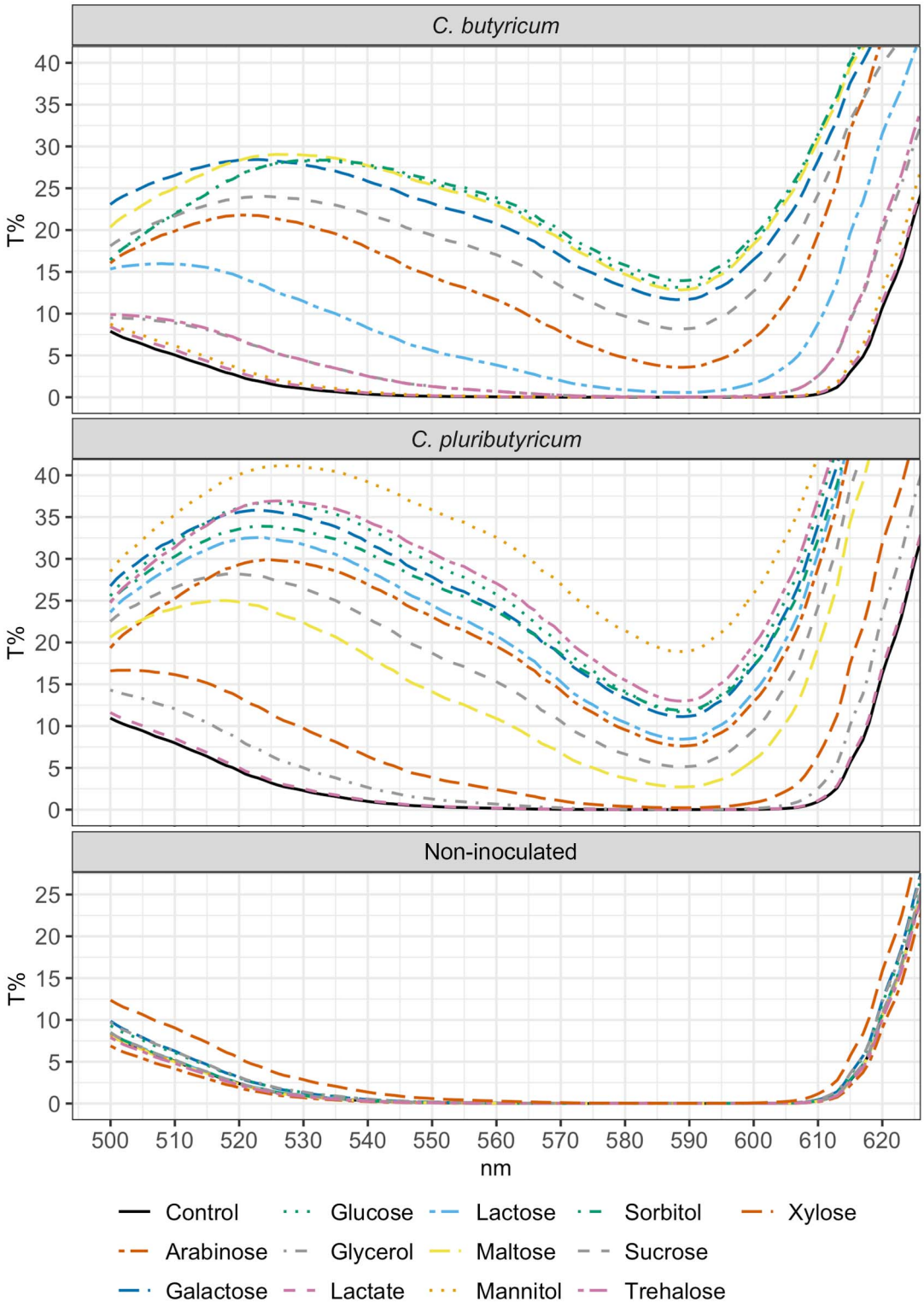


Figure 28. Spectra showing the average (n = 3) transmittance of light (500–630 nm; T%) in cultures of both *Clostridium* isolates in synthetic complete medium with added carbon sources. Control wells did not contain any added carbon source.

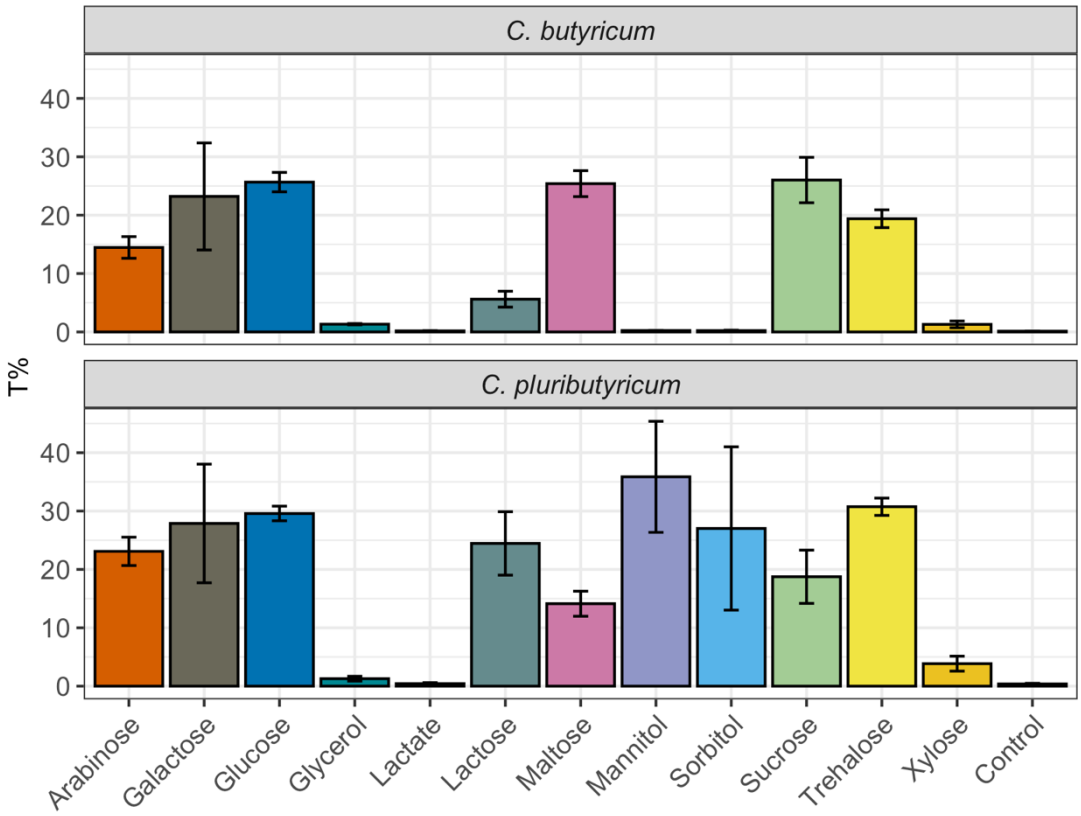


Figure 29. Transmittance (at 550 nm) of *Clostridium* cultures in synthetic complete medium with added carbon sources. Control wells did not contain any added carbon source. Error bars show standard deviation (n = 3).

In addition to the *in vitro* experiments, the gapseq programme was used to infer growth characteristics of both *Clostridium* isolates *in silico*. This was done using the full genome sequences that were generated from both isolates. Gapseq claims to predict and analyse metabolic pathways based on genome sequences and reportedly benefits from a lower error rate compared to other metabolic pathway analysis algorithms (Zimmermann, Kaleta and Waschina, 2021). The programme estimates an arbitrary growth rate (h^{-1}) in user-defined media based on genome data (Fig. 30). Using a range of virtual minimal media with the same added carbon sources as in the *in vitro* experiments revealed a remarkable correlation with the transmittance data in Fig. 29. If the transmittance data in Fig. 29 is taken as a proxy for growth, gapseq seems to correctly predict that, for example, growth of both species is better on glucose than on galactose, which is in turn better than on arabinose. Nevertheless,

gapseq also wrongly predicts that *C. butyricum* can grow as well on mannitol and sorbitol as it does on glucose, which did not reflect the *in vitro* data. The three carbon sources offering the lowest growth rate, as reported by gapseq, are glycerol, lactate, and xylose, which saw little to no observable increase in transmittance compared to the control with both isolates.

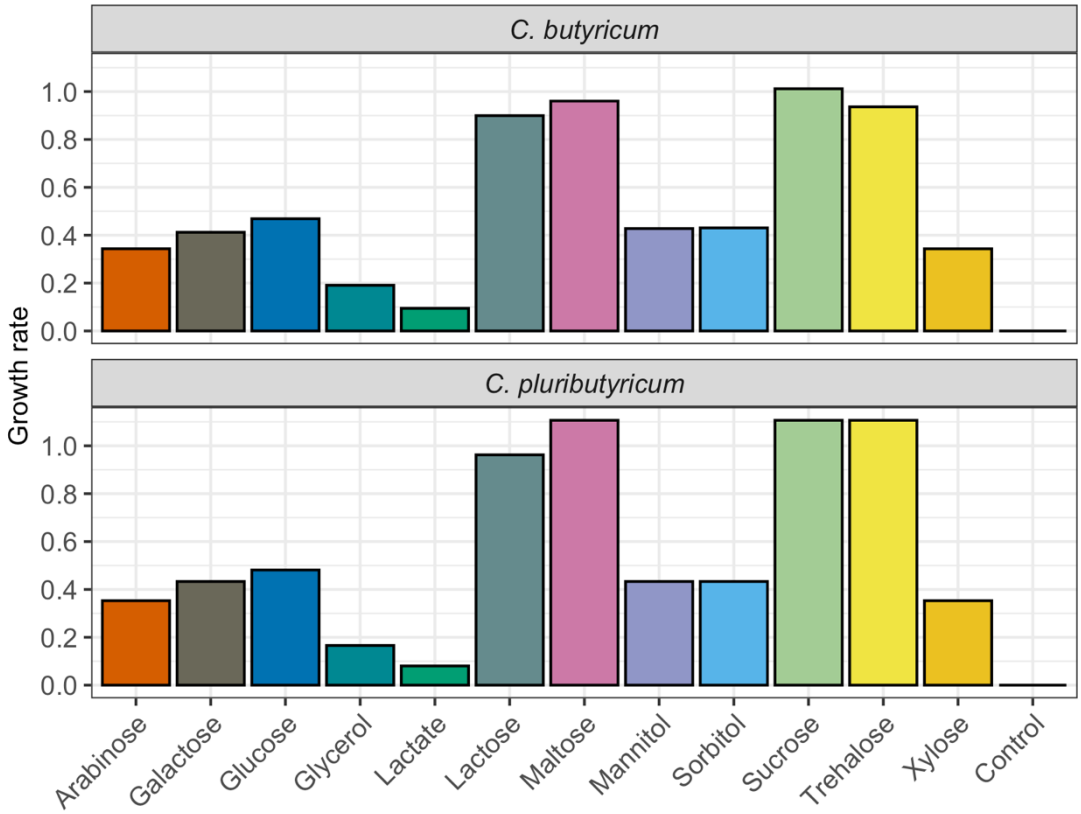


Figure 30. Gapseq prediction of *Clostridium* growth rates (h^{-1}) in 12 different minimal media with added carbon source, based on genome data.

5.2.2.1 PM1 Microplates

PM1 MicroPlates (Biolog, USA) were used to assess growth using a wider range of carbon sources. Here, growth was assessed via a change in the cultures' visible turbidity and their OD_{600}, using a plate reader. Out of 95 tested compounds (Tables 57–58), 17 carbon sources supported growth of both isolates. In addition, *C. butyricum* was also able to grow in the presence of adenosine, D-glucose-1-phosphate, D-glucose-6-phosphate, D-ribose, N-

Acetyl-β-D-Mannosamine as well as α-Methyl-D-Galactoside. Conversely, *C. pluributyricum* was able to grow on D-xylose as well as the 6-carbon sugar alcohols D-mannitol, D-sorbitol and dulcitol (also known as galactitol).

Table 57. Growth (+/–) of *Clostridium* isolates on PM1 MicroPlates
C. bu. = C. butyricum *C. pl. = C. pluributyricum*

Carbon source	*C. bu.*	*C. pl.*	Carbon source	*C. bu.*	*C. pl.*
1,2-Propanediol	–	–	**D-Glucose-1-Phosphate**	+	–
2-Aminoethanol	–	–	**D-Glucose-6-Phosphate**	+	–
2-Deoxy Adenosine	–	–	D-Glucuronic Acid	–	–
Acetic Acid	–	–	D-Malic Acid	–	–
Acetoacetic Acid	–	–	**D-Mannitol**	–	+
Adenosine	+	–	**D-Mannose**	+	+
Adonitol	–	–	**D-Melibiose**	+	+
α-Hydroxybutyric Acid	–	–	D-Psicose	–	–
α-Hydroxyglutaric Acid-ɣ-Lactone	–	–	**D-Ribose**	+	–
α-Keto-Butyric Acid	–	–	D-Saccharic Acid	–	–
α-Keto-Glutaric Acid	–	–	D-Serine	–	–
Bromosuccinic Acid	–	–	**D-Sorbitol**	–	+
Citric Acid	–	–	D-Threonine	–	–
D-Alanine	–	–	**D-Trehalose**	+	+
D-Aspartic Acid	–	–	**D-Xylose**	–	+
D-Cellobiose	+	+	D,L-α-Glycerol-Phosphate	–	–
D-Fructose	+	+	D,L-Malic Acid	–	–
D-Fructose-6-Phosphate	–	–	**Dulcitol**	–	+
D-Galactonic-Acid-ɣ-Lactone	–	–	Formic Acid	–	–
D-Galactose	+	+	Fumaric Acid	–	–
D-Galacturonic Acid	–	–	Glucuronamide	–	–
D-Glucosaminic Acid	–	–	Glycerol	–	–

Table 58. Growth (+/–) of *Clostridium* isolates on PM1 MicroPlates *(continued)*

Carbon source	*C. bu.*	*C. pl.*	Carbon source	*C. bu.*	*C. pl.*
Glycolic Acid	–	–	**Maltotriose**	+	+
Glycyl-L-Aspartic Acid	–	–	Methyl Pyruvate	–	–
Glycyl-L-Glutamic Acid	–	–	Mono Methyl Succinate	–	–
Glycyl-L-Proline	–	–	Mucic Acid	–	–
Glyoxylic Acid	–	–	**N-Acetyl-D-Glucosamine**	+	+
Inosine	+	+	**N-Acetyl-β-D-Mannosamine**	+	–
L-Alanine	–	–	p-Hydroxy Phenyl Acetic Acid	–	–
L-Alanyl-Glycine	–	–	Phenylethylamine	–	–
L-Arabinose	+	+	Proline	–	–
L-Asparagine	–	–	Propionic Acid	–	–
L-Aspartic Acid	–	–	Pyruvic Acid	–	–
L-Fucose	–	–	Succinic Acid	–	–
L-Galactonic Acid-ɣ-Lactone	–	–	**Sucrose**	+	+
L-Glutamic Acid	–	–	Thymidine	–	–
L-Glutamine	–	–	Tricarbalylic Acid	–	–
L-Lactic Acid	–	–	Tween 20	–	–
L-Lyxose	–	–	Tween 40	–	–
L-Malic Acid	–	–	Tween 80	–	–
L-Rhamnose	–	–	Tyramine	–	–
L-Serine	–	–	**Uridine**	+	+
L-Threonine	–	–	**α-D-glucose**	+	+
Lactulose	+	+	**α-D-Lactose**	+	+
m-Hydroxy Phenyl Acetic Acid	–	–	**α-Methyl-D-Galactoside**	+	–
m-Inositol	–	–	**β-Methyl-D-Glucoside**	+	+
m-Tartaric Acid	–	–	Negative control	–	–
Maltose	+	+			

5.2.2.2 Further metabolism predictions using gapseq

The gapseq software was used to predict the fermentation products (in mmol per gram of cell dry weight, mmol/gDW) of both *Clostridiu*m species in optimal medium (Table 59, see optimal medium in appendix). Comparing this with the results obtained from experiments, the model seems to accurately predict that *C. butyricum* produces more acetic acid and less butyric acid compared to *C. pluributyricum*. However, the program underestimates the production of butyric acid by *C. butyricum* and overestimates the production of acetic acid and propionic acid, which was not detected in significant amounts when culture supernatants were analysed. Gapseq predicted the production of formic acid by *C. pluributyricum* but not *C. butyricum,* which is a known producer of this VFA (Hagihara *et al.*, 2018). Production of formic acid was not observed in GC-FID analyses, as the used PERMABOND FFAP column was not suited for its detection.

The production of hydrogen and carbon dioxide was predicted for *C. butyricum* at a ratio of 1.7:1, respectively. In later experiments (Chapter 5.2.4.1), it was discovered that *C. butyricum* produced both gases at a ratio of about 2:1. *C. pluributyricum* also produced both gases, however the model only described the production of hydrogen.

Table 59. Fermentation products in mmol per gram of cell dry weight (mmol/gDW) predicted by gapseq for *C. butyricum* and *C. pluributyricum* under optimal growth conditions.

Metabolite	***C. butyricum***	***C. pluributyricum***
H^+	54.40	92.31
H_2	112.08	6.95
CO_2	66.13	–
Formic acid	–	60.76
Acetic acid	69.0	54.03
Butyric acid	1.21	14.80
Propionic acid	1.62	–
Succinic acid	0.03	–
5-methylthio-D-ribose	0.01	0.01

5.2.3 Acid production at various glucose concentrations

To optimise medium glucose concentration to achieve the highest production of VFAs, the effects of changing glucose concentrations in PYG medium were observed. The concentration of glucose was incrementally increased from 0 g/L to 100 g/L, which had significant effects on the pH of stationary cultures of both isolates (Fig. 31). Between 0 g/L and 2 g/L cultures showed either no growth or very limited growth, suggesting that the isolates were limited by a lack of glucose. At glucose concentrations higher than 10 g/L the pH of stationary *C. butyricum* cultures did not noticeably decrease further than about pH 4.5, suggesting that growth may be limited at this pH.

For *C. pluributyricum*, the pH decreased until the concentration of glucose was about 25 g/L, after which the final culture pH would be around 4.8.

While both isolates were capable of growing in medium containing 10% glucose, growth speed was noticeably decreased at glucose concentrations of 25 g/L and higher.

Increasing the glucose concentration from 5 g/L to 10 g/L, increased the yield of butyric acid in *C. butyricum* cultures by 60% (Fig. 32). However, no significant effect on the VFA yield of *C. pluributyricum*, and the yield of acetic acid in *C. butyricum* was observed.

Further increasing the concentration of glucose to 25 g/L had no effect on *C. butyricum* stationary culture pH but caused another significant decrease of pH in cultures of *C. pluributyricum*.

As slower growth speeds and no significant increase in VFA production had been observed in *C. pluributyricum* cultures when the glucose concentration was initially increased to 10 g/L, the production of VFAs by *C. pluributyricum* at 25 g/L glucose was not tested and all future cultures of the two isolates were grown in PYG medium containing 10 g/L glucose unless otherwise mentioned.

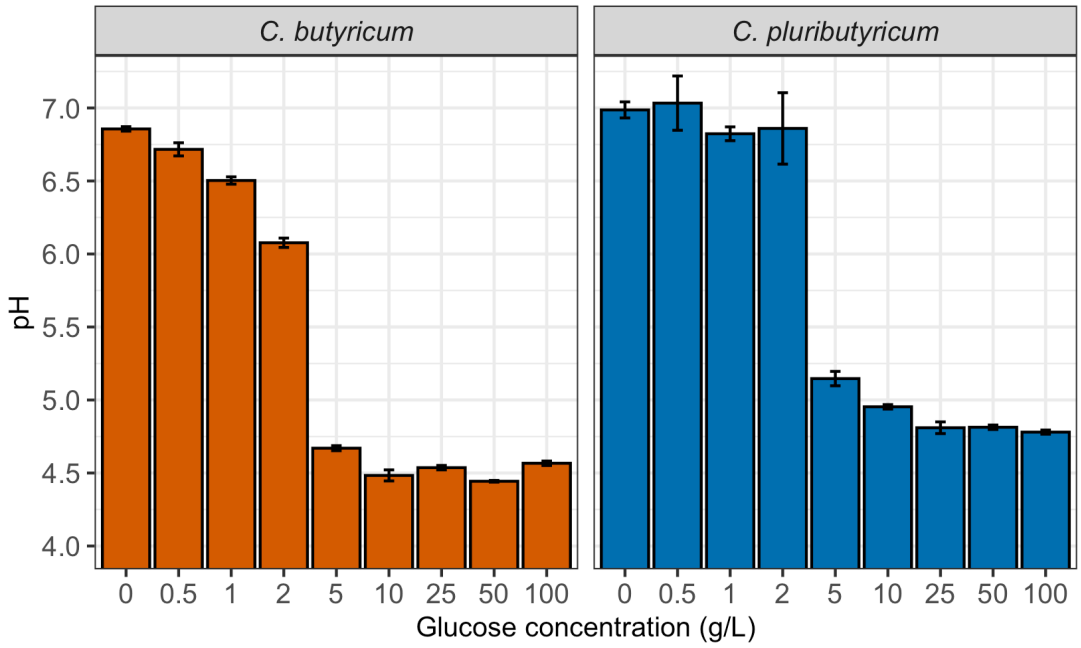


Figure 31. pH of *C. butyricum* and *C. pluributyricum* stationary cultures grown at various starting concentrations of glucose in the medium. Error bars show standard deviation (n = 3: n = 2 for *C. pluributyricum* at 100 g/L glucose, as one replicate was excluded due to lack of growth). Non-inoculated PYG medium adjusted to pH 7 was used as negative control, which maintained neutral pH.

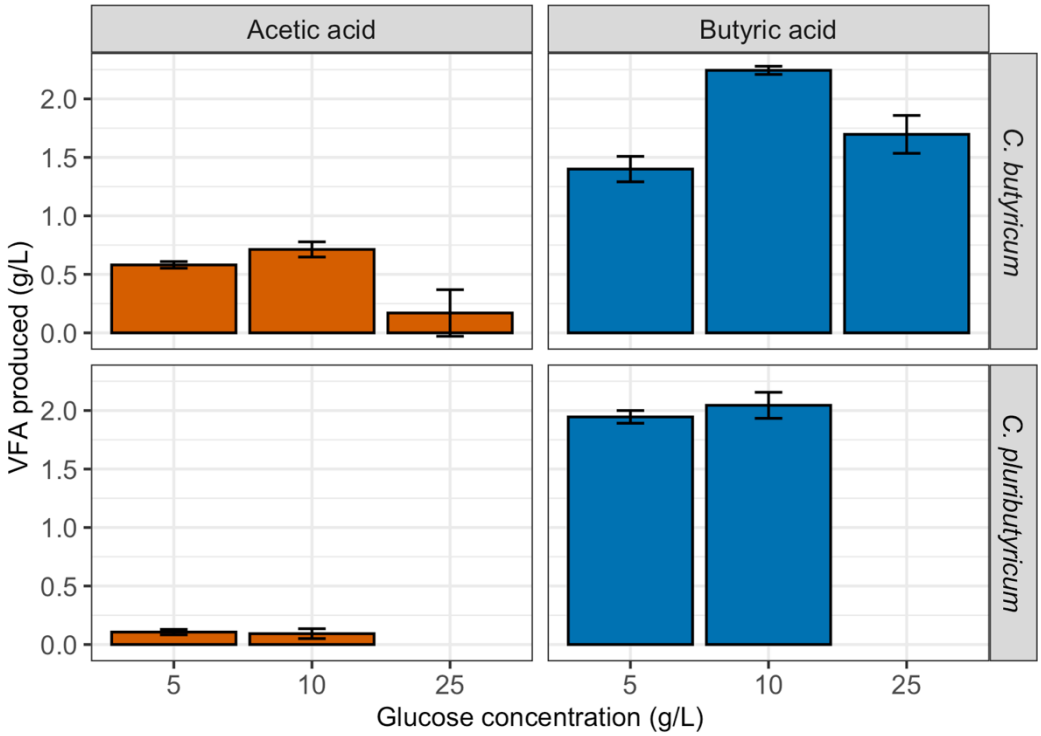


Figure 32. Concentration of acetic and butyric acid in stationary cultures, produced by *C. butyricum* and *C. pluributyricum* isolates grown at various starting concentrations of glucose in the medium. Cultures of *C. pluributyricum* were not cultured in medium containing 25 g/L glucose. Error bars show standard deviation (n = 3)

5.2.4 Gas production of *Clostridium* isolates

Gas production of *Clostridium* isolates was first observed through excessive foaming at the top of their cultures. The volume of gas being produced was then measured by first growing 2 mL cultures in 10 mL syringes. It was seen that cultures of *C. butyricum* produced about 3 mL of gas, while *C. pluributyricum* cultures produced significantly less, about 2.1 mL of gas (Fig. 33). Furthermore, *C. butyricum* cultures started growing exponentially and produced gas after around 6 hours, while the lag-phase of *C. pluributyricum* was around 48 hours before gas production was observed.

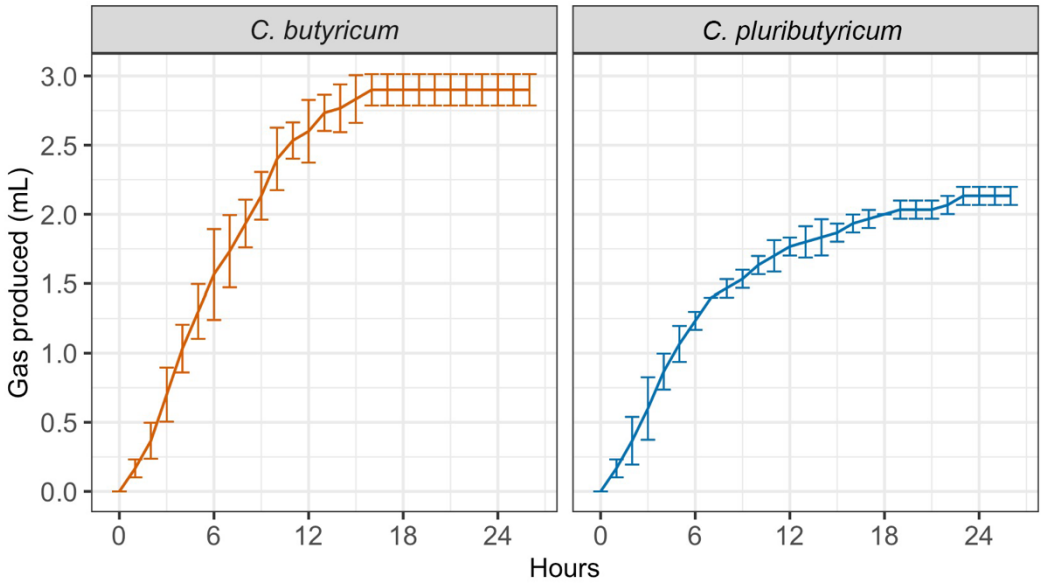


Figure 33. Gas production of the two *Clostridium* isolates in 2 mL syringe cultures (PYG medium with 5 g/L glucose). The *X*-axis describes time after gas production started. Error bars show standard deviation (n = 3).

Later, the cultures were grown at glucose concentrations of 10 g/L and the culture volume was increased to 3 mL to use the full range of the 10 mL syringes. Here, the difference in gas volume produced by *C. butyricum* and *C. pluributyricum* was no longer significant. Considering that culture volume in the syringe reactors had been increased by 50% from 2 mL to 3 mL, a proportionate increase of the amount of produced gas would be expected, even without considering the increase in glucose. Instead, the relative increase was only 10% for *C. butyricum* (albeit affected by large deviations between replicates) and 39% for *C. pluributyricum*.

This observation was affected by large variations in gas volumes in *C. butyricum* cultures, despite syringes being set up in identical replicates of 6. Nevertheless, even at the upper end of the standard deviation error bars, the increase compared to the average gas production in smaller cultures with less glucose would have been 47% at most. The average OD_{600} of *C. pluributyricum* syringe reactors was significantly lower (OD_{600} = 2.00 ± 0.48 SD) than that of *C. butyricum* cultures (OD_{600} = 3.23 ± 0.49 SD; p = 0.001).

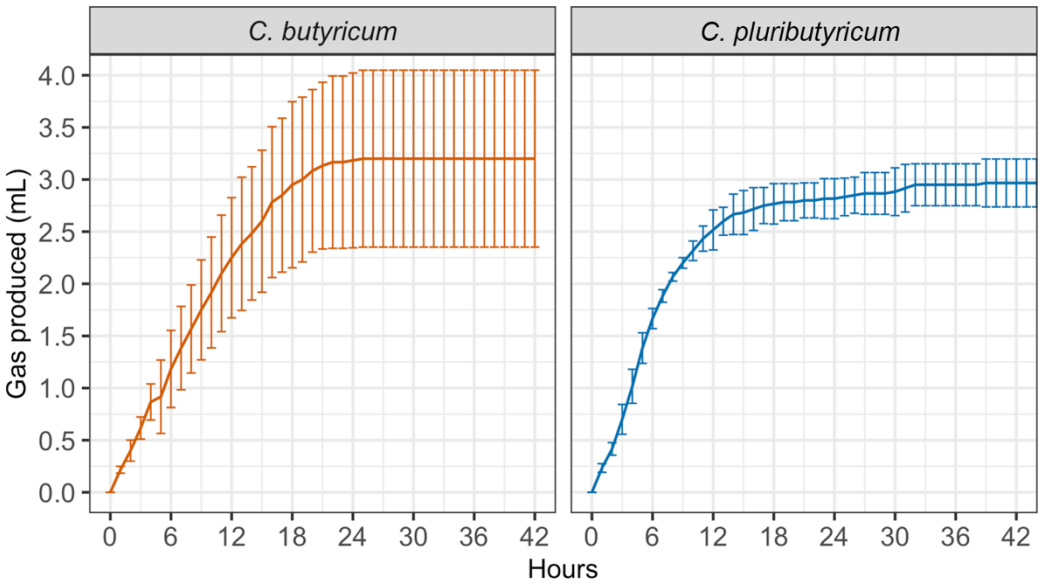


Figure 34. Gas production of the two *Clostridium* isolates in 3 mL syringe cultures (PYG medium with 10 g/L glucose). The *X*-axis describes time after gas production started. Error bars show standard deviation (n = 6).

5.2.4.1 Gas composition

Gas was collected from 500 mL cultures grown in a pure nitrogen atmosphere. The composition of gas samples was measured on a Cambridge Sensotec Rapidox 5100 portable multigas analyser that detects hydrogen, carbon dioxide, methane, and carbon monoxide. According to the manufacturer's handbook, the device uses infrared sensors for the detection of carbon dioxide, methane and carbon monoxide, a thermoconductivity detector to measure hydrogen, as well as an additional tuneable laser diode for methane. Only hydrogen and carbon dioxide were detected in the gas samples, matching reports by Lin *et al.* (2007). The analysis showed that gas produced by both isolates consisted of approximately two-thirds hydrogen and one-third carbon dioxide with no statistically significant difference between the isolates (p = 0.66). On average, gas produced by *C. butyricum* and *C. pluributyricum* contained 68% and 63% hydrogen, respectively.

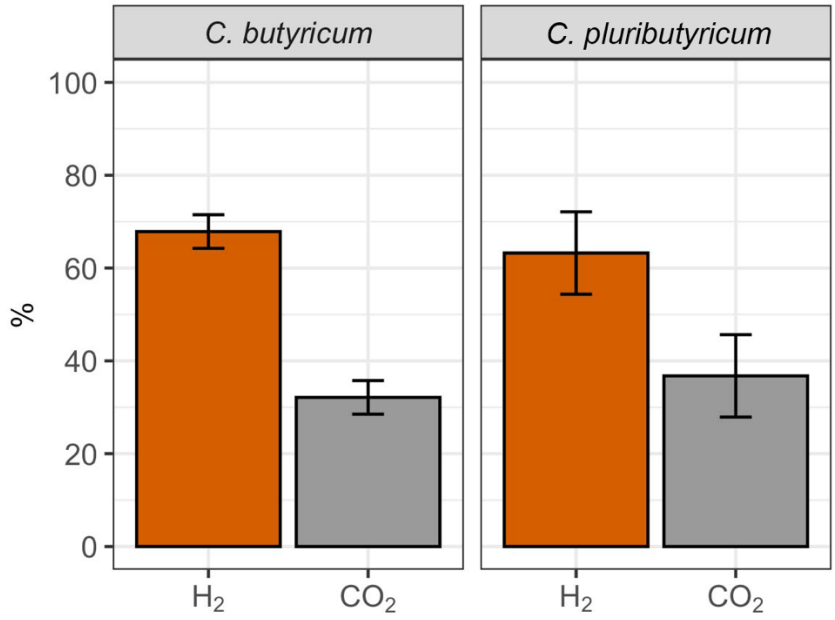


Figure 35. Composition of gas produced by 500 mL cultures of the two *Clostridium* isolates. Error bars show standard deviation (n = 3). So statistically significant difference between the hydrogen concentration in gas produced by the two isolates was observed.

5.3 Discussion

5.3.1 Isolate taxonomy

While the identity of “isolate 1” has undoubtedly been confirmed as a strain of *C. butyricum*, investigating the taxonomy of the second *Clostridium* isolate, which has been named *C. pluributyricum,* brought up many questions about its evolutionary history and its relationship to other species in the genus *Clostridium*.

First, the genome size of *C. pluributyricum* seems to be an outlier in the genus *Clostridium sensu stricto*, as it exceeds the largest reference genome in the genus (*C. saccharoperbutylacetonicum*, 6.67 Mb) by 724 kb and that of its closest relative (by ANI and POCP) *C. beijerinckii* by 1.44 Mb, however without a fully assembled genome sequence it is not known whether this size difference presents as an additional contiguous sequence (e.g. a chromosome or plasmids) or is evenly distributed. It is perhaps unsurprising that its closest relatives within the genus also have relatively large genomes, but it begs the question how the genome of *C. pluributyricum* came to be of such considerably bigger size. Qin *et al.* (2014) argues that genome size may be a stronger indicator of prokaryotic speciation than ANI or POCP values, as differences in genome size suggest that species have faced different environmental pressures and have taken different evolutionary pathways, which would be another argument that this isolate represents a new species. Even though bacteria can acquire DNA through horizontal gene transfer, bacterial genomes tend to decrease rather than increase in size during their evolution and the deletion of genes is generally more common than their acquisition (Mira, Ochman and Moran, 2001).

The loss of genes may also explain why members of the genus possess varying numbers of butyrate kinase gene orthologs. In a study on the distribution of the butyrate kinase pathway genes among butyrate-producing bacteria from the human colon, Louis *et al.* (2004) suggested that clostridial lineages experienced a progressive loss of butyrate kinase genes. If the evolutionary history of the studied *Clostridium* species is represented by their whole genome sequence phylogenetic tree, it suggests that the loss of

butyrate kinase genes may have occurred independently for various species and not within their most recent common ancestor (MRCA). This idea is further supported by the finding that *C. pluributyricum* and its closest relatives (e.g. *C. beijerinckii*) each possess two butyrate kinase orthologs that are however of different evolutionary lineages. Therefore, the MRCA of *C. pluributyricum*, *C. beijerinckii* and *C. drakei* may have possessed at least three butyrate kinase genes, different orthologs of which would have been lost in *C. pluributyricum* and *C. beijerinckii* after the divergence of their MRCA.

Another example of a *Clostridium* species that may have experienced the loss of its butyrate kinase genes is *C. tyrobutyricum.* As this species has previously been recognised for its high production of butyric acid, a BLAST search of its genome against the *buk* genes from *C. pluributyricum* and *C. butyricum* was carried out, however no orthologs were found. This agrees with results from Lee *et al.* (2016), who studied its genome and proteome, and found that it doesn't possess a butyrate kinase gene, which suggests that its butyric acid production is not through the phosphotransbutyrylase-butyrate kinase pathway, but strictly through acetate reassimilation by butyryl CoA-acetate transferase – a gene, which *C. pluributyricum* also possesses.

Furthermore, as the relatedness of several clostridial species was assessed in this analysis, their classification within the same genus was found to be at odds with proposed criteria for genus delimitation by 16S rRNA and POPC. Before Lawson and Rainey (2016) redefined the genus *Clostridium,* which is now defined by 16S rRNA similarity to *C. butyricum*, additional reclassifications had already been proposed to divide the genus further (Qin *et al.*, 2014). If the criteria were applied more strictly, the results presented here confirm that the creation of new genera would be justified for *C. acetobutylicum, C. perfringens* as well as *C. drakei* and its close relatives. Nevertheless, it has been acknowledged that the genus remains in constant flux, and that extensive reclassification, while technically warranted, may cause confusion, especially when applied to species relevant to the medical community (Lawson and Rainey, 2016).

5.3.2 Carbon source utilisation and fermentation products

Despite the close relationship of *C. pluributyricum* to *C. beijerinckii*, the two species differ in their ability to ferment certain carbon sources. Ikeda *et al.* (1988) described the phenotypic characteristics of 21 strains of *C. beijerinckii* to facilitate differentiation with *C. butyricum*. They reported that *C. beijerinckii* reacted weakly to mannitol and sorbitol but readily fermented inositol. In comparison, the spectrophotometric analyses indicated that cultures of *C. pluributyricum* produced most acid when subjected to mannitol and sorbitol, while growth in medium with inositol was not observed. Nevertheless, some characteristics overlapped, for example the inability to use ribose, which together with the ability to use mannitol and sorbitol can be used as a test to distinguish both *C. pluributyricum and C. beijerinckii* from *C. butyricum*. Among the carbon sources that Ikeda *et al.* tested with *C. butyricum*, glycerol was reported to produce a "weak reaction". In the tested *C. butyricum* isolate, growth from glycerol was not conclusively observed, however the spectrophotometric analysis showed a colour change of the indicator, distinct from the negative control and possible abiotic effects. Other similarities between *C. pluributyricum* and *C. beijerinckii* were seen in their ability to ferment xylose (Cynkin and Gibbs, 1958). While some strains of *C. butyricum* have been observed to use xylose as a carbon source (e.g. Heyndrickx *et al.*, 1991; Qiu, Xu and Ren, 2009; Rafineenia, 2013), this was not evident for the isolated strain. Both isolates were unable to utilise fucose, which is the main component of hydrolysed *Fucus vesiculosus* seaweed, followed by glucose, galactose and xylose (Rupérez, Ahrazem and Leal, 2002). Therefore, while these species have been isolated from seaweed digestate, they may not be able to fully contribute to the digestion of all present sugars.

5.3.3 Genome-based metabolic modelling

The gapseq software was used to infer carbon source usage from the isolates' genome sequences, with the aim to further support the validity of spectrophotometric measurements. Despite some false positive results, the overall overlap of the prediction overlapped closely with spectrophotometric results. In their paper upon releasing the software, Zimmermann, Kaleta and Waschina (2021) discussed the software's limitations, comparing it with other metabolic modelling models, namely as ModelSEED and CarveMe. Here, they stated that gapseq outperformed the other models in terms of providing more true positives and less false negatives, while underperforming slightly in regard to true negatives and false positives. From the 26 modelling experiments performed, 6 resulted in false positives, which is above their reported average occurrence of false positive of 11%.

For the prediction of fermentation products, the authors also reported that gapseq outperforms the other two tested models, however they also found that their software tends to underreport the production of lactic acid and formic acid. As production of both acids was not assessed for the isolates, the predicted production of formic acid for *C. pluributyricum*, together with literature reports that its close relative *C. beijerinckii* also produces formic acid (Cho, Shin and Kim, 2012) is therefore a strong indicator to suggest that *C. pluributyricum* does so as well. On the other hand, the production of lactic acid, which a known metabolite for *C. butyricum* but not *C. beijerinckii* (Ikeda *et al.*, 1988), has been wrongly predicted for *C. butyricum* and also likely for *C. pluributyricum*, which unlike *C. beijerinckii* possesses the lactate dehydrogenase gene.

Overall, while predictions have shown to be fallible, the results from gapseq indicate that it can be a powerful tool to predict the metabolic characteristics of novel isolates. For example, the isolation of organisms from digestate may be facilitated if metagenomics data can be used to suggest optimal growth conditions for yet uncultured species.

5.3.4 Acid production

Experiments in which the two isolates have been grown in increasing concentrations of glucose have shown that their lowest achievable culture pH slightly differs.

In cultures that have been acidified by the growth of *C. butyricum*, Chen *et al.* (2005) have measured that final culture pH reached about pH 4.5–5. Furthermore, in non-toxigenic strains of *C. butyricum*, limiting pH values for growth have been described to be acidulant-dependent, with limits of pH 4.1 in medium adjusted with hydrochloric acid, pH 4.7 in presence of lactic acid, pH 4.8 with acetic acid and up to pH 4.7 with citric acid, due to differing buffering capabilities of these acids (Ghoddusi, Sherburn and Aboaba, 2013).

In cultures of *C. beijerinckii*, the closest relative to *C. pluributyricum*, cultures reached pH 4.5 before growth stopped (Giraldeli, Fonseca and Reginatto, 2019). In comparison, cultures of *C. pluributyricum* consistently failed to reach pH values that were equal to those produced by *C. butyricum*, indicating that its pH tolerance may be lower than that of both *C. butyricum* and *C. beijerinckii*.

Interestingly, while both cultures of *C. butyricum* and *C. pluributyricum* experienced a drop in pH when glucose concentrations were increased from 5 g/L to 10 g/L, the concentration of acetic acid and butyric acid only increased in cultures of *C. butyricum*. This increase was skewed towards the production of butyric acid, which may be explained by the lower pH that is reached in these cultures, as *C. butyricum* tends to selectively produce butyric acid over acetic acid at decreasing pH (Heyndrickx *et al.*, 1987).

C. pluributyricum cultures saw a drop in pH without a significant change in measured acetic acid and butyric acid concentrations. This suggests that the drop in pH seen in *C. pluributyricum* cultures may be a result of the accumulation of other acids, such as formic acid or lactic acid. In cultures of *C. acetobutylicum*, the cessation of fermentation reactions in pH-uncontrolled medium has been associated with the accumulation of formic acid (Wang *et al.*, 2011).

When the two *Clostridium* isolates were initially grown at 5 g/L glucose, cultures of *C. pluributyricum* were recognised for their high concentration of butyric acid compared to cultures of *C. butyricum*. In contrast, *C. butyricum* produced more acetic acid than *C. pluributyricum*. As *C. pluributyricum* possesses a butyryl-CoA:acetate CoA-transferase gene, the low concentration of acetic acid may be the result of it being converted to butyric acid. In butyrate kinase-deficient species such as *C. tyrobutyricum* (Lee *et al.*, 2016), this is the main production pathway to produce butyric acid. Furthermore, in cultures of *C. tyrobutyricum*, butyrate is preferentially produced over acetate when the cultures are glucose-limited (Michel-Savin, Marchal and Vandecasteele, 1990). However, due to acid accumulation being the limiting factor in pH-uncontrolled batch cultures, the growth of *C. pluributyricum* in controlled, continuous cultures would reveal more information about its potential as butyric acid-producing organism.

Unlike the large increase in butyric acid production in *C. butyricum* after glucose levels were increased, it is not known why the butyric acid production of *C. pluributyricum* cultures did not significantly increase to surpass concentrations higher than 2 g/L. In *C. beijerinckii*, butyric acid concentrations of more than 3 g/L have been shown to start inhibiting both cell growth and hydrogen production (Kim *et al.*, 2008), however limits of growth in the presence of butyric acid have not been assessed for *C. pluributyricum*.

In cultures of *C. butyricum*, the concentrations of butyric acid are likely not to have reached toxic levels, as far higher concentrations have been reported in batch fermentation cultures (e.g. Zigová *et al.*, 1999; He *et al.*, 2005; Table 34). To completely inhibit *C. butyricum*, Heyndrickx *et al.* (1987) found that the addition of 35 g/L butyric acid to the growth medium were necessary.

Additionally, it is not known whether the two butyrate kinase genes in *C. pluributyricum* play multiple roles in the pathway. In *C. acetobutylicum*, it has been reported that its two butyrate kinase isozymes differ in their specificity for butyric acid (Sullivan, Cates and Bennett, 2010) and increased butyric acid production in single knockout mutants suggests that the less specific isozyme may also be involved in the production of acetic acid (Jang *et al.*, 2014). As

one of the butyrate kinase genes in *C. pluributyricum* is more distantly related to those in *C. acetobutylicum*, these findings are not transferrable and further study on their function is necessary, for example by purifying the produced enzymes and investigating their activity and kinetics, as in Huang *et al.* (2000). Furthermore, the growth of *C. pluributyricum*, which has generally been slower and with a longer lag phase than *C. butyricum* may also be due to a preferential production of butyric acid over acetic acid, which yields less ATP per mol of glucose. In *C. tyrobutyricum*, the production of butyric acid over acetic acid has been associated with slower growth rates (Michel-Savin, Marchal and Vandecasteele, 1990).

5.3.5 Gas production

5.3.5.1 Composition

The measured concentrations of hydrogen and carbon dioxide in the gas produced by *C. butyricum* were within the range of concentrations reported by various studies, which however highlighted that the concentration of hydrogen may be dependent on the studied strain as well as culturing conditions such as medium composition. Measuring the relative amount of H_2 in gas produced by various strains of *C. butyricum*, Heyndrickx *et al.* (1987), Chong *et al.* (2009), Zhao *et al.* (2010) and Beckers *et al.* (2015) measured concentrations of 61%, 60–75%, 58–73% and 59%, respectively. Heyndrickx *et al.* (1987) argues that the stoichiometrically expected composition of gas produced by *C. butyricum* should be around 57% H_2 and 33% CO_2, and that divergence from these values may be explained by the production of formic acid. According to Thauer *et al.* (1972), the production of formate through the pyruvate-formate-lyase reaction, in which pyruvate and coenzyme-A is converted into formate and acetyl-CoA, can act as an additional electron sink for stressed fermentations. Interestingly, the proposed 57/33 split accurately represents the 1.7:1 ratio of H_2 and CO_2 that was suggested by gapseq using *C. butyricum* genome data.

Levin, Pitt and Love (2004) explain that the theoretical maximum amount of hydrogen produced from a mol of glucose is 4 mol H_2 + 2 mol CO_2 when acetic acid is the end-product, i.e:

$$C_6H_{12}O_6 + 2\,H_2O \rightarrow 2\,CH_3COOH + 4\,H_2 + 2\,CO_2$$

Conversely, a mol of glucose can be converted into 2 mol H_2 + 2 mol CO_2 when butyric acid is the end-product, i.e:

$$C_6H_{12}O_6 + 2\,H_2O \rightarrow CH_2CH_2CH_2COOH + 2\,H_2 + 2\,CO_2$$

As a result, species such as *C. butyricum* and *C. pluributyricum*, which produce a mix of acetic acid and butyric acid, tend to produce higher yields of hydrogen than producers of propionate and reduced end-products (e.g. lactic acid, ethanol), such as *C. propionicum*. Therefore, even though *C. beijerinckii* is closely related to *C. pluributyricum* and produces acetic acid and butyric acid, its production of ethanol and butanol involves pathways that reduce NADH + H^+ to NAD^+, which prevents hydrogen from being liberated as gas. As a result, the concentration of hydrogen in gas produced by solventogenic species such as *C. beijerinckii* is expected to be lower than the results obtained from non-solventogenic *C. pluributyricum.* In gas produced by *C. beijerinckii*, Kim *et al.* (2008), Pan *et al.* (2008), Zhao *et al.* (2012) reported hydrogen concentrations of 38%, 50% and 16–48%, respectively.

Therefore, while the differences in H_2 concentrations comparing *C. butyricum* (68%) and *C. pluributyricum* (63%) were not statistically significant, differences may be expected due to the apparent higher production of acetic acid by *C. butyricum*. Still, as *C. butyricum* is a known producer of lactic acid and *C. pluributyricum* possesses the required genes to do so as well, the potential to produce even more hydrogen may be negated.

5.3.5.2 Volume

Regarding gas volumes, the initial observation that *C. butyricum* produced more gas than *C. tyrobutyricum* was first associated with the presence of multiple [FeFe] hydrogenase genes in its genome. In *C. perfringens* hydrogenase knockouts, hydrogen production was disrupted (Kaji *et al.*, 1999), while other studies have achieved increases in gas production when overexpressing [FeFe] hydrogenases in *C. paraputrificum* (Morimoto *et al.*, 2005) and *C. tyrobutyricum* (Jo *et al.*, 2010). Nevertheless, Klein *et al.* (2010) have shown that overexpression of either a *C. butyricum* hydrogenase or a native hydrogenase in *C. acetobutylicum* had no effect on gas volume and composition, however as *C. acetobutylicum* is a solventogenic species, limitations on gas production due to the production of reduced products may apply. Morimoto *et al.* (2005) also observed that hydrogenase overexpression had effects on VFA production and concluded that the over-oxidation of NADH caused a shift from lactic acid production to acetic acid production.

In flask cultures grown at 5 g/L glucose, *C. pluributyricum* primarily produced butyric acid, while *C. butyricum* produced a mix of both butyric and acetic acid. Therefore, another explanation for the increased gas production by *C. butyricum* compared to *C. pluributyricum* may be the increased liberation of hydrogen due to the higher production of acetic acid. The composition of gas produced by glucose-limited cultures was not measured, however due to the shift in *C. butyricum* cultures towards butyric acid production at 10 g/L it would be expected that cultures of *C. butyricum* produced relatively more hydrogen at 5 g/L glucose than at 10 g/L glucose, which might also explain the higher overall gas volume when compared to *C. pluributyricum* at lower glucose concentrations. As the cultures' total amount of acids other than acetic and butyric acid was not measured, the complete stoichiometry of their glucose fermentation is unknown.

Nevertheless, the increased gas production by *C. butyricum* may also be the result of higher culture densities of these cultures, which was only measured by determining OD_{600} for cultures growing in 10 g/L glucose and was significantly higher in cultures of *C. butyricum*. However, as particle size

and shape has an impact on optical density (Meyers, Furtmann and Jose, 2018), the OD_{600} of both species is not directly comparable and other methods such as CFU, cell counts or qPCR may be better suitable to determine culture density between organisms of different size.

5.3.6 Conclusion

C. butyricum and *C. pluributyricum* are both members of the genus *Clostridium sensu stricto*, which is characterised by fermenting sugars into products such as acetic acid, butyric acid and hydrogen.

While the newly isolated strain of *C. butyricum* is very similar to other strains of this species, the newly isolated species *C. pluributyricum* shows characteristics associated with its closest relative *C. beijerinckii*. This includes similarities such as sugar alcohol fermentation (i.e. mannitol, sorbitol), the inability to metabolise ribose and having two butyrate kinase genes as well as the genes to produce butyric acid from acetic acid and acetoacetate.

At the same time, *C. pluributyricum* also shares characteristics with *C. butyricum*, e.g. genes to produce lactic acid, high concentrations of hydrogen in produced gas and a lack of genes to produce solvents such as butanol. Furthermore, *C. pluributyricum* lacks the restriction modification systems found in *C. beijerinckii* (Dong *et al.*, 2010), which facilitates its genetic manipulation.

Other characteristics are unique to *C. pluributyricum*, like having a genome that is larger than any other species in the genus and containing the butyrate kinase *Cpl-c3-buk*, which stands out for a lack of close orthologs in immediate relatives.

Overall, the characteristics of these two isolates would make them suitable contributors to the acetogenesis stage in the anaerobic baffled reactor they were isolated from. Nevertheless, more can still be learned about the novel species *C. pluributyricum*, for example by growing it in pH-controlled fermenters, measuring its fermentation products in more detail and exploring its limits and potentials.

6 Genetic engineering of *Clostridium* isolates

6.1 Introduction

While the genus *Clostridium* comprises several species of industrial relevance, the tools to engineer members of this genus still lags behind well-studied model organisms (Joseph, Kim and Sandoval, 2018). Even though platform organisms such as *E. coli* are highly editable, the introduction of heterologous pathways is often problematic due to the burden of overexpression and metabolic drainage for product synthesis. As a result, these constraints drive the use of non-conventional microbial platforms with native and more robust biosynthesis pathways (Czajka *et al.*, 2017), which however requires the development of new synthetic biology tools. Yet, some of the advantages of using non-model organisms over a well-defined chassis such as *E. coli* have been described as the naturally present desirable traits as production platforms, increased flexibility to use different carbon sources, higher tolerances to feedstock substrates, end-products and bioprocess conditions such as pH, temperature and oxidative stress (Fatma, Schultz and Zhao, 2020). With regards to the utilisation of *Clostridium* as a non-model platform, Charubin *et al.* (2018) have pointed out that the genus is one of the broadest and most flexible groups of organisms in regards to substrate utilisation, which poses a unique advantage over non-model systems, as substrate costs can account for more than two-thirds of operational production costs for the production of chemicals from biomass.

Within the last 15 years, several advances have facilitated the engineering of *Clostridium* species. Most notable tool developments include the ClosTron gene knock-out system (Heap *et al.*, 2007), the pMTL80000 *E. coli* – *Clostridium* shuttle plasmids employed in this project (Heap *et al.*, 2009), a knock-out and knock-in system using allele-coupled exchange (Minton *et al.*, 2016) as well as the extension of CRISPR-Cas9 technology to the genus *Clostridium* by applying it in *C. acetobutylicum* (Bruder *et al.*, 2016).

The genetic toolkit (e.g. promoter libraries) for *Clostridium* has been described as lacking and many of the studies on toolkit development have

focussed on a few species such as *C. acetobutylicum*. While some studies shown that certain promoters are transferrable to related species, the full extent of this compatibility is not well studied. For example, Yang *et al.* (2017) developed a synthetic promoter library for *C. acetobutylicum* which also worked in *C. ljungdahlii*, however with much reduced strength.

Mordaka and Heap (2018) attempted to generate a promoter library by randomly mutating a known promoter. Here, they found that the newly designed promoters worked well in *E. coli* but not in *C. acetobutylicum* before reducing the random mutation rates, demonstrating an unexpected "stringency" of promoter sequences in *Clostridium*. In their study, they also compared different types of oxygen-independent reporters such as CreiLOV, PhiLOV, SNAP-tags and beta-glucuronidase (GusA), the latter of which performed best and was chosen to be used in this project as well.

6.1.1 Aims and objectives

As the number of promoters characterised for *C. butyricum* is lacking and none have ever been tested for the newly isolated *C. pluributyricum*, the first aim was to find a working promoter for the two species by constructing plasmids to express the GusA reporter under the endogenous *C. butyricum* / *C. pluributyricum* counterparts of the *C. acetobutylicum* thiolase promoter as well as the *C. sporogenes* ferredoxin promoter.

As previously shown (Fig. 18), the butyrate kinase and hydrogenase genes are part of the butanoate pathway. In *Clostridium* and other butyric acid-producing bacteria, butyrate kinase, which belongs to the family of transferases, metabolises the last step in the pathway, in which a phosphate group is transferred from butyryl phosphate to ADP, resulting in butyrate and ATP (Miller, 1978).

Within the butanoate pathway of *Clostridium*, [FeFe] hydrogenase is responsible for the production of hydrogen. Here, ferredoxin is reduced by ferredoxin-NAD^+ reductase, which transfers electrons from NADH to ferredoxin. The reduced ferredoxin then delivers the electron to the hydrogenase, resulting in the formation of H_2. While carrying out similar functions to [NiFe]

hydrogenases found in aerobic species such as *E. coli* and *Helicobacter pylori* (Vignais, Billoud and Meyer, 2001), [FeFe] hydrogenases in *Clostridium* species are more oxygen sensitive and cannot be reactivated after oxygen exposure, further highlighting the molecular mechanisms behind the strict anaerobic nature of Clostridia.

Due to the important role of these two enzymes in the production of butyric acid and hydrogen, each of the hydrogenase and butyrate kinase genes from *C. butyricum* and *C. pluributyricum* were overexpressed in engineered strains of both isolates and overexpression was confirmed using qPCR. Furthermore, to demonstrate the ability to overexpress these genes on a protein level, His-tagged constructs were created, and expression was assessed using Western blot.

Finally, the effects of *buk* and *hydA* overexpression in mutant strains was assessed by measuring the amount and composition of produced gas, as well as VFA concentrations in culture supernatant. Here, the aim was to test the hypothesis that the differences in *C. butyricum* and *C. pluributyricum* wild type cultures may have been due to the different number of *buk*/*hydA* gene orthologs within the genomes of these two species, the hypothesis being that an increased number of *buk* genes promotes the formation of butyric acid, while additional *hydA* genes increase hydrogen production.

6.2 Results

6.2.1 Finding native promoters for use in *C. butyricum* and *C. pluributyricum*

The sequences of both constitutive promoters included in the pMTL80000 *E. coli* – *Clostridium* shuttle plasmid kit (*Cac thl* and *Csp fdx*) were used in a BLAST analysis to find the sequence of the *C. acetobutylicum* thiolase gene (GenBank CP030018.1:263285-264463) and the *C. sporogenes* ferredoxin gene (GenBank CP011663.1:86476-86646). BLAST was then used to find the orthologs of both genes in *C. butyricum* and *C. pluributyricum.* Promoters in the upstream regions of the ferredoxin and thiolase genes in *C. butyricum* and *C. pluributyricum* were identified using BPROM (Tables 60 and 61).

Table 60. Thiolase genes from three *Clostridium* species and the promoters detected by Softberry BPROM in the non-coding region up to the next upstream gene.

Species	***C. acetobutylicum***	***C. butyricum***	***C. pluributyricum***
Gene length	392 aa	394 aa	392 aa
Nucleotide homology to *Cac thl*	100%	77%	75%
Length of upstream region	326 bp	236 bp	741 bp
Promoters in upstream region detected by BPROM	**Promoter 1:** Position: 225 -10 box (210) GGGTATAAT -35 box (190) TTGATA	**Promoter 1:** Position: 236 -10 box (192): AATTATAAT -35 box: TTAACG	**Promoter 1:** Position 518 -10 box (503) TATTAAAT -35 box (485) TTTCAA **Promoter 2:** Position 155 -10 box (140) TGTTATAAT -35 box (116) TTAAAG

Table 61. Ferredoxin genes from three *Clostridium* species and the promoters detected by Softberry BPROM in the non-coding region up to the next upstream gene.

Species	*C. sporogenes*	*C. butyricum*	*C. pluributyricum*
Gene length	56 aa	57 aa	57 aa
Nucleotide homology to *Csp fdx*	100%	85%	83%
Length of upstream region	81 bp	121 bp	163 bp
Promoters in upstream region detected by BPROM	**Promoter 1:** Position: 61 -10 box (42) TGCTATAAA -35 box (28) TGGTAA	**Promoter 1:** Position: 61 -10 box (40): ATATAAAAT -35 box (21): TTAAAT	**Promoter 1:** Position 128 -10 box (114) TGTTAAAAT -35 box (94) TTTAAA

6.2.2 GUS reporter system in *Clostridium*

The β-glucuronidase gene (*gusA*) from *E. coli* NEB 5-alpha was cloned into the pMTL80000 *E. coli–Clostridium* shuttle plasmids to evaluate expression under the endogenous promoters that were identified in *C. butyricum* (*Cbu thl* and *Cbu fdx*) and *C. pluributyricum* (*Cpl thl* and *Cpl fdx*) as well as under the ferredoxin promoter from *C. sporogenes*, which was included in the commercially available plasmid kit. Attempts to construct a plasmid containing the GUS reporter under the thiolase promoter from *C. acetobutylicum* were not successful as plasmid constructs in *E. coli* continued to be affected by recombination events.

In samples of bacteria that expressed the recombinant GUS reporter, addition of PNPG to pre-treated cells caused a colour change from clear to yellow within a few minutes. Fig. 36 shows this colour change as sample absorbance (ABS_{405}) relative to the culture's initial optical density (OD_{600}). This colour change is mediated through the enzymatic cleavage of PNPG (4-nitrophenyl β-D-glucuronide) by the expressed β-glucuronidase (GUS) enzymes, leaving behind 4-nitrophenyl that can be detected spectrophotometrically at around 405 nm.

It was found that the thiolase promoter from *C. butyricum* (*Cbu thl*) was the strongest in both *Clostridium* isolates, followed by the promoters *Cpl fdx*

and *Cbu thl*. The ferredoxin promoters from *C. pluributyricum* and *C. sporogenes* caused no expression in either species.

The plasmid containing the *Cbu fdx* promoter in front of its multiple cloning site was labelled pMTL8315B and was consequently used for all experiments where the aim was to overexpress recombinant proteins in *Clostridium* isolates. Initially, the reason for using the promoter with intermediate strength was to be able to later selectively increase or decrease the expression of recombinant proteins and observe possible effects, however the fine-tuning of expression was not investigated due to time constraints.

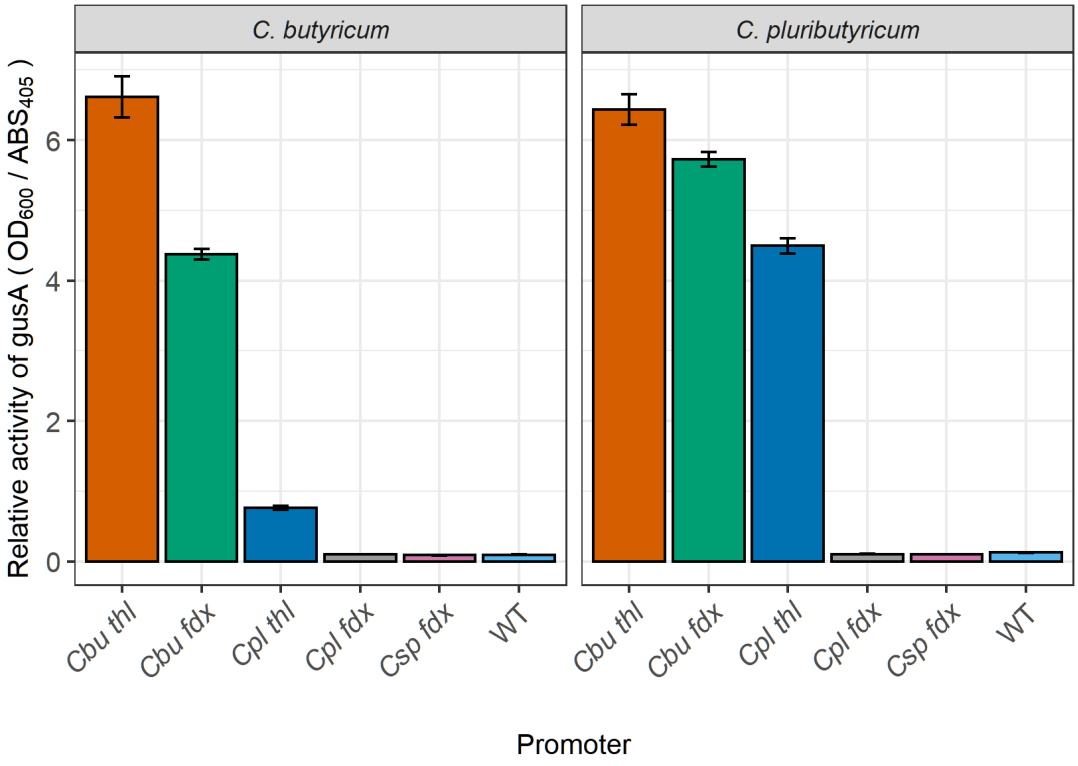


Figure 36. Relative activity of GUS reporter system when expressed in *Clostridium* mutants under five different promoters. Expression of the reporter gene *gusA* from *E. coli* was attempted under thiolase promoters from *C. butyricum* and *C. pluributyricum* as well as ferredoxin promoters from *C. butyricum, C. pluributyricum* and *C. sporogenes*. Error bars show standard deviation (n = 3)

Furthermore, a plasmid was constructed to express the GUS reporter under the *Pseudomonas putida* XylS promoter (Zwick, Lale and Valla, 2013), which is inducible using benzoic acid derivatives. This was tested in both species, using m-toluoic acid as the inducer at concentrations of 1 µM, 10 µM, 100 µM, 1 mM, however no expression was observed (results not shown).

6.2.3 RT-qPCR to confirm expression of plasmid-based butyrate kinase and hydrogenase genes in *Clostridium* mutants

To assess the expression of inserts in both *Clostridium* mutants, RT-qPCR was used to amplify the insert transcripts from cDNA libraries. The expression fold change compared to the empty plasmid control was calculated (Fig. 37) to confirm that the transformation of the *Clostridium* isolates with pMTL8315B plasmids induced the transcription of insert DNA to mRNA.

In *C. butyricum* mutants, it was seen that every mutant apart from that containing the pMTL8315B Cbu-c15-buk plasmid showed a considerably higher level of insert expression compared to the empty plasmid control. In *C. pluributyricum* mutants, every mutant exhibited a noticeably increased expression fold change compared to the control, with the lowest being 23.8 times higher (*Cpl-c67-hydA*, logFC = 1.37). Expression fold changes were generally highest (logFC > 4) in mutants expressing inserts that were not endogenous (e.g. ***Cpl****-c67-hydA* in *C. butyricum*).

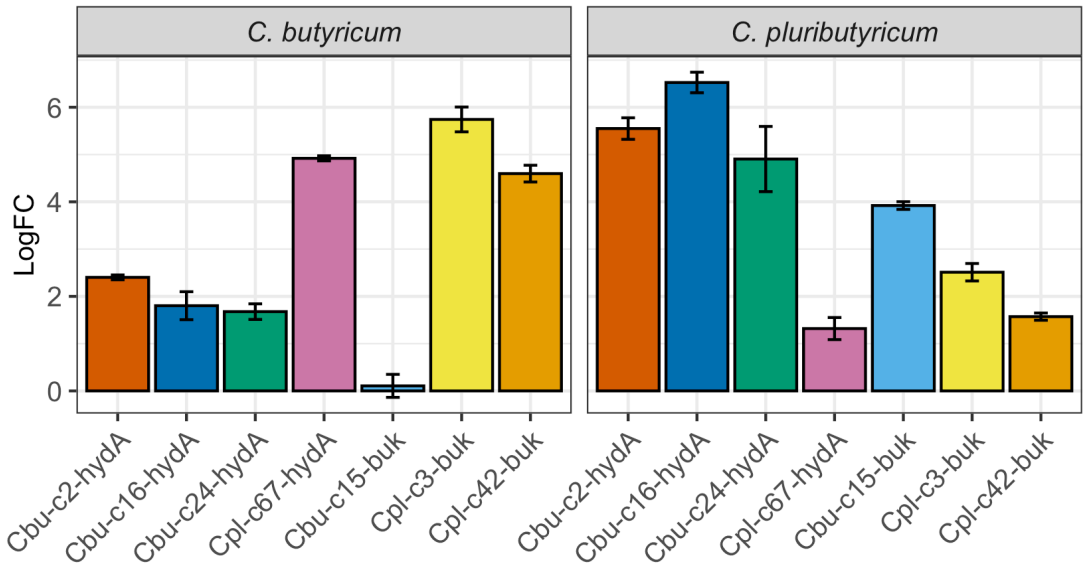


Figure 37. Expression fold change of insert genes in *Clostridium* mutants. cDNA libraries from *C. butyricum* and *C. pluributyricum* mutants were used in RT-qPCR reactions to amplify transcripts of genes on pMTL8315B plasmids. Cq values were used to determine the relative expression fold change (logFC) of genes transformed into *C. butyricum* and *C. pluributyricum* on the pMTL8315B plasmid. Expression is compared to the empty plasmid control and has been normalised against housekeeping genes *rrs* and *adk*. Error bars show standard deviation (n = 3)

The Cq values (Fig. 38) that were obtained when trying to amplify non-endogenous genes from the empty plasmid control's cDNA library were often very high (> 35), indicating that the target was not present. The high fold change values in non-endogenous genes can therefore be interpreted to be infinite, confirming that the gene is only being expressed in the mutant due to its presence on the plasmid. Inversely, when mutants (e.g. *C. butyricum* Cbu-c15-buk, *C. pluributyricum* Cpl-c67-hydA) showed a relatively low increase in expression fold change compared to the empty plasmid control it was likely due to already low Cq values when amplifying from control cDNA, indicating a high number of transcripts for the gene of interest even before attempting to induce overexpression.

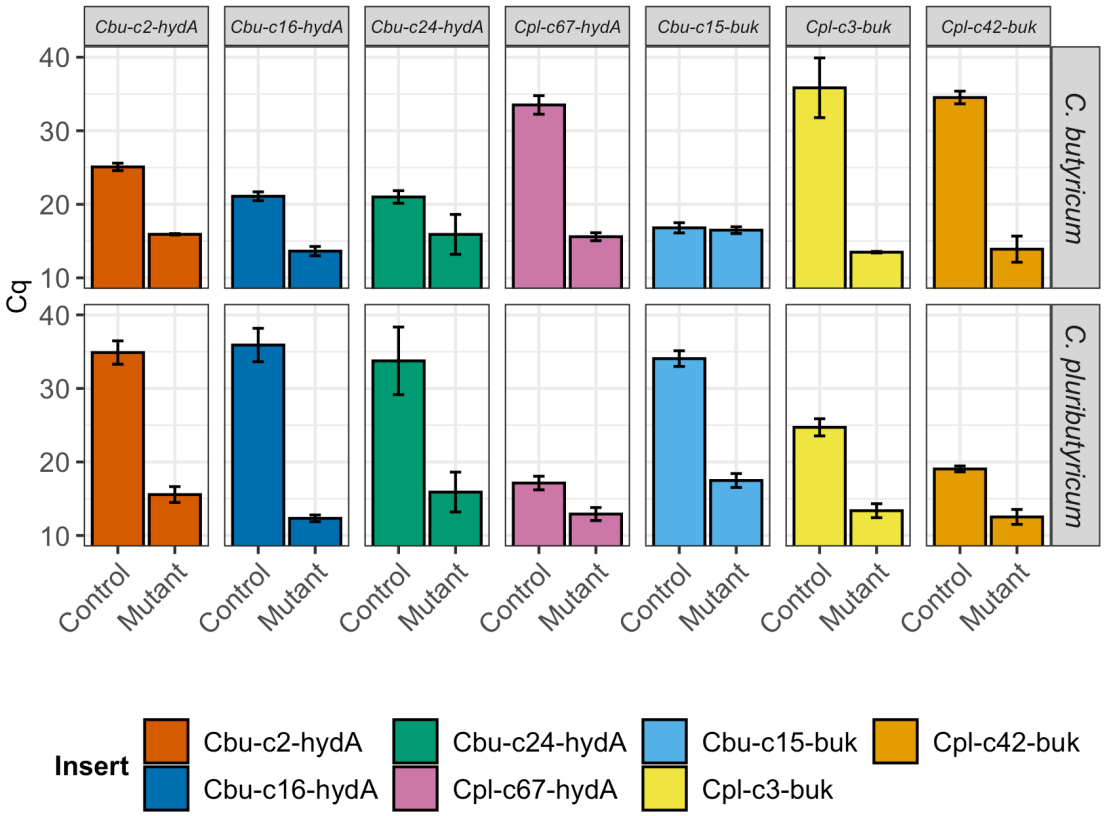


Figure 38. Average RT-qPCR Cq values in reactions to amplify insert genes. cDNA libraries from *C. butyricum* and *C. pluributyricum* mutants were used in RT-qPCR reactions to amplify transcripts of genes on pMTL8315B plasmids. Low Cq values mean high number of transcripts in sample. “Control” denotes empty plasmid control. Error bars show standard deviation (n = 3).

6.2.4 His-tagged genes and Western blot

As RT-qPCR only confirms the transcription from recombinant DNA to mRNA, the aim was to also show evidence for the production of recombinant buk/hydA protein as well. For this, the sequences of the inserts were modified to include a 6xHis-tag at the C-terminus. Proteins were extracted from *Clostridium* mutants and Western blot was used to detect the presence of His-tagged inserts in the samples. Western blots on solubilised cell lysates from *C. butyricum* (Fig. 39) and *C. pluributyricum* (Fig. 40) showed a fair amount of non-specific binding of anti-6-His antibody to native proteins, as indicated by the wild-type control. This was especially evident around 40 kDa, which overlapped with three of the genes that were to be overexpressed. However, these unspecific bands were generally fainter and all mutants showed highly visible (i.e. stained with anti-6-His antibody) bands at the expected molecular weight. Band patterns looked near identical in mutants of both species, indicating that the transformation with the pMTL8315B plasmids leads to the expression of recombinant butyrate kinases and hydrogenases in both *C. butyricum* and *C. pluributyricum*.

As polyhistidine tags can alter protein function and stability (Booth *et al.*, 2018), these mutants only served to prove that these genes can be successfully expressed. The production of recombinant protein in mutants expressing non-His-tagged versions, which were used in further experiments to measure gas and VFA production, is therefore strongly indicated but only confirmed on a mRNA transcript level, using RT-qPCR.

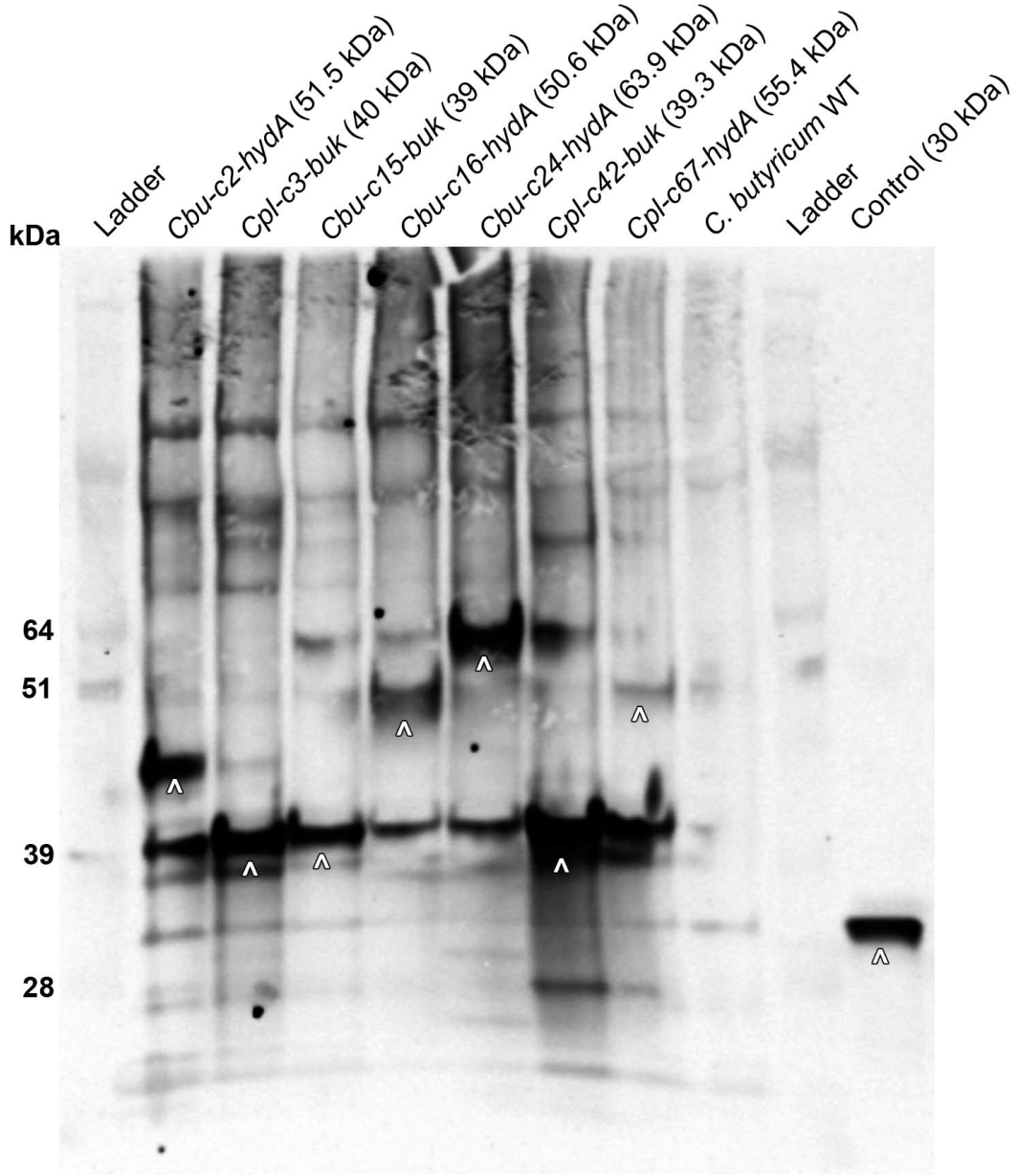


Figure 39. Western blot of solubilised protein extract from *C. butyricum* mutants. Western blotting was used to detect His-tagged proteins in cell lysates of *C. butyricum* mutants transformed with pMTL8315B plasmids containing *hydA* and *buk* genes with 6xHis modifications. Cell lysate pellets were solubilised using 8M urea prior to running them on an SDS-PAGE gel. White arrows show protein bands matching the expected molecular weight for the respective gene inserts. A purified his-tagged protein of known size was used as a control.

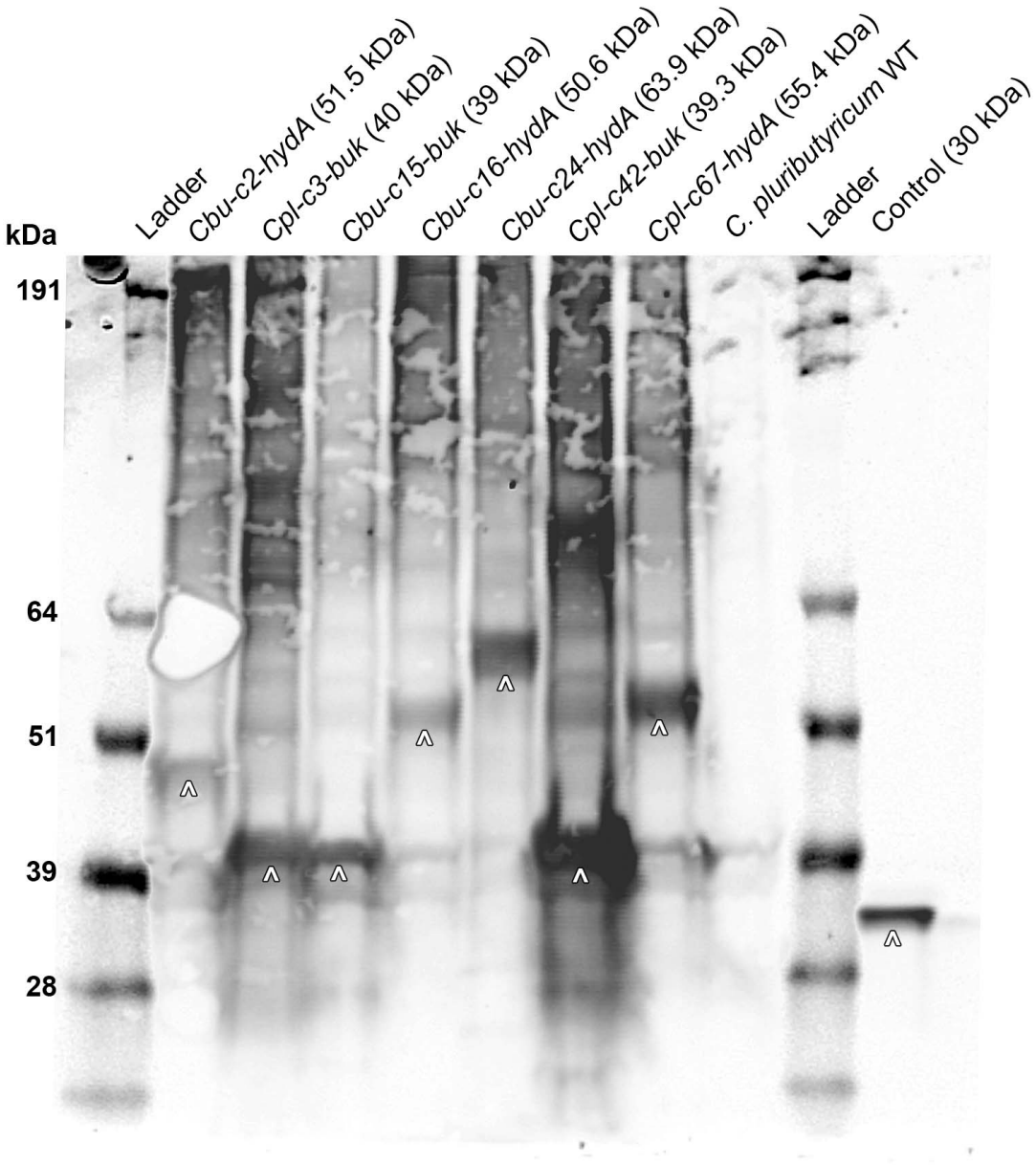


Figure 40. Western blot of solubilised protein extract from *C. pluributyricum* mutants. Western blotting was used to detect His-tagged proteins in cell lysates of *C. pluributyricum* mutants transformed with pMTL8315B plasmids containing *hydA* and *buk* genes with 6xHis modifications. Cell lysate pellets were solubilised using 8M urea prior to running them on an SDS-PAGE gel. White arrows show protein bands matching the expected molecular weight for the respective gene inserts. A purified his-tagged protein of known size was used as a control.

6.2.5 Gas production of *Clostridium* mutants

After several pMTL8315B plasmids containing various hydrogenase genes were constructed, the main aim was to test the effect of hydrogenase overexpression on gas production. Additionally, gas production of mutants harbouring plasmids for the overexpression of either of the three identified butyrate kinase genes was also assessed. The primary hypothesis was that overexpression of hydrogenases might lead to an increase in gas production, especially in *C. pluributyricum*, which naturally only expresses one hydrogenase gene while *C. butyricum* can express three different hydrogenases.

6.2.5.1 *C. butyricum* mutants

Assessing the gas production of *C. butyricum* mutants in 3 mL syringe cultures (Fig. 41) showed that none of the mutants produced a significantly higher amount of gas when compared to the empty plasmid control. Conversely, the mutant expressing the hydrogenase *Cbu-c16-hydA* produced a significantly lower volume of gas and grew to a very significantly lower OD_{600} when compared to the empty plasmid control ($p_{(Gas)}$ = 0.02, Fig. 42 A; $p_{(OD)}$ = 0.0002, Fig. 42 B). Furthermore, cultures of mutants expressing the hydrogenases *Cbu-c24-hydA* and *Cpl-c67-hydA* also showed significantly lower OD_{600} values compared to the empty plasmid control.

When adjusting the amount of produced gas for the differences in OD_{600}, only the mutant expressing the hydrogenase *Cbu-c24-hydA* stood out for having a significantly higher gas volume-to-OD ratio compared to the empty plasmid control (Fig. 42 C). Comparing the amount of produced gas with the OD_{600} of all cultures revealed a strong linear correlation (R^2 = 0.93) of culture optical density and gas volume (Fig. 42 D).

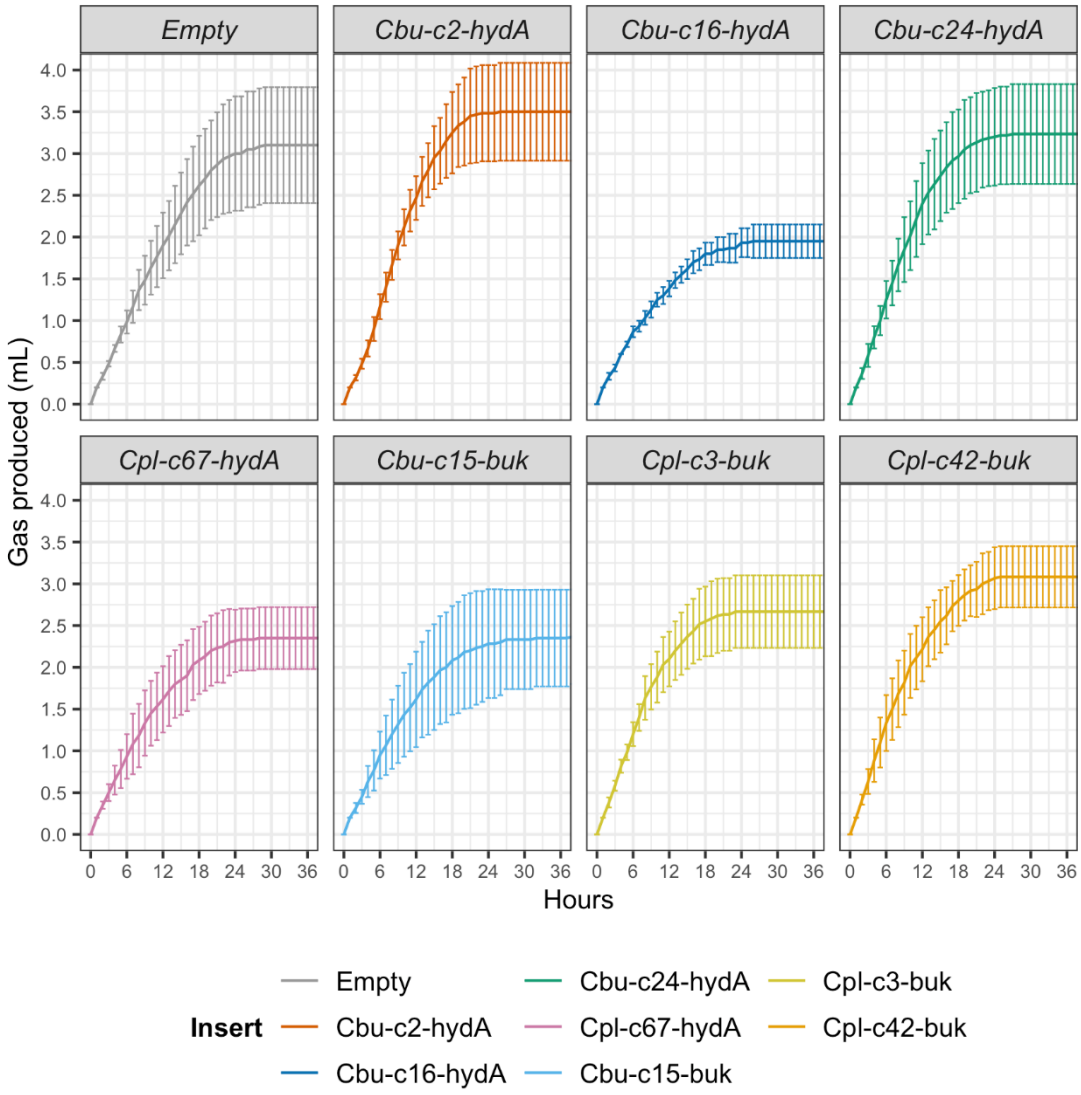


Figure 41. Gas production of *C. butyricum* syringe reactors. 3 mL cultures of *C. butyricum* mutants (pMTL8315B with 7 different inserts) cultures were grown in 10 mL syringes. The graphs show the amount of gas collected in the syringes up to 36 hours after gas production was first measurable (i.e. amount of gas produced passed 0.2 mL). Error bars show standard deviation (n = 6).

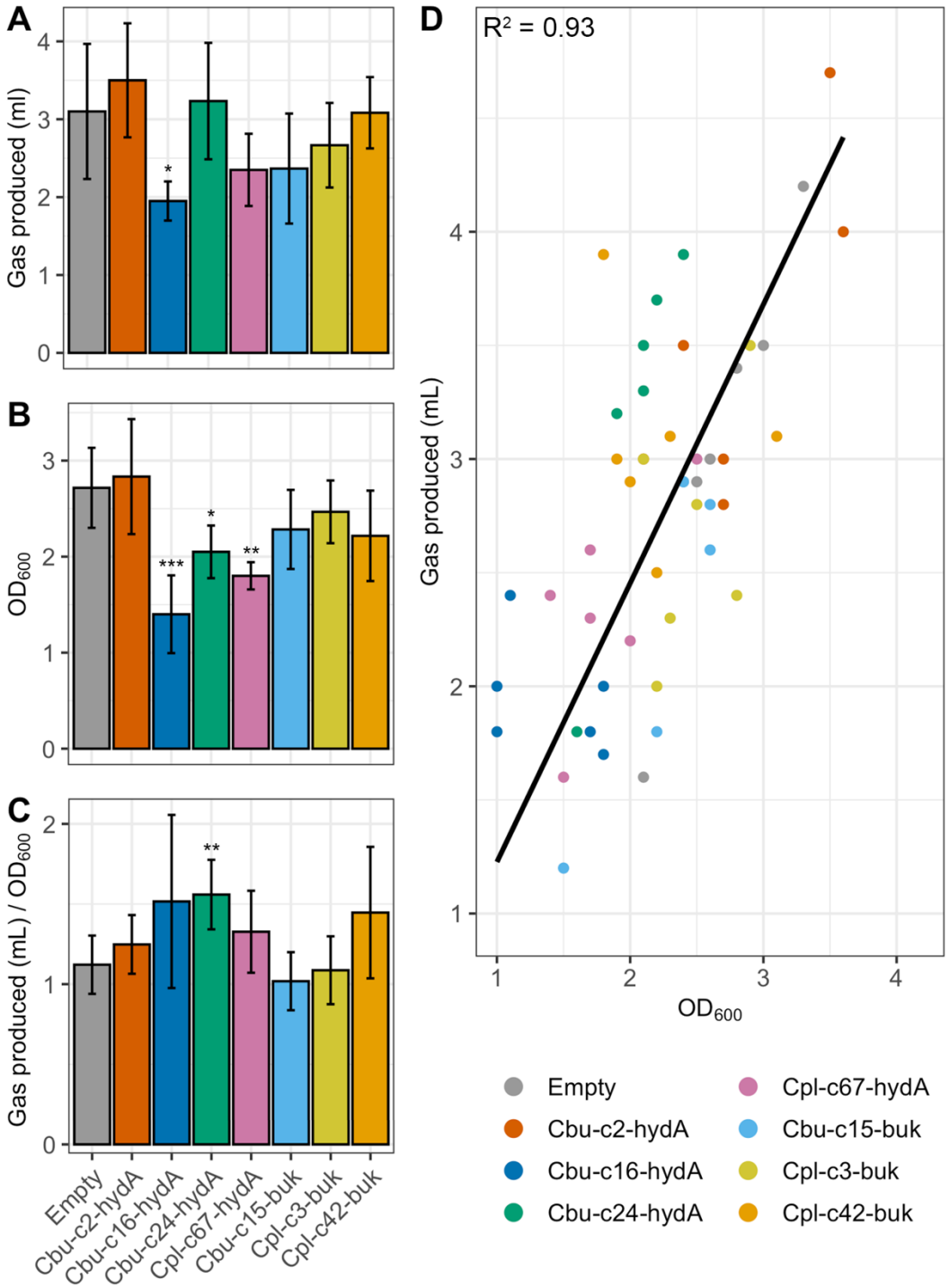


Figure 42. Gas production and OD_{600} of *C. butyricum* syringe reactors. 3 mL cultures of *C. butyricum* mutants (pMTL8315B with 7 different inserts) were grown in 10 mL syringes. Error bars show standard deviation (n = 6)
(A) shows the final amount of produced gas (mL) after cultures reached stationary phase.
(B) shows the optical density (OD_{600}) measured after gas production stopped.
(C) shows the amount of produced gas, divided by OD_{600}.
(D) compares the amount of produced gas against the culture OD_{600} of every individual sample, showing the correlation (adj. R^2 = 0.93) between culture density and produced gas. Asterisks show statistical significance compared to the empty pMTL8315B plasmid control. * = $p < 0.05$, ** = $p < 0.005$, *** = $p < 0.0005$.

6.2.5.2 *C. pluributyricum* mutants

When comparing the volume of gas produced by 3 mL syringe cultures of *C. pluributyricum* hydrogenase and butyrate kinase-expressing mutants to the empty plasmid control (Fig. 43), no significant differences in gas volume and culture OD_{600} could be observed (Fig. 44, A and B). Gas production correlated positively with OD_{600} (R^2 = 0.82), however not as strongly as in *C. butyricum* (Fig. 44 D). When the production of gas was adjusted for differences in culture density, no significant differences were observed either (Fig. 44 C).

As with the cultures of *C. butyricum*, this lack of significant differences appears amid the relatively high variation between the replicates, especially in the empty plasmid control (SD ± 0.76 mL gas produced), which was done in 9 replicates. Noticeably, one replicate of the mutant expressing *Cpl-c42-buk* unexpectedly stopped producing gas after 10 hours, only to restart gas production 20 hours later.

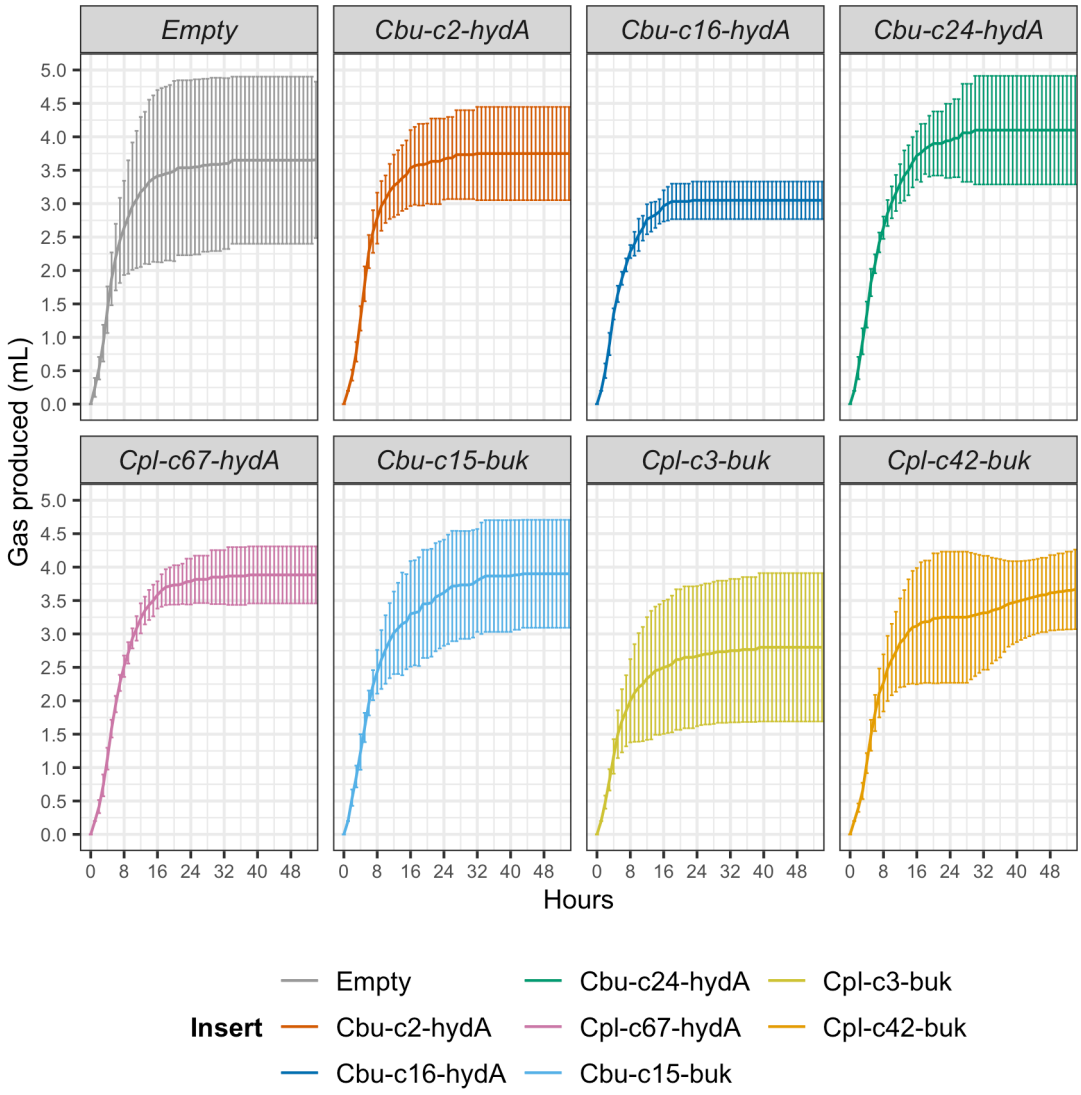


Figure 43. Gas production of *C. pluributyricum* syringe reactors. 3 mL cultures of *C. pluributyricum* mutants (pMTL8315B with 7 different inserts) cultures were grown in 10 mL syringes. Error bars show standard deviation, (n = 6; n = 8 for empty plasmid control, except at the final 54 h timepoint, where n = 9). The graphs show the amount of gas collected in the syringes up to 56 hours after gas production was first measurable (i.e. amount of gas produced passed 0.2 mL)

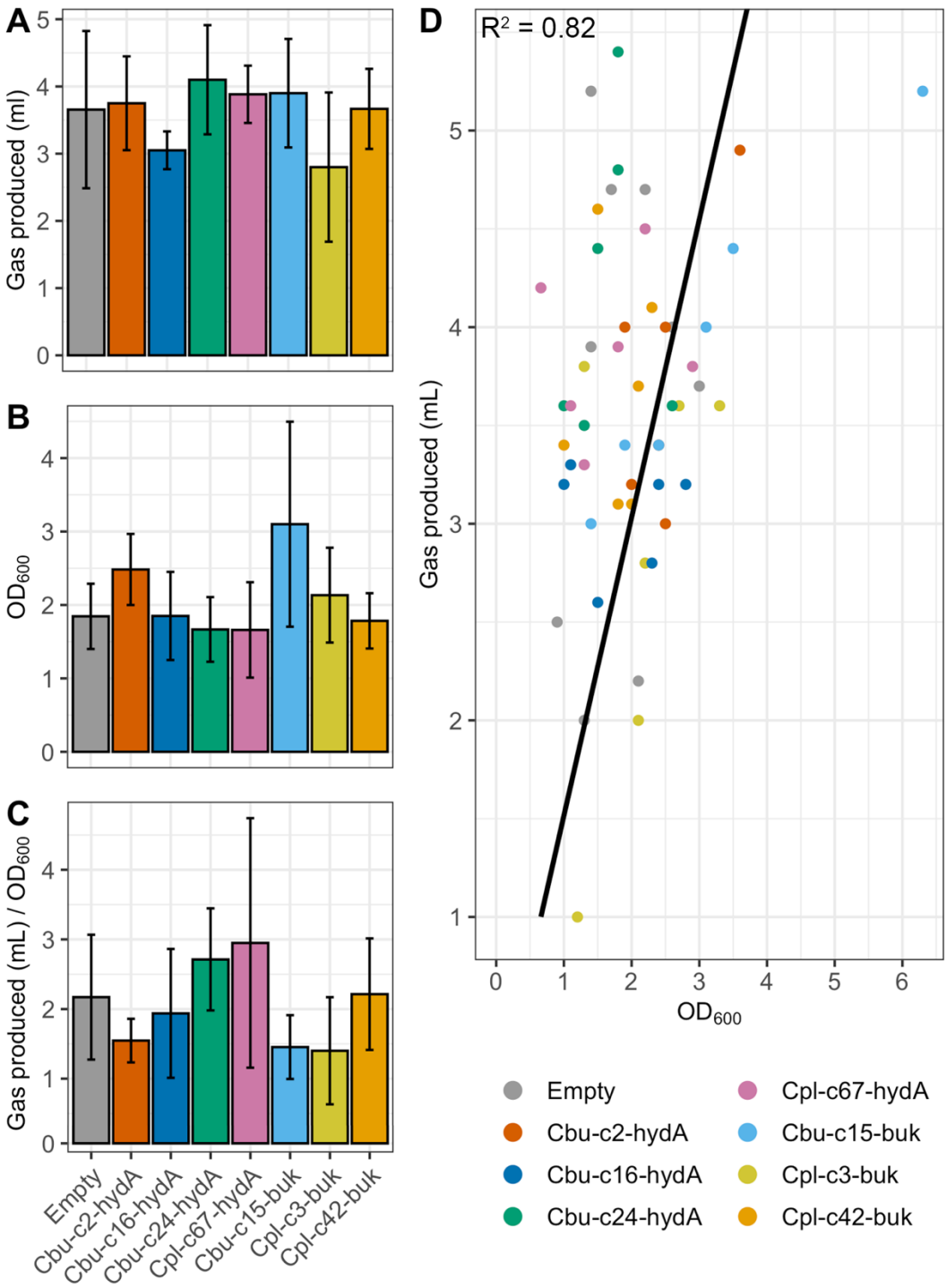


Figure 44. Gas production and OD_{600} of *C. pluributyricum* syringe reactors. 3 mL cultures of *C. pluributyricum* mutants (pMTL8315B with 7 different inserts) cultures were grown in 10 mL syringes. Error bars show standard deviation (n = 6; n = 9 for empty plasmid control).
(A) shows the final amount of produced gas (mL) after cultures reached stationary phase.
(B) shows the optical density (OD_{600}) measured after gas production stopped.
(C) shows the amount of produced gas, divided by OD_{600}.
(D) compares the amount of produced gas against the culture OD_{600} of every individual sample, showing the correlation (adj. R^2 = 0.82) between culture density and produced gas. No statistically significant differences were observed.

6.2.6 Composition of gas produced by *Clostridium* mutants

To measure whether the overexpression of hydrogenase genes has an impact on the relative amount of hydrogen produced by *Clostridium* mutants, those expressing *Cbu-c2-hydA*, *Cbu-c16-hydA*, *Cbu-c24-hydA* from *C. butyricum* and *Cpl-c67-hydA* from *C. pluributyricum* were grown in 500 mL cultures and the produced gas was collected in 1L gas bags for analysis.

Both isolates produced gas containing about 70–75 % hydrogen, with the other 25–30% being CO_2. While no differences between gas samples produced by hydrogenase mutants were observed in *C. pluributyricum*, the gas produced by *C. butyricum* mutants expressing *Cbu-c2-hydA* and *Cpl-c67-hydA* contained a significantly lower relative concentration of hydrogen (i.e. higher concentration of CO_2) when compared to the empty plasmid control, which especially contradicted initial expectations of an increase in hydrogen production when overexpressing the hydrogenase genes.

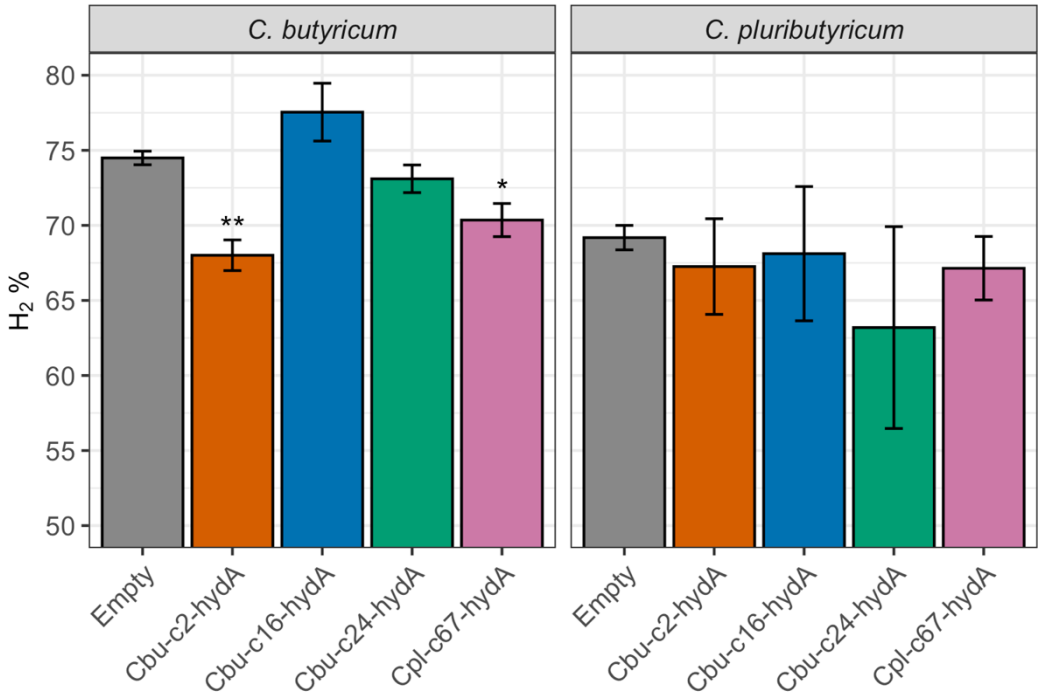


Figure 45. Hydrogen concentration in gas produced by *C. butyricum* and *C. pluributyricum* mutants. The remaining gas to 100% was made up of carbon dioxide. Error bars show standard deviation (n = 3). Asterisks above the bars denote statistical significance compared to the empty plasmid control (* = $p < 0.05$, ** = $p < 0.005$).

6.2.7 Volatile fatty acid production of *Clostridium* mutants

After the initial observation that *C. pluributyricum* produces more butyric acid than *C. butyricum* and metabolic modelling results that supported this observation based on genome data, the hypothesis arose that it may be because a difference in the number of detected butyrate kinase genes among the two species. Also, studies on other *Clostridium* species demonstrated an increase in VFA production when hydrogenase genes were overexpressed (Jo *et al.*, 2010). Therefore, the culture supernatants of *Clostridium* mutants were analysed to assess whether the expression of recombinant hydrogenases and butyrate kinases might have a measurable effect on VFA production. Here, the concentration of VFAs in supernatants from 25 mL cultures was measured using GC-FID. As with the experiments to measure culture gas production, the optical density of the individual replicates was recorded to account for possible differences in mutant growth (e.g. due to an increased metabolic burden as a result of expressing the plasmids).

6.2.7.1 C. butyricum mutants

Comparing the *hydA/buk*-overexpressing mutants with the empty plasmid control, there was a significantly increased average culture density (OD_{600}) for the hydrogenases *Cbu-c2-hydA*, *Cbu-c24-hydA*, *Cpl-c67-hydA* and the butyrate kinase mutant *Cpl-c3-buk* (Fig. 46). In cultures of mutants overexpressing *Cbu-c2-hydA* and *Cbu-c16*-hydA, culture pH was significantly lower compared to the empty plasmid control (Fig. 47). Culture density and the concentration of acetic and butyric acid correlated strongly, with denser cultures producing more VFAs (Fig. 48). Cultures of the mutant overexpressing *Cbu-c2-hydA* showed the highest average OD_{600} as well as a significantly increased concentration of acetic acid in the culture's supernatant (Fig. 49 A).

No significant differences between the concentration of butyric acid (Fig. 49 C) in culture supernatants were observed. When adjusting the concentration of VFAs for culture OD, none of the mutants appeared to produce a significantly higher amount of either butyric or acetic acid. To the contrary, the mutants expressing *Cpl-c67-hydA* and *Cpl-c3-buk* produced a significantly lower amount of acetic acid when adjusting for their significantly higher OD_{600} compared against the empty plasmid control. While mutant *Cbu-c16-hydA* produced a significantly lower amount of butyric acid compared to other mutants, this was not significantly different compared to the control, mainly due to the large variations seen between the control replicates.

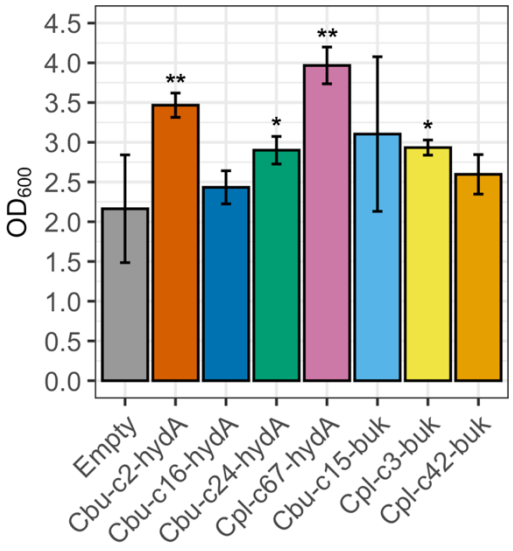


Figure 46. Optical density (OD_{600}) of *C. butyricum* mutant cultures before filtered supernatant was taken for GC-FID. Error bars show standard deviation (n = 6 for empty plasmid control and Cbu-c15-buk, n = 3 for all others). Asterisks show statistical significance compared to the empty pMTL8315B plasmid control (* = $p < 0.05$, ** = $p < 0.005$).

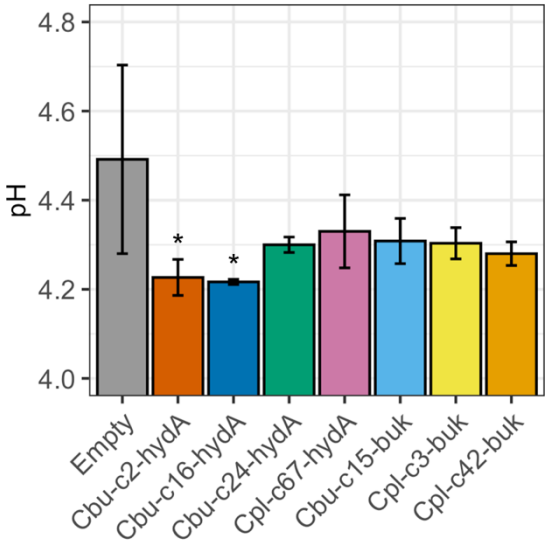


Figure 47. pH of filtered *C. butyricum* mutant culture supernatant taken for GC-FID. Error bars show standard deviation (n = 6 for empty plasmid control and Cbu-c15-buk, n = 3 for all others). Asterisks show statistical significance compared to the empty pMTL8315B plasmid control (* = $p < 0.05$)

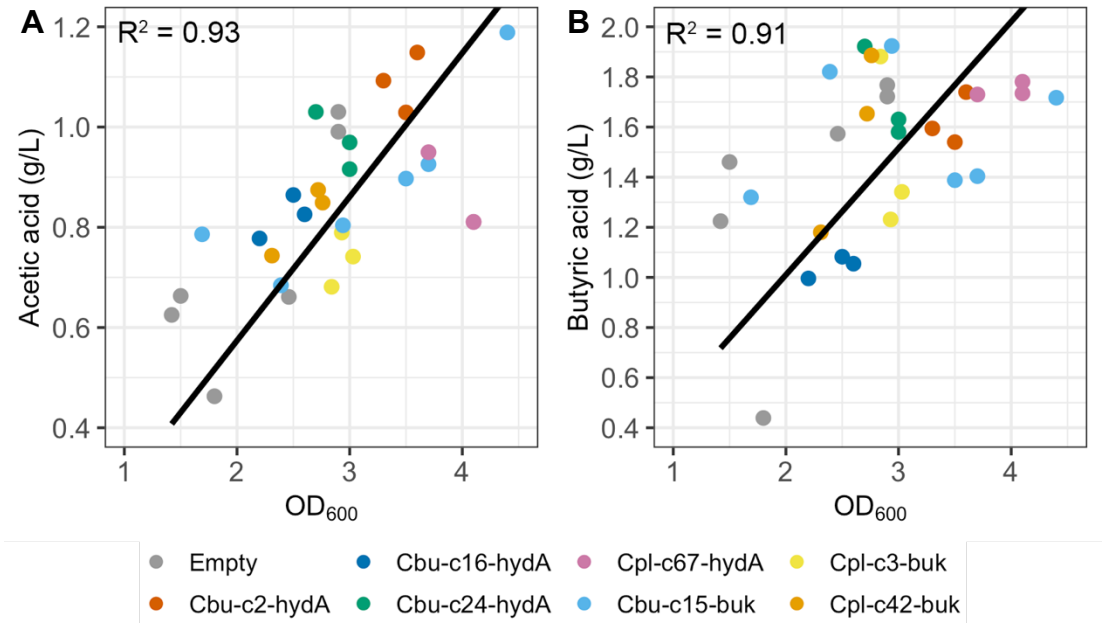


Figure 48. Optical density (OD_{600}) of *C. butyricum* mutant cultures (25 mL) plotted against the concentration of acetic acid (A) and butyric acid (B) from culture supernatant.

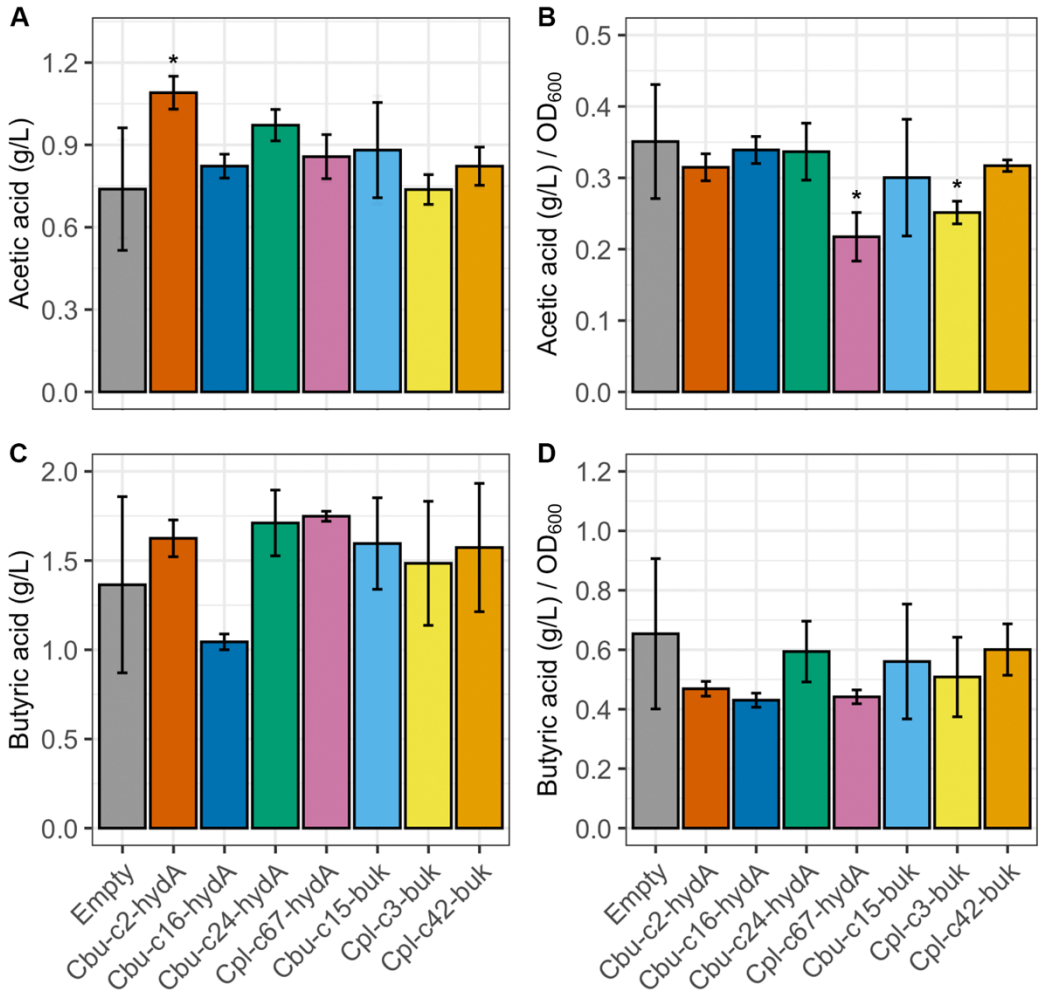


Figure 49. Concentrations of acetic acid (A) and butyric acid (C) in supernatants of *C. butyricum* mutants with 7 inserts as well as an empty plasmid control. Acid concentrations present in the supernatant of the non-inoculated control (0.04 g/L acetic acid and 0.03 g/L butyric acid) was subtracted from samples to reflect amounts produced by bacteria. Acid concentrations adjusted to culture OD_{600} are shown in panels B and D. Error bars show standard deviation (n = 6 for empty plasmid control and Cbu-c15-buk, n = 3 for all others). Asterisks show statistical significance compared to the empty pMTL8315B plasmid control (* = $p < 0.05$).

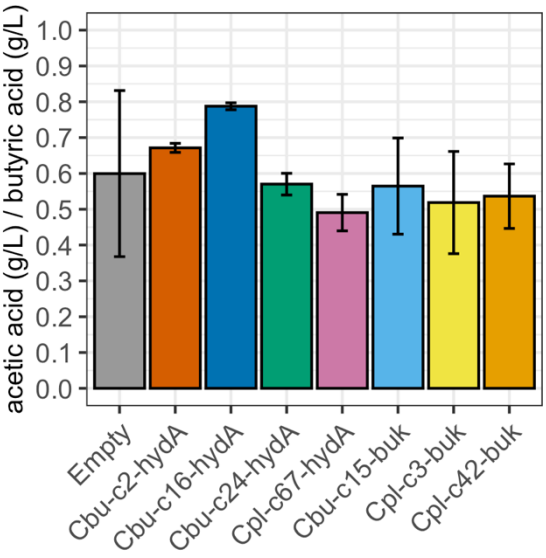


Figure 50. Ratio of produced acetic acid and butyric acid in *C. butyricum* mutant culture supernatant. Error bars show standard deviation (n = 6 for empty plasmid control and Cbu-c15-buk, n = 3 for all others). No statistically significant differences compared to the empty plasmid control were observed.

6.2.7.2 C. pluributyricum mutants

C. pluributyricum mutants that expressed one of the seven tested inserts did not show any significant differences in culture density (OD_{600}, Fig. 51) and culture pH (Fig. 52). The correlation between OD_{600} and the measured concentrations of VFAs (Fig. 53) was lower than in *C. butyricum*.

The mutant expressing the butyrate kinase *Cpl-c3-buk* was recognised as the only mutant producing a significantly higher yield of acetic acid (0.81 g/L ± 0.05 SD; p = 0.0006) compared to the empty plasmid control, while the mutant expressing the hydrogenase *Cbu-c2-hydA* produced a significantly lower amount (0.29 g/L ± 0.02 SD; p = 0.02; Fig. 54).

Still, when considering the ratio of acetic acid to butyric acid (Fig. 55), the production of VFAs was significantly more skewed towards acetic acid in mutants *Cpl-c3-buk* (0.48 ± 0.07 SD, p = 0.006), *Cbu-c16-hydA* (0.29 ± 0.04 SD, p = 0.02) and *Cbu-c24-hydA* (0.31 ± 0.01 SD, p = 0.02) compared to the empty plasmid control (0.16 ± 0.04 SD)

When adjusting for the OD_{600} value of the individual samples, the only sample producing a higher amount of acetic acid per OD_{600} was the mutant expressing the hydrogenase *Cbu-c24-buk* (Fig. 54 B). All mutants overexpressing either

of the four hydrogenases or three butyrate kinase genes showed a lower concentration of butyric acid in their culture's supernatant (Fig. 54 C), however amid high deviations between replicates of the empty plasmid control, this was not significant unless adjusted for OD_{600} (Fig. 54 D). When adjusting for culture density, hydrogenases *Cbu-c16-hydA, Cbu-c24-hydA* as well as all butyrate kinase mutants showed a significantly lower concentration of butyric acid per OD_{600}, compared to the empty plasmid control.

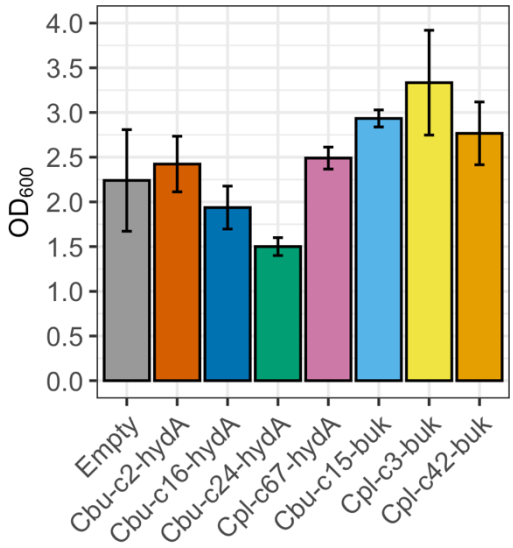


Figure 51. Optical density (OD_{600}) of *C. pluributyricum* mutant cultures before filtered supernatant was taken for GC-FID. Error bars show standard deviation (n = 3). No statistically significant differences compared to the empty plasmid control were observed.

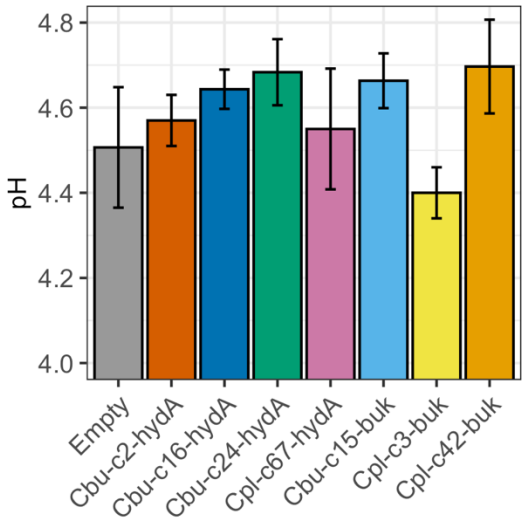


Figure 52. pH of filtered *C. pluributyricum* mutant culture supernatant taken for GC-FID. Error bars show standard deviation (n = 3). No statistically significant differences compared to the empty plasmid control were observed.

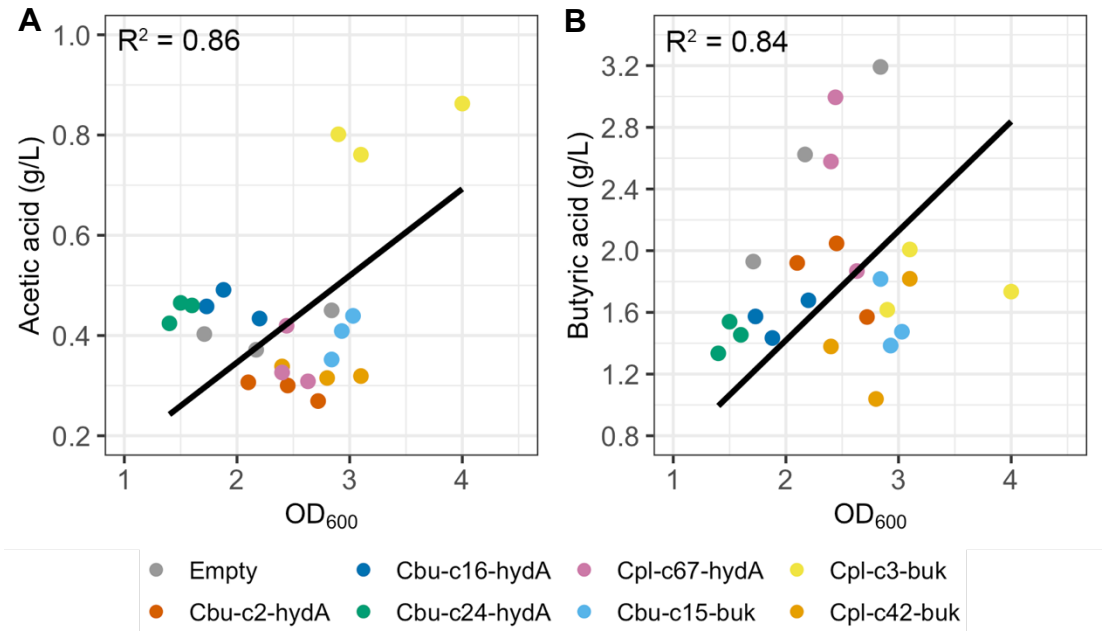


Figure 53. Optical density (OD_{600}) of *C. pluributyricum* mutant cultures (25 mL) plotted against the concentration of acetic acid (A) and butyric acid (B) from culture supernatant.

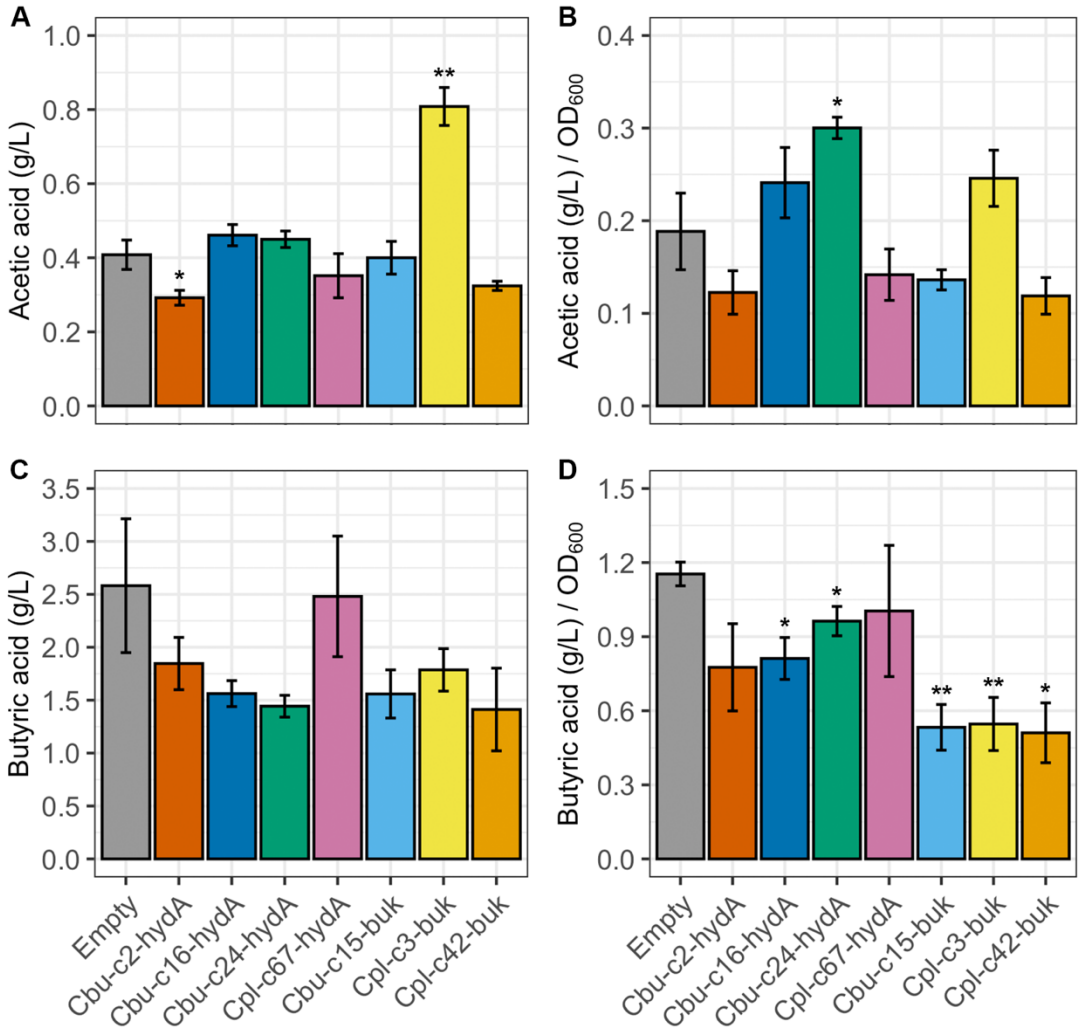


Figure 54. Concentrations of acetic acid (A) and butyric acid (C) in supernatants of *C. pluributyricum* mutants with 7 inserts as well as an empty plasmid control. Acid concentrations present in the supernatant of the non-inoculated control (0.04 g/L acetic acid and 0.03 g/L butyric acid) was subtracted from samples to reflect amounts produced by bacteria. Acid concentrations adjusted to culture OD_{600} are shown in panels B and D. Error bars show standard deviation (n = 3). Asterisks show statistical significance compared to the empty pMTL8315B plasmid control (* = $p < 0.05$; ** = $p < 0.005$).

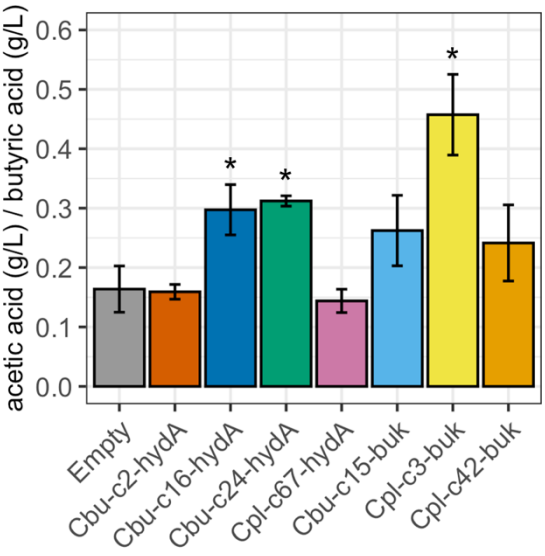


Figure 55. Ratio of produced acetic acid and butyric acid in *C. pluributyricum* mutant culture supernatant. Error bars show standard deviation (n = 3). Asterisks show statistical significance compared to the empty pMTL8315B plasmid control (* = $p < 0.05$).

6.3 Discussion

6.3.1 Promoters and recombinant gene expression

When the expression of *E. coli* β-glucuronidase (*gusA*) under the *C. butyricum* and *C. pluributyricum* thiolase / ferredoxin promoters was tested, expression was strongest under both promoters from *C. butyricum,* in either of the two isolates. As the portability of the *C. acetobutylicum* thiolase promoter to the moderately related *C. sporogenes* (91.6% 16S rRNA similarity) has been described, the functionality of *C. butyricum* promoters in the closely related *C. pluributyricum* was not necessarily surprising. Nevertheless, while functional, the *C. pluributyricum* thiolase promoter had very low activity in *C. butyricum*. Conversely, it was unexpected to see that the *C. butyricum* ferredoxin promoter evoked a stronger expression of *gusA* in *C. pluributyricum* than it did in *C. butyricum*, while the ferredoxin promoters of *C. pluributyricum* and *C. sporogenes* did not work at all in either species. Previously, the *C. sporogenes* ferredoxin promoter was used successfully in *C. difficile* (Zhang *et al.*, 2015) as well as *C. pluributyricum's* close relative *C. saccharoperbutylacetonicum* (Grosse-Honebrink *et al.*, 2021), which made this result yet more surprising. However, Mordaka and Heap (2018) reported that promoter sequences in *Clostridium* are very stringent, which may explain these results.

Generally, expression of *buk* and *hydA* genes under the chosen promoter was significantly increased, however the genes *Cpl-c67-hydA* in *C. pluributyricum* and *Cbu-c15-buk* in *C. butyricum* stood out for being highly expressed in the empty plasmid control mutants, meaning that the respective level of overexpression was low or insignificant, respectively. As *C. pluributyricum* and *C. butyricum* only possess one *hydA* / *buk* gene, respectively, it would be worthwhile to investigate whether the availability of multiple gene orthologs results in the lower expression of individual orthologs. A possible reason for this might be that multiple orthologs share roles within the cell's metabolism, which may be studied by knocking out one of the orthologs and monitoring the expression of the remaining gene(s). Conversely, as *C. butyricum* lacks other genes to produce butyric acid from acetic acid or acetoacetate, *Cbu-c15-buk* may be highly expressed as the sole enzyme in

this vital pathway. Nevertheless, despite lower levels of overexpression detected via RT-qPCR, protein lysates from mutants expressing His-tagged versions of the *buk* / *hydA* genes clearly produced corresponding bands on Western blots, indicating that the genes introduced via the pMTL80000 *E. coli* – *Clostridium* shuttle plasmids are being expressed.

6.3.2 Gas and volatile fatty acid production of *Clostridium* mutants

6.3.2.1 Gas volume

The growth of *Clostridium* and other bacteria in sealed syringes to measure the volume of their produced gas has been a long-standing common practice (e.g. Richards *et al.*, 1968; Sacks and Olson, 1979; Nche *et al.*, 1994; Hakalehto and Hänninen, 2012; Mumme *et al.*, 2014), and using this method with cultures of the two *Clostridium* isolates enabled the measurement of gas volumes produced by a large number of mutant cultures in parallel.

Nevertheless, the use of gas syringes also has some downsides as observed in comparisons with pressure-free systems that measure gas through liquid displacement. In the study by Nche, Nout and Rombouts (1994), the authors discuss that, due to the force that is required to lift the syringe plunger, gas volumes can only be measured after the produced gas is at sufficient pressure. In cultures of *C. perfringens*, this resulted in a 2-hour delay before gas production was observed, higher variability between replicates due to friction between plunger and syringe, and a generally lower measured gas volume.

Abdeshahian *et al.* (2014) studied the effect of culture headspace on hydrogen production of a novel *Clostridium* species and found that larger headspaces improved hydrogen production, confirming earlier observations in *C. acetobutylicum* (Oh *et al.*, 2009). Conversely, lower atmospheric pressures have been associated with the increased production of hydrogen and lower yields of reduced fermentation products (e.g. ethanol) from anaerobic sludge (Lee *et al.*, 2012), even though Kataoka *et al.* (1997) did not observe any

effects on hydrogen production when cultures of *C. butyricum* were grown in a vacuum.

In addition, *C. butyricum* and *C: beijerinckii* have been shown to produce hydrogen through other pathways such as nitrogenase-mediated nitrogen fixation (Chen, Toth and Kasap, 2001; Calusinska *et al.*, 2015), which was not studied here but could represent another variable, especially as the lack of headspace would have reduced the diffusion of N_2 from the anaerobic gas mixture into the medium compared to cultures grown in flasks.

Therefore, the lack of a headspace in syringe reactors and differences in culturing conditions must be kept in mind when comparing the results from these gas volume experiments to those in which gas composition and VFA production was measured, or results from other studies.

Nevertheless, significant differences between sets of samples were recorded despite high variability, at least in cultures of *C. butyricum*. Here, a significantly lower culture density was observed in three of the four hydrogenase mutants. This coincided in a significantly lower production of gas by the *Cbu-c16-hydA* mutant, which was however proportional to the decrease in OD_{600}, the reason of which is not known. In the case of the *Cbu-c24-hydA* mutant, the lower culture density with no significant changes in gas production meant that cultures produced an overproportional amount of gas when adjusting for OD_{600}.

In the comparable study by Morimoto *et al.* (2005), the overexpression of a *C. paraputrificum* hydrogenase resulted in the production of more gas without impacting growth. Using BLAST to search against the *C. butyricum* genome, the overexpressed gene from this study was found to be most closely related to *Cbu-c24-hydA* (73% query cover, 75.95% identity).

Still, when comparing the final optical density of cultures grown in either syringes or flasks (i.e. for gas composition and GC-FID analysis), those grown in flasks often reached a higher OD_{600}, and trends seen in syringe cultures were sometimes reversed, such as in cultures of the *C. butyricum* mutant expressing *Cpl-c67-hydA*.

It is very likely that the lower growth in syringes was due to the higher partial pressure of gas because of the lack of headspace. If these experiments were to be repeated, the use of pressure-free systems may help to reduce the high amount of variation between samples and increase hydrogen yield.

VFA concentration and the pH of syringe cultures was not compared to flask cultures, however the higher amount of dissolved hydrogen is likely to also affect the production of reduced products, as well as culture pH.

6.3.2.2 Gas composition

Due to time constraints, the composition of gas produced by *Clostridium* mutants was only assessed for mutants overexpressing hydrogenases. Significantly lower levels of hydrogen (i.e. higher amounts of CO_2) were seen in *C. butyricum* mutants overexpressing *Cbu-c2-hydA* and *Cpl-c67-hydA*. As discussed, the ratio of hydrogen to carbon dioxide is usually linked to the ratio between produced amounts of acetic acid and butyric acid, as well as the production of reduced fermentation products, which means that less hydrogen can be released as gas. As gas composition measurements required volumes of around 500 mL of gas to be collected per sample, these cultures were grown in 500 mL flasks. Culture conditions may therefore be different to the 25 mL cultures from which supernatant samples were taken for GC-FID analysis. Here, the two mutants that showed lower H_2 concentrations were seen to have significantly higher culture densities compared to the control. For the *C. butyricum* mutant expressing *Cbu-c2-hydA*, a significantly higher concentration of acetic acid was observed, which is usually associated with higher concentrations of hydrogen in produced gas, which is contrary to what was seen. Furthermore, the same effect was not observed in *C. pluributyricum* mutants. The overexpression of the his-tagged *hydA* genes was confirmed in the Western blots, which means that the actual expression of the non-His-tagged genes is likely as well, which is further demonstrated via qPCR. Therefore, the gas composition measurement results showing significantly lower H_2 concentrations for these mutants cannot be fully explained.

6.3.2.3 Volatile fatty acid production

One of the results that was most significant was the increased production of acetic acid by the *C. pluributyricum* mutant overexpressing its own butyrate kinase, *Cpl-c3-buk.* This was very unexpected as butyrate kinase is involved in the production of butyric acid, however a possible explanation is that the increased butyrate kinase activity results in a depletion of butyryl-CoA. As this is a necessary substrate in the production of butyric acid from acetic acid through butyryl-CoA:acetate CoA transferase, the accumulation of acetic acid is plausible. In order to test this hypothesis, the expression level of butyryl-CoA:acetate CoA transferase could be measured using RT-qPCR, as it may be affected in the mutant. However, this may not be completely conclusive as transcript levels may not necessarily indicate protein activity and other methods, such as cell-free extracts may be used to test the activity of this enzyme (Duncan *et al.*, 2002). To test the effect of this gene's activity on acetic acid concentrations in general, it could be knocked out to study the effects on acetic acid concentrations in the supernatant. Interestingly, the amount of butyric acid produced by all three *C. pluributyricum* butyrate kinase mutants was significantly lower when adjusted for OD, but the ratio of produced acetic and butyric acid was only skewed for the *Cpl-c3-buk* mutant. The reason why similar effects were not seen in *C. butyricum* mutants might be because *C. butyricum* does not possess the butyryl-CoA:acetate-CoA transferase to produce butyrate from acetate. Therefore, the overexpression of this gene in *C. butyricum* would be a way to test this hypothesis and potentially increase its production of butyric acid.

This ratio of acetic acid and butyric acid was also significantly higher for *C. pluributyricum* mutants overexpressing the *C. butyricum* hydrogenases *Cbu-c16-hydA* and *Cbu-c24-hydA*, meaning that they produced relatively higher amounts of acetic acid compared to butyric acid. Both mutants produced significantly lower amounts of butyric acid when adjusted for OD and the *Cbu-c24-hydA* mutant also produced more acetic acid when adjusted for OD. Overall this reflects results reported by Morimoto *et al.* (2005), who found that hydrogenase overexpression in *C. paraputrificum* resulted in an increase

of acetic acid production due to the over-oxidation of NADH. For the same reason, this may also have inhibited the production of reduced compounds such as butyric acid (see *C. pluributyricum Cbu-c16-hydA* and *Cbu-c24*-hydA) and possibly lactic acid. Conversely, Pyne *et al.* (2015) found that the antisense RNA knockdown of *hydA* in *C. pasteurianum* resulted in a 40% increase in the production of reduced metabolites such as butyric acid, explainable due to excess NADH.

Among all *C. butyricum* mutants, the average ratio of acetic/butyric acid was highest in mutants overexpressing *Cbu-c2-hydA* and *Cbu-c16-hydA*, however this was not significantly increased due to large variation in the control. As *C. butyricum* possesses multiple hydrogenase genes, the overexpression of one of them may have less of an effect compared to overexpression in *C. pluributyricum*.

6.3.3 Conclusion

The newly isolated strain of *C. butyricum* and the novel species *C. pluributyricum* have been genetically engineered for the first time. Overexpressing selected butanoate pathway genes resulted in significantly different phenotypes compared to the empty plasmid control, however challenges concerning variation between samples were observed. Some of the results are explainable by comparison with the literature, however further experiments are necessary to fully understand the effect of *hydA*/*buk* overexpression and engineer these species for their enhanced production of VFAs or hydrogen. The three newly identified promoters of different strengths can aid in the metabolic engineering of these isolates, and further approaches such as the creation of deletion mutants or integrating genes into their genome may give better comparisons to the wild type.

7 Conclusion

The results presented here have given insights into the inner workings of an industrial pilot-stage anaerobic baffled reactor by characterising its microbial community and isolating several species that have contributed to its function. Identifying patterns in the abundance of organisms along the reactor's seven chambers has allowed to unravel the four successive stages of anaerobic digestion and observe how each stage is reflected by noticeable changes in the metataxonomic dataset. Additionally, the results have demonstrated how process parameters such as loading rates or salinity can be inferred from the community composition. Compared to other available reports on the microbial communities in ABRs, this study is unique due to the reactor's size and the processing of grass and seaweed, two feedstocks that are considered sustainable sources of second- and third-generation biofuels, respectively. The comparison of the microbiota associated with the digestion of these two different types of lignocellulosic biomass has revealed differences that are largely characterised by a higher abundance of methanogens in grass digestate, indicating that the choice of feedstocks may be another key factor when the aim is to control this process for different desired end-products (e.g. VFAs, biogas).

Furthermore, the isolation of a new strain of *Clostridium butyricum* as well as a newly identified species within the genus *Clostridium sensu stricto* has extended the known limits on the genus's genome size and led to noteworthy observations about the evolutionary relationship of species within the genus. The identification of three working promoters with varying strengths and the subsequent genetic engineering of the two isolates to induce the plasmid-mediated overexpression of key genes in the butanoate pathway have laid the foundation for further studies on the species' metabolic capabilities. Such studies would benefit from a more detailed metabolic flux analysis to identify all fermentation products produced by the two species and possible pathway bottlenecks. In addition, the growth of these organisms at a larger scale and in more controlled conditions, using pH-adjusted continuous stirred-tank reactors would allow to assess their potential for industrial use.

Returning these isolates back to the reactor they were originally isolated from would be another opportunity to investigate how they contribute to the process of anaerobic digestion. In a recent study by Tukanghan *et al.* (2021), methane production in small scale (500 mL) batch-fed anaerobic digesters was enhanced by enriching their microbial consortia with *Clostridium* species. Applying similar methods to enrich the communities of different chambers in the industrial partner's ABR would allow to measure their persistence as well as resulting shifts in species abundance, which may steer the process into various directions. If the study on this reactor was to be expanded, it would greatly benefit from regular sampling of digestate and the recording of process parameters such as pH and temperature as well as methane production. To monitor the production of biogas and hydrogen, and investigate the connection between acetogen/methanogen abundance and gas composition, it may be beneficial to re-design the reactor to divide the headspace into separate chambers as well, so that gas produced from communities in different compartments can be sampled and analysed individually, similar to the design by Jürgensen *et al.* (2015).

By lifting the lid on the "black box" that describes the microbiome of bioreactors (Koch *et al.*, 2014), this project has contributed to improve our understanding of the multi-layered complexity found in the process of anaerobic digestion by focusing on the organisms that shape it.

References

Abdeshahian, P. *et al.* (2014) 'The production of biohydrogen by a novel strain Clostridium sp. YM1 in dark fermentation process', *International Journal of Hydrogen Energy*, 39(24), pp. 12524–12531. doi: 10.1016/j.ijhydene.2014.05.081.

Abt, B. *et al.* (2013) 'Genome sequence of the thermophilic fresh-water bacterium Spirochaeta caldaria type strain (h1t), reclassification of Spirochaeta caldaria, Spirochaeta stenostrepta, and Spirochaeta zuelzerae in the genus Treponema as Treponema caldaria comb. nov.', *Standards in Genomic Sciences*, 8(1), pp. 88–105. doi: 10.4056/sigs.3096473.

Acinas, S. G. *et al.* (2004) 'Divergence and Redundancy of 16S rRNA Sequences in Genomes with Multiple rrn Operons', *Journal of Bacteriology*, 186(9), pp. 2629–2635. doi: 10.1128/JB.186.9.2629-2635.2004.

Adekunle, K. F. and Okolie, J. A. (2015) 'A Review of Biochemical Process of Anaerobic Digestion', *Advances in Bioscience and Biotechnology*, 06(03), pp. 205–212. doi: 10.4236/abb.2015.63020.

Anderson, M. J. (2001) 'A new method for non-parametric multivariate analysis of variance', *Austral Ecology*, 26(1), pp. 32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x.

Anjuwon-Foster, B. R. and Tamayo, R. (2017) 'A genetic switch controls the production of flagella and toxins in Clostridium difficile', *PLoS Genetics*, 13(3), p. e1006701. doi: 10.1371/journal.pgen.1006701.

Apprill, A. *et al.* (2015) 'Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton', *Aquatic Microbial Ecology*, 75(2), pp. 129–137.

Bachmann, A., Beard, V. L. and McCarty, P. L. (1985) 'Performance characteristics of the anaerobic baffled reactor', *Water Research*. doi: 10.1016/0043-1354(85)90330-6.

Baeyens, J. *et al.* (2020) ‘Reviewing the potential of bio-hydrogen production by fermentation’, *Renewable and Sustainable Energy Reviews*, 131, p. 110023. doi: 10.1016/J.RSER.2020.110023.

Ban, Q. *et al.* (2013) ‘Linking performance with microbial community characteristics in an anaerobic baffled reactor’, *Applied Biochemistry and Biotechnology*, 169(6), pp. 1822–1836. doi: 10.1007/s12010-013-0105-6.

Barber, W. P. and Stuckey, D. C. (1999) ‘The use of the anaerobic baffled reactor (ABR) for wastewater treatment: a review’, *Water Research*, 33(7), pp. 1559–1578.

Beckers, L. *et al.* (2015) ‘Investigation of the links between mass transfer conditions, dissolved hydrogen concentration and biohydrogen production by the pure strain Clostridium butyricum CWBI1009’, *Biochemical Engineering Journal*, 98, pp. 18–28. doi: 10.1016/j.bej.2015.01.008.

Bertels, F. *et al.* (2014) ‘Automated reconstruction of whole-genome phylogenies from short-sequence reads’, *Molecular Biology and Evolution*, 31(5), pp. 1077–1088. doi: 10.1093/molbev/msu088.

Bertucci, M. *et al.* (2019) ‘Carbohydrate hydrolytic potential and redundancy of an anaerobic digestion microbiome exposed to acidosis, as uncovered by metagenomics’, *Applied and Environmental Microbiology*, 85(15). doi: 10.1128/AEM.00895-19/SUPPL_FILE/AEM.00895-19-S0001.PDF.

Bisanz, J. (2018) *jbisanz/qiime2R: Import qiime2 artifacts to R*, *GitHub*. Available at: https://github.com/jbisanz/qiime2R/ (Accessed: 31 December 2021).

Bogolitsyn, K. G. *et al.* (2017) ‘Isolation and Structural Characterization of Cellulose from Arctic Brown Algae’, *Chemistry of Natural Compounds*, 53(3), pp. 533–537. doi: 10.1007/s10600-017-2039-7.

Bolyen, E. *et al.* (2019) ‘Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2’, *Nature Biotechnology 2019 37:8*,

37(8), pp. 852–857. doi: 10.1038/s41587-019-0209-9.

Boone, D. R. (2015) ‘Methanobacterium’, in *Bergey’s Manual of Systematics of Archaea and Bacteria*. John Wiley & Sons, Ltd, pp. 1–8. doi: 10.1002/9781118960608.gbm00495.

Booth, W. T. *et al.* (2018) ‘Impact of an N-terminal polyhistidine tag on protein thermal stability’, *ACS Omega*, 3(1), pp. 760–768. doi: 10.1021/ACSOMEGA.7B01598/SUPPL_FILE/AO7B01598_SI_001.PDF.

Bouallagui, H. *et al.* (2005) ‘Bioreactor performance in anaerobic digestion of fruit and vegetable wastes’, *Process biochemistry*, 40(3–4), pp. 989–995.

Bowles, L. K. and Ellefson, W. L. (1985) ‘Effects of butanol on Clostridium acetobutylicum’, *Applied and Environmental Microbiology*, 50(5), pp. 1165–1170. doi: 10.1128/AEM.50.5.1165-1170.1985.

Braz, G. H. R. *et al.* (2019) ‘Organic overloading affects the microbial interactions during anaerobic digestion in sewage sludge reactors’, *Chemosphere*, 222, pp. 323–332. doi: 10.1016/j.chemosphere.2019.01.124.

Bruder, M. R. *et al.* (2016) ‘Extending CRISPR-Cas9 technology from genome editing to transcriptional engineering in the genus Clostridium’, *Applied and Environmental Microbiology*, 82(20), pp. 6109–6119. doi: 10.1128/AEM.02128-16/SUPPL_FILE/ZAM999117454SO1.PDF.

Bukin, Y. S. *et al.* (2019) ‘The effect of 16S rRNA region choice on bacterial community metabarcoding results’, *Scientific Data 2019 6:1*, 6(1), pp. 1–14. doi: 10.1038/sdata.2019.7.

Cabanettes, F. and Klopp, C. (2018) ‘D-GENIES: Dot plot large genomes in an interactive, efficient and simple way’, *PeerJ*, 2018(6), p. e4958. doi: 10.7717/peerj.4958.

Cai, H. *et al.* (2007) ‘Genotypic and phenotypic characterization of Lactobacillus casei strains isolated from different ecological niches suggests

frequent recombination and niche specificity', *Microbiology*, 153(8), pp. 2655–2665. doi: 10.1099/mic.0.2007/006452-0.

Cai, M. *et al.* (2009) 'Optimal production of polyhydroxyalkanoates (PHA) in activated sludge fed by volatile fatty acids (VFAs) generated from alkaline excess sludge fermentation', *Bioresource Technology*, 100(3), pp. 1399–1405. doi: 10.1016/j.biortech.2008.09.014.

Cai, M. *et al.* (2016) 'Metagenomic reconstruction of key anaerobic digestion pathways in municipal sludge and industrial wastewater biogas-producing systems', *Frontiers in Microbiology*, 7(MAY). doi: 10.3389/fmicb.2016.00778.

Callahan, B. J. *et al.* (2016) 'DADA2: high-resolution sample inference from Illumina amplicon data', *Nature methods*, 13(7), p. 581.

Callahan, B. J., McMurdie, P. J. and Holmes, S. P. (2017) 'Exact sequence variants should replace operational taxonomic units in marker-gene data analysis', *The ISME journal*, 11(12), p. 2639.

Calusinska, M. *et al.* (2015) 'Genome-wide transcriptional analysis suggests hydrogenase- and nitrogenase-mediated hydrogen production in Clostridium butyricum CWBI 1009', *Biotechnology for Biofuels*, 8(1), pp. 1–16. doi: 10.1186/S13068-015-0203-5/TABLES/1.

Caporaso, J. G. *et al.* (2011) 'Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample', *Proceedings of the national academy of sciences*, 108(Supplement 1), pp. 4516–4522.

Caporaso, J. G. *et al.* (2012) 'Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms', *The ISME journal*, 6(8), pp. 1621–1624.

Charubin, K. *et al.* (2018) 'Engineering Clostridium organisms as microbial cell-factories: challenges & opportunities', *Metabolic Engineering*, 50, pp. 173–191. doi: 10.1016/J.YMBEN.2018.07.012.

Chatellard, L., Trably, E. and Carrère, H. (2016) 'The type of carbohydrates specifically selects microbial community structures and fermentation patterns', *Bioresource technology*, 221, pp. 541–549. doi: 10.1016/J.BIORTECH.2016.09.084.

Chen, C. *et al.* (2016) 'Evaluation of COD effect on anammox process and microbial communities in the anaerobic baffled reactor (ABR)', *Bioresource Technology*, 216, pp. 571–578. doi: 10.1016/J.BIORTECH.2016.05.115.

Chen, J.-S., Toth, J. and Kasap, M. (2001) 'Nitrogen-fixation genes and nitrogenase activity in Clostridium acetobutylicum and Clostridium beijerinckii', *Journal of Industrial Microbiology and Biotechnology*, 27(5), pp. 281–286. doi: 10.1038/sj.jim.7000083.

Chen, J. *et al.* (2018) 'Whole-genome sequence of phage-resistant strain Escherichia coli DH5α', *Genome Announcements*. doi: 10.1128/genomeA.00097-18.

Chen, W. M. *et al.* (2005) 'Fermentative hydrogen production with Clostridium butyricum CGS5 isolated from anaerobic sewage sludge', *International Journal of Hydrogen Energy*, 30(10), pp. 1063–1070. doi: 10.1016/j.ijhydene.2004.09.008.

Cho, D. H., Shin, S. J. and Kim, Y. H. (2012) 'Effects of acetic and formic acid on ABE production by clostridium acetobutylicum and clostridium beijerinckii', *Biotechnology and Bioprocess Engineering*, 17(2), pp. 270–275. doi: 10.1007/s12257-011-0498-4.

Chong, M. L. *et al.* (2009) 'Optimization of biohydrogen production by Clostridium butyricum EB6 from palm oil mill effluent using response surface methodology', *International Journal of Hydrogen Energy*, 34(17), pp. 7475–7482. doi: 10.1016/j.ijhydene.2009.05.088.

Chong, S. C. and Boone, D. R. (2015) 'Methanoculleus', in *Bergey's Manual of Systematics of Archaea and Bacteria*. John Wiley & Sons, Ltd, pp. 1–5.

doi: 10.1002/9781118960608.gbm00505.

Chowdhury, H. and Loganathan, B. (2019) ‘Third-generation biofuels from microalgae: a review’, *Current Opinion in Green and Sustainable Chemistry*. Elsevier, pp. 39–44. doi: 10.1016/j.cogsc.2019.09.003.

Chowdhury, R. and Freire, F. (2015) ‘Bioenergy production from algae using dairy manure as a nutrient source: Life cycle energy and greenhouse gas emission analysis’, *Applied Energy*, 154(C), pp. 1112–1121. doi: 10.1016/j.apenergy.2015.05.045.

Cilia, V., Lafay, B. and Christen, R. (1996) ‘Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level’, *Molecular Biology and Evolution*, 13(3), pp. 451–461. doi: 10.1093/oxfordjournals.molbev.a025606.

Cirne, D. G. *et al.* (2007) ‘Hydrolysis and microbial community analyses in two-stage anaerobic digestion of energy crops’, *Journal of Applied Microbiology*, 103(3), pp. 516–527.

Ciuffreda, L., Rodríguez-Pérez, H. and Flores, C. (2021) ‘Nanopore sequencing and its application to the study of microbial communities’, *Computational and Structural Biotechnology Journal*, 19, pp. 1497–1511. doi: 10.1016/J.CSBJ.2021.02.020.

Cleenwerck, I., De Vos, P. and De Vuyst, L. (2010) ‘Phylogeny and differentiation of species of the genus Gluconacetobacter and related taxa based on multilocus sequence analyses of housekeeping genes and reclassification of Acetobacter xylinus subsp. sucrofermentans as Gluconacetobacter sucrofermentans (T’, *International Journal of Systematic and Evolutionary Microbiology*, 60(10), pp. 2277–2283. doi: 10.1099/ijs.0.018465-0.

Conklin, A., Stensel, H. D. and Ferguson, J. (2006) ‘Growth Kinetics and Competition Between Methanosarcina and Methanosaeta in Mesophilic

Anaerobic Digestion', *Water Environment Research*, 78(5), pp. 486–496. doi: 10.2175/106143006X95393/FORMAT/PDF.

Connon, S. A. and Giovannoni, S. J. (2002) 'High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates', *Appl. Environ. Microbiol.*, 68(8), pp. 3878–3885.

Cruz-Morales, P. *et al.* (2019) 'Revisiting the Evolution and Taxonomy of Clostridia, a Phylogenomic Update', *Genome Biology and Evolution*, 11(7), pp. 2035–2044. doi: 10.1093/gbe/evz096.

Cwyk, W. M. and Canale-Parola, E. (1979) 'Treponema succinifaciens sp. nov., an anaerobic spirochete from the swine intestine', *Archives of Microbiology*, 122(3), pp. 231–239. doi: 10.1007/BF00411285.

Cynkin, M. A. and Gibbs, M. (1958) 'Metabolism of pentoses by clostridia. II. The fermentation of C14-labeled pentoses by Clostridium per fringens, Clostridium beijerinckii, and Clostridium butylicum.', *Journal of Bacteriology*, 75(3), pp. 335–338. doi: 10.1128/JB.75.3.335-338.1958.

Czajka, J. *et al.* (2017) 'Synthetic biology for manufacturing chemicals: constraints drive the use of non-conventional microbial platforms', *Applied Microbiology and Biotechnology*, 101(20), pp. 7427–7434. doi: 10.1007/S00253-017-8489-9/FIGURES/2.

Davis, J. J. *et al.* (2020) 'The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities.', *Nucleic acids research*, 48(D1). doi: 10.1093/nar/gkz943.

Deangelis, K. M. *et al.* (2012) 'Anaerobic decomposition of switchgrass by tropical soil-derived feedstock-adapted consortia', *mBio*, 3(1), pp. 1–9. doi: 10.1128/mBio.00249-11.

Demirel, B. and Scherer, P. (2008) 'The roles of acetotrophic and hydrogenotrophic methanogens during anaerobic conversion of biomass to methane: a review', *Reviews in Environmental Science and Bio/Technology*,

7(2), pp. 173–190.

Dickson, E. M., Ryan, J. W. and Smulyan, M. H. (1976) ‘Systems considerations and transition scenarios for the hydrogen economy’, *International Journal of Hydrogen Energy*, 1(1), pp. 11–21. doi: 10.1016/0360-3199(76)90005-7.

Dong, H. *et al.* (2010) ‘Engineering Clostridium Strain to Accept Unmethylated DNA’, *PLOS ONE*, 5(2), p. e9038. doi: 10.1371/JOURNAL.PONE.0009038.

Dröge, S. *et al.* (2008) ‘Treponema isoptericolens sp. nov., a novel spirochaete from the hindgut of the termite Incisitermes tabogae’, *International Journal of Systematic and Evolutionary Microbiology*, 58(5), pp. 1079–1083. doi: 10.1099/ijs.0.64699-0.

Duncan, S. H. *et al.* (2002) ‘Acetate utilization and butyryl coenzyme A (CoA): Acetate-CoA transferase in butyrate-producing bacteria from the human large intestine’, *Applied and Environmental Microbiology*, 68(10), pp. 5186–5190. doi: 10.1128/AEM.68.10.5186-5190.2002.

Dwidar, M. *et al.* (2012) ‘The future of butyric acid in industry’, *The Scientific World Journal*, 2012.

Edgar, R. C. (2017) ‘Accuracy of microbial community diversity estimated by closed- and open-reference OTUs’, *PeerJ*, 2017(10), p. e3889. doi: 10.7717/PEERJ.3889/SUPP-2.

Egwari, L. O. and Rotimi, V. O. (1991) ‘Survival of bacteroides species in Amies transport medium stored at different temperatures.’, *The Central African journal of medicine*, 37(12), pp. 415–419.

El-Ashgar, N. M. *et al.* (2012) ‘Sol-Gel Thin Films Immobilized with Bromocresol Purple pH-Sensitive Indicator in Presence of Surfactants’, *ISRN Analytical Chemistry*, 2012, pp. 1–11. doi: 10.5402/2012/604389.

Ezaki, T. (2015) ‘Ruminococcus’, in *Bergey’s Manual of Systematics of Archaea and Bacteria*. John Wiley & Sons, Ltd, pp. 1–5. doi: 10.1002/9781118960608.gbm00678.

Faith, D. P. (1992) ‘Conservation evaluation and phylogenetic diversity’, *Biological conservation*, 61(1), pp. 1–10.

Fatma, Z., Schultz, J. C. and Zhao, H. (2020) ‘Recent advances in domesticating non-model microorganisms’, *Biotechnology Progress*, 36(5), p. e3008. doi: 10.1002/BTPR.3008.

Fayolle, F., Marchal, R. and Ballerini, D. (1990) ‘Effect of controlled substrate feeding on butyric acid production by Clostridium tyrobutyricum’, *Journal of Industrial Microbiology*, 6(3), pp. 179–183. doi: 10.1007/BF01577693.

Feijoo, G. *et al.* (1995) ‘Sodium inhibition in the anaerobic digestion process: Antagonism and adaptation phenomena’, *Enzyme and Microbial Technology*, 17(2), pp. 180–188. doi: 10.1016/0141-0229(94)00011-F.

Fernández, A. *et al.* (1999) ‘How stable is stable? Function versus community composition’, *Appl. Environ. Microbiol.*, 65(8), pp. 3697–3704.

Ferraris, L., Butel, M. J. and Aires, J. (2010) ‘Antimicrobial susceptibility and resistance determinants of Clostridium butyricum isolates from preterm infants’, *International Journal of Antimicrobial Agents*, 36(5), pp. 420–423. doi: 10.1016/j.ijantimicag.2010.07.005.

Fisher Costerton, J. W. (1956) ‘The Production of the Alpha Toxin of Clostridium perfringens in a Chemically Reproducible Medium’, *University of British Columbia, Archive of Master Theses (unpublished)*.

Fox, G. E., Wisotzkey, J. D. and Jurtshuk, P. (1992) ‘How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity’, *International Journal of Systematic Bacteriology*, 42(1), pp. 166–170. doi: 10.1099/00207713-42-1-166/CITE/REFWORKS.

Gabriel, C. L. (1928) ‘Butanol Fermentation Process’, *Industrial and Engineering Chemistry*. doi: 10.1021/ie50226a020.

Gawin, A., Valla, S. and Brautaset, T. (2017) ‘The XylS/Pm regulator/promoter system and its use in fundamental studies of bacterial gene expression, recombinant protein production and metabolic engineering’, *Microbial Biotechnology*. doi: 10.1111/1751-7915.12701.

Ghoddusi, H. B., Sherburn, R. E. and Aboaba, O. O. (2013) ‘Growth Limiting pH, Water Activity, and Temperature for Neurotoxigenic Strains of Clostridium butyricum’, *ISRN Microbiology*, 2013, pp. 1–6. doi: 10.1155/2013/731430.

Giordani, A. *et al.* (2021) ‘Performance and Microbial Community Analysis in an Anaerobic Hybrid Baffled Reactor Treating Dairy Wastewater’, *Water, Air, and Soil Pollution*, 232(10), pp. 1–16. doi: 10.1007/s11270-021-05348-0.

Giraldeli, L. D., Fonseca, B. C. and Reginatto, V. (2019) ‘Investigating how biomass hydrolysis derivatives inhibit H2 production by an isolated Clostridium beijerinckii’, *International Journal of Hydrogen Energy*, 44(29), pp. 14683–14693. doi: 10.1016/j.ijhydene.2019.04.136.

Glendinning, L. *et al.* (2021) ‘Metagenomic analysis of the cow, sheep, reindeer and red deer rumen’, *Scientific Reports*, 11(1). doi: 10.1038/s41598-021-81668-9.

Gomes, R. J. *et al.* (2018) ‘Acetic acid bacteria in the food industry: Systematics, characteristics and applications’, *Food Technology and Biotechnology*. University of Zagreb, pp. 139–151. doi: 10.17113/ftb.56.02.18.5593.

Graber, J. R., Leadbetter, J. R. and Breznak, J. A. (2004) ‘Description of Treponema azotonutricium sp. nov. and Treponema primitia sp. nov., the First Spirochetes Isolated from Termite Guts’, *Applied and Environmental Microbiology*, 70(3), pp. 1315–1320. doi: 10.1128/AEM.70.3.1315-

1320.2004.

Greenhalf, C. E. *et al.* (2012) ‘Thermochemical characterisation of straws and high yielding perennial grasses’, *Industrial Crops and Products*, 36(1), pp. 449–459. doi: 10.1016/j.indcrop.2011.10.025.

Griffin, M. E. *et al.* (1998) ‘Methanogenic population dynamics during start-up of anaerobic digesters treating municipal solid waste and biosolids’, *Biotechnology and bioengineering*, 57(3), pp. 342–355.

Grosse-Honebrink, A. *et al.* (2021) ‘Development of Clostridium saccharoperbutylacetonicum as a Whole Cell Biocatalyst for Production of Chirally Pure (R)-1,3-Butanediol’, *Frontiers in Bioengineering and Biotechnology*, 9, p. 355. doi: 10.3389/FBIOE.2021.659895/BIBTEX.

Gulhane, M. *et al.* (2017) ‘Study of microbial community plasticity for anaerobic digestion of vegetable waste in Anaerobic Baffled Reactor’, *Renewable Energy*, 101, pp. 59–66. doi: 10.1016/j.renene.2016.08.021.

Gunaseelan, V. N. (1997) ‘Anaerobic digestion of biomass for methane production: a review’, *Biomass and bioenergy*, 13(1–2), pp. 83–114.

Gupta, R. S. and Gao, B. (2009) ‘Phylogenomic analyses of clostridia and identification of novel protein signatures that are specific to the genus Clostridium sensu stricto (cluster I)’, *International Journal of Systematic and Evolutionary Microbiology*, 59(2), pp. 285–294. doi: 10.1099/ijs.0.001792-0.

Gwak, J.-H. *et al.* (2015) ‘Draconibacterium filum sp. nov., A New Member of the Prolixibacteraceae Isolated from Sediment of East Sea’, *Proceedings of the Korean Society of Microbiology Conference*, (4), pp. 174-174(1 pages). Available at: https://www.dbpia.co.kr/pdf/pdfView.do?nodeId=NODE07010685 (Accessed: 10 December 2021).

Hagihara, M. *et al.* (2018) ‘The impact of Clostridium butyricum MIYAIRI 588 on the murine gut microbiome and colonic tissue’, *Anaerobe*, 54, pp. 8–18.

doi: 10.1016/J.ANAEROBE.2018.07.012.

Hakalehto, E. and Hänninen, O. (2012) ‘Gaseous CO2 signal initiates growth of butyric acid-producing Clostridium butyricum in both pure culture and mixed cultures with Lactobacillus brevis’, *Canadian Journal of Microbiology*. NRC Research Press, pp. 928–931. doi: 10.1139/W2012-059.

Hastings, J. J. H. (1978) ‘Acetone—Butyl Alcohol Fermentation’, in *Economic Microbiology: Primary Products of Metabolism*, pp. 31–45. doi: 10.1016/b978-0-12-596552-1.50009-7.

He, G. Q. *et al.* (2005) ‘Batch and fed-batch production of butyric acid by Clostridium butyricum ZJUCB’, *Journal of Zhejiang University. Science. B*, 6(11), p. 1076. doi: 10.1631/JZUS.2005.B1076.

He, Q. *et al.* (2017) ‘Investigation of foaming causes in three mesophilic food waste digesters: Reactor performance and microbial analysis’, *Scientific Reports*, 7(1), pp. 1–10. doi: 10.1038/s41598-017-14258-3.

He, Q., Li, L. and Peng, X. (2017) ‘Early Warning Indicators and Microbial Mechanisms for Process Failure due to Organic Overloading in Food Waste Digesters’, *Journal of Environmental Engineering*, 143(12), p. 04017077. doi: 10.1061/(asce)ee.1943-7870.0001280.

Head, I. M., Saunders, J. R. and Pickup, R. W. (1998) ‘Microbial Evolution, Diversity, and Ecology: A Decade of Ribosomal RNA Analysis of Uncultivated Microorganisms’, *Microbial Ecology 1998 35:1*, 35(1), pp. 1–21. doi: 10.1007/S002489900056.

Heap, J. T. *et al.* (2007) ‘The ClosTron: a universal gene knock-out system for the genus Clostridium’, *Journal of microbiological methods*, 70(3), pp. 452–464.

Heap, J. T. *et al.* (2009) ‘A modular system for Clostridium shuttle plasmids’, *Journal of microbiological methods*, 78(1), pp. 79–85.

Heffernan, J. K. *et al.* (2020) ‘Enhancing CO2-Valorization Using Clostridium autoethanogenum for Sustainable Fuel and Chemicals Production’, *Frontiers in Bioengineering and Biotechnology*, 8, p. 204. doi: 10.3389/FBIOE.2020.00204.

Heyndrickx, M. *et al.* (1987) ‘Effect of various external factors on the fermentative production of hydrogen gas from glucose by Clostridium butyricum strains in batch culture’, *Systematic and Applied Microbiology*, 9(1–2), pp. 163–168. doi: 10.1016/S0723-2020(87)80072-3.

Heyndrickx, M., De Vos, P. and De Ley, J. (1991) ‘Fermentation of d-xylose by Clostridium butyricum LMG 1213t1 in chemostats’, *Enzyme and Microbial Technology*, 13(11), pp. 893–897. doi: 10.1016/0141-0229(91)90105-J.

Hierholtzer, A. and Akunna, J. C. (2012) ‘Modelling sodium inhibition on the anaerobic digestion process’. doi: 10.2166/wst.2012.345.

Holdeman, L. V., Cato, E. P. and Moore, W. E. C. (1971) ‘Eubacterium contortum (Prevot) comb. nov.: Emendation of Description and Designation of the Type Strain’, *International Journal of Systematic Bacteriology*, 21(4), pp. 304–306. doi: 10.1099/00207713-21-4-304.

Hollander, D. H. and Nell, E. E. (1954) ‘Improved preservation of Treponema pallidum and other bacteria by freezing with glycerol.’, *Applied microbiology*, 2(3), pp. 164–170. doi: 10.1128/aem.2.3.164-170.1954.

Hölzer, M. (2020) *Calculation of the Percentage of Conserved Proteins following Qin, Xie et al. (2014)*, *GitHub hoelzer/pocp*. Available at: https://github.com/hoelzer/pocp (Accessed: 24 December 2021).

Huang, K. X. *et al.* (2000) ‘Identification and characterization of a second butyrate kinase from Clostridium acetobutylicum ATCC 824’, in *Journal of Molecular Microbiology and Biotechnology*, pp. 33–38. Available at: www.caister.com/bacteria-plant (Accessed: 7 January 2022).

Huber, R. and Stetter, K. O. (2015) ‘Fervidobacterium’, in *Bergey’s Manual of*

Systematics of Archaea and Bacteria. John Wiley & Sons, Ltd, pp. 1–5. doi: 10.1002/9781118960608.gbm01269.

Hughes, K. A. and Nobbs, S. J. (2004) 'Long-term survival of human faecal microorganisms on the Antarctic Peninsula', *Antarctic Science*, 16(3), pp. 293–297. doi: 10.1017/S095410200400210X.

IEA (2012) 'Bio-based Chemicals - Value Added Products from Biorefineries', *IEA Bioenergy - Task42 Biorefinery*. Available at: https://www.ieabioenergy.com/wp-content/uploads/2013/10/Task-42-Biobased-Chemicals-value-added-products-from-biorefineries.pdf (Accessed: 2 January 2022).

IEA (2020) 'Outlook for biogas and biomethane. Prospects for organic growth. World Energy Outlook Special Report.', p. 93. Available at: https://www.iea.org/reports/outlook-for-biogas-and-biomethane-prospects-for-organic-growth (Accessed: 30 December 2021).

IEA (2021a) 'Renewables 2021', *International Energy Agency (IEA) Publications International.*, p. 167. Available at: www.iea.org/t&c/ (Accessed: 30 December 2021).

IEA (2021b) *Statistics report: Key World Energy Statistics 2021*, *IEA*.

Iino, T. *et al.* (2014) 'Description of Mariniphaga anaerophila gen. nov., Sp. nov., A facultatively aerobic marine bacterium isolated from tidal flat sediment, reclassification of the Draconibacteriaceae as a later heterotypic synonym of the Prolixibacteraceae and description of', *International Journal of Systematic and Evolutionary Microbiology*, 64(Pt_11), pp. 3660–3667. doi: 10.1099/ijs.0.066274-0.

Ikeda, T. *et al.* (1988) 'Phenotypic Characteristics in Distinguishing Clostridium butyricum from Clostridium beijerinckii', *Bifidobacteria and Microflora*, 7(1), pp. 56–60. doi: 10.12938/bifidus1982.7.1_56.

Intanoo, P., Chaimongkol, P. and Chavadej, S. (2016) 'Hydrogen and

methane production from cassava wastewater using two-stage upflow anaerobic sludge blanket reactors (UASB) with an emphasis on maximum hydrogen production', *International Journal of Hydrogen Energy*, 41(14), pp. 6107–6114. doi: 10.1016/J.IJHYDENE.2015.10.125.

Jang, Y. S. *et al.* (2014) 'Metabolic engineering of Clostridium acetobutylicum for butyric acid production with high butyric acid selectivity', *Metabolic Engineering*, 23, pp. 165–174. doi: 10.1016/J.YMBEN.2014.03.004.

Jawed, K. *et al.* (2020) 'Improved Butanol Production Using FASII Pathway in E. coli', *ACS Synthetic Biology*, 9(9), pp. 2390–2398. doi: 10.1021/ACSSYNBIO.0C00154/ASSET/IMAGES/LARGE/SB0C00154_0008.JPEG.

Jenkins, R. W. *et al.* (2013) 'Potential renewable oxygenated biofuels for the aviation and road transport sectors', in *Fuel*. doi: 10.1016/j.fuel.2012.08.019.

Jeswani, H. K., Chilvers, A. and Azapagic, A. (2020) 'Environmental sustainability of biofuels: A review: Environmental sustainability of biofuels', *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. doi: 10.1098/rspa.2020.0351.

Jiang, H. *et al.* (2018) 'The startup performance and microbial distribution of an anaerobic baffled reactor (ABR)treating medium-strength synthetic industrial wastewater', *Journal of Environmental Science and Health - Part A Toxic/Hazardous Substances and Environmental Engineering*. doi: 10.1080/10934529.2017.1368297.

Jiang, L. *et al.* (2010) 'Production of butyric acid from glucose and xylose with immobilized cells of clostridium tyrobutyricum in a fibrous-bed bioreactor', *Applied Biochemistry and Biotechnology*, 160(2), pp. 350–359. doi: 10.1007/s12010-008-8305-1.

Jo, J. H. *et al.* (2010) 'Molecular characterization and homologous overexpression of [FeFe]-hydrogenase in Clostridium tyrobutyricum JM1',

International Journal of Hydrogen Energy, 35(3), pp. 1065–1073. doi: 10.1016/J.IJHYDENE.2009.11.102.

Johnson, J. S. *et al.* (2019) 'Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis', *Nature Communications 2019 10:1*, 10(1), pp. 1–11. doi: 10.1038/s41467-019-13036-1.

Jones, D. T. and Woods, D. R. (1986) 'Acetone-butanol fermentation revisited', *Microbiological Reviews*. doi: 10.1128/mmbr.50.4.484-524.1986.

Joseph, R. C., Kim, N. M. and Sandoval, N. R. (2018) 'Recent developments of the synthetic biology toolkit for Clostridium', *Frontiers in Microbiology*, 9, p. 154.

Joyce, A. *et al.* (2018) 'Linking microbial community structure and function during the acidified anaerobic digestion of grass', *Frontiers in Microbiology*, 9(MAR), p. 540. doi: 10.3389/fmicb.2018.00540.

Jürgensen, L. *et al.* (2015) 'Hydrogen production using an anaerobic baffled reactor – Mass balances for pathway analysis and gas composition profiles', *International Journal of Hydrogen Energy*, 40(36), pp. 12154–12161. doi: 10.1016/J.IJHYDENE.2015.07.068.

Kaji, M. *et al.* (1999) 'The hydA gene encoding the H(2)-evolving hydrogenase of Clostridium perfringens: molecular characterization and expression of the gene', *FEMS microbiology letters*, 181(2), pp. 329–336. doi: 10.1111/J.1574-6968.1999.TB08863.X.

Kampmann, K. *et al.* (2012) 'Unexpected stability of Bacteroidetes and Firmicutes communities in laboratory biogas reactors fed with different defined substrates', *Appl. Environ. Microbiol.*, 78(7), pp. 2106–2119.

Kanehisa, M. and Goto, S. (2000) 'KEGG: Kyoto Encyclopedia of Genes and Genomes', *Nucleic Acids Research*, pp. 27–30. doi: 10.1093/nar/28.1.27.

Kaneko, M. *et al.* (2014) 'Quantitative analysis of coenzyme F430 in

environmental samples: A new diagnostic tool for methanogenesis and anaerobic methane oxidation', *Analytical Chemistry*, 86(7), pp. 3633–3638. doi: 10.1021/ac500305j.

Kasukabe, T., Rephaeli, A. and Honma, Y. (1997) 'An anti-cancer derivative of butyric acid (pivalyloxymethyl butyrate) and daunorubicin cooperatively prolong survival of mice inoculated with monocytic leukaemia cells', *British Journal of Cancer*. doi: 10.1038/bjc.1997.151.

Kataoka, N., Miya, A. and Kiriyama, K. (1997) 'Studies on hydrogen production by continuous culture system of hydrogen-producing anaerobic bacteria', *Water Science and Technology*, 36(6–7), pp. 41–47. doi: 10.1016/S0273-1223(97)00505-2.

Khan, M. A. *et al.* (2016) 'Optimization of process parameters for production of volatile fatty acid, biohydrogen and methane from anaerobic digestion', *Bioresource Technology*, 219, pp. 738–748. doi: 10.1016/J.BIORTECH.2016.08.073.

Kim, G. T. *et al.* (2005) 'Dissimilatory Fe(III) reduction by an electrochemically active lactic acid bacterium phylogenetically related to Enterococcus gallinarum isolated from submerged soil', *Journal of Applied Microbiology*, 99(4), pp. 978–987. doi: 10.1111/J.1365-2672.2004.02514.X.

Kim, J. K. *et al.* (2008) 'Hydrogen production conditions from food waste by dark fermentation with Clostridium beijerinckii KCTC 1785', *Biotechnology and Bioprocess Engineering*, 13(4), pp. 499–504. doi: 10.1007/s12257-008-0142-0.

Kleerebezem, R. *et al.* (2015) 'Anaerobic digestion without biogas?', *Reviews in Environmental Science and Biotechnology*, 14(4), pp. 787–801. doi: 10.1007/S11157-015-9374-6/FIGURES/9.

Klein, M. *et al.* (2010) 'Influence of hydrogenase overexpression on hydrogen production of Clostridium acetobutylicum DSM 792', *Enzyme and microbial*

technology, 46(5), pp. 384–390.

Klocke, M. *et al.* (2007) 'Microbial community analysis of a biogas-producing completely stirred tank reactor fed continuously with fodder beet silage as mono-substrate', *Systematic and applied microbiology*, 30(2), pp. 139–151.

Koch, C. *et al.* (2014) 'Microbiomes in bioenergy production: from analysis to management', *Current opinion in biotechnology*, 27, pp. 65–72.

Kudo, H., Cheng, K.-J. and Costerton, J. W. (1987) 'Interactions between Treponema bryantii and cellulolytic bacteria in the in vitro degradation of straw cellulose', *Canadian Journal of Microbiology*, 33(3), pp. 244–248.

Kumar, M. *et al.* (2012) 'Comparative economic assessment of ABE fermentation based on cellulosic and non-cellulosic feedstocks', *Applied Energy*. doi: 10.1016/j.apenergy.2011.12.079.

Kumar, S. *et al.* (2018) 'MEGA X: molecular evolutionary genetics analysis across computing platforms', *Molecular biology and evolution*, 35(6), pp. 1547–1549.

Kurade, M. B. *et al.* (2019) 'Acetoclastic methanogenesis led by Methanosarcina in anaerobic co-digestion of fats, oil and grease for enhanced production of methane', *Bioresource Technology*, 272, pp. 351–359. doi: 10.1016/j.biortech.2018.10.047.

Lagier, J. C. *et al.* (2015) 'Non contiguous-finished genome sequence and description of Clostridium jeddahense sp. nov', *Standards in Genomic Sciences*, 9(3), pp. 1003–1019. doi: 10.4056/sigs.5571026.

Lawson, P. A. *et al.* (2016) 'Reclassification of Clostridium difficile as Clostridioides difficile (Hall and O'Toole 1935) Prévot 1938', *Anaerobe*, 40, pp. 95–99. doi: 10.1016/j.anaerobe.2016.06.008.

Lawson, P. A. and Rainey, F. A. (2016) 'Proposal to restrict the genus Clostridium prazmowski to Clostridium butyricum and related species',

International Journal of Systematic and Evolutionary Microbiology, 66(2), pp. 1009–1016. doi: 10.1099/ijsem.0.000824.

Lee, D. J., Show, K. Y. and Su, A. (2011) 'Dark fermentation on biohydrogen production: Pure culture', *Bioresource Technology*, 102(18), pp. 8393–8402. doi: 10.1016/j.biortech.2011.03.041.

Lee, J. *et al.* (2016) 'Deciphering Clostridium tyrobutyricum Metabolism Based on the Whole-Genome Sequence and Proteome Analyses', *mBio*, 7(3). doi: 10.1128/MBIO.00743-16.

Lee, K. S. *et al.* (2012) 'Enhancing the performance of dark fermentative hydrogen production using a reduced pressure fermentation strategy', *International Journal of Hydrogen Energy*, 37(20), pp. 15556–15562. doi: 10.1016/J.IJHYDENE.2012.04.039.

Leggett, R. M. and Clark, M. D. (2017) 'A world of opportunities with nanopore sequencing', *Journal of Experimental Botany*, 68(20), pp. 5419–5429. doi: 10.1093/JXB/ERX289.

Lerm, S. *et al.* (2012) 'Archaeal community composition affects the function of anaerobic co-digesters in response to organic overload', *Waste Management*, 32(3), pp. 389–399. doi: 10.1016/j.wasman.2011.11.013.

Levin, D. B., Pitt, L. and Love, M. (2004) 'Biohydrogen production: Prospects and limitations to practical application', *International Journal of Hydrogen Energy*, 29(2), pp. 173–185. doi: 10.1016/S0360-3199(03)00094-6.

Lin, P. Y. *et al.* (2007) 'Biological hydrogen production of the genus Clostridium: Metabolic study and mathematical model simulation', *International Journal of Hydrogen Energy*, 32(12), pp. 1728–1735. doi: 10.1016/J.IJHYDENE.2006.12.009.

Lin, Y. *et al.* (2012) 'Performance and microbial community in hybrid anaerobic baffled reactor-constructed wetland for nitrobenzene wastewater', *Bioresource Technology*, 118, pp. 128–135. doi:

10.1016/J.BIORTECH.2012.05.056.

Liu, Y. *et al.* (2014) ‘Hydrogenispora ethanolica gen. nov., sp nov., an anaerobic carbohydrate-fermenting bacterium from anaerobic sludge’.

Louis, P. *et al.* (2004) ‘Restricted Distribution of the Butyrate Kinase Pathway among Butyrate-Producing Bacteria from the Human Colon’, *Journal of Bacteriology*, 186(7), pp. 2099–2106. doi: 10.1128/JB.186.7.2099-2106.2004.

Lozupone, C. and Knight, R. (2005) ‘UniFrac: a new phylogenetic method for comparing microbial communities’, *Appl. Environ. Microbiol.*, 71(12), pp. 8228–8235.

Łukajtis, R. *et al.* (2018) ‘Hydrogen production from biomass using dark fermentation’, *Renewable and Sustainable Energy Reviews*, 91, pp. 665–694. doi: 10.1016/J.RSER.2018.04.043.

Luo, G. *et al.* (2017) ‘Butyric acid fermentation in xylose and glucose by Clostridium tyrobutyricum’, *BioResources*, 12(2), pp. 2930–2940. doi: 10.15376/biores.12.2.2930-2940.

Macy, J. M., Ljungdahl, L. G. and Gottschalk, G. (1978) ‘Pathway of Succinate and Propionate Formation in Bacteroides fragilis’, *Journal of Bacteriology*, 134(1), pp. 84–91. doi: 10.1128/JB.134.1.84-91.1978.

Madden, T. (2013) ‘The BLAST sequence analysis tool’, *The BLAST Sequence Analysis Tool*.

Maddox, I. S. *et al.* (2000) ‘The cause of “acid crash” and “acidogenic fermentations” during the batch acetone-butanol-ethanol (ABE-) fermentation process’, *Journal of Molecular Microbiology and Biotechnology*, 2(1), pp. 95–100. Available at: www.caister.com/bacteria-plant (Accessed: 21 December 2021).

Madeira, F. *et al.* (2019) ‘The EMBL-EBI search and sequence analysis tools

APIs in 2019.', *Nucleic Acids Research*, 47(W1), pp. W636–W641. doi: 10.1093/NAR/GKZ268.

Martínez-Garćía, E. *et al.* (2015) 'SEVA 2.0: An update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities', *Nucleic Acids Research*. doi: 10.1093/nar/gku1114.

Maus, I. *et al.* (2020) 'Impact of process temperature and organic loading rate on cellulolytic/hydrolytic biofilm microbiomes during biomethanation of ryegrass silage revealed by genome-centered metagenomics and metatranscriptomics', *Environmental Microbiomes*, 15(1), pp. 1–21. doi: 10.1186/S40793-020-00354-X/TABLES/2.

McDowall, W. (2016) 'Are scenarios of hydrogen vehicle adoption optimistic? A comparison with historical analogies', *Environmental Innovation and Societal Transitions*, 20, pp. 48–61. doi: 10.1016/j.eist.2015.10.004.

Meier-Kolthoff, J. P. *et al.* (2021) 'TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes', *Nucleic Acids Research*, (1). doi: 10.1093/nar/gkab902.

Metcalf, D., Sharif, S. and Weese, J. S. (2010) 'Evaluation of candidate reference genes in Clostridium difficile for gene expression normalization', *Anaerobe*, 16(4), pp. 439–443. doi: 10.1016/j.anaerobe.2010.06.007.

Meyers, A., Furtmann, C. and Jose, J. (2018) 'Direct optical density determination of bacterial cultures in microplates for high-throughput screening applications', *Enzyme and Microbial Technology*, 118, pp. 1–5. doi: 10.1016/j.enzmictec.2018.06.016.

Michel-Savin, D., Marchal, R. and Vandecasteele, J. P. (1990) 'Control of the selectivity of butyric acid production and improvement of fermentation performance with Clostridium tyrobutyricum', *Applied Microbiology and Biotechnology*, 32(4), pp. 387–392. doi: 10.1007/BF00903770.

Milledge, J. J. and Harvey, P. J. (2018) 'Anaerobic Digestion and Gasification

of Seaweed', in *Grand Challenges in Biology and Biotechnology*. Springer, Cham, pp. 237–258. doi: 10.1007/978-3-319-69075-9_7.

Miller, T. L. (1978) 'The pathway of formation of acetate and succinate from pyruvate by Bacteroides succinogenes', *Archives of Microbiology 1978 117:2*, 117(2), pp. 145–152. doi: 10.1007/BF00402302.

Minton, N. P. *et al.* (2016) 'A roadmap for gene system development in Clostridium', *Anaerobe*, 41, pp. 104–112.

Mira, A., Ochman, H. and Moran, N. A. (2001) 'Deletional bias and the evolution of bacterial genomes', *Trends in Genetics*, 17(10), pp. 589–596. doi: 10.1016/S0168-9525(01)02447-7.

Montingelli, M. E., Tedesco, S. and Olabi, A. G. (2015) 'Biogas production from algal biomass: A review', *Renewable and Sustainable Energy Reviews*, 43, pp. 961–972.

Mordaka, P. and Heap, J. T. (2018) 'Stringency of synthetic promoter sequences in Clostridium revealed and circumvented by tuning promoter library mutation rates', *ACS synthetic biology*, 7(2), pp. 672–681.

Moriarty, P. and Honnery, D. (2019) 'Prospects for hydrogen as a transport fuel', *International Journal of Hydrogen Energy*, 44(31), pp. 16029–16037. doi: 10.1016/J.IJHYDENE.2019.04.278.

Morimoto, K. *et al.* (2005) 'Overexpression of a hydrogenase gene in Clostridium paraputrificum to enhance hydrogen gas production', *FEMS Microbiology Letters*, 246(2), pp. 229–234. doi: 10.1016/J.FEMSLE.2005.04.014.

Moset, V. *et al.* (2015) 'Mesophilic versus thermophilic anaerobic digestion of cattle manure: methane productivity and microbial ecology', *Microbial Biotechnology*, 8(5), pp. 787–800. doi: 10.1111/1751-7915.12271.

Mumm, R. H. *et al.* (2014) 'Land usage attributed to corn ethanol production

in the United States: sensitivity to technological advances in corn grain yield, ethanol conversion, and co-product utilization', *Biotechnology for biofuels*, 7(1), p. 61.

Mumme, J. *et al.* (2014) 'Use of biochars in anaerobic digestion', *Bioresource Technology*, 164, pp. 189–197. doi: 10.1016/J.BIORTECH.2014.05.008.

Murray, W. D. (1986) 'Symbiotic relationship of Bacteroides cellulosolvens and Clostridium saccharolyticum in cellulose fermentation', *Applied and Environmental Microbiology*, 51(4), pp. 710–714. doi: 10.1128/aem.51.4.710-714.1986.

Murray, W. D. and Khan, A. W. (1983) 'Growth requirements of Clostridium saccharolyticum, an ethanologenic anaerobe', *Canadian Journal of Microbiology*, 29(3), pp. 348–353. doi: 10.1139/m83-058.

Naik, S. N. *et al.* (2010) 'Production of first and second generation biofuels: A comprehensive review', *Renewable and Sustainable Energy Reviews*, 14(2), pp. 578–597. doi: 10.1016/J.RSER.2009.10.003.

Nche, P. F., Nout, M. J. R. and Rombouts, F. M. (1994) 'Gas producton by Clostridium perfringens as a measure of the fermentability of carbohydrates and processed cereal-legume foods', *Food Microbiology*, 11(1), pp. 21–29. doi: 10.1006/fmic.1994.1004.

Nesbø, C. L. *et al.* (2019) 'Genomic analysis of the mesophilic Thermotogae genus Mesotoga reveals phylogeographic structure and genomic determinants of its distinct metabolism', *Environmental Microbiology*, 21(1), pp. 456–470. doi: 10.1111/1462-2920.14477.

Nielsen, B. V *et al.* (2020) 'The effects of halogenated compounds on the anaerobic digestion of macroalgae', *Fermentation*. doi: 10.3390/FERMENTATION6030085.

Nygård, C. A. and Dring, M. J. (2008) 'Influence of salinity, temperature, dissolved inorganic carbon and nutrient concentration on the photosynthesis

and growth of Fucus vesiculosus from the Baltic and Irish Seas', *European Journal of Phycology*, 43(3), pp. 253–262. doi: 10.1080/09670260802172627.

Ofori-Boateng, C. and Kwofie, E. M. (2009) 'Water Scrubbing: A Better Option for Biogas Purification for Effective Storage', *World Applied Sciences Journal*, 5, pp. 122–125.

Oh, S. E. *et al.* (2009) 'Hydrogen production by Clostridium acetobutylicum ATCC 824 and megaplasmid-deficient mutant M5 evaluated using a large headspace volume technique', *International Journal of Hydrogen Energy*, 34(23), pp. 9347–9353. doi: 10.1016/j.ijhydene.2009.09.084.

Olson, E. S. *et al.* (2003) 'Ester fuels and chemicals from biomass', in *Applied Biochemistry and Biotechnology - Part A Enzyme Engineering and Biotechnology*. doi: 10.1385/ABAB:108:1-3:843.

Pagaling, E. *et al.* (2017) 'Assembly of microbial communities in replicate nutrient-cycling model ecosystems follows divergent trajectories, leading to alternate stable states', *Environmental microbiology*, 19(8), pp. 3374–3386. doi: 10.1111/1462-2920.13849.

Pan, C. M. *et al.* (2008) 'Fermentative hydrogen production by the newly isolated Clostridium beijerinckii Fanp3', *International Journal of Hydrogen Energy*, 33(20), pp. 5383–5391. doi: 10.1016/J.IJHYDENE.2008.05.037.

Parada, A. E., Needham, D. M. and Fuhrman, J. A. (2016) 'Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples', *Environmental microbiology*, 18(5), pp. 1403–1414.

Paster, B. J. and Canale-Parola, E. (1985) 'Treponema saccharophilum sp. nov., a large pectinolytic spirochete from the bovine rumen', *Applied and Environmental Microbiology*, 50(2), pp. 212–219. doi: 10.1128/aem.50.2.212-219.1985.

Patel, G. B. (2015) ‘Methanosaeta’, in *Bergey’s Manual of Systematics of Archaea and Bacteria*. John Wiley & Sons, Ltd, pp. 1–8. doi: 10.1002/9781118960608.gbm00513.

Patel, G. B. and Agnew, B. J. (1988) ‘Growth and butyric acid production by Clostridium populeti’, *Archives of Microbiology*, 150(3), pp. 267–271. doi: 10.1007/BF00407790.

Pei, A. Y. *et al.* (2010) ‘Diversity of 16S rRNA genes within individual prokaryotic genomes’, *Applied and Environmental Microbiology*, 76(12), pp. 3886–3897. doi: 10.1128/AEM.02953-09.

Peng, J. F. *et al.* (2013) ‘Spatial succession and metabolic properties of functional microbial communities in an anaerobic baffled reactor’, *International Biodeterioration & Biodegradation*, 80, pp. 60–65. doi: 10.1016/J.IBIOD.2013.02.006.

Pessôa, M. G. *et al.* (2019) ‘Newly isolated microorganisms with potential application in biotechnology’, *Biotechnology Advances*, 37(2), pp. 319–339. doi: 10.1016/J.BIOTECHADV.2019.01.007.

Pohlschroeder, M., Leschine, S. B. and Canale-Parola, E. (1994) ‘Spirochaeta caldaria sp. nov., a thermophilic bacterium that enhances cellulose degradation by Clostridium thermocellum’, *Archives of Microbiology 1994 161:1*, 161(1), pp. 17–24. doi: 10.1007/BF00248889.

Pollo, S. M. J., Zhaxybayeva, O. and Nesbø, C. L. (2015) ‘Insights into thermoadaptation and the evolution of mesophily from the bacterial phylum thermotogae’, *Canadian Journal of Microbiology*, 61(9), pp. 655–670. doi: 10.1139/cjm-2015-0073.

Popoff, M. R. and Truffaut, N. (1985) ‘Survey of plasmids in Clostridium butyricum and Clostridium beijerinckii strains from different origins and different phenotypes’, *Current Microbiology*, 12(3), pp. 151–156. doi: 10.1007/BF01567668.

Public Health England (2016) ‘UK Standards for Microbiology Investigations - Identification of Clostridium species’, *Standards Unit, Microbiology Services*, (3.2), pp. 1–30. Available at: https://www.gov.uk/uk (Accessed: 4 January 2022).

Pyne, M. E. *et al.* (2015) ‘Antisense-RNA-Mediated Gene Downregulation in Clostridium pasteurianum’, *Fermentation 2015, Vol. 1, Pages 113-126*, 1(1), pp. 113–126. doi: 10.3390/FERMENTATION1010113.

Qiao, J. T. *et al.* (2013) ‘Molecular characterization of bacterial and archaeal communities in a full-scale anaerobic reactor treating corn straw’, *Bioresource Technology*, 143, pp. 512–518. doi: 10.1016/j.biortech.2013.06.014.

Qin, Q. L. *et al.* (2014) ‘A proposed genus boundary for the prokaryotes based on genomic insights’, *Journal of Bacteriology*, 196(12), pp. 2210–2215. doi: 10.1128/JB.01688-14.

Qiu, J., Xu, J. and Ren, N. (2009) ‘Effect of initial substrate concentrations and pH on hydrogen production from xylose with Clostridium butyricum T4’, *Chinese Journal of Biotechnology*, 25(6), pp. 887–891. Available at: https://europepmc.org/article/med/19777817 (Accessed: 27 December 2021).

Qiu, Y.-L. *et al.* (2014) ‘Paludibacter jiangxiensis sp. nov., a strictly anaerobic, propionate-producing bacterium isolated from rice paddy field’, *Archives of microbiology*, 196(3), pp. 149–155.

Quast, C. *et al.* (2012) ‘The SILVA ribosomal RNA gene database project: improved data processing and web-based tools’, *Nucleic acids research*, 41(D1), pp. D590--D596.

Quigley, C. T. C. *et al.* (2020) ‘Bacterial Communities Show Algal Host (Fucus spp.)/Zone Differentiation Across the Stress Gradient of the Intertidal Zone’, *Frontiers in Microbiology*, 11, p. 2256. doi: 10.3389/FMICB.2020.563118/BIBTEX.

Rafineenia, R. (2013) ‘A Metabolic Model for Investigation of the Fermentative Hydrogen Production by Clostridium Butyricum W5 Grown on Xylose’, *Asian Journal of Experimental Biological Science*, 4(3), pp. 472–475.

Rambaut, A. (2012) *rambaut/figtree*, *GitHub*. Available at: https://github.com/rambaut/figtree (Accessed: 31 December 2021).

Rephaeli, A., Zhuk, R. and Nudelman, A. (2000) ‘Prodrugs of butyric acid from bench to bedside: Synthetic design, mechanisms of action, and clinical applications’, *Drug Development Research*. doi: 10.1002/1098-2299(200007/08)50:3/4<379::AID-DDR20>3.0.CO;2-8.

Richards, E. A., Steggerda, F. R. and Murata, A. (1968) ‘Relationship of bean substrates and certain intestinal bacteria to gas production in the dog.’, *Gastroenterology*, 55(4), pp. 502–509. doi: 10.1016/s0016-5085(19)34026-0.

Richter, M. and Rosselló-Móra, R. (2009) ‘Shifting the genomic gold standard for the prokaryotic species definition’, *Proceedings of the National Academy of Sciences of the United States of America*, 106(45), pp. 19126–19131. doi: 10.1073/pnas.0906412106.

Romero Aguilar, M. A. *et al.* (2013) ‘Effect of HRT on hydrogen production and organic matter solubilization in acidogenic anaerobic digestion of OFMSW’, *Chemical Engineering Journal*, 219, pp. 443–449. doi: 10.1016/J.CEJ.2012.12.090.

Roque, B. M. *et al.* (2021) ‘Red seaweed (Asparagopsis taxiformis) supplementation reduces enteric methane by over 80 percent in beef steers’, *PLoS ONE*, 16(3 March), p. e0247820. doi: 10.1371/journal.pone.0247820.

Ross, A. B. *et al.* (2008) ‘Classification of macroalgae as fuel and its thermochemical behaviour’, *Bioresource Technology*, 99(14), pp. 6494–6504. doi: 10.1016/j.biortech.2007.11.036.

Rossi-Tamisier, M. *et al.* (2015) ‘Cautionary tale of using 16s rRNA gene sequence similarity values in identification of human-associated bacterial

species', *International Journal of Systematic and Evolutionary Microbiology*, 65(6), pp. 1929–1934. doi: 10.1099/ijs.0.000161.

Rupérez, P., Ahrazem, O. and Leal, J. A. (2002) 'Potential antioxidant capacity of sulfated polysaccharides from the edible marine brown seaweed Fucus vesiculosus', *Journal of Agricultural and Food Chemistry*, 50(4), pp. 840–845. doi: 10.1021/jf010908o.

Russelle, M. P. *et al.* (2007) 'Comment on "Carbon-negative biofuels from low-input high-diversity grassland biomass".', *Science (New York, N.Y.)*. Science. doi: 10.1126/science.1139388.

Sacks, L. E. and Olson, A. C. (1979) 'Growth of Clostridium perfringens strains on α-galactosides', *Journal of Food Science*, 44(6), pp. 1756–1760. doi: 10.1111/j.1365-2621.1979.tb09133.x.

Schellenberg, J. *et al.* (2011) 'Pyrosequencing of chaperonin-60 (cpn60) amplicons as a means of determining microbial community composition.', *Methods in Molecular Biology (Clifton, N.J.)*, 733, pp. 143–158. doi: 10.1007/978-1-61779-089-8_10.

Schievano, A. *et al.* (2014) 'Can two-stage instead of one-stage anaerobic digestion really increase energy recovery from biomass?', *Applied energy*, 124, pp. 335–342.

Sebald, M. (1994) 'Genetic basis for antibiotic resistance in anaerobes', *Clinical Infectious Diseases*, 18, pp. S297–S304. doi: 10.1093/clinids/18.Supplement_4.S297.

Seemann, T. (2014) 'Prokka: rapid prokaryotic genome annotation', *Bioinformatics (Oxford, England)*, 30(14), pp. 2068–2069. doi: 10.1093/BIOINFORMATICS/BTU153.

Setlow, P. (2014) 'Spore Resistance Properties', *Microbiology Spectrum*, 2(5). doi: 10.1128/microbiolspec.tbs-0003-2012.

Siegert, M. *et al.* (2014) ‘The presence of hydrogenotrophic methanogens in the inoculum improves methane gas production in microbial electrolysis cells’, *Frontiers in Microbiology*, 5(DEC), p. 778. doi: 10.3389/FMICB.2014.00778/BIBTEX.

Smith, M. R. (1983) ‘Reversal of 2-bromoethanesulfonate inhibition of methanogenesis in Methanosarcina sp.’, *Journal of Bacteriology*, 156(2), pp. 516–523. doi: 10.1128/jb.156.2.516-523.1983.

Sogin, M. L. *et al.* (2006) ‘Microbial diversity in the deep sea and the underexplored “rare biosphere”’, *Proceedings of the National Academy of Sciences of the United States of America*, 103(32), pp. 12115–12120. doi: 10.1073/pnas.0605127103.

Solovyev, V. and Salamov, A. (2011) ‘Automatic annotation of microbial genomes and metagenomic sequences’, in *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*.

Stackebrandt, E. and Ebers, J. (2006) ‘Taxonomic parameters revisited: tarnished gold standards’, *Microbiology Today*, 33, pp. 152–155. Available at: https://socgenmicrobiol.org.uk/pubs/micro_today/pdf/110602.pdf (Accessed: 24 December 2021).

Stanton, T. B. and Canale-Parola, E. (1980) ‘Treponema bryantii sp. nov., a rumen spirochete that interacts with cellulolytic bacteria’, *Archives of Microbiology*, 127(2), pp. 145–156. doi: 10.1007/BF00428018.

Steele, D. B. and Stowers, M. D. (1991) ‘Techniques for selection of industrially important microorganisms’, *Annual Review of Microbiology*, pp. 89–106. doi: 10.1146/annurev.mi.45.100191.000513.

Stim-Herndon, K. P., Petersen, D. J. and Bennett, G. N. (1995) ‘Characterization of an acetyl-CoA C-acetyltransferase (thiolase) gene from Clostridium acetobutylicum ATCC 824’, *Gene*. doi: 10.1016/0378-1119(94)00838-J.

Stothard, P. (2000) ‘The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences.’, *BioTechniques*, 28(6). doi: 10.2144/00286ir01.

Su, C. *et al.* (2019) ‘Effects of hydraulic retention time on the performance and microbial community of an anaerobic baffled reactor-bioelectricity Fenton coupling reactor for treatment of traditional Chinese medicine wastewater’, *Bioresource Technology*, 288, p. 121508. doi: 10.1016/j.biortech.2019.121508.

Sullivan, L., Cates, M. S. and Bennett, G. N. (2010) ‘Structural correlations of activity of Clostridium acetobutylicum ATCC 824 butyrate kinase isozymes’, *Enzyme and Microbial Technology*, 46(2), pp. 118–124. doi: 10.1016/J.ENZMICTEC.2009.10.001.

Suryawanshi, P. C., Chaudhari, A. B. and Kothari, R. M. (2010) ‘Thermophilic anaerobic digestion: The best option for waste treatment’, *Critical Reviews in Biotechnology*, pp. 31–40. doi: 10.3109/07388550903330505.

Tadasse, G. *et al.* (2016) ‘Drivers and triggers of international food price spikes and volatility’, in *Food price volatility and its implications for food security and policy*. Springer, Cham, pp. 59–82.

Tao, Y. *et al.* (2007) ‘High hydrogen yield from a two-step process of dark- and photo-fermentation of sucrose’, *International Journal of Hydrogen Energy*, 32(2), pp. 200–206. doi: 10.1016/J.IJHYDENE.2006.06.034.

Taylor, M. M. (1972) ‘Eubacterium fissicatena sp.nov. isolated from the alimentary tract of the goat.’, *Journal of General Microbiology*, 71(3), pp. 457–463. doi: 10.1099/00221287-71-3-457.

Thauer, R. K. (1998) ‘Biochemistry of methanogenesis: a tribute to Marjory Stephenson:1998 Marjory Stephenson Prize Lecture’, *Microbiology*, 144(9), pp. 2377–2406. doi: 10.1099/00221287-144-9-2377.

Thauer, R. K. (2019) ‘Methyl (Alkyl)-Coenzyme M Reductases: Nickel F-430-

ContainingEnzymes Involved in Anaerobic Methane Formation and in Anaerobic Oxidationof Methane or of Short Chain Alkanes', *Biochemistry*, 58(52), p. 5198. doi: 10.1021/ACS.BIOCHEM.9B00164.

Thauer, R. K., Kirchniawy, F. H. and Jungermann, K. A. (1972) 'Properties and Function of the Pyruvate-Formate-Lyase Reaction in Clostridiae', *European Journal of Biochemistry*, 27(2), pp. 282–290. doi: 10.1111/J.1432-1033.1972.TB01837.X.

Tilche, A. and Vieira, S. M. M. (1991) 'Discussion report on reactor design of anaerobic filters and sludge bed reactors', in *Water Science and Technology*. doi: 10.2166/wst.1991.0225.

Tilman, D., Hill, J. and Lehman, C. (2006) 'Carbon-negative biofuels from low-input high-diversity grassland biomass', *Science*, 314(5805), pp. 1598–1600.

Tomei, M. C. *et al.* (2009) 'Modeling of Anaerobic digestion of sludge', *Critical Reviews in Environmental Science and Technology*. Taylor & Francis Group, pp. 1003–1051. doi: 10.1080/10643380801977818.

Touw, J. J. A., van Steenbergen, T. J. M. and de Graaff, J. (1982) 'Butyrate: a cytotoxin for Vero cells produced by Bacteroides gingivalis and Bacteroides asaccharolyticus', *Antonie van Leeuwenhoek 1982 48:4*, 48(4), pp. 315–325. doi: 10.1007/BF00418285.

Tran, P. T. *et al.* (2017) 'Anaerobic Baffled Reactor in Treatment of Natural Rubber Processing Wastewater: Reactor Performance and Analysis of Microbial Community'. doi: 10.2965/jwet.17-038.

Tukanghan, W. *et al.* (2021) 'Symbiotic Bacteroides and Clostridium-rich methanogenic consortium enhanced biogas production of high-solid anaerobic digestion systems', *Bioresource Technology Reports*, 14, p. 100685. doi: 10.1016/J.BITEB.2021.100685.

Udaondo, Z., Duque, E. and Ramos, J. L. (2017) 'The pangenome of the

genus Clostridium', *Environmental Microbiology*, 19(7), pp. 2588–2603. doi: 10.1111/1462-2920.13732.

Ueki, A. *et al.* (2006) 'Paludibacter propionicigenes gen. nov., sp. nov., a novel strictly anaerobic, Gram-negative, propionate-producing bacterium isolated from plant residue in irrigated rice-field soil in Japan', *International journal of systematic and evolutionary microbiology*, 56(1), pp. 39–44.

Ungkulpasvich, U. *et al.* (2021) 'Symbiotic chitin degradation by a novel anaerobic thermophilic bacterium Hydrogenispora sp. UUS1-1 and the bacterium Tepidanaerobacter sp. GT38', *Enzyme and Microbial Technology*, 144, p. 109740. doi: 10.1016/j.enzmictec.2020.109740.

Vandecasteele, S. J. *et al.* (2001) 'Quantification of expression of Staphylococcus epidermidis housekeeping genes with Taqman quantitative PCR during in vitro growth and under different conditions', *Journal of Bacteriology*, 183(24), pp. 7094–7101. doi: 10.1128/JB.183.24.7094-7101.2001.

Vandesompele, J. *et al.* (2002) 'Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes.', *Genome biology*, 3(7), p. research0034.1. doi: 10.1186/gb-2002-3-7-research0034.

Vanegas, C. H. and Bartlett, J. (2013) 'Green energy from marine algae: biogas production and composition from the anaerobic digestion of Irish seaweed species', *Environmental technology*, 34(15), pp. 2277–2283.

Větrovský, T. and Baldrian, P. (2013) 'The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses', *PLoS ONE*, 8(2), p. e57923. doi: 10.1371/journal.pone.0057923.

Vignais, P. M., Billoud, B. and Meyer, J. (2001) 'Classification and phylogeny of hydrogenases', *FEMS Microbiology Reviews*, 25(4), pp. 455–501. doi: 10.1111/j.1574-6976.2001.tb00587.x.

Vrbová, V. and Ciahotný, K. (2017) ‘Upgrading Biogas to Biomethane Using Membrane Separation’, *Energy and Fuels*, 31(9), pp. 9393–9401. doi: 10.1021/ACS.ENERGYFUELS.7B00120.

De Vrieze, J. *et al.* (2018) ‘The active microbial community more accurately reflects the anaerobic digestion process: 16S rRNA (gene) sequencing as a predictive tool’, *Microbiome*, 6(1), p. 63.

Wachsman, J. T. and Barker, H. A. (1954) ‘Characterization of an orotic acid fermenting bacterium, Zymobacterium oroticum, nov. gen., nov. spec.’, *Journal of Bacteriology*, 68(4), pp. 400–404. doi: 10.1128/jb.68.4.400-404.1954.

Wagner, D. (2020) ‘Methanosarcina’, in *Bergey’s Manual of Systematics of Archaea and Bacteria*. John Wiley & Sons, Ltd, pp. 1–23. doi: 10.1002/9781118960608.gbm00519.pub2.

Wainaina, S. *et al.* (2019) ‘Bioengineering of anaerobic digestion for volatile fatty acids, hydrogen or methane production: A critical review’, *https://doi.org/10.1080/21655979.2019.1673937*, 10(1), pp. 437–458. doi: 10.1080/21655979.2019.1673937.

Wang, L. *et al.* (2019) ‘Metabolic engineering of Escherichia coli for the production of butyric acid at high titer and productivity’, *Biotechnology for Biofuels*, 12(1). doi: 10.1186/S13068-019-1408-9.

Wang, S. *et al.* (2011) ‘Formic acid triggers the “acid crash” of acetone-butanol-ethanol fermentation by Clostridium acetobutylicum’, *Applied and Environmental Microbiology*, 77(5), pp. 1674–1680. doi: 10.1128/AEM.01835-10.

Weiland, P. and Rozzi, A. (1991) ‘The start-up, operation and monitoring of high-rate anaerobic treatment systems: discusser’s report’, *Water Science and Technology*, 24(8), pp. 257–277.

Weisburg, W. G. *et al.* (1991) ‘16S ribosomal DNA amplification for

phylogenetic study.', *Journal of Bacteriology*, 173(2), p. 697. doi: 10.1128/JB.173.2.697-703.1991.

Whitman, W. (editor) (2015) 'Alkaliflexus', in *Bergey's Manual of Systematics of Archaea and Bacteria*. Wiley, pp. 1–3. doi: 10.1002/9781118960608.gbm00239.

Wickham, H. (2016) *Ggplot2: Elegrant graphics for data analysis*, *Springer-Verlag New York*.

Woese, C. R., Kandler, O. and Wheelis, M. L. (1990) 'Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya', *Proceedings of the National Academy of Sciences of the United States of America*, 87(12), pp. 4576–4579. doi: 10.1073/pnas.87.12.4576.

Wu, Z. and Yang, S. T. (2003) 'Extractive fermentation for butyric acid production from glucose by Clostridium tyrobutyricum', *Biotechnology and Bioengineering*, 82(1), pp. 93–102. doi: 10.1002/bit.10542.

Xia, Y. *et al.* (2014) 'Thermophilic microbial cellulose decomposition and methanogenesis pathways recharacterized by metatranscriptomic and metagenomic analysis', *Scientific Reports*, 4(1), pp. 1–9. doi: 10.1038/srep06708.

Yang, G. *et al.* (2017) 'Rapid Generation of Universal Synthetic Promoters for Controlled Gene Expression in Both Gas-Fermenting and Saccharolytic Clostridium Species', *ACS Synthetic Biology*, 6(9), pp. 1672–1678. doi: 10.1021/acssynbio.7b00155.

Yarza, P. *et al.* (2014) 'Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences', *Nature Reviews Microbiology*, 12(9), pp. 635–645. doi: 10.1038/nrmicro3330.

Ye, J. *et al.* (2012) 'Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction.', *BMC bioinformatics*, 13, p. 134. doi: 10.1186/1471-2105-13-134.

Zaremba-Niedzwiedzka, K. *et al.* (2017) ‘Asgard archaea illuminate the origin of eukaryotic cellular complexity’, *Nature*, 541(7637), pp. 353–358. doi: 10.1038/nature21031.

Zhang, J. *et al.* (2011) ‘Performance and spatial community succession of an anaerobic baffled reactor treating acetone-butanol-ethanol fermentation wastewater’, *Bioresource Technology*, 102(16), pp. 7407–7414. doi: 10.1016/j.biortech.2011.05.035.

Zhang, J. *et al.* (2017) ‘Three-stage anaerobic co-digestion of food waste and horse manure’, *Scientific Reports*, 7(1), pp. 1–10. doi: 10.1038/s41598-017-01408-w.

Zhang, M. and Shi, E. (2021) ‘The spatial community succession of an anaerobic baffled reactor with the variation of hydraulic retention time’, *Environmental Technology and Innovation*, 22, p. 101497. doi: 10.1016/j.eti.2021.101497.

Zhang, X. *et al.* (2009) ‘Application of (R)-3-Hydroxyalkanoate Methyl Esters Derived from Microbial Polyhydroxyalkanoates as Novel Biofuels’, *Biomacromolecules*, 10(4), pp. 707–711. doi: 10.1021/BM801424E.

Zhang, Y., Grosse-Honebrink, A. and Minton, N. P. (2015) ‘A universal mariner transposon system for forward genetic studies in the genus clostridium’, *PLoS ONE*, 10(4). doi: 10.1371/JOURNAL.PONE.0122411.

Zhang, Z. *et al.* (2020) ‘Rumen Methanogenesis, Rumen Fermentation, and Microbial Community Response to Nitroethane, 2-Nitroethanol, and 2-Nitro-1-Propanol: An In Vitro Study’, *Animals : an Open Access Journal from MDPI*, 10(3). doi: 10.3390/ANI10030479.

Zhao, P. *et al.* (2010) ‘Hydrogen production characteristics from dark fermentation of maltose by an isolated strain F.P 01’, in *International Journal of Hydrogen Energy*. Pergamon, pp. 7189–7193. doi: 10.1016/j.ijhydene.2009.12.188.

Zhao, X. *et al.* (2012) ‘The effects of metal ions and l-cysteine on hydA gene expression and hydrogen production by Clostridium beijerinckii RZF-1108’, *International Journal of Hydrogen Energy*, 37(18), pp. 13711–13717. doi: 10.1016/J.IJHYDENE.2012.02.144.

Zhou, J. J. *et al.* (2020) ‘Metabolism, morphology and transcriptome analysis of oscillatory behavior of Clostridium butyricum during long-term continuous fermentation for 1,3-propanediol production’, *Biotechnology for Biofuels*, 13(1), pp. 1–18. doi: 10.1186/s13068-020-01831-8.

Zhou, L. Y. *et al.* (2019) ‘Maribellus luteus gen. Nov., sp. nov., a marine bacterium in the family prolixibacteraceae isolated from coastal seawater’, *International Journal of Systematic and Evolutionary Microbiology*, 69(8), pp. 2388–2394. doi: 10.1099/ijsem.0.003495.

Zhou, Z., Meng, Q. and Yu, Z. (2011) ‘Effects of methanogenic inhibitors on methane production and abundances of methanogens and cellulolytic bacteria in in vitro ruminal cultures’, *Applied and Environmental Microbiology*, 77(8), pp. 2634–2639. doi: 10.1128/AEM.02779-10.

Ziganshin, A. M. *et al.* (2013) ‘Microbial community structure and dynamics during anaerobic digestion of various agricultural waste materials’, *Applied microbiology and biotechnology*, 97(11), pp. 5161–5174.

Ziganshin, A. M. *et al.* (2019) ‘Spatial separation of metabolic stages in a tube anaerobic baffled reactor: reactor performance and microbial community dynamics’, *Applied Microbiology and Biotechnology*, 103(9), pp. 3915–3929. doi: 10.1007/s00253-019-09767-2.

Zigová, J. *et al.* (1999) ‘Butyric acid production by Clostridium butyricum with integrated extraction and pertraction’, *Process Biochemistry*, 34(8), pp. 835–843. doi: 10.1016/S0032-9592(99)00007-2.

Zimmermann, J., Kaleta, C. and Waschina, S. (2021) ‘gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate

metabolic models', *Genome Biology 2021 22:1*, 22(1), pp. 1–35. doi: 10.1186/S13059-021-02295-1.

Zwick, F., Lale, R. and Valla, S. (2013) 'Regulation of the expression level of transcription factor XylS reveals new functional insight into its induction mechanism at the Pm promoter', *BMC Microbiology*, 13(1), pp. 1–12. doi: 10.1186/1471-2180-13-262.

Appendix

Butyrate kinase Clustal Omega multiple alignment

* = fully conserved residue
: = strongly similar properties
. = weakly similar properties

```
CLUSTAL O(1.2.4) multiple sequence alignment

Clostridium_acetobutylicum_DSM_1731_buk2         MKFKLLTINPGSTSTKIAVFENEKEILSETLRHSSKELEAYKNIYEQFEFRKDTILKVLK    60
Clostridium_saccharoperbutylacetonicum_buk2      MSYKLLIINPGSTSTKISVYDGEREIFGETLRHSAEEIGKFEHIFDQQNFRTEVILKTLE    60
Clostridium_beijerinckii_buk2                    MSHKLLIINPGSTSTKLAVYEDEKEIFEETLRHSSEEIGKFKRVAEQQSFRTSIILKILQ    60
Clostridium_saccharobutylicum_buk2               MKYKLLIINPGSTSTKISVYHDEKQIFQETLRHSSDEIGNFKHVVDQQNFRTEAILKILE    60
Clostridium_acetobutylicum_DSM_1731_buk1         -MYRLLIINPGSTSTKIGIYDDEKEIFEKTLRHSAEEIEKYNTIFDQFQFRKNVILDALK    59
Clostridium_carboxidivorans_P7_buk1              MSYKLLILNPGSTSTKIGVYDGENEILEETLRHSSEEIEKYATIYDQFEFRKEVILKVLK    60
Clostridium_drakei_buk1                          MSYKLLILNPGSTSTKIGVYDGENQILEETLRHSSEEIEKYATVYDQFEFRKEVILKVLK    60
Clostridium_scatologenes_buk1                    MSYKLLILNPGSTSTKIGVYDGENQILEETLRHSSEEIEKYATVYDQFEFRKEVILKVLK    60
Clostridium_perfringens_ATCC_13124_buk1          MAYKLLIINPGSTSTKIGVYEGEKEILEETLRHSAEEILKYDTIFDQLDFRKEVILKVLK    60
Clostridium_butyricum_buk1                       MAYKLLIINPGSTSTKIGVYEDEKELFEETLRHTNEELKQFDAIFDQFQFRKDVILKVLS    60
Clostridium_butyricum_ Cbu-c15-buk               MAYKLLIINPGSTSTKIGVYEDEKELFEETLRHTNEELKQFDAIFDQFQFRKDVILKVLS    60
Clostridium_saccharobutylicum_buk1               MSYKLLIINPGSTSTKIGVYENEKELFEETLRHTNEEIKRYDTIYDQFGFRKDVILKVLS    60
Clostridium_beijerinckii_buk1                    MSYKLLIINPGSTSTKIGVYENEKELFEETLRHTNEEIKRYDTIYDQFEFRKELILNILK    60
Clostridium_saccharoperbutylacetonicum_buk1      MSYKLLIINPGSTSTKIGVYENEKELFEETLRHTNEEIKRYDTIYDQFGFRKEVILNVLK    60
Clostridium_pluributyricum_Cpl-c42-buk           MSYKLLIINPGSTSTKIGVYEDEKELFEETLRHTNEEIKRYDTIYDQFGFRKEVILNVLK    60
Clostridium_carboxidivorans_P7_buk3              -MYKILAINPGSTSTKIAIYDDTEELFKTTIEHSSEEVKKYENIADQYSMRYEAIMKFLK    59
Clostridium_drakei_buk3                          -MYKILAINPGSTSTKIAIYDDTEELFKTSIEHSSEEVKKYENIADQYAMRYEAIMKFLK    59
Clostridium_scatologenes_buk3                    -MYKILAINPGSTSTKIAIYADTEELFKTSIEHSSEEVKKYENIADQYAMRYEAIMKFLK    59
Clostridium_fermenticellae_buk1                  MGYKILVINPGSTSTKIALYENEKELFNVTLEHPIEEIEKYNKIADQFDMRKKVVLSFLK    60
Clostridium_pluributyricum_Cpl-c3-buk            MNYKVLAINPGSTSTKIALYENEKEVFCKTLDHPAEEIEKYNKIADQFDMRKEIVLSFLK    60
Clostridium_fermenticellae_buk2                  MSYKILAINPGSTSTKIALYKDEKEVFCKTLDHLVEEIERYDNVADQFDMRKDIVLSFLK    60
Clostridium_carboxidivorans_P7_buk2              MSYKILAINPGSTSTKIALYEDEKEIFCKTLEHPVEQIEKYENVADQFDMRKEVVLSFLK    60
Clostridium_drakei_buk2                          MSYKILAINPGSTSTKIALYEDEKEIFCKTLEHPVEQIEKYENTADQFDMRKEVVLSFLK    60
Clostridium_scatologenes_buk2                    MSYKILAINPGSTSTKIALYEDEKEIFCKTLEHPVEQIEKYENTADQFDMRKEVVLSFLK    60
                                                   .::* :********:.:: . .:::  :: *  .::  :    :*  :* . ::. *.
```

```
Clostridium_acetobutylicum_DSM_1731_buk2          DKNFNIQNIDAVVGRGGLLKPIVGGTYKVNEKMLKDLKAGVQGEHASNLGGIIANSIAEA   120
Clostridium_saccharoperbutylacetonicum_buk2       DNKIDIQNLDAIVGRGGLLKPILSGTYRVNENMLKDLKESIRGEHASNLGAIIANEIANS   120
Clostridium_beijerinckii_buk2                     DNKIDIKEMDAVVGRGGLLKPIISGTYNVNDKMLKDLKENVQGEHASNLGAIIANEIASS   120
Clostridium_saccharobutylicum_buk2                SKKIDIQEIDAVVGRGGLLKPIISGTYNINDKMLEDLKDNIRGEHASNLGAIIASKIATD   120
Clostridium_acetobutylicum_DSM_1731_buk1          EANIEVSSLNAVVGRGGLLKPIVSGTYAVNQKMLEDLKVGVQGQHASNLGGIIANEIAKE   119
Clostridium_carboxidivorans_P7_buk1               EKNFDINTLDGVVGRGGLLKPIESGTYKVNDAMLEDLKVGVQGQHASNLGGIIANEIGKS   120
Clostridium_drakei_buk1                           EKNFDINTLDGVVGRGGLLKPIESGTYKVNDAMLEDLKVGVQGEHASNLGGIIANEIGKS   120
Clostridium_scatologenes_buk1                     EKNFDINTLDGVVGRGGLLKPIASGTYKVNDAMVEDLKAGVRGEHASNLGGIIASEIGKS   120
Clostridium_perfringens_ATCC_13124_buk1           EKGIDINELDAVVGRGGMLKPIEGGTYEVNEAMVEDLKIGVQGPHASNLGGILSNEIAKE   120
Clostridium_butyricum_buk1                        EKNFDIKTLSAVVGRGGMLKPVEGGTYAVNDAMVEDLKVGVQGPHASNLGGILARSIADE   120
Clostridium_butyricum_ Cbu-c15-buk                EKNFDIKTLSAVVGRGGMLKPVEGGTYAVNDAMVEDLKVGVQGPHASNLGGILARSIADE   120
Clostridium_saccharobutylicum_buk1                EKNFDLKTLSAVVGRGGMLKPVEGGTYAVNDAMVEDLKVGVQGPHASNLGGILARSIGDE   120
Clostridium_beijerinckii_buk1                     EKNFDIKTLSAIVGRGGMLKPVEGGTYSVNDAMVEDLKVGVQGPHASNLGGIIAKSIGDE   120
Clostridium_saccharoperbutylacetonicum_buk1       EKNFDIKTLSAVVGRGGMLKPVEGGTYAVNDAMVEDLKAGVQGPHASNLGGIIAKSIGDE   120
Clostridium_pluributyricum_ Cpl-c42-buk           EKNFDITTLSAIVGRGGMLKPVEGGTYAVNDAMVEDLKVGVQGPHASNLGGIIAKSIGDE   120
Clostridium_carboxidivorans_P7_buk3               EVDFDVKALSAVVGRGGILPPVKSGAYRVNDSMVERLAKRPVVEHASNLGAIISYAIAKP   119
Clostridium_drakei_buk3                           EVNFDVKDLSAVVGRGGILPPVKSGAYRVNDSMVERLSKRPVVEHASNLGAIISYAIAKP   119
Clostridium_scatologenes_buk3                     EVNFDVKDLSAVVGRGGILPPVKSGAYRVNDSMVERLSKRPVVEHASNLGAIISYAIAKP   119
Clostridium_fermenticellae_buk1                   EKGIEINKLDAVVGRGGMLPKVQSGAYKVNKLMINRLRNNPVVEHASNLGALVAYEIANS   120
Clostridium_pluributyricum_ Cpl-c3-buk            ENGYEISELSAVVGRGGMVPSVESGAYEVNEIMVDRLKNNPVIEHASNLGALIAYEIASS   120
Clostridium_fermenticellae_buk2                   NNGYEVKELDAVVGRGGMVPKVKSGAYKVNETMVDRLKNNPIVGHASNLGALIAYEIANS   120
Clostridium_carboxidivorans_P7_buk2               QNGYEVKELAAVVGRGGMVPKVKSGAYKVNETMVDRLKNNPVVEHASNLGALIAYEIANS   120
Clostridium_drakei_buk2                           QNGYEVKELAAVVGRGGMVPKVKSGAYKVNETMVDRLKNNPVVEHASNLGALIAYEIASS   120
Clostridium_scatologenes_buk2                     QNGYEVKELAAVVGRGGMVPKVKSGAYKVNETMVDRLKNNPVVEHASNLGALIAYEIANS   120
                                                  .   ::  : .:*****::  : .*:* :*. *:. *      ******.:::  *.
```

```
Clostridium_acetobutylicum_DSM_1731_buk2          FGVSAYIVDPVVVDEMEDIARFSGIPELPRKSIFHALNQKAVAKRYAKESERDYEDLNII   180
Clostridium_saccharoperbutylacetonicum_buk2       IGKTAFIVDPVVVDEMDDIARFSGIPELPRKSIFHALNQKAVAKKYARESGLNYEDINVI   180
Clostridium_beijerinckii_buk2                     IGKPAFIVDPVVVDEMEELAKVSGMPEIPRKSIFHALNQKAVAKRYAKESSKKYEDINII   180
Clostridium_saccharobutylicum_buk2                IQKPAFIVDPVVVDEMEEIARISGISEIRRKSIFHALNQKAVAKRYAKERSRNYEDVNVI   180
Clostridium_acetobutylicum_DSM_1731_buk1          INVPAYIVDPVVVDELDEVSRISGMADIPRKSIFHALNQKAVARRYAKEVGKKYEDLNLI   179
Clostridium_carboxidivorans_P7_buk1               INKPAFIVDPVVVDELDEAARISGMPEIERISIFHALNQKAVAKRYAKENNKKYDELNLV   180
Clostridium_drakei_buk1                           INKPAFIVDPVVVDELDEAARISGMPEIERKSKFHALNQKAVAKRYAKENSKKYDELNVI   180
Clostridium_scatologenes_buk1                     INKPSFIVDPVVVDELDEAARISGMPEIERTSIFHALNQKAVAKRYAKENGKKYDELNVI   180
Clostridium_perfringens_ATCC_13124_buk1           IGKRAFIVDPVVVDEMEDVARLSGVPELPRKSKFHALNQKAVAKRYAKEHNTSYEDVNLI   180
Clostridium_butyricum_buk1                        IGVPSFIVDPVVTDELADVARLSGTPDIPRKSKFHALNQKAVAKRYGKESGKGYENLNLV   180
Clostridium_butyricum_ Cbu-c15-buk                IGVPSFIVDPVVTDELADVARLSGTPDIPRKSKFHALNQKAVAKRYGKESGKGYENLNLV   180
Clostridium_saccharobutylicum_buk1                LKIPSFIVDPVVTDELADVARLSGTPDIPRKSKFHALNQKAVAKRYGKESGKGYENLNLV   180
Clostridium_beijerinckii_buk1                     LNIPSFIVDPVVTDELADVARLSGVPELPRKSKFHALNQKAVAKRYGKESKKGYENLNLV   180
Clostridium_saccharoperbutylacetonicum_buk1       LNIPSFIVDPVVTDELAEVARLSGVPELPRKSKFHALNQKAVAKRYGKESGQGYENLNLV   180
Clostridium_pluributyricum_ Cpl-c42-buk           LNVPSFIVDPVVTDELADVARPSGVPELPRKSKFHALNQKAVAKRYGKESEKGYENLNLV   180
Clostridium_carboxidivorans_P7_buk3               LNIPAFIYDSVAVDEFEDIARISGLADIKRESFIHALNMRAAAIKTAKKLGKPYEQCNLV   179
Clostridium_drakei_buk3                           LNIPAFIYDSVAVDEFEDIARISGLADIKRESFIHALNMRAAAIKTAKKLGKPYEQCNLV   179
Clostridium_scatologenes_buk3                     LNIPAFIYDSVAVDEFEDIARISGLADIKRESFIHALNMRAAAIKTAKKLGKPYEQCNLV   179
Clostridium_fermenticellae_buk1                   IGVLAYIYDSVRVDEMNDIARISGMSDIPRESTSHALNSRAMAIKAAKKYGKKYKDMNFI   180
Clostridium_pluributyricum_ Cpl-c3-buk            IGVSAYIYDSVRVDELRDIARISGMPDIPRASTSHALNSRAMAMKVAAKYGKKYSDMNFV   180
Clostridium_fermenticellae_buk2                   IGVSAYIYDSVRVDELKDIARISGMPDIPRTSTSHALNTRAMAMKVAKKYGKKYSDMNFI   180
Clostridium_carboxidivorans_P7_buk2               IGVSAYIYDSVRVDELEDIARISGMPDIPRTSTSHALNTRAMAMKVAKKYGKKYSDMNFI   180
Clostridium_drakei_buk2                           IGVPAYIYDSVRVDELEDIARVSGMPDIPRTSTSHALNTRAMAMKVAKKYGKKYSDMNFI   180
Clostridium_scatologenes_buk2                     IGVPAYIYDSVRVDELEDIARISGMPDIPRTSTSHALNTRAMAMKVAKKYGKKYSDMNFI   180
                                                  :   ::* * * .**: : :: **  :: * *  **** :* * : . :    *.: *.:
```

```
Clostridium_acetobutylicum_DSM_1731_buk2          VAHMGGGVSVGAHKNGKIIDVNNALDGEGAFSPERSGNLPSGDLVRLCFSGKYTEDEILK   240
Clostridium_saccharoperbutylacetonicum_buk2       VAHMGGGVSVGAHKRGRIIDVNNALDGEGAFSPERSGGVPAGDLARMCFSGKYTLEEILK   240
Clostridium_beijerinckii_buk2                     VVHMGGGVSVGAHKKGKIVDVNNCLDGDGPFSPERSGGVPVGDLVRLCFSGKYTLGEVLK   240
Clostridium_saccharobutylicum_buk2                VAHMGGGVSVGAHRNGKIIDVNNALDGDGAFSPERSGGVPAGDLVRMCFSGKHTMQEVLK   240
Clostridium_acetobutylicum_DSM_1731_buk1          VVHMGGGTSVGTHKDGRVIEVNNTLDGEGPFSPERSGGVPIGDLVRLCFSNKYTYEEVMK   239
Clostridium_carboxidivorans_P7_buk1               VTHMGGGVTVGAHKKGRVVDVANGLDGDGPFSPERTGGLPVGGLIKLCYSGKYTLEEMKK   240
Clostridium_drakei_buk1                           VTHMGGGVSVGAHKEGKVVEVSNALDGDGPFSPERTGGLPVGSLIELCYSGKYTIEEMKK   240
Clostridium_scatologenes_buk1                     VTHMGGGVSVGAHKKGRVVEVSNALDGDGPFSPERTGGLPVGSLIKLCYSGKYTIEEMKK   240
Clostridium_perfringens_ATCC_13124_buk1           VVHMGGGVSVGAHRKGRVIDVNNALDGDGPFSPERAGGVPSGELLEMCFSGKYSKEEVYK   240
Clostridium_butyricum_buk1                        VVHMGGGVSVGAHNHGKVVDVNNALDGDGPFSPERAGSVPAGDLIKMCFSGKYSESEVYS   240
Clostridium_butyricum_ Cbu-c15-buk                VVHMGGGVSVGAHNHGKVVDVNNALDGDGPFSPERAGSVPAGDLIKMCFSGKYSESEVYS   240
Clostridium_saccharobutylicum_buk1                VVHMGGGVSVGAHNHGKVVDVNNALDGDGPFSPERAGSVPAGDLIKMCFSGKYSESEVYT   240
Clostridium_beijerinckii_buk1                     VVHMGGGVSVGAHNHGKVVDVNNALDGDGPFSPERAGSVPIGDLIKMCFSGKYSESEVYV   240
Clostridium_saccharoperbutylacetonicum_buk1       VVHMGGGVSVGAHNHGKVVDVNNALDGDGPFSPERAGSVPIGDLIKMCFSGKYSEAEVYG   240
Clostridium_pluributyricum_ Cpl-c42-buk           VVHMGGGVSVGAHNHGKVVDVNNALDGDGPFSPERAGSVPVGDLIKMCFSGQYSESEVYG   240
Clostridium_carboxidivorans_P7_buk3               VAHLGGGISLTVHKGGKMIDAV--TDEEGPFSPERSGRVPCKRLIEMCYKN--DERTMKK   235
Clostridium_drakei_buk3                           VAHLGGGISLTVHKGGKMIDAV--TDEEGPFSPERSGRVPCKRLIEMCYKN--DERTMKK   235
Clostridium_scatologenes_buk3                     VAHLGGGISLTVHKGGKMIDAV--TDEEGPFSPERSGRVPCKRLIEMCYKN--DERTMKK   235
Clostridium_fermenticellae_buk1                   VAHLGGGISLSVHEKGKMVDVM--PDDDGPFSPERAGKLQCKALIDLCYSGKFDKETMQK   238
Clostridium_pluributyricum_ Cpl-c3-buk            VAHLGGGISVSVHEKGRMVDIL--ADDEGPFSPERAGKVPCNPLIELCYSGKYDKNTMKK   238
Clostridium_fermenticellae_buk2                   VAHLGGGISVNVHRKGQMVDIM--ADDEGPFSPERAGRVPCNALIDLCYSGKYDKKTMKK   238
Clostridium_carboxidivorans_P7_buk2               VAHLGGGISVNVHRKGQMVDIM--ADDEGPFSPERAGKVPCNALIDLCYSGKFDKKTTKK   238
Clostridium_drakei_buk2                           VAHLGGGISVNVHRKGQMVDIM--ADDEGPFSPERAGKVPCNALIDLCYSGKFDKKTMKK   238
Clostridium_scatologenes_buk2                     VAHLGGGISVNVHRKGKMVDIM--ADDEGPFSPERAGKVPCNALIDLCYSGKFDKKTTKK   238
                                                  *.*:*** :: .*. *::::     * :* *****:* :    *  :*:..
```

```
Clostridium_acetobutylicum_DSM_1731_buk2          KITGKGGFVAYHGTNNALDVQNAALEGDYDAKMTYNAMGYQVAKDIGSAAAVLDGKVDCI   300
Clostridium_saccharoperbutylacetonicum_buk2       KITGKGGFVAYLNTNDARIVENAALDGDAKAKLVHDAMGYQVAKDIGAAAVVLNGKVDAI   300
Clostridium_beijerinckii_buk2                     KITGKGGFVDYLDTNDARIVEKATLKGDPKAKLIHDAMGYQVAKSIGAAAAVLNGKVDAI   300
Clostridium_saccharobutylicum_buk2                KITGKGGFVAYLDTNDARIVENAALKGDRKAKLIHDAMGYQVAKEIGAAATVLNGKVDAI   300
Clostridium_acetobutylicum_DSM_1731_buk1          KINGKGGVVSYLNTIDFKAVVDKALEGDKKCALIYEAFTFQVAKEIGKCSTVLKGNVDAI   299
Clostridium_carboxidivorans_P7_buk1               KISGKGGIVAYLNTNDFREVEQKAESGDKKAKLVFDAFILQVGKEIGKCAAVLHGKVDAL   300
Clostridium_drakei_buk1                           KISGKGGVVAYLNTNDFREVEQKSQ-TDKKAKLVFDAFILQVGKEIGKCAAVLHGKVDAI   299
Clostridium_scatologenes_buk1                     KISGKGGIVAYLNTNDFREVEQKAQ-TDKKAKLVFDAFILQVGKEIGKCAAVLHGKVDAI   299
Clostridium_perfringens_ATCC_13124_buk1           KLVGKGGFVAYANTNDARDLIKLSQEGDEKGSLIFNAFIYQIAKEIGSMAVVLDGEVDAI   300
Clostridium_butyricum_buk1                        KIVGKGGFVGYLNTNDVKGTIDKMEAGDKECENIYKAFLYQITKAIGEMSAALNGKVDQI   300
Clostridium_butyricum_ Cbu-c15-buk                KIVGKGGFVGYLNTNDVKGTIDKMEAGDKECENIYKAFLYQITKAIGEMSAALNGKVDQI   300
Clostridium_saccharobutylicum_buk1                KVVGKGGFVGYLNTNDVKGTIDKMEAGDKECEKIYKAFLYQITKAIGEMSAALNGKVDQI   300
Clostridium_beijerinckii_buk1                     KTVGKGGFVGYLDTNDVKGVIDKMEAGDKECEVVYKAFVYQIAKSIGEMSVVLEGKIDQI   300
Clostridium_saccharoperbutylacetonicum_buk1       KVVGKGGFVGYLNTNDVKGVIDKMEAGDKECEGIYKAFVYQIAKAIGEMSVVLEGKVDQI   300
Clostridium_pluributyricum_ Cpl-c42-buk           KVVGKGGFVGYLNTNDVKGVIDSMEAGDKKCEEIYKAFVYQISKAIGEMSVVLEGKVDQI   300
Clostridium_carboxidivorans_P7_buk3               KIRGDGGLISYLGTNSALDVEKRIENGDAEAKLVYEAMAYQIAKAIGELATVVKGKVDAV   295
Clostridium_drakei_buk3                           KIRGDGGLISYLGTNSALDVEKRIENGDAEAKLVYEAMAYQIAKSIGELATVVKGKVDAV   295
Clostridium_scatologenes_buk3                     KIRGDGGLISYLGTNSALDVEKRIENGDAEAKLVYEAMAYQIAKSIGELATVVKGKVDAV   295
Clostridium_fermenticellae_buk1                   KLRGNGGLKDYLGTKDAREVEKMIKNGNKKARLVYEAMAYQIAKGIGELATVVEGKVDLI   298
Clostridium_pluributyricum_ Cpl-c3-buk            KLRGNGGLKAYLNTVDAREVEKMIENEDENAKLIYEAMAYQVAKGIGELATVVDGKVDAI   298
Clostridium_fermenticellae_buk2                   KLRGNGGLKAYLNTVDAREVEKMIKNGDERAKLVYEAMAYQIAKGIGELSTVVEGKVDVI   298
Clostridium_carboxidivorans_P7_buk2               KLRGNGGLKAYLNTVDAREVERMIESGDEKAKLVYEAMAYQVAKGIGELATVVEGKVDAI   298
Clostridium_drakei_buk2                           KLRGNGGLKAYLNTVDAREVEKMIENGDEKAKLVYEAMAYQVAKGIGELATVVEGKVDAI   298
Clostridium_scatologenes_buk2                     KLRGNGGLKAYLNTVDAREVEKMIENGDEKAKLVYEAMAYQVAKGIGELATVVEGKVDAI   298
                                                  *  *.**.  * .* .          :      ..*:  *: * **  :..:.*::* :
```

```
Clostridium_acetobutylicum_DSM_1731_buk2          ILTGGIAYNKLMTDFIAKKVSFIAPITIYPGEDEMLALAEGTLRVLSGQEEAKKYK--     356
Clostridium_saccharoperbutylacetonicum_buk2       ILTGGMAYGKPIVNFISEKVRFIAPIVIYPGEDEMLALAEGVTRVLDGEEEVKDYK--     356
Clostridium_beijerinckii_buk2                     ILTGGMAYGKPIIDVISKRVSFIAPIVVYPGEDEMLALAQGVIRTLNGEEAVGEYI--     356
Clostridium_saccharobutylicum_buk2                ILTGGMAYGKPMVDFISKKVGFIAPIVVYPGEDEMLALAQGALRVLNGEEKVREYE--     356
Clostridium_acetobutylicum_DSM_1731_buk1          ILTGGIAYNEHVCNAIEDRVKFIAPVVRYGGEDELLALAEGGLRVLRGEEKAKEYK--     355
Clostridium_carboxidivorans_P7_buk1               ILTGGIAYSKTVTAAIKDMVEFIAPVVVYPGEDELLALAQGGLRVLGGEEQAKEYK--     356
Clostridium_drakei_buk1                           ILTGGIAYSKVVTAELKDMIEFIAPVVVYPGEDELLALAQGGLRVLSGEEQAKEYK--     355
Clostridium_scatologenes_buk1                     ILTGGIAYSKAVTAAIKDMIEFISPVVVYPGEDELLALAQGGLRVLSGEEQAKEYK--     355
Clostridium_perfringens_ATCC_13124_buk1           VLTGGIAYSDYVTNAINKKVKWIAPMVVYGGEDELLALAQGAIRVLDGVEEAKIYK--     356
Clostridium_butyricum_buk1                        VLTGGIAYSPTLVPDLKANVEWIAPVTVYPGEDELLALAQGAIRVLDGEEKAKVY---     355
Clostridium_butyricum_Cbu-c15-buk                 VLTGGIAYSPTLVPDLKANVEWIAPVTVYPGEDELLALAQGAIRVLDGEEKAKVY---     355
Clostridium_saccharobutylicum_buk1                ILTGGIAYSPTLVPDLKSNVEWIAPVTVYPGEDELLALAQGAIRVLDGEEEAKVY---     355
Clostridium_beijerinckii_buk1                     IFTGGIAYSPTLVPDLITKVEWIAPVTVYPGEDELLALAQGAIRVLDGEEEAKIY---     355
Clostridium_saccharoperbutylacetonicum_buk1       IFTGGIAYSPTLVPDLKEKVEWIAPVTVYPGEDELLALAQGAIRVLDGEEKAKVY---     355
Clostridium_pluributyricum_ Cpl-c42-buk           IFTGGIAYSPTLVPDLKAKVEWIAPVTVYPGEDELLALAQGAIRVLDGEEQARIY---     355
Clostridium_carboxidivorans_P7_buk3               VITGGIAYSKMMTGWIKERVEFIAPVEILPGENELESLALGTLRVIKGEEKAHEYDLD     353
Clostridium_drakei_buk3                           VITGGIAYSKMMTEWVKERVEFIAPVEILPGENELESLALGALRVIKGEEKAHEYDLD     353
Clostridium_scatologenes_buk3                     VITGGIAYSKMMTEWVKERVEFIAPVEILPGENELESLALGALRVIKGEEKAHEYDLD     353
Clostridium_fermenticellae_buk1                   IITGGIAYSNMLTDWIKKRVEFIAPVEIMPGENEMESLALGTLRVLKGEENAREYKE-     355
Clostridium_pluributyricum_ Cpl-c3-buk            VITGGIAYSKMITSWIKKRVEFIGTVEIMPGENEMESLSLGVLRVLKGEETAKEYLER     356
Clostridium_fermenticellae_buk2                   VITGGIAYSDMITRWIKKRVEFIAPVEIMPGENEMESLALGILRVLKGEEEAKEYIEE     356
Clostridium_carboxidivorans_P7_buk2               VITGGIAYSDMITNWIKKRVEFIAPVEIMPGENEMESLALGTLRVLKGEEEAREYVE-     355
Clostridium_drakei_buk2                           VITGGIAYSDMITNWIKKRVQFIAPVEIMPGENEMESLALGTLRVLKGEEEAREYVE-     355
Clostridium_scatologenes_buk2                     VITGGIAYSDMITNWIKKRVQFIAPVEIMPGENEMESLALGTLRVLKGEEEAREYVE-     355
                                                  ::***:**.  :   :   : :*. :    **:*: :*: *  *.: * * .  *
```

Gapseq optimal medium predictions

“maxFlux”: Defines the maximum uptake rate of the compound in mmol/gDW/hr

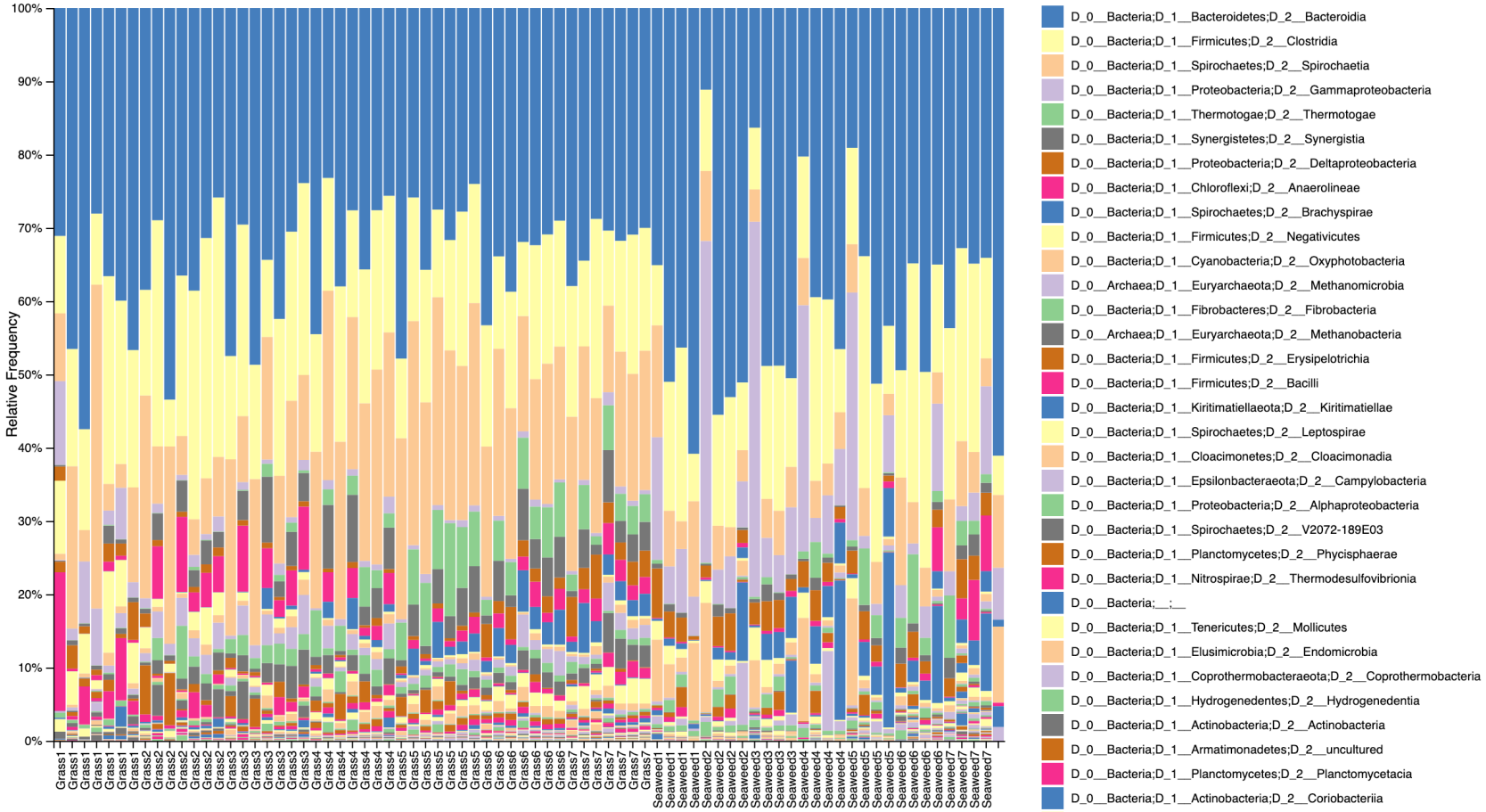
Appendix Table 1. Gapseq output, predicting an optimal medium for *C. butyricum*

compounds	name	maxFlux
cpd00001	H2O	100
cpd00971	Sodium	10
cpd00205	K+	10
cpd00099	Cl-	10
cpd10515	Fe2+	10
cpd10516	Fe3+	0.1
cpd00009	Phosphate	10
cpd00149	Cobalt	10
cpd00254	Mg	10
cpd00244	Nickel	1
cpd00048	Sulfate	10
cpd00239	H2S	1
cpd00034	Zn2+	10
cpd00058	Cu2+	10
cpd00030	Mn2+	10
cpd00063	Ca2+	10
cpd00528	N2	25
cpd00027	D-Glucose	5
cpd00082	D-Fructose	5
cpd00208	Lactose	2.5
cpd90003	starch (n=27, 3xalpha1-6, 23xalpha1-4)	0.18
cpd00382	Melitose	1.67
cpd00076	Sucrose	2.5
cpd00122	N-Acetyl-D-glucosamine	5
cpd00054	L-Serine	0.1
cpd00066	L-Phenylalanine	0.1
cpd00069	L-Tyrosine	0.1
cpd00644	Pantothenic acid	0.025
cpd00305	Thiamin	0.025
cpd00159	L-Lactate	15
cpd00029	Acetate	20
cpd00067	H+	35

Appendix Table 2. Gapseq output, predicting an optimal medium for *C. pluributyricum*

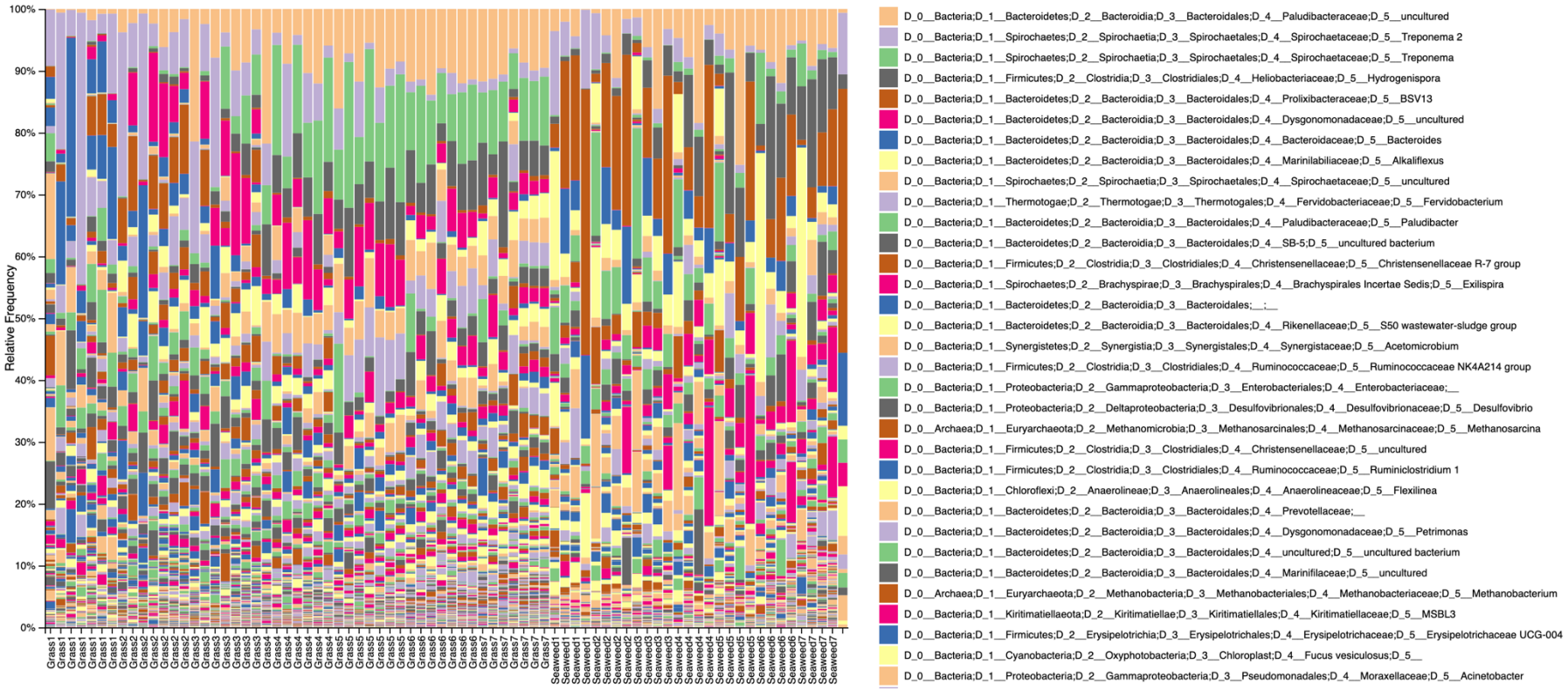
compounds	name	maxFlux
cpd00001	H2O	100
cpd00971	Sodium	10
cpd00205	K+	10
cpd00099	Cl-	10
cpd10515	Fe2+	10
cpd10516	Fe3+	0.1
cpd00009	Phosphate	10
cpd00149	Cobalt	10
cpd00254	Mg	10
cpd00244	Nickel	1
cpd00048	Sulfate	10
cpd00239	H2S	1
cpd00034	Zn2+	10
cpd00058	Cu2+	10
cpd00030	Mn2+	10
cpd00063	Ca2+	10
cpd00528	N2	25
cpd00027	D-Glucose	5
cpd00082	D-Fructose	5
cpd00208	Lactose	2.5
cpd90003	starch (n=27, 3xalpha1-6, 23xalpha1-4)	0.18
cpd00076	Sucrose	2.5
cpd00122	N-Acetyl-D-glucosamine	5
cpd00054	L-Serine	0.1
cpd00069	L-Tyrosine	0.1
cpd00644	Pantothenic acid	0.025
cpd00305	Thiamin	0.025
cpd00159	L-Lactate	15
cpd00029	Acetate	20
cpd00036	Succinate	10
cpd01020	Gallate	10
cpd00313	Quercetin	10
cpd00067	H+	75

QIIME 2 taxonomic bar plots to show variation between individual digestate samples (Class)



Appendix Figure 1. Taxonomic bar plots showing the relative average abundance of sequences and their association by class. Showing Top 33 taxa by abundance, grouped by feedstock and chamber (e.g. chamber 1 containing grass feedstock = Grass1), sorted by descending relative abundance.

QIIME 2 taxonomic bar plots to show variation between individual digestate samples (Species)



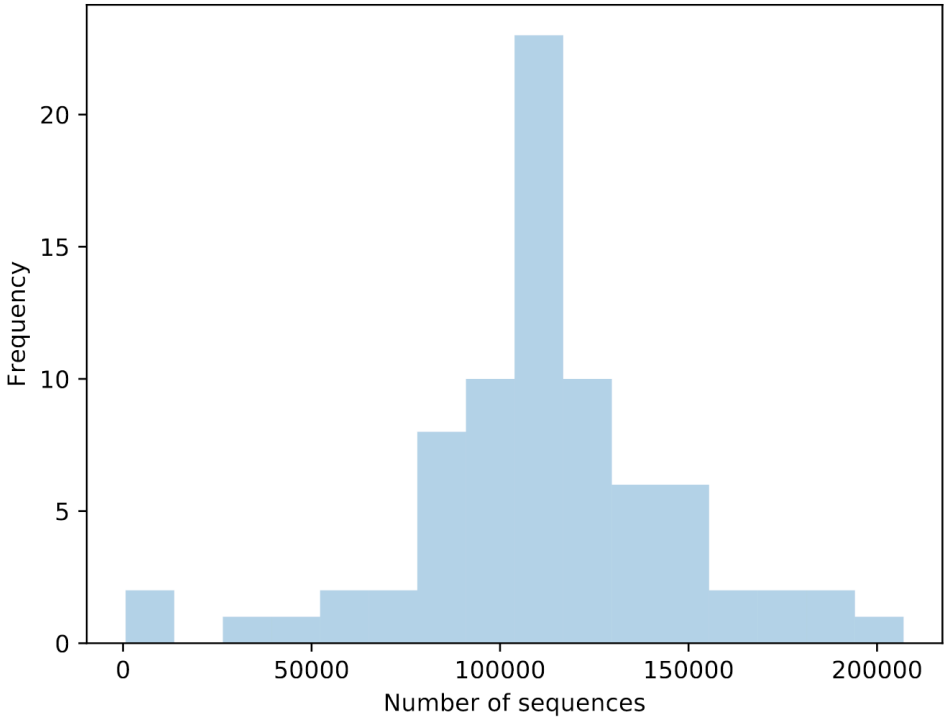
Appendix Figure 2. Taxonomic bar plots showing the relative average abundance of sequences and their association by species. Showing Top 33 taxa by abundance, grouped by feedstock and chamber (e.g. chamber 1 containing grass feedstock = Grass1), sorted by descending abundance

Data on the number of sequencing reads per digestate sample

Appendix Table 3. Number of reads per sample before and after DADA2 denoising

sampling date	feedstock	chamber	input	filtered	denoised	merged	non-chimeric
01062017	seaweed	1	64357	56540	56540	55647	55645
01062017	seaweed	2	50182	43446	43446	42343	42343
01062017	seaweed	3	54017	47347	47347	46185	46185
01062017	seaweed	4	2535	2179	2179	2023	2023
01062017	seaweed	5	101339	89391	89391	87464	87289
01062017	seaweed	6	119004	102548	102548	96720	89894
01062017	seaweed	7	110592	95063	95063	91000	82456
01112017	grass	1	84566	68893	68893	67155	66599
01112017	grass	2	111448	91361	91361	87650	86190
01112017	grass	3	66206	59178	59178	58000	57452
01112017	grass	4	106569	90756	90756	86410	79626
01112017	grass	5	146162	126383	126383	121743	111888
01112017	grass	6	143832	102155	102155	98311	89446
01112017	grass	7	109929	93481	93481	90127	88157
07042017	seaweed	1	102757	88240	88240	85557	82742
07042017	seaweed	2	115071	102629	102629	100367	98323
07042017	seaweed	3	97691	80883	80883	78353	76570
07042017	seaweed	4	112629	87743	87743	85100	84198
07042017	seaweed	5	100417	79429	79429	76965	75158
07042017	seaweed	6	116934	98409	98409	96381	94439
07042017	seaweed	7	99875	78245	78245	75754	75415
07122017	grass	1	132197	106213	106213	104995	104970
07122017	grass	2	104359	83339	83339	81689	80685
07122017	grass	3	110712	88721	88721	87275	87159
07122017	grass	4	107032	78188	78188	77097	76999
07122017	grass	5	107293	79367	79367	78363	78338
07122017	grass	6	83064	60372	60372	58854	58794
07122017	grass	7	34012	28342	28342	27006	26463
08112017	grass	1	139092	112967	112967	110100	106746
08112017	grass	2	172361	137439	137439	134601	131873
08112017	grass	3	134740	111792	111792	108025	104831
08112017	grass	4	159007	120849	120849	119212	119194
08112017	grass	5	117923	90255	90255	88219	88013
08112017	grass	6	110256	82960	82960	80439	77512
08112017	grass	7	126207	77519	77519	75863	74494
13122017	grass	1	147022	107163	107163	104924	99504
13122017	grass	2	132093	93015	93015	91165	91134

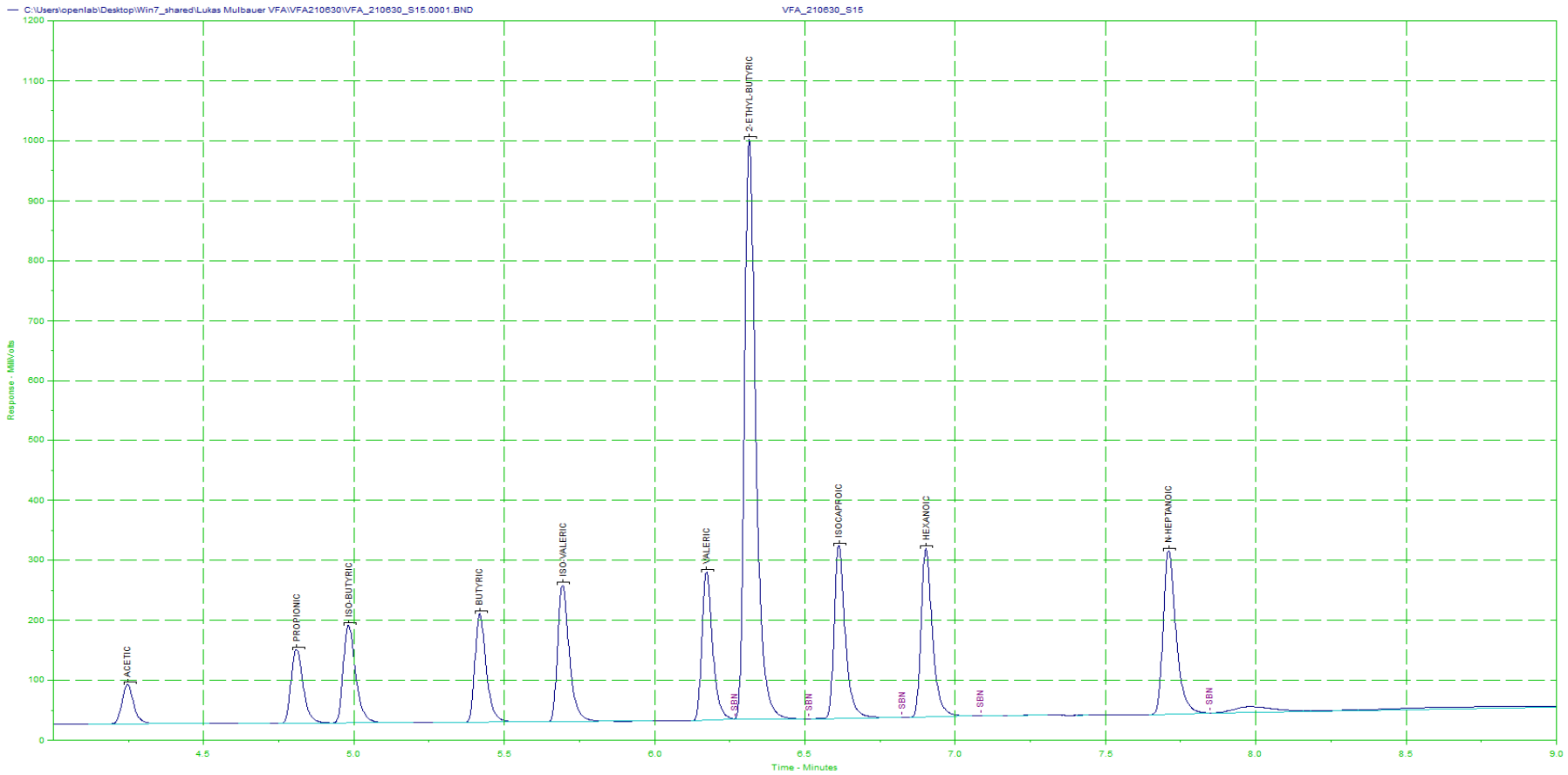
13122017	grass	3	115034	79178	79178	77405	72468
13122017	grass	4	106911	81625	81625	80526	80436
13122017	grass	5	115573	82764	82764	81800	81751
13122017	grass	6	95819	76098	76098	74494	74416
13122017	grass	7	105345	81337	81337	79442	79360
15032017	seaweed	1	126910	99695	99695	96007	87397
15032017	seaweed	2	152325	130364	130364	126097	121095
15032017	seaweed	3	115463	89705	89705	87508	84416
15032017	seaweed	4	106689	92621	92621	90614	89344
15032017	seaweed	5	81680	67869	67869	65917	61835
15032017	seaweed	6	107064	87483	87483	84671	83977
15032017	seaweed	7	100983	75401	75401	72745	66316
15112017	grass	1	117291	80044	80044	78109	78046
15112017	grass	2	125368	79800	79800	77936	77936
15112017	grass	3	115280	80919	80919	78718	76657
15112017	grass	4	108231	79013	79013	77509	77496
15112017	grass	5	92513	69512	69512	66495	65024
15112017	grass	6	87420	52752	52752	49747	40085
15112017	grass	7	85675	73886	73886	71144	70493
16022017	seaweed	1	107083	97364	97364	95277	90066
16022017	seaweed	2	100675	84930	84930	83623	82229
16022017	seaweed	3	97906	78540	78540	76669	72250
16022017	seaweed	4	67288	60563	60563	59242	57102
16022017	seaweed	5	80087	70337	70337	68389	66179
16022017	seaweed	6	111151	93617	93617	90705	89106
16022017	seaweed	7	121234	97014	97014	93898	92828
22112017	grass	1	86348	69448	69448	67918	67835
22112017	grass	2	207029	157574	157574	154766	154652
22112017	grass	3	148697	107358	107358	105251	105251
22112017	grass	4	181587	130131	130131	127703	127703
22112017	grass	5	175363	145482	145482	143668	143668
22112017	grass	6	158189	133242	133242	131247	131241
22112017	grass	7	187888	143955	143955	141178	141178
29112017	grass	1	121354	101976	101976	99364	97777
29112017	grass	2	137006	104100	104100	102258	101519
29112017	grass	3	151740	103275	103275	99825	86782
29112017	grass	4	106262	77517	77517	76055	76052
29112017	grass	5	122771	86337	86337	84227	80497
29112017	grass	6	89346	72223	72223	69795	68738
29112017	grass	7	136321	113182	113182	111172	111172
Negative			681	589	589	512	512



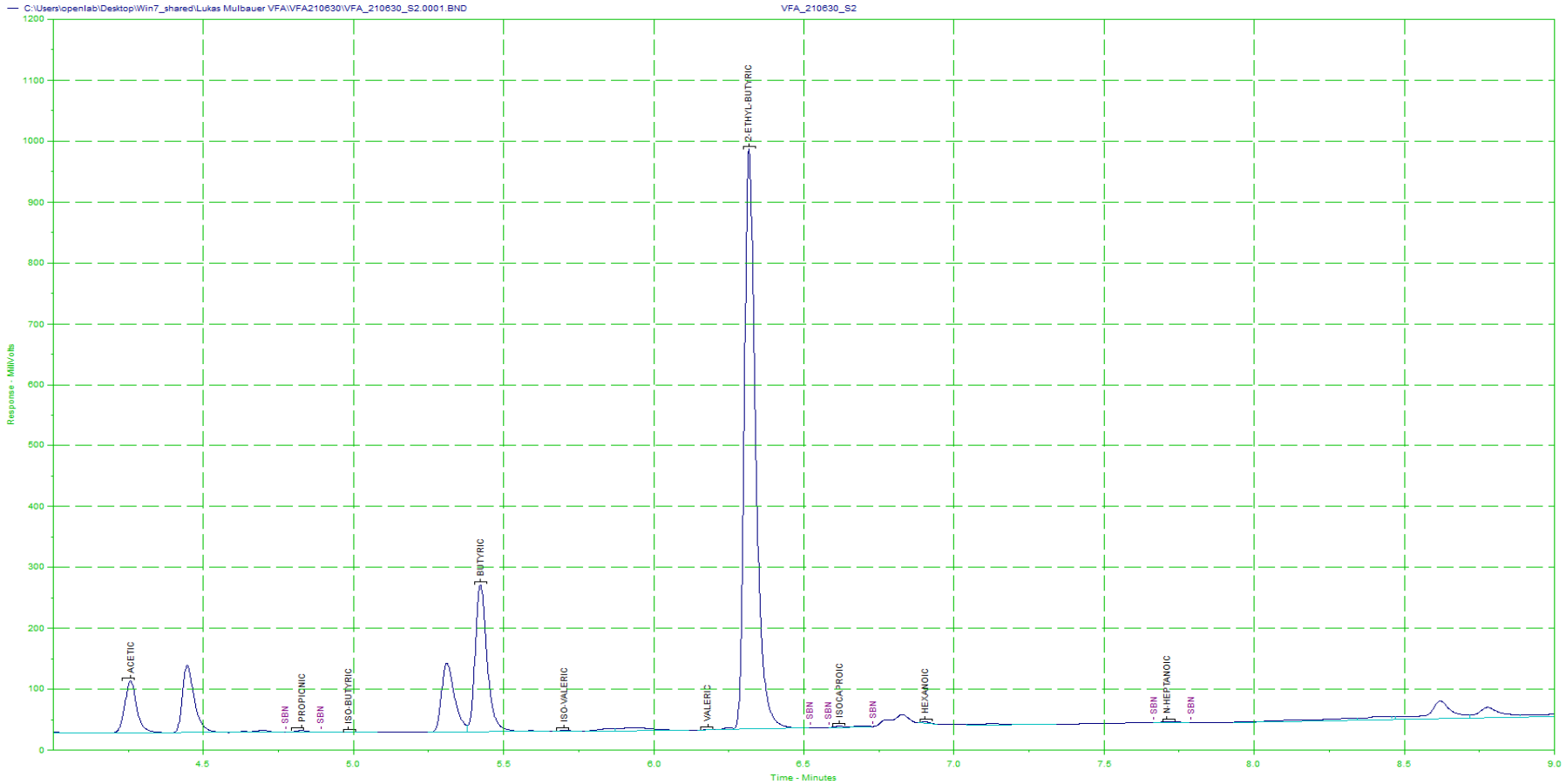
Appendix Figure 3. Histogram showing the number of digestate samples (Frequency) grouped by the number of input sequences generated through Illumina sequencing

Appendix Table 4. Sequence count statistics for digestate samples	
Minimum:	681
Median:	110424.0
Mean:	111423.871795
Maximum:	207029
Total:	8691062

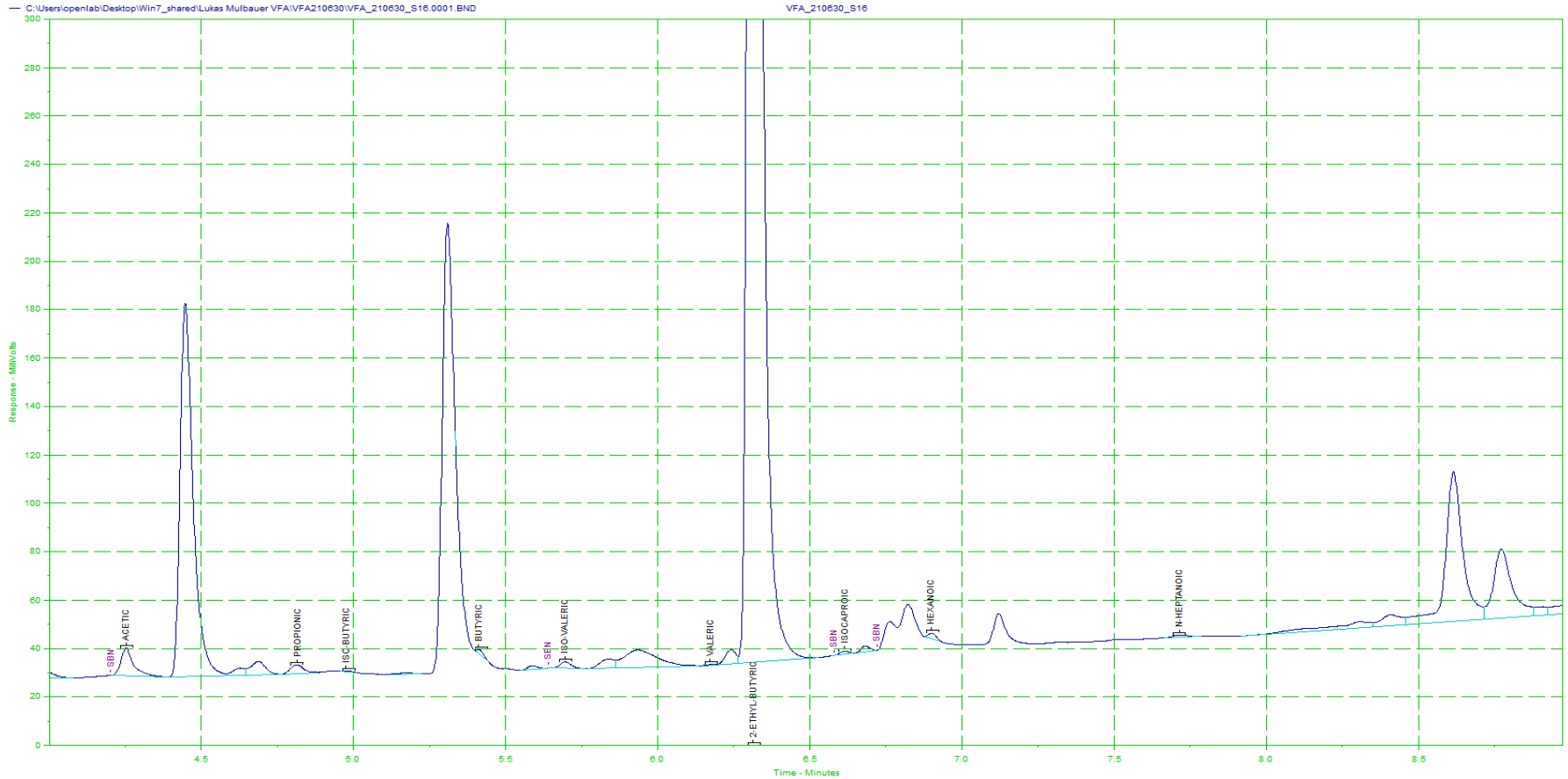
Examples of GC-FID traces



Appendix Figure 4: GC-FID trace of standard VFA mix as described in Table 11.



Appendix Figure 5: GC-FID trace of supernatant from a culture of *C. butyricum* pMTL8315B (empty plasmid control).



Appendix Figure 6: GC-FID trace of non-inoculated PYG medium (note the difference in Y-axis scale compared to Appendix Fig. 4 and 5).